

|        |        |
|--------|--------|
| 申<br>报 | 系列：教师  |
|        | 专业：生物学 |
|        | 职称：副教授 |

## 业绩成果材料

（申报人的业绩成果材料包括论文、科研项目、获奖以及其他成果等）

单 位（二级单位） 生命科学学院

姓 名 张明月

材料核对人：

单位盖章：

核对时间：

华南农业大学制

## 目 录

### 一、科研项目

1. 主持：关于国家自然科学基金青年项目的资助项目计划书及有关佐证材料 ..... 1
2. 主持：关于深圳市科技局科技重大专项子项目的立项合同书 ..... 13
3. 主持：关于江苏省自然科学基金的立项合同及有关佐证材料 ..... 28
4. 主持：关于中国博士后基金一等资助的立项合同及有关佐证材料 ..... 44
5. 主参：广东省农业农村厅乡村振兴战略专项种业振兴行动项目合同书 ..... 46

### 二、论文、著作等

1. 检索证明 ..... 67
2. 以第一作者发表本专业论文情况
  - 2.1. 发表 Nature Communications 期刊论文 “Genome-wide association studies provide insights into the genetic determination of fruit traits of pear” ..... 71
  - 2.2. 发表 Horticultural Plant Journal 期刊论文 “Genomic selection of eight fruit traits in pear” ..... 81
  - 2.3. 发表 Horticulture Research 期刊论文 “Pear genetics: Recent advances, new prospects, and a roadmap for the future” ..... 90
  - 2.4. 发表 BMC Biology 期刊论文 “Rearrangement and domestication as drivers of Rosaceae mitogenome



|   |     |
|---|-----|
| plasticity” .....   | 116 |
| 2.5. 发表 BMC Genomics 期刊论文 “Comparison of multiple algorithms to reliably detect structural variants in pears” .....   | 135 |
| 2.6. 发表 BMC Genomics 期刊论文 “Contrasting genetic variation and positive selection followed the divergence of NBS-encoding genes in Asian and European pears” .....              | 150 |
| 2.7. 发表 Plant Journal 期刊论文 “Natural variations in the PbCPK28 promoter regulate sugar content through interaction with PbTST4 and PbVHA-A1 in pear” ..                        | 166 |
| 2.8. 发表 Tree Genetics & Genomics 期刊论文 “Genome-wide genetic diversity and IBD analysis reveals historic dissemination routes of pear in China” .....                           | 187 |
| 2.9. 发表 Gene 期刊论文 “Identification of key genes related to seedlessness by genome-wide detection of structural variation and transcriptome analysis in 'Shijiwuhe' pear” ..... | 199 |
| 3. 以通讯作者发表本专业论文情况   |     |
| 3.1. 发表 BMC Plant Biology 期刊论文 “The pear genomics database (PGDB): a comprehensive multi-omics research platform for Pyrus spp.” .....  | 210 |

### **三、科研成果**

#### 1. 科技奖励证书

|                                |     |
|--------------------------------|-----|
| 1.1. 神农中华农业科技奖优秀创新团队奖（排名 14） . | 218 |
|--------------------------------|-----|

### **四、其他业绩**

#### 1. 个人荣誉

|  |     |
|--|-----|
| 1.1. 中国植物生理与植物分子生物学学会第二届“优秀女青年”奖 ..... | 219 |
|--|-----|

**【佐证材料切记与目录页所列页码对应, 不要用图片格式的材料  
进行打印。】**



|        |                    |
|--------|--------------------|
| 项目批准号  | 31901978           |
| 申请代码   | C150102            |
| 归口管理部门 |                    |
| 依托单位代码 | 21009508A1649-1247 |



3 1901978 1005 307

# 国家自然科学基金委员会 资助项目计划书

资助类别：青年科学基金项目

亚类说明：

附注说明：

项目名称：基于梨全基因组关联分析筛选调控果实发育期的BRI1基因功能及其作用机制研究

直接费用：24万元 执行年限：2020.01-2022.12

负责人：张明月

通讯地址：江苏省 南京市 南京农业大学 卫岗一号 园艺学院

邮政编码：210095 电 话：(025)84396485

电子邮件：myzhang@njau.edu.cn

依托单位：南京农业大学

联系人：赵珩 电 话：025-84399807

填表日期：2019年08月27日

国家自然科学基金委员会制



## 国家自然科学基金委员会资助项目计划书填报说明

- 一、项目负责人收到《关于国家自然科学基金资助项目批准及有关事项的通知》（以下简称《批准通知》）后，请认真阅读本填报说明，参照国家自然科学基金相关项目管理办法及《国家自然科学基金资助项目资金管理办法》（请查阅国家自然科学基金委员会官方网站首页“政策法规”栏目），按《批准通知》的要求认真填写和提交《国家自然科学基金委员会资助项目计划书》（以下简称《计划书》）。
- 二、填写《计划书》时要求科学严谨、实事求是、表述清晰、准确。《计划书》经国家自然科学基金委员会相关项目管理部门审核批准后，将作为项目研究计划执行和检查、验收的依据。
- 三、《计划书》各部分填写要求如下：
  - （一）简表：由系统自动生成。
  - （二）摘要及关键词：各类获资助项目都必须填写中、英文摘要及关键词。
  - （三）项目组主要成员：计划书中列出姓名的项目组主要成员由系统自动生成，与申请书原成员保持一致，不可随意调整。如果批准通知中“项目评审意见及修改意见表”中“对研究方案的修改意见”栏目有调整项目组成员相关要求的，待项目开始执行后，按照项目成员变更程序另行办理。
  - （四）资金预算表：根据批准资助的直接费用，按照《国家自然科学基金项目预算表编制说明》填报资金预算表和预算说明书。国家重大科研仪器研制项目、重大项目还应按照预算评审后批复的直接费用各科目金额填报资金预算表、预算说明书及相应的预算明细表。
  - （五）正文：
    1. 面上项目、青年科学基金项目、地区科学基金项目：如果《批准通知》中没有修改要求的，只需选择“研究内容和研究目标按照申请书执行”即可；如果《批准通知》中“项目评审意见及修改意见表”中“对研究方案的修改意见”栏目明确要求调整研究期限和研究内容等的，须选择“根据研究方案修改意见更改”并填报相关修改内容。
    2. 重点项目、重点国际（地区）合作研究项目、重大项目、国家重大科研仪器研制项目：须选择“根据研究方案修改意见更改”，根据《批准通知》的要求填写研究（研制）内容，不得自行降低、更改研究目标（或仪器研制的技术性能与主要技术指标以及验收技术指标）或缩减研究（研制）内容。此外，还要突出以下几点：
      - （1）研究的难点和在实施过程中可能遇到的问题（或仪器研制风险），拟采用的研究（研制）方案和技术路线；
      - （2）项目主要参与者分工，合作研究单位之间的关系与分工，重大项目还需说明课题之间的关联；
      - （3）详细的年度研究（研制）计划。



3. 国家杰出青年科学基金、优秀青年科学基金和海外及港澳学者合作研究基金项目：须选择“根据研究方案修改意见更改”，按下列提纲撰写：
  - (1) 研究方向；
  - (2) 结合国内外研究现状，说明研究工作的学术思想和科学意义（限两个页面）；
  - (3) 研究内容、研究方案及预期目标（限两个页面）；
  - (4) 年度研究计划；
  - (5) 研究队伍的组成情况。
4. 国家自然科学基金基础科学中心项目：须选择“根据研究方案修改意见更改”，应当根据评审委员会和现场考察专家组的意见和建议，进一步完善并细化研究计划，作为评估和验收的依据。按下列提纲撰写：
  - (1) 五年拟开展的研究工作（包括主要研究方向、关键科学问题与研究内容）；
  - (2) 研究方案（包括骨干成员之间的分工及合作方式、学科交叉融合研究计划等）；
  - (3) 年度研究计划；
  - (4) 五年预期目标和可能取得的重大突破等；
  - (5) 研究队伍的组成情况。
5. 对于其他类型项目，参照面上项目的方式进行选择和填写。



简表

|        |           |                                       |     |   |      |                        |     |               |
|--------|-----------|---------------------------------------|-----|---|------|------------------------|-----|---------------|
| 申请者信息  | 姓 名       | 张明月                                   | 性 别 | 女 | 出生年月 | 1988年03月               | 民 族 | 汉族            |
|        | 学 位       | 博士                                    |     |   | 职称   | 讲师                     |     |               |
|        | 是否在站博士后   | 否                                     |     |   | 电子邮件 | myzhang@njau.edu.cn    |     |               |
|        | 电 话       | (025)84396485                         |     |   | 个人网页 |                        |     |               |
|        | 工 作 单 位   | 南京农业大学                                |     |   |      |                        |     |               |
|        | 所 在 院 系 所 | 园艺学院                                  |     |   |      |                        |     |               |
| 依托单位信息 | 名 称       | 南京农业大学                                |     |   |      |                        | 代码  | 21009508A1649 |
|        | 联 系 人     | 赵珩                                    |     |   | 电子邮件 | zhaoheng@njau.edu.cn   |     |               |
|        | 电 话       | 025-84399807                          |     |   | 网站地址 | http://www.njau.edu.cn |     |               |
| 合作单位信息 | 单 位 名 称   |                                       |     |   |      |                        |     |               |
|        |           |                                       |     |   |      |                        |     |               |
| 项目基本信息 | 项 目 名 称   | 基于梨全基因组关联分析筛选调控果实发育期的BRI1基因功能及其作用机制研究 |     |   |      |                        |     |               |
|        | 资 助 类 别   | 青年科学基金项目                              |     |   |      | 亚 类 说 明                |     |               |
|        | 附 注 说 明   |                                       |     |   |      |                        |     |               |
|        | 申 请 代 码   | C150102:果树种质资源与遗传育种学                  |     |   |      |                        |     |               |
|        | 基 地 类 别   |                                       |     |   |      |                        |     |               |
|        | 执 行 年 限   | 2020.01-2022.12                       |     |   |      |                        |     |               |
|        | 直 接 费 用   | 24万元                                  |     |   |      |                        |     |               |



## 项目摘要

### 中文摘要:

我国梨的栽培品种有近100个，以中、晚熟品种居多，集中在八九月份上市，导致价格普遍偏低，因而选育早熟的优质梨品种成为我国梨品种结构调整的重点和主要育种目标之一，但是关于梨的发育期调控机制研究相对较少，机制不明，无法有效的指导育种。本研究前期基于312份砂梨重测序数据，开展全基因组关联分析，筛选到与发育期显著相关联的候选基因BRI1（编码油菜素内酯受体），利用梨的瞬时转化、番茄稳定遗传转化等现代分子生物学方法验证BRI1调控果实发育期的功能，借助膜蛋白酵母双杂交、转录组基因共表达、Hi-C等技术预测并鉴定互作基因，明确BRI1调控机制；筛选早熟、晚熟品种中的差异遗传变异位点，发掘功能基因序列特征，开发分子检测标记，为筛选早熟梨品种提供高效的育种技术。研究有望拓展梨果实发育期调控的基础理论研究，为梨果实早熟性的遗传改良提供基因资源和分子选择标记，具有理论和实践意义。

### Abstract:

There are near 100 pear cultivars in our country, and most of them are medium maturing and late maturing cultivars. Concentrated pear mature period in August and September induced the generally low price. Therefore, breeding for early maturing and good quality pears becomes the keynote of fruit structural adjustment and one of the main breeding goals. However, the mechanism to regulate fruit development period is comparatively seldom studied and is not clearly known in pears. It is difficult to guide breeding. This project is based on the re-sequencing data of natural population from 312 sand pears, and the genome wide association study was carried out. One significant candidate gene BRI1 encoding brassinosteroids receptor was predicted to associate with trait of fruit development period. Modern molecular biology methods such as transient expression in pears and transformation in tomatoes will be applied to verify the function of BRI1 for regulating fruit development period. The interaction mechanism of BRI1 will be clarified by Yeast two-hybrid system, gene co-expression by RNA-seq data and Hi-C technology. Furthermore, the genetic difference of BRI1 in early maturing and late maturing cultivars will be dug out by gene sequences. Development of molecular markers will help to improve the efficiency of early maturing pear breeding. This research will expend the theory basis of fruit development period, and provide gene resource for genetic modification and molecular markers in the trait of early maturing. It has significant meaning both in theory and in practice.

**关键词(用分号分开):** 梨; 关联分析; 果实发育期; BRI1; 基因功能

**Keywords(用分号分开):** Pear; Association study; Fruit development period; BRI1; Gene function



项目组主要成员

| 编号  | 姓名  | 出生年月    | 性别 | 职称 | 学位 | 单位名称   | 电话             | 证件号码               | 项目分工  | 每年工作<br>时间<br>(月) |
|-----|-----|---------|----|----|----|--------|----------------|--------------------|-------|-------------------|
| 1   | 张明月 | 1988.03 | 女  | 讲师 | 博士 | 南京农业大学 | (025) 84396485 | 210902198803041023 | 项目负责人 | 8                 |
| 总人数 |     | 高级      |    |    | 中级 | 初级     | 博士后            |                    | 博士生   | 硕士生               |
|     |     |         |    |    |    |        |                |                    |       |                   |





## 国家自然科学基金项目直接费用预算表（定额补助）

项目批准号：31901978

项目负责人：张明月

金额单位：万元

| 序号 | 科目名称                 | 金额      |
|----|----------------------|---------|
| 1  | 项目直接费用合计             | 24.0000 |
| 2  | 1、设备费                | 0.3000  |
| 3  | (1)设备购置费             | 0.0000  |
| 4  | (2)设备试制费             | 0.0000  |
| 5  | (3)设备升级改造与租赁费        | 0.3000  |
| 6  | 2、材料费                | 7.2000  |
| 7  | 3、测试化验加工费            | 5.8000  |
| 8  | 4、燃料动力费              | 0.8000  |
| 9  | 5、差旅/会议/国际合作与交流费     | 2.3000  |
| 10 | 6、出版/文献/信息传播/知识产权事务费 | 3.4000  |
| 11 | 7、劳务费                | 4.2000  |
| 12 | 8、专家咨询费              | 0.0000  |
| 13 | 9、其他支出               | 0.0000  |



## 预算说明书（定额补助）

（请按照《国家自然科学基金项目预算表编制说明》等的有关要求，对各项支出的主要用途和测算理由，以及合作研究外援资金、单价 $\geq 10$ 万元的设备费等内容进行必要说明。）

**项目直接费用——预算金额总计：24.00万元**

### 1、设备费——预算金额：0.30万元

主要包括设备改造与租赁费，是指现有仪器部分零部件的更新、保养和维修等费用，包含服务器计算与存储设备的升级更新、电泳槽的电极更换、植物培养箱等灯管更换、移液器的校准和维修等。按0.10万元/年计算， $0.10\text{万元/年} \times 3\text{年} = 0.30\text{万元}$ 。

### 2、材料费——预算金额：7.20万元

实验常规分子生化试剂：包括高保真DNA聚合酶、Taq酶、连接酶、限制性内切酶、琼脂糖、DNAMarker、聚丙烯酰胺、T载体、琼脂粉、酵母培养基、胰蛋白胨、酵母提取物、2,4-表油菜素内酯等，平均单价为250元/个（瓶），需消耗80个（瓶）；液态氮气，用于植物叶片、果实的速冻、保存，提取DNA、RNA样品时需要液氮速冻研磨，单价10元/升 $\times 1000$ 升。常规分子生化试剂购买统计： $250\text{元/瓶} \times 80\text{瓶} + 10\text{元/升} \times 1000\text{升} = 3.00\text{万元}$ 。

试剂盒：DNA提取试剂盒、质粒提取试剂盒、RNA提取试剂盒、DNA回收试剂盒、PCR产物回收纯化试剂盒、载体转化试剂盒等，平均单价400元/盒，共需消耗试剂盒25个；实时荧光定量PCR试剂盒每个5ml（500孔检测），需要约5个，每个0.10万元；高保真PCR试剂盒（500次），需要2盒，1000元每盒。试剂盒购买统计： $400\text{元/盒} \times 25\text{盒} + 1000\text{元/盒} \times 5\text{盒} + 1000\text{元/盒} \times 2\text{盒} = 1.70\text{万元}$ 。

耗材费用：具体包括移液器吸头，每包（1000根吸头）平均单价120元，PCR管每包（500个）单价200元，冻存管每包（500支）单价150元，离心管每包（500个）平均单价150元；按照塑料制品平均单价150元/包，共需消耗100包；三角瓶平均单价20元/个，螺旋盖玻璃试剂瓶平均单价26元/个，量筒、烧杯平均单价24元/个，培养皿30元/包（10副），按照玻璃器皿的平均单价为25元/件、包，需消耗200件；实验镊子、剪刀、测量尺、标签纸、自封袋、胶皮手套、一次性口罩、封口膜、塑料袋、蛭石、花盆、记号笔、计时器等其他各类耗材用具，0.5万元。耗材购买费用： $150\text{元/包} \times 100\text{包} + 25\text{元/个} \times 200\text{个} + 0.5\text{万元} = 2.50\text{万元}$ 。

**3、测试化验加工费——预算金额5.80万元**

部分样品的高通量测序数据补充：按照诺禾致源公司100元/Gb的价格，补充高通量测序20个品种，按照梨单个品种基因组大小为500Mb计算，使测序的覆盖度达到10fold计算，补充第二代测序费用：100元/Gb×20个×500Mb/1Gb×10fold=1.00万元。

引物合成及DNA测序：用于基因表达定量分析、基因克隆、瞬时表达、转模式植物验证等都需要合成引物，DNA测序：300对引物，每对引物平均60个碱基，每个碱基一元；按15元/测序反应，按2000次计，DNA测序费用：1元/bp×60bp/对引物×300对+15元/次×2000次=4.80万元。

**4、燃料动力费——预算金额0.80万元**

燃料动力费预算用于实验室大型仪器设备，包括大型服务器计算运行、超低温冰箱运行、组织培养箱、温室等的电力消耗费用，每年约需0.267万元，共计：0.267万元/年×3年=0.80万元。

**5、差旅/会议/国际合作与交流——预算金额2.30万元**

主要用于参加国内学术会议。参加国内与果树分子生物学相关的学术会议5人，会议注册费平均2000元/人，往返交通费平均1000元/人次，住宿费按每人每天300元计算，公杂费按每人每天100元计，每人出差平均以4天计算。总计：(2000元/人+1000元/人+300元×4天+100元×4天)×5人=2.30万元。

**6、出版/文献/信息传播/知识产权事务费——预算金额3.40万元**

预计发表SCI论文2篇，中文论文3篇，出版专利2个；SCI论文版面费等平均费用每篇0.8万元/篇计算；国内期刊按照0.2万元/篇计算；专利申请费平均每个专利0.6万元。总计：0.8万元×2篇+0.2万元/篇×3篇+0.6万元/个×2个=3.40万元。

**7、劳务费——预算金额4.20万元**

参加项目的博士研究生（2人）和硕士研究生（3人）的劳务生活补助，按照每人每年工作10个月计算，博士每人每月发放400元，硕士每人每月发放200元，劳务费计算：400元/人×2人×30月+200元/人×3人×30月=4.20万元。

**8、专家咨询费——无****9、其他支出——无**

项目负责人签字：

科研部门公章：

财务部门公章：



## 报告正文

研究内容和研究目标按照申请书执行。



## 国家自然科学基金资助项目签批审核表

|   |  |   |     |     |     |     |     |
|---|--|---|-----|-----|-----|-----|-----|
| <p>我接受国家自然科学基金的资助，将按照申请书、项目批准意见和计划书负责实施本项目（批准号：31901978），严格遵守国家自然科学基金委员会关于资助项目管理、财务等各项规定，切实保证研究工作时间，认真开展研究工作，按时报送有关材料，及时报告重大情况变动，对资助项目发表的论著和取得的研究成果按规定进行标注。</p> <p>项目负责人（签章）：<br/>年 月 日</p> |  | <p>我单位同意承担上述国家自然科学基金项目，将保证项目负责人及其研究队伍的稳定和研究项目实施所需的条件，严格遵守国家自然科学基金委员会有关资助项目管理、财务等各项规定，并督促实施。</p> <p>依托单位（公章）<br/>年 月 日</p> |     |     |     |     |     |
| 本<br>栏<br>目<br>由<br>基<br>金<br>委<br>填<br>写   | <p>科学处审查意见：</p>                            |   |     |     |     |     |     |
|   | <p>建议年度拨款计划（本栏目为自动生成，单位：万元）：</p>           |   |     |     |     |     |     |
|   | 年度   | 总额  | 第一年 | 第二年 | 第三年 | 第四年 | 第五年 |
|   | 金额   |   |     |     |     |     |     |
|   | <p>科学部审查意见：</p> <p>负责人（签章）：<br/>年 月 日</p>  |   |     |     |     |     |     |
| 本<br>栏<br>目<br>主<br>要<br>用<br>于<br>重<br>大<br>项<br>目<br>等  | <p>相关局室审核意见：</p> <p>负责人（签章）：<br/>年 月 日</p> |   |     |     |     |     |     |
|   | <p>委领导审批意见：</p> <p>委领导（签章）：<br/>年 月 日</p>  |   |     |     |     |     |     |

## 证明

南京农业大学园艺学院工作人员张明月同志承担科研情况：主持国家自然科学基金青年项目“基于梨全基因组关联分析筛选调控果实发育期的BRI1基因功能及其作用机制研究”（项目批准号：31901978），资助经费：24万元，经费本号：130-6S01180006，执行年限：2020.01-2022.12。





20240831194744001

深圳市科技计划项目合同书

项目编号: KCXFZ20240903093007010 计划年度: 2024  
项目类别: 科技重大专项(可持续发展专项) 计划类别: 创新创业专项  
下达文号: 深科创资(2024)49号 资金类别: 深圳市科技研发资金

## 深圳市科技计划项目合同书 (通用项目)

项目名称: 重202410045 优异玉米新品种培育技术研究与应用示范  
申报时间: 2024-08-31  
实施期间: 2024-12-10至2027-12-09  
管理单位(甲方): 深圳市科技创新局  
承担单位(乙方): 中山大学·深圳 2025.1.10 (盖章)  
通讯地址: 深圳市光明区公常路66号  
项目负责人: 杨芳 联系电话: 18040518081  
项目联系人: 杨芳 联系电话: 18040518081

深圳市科技创新局制  
二〇二四年二月



## 一、研究内容和任务

(一) 玉米矮化耐密抗倒伏优异基因挖掘研究。

(二) 玉米优异基因指标测定和功能验证研究。

(三) 玉米优异基因分子调控机理研究。

(四) 玉米优异种质材料的创制和新品种培育。

本项目旨在解决玉米耐密种植中株型优化和抗倒伏性的科学与产业问题，通过以下研究内容实现：

### 1. 玉米矮化耐密抗倒伏优异基因挖掘研究

(1) 利用突变体semi-dwarf plant1 (sdp1)与自交系构建分离群体，已克隆关键基因SDP1，并初步验证了其生物学功能。

(2) 利用密植模拟条件下产生的株高、茎秆等表型及相应的转录组数据，通过共表达网络分析筛选5个候选基因 (bHLH2a/2b, TCP2a/2b/2c)，并验证其在密植响应中的功能。

### 2. 玉米优异基因指标测定和功能验证

(1) 针对以上克隆的6个株型及密植相关功能基因 (目前已鉴定出SDP1, bHLH2a/2b, TCP2a/2b/2c)，利用或创制相关遗传材料 (功能缺失突变体或转基因)，全面验证其生物学功能。

(2) 利用以上相关遗传材料，进行不同密度的田间种植，调查其密植响应，包括茎秆粗细、株型、产量等主要农艺性状变化，茎秆细胞形态、叶绿素含量、净光合效率等细胞学及生理学指标变化，明确各基因响应密植的主要作用方式；基于各项指标测定，明确调控玉米矮化耐密抗倒伏的优异基因。

### 3. 玉米优异基因分子调控机理研究

(1) 应用RNA-seq和ChIP-seq技术鉴定6个转录因子的下游调控靶基因，筛选互作蛋白，构建分子调控网络。

(2) 利用CRISPR编辑技术获得突变体，明确生物学功能，鉴定各基因间遗传互作关系。

### 4. 玉米优异种质材料的创制和新品种培育

(1) 利用自然群体数据挖掘主效功能基因相关的优异单倍型，开发分子标记，改良自交系和玉米品种，提高耐密性。

(2) 基于功能基因作用机制，创制新种质，培育耐密新品种。

(3) 开展田间测试与示范，评估育种价值，建立示范工程。

### 5. 项目创新点

(1) 项目围绕玉米密植响应这一认知欠缺且复杂的生物学现象，同时开展多个功能基因的遗传与分子机制研究，将为密植响应这一复杂性状的分子育种提供重要理论。同时，研究还将创制的耐密新种质用于密植增产的育种效应评估。因此，本研究结果具有理论与应用创新。

(2) 本研究从茎秆响应密植中的精细表型与基因差异表达相结合，快速获得了多个 (5个) 影响株型、密植响应的功能基因，其中3个基因单独表现的功能并不强，很难从正向遗传学途径被鉴定到。因此，项目具有高效鉴定复杂性状功能基因的方法创新。

6. 涉及的知识产权包括：新克隆基因的专利申报和创制的耐密新种质及新品种的专利保护。



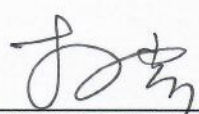
## 二、研究成果及主要指标（考核指标）

| 成果形式                          |  | 成果数量    |
|-------------------------------|--|---------|
| 实施期结束时项目所处阶段                  |  | 中试前期    |
| 项目所培养的人才数（博士/硕士/工程师/技术工人）     |  | 0/3/0/0 |
| 项目所新增的科研助理人数                  |  | 0       |
| 项目所产生的专利申请数（发明专利/实用新型/PCT）    |  | 8/0/0   |
| 项目所产生的专利授权数（发明专利/实用新型/PCT）    |  | 0/0/0   |
| 项目所产生的软件著作权数                  |  | 0       |
| 项目所发表的论文（论文总数/SCI检索数量/EI检索数量） |  | 3/0/0   |
| 项目所累计新增销售收入（万元）               |  | 1200.00 |
| 其他主要技术指标、经济指标、社会效益。（限1000字之内） | <p>(1) 克隆及鉴定调控玉米矮化、株型、抗倒伏及密植响应的关键基因4-5个，明确其生物学功能，并解析其分子机制；</p> <p>(2) 鉴定关键基因优异单倍型，创制矮化、耐密、抗倒伏新种质8-10份；</p> <p>(3) 利用株型优良新种质培育或改良骨干自交系及品种株型、抗倒伏性等，获得耐密增产新（改良）品种3-5个，产量提高&gt;5%；</p> <p>(4) 申请相关发明专利≥8个；</p> <p>(5) 发表高水平学术论文≥3篇；</p> <p>(6) 实现销售收入≥1200万元；</p> <p>(7) 在深圳市对口帮扶地区建立应用示范工程1项，示范基地面积200亩以上，增产提高5%以上；</p> <p>(8) 培养农业领域专业研究生3名。</p> |         |

## 三、项目计划进度和阶段目标

| 项目实施期间：2024-12-10至2027-12-09 |                       |   |
|------------------------------|-----------------------|---|
| 阶段                           | 起止时间                  | 研究内容与预期目标   |
| 第一阶段（半年）                     | 2024-12-10至2025-06-30 | <p>(1) 基于前期对突变体sdp1的克隆结果，利用等位突变体验证基因功能，同时创制SDP1带GFP标签的转基因，用于互补验证，同时开展分子机制的解析；</p> <p>(2) 基于密植响应候选基因（bHLH2a/2b, TCP2a/2b/2c）初步表型，调查其在密植条件下的株高、开花期、产量等相关农艺性状，全面鉴定其茎秆细胞学表型、叶绿素含量、净光合效率等细胞学及生理学指标变化，明确其在不同密度种植条件下的生物学功能；</p> <p>(3) 根据以上突变体表型，明确各基因的遗传效应；</p> <p>(4) 利用农艺性状和基因型数据完整的自然群体，鉴定显著影响株型、抗倒伏性等性状的关键基因的优异单倍型，设计分子标记，检测骨干自交系中该位点的使用情况，并对不含该优异单倍型的自交系进行改良；</p> <p>(5) 对功能明确、调控密植相关性状的重要基因及遗传材料进行应用专利保护申请。</p>   |
| 第二阶段（半年）                     | 2025-07-01至2025-12-31 | <p>(1) 重点开展功能主效基因的分子机制研究：基于以上候选基因均为转录因子，将完成与表型相对应的RNA-seq和ChIP-seq，鉴定潜在的下游调控靶基因；</p> <p>(2) 利用不同组织表达谱数据和密植下茎秆差异表达谱数据，以及同源基因功能信息等，进一步筛选可能响应密植的下游靶基因；</p> <p>(3) 同时通过IP-mass或者酵母cDNA文库筛选其各自的互作蛋白；结合不同组织表达谱数据、前期产生的密植下茎秆差异表达谱数据和同源基因功能信息等，进一步筛选功能可能相关的互作蛋白；</p> <p>(4) 将显著影响株型、抗倒伏性的突变位点设计分子标记，用于改良株型不够紧凑或不耐密植的骨干自交系；</p> <p>(5) 基于关键基因的遗传作用方式，利用Crispr编辑技术创制新种质，筛选出矮化、株型紧凑、抗倒伏性良好等耐密性状，开展耐密新种质鉴定用于培育耐密新品种；</p> <p>(6) 对功能明确、调控密植相关性状的重要基因及遗传材料进行应用专利保护申请。</p> |
| 第三阶段（半年）                     | 2026-01-01至2026-06-30 | <p>(1) 针对以上鉴定到的候选下游靶基因和互作蛋白，利用Crispr编辑技术或EMS突变体库获得相应突变体，开展表型鉴定；</p> <p>(2) 完善以上各基因调控下游靶基因和互作蛋白的分子调控关系验证，明确上下游的调控关系；</p> <p>(3) 将株高、抗倒伏性等株型优良的遗传材料（突变体等）种植于不同密度，测试其密植育种效应；</p> <p>(4) 继续完成骨干自交系的鉴定、改良、品种组配和改良效应评估，包括材料改良前后在高低不同种植密度下的农艺性状和产量性状的鉴定；</p> <p>(5) 对功能明确、调控密植相关性状的重要基因及遗传材料进行应用专利保护申请。</p>  |



| 阶段   | 起止时间  | 研究内容与预期目标  |
|--|---|--|
| 第四阶段（半年）   | 2026-07-01至2026-12-30   | (1) 利用下游基因和互作蛋白突变的不同等位变异突变体，调查其各种表型性状和密度响应关键部位的细胞学特征，明确各基因的生物学功能；<br>(2) 将以上有显著表型的突变体，种植于不同密度，全面调查和对比田间农艺性状、净光合效率等生理指标，明确响应密植的功能下游靶基因或互作蛋白；<br>(3) 通过遗传杂交，构建不同组合突变体，鉴定基因直接的遗传互作关系；其与以上6个株型和密植响应相关的功能基因间的遗传互作关系；<br>(4) 基于以上各基因之间的分子调控和遗传互作关系，构建这些功能基因共同调控玉米株型相关性状和密植响应过程中的分子调控网络；<br>(5) 继续完成改良骨干自交系和杂交种的密植育种效应评估；<br>(6) 对收获的合格玉米通过合作单位销售渠道进行销售；<br>(7) 对改良效应明显的密植优良遗传材料进行应用专利保护申请；<br>(8) 整理研究结果，准备发表论文。 |
| 第五阶段（半年）   | 2027-01-01至2027-06-30   | (1) 继续完成改良杂交种的密植育种效应评估；<br>(2) 利用改良效应显著的品种在深圳市对口扶贫地区，即项目科技应用合作单位开展田间测试与示范，并调查田间表现与测产；<br>(3) 对收获的合格玉米通过合作单位销售渠道进行销售；<br>(4) 整理研究结果，准备发表论文。   |
| 第六阶段（半年）   | 2027-07-01至2027-12-09   | (1) 继续完成改良杂交种的密植育种效应评估；<br>(2) 继续开展田间测试与示范，调查田间表现与测产；<br>(3) 对收获的合格玉米通过合作单位销售渠道进行销售；<br>(4) 整理项目相关数据，准备验收结题。   |
| 中期评估、阶段性考核指标（限1000字之内）   | 中期评估、阶段性考核指标如下：<br>(1) 克隆及鉴定调控玉米矮化、株型相关、抗倒伏及密植响应基因 3 个，明确其生物学功能；<br>(2) 鉴定关键基因优异单倍型或创制矮化、耐密或株型紧凑、抗倒伏新种质 5 份；<br>(3) 利用株型优良新种质培育或改良骨干自交系及品种株型、抗倒伏性等，获得耐密增产新（改良）品种 3 个，产量提高5%；<br>(4) 申请相关发明专利≥3 个；<br>(5) 发表高水平学术论文≥1 篇。 |  |
| 此表作为项目过程管理（包括但不限于年度报告、阶段性考核、随机抽查）的重要依据。甲方可以根据检查评估结果视情况对项目进行处理。<br><br>本人已知悉上述事项。<br><br>项目负责人（签字）：  |   |  |

说明：1.科技计划项目相关管理办法和申报指南等没有特别要求的，科技计划项目实施周期一般为1~5年。项目实施的起点时间应当在项目立项当年内。

2.需要进行中期评估、阶段性考核的，应明确考核阶段，填写与“二、研究成果及主要指标”相对应的可量化可考核指标。

20240831194744001



## 四、项目经费预算(单位: 万元)

| 财政资助总额 | 500.00   |          |           |        |
|--------|--|----------|-----------|--------|
| 年度拨款计划 | 首笔拨付资助金额的50%，中期评估通过后拨付资助金额的50%。                                      |          |           |        |
| 序号     | 经费支出类别(A)  | 财政资助额(B) | 项目自筹经费(C) | 总经费(D) |
| 01     | 合计（直接费用+间接费用）  | 500.00   | 50.00     | 550.00 |
| 02     | 一、直接费用（03+04+05）   | 436.48   | 50.00     | 486.48 |
| 03     | 设备费  | 0.00     | 0.00      | 0.00   |
|        | (1)购置设备费   | 0.00     | 0.00      | 0.00   |
|        | (2)试制设备费   | 0.00     | 0.00      | 0.00   |
|        | (3)设备改造与租赁费  | 0.00     | 0.00      | 0.00   |
| 04     | 业务费（包括材料费、测试化验加工费、燃料动力费、出版/文献/信息传播/知识产权事务费、差旅费、会议费、国际合作与交流费以及其他相关费用） | 336.48   | 40.00     | 376.48 |
| 05     | 人力资源费（包括人员费、劳务费、专家咨询费等）  | 100.00   | 10.00     | 110.00 |
| 06     | 二、间接费用（07+08+09）   | 63.52    | 0.00      | 63.52  |
| 07     | 单位水电气暖等消耗  | 14.98    | 0.00      | 14.98  |
| 08     | 管理费用   | 19.04    | 0.00      | 19.04  |
| 09     | 绩效支出   | 29.50    | 0.00      | 29.50  |
| 备注     |  |          |           |        |

说明:

1.本表作为资金管理依据。 $D=B+C$ ； $B06 \leq (B02-B03) \times 30\%$ ,  $B06 \leq (B02-B03) \times 60\%$  (数学等纯理论基础研究项目)； $D06 \leq (D02-D03) \times 30\%$ ,  $D06 \leq (D02-D03) \times 60\%$  (数学等纯理论基础研究项目)；项目承担单位为企业的， $C01 > 0$ 。科目经费预算比例参照《深圳市科技研发资金管理办法》及深圳市科技计划项目相关经费预算编制指引执行。

2.鼓励先行投入项目研发，可追溯确认前期预研和筹备的经费投入，作为项目自筹部分确定项目预算，追溯期从项目申报之日起最长不超过6个月。

3.项目承担单位在提供可支持科研活动的项目设备证明后,已有设备可按现值计入自筹经费,使用财政资金购置的设备除外;同一项目设备可以用于不同科技项目,但不能重复计入不同项目的经费预算。

4.非预算单位可以使用科技研发资金列支人员费,但不得超过市统计部门公布的科学研究和技术服务业从业人员工资水平。预算单位不可以列支。

5.对项目聘用人员,可以使用科技研发资金列支劳务费(含社会保险补助、住房公积金),但不得超过市统计部门公布的科学研究和技术服务业从业人员平均工资水平。

6.间接费用实行总额控制,由项目承担单位统筹安排使用,并做好使用记录,可以全部用于绩效支出,但绩效支出只能用于本合同的项目组成员。

7.当财政资助资金少于申请资助资金时,项目承担单位为企业法人的,项目总经费不变,自筹经费相应补足;项目承担单位为非企业法人的,项目总经费和自筹经费部分可以调整,但调整后的总经费低于原定总经费60%的,应当向甲方提出书面申请。



### 资金预算确认及自筹经费承诺

我单位申报的重202410045 优异玉米新品种培育技术研究与应用示范项目计划总经费投入550.00万元。其中，申请市财政资助500.00万元、自筹经费50.00万元。根据深圳市科技研发资金和科技计划项目管理的有关规定，本单位的项目资金预算编列完成，自筹资金按时到位，承诺对获得的市财政资助资金不列支市科技研发资金使用负面清单事项。

项目负责人（签字）：

单位财务负责人（签字）：

张斌

## 项目拟购置、试制设备清单（单位：万元）


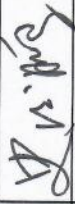
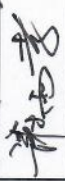

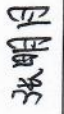

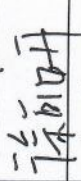
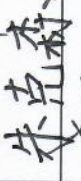
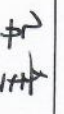

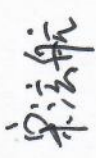
| 序号          | 仪器设备名称（品目） | 设备分类 | 数量/单位 | 单价（万元） | 经费来源 |
|-------------|------------|------|-------|--------|------|
| 购置设备费合计（万元） |            | 0.00 |       |        |      |
| 试制设备费合计（万元） |            | 0.00 |       |        |      |
| 设备费总计（万元）   |            | 0.00 |       |        |      |

说明：

- 1.设备分类：购置、试制。
2. 单价50万元及以上的设备费应当提供基本测算依据及明细说明。单项20万以上设备仪器和软件的购置，应当单独列示。其他可以按照品目合并列示。
- 3.项目承担单位属预算管理单位的，应当按要求另行编制政府采购计划。
4. 以甲方资助经费购置的大型科学仪器设施设备，应当按照有关规定进行评议，并在甲方指定的共享平台对外开放，但是，涉及国家安全等不宜公开的除外。



## 五、项目组成员

| 序号 | 角色    | 姓名  | 证件号码               | 联系电话        | 职称    | 学历    | 在项目中承担的任务  | 所在单位           | 签名  |
|----|-------|-----|--------------------|-------------|-------|-------|--|----------------|---|
| 1  | 项目负责人 | 杨芳  | 510126197702220624 | 18040518081 | 教授    | 博士研究生 | 项目设计统筹、实施推进、及时跟踪各子任务的实施情况，及时召集讨论、解决问题，确保后续工作顺利结题，负责项目年度总结及结题报告等。 | 中山大学·深圳        |    |
| 2  | 主要成员  | 史俊鹏 | 430424199004154239 | 13581863941 | 副教授   | 博士研究生 | 统筹相关数据分析，完成重要基因优异单倍型鉴定等。   | 中山大学·深圳        |    |
| 3  | 主要成员  | 谢雨孝 | 320102198806284612 | 15050670628 | 无     | 博士研究生 | 负责品种测试、推广、销售等工作  | 深圳农科玉种业有限公司    |    |
| 4  | 主要成员  | 陈舜权 | 44512119881004513X | 18927279663 | 高级农艺师 | 博士研究生 | 玉米新种质选育与新品种培育及改良，并完成育种效应评估。                                      | 深圳市作物分子设计育种研究院 |    |
| 5  | 主要成员  | 张明月 | 210902198803041023 | 15005167626 | 副教授   | 博士研究生 | 玉米群体耐密抗倒等性状表型鉴定、数据相关分析、单倍型鉴定、功能鉴定等。                              | 华南农业大学         |    |
| 6  | 其他成员  | 洪坤磊 | 411524199011161114 | 18123617735 | 工程师   | 硕士研究生 | 玉米品种试验、推广与销售。  | 深圳农科玉种业有限公司    |    |
| 7  | 其他成员  | 徐可妍 | 33250120000902004X | 18867840056 | 无     | 本科    | 遗传分析，分子机制相关内容。   | 中山大学·深圳        |   |
| 8  | 其他成员  | 朱克森 | 362329199606046112 | 18296117305 | 无     | 博士研究生 | 玉米新种质鉴定、农艺性状及育种效应评估。   | 中山大学·深圳        |  |
| 9  | 其他成员  | 姜艺  | 32062120000615332X | 19825307474 | 无     | 本科    | 玉米新种质鉴定与改良等。   | 中山大学·深圳        |  |
| 10 | 其他成员  | 刘欣桐 | 442000199808238404 | 18022126079 | 无     | 硕士研究生 | 新种质鉴定与改良等。   | 中山大学·深圳        |  |
| 11 | 其他成员  | 宋滋航 | 612522200204030015 | 17342351597 | 无     | 本科    | 新种质鉴定与改良效应评估等。   | 中山大学           |  |

| 序号 | 角色   | 姓名  | 证件号码                   | 联系电话            | 职称 | 学历    | 在项目中承担的任务                 | 所在单位           | 签名  |
|----|------|-----|------------------------|-----------------|----|-------|---------------------------|----------------|-----|
| 12 | 其他成员 | 袁书婷 | 130128199<br>210150026 | 133151745<br>38 | 无  | 博士研究生 | 玉米新种质鉴定与选育。               | 深圳市作物分子设计育种研究院 | 袁书婷 |
| 13 | 其他成员 | 王志奎 | 411526199<br>102150413 | 182569445<br>28 | 无  | 本科    | 各类突变体及种质田间表型鉴定、农艺性状调查。    | 深圳市作物分子设计育种研究院 | 王志奎 |
| 14 | 其他成员 | 林志明 | 441422199<br>903163710 | 152191097<br>95 | 无  | 本科    | 田间材料种植、鉴定、筛选, 育种效应评估等。    | 深圳市作物分子设计育种研究院 | 林志明 |
| 15 | 其他成员 | 李青云 | 610525199<br>407123723 | 156103641<br>98 | 无  | 博士研究生 | 数据分析、基因型鉴定、单倍型分析等。        | 中山大学·深圳        | 李青云 |
| 16 | 其他成员 | 杨娟  | 620123199<br>110073224 | 130513778<br>65 | 无  | 博士研究生 | 关键基因机制解析、玉米新种质鉴定、育种效应评估等。 | 华南农业大学         | 杨娟  |

是否同意将项目负责人及主要成员信息用于自动申请加入市科技评审专家库: 是

说明:

- 1.项目负责人应为项目承担单位全职人员。
- 2.主要成员最多4人。
- 3.其他成员按实际情况列明。



# 六、合作单位 (选填)

| 承担单位/合作单位 | 单位名称 (盖章)            | 统一信用代码             | 任务分工   | 自筹经费<br>出资额<br>(万元) | 自筹经费<br>分摊金额<br>(万元) | 市财政资<br>助额分配<br>(万元) |
|-----------|----------------------|--------------------|--|---------------------|----------------------|----------------------|
| 承担单位      | 中山大学·深圳<br>2025.1.20 | 12440300MB2C47612Y | 负责项目实施及协调, 承担玉米矮化、耐密植、抗倒伏优异基因挖掘3-4个, 开展相关机理研究, 发表高水平论文≥3篇, 申请发明专利≥3件, 创制玉米矮化、耐密植新种质4-5份, 改良获得耐密高产产品种2-3个; 负责在深圳市对口帮扶地区建立应用示范工程1项, 面积≥200亩。 | 0.00                | 0.00                 | 225.00               |
| 合作单位      | 深圳市作物分子设计育种研究院       | 12440300051512030A | 负责玉米矮化、耐密、抗倒伏优异种质材料的收集、创制和新品种培育、改良等研究, 培育矮化、耐密、抗倒伏玉米新品种2-3个, 并申报国家植物新品种权; 创制玉米矮化、耐密植、抗倒伏新种质2-3份, 申请专利≥4件。                                  | 0.00                | 0.00                 | 175.00               |
| 合作单位      | 深圳农科种业有限公司           | 914403006939986340 | 负责在深圳市对口帮扶地区开展矮化耐密抗倒伏优异玉米新品种的示范推广, 实现销售收入≥1200万元。  | 50.00               | 50.00                | 50.00                |
| 合作单位      | 华南农业大学               | 124400004554165634 | 负责玉米矮化、耐密、抗倒伏优异种质创制及品种改良等研究; 承担玉米矮化、耐密植、抗倒伏相关基因挖掘2个, 创制或改良矮化耐密新种质2份, 申请专利≥1件。  | 0.00                | 0.00                 | 50.00                |
|           |                      |                    | 合计   | 50.00               | 50.00                | 500.00               |

说明:

- 1.乙方和合作单位应明确各自任务分工、自筹经费的分摊和财政资金的分配。
- 2.在财政资金分配上, 乙方所获得的金额应不低于单个合作单位所获得金额。
- 3.乙方负有项目实施的主要责任。
- 4.自筹经费出资额是指项目承担单位或者合作单位投入的项目自筹资金。自筹经费分配金额是指项目承担单位或者合作单位获得的项目自筹资金。

金。



出现以下情形的，乙方应当在项目实施期限届满之前，及时报甲方备案：

- （一）单位名称发生变更；
- （二）项目组主要成员、其他成员发生变更；
- （三）在项目总经费不减少且不超过间接费用控制限额的前提下，设备费、间接费用、设备清单发生调整。

乙方的法定代表人、通讯地址、联系人、联系电话等信息发生变化时，应及时在深圳市科技业务管理系统办理有关信息更新手续。

**第七条** 出现下列情形之一的，乙方应当在项目实施期限届满之前向甲方提出变更申请：

- （一）项目负责人因工作调动、伤病、死亡或者其他重大原因无法继续履行工作职责，确需变更项目负责人的；
- （二）项目负责人工作发生调动，确需变更项目承担单位的，拟变更的项目承担单位应当符合该项目申请指南的申请条件、具备继续实施项目的能力和科研条件，且应当经原项目承担单位与拟变更项目承担单位协商一致；
- （三）因客观原因导致项目实施进度被迫延迟，确需申请变更实施期限的（延期单次不超过1年、总计不超过2次、总延期时长不超过原项目实施期的一半）；
- （四）承担单位经协商，确需变更合作单位或资金分配的；
- （五）其他需要甲方批准变更的情形。

**第八条** 出现下列情形之一的，甲方有权终止项目：

- （一）项目实施过程中，经证明技术路线不合理、不可行且无替代方案，导致项目无法完成的；
- （二）因项目研究开发的关键技术已由第三方公开，或者市场发生重大变化，使研究开发工作成为不必要的；
- （三）乙方因经营异常等导致对项目实施产生重大影响或者已不具备履行科技计划项目能力的；
- （四）项目实施过程中被责令限期整改，未按期完成整改或者整改未达到要求的；
- （五）不遵守合同书（任务书）规定，未履行合同书（任务书）约定的主要义务的；
- （六）乙方及合作单位、项目负责人及项目组主要成员在项目实施、经费使用、科研诚信和科技伦理、安全责任、知识产权侵权、研发成果剽窃等方面出现性质恶劣、影响较大、涉及金额较大等重大违法违规行为的；
- （七）在项目立项、项目实施和验收过程中，发生违反中介相关申报承诺、贿赂或者利益交换等不正当行为的；
- （八）失信联合惩戒对项目有重大影响的；
- （九）导致项目无法实施的其他情形。

项目终止的，甲方按程序停止后续拨款，委托会计师事务所进行项目资金专项审计，确定应退回资金额度（含结余资金、使用不合规资金），乙方应当按照甲方终止通知的要求及时退回应退资金及利息。

**第九条** 乙方在本合同规定的项目实施期限届满之日起6个月内，应当主动向甲方提出项目验收申请，并按照国家 and 省有关规定，在提交验收时一并提交科技报告及其他材料。

超出本合同规定的项目实施期限届满之日起6个月仍未申请验收的，甲方有权终止项目，并自验收期限届满之日起，最长三年内不予受理和立项项目负责人参与的市科技计划项目，不推荐其申报国家、广东省科技计划项目，不授予市科技奖励，不提名国家、广东省科技奖励。乙方为企业的，按照对项目负责人的处理方式，对乙方进行处理。

**第十条** 验收结论为通过，且乙方和项目负责人无不良科研信用记录，结余资金及利息留归



九、合同签约各方

甲方（盖章）：深圳市科技创新局

授权代表（签字）：黄薇 日期：2025年01月15日

经办人（签字）：廖敬扬、方一臻

乙方（盖章）：中山大学·深圳

法定代表人（签字）：高松 日期：2025年01月15日

项目负责人（签字）：杨芳

开户单位名称：中山大学·深圳

开户银行名称：中国农业银行深圳市分行

开户银行账号：41000500040085564

# 江苏省科技项目合同

计划类别 基础 research 计划（自然科学基金）--青年基金项目

项目编号 BK20180529

项目名称 基于 GWAS 的梨果实抗坏血酸代谢基因挖掘及调控网络研究

项目类别 无

起止年限 2018 年 7 月 至 2021 年 6 月

项目负责人 张明月 电话及手机 15005167626 025-84396485

项目联系人 赵珩 电话及手机 13815860856 025-84399807

承担单位 南京农业大学

单位地址 江苏南京卫岗 1 号 邮政编码 210095

项目主管部门 南京农业大学

江苏省科学技术厅

二〇一八

**委托单位（甲方）：** 江苏省科学技术厅

法定代表人： 王秦

地址： 南京市北京东路 39 号

邮政编码： 210008

**承担单位（乙方）：**

承担单位： 南京农业大学

法定代表人： 周光宏

地址： 江苏南京卫岗 1 号 邮政编码： 210095

项目负责人： 张明月

电话： 15005167626 传真：           

电子邮件： kjcxmk@njau.edu.cn

**保证单位（丙方、项目主管部门）：** 南京农业大学

法定代表人（科技局局长）： 周光宏

地址： 江苏省 南京市 玄武区 南京农业大学 卫岗一号 邮政编码：  
210095

甲方批准由乙方承担省科技计划《基于 GWAS 的梨果实抗坏血酸代谢基因挖掘及调控网络研究》项目的研究开发或建设任务。依据《中华人民共和国合同法》的规定，为明确甲、乙、丙三方的权利和责任，保证项目的顺利实施和科研经费的合理使用，签订本合同。

## 一、项目的目标和主要研究内容

要解决的主要技术难题和问题，项目研究的创新点和内容等。

- 1、获得 312 份砂梨全基因组的基因型数据，查找全基因组范围的变异信息
- 2、利用 312 份砂梨自然群全基因组关联分析方法，挖掘砂梨果实维生素 C 代谢的关键候选基因
- 3、通过草莓稳定转化和梨果实的瞬时转化，明确基因在果实中所行使的调控功能
- 4、通过基因组的三维空间结构互作 Hi-C 技术和酵母双杂交技术，明确维生素 C 代谢相关基因的调控网络
- 5、通过筛选与维生素 C 含量相关的变异位点，开发分子标记，应用于分子标记辅助育种



## 二、项目验收内容和考核指标

包括 1、主要技术指标：如形成的专利、新技术、新产品、新品种、新装置、论文专著等数量、指标及其水平等；2、主要经济指标：如技术及产品所形成的市场规模、效益等；3、项目实施中形成的示范基地、中试线、生产线及其规模等；4、其他应考核的指标。

获得 312 份砂梨全基因组的基因型数据，查找全基因组范围的变异信息利用 312 份砂梨自然群全基因组关联分析方法，挖掘砂梨果实维生素 C 代谢的关键候选基因通过草莓稳定转化和梨果实的瞬时转化，明确基因在果实中所行使的调控功能通过基因组的三维空间结构互作 Hi-C 技术和酵母双杂交技术，明确维生素 C 代谢相关基因的调控网络通过筛选与维生素 C 含量相关的变异位点，开发分子标记，应用于分子标记辅助育种 1、明确与梨果实维生素 C 代谢关联的重要变异位点获得控制梨果维生素 C 代谢相关的重要功能基因探究目的基因在梨果实和草莓果实中的时空表达特征初步明确梨果维生素 C 代谢相关的基因调控网络开发高维生素 C 梨品种的分子标记，建立分子辅助育种技术体系 2、筛选高维生素 C 含量的优质梨品种，促进梨产业健康稳定发展开发应用于苗期筛选高维生素 C 含量的分子标记，快速鉴定优质品种 3、研究结果申请专利 2 项，发表学术论文 5 篇，其中 SCI 论文 2 篇，中文核心期刊 3 篇。4、改良梨果实营养品质，推动梨产业稳定健康发展；培养 2 名博士研究生，培养 2 名硕士研究生

### 三、项目进度及考核指标

| 时间                             | 考核指标   |
|--------------------------------|--|
| 2018 年 7 月<br>至<br>2018 年 12 月 | 在已获得的 312 份砂梨自然群体高通量测序数据的基础上,对个别不满足测序质量要求的品种进行补充、测序;                                       |
| 2019 年 1 月<br>至<br>2019 年 6 月  | 质量评价、序列比对、SNP 查找;结合补充数据,利用一般线性模型与混合线性模型进行全基因组关联分析;对关联区域进行连锁不平衡分析、基因查找、功能注释,获得维生素 C 相关候选基因; |
| 2019 年 7 月<br>至<br>2019 年 12 月 | 对候选基因进行基因表达量验证、基因亚细胞定位;构建超表达载体,转模式植物草莓,观察 T1 代转基因果实发育期的变化;                                 |
| 2020 年 1 月<br>至<br>2020 年 6 月  | 构建候选基因的过量表达和 RNAi 转化载体,观察对梨果实发育期的影响;   |
| 2020 年 7 月<br>至<br>2020 年 12 月 | 对关联分析候选区域进行分子标记开发,应用于杂交后代的苗期早期筛选;对上一年度的实验数据结合本年情况进行补充、重复、核验等;                              |

|                               |              |
|-------------------------------|--------------|
| 2021 年 1 月<br>至<br>2021 年 6 月 | 完成专利申报、论文发表。 |
|-------------------------------|--------------|

#### 四、项目承担单位、参加单位及主要研究开发人员

(一) 项目承担单位、参加单位

|            |        |  |           |
|------------|--------|--|-----------|
| 项目承担单位     | 南京农业大学 |  |           |
| 项目合作<br>单位 | 单位名称   |  | 国家或<br>地区 |
|            | 单位一    |  |           |
|            | 单位二    |  |           |
|            | 单位三    |  |           |
|            | 单位四    |  |           |
|            | 单位五    |  |           |
| 境外合作<br>单位 | 单位一    |  |           |
|            | 单位二    |  |           |

(二) 项目主要研究开发人员

| 姓名                    | 性别 | 出生年份 | 职称 | 学位 | 从事专业 | 本项目工作时间(%) | 工作单位   | 项目分工                | 国别 | 身份证件号码             |
|-----------------------|----|------|----|----|------|------------|--------|---------------------|----|--------------------|
| 项目负责人                 |    |      |    |    |      |            |        |                     |    |                    |
| 张明月                   | 女  | 1988 | 中级 | 博士 | 果树学  | 90         | 南京农业大学 | 全基因组关联分析；梨转化；分子标记开发 | 中国 | 210902198803041023 |
| 项目骨干（不超过 5 人，不含项目负责人） |    |      |    |    |      |            |        |                     |    |                    |
| 谷超                    | 男  | 1985 | 副高 | 博士 | 果树学  | 60         | 南京农业大学 | 梨瞬时转化、草莓稳定遗传转化      | 中国 | 340721198508144836 |
|                       |    |      |    |    |      |            |        |                     |    |                    |
|                       |    |      |    |    |      |            |        |                     |    |                    |
|                       |    |      |    |    |      |            |        |                     |    |                    |
|                       |    |      |    |    |      |            |        |                     |    |                    |
| 参加人员（不超过 10 人）        |    |      |    |    |      |            |        |                     |    |                    |
| 明美玲                   | 女  | 1994 | 其他 | 学士 | 果树学  | 70         | 南京农业大学 | 参与梨瞬时转化；草莓稳定遗传转化    | 中国 | 36078219941018542X |
| 汪润泽                   | 男  | 1990 | 其他 | 学士 | 果树学  | 70         | 南京农业大学 | 参与全基因组关联分析模型        | 中国 | 34082619900901561X |



## 五、项目经费预算

### （一）项目经费来源预算

经费单位：万元

|           | 预算数 | 2018 年 | 2019 年 | 2020 年 | 备注 |
|-----------|-----|--------|--------|--------|----|
| 合计        | 20  | 20     | 0      | 0      |    |
| 1、省拨款     | 20  | 20     | 0      | 0      |    |
| 2、部门、地方配套 | 0   | 0      | 0      | 0      |    |
| 3、承担单位自筹  | 0   | 0      | 0      | 0      |    |
| 4、其他来源    | 0   | 0      | 0      | 0      |    |

### （二）项目经费支出预算

经费单位：万元

|                     | 预算数  | 其中：省拨款 | 备注   |
|---------------------|------|--------|--|
| （一）直接费用             | 18.6 | 18.6   |  |
| 1、设备费               | 0.3  | 0.3    | 主要包括设备改造与租赁费，其中主要是指现有仪器的部分零部件的保养和维修等费用，包括电泳槽电极更换、植物培养箱灯管更换、移液器校准和维修等 |
| 2、材料费/测试化验加工费/燃料动力费 | 13.2 | 13.2   | 用于实验常规分子生化试剂的购置、DNA 和 RNA 提取   |

|                   |      |      |                                |
|-------------------|------|------|--------------------------------|
|                   |      |      | 等试剂盒购买、耗材购买                    |
| 3、差旅费/会议费/国际合作交流费 | 1.5  | 1.5  | 主要用于学术交流、采样、田间交通差旅费用           |
| 4、劳务费/专家咨询费       | 1.2  | 1.2  | 参加项目的研究生补助和试验地工人劳务支出；学术讲座专家咨询费 |
| 5、其他费用            | 2.4  | 2.4  | 用于SCI论文版面费、中文核心期刊面板费、专利申请费用    |
| (二) 间接费用          | 1.4  | 1.4  |                                |
| 6、管理费             | 0.42 | 0.42 | 项目管理等费用                        |
| 7、绩效支出            | 0.98 | 0.98 | 绩效支出等其他用途                      |
| 合计                | 20   | 20   |                                |



## 六、其他条款

### （一）缔约各方的权利、义务

第一条 缔约各方均应共同遵守国家、省有关科技计划与经费管理的规定，严格遵守并认真履行本合同的各项条款。

甲方应按合同约定的金额提供项目研究开发经费，有权监督、检查合同履行情况。合同履行期间，甲方有权直接组织或委托丙方检查、监督乙方对本合同的履行情况。乙方完成项目研究开发任务后，由甲方负责进行验收。

乙方应严格履行合同义务，为项目实施提供承诺的技术与条件保障，以及财务管理、成果管理、科技档案管理服务合同约定的其他义务。项目申请验收前乙方应按照规定提交科技报告，未提交科技报告的项目不予验收。乙方应加强项目实施成果的转化，自项目验收后一年内未实施转化的项目，甲方有权责成乙方将成果交省内技术产权交易机构挂牌转让。

丙方应按合同约定的金额提供项目配套经费，并进行相关的协调和监督。

第二条 甲方有权根据乙方项目计划进度完成情况决定是否拨付后续经费。乙方使用项目经费应按照合同约定的支出范围执行，保证专款专用，并实行单独核算，严禁弄虚作假、截留和挪用项目经费等违反财经纪律的行为。

第三条 甲、乙、丙各方对项目合同及其他技术资料负有保密责任。

### （二）违约责任

第四条 甲方未能按合同约定的经费数提供经费，导致乙方研究开发工作延误的，应允许合同规定的研究开发工作完成期限相应顺延。

第五条 因乙方原因，导致研究开发工作未能达到合同约定指标的，乙方应采取措施尽快使项目达到合同预定要求，并承担由此而增加的费用。

第六条 乙方无正当理由未履行合同时，甲方有权停拨、追缴部分或全部省拨经费，由此造成的经济损失由乙方承担。对乙方在申报和实施项目中的失信行为，甲方将根据省科技信用管理的有关规定记入不良信用记录，并报送至省公共信用信息平台，列入乙方的社会信用记录。

第七条 乙方违反经费使用规定或经甲方检查确认计划进度不符合合同约定的，甲方有权减拨或停拨后续经费；情节严重的，甲方有权终止合同，乙方应返还甲方已拨付的全部经费。

第八条 乙方因不可抗力不能履行合同义务时，可以免除违约责任，但应及时通知甲、丙方，并在合理的期限内出具因不可抗力导致合同不能履行的证明。

第九条 在履行本合同过程中，确因在现有水平和条件下难以克服的技术困难，导致研究开发部分或全部失败造成损失的，经甲方确认风险责任后，甲方在其拨款额度范围内承担损失。

### （三）合同的变更、解除和争议解决

第十条 合同的变更或解除，须经缔约各方协商一致，并签署书面文件。

第十一条 发生下列情况之一的，缔约方应当协商变更或解除合同：

(1) 由于不可抗力或意外事故导致合同无法履行或部分无法履行；  
(2) 由于项目目标已被他人先行实现，有关成果已被申请专利或公开，继续履行合同已无必要；

(3) 由于乙方未按合同要求履行合同，或是由于其他原因，导致项目在检查或评估中被淘汰的。

第十二条 合同一方发生合并、分立或更名时，由变更后的单位继受或分别继受变更一方在合同中的权利义务。

第十三条 合同在履行过程中发生争议的，缔约各方应通过友好协商的方式解决。如协商不成时，缔约各方有权向人民法院起诉或仲裁机构申请仲裁，但在有关司法、仲裁结果生效之前，乙方有义务按照甲方要求继续履行或终止履行本合同。

#### (四) 附 则

第十四条 项目任务书、可行性论证报告作为合同附件。项目如涉及多家（包含两家）单位参加，乙方应在签订本合同前与有关单位就合作任务和知识产权分配等问题签订有关合同或协议（仅委托其他单位进行常规试验、提供社会化科技服务和少量辅助科研工作的情况除外），同时作为本合同的附件。

第十五条 有关合同的未尽事宜，按照国家、省有关科技计划与经费管理的规定执行。

第十六条 本合同正本一式六份（甲、乙、丙方各执2份），自缔约各方签章后生效。

第十七条 本合同的解释权归甲方享有。

### 七、附加条款

第十八条 省自然科学基金青年基金项目负责人在申报该类项目前，必须是未主持承担过省级或国家级科技计划项目（含国家自然科学基金项目）。否则，一经发现，按弄虚作假处理，取消该项目立项，并追究项目负责人、承担单位和科技主管部门责任。

## 八、签订合同各方

甲方：

法定代表人或委托代理人（签字）

项目主管处室负责人（签字）

项目主管处室经办人（签字）

公 章

年 月 日

乙方：

承担单位法定代表人或委托代理人（签字）

项目负责人（签字）

公 章

开户银行、帐号

年 月 日

丙方：

法定代表人或委托代理人（签字）

公 章

年 月 日

**说明：**

1、本合同适用于省基础研究、重点研发、创新能力建设、国际科技合作、软科学研究等计划。

2、合同条款中所有空项都需如实填写，确无此项的，请在该栏中打“/”或在空白处写“无”。

3、乙方盖章必须是单位公章，部门章无效。

## 证明

南京农业大学园艺学院工作人员张明月同志承担科研情况：主持江苏省自然科学基金青年基金项目“基于 GWAS 的梨果实抗坏血代谢基因挖掘及调控网络研究”（项目批准号：BK20180529），资助经费：20 万元，经费本号：130-6S01180006，执行年限：2018 年 7 月-2021 年 6 月。



主持博士后面上一等资助



## 证明

南京农业大学园艺学院工作人员张明月同志承担科研情况：主持第 63 批中国博士后科学基金面上项目“基于 GWAS 的梨果实发育期调控基因挖掘及功能研究”，获一等资助（项目批准号：2018M630565），资助经费：8 万元，经费本号：440-82171321，执行年限：2018 年-2020 年。



21  
合同编号: 2024-NPY-00-045

## 2024年省级乡村振兴战略专项资金 种业振兴行动项目

### 合 同 书

项目名称: 玉米基因编辑技术研究及种质创制

项目管理单位(甲方): 广东省农业农村厅

项目牵头承担单位(乙方): 华南农业大学

项目推荐(主管)单位(丙方):

项目负责人: 王海洋

联系电话: 13552603750

项目联系人: 孔德鑫

联系电话: 13668985185

广东省农业农村厅制



第一条 为保障 2024 年省级乡村振兴战略专项资金种业振兴行动项目顺利实施，按时保质保量完成项目任务，根据《中华人民共和国民法典》《广东省省级财政专项资金管理办法（修订）》（粤府〔2023〕34 号）、《广东省农业农村厅种业振兴行动专项资金管理办法（试行）》等文件有关规定，经甲、乙、丙三方协商一致，签署本合同书。

第二条 甲方的权利义务：本合同履行过程中，甲方有权对乙方项目的实施情况和资金到位、使用情况进行监督、检查，提出改进要求。

第三条 乙方的权利义务：

1.按财政资金管理规定，对甲方核拨的资金做到专款专用，单独列账，并随时配合甲方进行监督检查。

2.认真填写本合同书《项目任务书》，《项目任务书》的内容应与乙方的《项目申报书》保持一致。

3.严格按照本合同书及合同书《项目任务书》的要求及时完成项目建设内容，项目实施完成后，按照本合同验收报告模版要求提交验收报告。

4.按照《广东省农业农村厅种业振兴行动专项资金管理办法（试行）》规定，按期（每年 6 月 30 日、12 月 31 日）向甲方、丙方书面报告项目实施进展及资金使用情况等内容。

5.乙方需保留与所有参与单位的合作实施协议和相关财务凭证，并向甲方备案。

第四条 丙方的权利义务：

1.为乙方项目实施提供必要的条件保障。

2.负责对项目承担单位的实施条件、能力以及财务管理规范进行审查，对推荐项目的实施场地、申报资料等进行真实性审核，并监督项目实施、资金预算执行情况。

3.协助甲方完成对财政资金投资 500 万元（含）以上【科研项目财政资金在 200 万元（含）以上】项目验收等工作；负责完成对财政资金投资 500 万元以下（科研项目财政资金 200 万元以下）项目验收工作，并及时向甲方报告情况。

第五条 本项目资金不得用于以下方向：1.行政事业单位基本支出；2.各项奖金、津贴、福利补助、职工工资、奖励绩效等；3.企业担保金和弥补企业亏损；4.修缮楼堂馆所以及建造职工住宅；5.弥补单位预算支出缺口和偿还债务；6.购买交通工具及通讯设备；7.形成地方政府债务的支出；8.购买理财产品、发放借款及平衡预算等。

第六条 项目验收。项目验收及结果处理严格执行《广东省农业农村厅专项资金项目验收管理办法（试行）》（粤农农办〔2023〕73 号）的规定。乙方应在项目完成后 3 个月内，提出验收申请。申请验收除了规定材料外，还应该提交项目审计报告或者经费决算表，其中财政专项资金 50 万元以下的项目，需提交由项目承担单位财务部门出具的经费决算表，财政专项资金 50 万元（含）以上的项目，需提交由项目承担单位委托会计师事务所出具的审计报告。对财政资金投资 500 万元（含）以上【科研项目财政资金在 200 万元（含）以上】的项目，及乙方直接向甲方申报的项目，由甲方负责组织验收；对财政资金投资 500 万元以下（科研项目财政资金 200 万元以下）的项目，由项目推荐（主管）单位（丙方）负责验收，验收单位向甲方提交验收材料，甲方对验收材料进行审核确认。

第七条 在履行本合同的过程中，如出现相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。因非不可抗力因素导致的项目未履行或未履行完毕，或因乙方责任造成项目不能继续开展的，甲方有权终止项目合同，收回尚未使用和使用不符合规定的财政经费。



第八条 在履行本合同的过程中,乙方发现可能导致项目整体或部分失败的情形时,应及时通知甲方,并采取适当措施减少损失,没有及时通知并采取适当措施,致使损失扩大的,应当就扩大的损失承担责任。

第九条 实施项目所获得的科技成果(项目成果)归属、成果转让和实施技术成果所产生的经济利益的分享,按照国家和广东省有关规定执行。项目研究成果应向省农业农村厅进行登记、备案,对外发布前应征求省农业农村厅的意见。

第十条 本合同在履行过程中发生的任何争议,由甲乙丙三方友好协商解决。

第十一条 本合同未尽事宜,各方同意按照《广东省省级财政专项资金管理办法(修订)》(粤府〔2023〕34号)、《广东省农业农村厅种业振兴行动专项资金管理办法(试行)》履行。

项目管理单位(甲方) (盖章): 广东省农业农村厅

法定代表人(或授权代表) (签章):

签订日期: 2025年2月19日

项目牵头承担单位(乙方) (盖章): 华南农业大学

法定代表人(或授权代表) (签章):

项目负责人(签章): 王海洋

签订日期: 2025年1月14日

乙方推荐(主管)单位(丙方) (盖章):

法定代表人(或授权代表) (签章):

签订日期: 年 月 日

# 项目任务书

## 填写说明

一、本项目任务书由乙方填写。

二、本项目任务书所列内容应实事求是填写，表达要明确、严谨。对填写不符合要求的，或填报内容出现虚报夸大、不切实际的，将退回项目承担单位修改。

三、项目任务书规定的项目考核指标、建设内容和绩效目标必须依据《项目申报书》填写，应遵循明确、量化、可考核的原则，其中技术指标应明确项目完成时达到的关键技术参数及预期可以形成的发明专利、标准、新技术、新产品、新装置、论文、专著等的数量。项目申报指南对项目技术、经济和成果等指标有明确要求的，应符合项目申报指南的要求，相关专项管理办法有特别规定的，应符合相关规定。

四、《项目申报书》及申报指南是本项目任务书填报的重要依据，项目任务书填报不得修改考核指标、绩效目标、资金预算等内容。《项目申报书》、申报指南和本项目任务书将共同作为项目过程管理、综合绩效评价（验收）和监督评估的重要依据。

五、省财政资金支出的预算计划应按照国家及省相关规定执行。

六、表格栏目不够可自行增加。



## 一、目的及意义

玉米是我国第一大农作物，也是广东省重要的粮经兼用作物。广东是我国农业和人口大省，目前的粮食自给率却只有约 33%。广东省年玉米种植面积约 200 万亩左右，是省内粮食生产供给的重要组成部分，玉米的充足优质供应对保证广东乃至全国的粮食安全至关重要。长期的玉米育种和生产实践表明，耐密性提升和杂种优势高效利用是提高玉米产量的关键技术措施。玉米品种耐密性和其耐密株型密切相关，而耐密株型的组成中至少包括了半矮秆、抗倒伏、紧凑叶夹角、早花等，其中，较小的茎叶夹角可以减少玉米植株之间的相互遮荫，利于通风透光，利于密植。株高和穗位高会影响玉米植株重心，进而影响玉米的抗倒伏能力和收获指数，影响“源-库”平衡，适当降低株高有利于玉米的密植和高产。而合理的开花期是玉米合理避害，并适应不同生态区及现代栽培制度的关键。

杂种优势的高效利用是提高玉米产量的关键技术措施。杂种优势是生物界的普遍现象，也是提高农作物产量的一种有效途径。从上世纪 60 年代我国第一个单交种诞生到如今，基本以十年为一个单位，发生了至少 6 次大规模的单交种更新换代。目前我国生产上 99% 以上的玉米品种都是基于杂种优势的杂交种。玉米杂交种比开放授粉品种的产量可提高 50% 以上，比亲本自交系普遍可提高 100% 以上。因此，杂种优势的提高是玉米增产的有效途径。另外，提高玉米籽粒品质是提高玉米营养和经济价值的重要要求。玉米品质育种的具体内容包括改良玉米籽粒中的淀粉、蛋白质、油分等成分的含量和组成，使其更适合人类和家畜的营养需要或者服务于特殊的工业用途。其中高蛋白玉米改良是当前生产上重要的需求方向。因此，以玉米株高、开花期、蛋白品质及杂种优势利用为重心，开展对应的遗传基础研究、育种技术开发及种质创制，是培育高产优质玉米品种，提高我省粮食自给能力的重要途径。尽管目前诸多农艺性状方面均已有一定的调控基因被克隆，但大多数基因仅仅停留在理论研究层面，几乎没有育种应用。并且对于少数几个育种应用的基因大多还是基于传统的分子标记辅助选择等方法，存在周期长、效率低、精准度差等问题。

基因编辑 (gene editing)，又称基因组编辑 (genome editing)，是一种能够精确对生物体基因组特定目标基因进行修饰的基因工程技术。2016 年，以碱基编辑技术为代表的编辑工具率先实现将 CRISPR-Cas 系统从切割 DNA 的“剪刀”变为能改



写特定碱基的“铅笔”，打开了精准基因组编辑的大门。基因编辑系列衍生技术的开发为提高玉米等作物育种效率提供了重要的基础支撑。目前，世界范围内已经系统建立了多样化的基因编辑技术体系，开发了多重编辑、DNA 片段的精准插入或删除、DNA 大片段操作、基因表达精细调控等多层次编辑技术。值得一提的是，针对传统作物定向改良育种周期长、成本高、效率低的问题，项目组的前期巧妙地将单倍体诱导与基因编辑技术结合起来，开发了“单倍体介导的基因编辑技术（IMGE）”，2 代内可获得性状精准改良的育种纯系，大大加快玉米育种进程。该技术不受待改良材料遗传背景和遗传转化能力的限制，普适性广；精准高效，可提升作物定向改良效率 3 倍以上，并且改良后的材料不携带基因编辑载体，不含有外源转基因成分，可打破基因编辑育种的遗传转化和载体去除两大技术屏障，增加了基因编辑育种的灵活性及其产品开发的便捷性；且操作过程简便，在单倍体诱导过程中即可进行基因编辑，“能授粉就能做基因编辑”，对育种家异常友好。被业界同行认为有望成为“下一代作物育种技术”。尽管世界范围内基因编辑技术已有了长足的发展，但在我国种质改良中的应用还相对有限，尤其是在玉米改良研究方面的进展还较慢。

玉米株高、开花期、蛋白品质改良和杂种优势高效利用对我省乃至我国玉米安全生产至关重要，为了高效的进行玉米种质改良，项目组成员前期结合基因编辑技术，鉴定了可调控玉米株高/穗位高的 *ZmSPL12*、*ZmTCP4/14/25*、*Br2* 等基因，调控玉米开花期的 *ZmDBB6* 等基因，调控玉米产量的 *ZmKOB1* 基因，调控玉米蛋白品质的 *ZmACP1*、*ZmSK1*、*ZmCRR5* 等基因，构建了玉米杂种优势研究的遗传群体，初步鉴定了调控玉米杂种优势的关键基因。针对高产优质玉米改良的重大产业需求，本项目中，我们拟利用前期鉴定控制株高、生育期、产量及品质等基因及突变材料基础上，结合转录组、代谢组、蛋白质等多组学分析，解析其性状调控遗传网络；结合基因编码区及非编码区基因编辑技术和 IMGE 技术，创制目标性状突出且综合性状优良的玉米新种质；并培育高产优质的强优势玉米杂交种，为我省乃至全国玉米安全高效生产提供理论、基因及种质方面的支撑。



## 二、项目建设内容

详细说明项目建设内容（项目需求或项目建设任务，包含项目参与单位的建设内容）。

针对广东玉米产业在种质创新和新品种培育中的技术瓶颈问题，本项目按照指南建设内容要求，分别设立了三大课题：课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析；课题 2：玉米杂种优势多组学分析与杂种优势基因挖掘；课题 3：高产优质玉米种质创新与新品种培育（课题 1 负责人：牵头承担单位孔德鑫教授；课题 2 负责人：参与单位 2 卢洪研究员；课题 3 负责人：参与单位 1 李坤博士），为我省乃至全国玉米安全高效生产提供理论、基因及种质方面的支撑。

项目牵头承担单位：华南农业大学

项目参与单位 1：广东省农业科学院作物研究所

项目参与单位 2：中国农业科学院深圳农业基因组研究所（岭南现代农业科学与技术广东省实验室深圳分中心）

### 课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析

课题 1 负责人：孔德鑫

参与人员：王海洋、柳青、张晓明、徐国强、姚旻昊、高永盟、赖晓君、李坤

主要研究内容如下：

#### （1）玉米株高、生育期和产量调控基因突变体材料创制与多组学分析

针对目前制约广东省玉米育种和产业化发展的瓶颈问题，本项目团队在前期广泛收集了玉米优良自交系、杂交种骨干亲本，开展了多年多点田间条件下各类农艺性状的表型鉴定与评价，并利用群体遗传学结合全基因组关联分析、转录组分析、eQTL、TWAS 分析等技术鉴定到一批调控玉米株高（*ZmSPL12*、*ZmTCP4/14/25*、*Br2*）、生育期（*ZmDBB6*）和产量（*ZmKOB1*）的候选基因。

前期我们利用基因编辑技术创制了这些基因的突变体材料，在基因型和表型鉴定基础上，本项目将利用 RNA-seq、DAP-seq、IP-MS 等多组学技术结合酵母单杂交、EMSA、酵母双杂交、双分子荧光互补、免疫共沉淀、玉米原生质体瞬时转化等实验鉴定其直接下游靶基因和互作蛋白，构建其分子调控网络。为揭示这些基因在控制玉米株高、生育期、产量方面的生物功能提供理论依据。

#### （2）调控玉米品质性状的基因编辑突变体材料创制与多组学分析

本项目团队前期利用 QTL 定位并克隆了调控玉米籽粒品质关键基因 *ZmACP1*（编码酸性磷酸酶 1）以及风味和产量平衡关键基因 *ZmSK1*（编码莽草酸激酶 1）和 *ZmCRR5*（编码一个糖苷水解酶蛋白）。为了深入研究这些基因影响玉米品质和风味的生物功能，本项目已利用基因编辑技术创制了这些基因的突变材料。通过表型鉴定发现，在 *ZmACP1* 基因敲除系籽粒组织中，腺苷和烟酰胺核糖含量显著降低；而 *ZmSK1* 基因敲除系中，其籽粒奎宁酸含量及产量显著高于野生型，有较大育种价值。在基因型和表型鉴定基础上，本项目将利用蛋白组、代谢组等多组学技术筛选 *ZmACP1*、*ZmSK1*、*ZmCRR5* 编码蛋白的催化底物及其互作调控因子，并进一步利用



酵母双杂交、双分子荧光互补、免疫共沉淀等技术鉴定 *ZmACP1*、*ZmSK1*、*ZmCRR5* 的候选互作蛋白。

## 课题 2：玉米杂种优势多组学分析与杂种优势基因挖掘

课题 2 负责人：卢洪

课题参与人员：张明月、郭小萍、余映宏、杜国华、马亚影、尹翼亮、何佳甜、许靓蕊、王沛东、李平、徐淑环

主要研究内容如下：

### (1) 玉米杂种群体及亲本大田表型鉴定，获得产量、株高、生育期等相关杂种优势表型数据

对我国近 1200 份育种材料大田评价和基因组分析的基础上，获得与株高、穗位高、穗位系数（穗位高/株高的比例）、生育期、雄花分枝数、产量因子等一系列表型性状紧密关联的 QTL，对重要候选基因的启动子区域进行基因编辑，产生系列突变体材料（*knr2*、*cct1*、*hsf21*、*dw3*、*gib2*、*srt2*），选择能够代表各杂种优势群的优异育种材料与三个骨干系（分别代表 SS、NSS、Iodent 群体）进行杂交，组建杂交组合。在此基础上对这些杂交种群体及亲本进行多点大田表型试验，全面评估在不同环境条件下的农艺性状、抗性、产量和潜在的育种价值，获得产量、株高、生育期等相关性状的杂种优势表型数据。

### (2) 高通量检测获得基因组数据、转录组数据和代谢组数据

对杂交种群体（由 *knr2*、*cct1*、*hsf21*、*dw3*、*gib2*、*srt2* 突变体材料和骨干自交系构建）所有亲本进行基因组重测序，获得基因组数据；选取杂种优势最显著的杂交组合，对这些杂交组合的亲本及杂交种进行转录组测序及代谢组检测，获得转录组数据和代谢组数据。

### (3) 组学生物信息学分析，挖掘杂种优势基因，解析杂种优势的分子机理

整合 14 份代表性的骨干自交系（B73、Mo17、郑 58、黄早四、178、昌 7-2、OH43、京 92、京 724、丹 340、A632、478、PH207、S37）的基因组、骨干自交系构建的双列杂交群体转录组、优异杂交群体代谢组和群体表型组信息，建立多组学数据库，开展 GWAS 分析、转录组关联分析（TWAS）和 eQTL 分析，挖掘杂种优势相关基因，并在多组学水平上探索杂种优势的分子机理。

## 课题 3：高产优质玉米种质创新与新品种培育

课题 3 负责人：李坤

参与人员：李高科、于永涛、黄君、李春艳、祁喜涛、肖颖妮、卢文佳、李武

主要研究内容如下：

### (1) 高产优质新种质的创制

对于已经挖掘的控制玉米株高（*ZmSPL12*、*ZmTCP4/14/25*、*Br2*）、生育期（*ZmDBB6*）、产量（*ZmKOB1*）和品质形成（*ZmACP1*、*ZmSK1*、*ZmCRR5*）的关键基因，本项目前期获得了大量变异位点的突变体，在对这些突变体精准鉴评的基础



上, 利用优异突变体转育到其他背景骨干自交系中 (*GQ1*, *Qun01X01* 等), 创制高产优质新种质资源。同时开发这些突变体的快速检测技术, 满足后续高通量检测需要。本项目利用项目组成熟的常规回交技术、单倍体技术, 快速获得包含优异变异的纯合种质, 进一步筛选和优化具有目标性状的种质材料。

### (2) 高配合力自交系筛选与组合鉴定

利用创制的优异种质进行大量杂交组配, 同时结合分子标记技术对突变体的组合形式 (杂合、纯合) 进行鉴定, 同时结合杂交种优势群, 对含有优异突变体的种质进行杂交, 组配优异组合, 考察不同杂交组合后代的表现形式, 对株高、产量、生育期、品质 (奎宁酸、果糖、蔗糖腺苷等) 等性状进行鉴定。筛选表现优异的组合及高配合力亲本, 进行知识产权保护和苗头组合的鉴定。

### (3) 新品种的培育与测试

将优异组合在不同生态区, 不同季节进行大量实验, 评估其产量、品质、抗逆性等性状表现, 通过多代选育和多点测试, 筛选出适应不同生态区域、具备高产优质特性的新品种, 将选育出的新品种进行登记、保护和推广。

#### 项目任务实施分工分述如下:

项目主持单位: 主要负责整个项目的协调、组织工作; 产量、株高和生育期等突变体材料创制; 参与建立杂种优势研究与利用多组学数据库 1 个并挖掘杂种优势基因 1 个; 申请植物新品种权保护证书 2-3 个, 审定玉米新品种 1 个。

项目参与单位 1: 主要负责创制高产、优质等重要性状玉米突变体材料 2-3 个; 申请植物新品种权保护证书 4-6 个, 审定玉米新品种 2 个。

项目参与单位 2, 创制高产、优质等重要性状玉米突变体材料 1-2 个; 主导建立杂种优势研究与利用多组学数据库 1 个; 挖掘杂种优势基因 2 个。

备注: 项目建设内容 (项目需求或项目建设任务) 按《项目申报书》内容填写。



### 三、项目绩效目标

主要说明项目实施后，预期达到的目标和产生的效果，相关表述应量化。

#### （一）总体目标

针对制约广东省玉米育种和产业化发展的瓶颈问题，围绕玉米株高、生育期、产量及品质等重要性状，在前期利用基因编辑技术创制突变体材料基础上，综合利用基因组、转录组、代谢组等多组学技术和相关分子遗传技术系统解析控制玉米株高、生育期、品质、产量及杂种优势等性状形成的遗传基础和分子网络，结合基因编码区及非编码区基因编辑技术和 IMGE 技术，创制目标性状突出且综合性状优良的玉米新种质，并培育高产优质的强优势玉米杂交种，为我省乃至全国玉米安全高效生产提供理论、基因及种质方面的支撑。

#### （二）总体考核指标

- （1）利用基因编辑技术创制高产、优质等重要性状玉米突变体材料 8-10 个；
- （2）建立杂种优势研究与利用的多组学数据库 1 个；
- （3）申请植物新品种权保护证书 6-8 个，成为生物育种的重要基础材料；
- （4）申请审定优异玉米新品种 3 个（已进入区域实验），具有商业化推广潜力。

#### （三）年度考核指标

##### 1. 2025 年绩效指标

- （1）创制株高、生育期、产量、品质等重要性状玉米突变体材料 4-5 个；
- （2）建立杂种优势研究与利用的多组学数据库 1 个；
- （3）培育出通过国家或者广东省审定的高产耐密、优质、玉米新品种 1-2 个以上；
- （4）申请新品种保护证书 3-4 个。

##### 2. 2026 年绩效指标

- （1）创制株高、生育期、产量、品质等重要性状玉米突变体材料 4-5 个；
- （2）培育出通过国家或者广东省审定的高产耐密、优质、玉米新品种 1-2 个。
- （3）申请新品种保护证书 3-4 个。

#### 四、项目进度安排

详细说明各阶段的工作内容和时间安排情况。

##### 2025 年:

(1) 对前期已经完成表型和基因型鉴定的约 30 份突变体材料进行新种质创制及完成部分种质的 DUS 测试;

(2) 挖掘产量和品质方面表现优异的突变体材料, 创制高产、优质等重要性状玉米突变体材料 4-5 个;

(3) 在 3 个地点种植玉米杂交种群体及其亲本, 在植物生长的不同阶段进行取样。对于少数重要材料及杂交种, 计划取样 4 次, 每次 3 次重复, 每份材料共取得 12 份样本, 对其开展转录组和代谢组检测。

(4) 利用杂交、回交、单倍体等技术方法, 考察突变体在不同遗传背景下的表型, 并配置杂交组合; 在不同的生长阶段收集杂交种群体及其亲本表型数据, 获得表型组、基因组、转录组和代谢组数据, 开展大规模生物信息学分析, 定位及挖掘 2 个杂种优势相关基因。

(5) 在多个测试开展新品种权的比较测试, 申请植物品种保护证 3-4 份;

##### 2026 年

(1) 纯化, 创制高产, 高品质突变体材料 1-2 份;

(2) 考察不同杂交组合在不同区域的产量和品质表现; 建立 1 个玉米杂种优势多组学数据库, 包括表型组、基因组、转录组、代谢组; 开展进行多组学分析, 解析杂种优势分子机理

(3) 对前期定位的杂种优势基因进行功能验证, 探讨关键基因的分子机理, 开发多个分子标记。

(4) 申请植物品种权保护证书 1-2 份, 申请审定优异玉米新品种 2 个;

(5) 整合研究结果, 完成项目结题;

备注: 项目绩效目标按《项目申报书》内容填写。



## 五、项目主要合作、参与单位(含牵头承担单位)

| 单位名称                                      | 单位性质 | 统一社会信用代码           | 通讯地址                 |
|---|------|--------------------|----------------------|
| 华南农业大学                                    | 高等院校 | 124400004554165634 | 广东省广州市天河区五山路483号     |
| 广东省农业科学院作物研究所                             | 科研院所 | 12440000455861755H | 广东省广州市天河区五山路金颖西二街18号 |
| 中国农业科学院深圳农业基因组研究所(岭南现代农业科学与技术广东省实验室深圳分中心) | 科研院所 | 12440300088261482M | 广东省深圳市大鹏新区鹏飞路7号      |

## 六、项目组主要成员(含项目负责人)

| 姓名  | 性别 | 身份证号               | 单位  | 职称/职务 | 电话          |
|-----|----|--------------------|---|-------|-------------|
| 王海洋 | 男  | 561962265          | 华南农业大学                                    | 正高级   | 13552603750 |
| 孔德鑫 | 男  | 413029198005052415 | 华南农业大学                                    | 正高级   | 13668985185 |
| 柳青  | 女  | 231102198610061424 | 华南农业大学                                    | 副高级   | 13155656503 |
| 张明月 | 女  | 210902198803041023 | 华南农业大学                                    | 副高级   | 15005167626 |
| 黄君  | 男  | 500381198601258015 | 华南农业大学                                    | 副高级   | 18819363706 |
| 李坤  | 男  | 341221199105064151 | 广东省农业科学院作物研究所                             | 中级    | 18621347676 |
| 李高科 | 男  | 511023197903262675 | 广东省农业科学院作物研究所                             | 正高级   | 13922215804 |
| 于永涛 | 男  | 132923197608162419 | 广东省农业科学院作物研究所                             | 正高级   | 18302096319 |
| 卢洪  | 男  | 420111197103185015 | 中国农业科学院深圳农业基因组研究所(岭南现代农业科学与技术广东省实验室深圳分中心) | 正高级   | 13910554761 |



|     |   |                    |   |     |             |
|-----|---|--------------------|---|-----|-------------|
| 杜国华 | 男 | 360731198512163833 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 中级  | 13530957241 |
| 马亚影 | 女 | 41140319920310634X | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 其他  | 15880035760 |
| 李春艳 | 女 | 41272419760710544X | 广东省农业科学<br>院作物研究所   | 副高级 | 13416486118 |
| 祁喜涛 | 男 | 41052619771127821X | 广东省农业科学<br>院作物研究所   | 副高级 | 13826259675 |
| 肖颖妮 | 女 | 362203198702066128 | 广东省农业科学<br>院作物研究所   | 副高级 | 18826067126 |
| 卢文佳 | 男 | 440106196712202093 | 广东省农业科学<br>院作物研究所   | 正高级 | 13503005389 |
| 李武  | 男 | 511221198105161812 | 广东省农业科学<br>院作物研究所   | 正高级 | 13602443815 |
| 张晓明 | 男 | 230204199205300016 | 华南农业大学  | 其他  | 18745683447 |
| 徐国强 | 男 | 340823199509091855 | 华南农业大学  | 其他  | 13922039061 |
| 姚旻昊 | 女 | 230202199408211029 | 华南农业大学  | 其他  | 18245299987 |
| 高永盟 | 男 | 341122199712063039 | 华南农业大学  | 其他  | 19864377065 |
| 赖晓君 | 女 | 440582199902263067 | 华南农业大学  | 其他  | 15362939350 |
| 郭小萍 | 女 | 511521199509212905 | 华南农业大学  | 其他  | 18408271754 |
| 余映宏 | 男 | 441621199910306438 | 华南农业大学  | 其他  | 13018611086 |
| 袁亚腾 | 男 | 410403200008265657 | 华南农业大学  | 其他  | 13071709350 |
| 尹翼亮 | 男 | 370982199906068035 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 其他  | 17346716036 |

|     |   |                    |   |     |             |
|-----|---|--------------------|---|-----|-------------|
| 何佳甜 | 男 | 142702199902022423 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 其他  | 13353596820 |
| 许靓蕊 | 女 | 142729199904033329 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 其他  | 18434889094 |
| 王沛东 | 男 | 211223198811231016 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 副高级 | 18742209823 |
| 李平  | 男 | 132628195910133711 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 其他  | 13691239434 |
| 徐淑环 | 女 | 132628196210013721 | 中国农业科学院<br>深圳农业基因组<br>研究所(岭南现代<br>农业科学与技术<br>广东省实验室深<br>圳分中心) | 其他  | 15233415473 |



## 七、资金使用预算

主要说明资金使用的范围或方向及资金使用进度安排（包含参与单位内容）。

### （一）省级财政资金预算主要用途项：

目经费 500.00 万元，包括间接经费 28.80 万元，用于以下三大课题的支出：

课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析（负责人：孔德鑫教授），220.00 万元，包括间接经费 13.20 万元，主要用于基因克隆及功能分析时所需材料费、测试化验加工费、燃料动力费、差旅费、出版/文献/信息传播/知识产权事务费、劳务费、专家咨询 费等支出；、

课题 2：玉米杂种优势多组学分析与杂种优势基因挖掘（负责人：卢洪研究员），120.00 万元，包括间接经费 6.00 万元，主要用于技术体系创建所需材料费、测试 化验加工费、燃料动力费、差旅费、出版/文献/信息传播/知识产权事务费、劳务费、专家咨询费等支出；

课题 3：高产优质玉米种质创新与新品种培育（负责人：李坤博士），160 万元，包括间接经费 9.60 万元，主要用于材料费、测试化验加工费、差旅费、 劳务费、材料种植、品种权登记、品种审定等。

### （一）直接费用

1.设备费（应当对仪器设备购置进行重点说明）经费预算 0 万元，占总经费预算 0 %。

### 2.材料费

经费预算 111.30 万元，占总经费预算 22.26 %。研究内容：

（1）名称：课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析；课题 2：玉米杂种优势多组学分析与杂种优势基因挖掘；课题 3：高产优质玉米种质创新与 新品种培育。

（2）与项目（课题）任务的相关性说明：应说明该项支出的主要用途，用于支撑 保障研究的哪部分内容，拟支出内容的必要性。

玉米株高、生育期和产量调控基因突变体创制实验；调控玉米品质性状的基因编 辑突变体创制研究所涉及的实验；玉米杂种群体及亲本大田表型鉴定实验；高产优质 新种质创制、高配合力自交系筛选与组合鉴定，新品种权和品种测试实验等实验所需 的材料费。

### （3）测算方法（依据）：

①采购各种工具酶，共 6.81 万元。包括：KOD FXNeo, 10 支×1700 元/个=1.70 万元；Phanta Max Super-Fidelity DNA Polymerase 等，均价 261 元/个，共 100 个，小计 2.61 万元，2×T5 Super PCR Mix (Colony)，单价 2500 元/盒，共 10 盒，小计 2.50 万元。

②采购各种分子生物学实验所需各种试剂盒，共计 71.61 万元。包括：TransStart Top Green qPCR SuperMix, 单价 7000 元/盒，共 6 盒，小计 4.20 万元，NEBNext Ultra II DNA 文库制备试剂盒，单价 5400，共 4 盒，小计 2.15 万元，2×T5 Super PCR Mix (Colony),单价 2500 元/包，共 20 包，



小计 5.00 万元, Anti-GFP 抗体(ab290), 单价 5137 元/管, 共 4 管, 小计 2.06 万元, Yeast Transformation System, 单价 4500 元/个, 共 4, 小计 1.80 万元, 荧光定量试剂盒, 单价 3000 元/盒, 50 盒, 小计 15.00 万元, DNA 提取试剂盒, 单价 500 元/盒, 200 盒, 小计 10.00 万元, RNA 提取试剂盒, 单价 2500 元/盒, 50 盒, 小计 12.50 万元, 双荧光素酶报告基因检测试剂盒, 750 元/套, 60 套, 小计 4.50 万元, 质粒提取试剂盒, 单价 200 元/盒, 300 盒, 小计 6.00 万元, EMSA 探针试剂盒, 单价 1500 元/盒, 40 盒, 小计 6.00 万元, 感受态, 240 元/包, 100 包, 小计 2.40 万元。

③采购各种化学试剂, 药品, 共计 17.08 万元。包括: X- $\alpha$ -gal, 单价 4224 元/袋, — 89 — 共 10 袋, 小计 4.22 万元, Aureobasidin A, 单价 4700 元/袋, 共 10 袋, 小计 4.70 万元, Tris, 单价 1500 元/瓶, 共 30 瓶, 小计 4.50 万元, 抗体 (Anti-GFP), 单价 2500 元, 10 支, 小计 2.50 万元。各种培养基, 小计 1.16 万元。

④采购各种农资耗材。共计 15.80 万元。化肥 200 元/袋, 250 袋, 小计 5.00 万元; 营养土 200 元/袋, 250 袋, 小计 5.00 万元; 农药 50 元/瓶, 1000 瓶, 小计 5.00 万元; 纱网袋, 牛皮纸袋插牌, 标签, 挂牌, 2 批, 约 4000 元/批, 小计 0.80 万元。

### 3.测试化验加工费

经费预算 161.00 万元, 占总经费预算 32.20 %。研究内容:

(1) 名称: 课题 1: 玉米重要性状基因编辑突变体材料创制与多组学分析; 课题 2: 玉米杂种优势多组学分析与杂种优势基因挖掘; 课题 3: 高产优质玉米种质创新与新品种培育。

(2) 与项目(课题)任务的相关性说明: 应说明该项支出的主要用途, 用于支撑保障研究的哪部分内容, 拟支出内容的必要性。控制玉米株高、生育期和产量、品质性状等关键基因的分子调控机制研究所涉及各种测序、遗传转化、转录组、代谢组、蛋白质组等多组学实验; 玉米杂种优势分析与杂种优势基因挖掘研究所涉及的基因组、转录组、代谢组等实验。

(3) 测算方法(依据): 转录组测序及分析, 单价 600 元/, 共 800 个, 小计 48.00 万元。基因组 DNA 建库测序及分析, 单价 100000 元/例, 共 3 例, 小计 30.00 万元。玉米遗传转化, 单价 15000 元/, 共 17 个, 小计 25.5 万元。引物合成, 单价 7500 元, 共 20 条, 小计 15.00 万元。DNA 测序, 单价 8500 元, 共 20 次, 小计 17.00 万元。激光共聚焦观察, 单价 200 元/小时, 共 135 小时, 小计 2.70 万元。酵母文库构建, 单价 30000 元/个, 共 1 个, 小计 3.00 万元。代谢组检测, 单价 5000 元, 共 30, 小计 15.00 万元。DAP-seq 测序及分析, 单价 1000 元, 共 8 个, 小计 0.80 万元。植物激素测定, 单价 2000 元, 共 20 个, 小计 4.00 万元。

### 4.燃料动力费

经费预算 4.00 万元, 占总经费预算 0.8 %。研究内容:

(1) 名称: 玉米重要性状基因编辑突变体材料创制与多组学分析; 课题 2: 玉米杂种优势多组



学分析与杂种优势基因挖掘；课题 3：高产优质玉米种质创新与新品种 培育。

(2) 与项目(课题)任务的相关性说明：应说明该项支出的主要用途，用于支撑 保障研究的哪部分内容，拟支出内容的必要性。玉米各种种质材料、种植所需的田间管理的机器设备所需电费，实验室研究使用 的超速离心机、制冰机、烘箱、冰箱等需要的电费。

(3) 测算方法(依据)：超低温冰箱，单价 10 元/天，共 400，小计 0.40 万元。每年秸秆粉碎、耕地、耘田、打药、收获及运输使用田间农机具预计支出燃油费，单价 300 元/50L,共 300，小计 3.00 万元。用于试验研究及播种前催芽及收获后烘干等使用的超速离心机、制冰机、烘箱、催芽机、种子烘干机及其他实验设备运行预计耗电，单价 3000 元/年，共 2，小计 0.60 万元。

#### 5.出版/文献/信息传播/知识产权事务费

经费预算 26.75 万元，占总经费预算 5.35 %。研究内容：

(1) 名称：课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析；课题 2：玉米杂种优势多组学分析与杂种优势基因挖掘；课题 3：高产优质玉米种质创新与 新品种培育。

(2) 与项目(课题)任务的相关性说明：应说明该项支出的主要用途，用于支撑 保障研究的哪部分内容，拟支出内容的必要性。玉米株高、生育期、产量及品质等关键基因突变体及多组学分析研究，杂种优势 基因挖掘等成果涉及的专利申请费、论文发表。高产优质新种质创制，新品种权登记 及新品种审定所需的费用。

(3) 测算方法(依据)：论文版面费，单价 18000 元/篇，预计 5 篇，小计 9.00 万元。专利代理费(国内)，单价 6000 元/件，预计 5 件，小计 3.00 万元。新品种审定，单价 30000 元/个，预计 3 个，小计 9.00 万元。 新品种权，单价 7000 元/个，预计 7 个，小计 4.90 万元。邮寄项目研究相关的种子、材料、文件等支出小计 0.85 万元。

6.会议/差旅/国际合作与交流费(科研人员结合科研活动实际需要编制预算并按 规定统筹安排使用，其中不超过直接费用 10%的，不需要提供预算测算依据。超过直接费用 10%的，应说明测算过程和依据) 经费预算 26.85 万元，占总经费预算 5.37 %。研究内容：

(1) 名称：课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析；课题 2：玉米杂种优势多组学分析与杂种优势基因挖掘；课题 3：高产优质玉米种质创新与 新品种培育。

(2) 与项目(课题)任务的相关性说明：应说明该项支出的主要用途，用于支撑 保障研究的哪部分内容，拟支出内容的必要性。玉米株高、生育期、产量及品质等关键基因突变体、杂种优势群体材料、高产优 质新种质，新品种测试等材料基地种植、表型鉴定、授粉等所需差旅，课题相关的学 术交流会等。

(3) 测算方法(依据)：

①课题 1：玉米重要性状基因编辑突变体材料创制与多组学分析(10.2 万元)；海南，廊坊基地差旅，4000 元/人次，12 人次/年，2 年共 4.80 万元；住宿费 1000 元/次，共 10 次，2 年共 2 万



元, 市内交通, 5000 元/年, 2 年共 1.00 万元。种植基地运送材料, 3000 元, 8 次, 2 年共 2.4 万元

②课题 2: 玉米杂种优势多组学分析与杂种优势基因挖掘 (7.65 万元) 海南, 惠州, 肇庆差旅, 4000 元/人次, 5 人次/年, 2 年共 4.00 万元; 住宿费 400 元/次, 共 6 次, 2 年共 0.24 万元, 餐费及交通补贴 180 元/次, 共 20 人, 2 年共 0.36 万元。市内交通, 5000 元/年, 2 年共 1.00 万元。种植基地运送材料, 3000 元, 4 次, 2 年共 1.20 万元。国内会议注册 2000 元, 2 人, 共 0.40 万元, 住宿费 450 元/天, 共 0.45 万元

③课题 3: 名称: 高产优质玉米种质创新与新品种培育 (9.00 万) 甘肃, 海南, 惠州, 肇庆调查取样差旅, 4000 元/人次, 7 人次/年, 2 年共 2.80 万元; 餐费及交通补贴 180 元/次, 共 25 人, 2 年共 0.45 万元。市内交通, 5000 元/年, 2 年共 0.50 万元。种植基地运送材料, 3000 元, 8 次, 2 年共 2.40 万元 国内会议注册 2000 元, 5 人, 共 1.00 万元, 机票费 3500 元/人, 4 人, 共 1.40 万元, 住宿费 450 元/天, 10 天, 共 0.45 万。

7.培训费 经费预算 0 万元, 占总经费预算 0%。研究内容: 无

8.劳务费 (劳务费预算无比例限制。参与项目 (课题) 实施的研究生、博士后、访问学者以及项目 (课题) 聘用的人员、科研辅助人员等, 均可开支劳务费。项目 (课题) 聘用人员的劳务费开支标准, 参照当地科学研究和技术服务业从业人员平均工资水平, 根据其在项目 (课题) 实施中承担的工作任务确定, 其社会保险补助纳入劳务费科目中列支)。

经费预算 119.10 万元, 占总经费预算 23.82%。研究内容:

(1) 名称: 课题 1: 玉米重要性状基因编辑突变体材料创制与多组学分析; 课题 2: 玉米杂种优势多组学分析与杂种优势基因挖掘; 课题 3: 高产优质玉米种质创新与新品种培育。

(2) 与项目 (课题) 任务的相关性说明: 应说明该项支出的主要用途, 用于支撑保障研究的哪部分内容, 拟支出内容的必要性。玉米株高、生育期、产量及品质等关键基因突变体、杂种优势群体材料、高产优质新种质, 新品种测试等材料基地种植、表型鉴定、授粉等所需差旅, 课题相关的人工费, 研究生三助补贴。

(3)测算方法 (依据): 参与项目博士生 10 人×2000 元×8 月=16.00 万元, 硕士 30 人×1500 元×8 月=36.00 万元, 博士后 5 人×10000 元×8 月=40.00 万元, 合同工 5 人×6000 元×8 月=24.00 万元, 临时工 5 人×775 元×8 月=3.10 万元, 劳务费共计 119.10 万元。

9.专家咨询费 经费预算 1.70 万元, 占总经费预算 0.34%。研究内容:

(1) 名称: 课题 1: 玉米重要性状基因编辑突变体材料创制与多组学分析; 课题 2: 玉米杂种优势多组学分析与杂种优势基因挖掘; 课题 3: 高产优质玉米种质创新与新品种培育。

(2) 与项目 (课题) 任务的相关性说明: 应说明该项支出的主要用途, 用于支撑保障研究的哪部分内容, 拟支出内容的必要性。玉米株高、生育期、产量及品质等关键基因突变体、杂种优势群体

材料、高产优质新种质，新品种测试等材料基地种植、表型鉴定、授粉等所需差旅，课题相关的学术交流及专家咨询等。

(3) 测算方法(依据): 年度汇报 1 次, 中期检查 1 次, 验收 1 次, 每次 5 人, 1000 元/人, 共 1.50 万元; 研究期间专家学术交流, 项目咨询, 1 人次, 2000 元/人, 共 0.20 万元。

10. 基本建设费 经费预算 0 万元, 占总经费预算 0% 研究内容: 无

11. 其他费用:

(1) 名称: 课题 1: 玉米重要性状基因编辑突变体材料创制与多组学分析; 课题 2: 玉米杂种优势多组学分析与杂种优势基因挖掘; 课题 3: 高产优质玉米种质创新与新品种培育。

(2) 与项目(课题)任务的相关性说明: 应说明该项支出的主要用途, 用于支撑保障研究的哪部分内容, 拟支出内容的必要性。项目结题时需委托会计事务所进行审计; 承担该项目的其他设备(非从此项目购置的)如超低温冰箱、电击仪、灭菌锅等, 会损耗, 产生维修费用, 实验材料种植所需的地租等。

(3) 测算方法(依据): 用于实验室种植实验材料用地,  $2500 \text{ 元/年} \times 50 \text{ 亩} = 12.5 \text{ 万元}$ , 其他维修  $5000 \text{ 元} \times 16 \text{ 次} = 8 \text{ 万元}$ ; 共 20.5 万元。

(二) 间接费用 经费预算 28.8 万元, 占总经费预算 5.76%。项目承担单位按到位经费 6%提取管理费。



## 八、保障措施（说明围绕完成项目任务、目标所要采取的具体措施）。

（1）本项目的牵头单位为华南农业大学，项目负责人是华南农业大学王海洋教授。项目负责人将认真落实项目执行方的主体责任，对项目总体负责，统筹本项目的方案制定及具体实施。

（2）项目负责人根据项目任务、总目标及任务特点，采用项目、课题和子任务三级分层管理机制，将项目目标合理分配到各参与骨干人员，形成课题负责人、课题组长和子任务承担人分级负责的内部组织管理方式。通过密切协作，充分发挥课题团队的各自优势，定期开展合作交流，提升团队科研水平。

（3）项目组定期对建设进展开展自查，并接受专家委员会的定期考核，评价本项目所采用的各项技术方案和研究方法的可行性、督查项目实施和研究进度合理性，对查出的问题及时纠正。

（4）项目负责人根据任务承担人的承担工作量和绩效指标的情况，合理分配费用。按“花钱必问效，无效必问责”的原则，实行资助经费与科研产出挂钩，建立年度考核、以评促研的运行机制。同时在本课题组内部设立财务管理小组，负责专项经费的预算、使用及全程监控。

（5）项目运行采用目标考核制，通过年度检查、中期检查等方式监督研究内容落实情况。项目每年至少召开一次项目进展年会或学术交流会，由聘期的专家组对各课题和各子任务的研究进展进行指导和评价。同时，鼓励课题内部充分利用视频会议方式，在项目实施过程中充分交流科研进展，鼓励研究材料、数据和研究结果在项目内及时共享。

（6）严格财会制度，加强对项目资金的管理，实行专人、专账管理，保证专款专用。资金使用方面，严格按计划开支。根据本项目实施计划，对于需拨付项目区的资金和物资，按实施进度由主持单位拨付项目区，并取回凭证，以备审计。同时加强追踪管理，督促基层单位按计划使用资金。

## 检索证明

根据委托人提供的论文材料, 委托人华南农业大学生命科学学院 Mingyue Zhang 10 篇论文收录情况如下表。

| 序号 | 论文名称  | 发表刊物及发表的年月卷期/页码等   | 作者排名   | 论文等级 | 作者文中单位 | 收录情况 | 影响因子   | 中科院大类分区                          | 引用                 |
|----|---|--|--------|------|--------|------|--|----------------------------------|--------------------|
| 1  | Genome-wide association studies provide insights into the genetic determination of fruit traits of pear | NATURE COMMUNICATIONS<br>出版年: 2021<br>出版日期: FEB 18<br>卷期: 12 1 页码: -<br>文献号: 1144<br>文献类型: Article | 第一作者   | T2类  | 南京农业大学 | SCI  | IF2-year=17.694<br>IF5-year=17.764<br>(2021) | 综合性期刊 1 区<br>Top 期刊: 是<br>(2021) | SCI 核心合集<br>总引: 72 |
| 2  | The pear genomics database (PGDB): a comprehensive multi-omics research platform for Pyrus spp.         | BMC PLANT BIOLOGY<br>出版年: 2023<br>出版日期: SEP 16<br>卷期: 23 1 页码: -<br>文献号: 430<br>文献类型: Article      | 共同通讯作者 | A类   | 山东农业大学 | SCI  | IF2-year=4.3<br>IF5-year=5.2<br>(2023)       | 生物学 2 区<br>Top 期刊: 是<br>(2023)   | SCI 核心合集<br>总引: 3  |
| 3  | Rearrangement and domestication as drivers of Rosaceae mitogenome plasticity                            | BMC BIOLOGY<br>出版年: 2022<br>出版日期: AUG 19<br>卷期: 20 1 页码: -<br>文献号: 181<br>文献类型: Article            | 并列第一作者 | A类   | 南京农业大学 | SCI  | IF2-year=5.4<br>IF5-year=7.1<br>(2022)       | 生物学 2 区<br>Top 期刊: 否<br>(2022)   | SCI 核心合集<br>总引: 38 |



|   |  |  |        |      |        |  |  |                                 |                    |
|---|--|--|--------|------|--------|--|--|---------------------------------|--------------------|
| 4 | Pear genetics: Recent advances, new prospects, and a roadmap for the future  | HORTICULTURE RESEARCH<br>出版年: 2022<br>出版日期: JAN 5<br>卷期: 9 页码: -<br>文献号: uhab040<br>文献类型: Review | 并列第一作者 | T2 类 | 南京农业大学 |  | IF2-year=8.7<br>IF5-year=9.0<br>(2022)     | 农林科学 1 区<br>Top 期刊: 是<br>(2022) | SCI 核心合集<br>总引: 38 |
| 5 | Genomic selection of eight fruit traits in pear  | HORTICULTURAL PLANT JOURNAL<br>出版年: 2024<br>出版日期: MAR<br>卷期: 10 2 页码: 318-326<br>文献类型: Article   | 并列第一作者 | T2 类 | 南京农业大学 |  | IF2-year=6.2<br>IF5-year=6.1<br>(2024)     | 农林科学 1 区<br>Top 期刊: 是<br>(2025) | SCI 核心合集<br>总引: 8  |
| 6 | Comparison of multiple algorithms to reliably detect structural variants in pears  | BMC GENOMICS<br>出版年: 2020<br>出版日期: JAN 20<br>卷期: 21 1 页码: -<br>文献号: 61<br>文献类型: Article          | 并列第一作者 | A 类  | 南京农业大学 |  | IF2-year=3.969<br>IF5-year=4.478<br>(2020) | 生物学 2 区<br>Top 期刊: 是<br>(2020)  | SCI 核心合集<br>总引: 16 |
| 7 | Contrasting genetic variation and positive selection followed the divergence of NBS-encoding genes in Asian and European pears | BMC GENOMICS<br>出版年: 2020<br>出版日期: NOV 19<br>卷期: 21 1 页码: -<br>文献号: 809                          | 并列第一作者 | A 类  | 南京农业大学 |  | IF2-year=3.969<br>IF5-year=4.478<br>(2020) | 生物学 2 区<br>Top 期刊: 是<br>(2020)  | SCI 核心合集<br>总引: 11 |

|    |   |  |        |     |        |     |  |                                |                    |  |  |  |
|----|---|--|--------|-----|--------|-----|--|--------------------------------|--------------------|--|--|--|
|    | 文献类型: Article   |  |        |     |        |     |  |                                |                    |  |  |  |
| 8  | Natural variations in the PbCPR28 promoter regulate sugar content through interaction with PbTST4 and PbVHA-A1 in pear                              | PLANT JOURNAL<br>出版年: 2023<br>出版日期: APR<br>卷期: 114 1 页码: 124-141<br>文献类型: Article                  | 并列第一作者 | T2类 | 南京农业大学 | SCI | IF2-year=6.2<br>IF5-year=7.2<br>(2023)     | 生物学 1 区<br>Top 期刊: 是<br>(2023) | SCI 核心合集<br>总引: 11 |  |  |  |
| 9  | Genome-wide genetic diversity and IBD analysis reveals historic dissemination routes of pear in China   | TREE GENETICS & GENOMES<br>出版年: 2022<br>出版日期: FEB<br>卷期: 18 1 页码: 145<br>文献号: 145<br>文献类型: Article | 并列第一作者 | B类  | 南京农业大学 | SCI | IF2-year=2.4<br>IF5-year=2.4<br>(2022)     | 生物学 3 区<br>Top 期刊: 否<br>(2022) | SCI 核心合集<br>总引: 7  |  |  |  |
| 10 | Identification of key genes related to seedlessness by genome-wide detection of structural variation and transcriptome analysis in 'Shijiwuhe' pear | GENE<br>出版年: 2020<br>出版日期: MAY 15<br>卷期: 738 页码: -<br>文献号: 144480<br>文献类型: Article                 | 并列第一作者 | B类  | 南京农业大学 | SCI | IF2-year=3.688<br>IF5-year=3.329<br>(2020) | 生物学 3 区<br>Top 期刊: 否<br>(2020) | SCI 核心合集<br>总引: 3  |  |  |  |

说明: 论文等级和中科院大类分区按《华南农业大学学术论文评价方案(试行)》划分。

报告免责声明:如未盖章,报告无效

华南农业大学图书馆SCAULIB202519125





ARTICLE



<https://doi.org/10.1038/s41467-021-21378-y>

OPEN

# Genome-wide association studies provide insights into the genetic determination of fruit traits of pear

Ming-Yue Zhang<sup>1,7</sup>, Cheng Xue<sup>1,2,7</sup>, Hongju Hu<sup>3,7</sup>, Jiaming Li<sup>1,7</sup>, Yongsong Xue<sup>1,7</sup>, Runze Wang<sup>1</sup>, Jing Fan<sup>3</sup>, Cheng Zou<sup>4</sup>, Shutian Tao<sup>1</sup>, Mengfan Qin<sup>1</sup>, Bing Bai<sup>1</sup>, Xiaolong Li<sup>1</sup>, Chao Gu<sup>1</sup>, Shan Wu<sup>5</sup>, Xu Chen<sup>2</sup>, Guangyan Yang<sup>1</sup>, Yueyuan Liu<sup>1</sup>, Manyi Sun<sup>1</sup>, Zhangjun Fei<sup>5,6</sup>✉, Shaoling Zhang<sup>1</sup>✉ & Jun Wu<sup>1</sup>✉

Pear is a major fruit tree crop distributed worldwide, yet its breeding is a very time-consuming process. To facilitate molecular breeding and gene identification, here we have performed genome-wide association studies (GWAS) on eleven fruit traits. We identify 37 loci associated with eight fruit quality traits and five loci associated with three fruit phenological traits. Scans for selective sweeps indicate that traits including fruit stone cell content, organic acid and sugar contents might have been under continuous selection during breeding improvement. One candidate gene, *PbrSTONE*, identified in GWAS, has been functionally verified to be involved in the regulation of stone cell formation, one of the most important fruit quality traits in pear. Our study provides insights into the complex fruit related biology and identifies genes controlling important traits in pear through GWAS, which extends the genetic resources and basis for facilitating molecular breeding in perennial trees.

<sup>1</sup>Centre of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, Jiangsu, China. <sup>2</sup>Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, China. <sup>3</sup>Wuchang Sand Pear Germplasm, Hubei Academy of Agricultural Sciences, Wuhan, Hubei, China. <sup>4</sup>Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. <sup>5</sup>Boyce Thompson Institute, Cornell University, Ithaca, NY, USA. <sup>6</sup>USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA. <sup>7</sup>These authors contributed equally: Ming-Yue Zhang, Cheng Xue, Hongju Hu, Jiaming Li, Yongsong Xue. ✉email: [zf25@cornell.edu](mailto:zf25@cornell.edu); [slzhang@njau.edu.cn](mailto:slzhang@njau.edu.cn); [wujun@njau.edu.cn](mailto:wujun@njau.edu.cn)

Pear has been cultivated by human for at least 3000 years<sup>1</sup>. However, traditional breeding procedure of pear normally takes at least 13–15 years because of its long juvenile stage. Molecular breeding is an efficient way to help improve cultivars using pre-selection from large numbers of diverse individuals to identify and introduce precise genetic regions that can be expected to confer desired phenotypic benefits. For example, previous work identified 31 quantitative trait loci (QTLs) for 11 traits based on a pear linkage map constructed with 3241 markers<sup>2</sup>. Amplified fragment length polymorphism (AFLP) markers were identified to be linked to a scab resistance gene in an inter-specific hybrid pear<sup>3</sup>, and fire-blight resistance-related QTLs were identified in European pears<sup>4,5</sup>. Pear harvest time and skin color-related QTLs were also investigated in Japanese pears<sup>6</sup>. However, pear research and breeding have been limited by factors such as the time-consuming process of constructing cross populations for genetic mapping, and restricted genetic diversity.

With the completion of pear whole-genome sequencing, analyzing large natural populations through high-throughput sequencing and genome-wide association studies (GWAS) provides tremendous opportunities for exploring variation alleles and key genes underlying important traits, which can help to develop molecular breeding markers and reveal genetic mechanisms of these traits. In plants, GWAS was initially performed in cereal crops, such as maize, for which flowering time was found to be associated with *Dwarf8* polymorphisms<sup>7</sup>. In rice, GWAS was performed for 14 traits using a population consisting of 517 landraces with 3.6 million single nucleotide polymorphisms (SNPs)<sup>8</sup>. GWAS was also performed in rice to rapidly identify genes influencing agronomic traits using the gene-based association analysis<sup>9</sup>. Despite the wide applications of GWAS in cereal and other annual crops, only a few GWAS investigations have been conducted for perennial fruit trees till now. One addressed 129 peach accessions and explored 12 fruit and flower-related traits<sup>10</sup>. A second study explored pathogen resistance in apricot<sup>11</sup>. Recently, a 200K array was developed to facilitate GWAS of two traits in pear<sup>12</sup>. Therefore, additional GWAS with large population size, wide genetic background, and high-density SNPs are necessary to better understand the genetic architecture of fruit quality and phenological traits for fruit tree crops.

The skin color and the contents of sugars, acids, and stone cells are important pear fruit quality traits and also the main targets for pear fruit improvement. However, studies of fruit quality traits in pears are relatively limited compared with other fruit trees such as apple and grape, with only a few reports on pear fruit skin color<sup>13</sup>, sugar<sup>14</sup>, organic acid<sup>15</sup>, and stone cells<sup>16,17</sup>. Pear fruits have special brachysclereid cells called stone cells with secondary-thickened cell walls that are formed from parenchyma cells via the deposition of lignins and celluloses<sup>18</sup>. Recently it was determined that the lignin forms present in pear stone cells are primarily guaiacyl-lignin (G-lignin) and syringyl-lignin (S-lignin) monomers<sup>19</sup>.

In this study, we carry out whole-genome resequencing with an average of around ten-fold coverage on an association panel comprising 312 sand pear (*Pyrus pyrifolia*) accessions collected from China, Korea, and Japan, and revealed their population structure. Meanwhile, landrace and cultivated sand pears are explored to identify selective sweep signatures in the genome during pear improvement. We then perform GWAS of 11 agronomic traits for which phenotype data are collected over three successive years, to identify the associated genome loci and candidate functional genes. A gene involved in regulating lignin contents is identified to function in stone cell formation in pear fruits, and it is verified in both transgenic pear and *Arabidopsis*. Genome loci associated with agronomic traits of pear and the identified gene involved in pear stone cell formation provide

valuable information for fruit trees breeding, and for deeper understanding of the fruit biology.

## Results and discussion

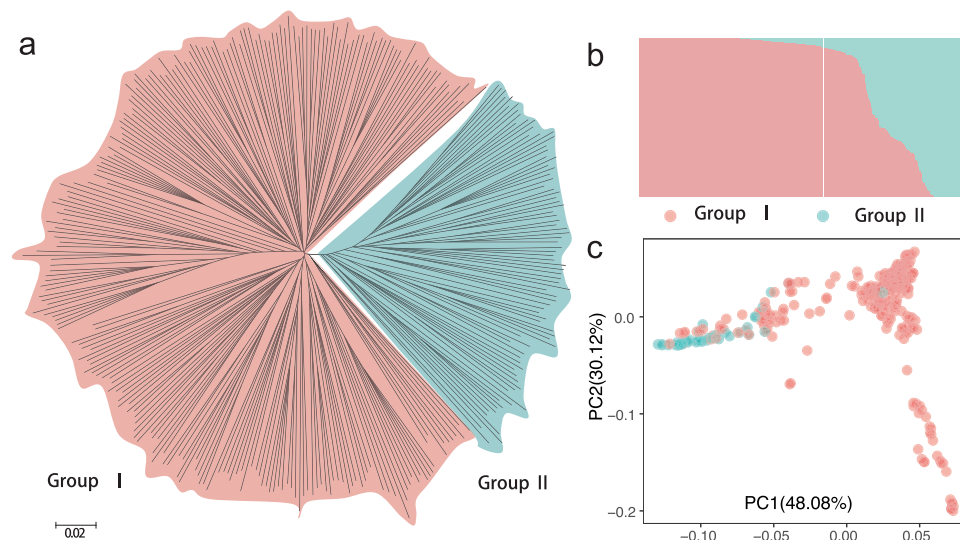
### Sequencing, variants, and population structure of sand pears.

We analyzed a germplasm diversity panel comprising 312 sand pears (*P. pyrifolia*), of which 231 were landraces and 81 were improved pear cultivars (Supplementary Data 1). Genomes of these accessions were sequenced using the Illumina technology, and a total of 2.15 Tb of raw sequence data were obtained. After removing low-quality and adapter sequences, an average of about ten-fold coverage data were obtained for each accession and used for SNP calling. The average mapping rate of the cleaned reads to the reference genome of *Pyrus bretschneideri*<sup>20</sup> was 73.5% (Supplementary Data 1), higher than that in a previous study of 113 pears from different species<sup>21</sup>.

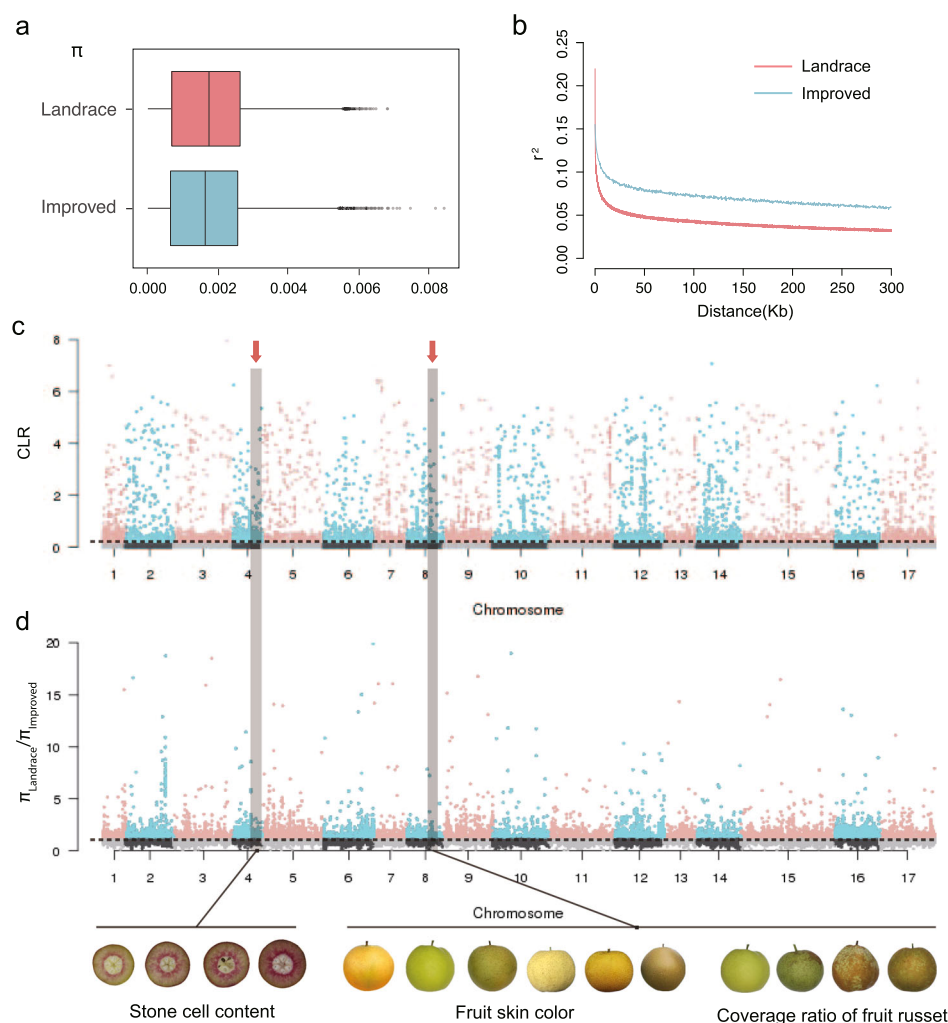
A total of 3.40 million SNPs with missing rate of no more than 30% and a minor allele frequency (MAF)  $\geq 0.03$  were obtained and used for downstream analyses. We randomly selected 3942 SNP loci for Sanger sequencing, which revealed a high accuracy of the called SNPs (97.6%; Supplementary Data 2). Using these high-quality SNPs, we first inferred phylogenetic relationships of the sand pears, which suggested that the sand pears we analyzed were divided into two major subpopulations, one comprising accessions from China (Group I) and the other comprising accessions from Japan and the Korean peninsula (Group II) (Fig. 1a). This was largely consistent with the results obtained from the population structure and  $F_{ST}$  analysis (Fig. 1b and Supplementary Figs. 1–3) and the principal component analysis (PCA) (Fig. 1c).

**Selective sweep signals during pear improvement.** The nucleotide diversity ( $\pi$ ) of the entire sand pear population here was  $1.73 \times 10^{-3}$ , and that of the landrace ( $1.75 \times 10^{-3}$ ) was higher than that of the improved pears ( $1.70 \times 10^{-3}$ ) (Fig. 2a). As expected, the diversity level we detected here for sand pear was substantially lower compared to that ( $5.5 \times 10^{-3}$ ) of combined Asian and European cultivated and wild pears reported previously<sup>21</sup>. This previous study<sup>21</sup> indicated that similar diversity was preserved in pears during the domestication process, which is different from the strong selection in annual crops, and during the improvement process, the diversity reduction in pear was slightly lower than that in annual crops such as soybean and rice<sup>22,23</sup>. At the transcriptome level,  $\pi$  of pear landrace group was also higher than that in the improved group, but  $\pi$  remains the same level from wild group to landrace group<sup>24</sup>. The linkage disequilibrium (LD) decay of the improved pears was less rapid than that of landraces (Fig. 2b). What was more, the LD decay in pears was much more rapid than that in annual crops like rice<sup>23</sup> and soybean<sup>25</sup>, but similar to that in apple<sup>26</sup>, mainly due to the long generation time and self-incompatibility.

We identified candidate selective sweeps during pear improvement based on composite likelihood rates and nucleotide diversity throughout the genome for all of the accessions in the germplasm diversity panel of sand pears, which in total covered 11.1 Mb (2.89%; Supplementary Data 3) of the pear genome (Fig. 2c) and harbored 1417 genes (Supplementary Data 4), of which some were known to be related to metabolisms of sugars, organic acids, amino acids, lignin, as well as resistance and photosynthesis. We previously found that during the domestication process, genes related to metabolisms of sugars, organic acids, lignin were also under selection<sup>21</sup>. However, the fruit size-related genes were enriched in domestication sweeps<sup>21</sup> while no such genes were found in selective sweep regions during pear improvement. We conclude that traits of lignin, sugar, and acid contents were under



**Fig. 1 Phylogenetic relationship and population structure of 312 sand pears.** **a** Phylogenetic tree of sand pears. **b** Population structure of sand pears.  $K$  value is 2. **c** PCA of sand pears. Group I with red background corresponds to the Chinese group, and Group II with blue background indicates the Japanese and Korea group.



**Fig. 2 Diversity and selective sweep of landrace and improved population.** **a** Nucleotide diversity ( $\pi$ ) of sand pears in landrace ( $n = 231$ ) and improved ( $n = 81$ ) groups. For each box plot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median. The whisker represents  $1.5 \times$  interquartile range of the lower or upper quartile. **b** LD in the landrace and improved groups. **c** Selective sweep regions during sand pear improvement (landrace vs improved groups) detected by CLR. The horizontal dashed line refers to the top 5% threshold of CLR scores. **d** Selective sweep regions during sand pear improvement (landrace vs improved groups) by  $\pi$  ratios. The horizontal dashed line refers to the top 15% threshold of  $\pi_{\text{Landrace}}/\pi_{\text{Improved}}$ . Fruit quality-related traits, stone cell content, coverage ratio of fruit russet, and fruit skin color, are demonstrated in the combined selective sweep regions.



continuous selection during both domestication and improvement. It was also reported that lignin biosynthesis-related genes were in the selective sweep signals during domestication and improvement at the transcriptome level<sup>24</sup>.

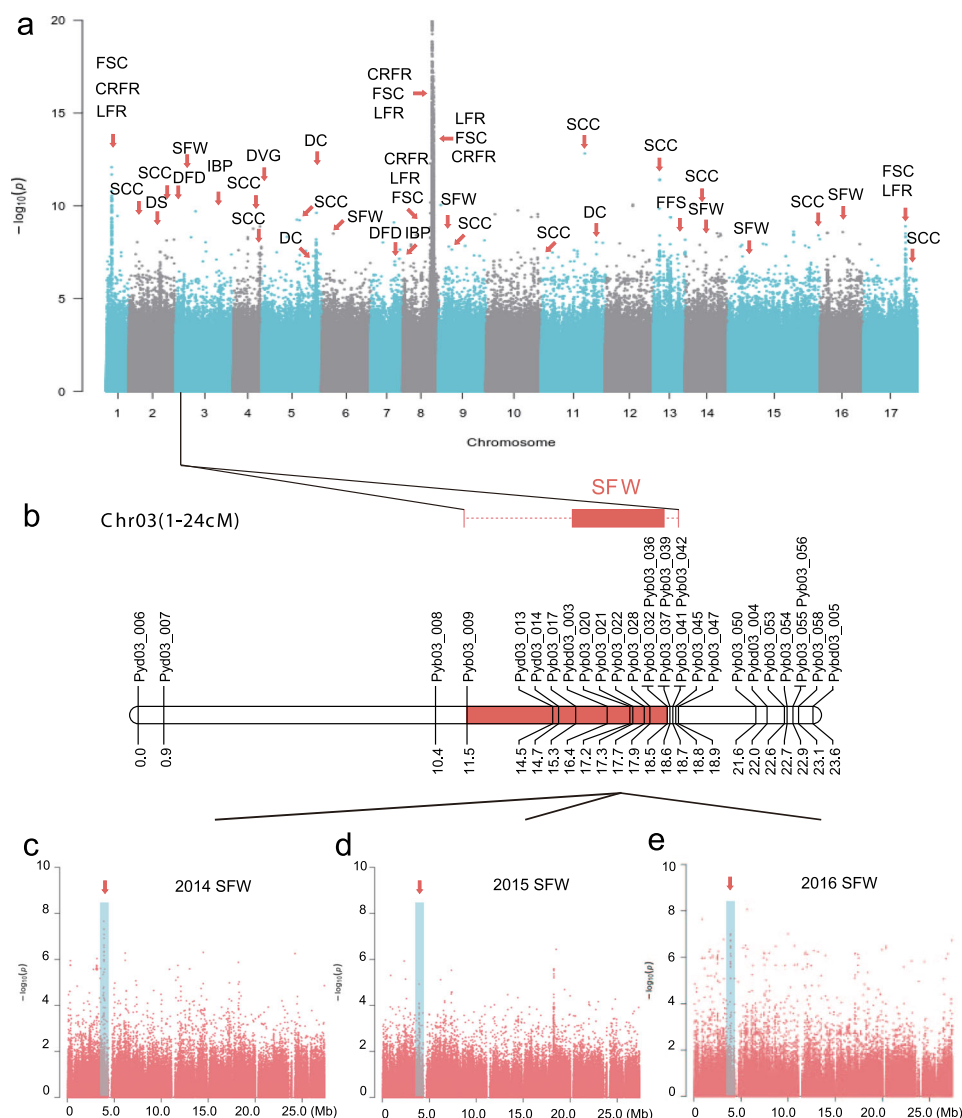
**GWAS of fruit quality and phenological traits.** A total of 11 agronomic traits were investigated in the 312 sand pears for three consecutive years (Supplementary Data 5 and Supplementary Fig. 4), including three fruit phenological traits (initial bloom period, days of fruit development, and days of vegetative growth) and eight fruit quality traits (single fruit weight, stone cell content, fruit skin color, coverage ratio of fruit russet, location of fruit russet, furrows on fruit surface, direction of carpogonium, and direction of sepal). We identified five candidate loci for the three phenological traits of sand pears and 37 candidate loci for the eight fruit quality traits (Fig. 3a, Supplementary Data 6, 7, and Supplementary Figs. 5–15).

Single fruit weight is very important for fruit production, and our GWAS identified five lead associated SNPs with this trait. Of

particular note, one of these candidate SNPs located within a region on chromosome 3 which was previously reported to have a high LOD score for single fruit weight in a QTL study<sup>2</sup> (Fig. 3b–e). Another candidate SNP (Chr9\_2479073) is located within a gene (*Pbr001726*) encoding a stachyose synthase precursor, suggesting a plausible inference about the impact of this gene on starch accumulation and thus single fruit weight.

Fruit skin color has been a major focus in modern breeding programs for many fruit tree species. Our GWAS results revealed three candidate loci for this trait, and we also found that the skin color-associated loci were in regions under selection during the improvement process for sand pears (Figs. 2c, d and 3a). It was notable that eight of the candidate genes for this trait were annotated as *MYB* transcription factors, which have been widely reported to regulate fruit skin and flesh colors such as in apple<sup>27</sup>, strawberry<sup>28</sup>, kiwifruit<sup>29</sup>, and tomato<sup>30</sup>.

The furrows on fruit surface affect fruit appearance, but the molecular basis of furrow development remains poorly understood. Our GWAS of furrows on fruit surface trait identified one lead associated SNP on chromosome 13 (Chr13 11586174)



**Fig. 3 GWAS of 11 traits in 312 sand pears.** **a** Manhattan plot of GWAS for 11 traits in sand pears. IBP initial bloom period, DFD days of fruit development, DVG days of vegetative growth, FSC fruit skin color, CRFR coverage ratio of fruit russet, LFR location of fruit russet, FFS furrows on fruit surface, DC direction of carpododium, DS direction of sepal, SFW single fruit weight, SCC stone cell content. **b** QTL for trait of single fruit weight in the population of 'Bayuehong' × 'Dangshansuli'. **c–e** Manhattan plots of chromosome 3 of GWAS for trait of single fruit weight in the years 2014 (**c**), 2015 (**d**), and 2016 (**e**).

(Fig. 3a and Supplementary Data 7). Furthermore, our GWAS identified one SNP (Chr8\_14112990) that was significantly associated with both the coverage ratio of fruit russet and the location of fruit russet, and the associated candidate gene encoded an endoglucanase (Fig. 3a and Supplementary Data 7). The candidate region of coverage ratio of fruit russet and the location of fruit russet was in the selective region, suggesting that these two fruit russet traits could have been under selection during sand pear improvement.

The phenological trait, days of fruit development, dramatically affects fruit maturation, and the appropriate regulation of the pear fruit maturation period could substantially contribute to efficient shipping and storage of pears. Two candidate genome loci associated with this trait were identified, one of them was corresponded to the predicted gene *Pbr013897* (Fig. 3a and Supplementary Data 7), which encodes a gibberellin 20-oxidase. It has been reported that gibberellin 20-oxidase is involved in fruit growth in apple<sup>31</sup>, and it could also regulate the flowering and stolon differentiation in strawberry<sup>32</sup>.

**A stone cell-related gene from an uncharacterized family.** High accumulation of stone cells significantly affects the quality of fruit and is particularly prevalent in sand pear fruit flesh, wherein they distribute in a mosaic pattern<sup>33</sup>. These brachysclereid cells have secondary thickened cell walls and ultimately become parenchyma cells with lignin and cellulose accumulation<sup>18</sup>. Our GWAS identified 12 candidate loci containing a total of 35 SNPs on chromosomes 2, 4, 5, 9, 11, 13, 14, 15, and 17 that were associated with stone cell content (Figs. 3a and 4a, b, and Supplementary Data 7). Since stone cells in pear are mainly produced at early stages of fruit development<sup>16</sup>, we next examined the fruit development RNA-Seq data that we previously generated for four pear cultivars ('Dangshansuli', 'Hosui', 'Yali', and 'Starkrimson')<sup>20,34</sup> to identify genes in these loci that were highly expressed in the early fruit development. Among the identified candidate genes, in addition to previously characterized lignin regulating genes such as those encoding cinnamate 4-hydroxylases<sup>35</sup>, *PbrSTONE* (lignin-related stone cell formation; *Pbr005042*) gene was of particular interest. This gene showed no characterized functions from other organisms, and its encoded protein contained a conserve domain (DUF1223) with an unknown function.

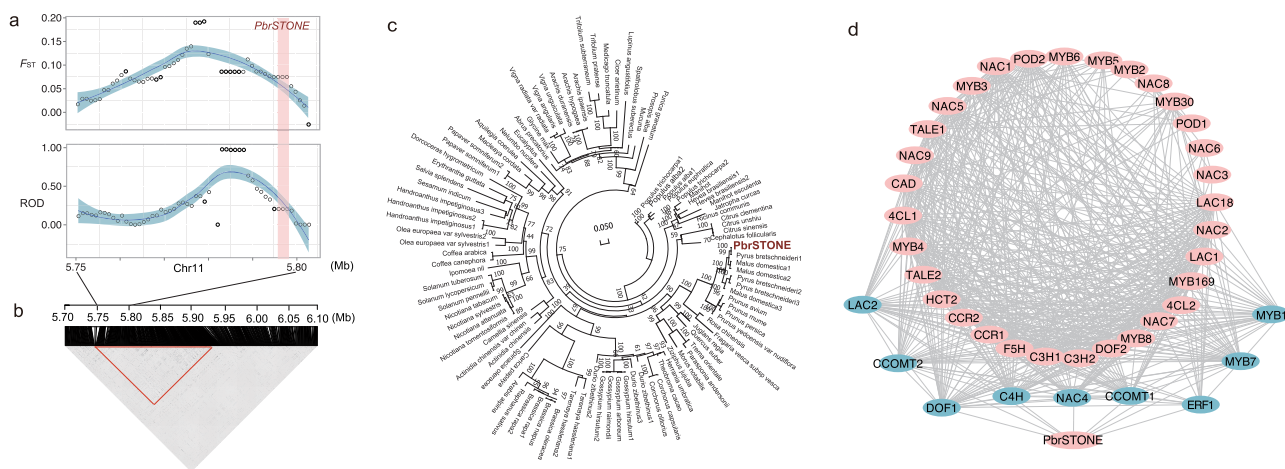
We measured stone cell contents in 312 sand pear accessions, which revealed that the landrace accessions had significantly

more stone cells than the improved cultivars ( $p$ -value =  $3.69E-15$ ) (Supplementary Fig. 16). Genome region of *PbrSTONE* had high  $F_{ST}$  and reduction of diversity (ROD) values indicated that it has been under selection (Fig. 4a, b). Furthermore, transcript of *PbrSTONE* was abundantly expressed at early stages of fruit development (Supplementary Fig. 17). We constructed a phylogenetic tree using *PbrSTONE* and its homologs in 81 eudicot species, and found *PbrSTONE* and its homologous genes from *Malus domestica*, *Prunus avium*, *Prunus persica*, and *Juglans regia* clustered in one group (Fig. 4c). We also carried out gene co-expression network analysis to find candidate genes from the lignin biosynthetic pathway and transcription factors correlated with *PbrSTONE*, and identified seven genes including *PbrLAC2* that had similar expression patterns with *PbrSTONE* (Fig. 4d).

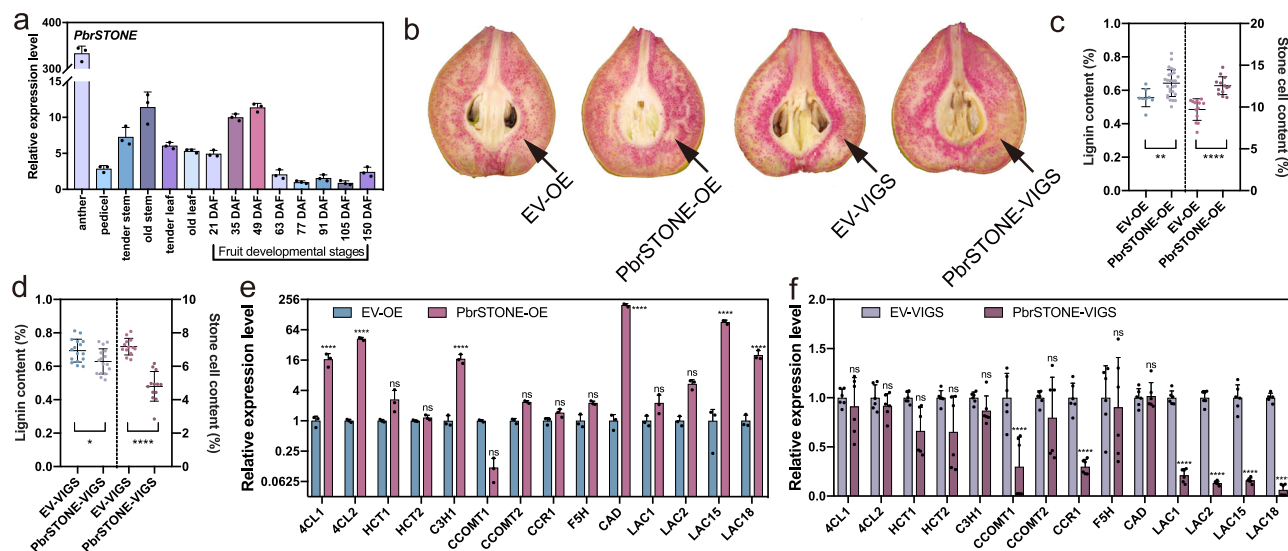
We then conducted a qPCR-based analysis of *PbrSTONE* gene in different tissues of 'Dangshansuli' plants, which revealed relatively high expression levels of *PbrSTONE* in anthers, stems, and young fruits (from 21 days after flowering (DAF) to 49 DAF) (Fig. 5a), but much lower levels in pedicels and leaves, as well as in older fruits (after 63 DAF) (Fig. 5a). The expression pattern of *PbrSTONE* during pear fruit development was consistent with the fact that lignin deposition in stone cells generally occurs during the early stages of fruit development<sup>16</sup>.

To further elucidate the functional role of *PbrSTONE* in lignin biosynthesis of stone cells, the *PbrSTONE* overexpression or silencing constructs were agro-infiltrated into 'Dangshansuli' pear fruit at 35 DAF. We confirmed the expression of *PbrSTONE* (Supplementary Fig. 18a) at 10 days post infiltration, and also observed a dramatic change of the lignin staining signal at the infiltration sites. Overexpressing *PbrSTONE* significantly increased stone cell and lignin contents, while silencing *PbrSTONE* decreased them (Fig. 5b–d). The most striking observation from these transient expression experiments was that regulating *PbrSTONE* expression caused a significant change of the expression of lignin biosynthesis-related genes, particularly the expression of LACs whose homologous genes in *Arabidopsis* are known to be involved in lignin polymerization<sup>36,37</sup> (Fig. 5e, f).

We next generated transgenic *Arabidopsis* lines stably expressing *PbrSTONE* under the control of the 35S promoter. We confirmed the constitutive expression of *PbrSTONE* in stems of 4-week-old plants from three overexpression lines (*PbrSTONE*-OE) (Supplementary Fig. 18b). In  $T_3$  generation homozygous plants, qPCR analysis of 4-week-old wild-type (WT) and *PbrSTONE*-OE



**Fig. 4** Stone cell content-related gene *PbrSTONE*. **a** Distribution of  $F_{ST}$  and ROD in the *PbrSTONE* gene region in landrace and improved groups. The shaded region in pink indicates the location of *PbrSTONE*, and the shade region in blue presents the smooth local regression by 'loess'. **b** LD in the *PbrSTONE* gene region. **c** Phylogenetic tree of homologous genes of *PbrSTONE* in 81 plant species. **d** Co-expression network of *PbrSTONE* with lignin related genes. Source data underlying (c, d) are provided as a Source Data file.



**Fig. 5 Functional validation of *PbrSTONE* in controlling stone cell content and lignin of pears.** **a** Expression profiles of *PbrSTONE* in different tissues of ‘Dangshansuli’ plants determined by q-RT-PCR;  $n = 3$  biologically independent samples. **b** Transient assays using *PbrSTONE* overexpression and silencing constructs in ‘Dangshansuli’ fruit at 35 days after flowering (DAF). Images were taken at 10 days after agro-infiltration. **c, d** Lignin and stone cell contents in the flesh tissue around the infiltration sites. More than six fruits were injected with each construct in an experiment that was repeated three times. EV empty vector, OE overexpression, VIGS virus-induced gene silencing. For lignin content,  $n = 8, 27, 16$ , and 16 biologically independent samples for EV-OE, *PbrSTONE*-OE, EV-VIGS, and *PbrSTONE*-VIGS, respectively. For stone cell content,  $n = 13$  biologically independent samples for all lines. **e, f** Expression levels of genes encoding enzymes involved in monolignol biosynthesis and polymerization in the flesh tissue around the infiltration site using qPCR analysis;  $n = 3$  biologically independent samples for EV-OE and *PbrSTONE*-OE, and  $n = 6$  for EV-VIGS and *PbrSTONE*-VIGS. Data presented are mean  $\pm$  SD;  $p$ -values were determined by two-tailed Student’s  $t$ -test (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ; ns, not significant). Source data are provided as a Source Data file.

plants revealed that the expression of the *Arabidopsis* homologs of the pear lignin biosynthesis genes *PbrC3H1* (*AtC3H1*) and *PbrLAC15/18* (*AtLAC17*) were significantly induced by transgenic overexpression of *PbrSTONE* (Fig. 6a). We analyzed a variety of phenotypes and found several major differences between WT and *PbrSTONE*-OE plants. Specifically, although the final height and diameter of the primary inflorescence stem of the transgenic lines did not differ from the WT at 8 weeks (Supplementary Fig. 19), the transgenic plants had significantly increased lignin content and decreased root length compared to the WT (Fig. 6b and Supplementary Fig. 19). Moreover, the transgenic plants had significantly increased accumulation of G-lignin monomers compared to WT plants, whereas no differences were observed for the accumulation of S-lignin monomers (Fig. 6c).

Toluidine blue O staining showed that the *PbrSTONE*-OE stems contained significantly higher numbers of interfascicular fiber (IF) cell layers than WT plants (Fig. 6d–f). Both the vessels and IF cell tissues of the *PbrSTONE*-OE lines exhibited stronger chemical staining and auto-fluorescence signals for lignin components than did the WT tissues (Fig. 6g–l, q, r), further supporting that *PbrSTONE* expression dramatically alters lignin metabolism. These results support that *PbrSTONE* functions in both IF formation and lignin deposition in vessels and in IF cell tissues.

To examine anatomical features of the secondary cell walls of the multiple *PbrSTONE*-OE lines, we conducted transmission electron microscopy (TEM) analysis. We observed that the secondary cell walls of the vessels in *PbrSTONE*-OE lines were significantly thicker than in WT plants, whereas no such difference in secondary cell wall thickness was observed when examining the IF tissue (Fig. 6m–p, s, t).

In order to better understand how *PbrSTONE* regulates lignin formation in stone cells, BiFC-LUC assay was conducted with candidate co-expressed genes to identify potential *PbrSTONE* interacting proteins. *PbrSTONE* was found to interact with

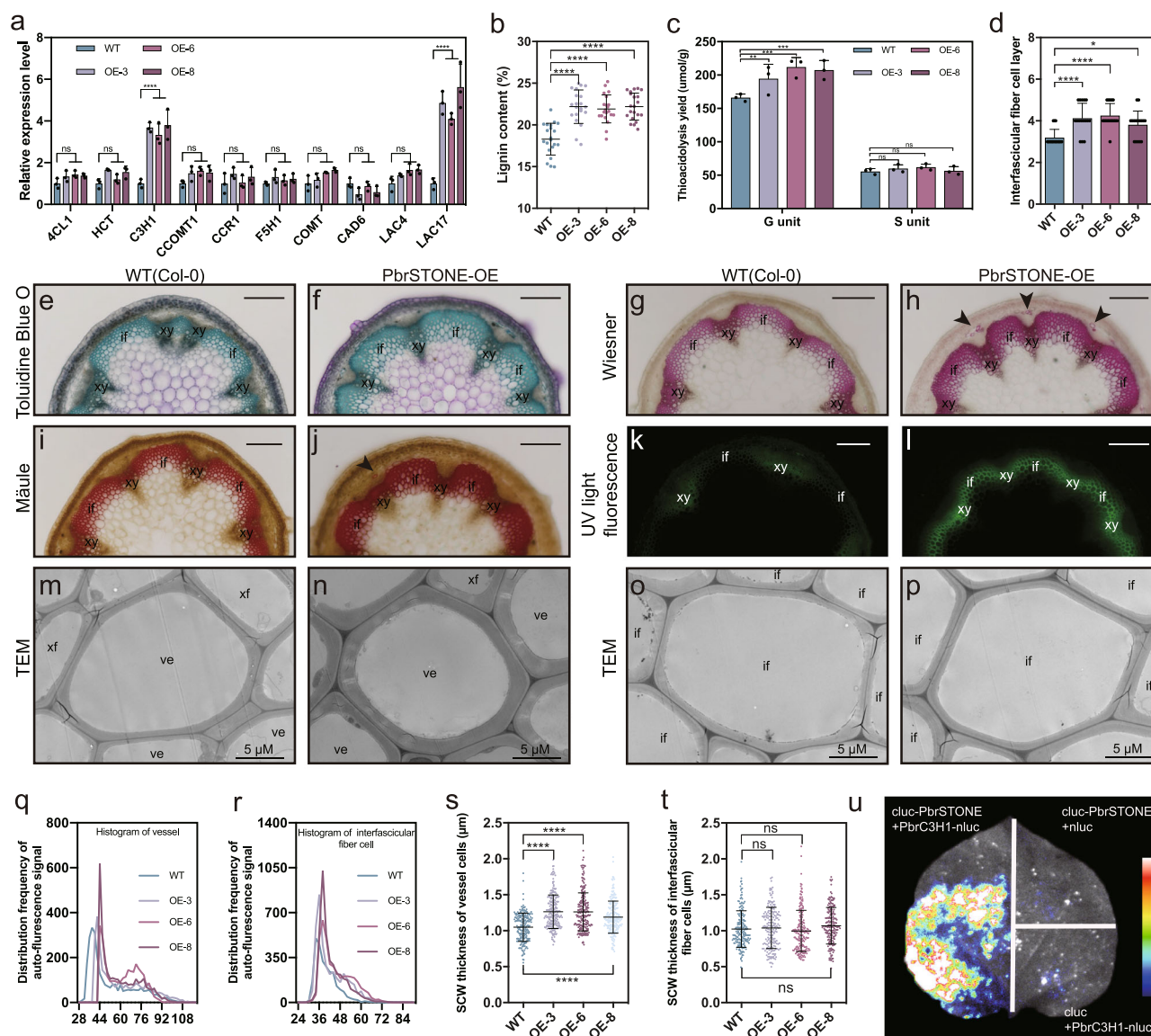
*PbrC3H1* (Fig. 6u), which was positively regulated by a positive regulator of lignification and stone cell development, *PbrMYB169* (ref. 17). The homologous gene of *PbrC3H1* in *Arabidopsis*, *AtREF8* (AT2G40890.1), encodes a coumarate 3-hydroxylase (C3H), a P450-dependent monooxygenase, and is involved in lignin and flavonoid biosynthesis<sup>38</sup>. *PbrSTONE* was cytoplasm and cytomembrane localized as determined by *PbrSTONE*-GFP fusion protein in transgenic *Arabidopsis* root (Supplementary Fig. 20), consistent with the previously reported subcellular localization of C3H<sup>39</sup>. Taken together, these results suggested that *PbrSTONE* may take part in lignification of stone cells via interacting with *PbrC3H1* to form a protein complex.

## Methods

**Agronomic trait measurement.** All of the measurements for the agronomic traits were based on the Description and data standard for pear (*Pyrus spp.*)<sup>40</sup>. Data were collected in 2014, 2015, and 2016 (some traits were recorded in 2 years because of the unpredictable dropped fruits). Single fruit weight was recorded as the average weight of ten fruits from each sample. Fruit skin color was recorded as ‘1’ to ‘6’ according to the Supplementary Fig. 21. Coverage ratio of fruit russet was defined as the ratio of russet area to fruit skin area in the mature fruits. Data were recorded as ‘1’, ‘3’, ‘5’, ‘7’ (Supplementary Fig. 22). Location of fruit russet was recorded as ‘1’, ‘2’, ‘3’, and ‘4’ (Supplementary Fig. 23). Furrows on mature fruit surface were recorded as ‘0’ and ‘1’ (Supplementary Fig. 24). Direction of carpodium was recorded as ‘1’ and ‘2’ according to the Supplementary Fig. 25. Direction of sepal was recorded as ‘1’, ‘2’, and ‘3’ according to the Supplementary Fig. 26. For days of fruit development, days of vegetative growth, and initial bloom period, data were recorded as the calendar date and were calculated as the number of days from the earliest date to the last one. The stone cell contents were measured according to Tao et al.<sup>33</sup>. Briefly, pulp of mature fruits was collected, stored at  $-20^{\circ}\text{C}$  for 24 h, and then homogenized with juice extractor. The sample was centrifuged, and after centrifugation the supernatant was removed. The resulting pellets (stone cells) were re-suspended in water, vortexed, and centrifuged. These washing steps were repeated several times until the supernatant was transparent. The collected stone cells were oven-dried and then weighed three times.

**Genome re-sequencing.** Young leaves from a total of 312 sand pear accessions (Supplementary Data 1) were collected from the Wuchang Sand Pear Germplasm collection (Hubei Academy of Agricultural Sciences, China). DNA was extracted





**Fig. 6 Functional validation of *PbrSTONE* in *Arabidopsis*.** **a** Expression levels of genes encoding enzymes involved in monolignol biosynthesis and polymerization in WT and *PbrSTONE* transgenic *Arabidopsis* plants using qPCR analysis;  $n = 3$  biologically independent samples. **b**, **c** Lignin analysis for WT and *PbrSTONE* transgenic *Arabidopsis* plants;  $n = 21$  biologically independent samples for **b**, and  $n = 3$  for **c**. **d** Layers of stem interfacial fiber cells in transgenic *Arabidopsis* plants expressing *PbrSTONE*;  $n = 16$  biologically independent samples. **e**, **f** Cross-sections of 8-week-old inflorescence stems stained with Toluidine blue O. **g**–**j** Cross-sections of 8-week-old inflorescence stems stained with Mäule and Wiesner reagent, which bind to S-lignin monomer and G-lignin monomer, respectively. **k**, **l** Lignin auto-fluorescence of cross-sections detected under UV light. Scale bars: 200  $\mu\text{m}$  in **k**–**r**. Bar in **k**–**r** = 200  $\mu\text{m}$ ; xy, xylem; if, interfacial fiber cell. The transgenic and WT plants were cultivated two times and cross-sections were taken from five independent plants which showed similar results. A representative picture from each line is shown. **m**–**p** Transmission electron micrographs of the cross-section of cells in **e**; ve, vessel; xf, xylary fiber; if, interfacial fiber cells. **q**, **r** Statistical analysis of lignin auto-fluorescence signal in the vessel cells and interfacial fiber cells. More than 12 samples in each line were analyzed. **s**, **t** Statistical analysis of secondary cell wall thickness in the vessel cells and interfacial fiber cells. More than 50 cells in each line were analyzed. **u** Firefly luciferase complementation assay in young *Nicotiana benthamiana* leaves. Data presented are mean  $\pm$  SD;  $p$ -values were determined by two-tailed Student's  $t$ -test ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ,  $****p < 0.0001$ ; ns, not significant). Source data underlying (**a**–**t**) are provided as a Source Data file.

using plant genomic DNA kits from TIANGEN<sup>®</sup>. DNA quality was checked by NanoPhotometer<sup>®</sup> spectrophotometer (IMPLEN) and 1% agarose gels, and DNA concentration was measured by Qubit<sup>®</sup> DNA Assay Kit in a Qubit 2.0 Fluorometer (Life Technologies). NEB Next Ultra DNA Library Prep Kit for Illumina<sup>®</sup> (NEB) was used to generate the sequencing libraries. The cBot Cluster Generation System of the HiSeq 4000 PE Cluster Kit (Illumina) was used for clustering of the indexed libraries, and the Illumina HiSeq 4000 system was used to sequence the DNA libraries (150 bp paired-end reads).

**Data filtering and SNP calling.** Quality of the 150-bp paired-end reads was checked by FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>),

and the reads were then processed using Trimmomatic<sup>41</sup> to trim the adapters and sequences with average quality per base below 20. The trimmed reads with length of at least 40 bp were mapped against the reference pear genome<sup>20</sup> (*P. bretschneideri*) using BWA aln<sup>42</sup> with parameters ‘-n 0.04 -o 1 -e 2’. Genome Analysis Toolkit (4.1.4) was used for SNP calling with unified genotyper method<sup>43</sup>, and the SNPs were filtered using criteria of MAF  $\geq 0.03$  and missing rate no more than 30%.

**$F_{ST}$  and ROD analysis and selective sweep detection.**  $F_{ST}$  and ROD were calculated using vcftools<sup>44</sup> with a window size of 10 kb and a window step size of 5 kb, and ‘gplots’ (<https://cran.r-project.org/web/packages/gplots/>) in the R packages was applied for presentation. RNA-Seq data of landrace and improved pears<sup>24</sup> were

used for *PbrSTONE* selection analysis. Candidate selective sweeps were first detected using the SweepD's composite likelihood rate (CLR) test of the site frequency spectra<sup>45</sup> with a window size of 10 kb, windows with top 5% CLR scores were considered candidate regions under a selective sweep. These regions were further filtered using  $\pi$  ratios, and only those having top 15%  $\pi$  ratios were considered the final selective sweeps. GWAS associated loci within 550 bp up or downstream of the identified selective sweeps were extracted. We also extracted genes in the selective sweep regions based on the gene annotation of the 'Dangshansuli' genome covered by sweeps.

**Genome-wide association studies.** The pear population structure was calculated by STRUCTURE<sup>46</sup> with the simulated population structure (K value) ranging from 2 to 6. GWAS were performed using both TASSEL<sup>47</sup> and EMMAX programs<sup>48</sup>. Candidate associated loci identified by both TASSEL and EMMAX were considered further in our study, as well as those identified in one program but with genes annotated as known to be related to the corresponding traits. TASSEL was also used for the phylogenetic tree (Neighbor-Joining method), PCA, kinship matrix of the population, and LD analyses. General linear model (GLM) and mixed linear model (MLM) used for GWAS analyses were compared based on the Q-Q plots, and we finally found MLM-PCA-kinship was the best for our sand pear population. The  $R^2$  of LD was visualized using the LDheatmap R package<sup>49</sup>. The cutoff was used as  $P = 1E-5$ . The genomic inflation factor was calculated by R package of GenABEL<sup>50</sup> with regression method.

**Gene annotation and statistical analysis.** We used the annotated information of pear genes from KEGG and Swissport database. Specifically, the DNA and amino acid sequences of *PbrSTONE* were used for BLAST analysis against the NCBI nucleotide, protein and conserved domain databases. We also used hidden Markov model analysis to compare *PbrSTONE* amino acid sequences to the Pfam domain database<sup>51</sup>. We used Student's *t*-tests or one-way ANOVA to compare statistical differences using GraphPad prism 8.0 software (GraphPad Software).

**Gene co-expression analysis and phylogeny of *PbrSTONE*.** Genes (*4CL1*, *4CL2*, *C3H1*, *C3H2*, *CAH*, *CAD*, *CCOMT1*, *CCOMT2*, *CCR1*, *CCR2*, *F5H*, *HCT1*, *HCT2*, *POD1*, *POD2*, *LAC1*, *LAC2*, and *LAC18*) from lignin biosynthetic pathway and transcription factors (*LAC1*, *LAC2*, *LAC18*, *MYB1-8*, *MYB24*, *MYB30*, *MYB169*, *NAC1-9*, *ERF1*, *DOF1*, *DOF2*, *TALE1*, and *TALE2*) were selected as the candidate genes for co-expression network analysis with *PbrSTONE* using WGCNA<sup>52</sup>, using the fruit gene expression data of six developmental stages of five pear subspecies from our previous study<sup>35</sup>. The correlation relationship was visualized using Cytoscape (<http://www.cytoscape.org>).

We performed 'blastp' of the *PbrSTONE* protein sequence against the NCBI nr database, and obtained 99 proteins from 81 eudicot species with an *e*-value  $<1E-5$  and sequence identity above 60%. All of the identified proteins were aligned with CLUSTAL<sup>53</sup> (version 2.1), and the Maximum Likelihood method in IQ-TREE<sup>54</sup> (multicore version 1.6.10) was applied for the phylogenetic tree construction with 1000 bootstraps.

**Gene expression analysis using qPCR.** cDNA was synthesized using one-step gDNA removal and cDNA synthesis kit (Transgen). qPCR was performed using the LightCycler 480 SYBR GREEN Master system (Roche). Primers are shown in Supplementary Data 8. The analysis was conducted using three biological and three technical repeats. Relative expression levels of each gene were calculated using the  $2^{-\Delta\Delta C_p}$  method. *PbrGAPDH* and *Atactin/AtEF1 $\alpha$*  were used as reference genes for pear and *Arabidopsis*, respectively.

**Transient transformation of *PbrSTONE* in pear fruit flesh.** To transiently overexpress *PbrSTONE*, the full-length coding sequence of *PbrSTONE* was fused in frame to the N-terminus of GFP under the control of CaMV 35S promoter in the binary vector pCAMBIA1302 (p1302) to form the fusion vector 35S::PbrSTONE-GFP. For TRV virus-induced gene silencing (VIGS), the partial coding sequences of *PbrSTONE* (1–276 bp) was amplified using primers listed in Supplementary Data 8 and the amplified fragment was inserted into the vector TRV2. The fusion constructs and the control vector (p1300) were transferred into *Agrobacterium tumefaciens* strain GV3101 by the freeze-thaw method. For transiently transformed pear fruit flesh, the cells were infiltrated into 'Dangshansuli' fruit flesh at 35 °DAF using needleless syringes. The transformed fruits were placed in the dark at 22 °C overnight and then incubated in a growth chamber at 22 °C under 16 h photoperiod for 10 days before being examined and imaged<sup>16</sup>. Six fruits were injected with each construct in an experiment that was repeated three times independently. The acetyl bromide-based method was used to detect lignin in fruit flesh tissue. The lignin content was shown as a percentage (calculated lignin content/ dry weight of stone cells  $\times 100\%$ )<sup>16</sup>.

**Arabidopsis transformation.** *Agrobacterium tumefaciens* containing the vector 35S::PbrSTONE-GFP were transformed into *Arabidopsis* Col-0 plants using using the floral dip method<sup>55</sup>. Expression levels of *PbrSTONE* in primary inflorescence stem of two-week-old T<sub>1</sub> plants were analyzed by semi-quantitative RT-PCR using

primers listed in Supplementary Data 8. Relatively high levels of *PbrSTONE* expression were measured in three transgenic lines (OE-3, OE-6, and OE-8), which were selected to generate T<sub>3</sub> homozygous plants for the subsequent analyses. Root length of each transgenic line and the WT control were measured at 7 days after germination on agar medium. Subsequently the plants were transplanted in a glasshouse with a 16 h photoperiod. The height and diameter of primary inflorescence stem were determined for each transgenic line and the WT control at 8 weeks post germination.

**Lignin analysis.** For WT and T<sub>3</sub> transgenic *Arabidopsis* plants, the basal 10 cm of the primary inflorescence stem was chopped and pooled together from five plants. Samples of stem pieces were extracted sequentially for water, ethanol, chloroform, and acetone to gain purified cell wall residue, then dried for the subsequent lignin analysis. Acetyl bromide-soluble lignin content was estimated according to a standard procedure<sup>56</sup>. The lignin composition was determined by thioacidolysis<sup>57</sup>. The released lignin monomers were derivatized with *N,O*-bis (trimethylsilyl) acetamide and quantified by GC-MS as their trimethylsilylated derivatives. The analytical data were presented as the mean values obtained from the same amount of starting material harvested from the pooled xylem tissues of five plants.

**Microscopy.** For light microscopy analysis, the basal 1 cm of primary inflorescence stems of 8-week-old *Arabidopsis* WT and *PbrSTONE* transgenic plants were harvested and fixed in formalin-acetic acid-alcohol fixative (1.9% formaldehyde solution [v/v], 63% ethyl alcohol [v/v], 5% glacial acetic acid [v/v], 30.1% water [v/v]) at 4 °C for 7 days. Cross sections (100  $\mu$ m thickness) of inflorescence stems were sectioned with a Leica VT1000S vibratome (Leica Mikrosysteme, Germany), stained with Toluidine Blue O, Mäule and Wiesner reagent separately, and then observed with Nikon Ni-U microscope (Nikon, Japan)<sup>58</sup>. Nikon Ti-E microscope (Nikon, Japan) was used to visualize the lignin auto-fluorescence under UV light (excitation at 355/25). Fluorescence intensity was measured on the images with ImageJ software.

TEM analyses were performed with the stem samples that were fixed in a fixative (2.5% [v/v] glutaraldehyde, 50 mM sodium phosphate [pH 7.2], and 4% [v/v] formaldehyde) for 12 h and further fixed by 0.5% osmic acid<sup>59</sup>. The tissues were then dehydrated and embedded in LR white resin (Polysciences, Warrington, PA, USA). Ultrathin sections were stained with uranyl acetate and lead citrate. The secondary wall thickness of vessel cells was gained on the images of TEM with ImageJ.

**Bimolecular fluorescence complementation analysis.** pCAMBIA1300-NLuc and pCAMBIA1300-CLuc were used for bimolecular fluorescence complementation (BiFC) analysis<sup>60</sup>. Full-length coding sequences of *PbrSTONE* and *PbrC3H1* were amplified using the primers listed in Supplementary Data 8 and then inserted individually into the two plasmids and transferred into *A. tumefaciens* strains GV3101. Equal amounts of *PbrSTONE* and *PbrC3H1* combinations were co-infiltrated into the young leaves of *Nicotiana benthamiana* plants. On the second day post-infiltration, 1 mM luciferin was sprayed onto the leaves, and the samples were incubated in the dark for 5–10 min. LUC images were captured using a low light cooled CCD imaging apparatus (Tanon 5200 Multi).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. Raw genome resequencing reads have been deposited into the NCBI sequence read archive (SRA) under BioProject accession [PRJNA563813](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA563813). Sequence data of *PbrSTONE* and *PbrC3H1* can be found at GenBank with the accession numbers [MT711883](https://www.ncbi.nlm.nih.gov/nuclot/MT711883) and [MT711884](https://www.ncbi.nlm.nih.gov/nuclot/MT711884), respectively. SNPs identified in this study are available at Zenodo (<https://zenodo.org/deposit/3971245>). Figures 1, 2, 3a, 3c, 3d, 3e and 4b are generated from the SNPs data using parameters specified in 'Methods'. Source data are provided with this paper.

Received: 9 September 2019; Accepted: 25 January 2021;  
Published online: 18 February 2021

## References

- Rom, R. C. & Robert, F. C. *Rootstocks for Fruit Crops* (Wiley, 1987).
- Wu, J. et al. High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *J. Exp. Bot.* **65**, 5771–5781 (2014).

3. Cho, K. H. et al. Development of AFLP and CAPS markers linked to the scab resistance gene, *Ryn2*, in an inter-specific hybrid pear (*Pyrus* spp.). *J. Hortic. Sci. Biotechnol.* **84**, 619–624 (2009).
4. Dondini, L. et al. Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. *Mol. Breed.* **14**, 407–418 (2004).
5. PMF, L. Roux et al. Redefinition of the map position and validation of a major quantitative trait locus for fire blight resistance of the pear cultivar ‘Harrow Sweet’ (*Pyrus communis* L.). *Plant Breed.* **131**, 656–664 (2012).
6. Yamamoto, T. et al. Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci.* **64**, 351–361 (2014).
7. Thornberry, J. M. et al. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**, 286–289 (2001).
8. Huang, X. H. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–976 (2010).
9. Yano, K. et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* **48**, 927–934 (2016).
10. Cao, K. et al. Genome-wide association study of 12 agronomic traits in peach. *Nat. Commun.* **7**, 13246 (2016).
11. Mariette, S. et al. Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). *N. Phytol.* **209**, 773–784 (2016).
12. Li, X. et al. Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol. J.* **17**, 1582–1594 (2019).
13. Yao, G. F. et al. Map-based cloning of the pear gene MYB114 identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant J.* **92**, 437–451 (2017).
14. Li, J. M. et al. A new insight into the evolution and functional divergence of SWEET transporters in Chinese white pear (*Pyrus bretschneideri*). *Plant Cell Physiol.* **58**, 839–850 (2017).
15. Xu, L. L., Qiao, X., Zhang, M. Y. & Zhang, S. L. Genome-wide analysis of aluminum-activated malate transporter family genes in six rosaceae species, and expression analysis and functional characterization on malate accumulation in Chinese white pear. *Plant Sci.* **274**, 451–465 (2018).
16. Xue, C. et al. PbrmiR397a regulates lignification during stone cell development in pear fruit. *Plant Biotechnol. J.* **17**, 103–117 (2019).
17. Xue, C. et al. PbrMYB169 positively regulates lignification of stone cells in pear fruit. *J. Exp. Bot.* **70**, 1801–1814 (2019).
18. Smith, W. W. The course of stone cell formation in pear fruits. *Plant Physiol.* **10**, 587–611 (1935).
19. Cai, Y. P. et al. Study of the structure and biosynthetic pathway of lignin in stone cells of pear. *Sci. Hortic.* **125**, 374–379 (2010).
20. Wu, J. et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* **23**, 396–408 (2013).
21. Wu, J. et al. Diversification and independent domestication of Asian and European pears. *Genome Biol.* **19**, 77 (2018).
22. Zhou, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408–414 (2015).
23. Xu, X. et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* **30**, 105–111 (2012).
24. Li, X. L. et al. Comparative transcriptomic analysis provides insight into the domestication and improvement of pear (*P. pyrifolia*) fruit. *Plant Physiol.* **180**, 435–452 (2019).
25. Lam, H. M. et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **42**, 1053–1059 (2010).
26. Duan, N. B. et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat. Commun.* **8**, 249 (2017).
27. An, J. P. et al. R2R3-MYB transcription factor MdMYB23 is involved in the cold tolerance and proanthocyanidin accumulation in apple. *Plant J.* **96**, 562–577 (2018).
28. Schaart, J. G. et al. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria x ananassa*) fruits. *New Phytol.* **197**, 454–467 (2013).
29. Ampomah-Dwamena, C. et al. A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol.* **221**, 309–325 (2019).
30. Dal Cin, V. et al. Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *Plant Cell* **23**, 2738–2753 (2011).
31. Kusaba, S. et al. Characterization of gibberellin 20-oxidase gene in apple. *Acta Hortic.* **538**, 605–608 (2000).
32. Tenreira, T. et al. A specific gibberellin 20-oxidase dictates the flowering-running decision in diploid strawberry. *Plant Cell* **29**, 2168–2182 (2017).
33. Tao, S. T. et al. Anatomy, ultrastructure and lignin distribution of stone cells in two *Pyrus* species. *Plant Sci.* **176**, 413–419 (2009).
34. Zhang, M. Y. et al. Distinct transcriptome profiles reveal gene expression patterns during fruit development and maturation in five main cultivated species of pear (*Pyrus* L.). *Sci. Rep.* **6**, 28130 (2016).
35. Lewis, A. N. G. et al. Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* **61**, 221–294 (2002).
36. Berthet, S. et al. Disruption of LACCASE4 and 17 results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* **23**, 1124–1137 (2011).
37. Zhao, Q. et al. LACCASE is necessary and nonredundant with PEROXIDASE for lignin polymerization during vascular development in *Arabidopsis*. *Plant Cell* **25**, 3976–3987 (2013).
38. Franke, R. et al. Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of *Arabidopsis*. *Plant J.* **30**, 47–59 (2002).
39. Fraser, C. M. & Chapple, C. The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis Book* **9**, e0152 (2011).
40. Cao, Y. -F., Liu, F., Hu, H. & Zhang, B. *Description and Data Standard for Pear* (*Pyrus* spp.). (China Agriculture Press, 2006).
41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
42. Li, H. et al. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
43. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
44. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
45. Pavlidis, P., Zivkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
46. Pritchard, J. K. et al. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
47. Zhang, Z. W. et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).
48. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
49. Ji-Hyung Shin, S. B. et al. LDheatmap: graphical display of pairwise linkage disequilibrium between SNPs. *J. Stat. Softw.* **16**, Code Snippet 3 (2006).
50. Aulchenko, Y. S. et al. Genome-wide rapid association using mixed model and regression: a fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis. *Genetics* **177**, 577–585 (2007).
51. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, 427–432 (2019).
52. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
53. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
54. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).
55. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
56. Acker, R. V. et al. Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnol. Biofuels* **6**, 1–17 (2013).
57. Lapierre, C., Pollet, B. & Rolando, C. New insights into the molecular architecture of hardwood lignins by chemical degradative methods. *Res. Chem. Intermediat.* **21**, 397–412 (1995).
58. Mitra, P. P. & Loque, D. Histochemical staining of *Arabidopsis thaliana* secondary cell wall elements. *J. Vis. Exp.* e51381 (2014).
59. Vanholme, R. et al. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* **341**, 1103–1106 (2013).
60. Chen, H. et al. Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiol.* **146**, 368–376 (2008).

## Acknowledgements

This work was supported by the National Key Research and Development Program (2018YFD1000200), the National Science Foundation of China (31725024, 31901978, 31801835, 31901983), the Earmarked Fund for the China Agriculture Research System (CARS-28), the Earmarked Fund for Jiangsu Agricultural Industry Technology System JATS[2020]401, the Natural Science Foundation of Jiangsu Province (SBK2018043235), the China Postdoctoral Science Foundation (2018M630565), and the US National Science Foundation (IOS-1855585). We thank Qingwen Liu, Lun Liu,



Jiaying Zhang for providing pear photos and other contributions to this work. This project was supported by the Bioinformatics Center of Nanjing Agricultural University.

### Author contributions

J.W., S.Z., and Z.F. designed the project. M.Y.Z. performed the data analysis. C.X. and Y.X. performed the functional gene verification with assistance from X.C., H.H., J.F., M.Q. and B.B. collected the phenotype data. J.L., R.W., C.Z., S.T., C.G., X.L., S.W., G.Y., Y.L., and M.S. gave suggestions of the analysis. M.Y.Z., C.X., and J.W. wrote and revised the paper. Z.F. provided guidance on data analysis, and revised the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21378-y>.

**Correspondence** and requests for materials should be addressed to Z.F., S.Z. or J.W.

**Peer review information** *Nature Communications* thanks Pere Arús, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021



## Genomic selection of eight fruit traits in pear

Manyi Sun<sup>a,1</sup>, Mingyue Zhang<sup>b,1</sup>, Satish Kumar<sup>c</sup>, Mengfan Qin<sup>a</sup>, Yueyuan Liu<sup>a</sup>,  
Runze Wang<sup>a</sup>, Kaijie Qi<sup>a</sup>, Shaoling Zhang<sup>a</sup>, Wenjing Chang<sup>a</sup>, Jiaming Li<sup>a</sup>, and Jun Wu<sup>a,d,\*</sup>

<sup>a</sup> College of Horticulture, State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

<sup>b</sup> State Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an, Shandong 271018, China

<sup>c</sup> The New Zealand Institute for Plant and Food Research Limited, Private Bag 1401, Havelock North 4157, New Zealand

<sup>d</sup> Zhongshan Biological Breeding Laboratory, Nanjing, Jiangsu 210014, China

Received 25 November 2022; Received in revised form 11 January 2023; Accepted 3 April 2023

Available online 27 April 2023

### A B S T R A C T

Genomic selection (GS) has the potential to improve selection efficiency and shorten the breeding cycle in fruit tree breeding. In this study, we evaluated the effect of prediction methods, marker density and the training population (TP) size on pear GS for improving its performance and reducing cost. We evaluated GS under two scenarios: (1) five-fold cross-validation in an interspecific pear family; (2) independent validation. Based on the cross-validation scheme, the prediction accuracy (PA) of eight fruit traits varied between 0.33 (fruit core vertical diameter) and 0.65 (stone cell content). Except for single fruit weight, a slightly better prediction accuracy (PA) was observed for the five parametrical methods compared with the two non-parametrical methods. In our TP of 310 individuals, 2 000 single nucleotide polymorphism (SNP) markers were sufficient to make reasonably accurate predictions. PAs for different traits increased by 18.21%–46.98% when the TP size increased from 50 to 100, but the increment was smaller (–4.13%–33.91%) when the TP size increased from 200 to 250. For independent validation, the PAs ranged from 0.11 to 0.45 using rrBLUP method. In summary, our results showed that the TP size and SNP numbers had a greater impact on the PA than prediction methods. Furthermore, relatedness among the training and validation sets, and the complexity of traits should be considered when designing a TP to predict the test panel.

**Keywords:** Pear; *Pyrus*; Prediction method; TP size; SNP marker number

### 1. Introduction

Pear (*Pyrus*), the third most important fruit tree species in the world, belongs to the Rosaceae family (Potter et al., 2007; He et al., 2022). Annual worldwide pear production is 25.66 million tons (2021, FAOSTAT). Due to the long juvenile period (6–12 years) (van Nocker and Gardiner, 2014) and self-incompatibility

(Claessen et al., 2019), pear breeding is a costly and time-consuming process. In addition, fruit quality (Zhang et al., 2022), which is a target of breeding performance, can only be measured late in the breeding programs.

Marker-assisted selection (MAS), which is suitable for monogenic and oligogenic traits, however, it may not perform well for polygenic traits in pear breeding. With the development of

<sup>1</sup> These authors contributed equally to this work.

\* Corresponding author.

E-mail address: [wujun@njau.edu.cn](mailto:wujun@njau.edu.cn)

Peer review under responsibility of Chinese Society of Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS)

<https://doi.org/10.1016/j.hpj.2023.04.008>

2468-0141/Copyright © 2023 Chinese Society for Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS). Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

high-throughput genotyping technologies, a large number of molecular markers have become available, so genomic selection (GS) can be applied to overcome the limitation of MAS in pear breeding projects. GS, a molecular breeding tool, constructs prediction models by using genome-wide markers and phenotypic data of the training population (TP) (Meuwissen et al., 2001) and uses these constructed models to predict breeding values of another population with only genotypic information. Therefore, GS can improve traditional breeding strategies through reducing large-scale phenotyping, cost and resources, shortening breeding cycles and accelerating breeding progress (Heffner et al., 2010; Nakaya and Isobe, 2012).

The prediction accuracy (PA) of GS (Wiggans et al., 2017) is generally measured as the Pearson's correlation coefficient between the observed breeding value and the genomic estimated breeding value (GEBV) (Desta and Ortiz, 2014). Factors affecting GS accuracy include genetic architecture (Desta and Ortiz, 2014), size and genetic composition of the TP (Spindel et al., 2015), number of markers (Wang et al., 2018b), statistical models (Wang et al., 2018b), population structure (Stewart-Brown et al., 2019), genetic distance between TP and the validation population (VP) (Spindel et al., 2015), and trait heritability (Stich and van Inghelandt, 2018). The effects of TP size, GS model, marker density and trait on the PA were evaluated in rice: 7 142 markers [1 single nucleotide polymorphism (SNP) for each 0.2 cM] were found to be efficient for genomic prediction, and no prediction method proved more suitable for the three traits (Spindel et al., 2015). Application of genomic prediction on 'Zhikong' scallop showed that heritability, TP size, and marker density had a large effect on the prediction performance (Wang et al., 2018b). In addition, methods of designing the TP set and genetic background of the TP were all crucial for an accurate genomic prediction (Stich and van Inghelandt, 2018).

Compared with annual crops and livestock species, very few GS studies have been reported in fruit tree species (Minamikawa et al., 2017; Hardner et al., 2019). In pear breeding, the effect of the different methods and genetic background of the TP on PA has been explored using a cross-validation approach (Minamikawa et al., 2018; Kumar et al., 2019). However, there are no reports on the effect of TP size and marker density on the performance of the prediction model in an independent breeding population. Our objectives were to evaluate the effects of parametrical and non-parametrical GS methods, marker density and TP size on the PA using cross-validations, and an independent population.

## 2. Materials and methods

### 2.1. Materials

The plant materials used for cross-validation consisted of 310 individuals, generated from a cross between 'Niitaka' (*Pyrus pyrifolia*) and 'Hongxiangsu' (*P. sinkiangensis* × *P. bretschneideri*) (NH), planted at the Jiangpu Horticulture Experimental Station of Nanjing Agricultural University, Nanjing, Jiangsu, China. 'Niitaka', the maternal parent, has a round shape and high fruit weight (450–500 g). 'Hongxiangsu', the paternal parent, has an oval or spindle shape and low fruit weight (~200 g). Since the two parents have large phenotypic variance, the hybrid population is

suitable for studying the genetic regularity of pear fruit shape and size. We randomly collected five to eight mature fruits of each of 310 trees for trait measurement in successive 3 years (2015–2017). Five young leaves from each of 310 individuals were randomly collected and quickly frozen by liquid nitrogen for further DNA extraction and sequencing. In addition, 31 seedlings from an F<sub>1</sub> generation of 'Cuiguan' (*P. pyrifolia*) × 'Starkrimson' (*Pyrus communis*) (CS) (Li et al., 2019) were used as a test population for independent validation of GS models developed from the NH family.

### 2.2. Trait phenotyping

The measured fruit traits included fruit transverse diameter (TD), fruit vertical diameter (VD), single fruit weight (SFW), fruit core vertical and transverse diameters (FCVD, FCTD), fruit shape index (FSI), stone cell content (SCC) and soluble solid content (SSC). We randomly selected five to eight mature fruits for trait measurement. TD, VD, FCTD and FCVD were measured (mm) by using digital Vernier caliper (Greener, China); FSI was calculated as the ratio of VD to TD, and SFW was weighed using digital scales (g) (G&G Electronic Scale T2000, Powerhouse®). SCC of fruit flesh was extracted using hydrochloric acid and weighed using digital scales (g · kg<sup>-1</sup>) (G&G Electronic Scale T2000, Powerhouse®). SSC was measured using a refractometer (Atago, PAL-1, Japan). Collection and analysis of the phenotypic data for the CS population were as described previously (Li et al., 2019).

### 2.3. DNA extraction and genotypic data preparing

Leaves of 310 NH individuals and their parents ('Niitaka' and 'Hongxiangsu') were collected for DNA extraction and sequencing. DNA extraction and library preparation of 176 NH individuals was conducted using Plant DNA Isolation kit (DE-0611, <http://www.foregene.com>) and Truseq Nano DNA HT Sample preparation Kit (Illumina, USA), respectively. An Illumina HiSeq2500 platform was then used for generating 150 bp paired-end reads sequencing. Protocols for DNA extraction and genotyping-by-sequencing (GBS) library preparation of the other 134 NH individuals were the same as previously reported (Zhou et al., 2016).

The raw sequence of each individual was trimmed by Trimmomatic (version: 0.39) (Bolger et al., 2014). Then, clean reads of each individual were aligned to the ('Dangshansuli 1.0') reference genome using BWA (Burrows-Wheeler-Alignment) software (version: 0.7.16) (Li and Durbin, 2009) with parameter: mem -t 4 -k 32 -M. We performed SNP calling by using BCFtools (version 1.9) (Li, 2011), and low-quality SNPs were filtered out when DP (read depth) < 4. Then, the two SNP datasets of the 176 and 134 NH individuals were merged for further filtering out SNPs with missing-rate > 10% and minor allele frequency (MAF) < 0.05 by using VCFtools (version 0.1.16) (Danecek et al., 2011). Finally, a total of 178 468 SNPs were retained from sequencing of the training population seedlings. The independent VP of 31 CS individuals was genotyped using the 200K SNP chip (Li et al., 2019), and a set of 861 common SNPs between the TP of NH and the CS population was used to make an independent validation. The common SNPs in the training population and the independent VP were converted into a numeric format using NGSEP software



(Duitama et al., 2014) and missing data were imputed with the mean algorithm that imputes mean value for each marker through the package rrBLUP (Endelman, 2011) in R.

2.4. Heritability and two prediction scenarios

Narrow sense heritability ( $h^2$ ) was calculated for each trait by using the function kin.ship in the rrBLUP package (Endelman, 2011). An estimate of  $h^2$  of each trait was calculated as the ratio of genetic variance ( $\sigma_g^2$ ) to the phenotypic variance ( $\sigma_g^2 + \sigma_e^2$ ), where  $\sigma_g^2$  and  $\sigma_e^2$  were the estimated additive genetic variance and the residue variance, respectively. Two validation scenarios were evaluated in this study: (1) five-fold cross-validation within the TP; (2) independent validation, where the NH population in scenario 1 was used as a TP to predict fruit traits of seedlings in the CS population.

2.5. Five-fold cross validation

Cross-validation as a method to obtain PA for a particular trait has been used in a range of studies (Spindel et al., 2015; Lozada et al., 2019). In this study, five-fold cross-validation was used to measure the PA of eight fruit traits. The NH training population was divided into five equal-sized sets. One set (62 lines) was extracted as a VP and the remaining four sets (248 lines) as TP used for developing the prediction model. Phenotype and genotype data of the TPs and genotype data of the VP was used to predict the GEBV value of VP. The real phenotype data and GEBV value of VP were used to calculate the PA. Eventually, all sets were predicted, and the mean PA (for each trait) was recorded. For SCC prediction, only 278 lines having SCC data were used for five-fold cross-validation and further analysis.

2.6. Factors affecting GS accuracy

Five parametrical and two non-parametrical prediction methods were applied to calculate the GEBV: ridge regression best linear unbiased prediction (RR-BLUP) was applied using the R package rrBLUP (Endelman, 2011); BayesA, BayesB, BayesC, Bayes ridge regression (BRR) and reproducing kernel Hilbert spaces

regression (RKHS) methods were implemented in the R package BGLR (Perez and de los Campos, 2014). The sparse neural networks (SNN) method was implemented in the R package snnR (Wang et al., 2018a).

To evaluate the effect of marker number on the PA for different traits, 13 random subsets of markers with different numbers (100, 200, 400, 600, 800, 1 000, 2 000, 5 000, 10 000, 50 000, 100 000, 150 000, 178 468) were selected from the NH training population SNP dataset, and the rrBLUP method was used to make a five-fold cross-validation with 100 replicates, and the mean accuracy of each subset was recorded.

For evaluating the effect of TP size on PA, five different TP subsets (50, 100, 150, 200 and 250 lines) were randomly selected and a VP with constant size (50) was selected from the rest of the samples. The phenotype and all SNPs (178 468) of TP subsets were used for building prediction models and further predicting the GEBV value of VP. The real phenotype data and GEBV value of VP were used to calculate the PA. Each TP subset was randomly selected with 500 replicates. Finally, the mean PA of each TP size was recorded for each trait.

2.7. Independent validation of GS

To evaluate the relatedness between the TP and VP, the additive kinship matrix of 341 seedlings (310 TP seedlings and 31 VP seedlings) was constructed by ‘A.mat’ function of the rrBLUP/R package (Endelman, 2011), and a principal component analysis (PCA) was performed using the PLINK software (version: 1.9) (Purcell et al., 2007) with default parameters. The GS model developed from the NH TP was used to predict the GEBV of the CS VP with seven prediction methods, and the performance of the independent validation was evaluated by PA.

3. Results

3.1. Trait heritability and distribution

$h^2$  and the trait distribution of the TP (first generation family of ‘Niitaka’ × ‘Hongxiangsu’) and VP (first generation family of

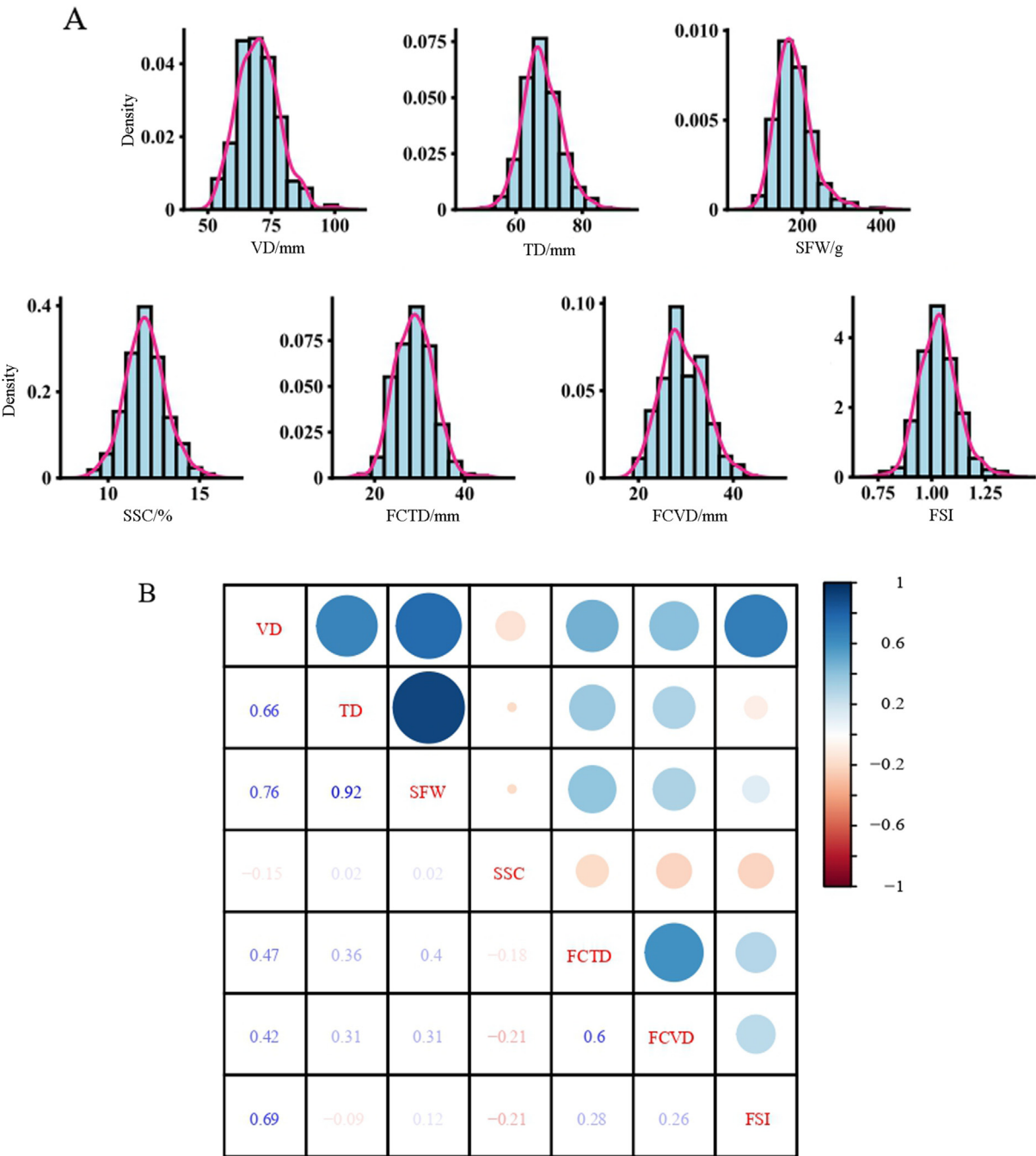
Table 1 Heritability ( $h^2$ ) and fruit trait summary of the training and validation populations

| Population  | No. lines | Trait                       | Mean   | Min   | Max    | $h^2$ |
|---|-----------|-----------------------------|--------|-------|--------|-------|
| Training population (‘Niitaka’ × ‘Hongxiangsu’)   | 310       | VD/mm                       | 69.51  | 51.53 | 101.01 | 0.57  |
|   |           | TD/mm                       | 67.58  | 49.76 | 88.60  | 0.39  |
|   |           | SFW/g                       | 177.64 | 74.30 | 406.83 | 0.37  |
|   |           | SSC/%                       | 11.97  | 8.90  | 15.80  | 0.57  |
|   |           | FCTD/mm                     | 28.84  | 16.41 | 45.06  | 0.56  |
|   |           | FCVD/mm                     | 29.54  | 18.88 | 44.87  | 1.00  |
|   |           | FSI                         | 1.03   | 0.76  | 1.35   | 0.78  |
|   |           | SCC/(g · kg <sup>-1</sup> ) | 4.15   | 0.07  | 11.72  | 0.53  |
|   |           | —                           | —      | —     | —      | —     |
| Validation population (‘Cuiguan’ × ‘Starkrimson’) | 31        | VD/mm                       | 69.28  | 37.13 | 87.40  | —     |
|   |           | TD/mm                       | 68.55  | 49.52 | 83.66  | —     |
|   |           | SFW/g                       | 186.67 | 94.70 | 306.63 | —     |
|   |           | SSC/%                       | 11.67  | 10.62 | 13.39  | —     |
|   |           | FCTD/mm                     | 30.25  | 20.91 | 40.79  | —     |
|   |           | FCVD/mm                     | 33.48  | 23.34 | 47.73  | —     |
|   |           | FSI                         | 1.01   | 0.75  | 1.15   | —     |

Note: Narrow sense heritability was calculated using the formula:  $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$  ( $\sigma_g^2$ : the estimated additive genetic variance;  $\sigma_e^2$ : the residue variance). VD: vertical diameter; TD: transverse diameter; SFW: single fruit weight; SSC: soluble solid content; FCTD: fruit core transverse diameter; FCVD: fruit vertical diameter; FSI: fruit shape index; SCC: stone cell content.

‘Cuiguan’ × ‘Starkrimson’) are presented in Table 1 and Fig. 1, a. In the training population,  $h^2$  for the measured traits ranged from 0.37 to 1.00. SFW and TD showed low heritability (0.37 and 0.39, respectively), but moderate–high heritability was observed for the other traits, with FCVD having the highest heritability (1.00), followed by FSI (0.78). Correlations between traits are presented in Fig. 1, b. VD, TD and SFW showed significantly ( $P$ -value = 0)

positive correlation with each other (Tables S1 and S2). FCTD was highly correlated with FCVD (0.60,  $P$ -value = 0), and FSI had positive correlation with VD (0.69,  $P$ -value = 0). No significant correlation appeared between SSC and other traits (Tables S1 and S2). Similar correlation trends were also observed in the VP (Table S3). Except for SSC, other traits had significant correlation with one or more traits (Tables S3 and S4). Fruit size (measured by VD



**Fig. 1 The summary of seven traits in training population**

(A) The distribution of seven traits in training population. Pink lines represent kernel density estimation. (B) The correlation of seven traits in training population. Color scale represents Pearson correlation coefficient (PCC). VD, vertical diameter; TD, transverse diameter; SFW, single fruit weight; SSC, soluble solid content; FCTD, fruit core transverse diameter; FCVD, fruit core vertical diameter; FSI, fruit shape index.

Table 2 Prediction accuracy for eight traits with seven methods

| Trait | rrBLUP | BayesA | BayesB | BayesC | BRR  | RKHS | SNN  |
|-------|--------|--------|--------|--------|------|------|------|
| VD    | 0.48   | 0.49   | 0.48   | 0.46   | 0.47 | 0.49 | 0.47 |
| TD    | 0.45   | 0.46   | 0.44   | 0.42   | 0.44 | 0.44 | 0.40 |
| SSC   | 0.60   | 0.59   | 0.58   | 0.57   | 0.60 | 0.57 | 0.56 |
| SFW   | 0.46   | 0.45   | 0.45   | 0.43   | 0.46 | 0.47 | 0.46 |
| FSI   | 0.57   | 0.57   | 0.55   | 0.54   | 0.55 | 0.55 | 0.57 |
| FCVD  | 0.43   | 0.42   | 0.41   | 0.40   | 0.43 | 0.39 | 0.33 |
| FCTD  | 0.50   | 0.52   | 0.50   | 0.50   | 0.50 | 0.48 | 0.48 |
| SCC   | 0.63   | 0.61   | 0.62   | 0.55   | 0.62 | 0.65 | 0.57 |

Note: VD: vertical diameter; TD: transverse diameter; SFW: single fruit weight; SSC: soluble solid content; FCTD: fruit core transverse diameter; FCVD: fruit vertical diameter; FSI: fruit shape index; SCC: stone cell content; BRR: Bayes ridge regression; RKHS: Reproducing kernel Hilbert spaces regression; SNN: Sparse neural networks.

and TD) showed significant ( $P$ -value < 0.01) correlation with fruit core size (measured by FCVD and FCTD).

3.2. Evaluating PA across eight fruit traits with seven prediction methods

A total of seven prediction methods that have been widely used in rice, maize and horticultural crops, were chosen for analyzing the pear data. The seven methods consisted of one parametric method: rrBLUP, four linear, parametric; four Bayesian methods: BayesA, BayesB, BayesC, BRR; one non-linear, non-parametric method: RKHS and one non-parametric neural networks method: SNN.

Five-fold cross-validation was performed using whole training population datasets (310 lines, 178 468 SNPs) to predict fruit size (VD, TD and SFW), core size (FCVD and FCTD), FSI and SSC. Based on the optimal selection models, SSC had the highest PA (0.60), followed by FSI (0.57), FCTD (0.52), VD (0.49), SFW (0.47), TD (0.46) and FCVD (0.43) (Table 2), and the PA of eight traits (except for FCVD) using seven prediction methods are greater than 0.40. The

PA of SCC was 0.65 (278 lines, 178 440 SNPs) using the RKHS method. The performance of these seven prediction methods exhibited slight differences for each of the eight traits: the PA difference between the seven methods varied from 0.03 (VD and FSI) to 0.10 (FCVD and SCC) (Table 2). Compared with five parametric methods, the two non-parametric methods displayed slightly lower performance especially for FCVD (Table 2).

3.3. Effect of marker number and training population size

In order to determine the effect of the number of markers and the necessary number of markers for genomic prediction in a pear population, 13 different sized SNP subsets were randomly selected from the 178 468 SNPs dataset. Five-fold cross-validation was performed for each subset of SNPs. When the number of SNPs increased from 100 to 2 000, a high increase in PA was observed for six traits (VD, TD, SSC, SFW, FSI and FCTD) and for FCVD, the PA slightly increased (from 0.36 to 0.43). The PAs had slightly change when the SNP number increased from 2 000 to 178 468 (Table S5; Fig. 2). For SCC, a high increase of PA (from 0.46 to 0.61) was

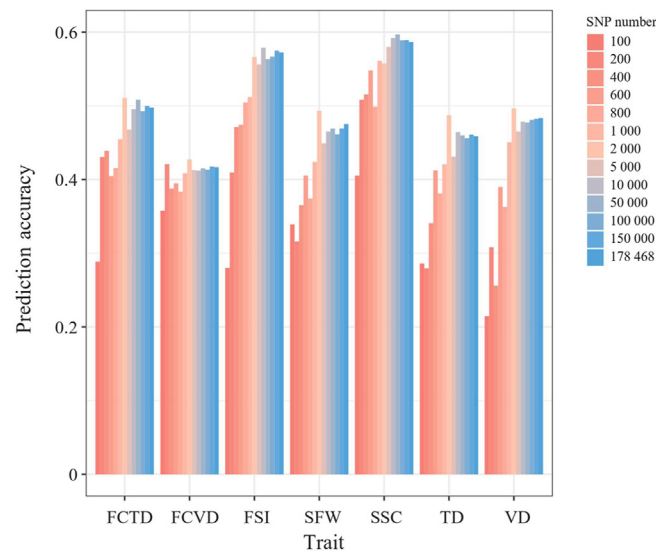


Fig. 2 Effect of SNP numbers on prediction accuracy of genomic selection for seven fruit traits

VD, vertical diameter; TD, transverse diameter; SSC, soluble solid content; SFW, single fruit weight; FSI, fruit shape index; FCVD, fruit core vertical diameter; FCTD, fruit core transverse diameter. Each SNP subset was randomly selected from 178 468 SNPs. The rrBLUP method was used to make prediction.

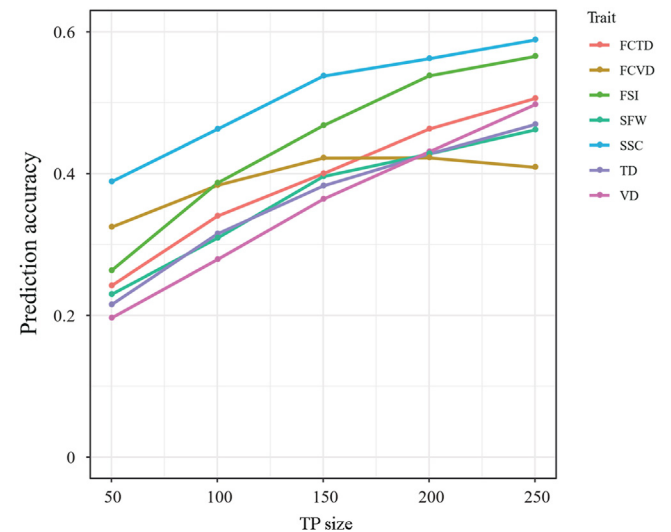
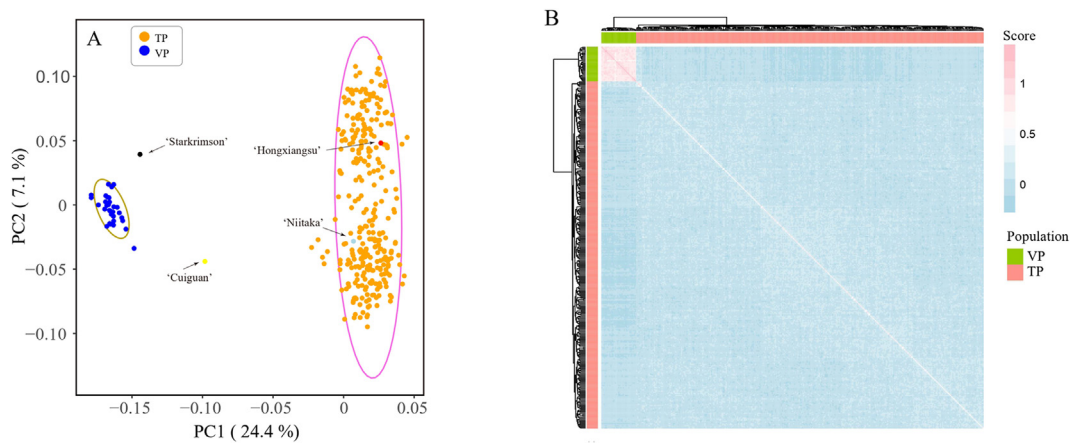


Fig. 3 Effect of training population (TP) size on prediction accuracy of genomic selection for seven fruit traits

VD, vertical diameter; TD, transverse diameter; SSC, soluble solid content; SFW, single fruit weight; FSI, fruit shape index; FCVD, fruit core vertical diameter; FCTD, fruit core transverse diameter.





**Fig. 4 Genetic information between Training population (TP) and validation population (VP)**  
(A) The principal-component analysis for training population and validation population. Orange point represents the individual of TP; blue point represents the individual of VP. (B) Heatmap of the genomic relationship matrix for training population and validation population. Plink bar represents the individual of TP and green bar represents the individual of VP. TP: ‘Niitaka’ × ‘Hongxiangsu’; VP: ‘Cuiguan’ × ‘Starkrimson’.

**Table 3 Prediction accuracy of independent validation with seven methods**

| Trait | rrBLUP | BRR   | BayesA | BayesB | BayesC | RKHS  | SNN   |
|-------|--------|-------|--------|--------|--------|-------|-------|
| VD    | 0.26   | −0.01 | −0.04  | −0.02  | −0.08  | 0.18  | 0.18  |
| TD    | 0.27   | 0.13  | 0.08   | 0.14   | 0.15   | −0.03 | −0.14 |
| SFW   | 0.36   | −0.04 | 0.07   | 0.28   | 0.01   | 0.09  | 0.05  |
| SSC   | 0.34   | 0.34  | 0.38   | 0.39   | 0.33   | 0.49  | −0.12 |
| FSI   | 0.11   | −0.25 | −0.08  | 0.05   | −0.21  | 0.02  | 0.04  |
| FCVD  | 0.21   | 0.29  | 0.23   | 0.23   | 0.21   | 0.21  | −0.01 |
| FCTD  | 0.45   | 0.43  | 0.48   | 0.44   | 0.39   | 0.36  | 0.04  |

Note: VD: vertical diameter; TD: transverse diameter; SFW: single fruit weight; SSC: soluble solid content; FCTD: fruit core transverse diameter; FCVD: fruit vertical diameter; FSI: fruit shape index; BRR: Bayes ridge regression; RKHS: Reproducing kernel Hilbert spaces regression; SNN: Sparse neural networks.

observed when SNP numbers increased from 100 to 2 000, and plateaued when SNP numbers were higher than 2 000 (Fig. S1, a). These results indicate that the PA of all traits reached a plateau point when the SNP numbers increased to 2 000 in this training set.

To evaluate the impact of TP size on the PA of the models, five training population sizes were chosen (50, 100, 150, 200 and 250) (Fig. 3; Table S6), whilst the VP size was held constant with 50 samples for seven traits. With TP size increasing from 50 to 250, the PA increased by 153.54%, 118.32%, 101.18%, 51.53%, 26.09%, 114.99% and 109.34% for VD, TD, SFW, SSC, FCTD, FSI and FCVD, respectively (Table S6). When the TP size increased from 50 to 100, PA dramatically increased (from 18.21% to 46.98%) (Table S7) for seven traits (VD, TD, SFW, SSC, FCVD, FCTD and FSI). The PA of SSC, FCTD and FSI increased slightly (from −4.13% to 10.45%), but high increases were still observed for VD (33.91%), TD (19.28%), SFW (14.71%) and FCVD (17.84%) when the TP size increased from 200 to 250. For SCC, PA increased by 29.58% when TP size increased from 50 to 200, with a smaller but still notable increase (19.15%) when TP size increased from 50 to 100 (Fig. S1, b).

3.4. Assessment of population structure and independent validation

The population structure of TP and VP was evident from the PCA and the additive kinship matrix (Fig. 4, b). The first and the second

principal component explained 24.40% and 7.10% of the variance, respectively (Fig. 4, a). The lack of strong genetic relationships (from −0.27 to 0.14) between the TP and the VP is evidence of large genetic diversity between them (Table S8). The estimated kinship values varied between −0.26 and 1.08 in the TP of NH, and 46.39% of values were lower than 0 (Table S8). In the VP, estimated kinship values ranged from 0.59 to 1.38, indicating that the individuals of the VP are relatively more related (van Raden, 2008).

The TP (NH, 310 lines) was used to predict the VP (CS, 31 lines). The results of independent validation showed that the PA for all seven traits (excluding SCC) were lower compared with the five-fold cross-validation of training population (Table 3). The seven prediction methods performed differently, and no method was suitable for all seven traits. The rrBLUP method performed best for VD (0.26), TD (0.27), SFW (0.36) and FSI (0.11). For SSC, the highest PA was 0.49 using the RKHS method. BRR and BayesA were the best prediction methods for FCVD (0.29) and FCTD (0.48), respectively (Table 3).

4. Discussion

4.1. Five-fold cross-validation

In this research, we evaluated the effect of non-parametric (RKHS and SNN) and parametric (rrBLUP, BayesA, BayesB,

BayesC, and BRR) methods on PA with seven fruit traits. PAs of the seven methods were very similar for all traits except FCVD. This result was consistent with a previous report that indicated a slightly higher PA associated with nine parametric methods compared to three non-parametric methods (Howard et al., 2014).

The effect of increasing the marker density on the PA was assessed with 13 different marker numbers. Increasing the marker number from 100 to 2 000 had an obvious effect on the PAs of all traits, but no meaningful increase in PA was observed when the marker set increased from 2 000 to 178 486. This suggests that a small fraction of the total markers may be sufficient for accurate prediction. This marker number was higher than the plateau point of a previous study using a small population (Heffner et al., 2011), and this number may increase in large TPs (Norman et al., 2018) as increasing the marker numbers might capture more genetic information, and potentially increase PAs. In practice, PA usually reached a plateau with the increasing marker number (Lorenz et al., 2012; Heslot et al., 2014). In order to make use of more markers, the TP size could be increased for obtaining the additional genetic information, otherwise a relatively small TP size would limit any positive effect of the increased marker density (Muir, 2007).

As expected, the TP size had an obvious effect on PA, and it had a relatively larger effect on the PA than either the marker number or the prediction method (Meuwissen, 2009; Riedelsheimer et al., 2013). The PA increased greatly (from 18.21% to 47.00%) when the TP size increased from 50 to 100 and tended to plateau for a TP size of 200. This result was consistent with a study on wheat (Arruda et al., 2015), which showed that PA increased by 11.20% when the training population increased from 96 to 144, reaching a plateau between 192 and 218 individuals. For VD, TD, SFW and FCVD, a larger population size could still increase the PA. In general, increasing the TP size can achieve higher PA by improving the estimation of marker effects (Heffner et al., 2011). Generally, our result supported earlier reports that the TP size, marker density and prediction methods should all be considered, especially TP size, for improving PA.

#### 4.2. Independent validation

The purpose of GS is to use a training set to predict the unobserved breeding value of a selected population. To mimic this scenario, we evaluated PAs using a TP (NH;  $N = 310$  seedlings) to predict fruit traits of another bi-parental population (CS;  $N = 31$  seedlings). Compared with the five-fold cross-validation scheme, the PA of seven traits decreased by different magnitude (from 9.81% to 80.07%, rrBLUP method) in independent validation (Table S9). This may be caused by the lower kinship coefficients between the TP and VP. The TP consisted of F1 hybrids of 'Nuitaka' (*P. pyrifolia*) and 'Hongxiangsu' (*P. sinkiangensis* × *P. bretschneideri*), and the VP comprised first-generation hybrids of 'Cuiguan' (*P. pyrifolia*) and 'Starkrimson' (*P. communis* L.). The genetic distance between the male parents of the TP and VP could have led to the decrease of PA (Lozada et al., 2019). Generally, the PA across populations that have low relatedness is lower in comparison to that with high relatedness or within a population (Pszczola et al., 2012). Lower PAs across populations than within populations have also been observed in

apple (Kumar et al., 2012; Muranty et al., 2015), grapevine (Fodor et al., 2014) and maize (Zhao et al., 2012).

In addition to the low relatedness between the TP and VP, a large decrease (80.07%) in the PA of FSI, compared with VD (44.26%) and TD (39.11%) (Table S9), could have been caused by the trait complexity of FSI. FSI, as the ratio of VD/TD, had more complexity than either VD or TD alone because the variance of FSI is affected by the underlying variability of both VD and TD. The rrBLUP method performed better than other methods for VD, TD, SFW and FSI, whereas the RKHS method that performed best for SSC, and BRR was the best prediction method for FCVD. Generally, rrBLUP is better for highly polygenic traits, while the Bayesian methods showed better performance for traits with major genes (Meuwissen et al., 2001), and RKHS and SNN were more suitable for traits under non-additive genetic architectures (Heslot et al., 2012). Both parametric and non-parametric methods should be tried to cross-confirm the results, if the genetic architecture of agronomic traits are unclear.

#### 4.3. The future of GS in pear

The implementation of GS in pear breeding programs shows promise. For the eight traits sampled in this study, 2 000 markers and 250 samples may be sufficient to get an accurate prediction. Based on the GS of fruit shape, fruit size and SCC in NH population, we can quickly select for round-shaped, large-fruited and low stone cell containing progenies. Phenotypic variance is affected by the outcome of many processes, including genotype, environment and genotype × environment interaction (Millet et al., 2019): environment can significantly affect fruit traits such as vitamin C (San Jose et al., 2014), and sugar and flavonoid content (Jiang et al., 2020). Depending on the magnitude of genotype-by-environment interactions, the PAs of GS can be affected when the training and validation accessions are observed independently in different environmental conditions. To improve stability, repeatability and the genetic gain from GS, continued optimization of GS models by incorporating fixed effects and accumulating high quality data across years and multiple environments would be required for further application of GS in pear. Currently, pear improvement is mostly based on pedigree breeding or selecting candidates by phenotype: these methods have a long breeding cycle and are costly. With the development of high-density pear SNP arrays (Li et al., 2019; Montanari et al., 2019), lots of cheap molecular markers can capture more genetic information in pear breeding populations. As a result, we predict that in the near future, GS will become a reliable and cost-effective means of replacing phenotypic selection in pear breeding (Saito, 2016).

### 5. Conclusions

We used a single F<sub>1</sub> pear hybrid population consisting of 310 seedlings to evaluate the factors affecting the PA of GS. TP size and the number of markers were strongly influenced on PA. Although different methods appeared very similar in their PAs by cross-validation, combining the results of several methods is recommended before the genetic architecture can be confirmed.

Our results would be helpful in designing the TP, select VP, and determine the marker numbers and TP size for GS in pear.

## Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This article was supported by the National Key Research and Development Program (Grant No. 2022YFD1200503), Jiangsu Agricultural Science and Technology Innovation Fund [Grant No. CX(22)3043], the Earmarked Fund for China Agriculture Research System (Grant No. CARS-28), the Earmarked Fund for Jiangsu Agricultural Industry Technology System (Grant No. JATS [2022] 454). We thank the high-performance computing platforms of the Bioinformatics Center of Nanjing Agricultural University for supporting this project. Also, thanks to Cath Kington Vincent Bus and Janine Johnson for helpful comments and suggestions.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.hpj.2023.04.008>.

## REFERENCES

- Arruda, M.P., Brown, P.J., Lipka, A.E., Krill, A.M., Thurber, C., Kolb, F.L., 2015. Genomic selection for predicting fusarium head blight resistance in a wheat breeding program. *Plant Genome*, 8.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30: 2114–2120.
- Claessen, H., Keulemans, W., Van de Poel, B., De Storme, N., 2019. Finding a compatible partner: self-incompatibility in European pear (*Pyrus communis*); molecular control, genetic determination, and impact on fertilization and fruit set. *Front Plant Sci*, 10: 407.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics*, 27: 2156–2158.
- Desta, Z.A., Ortiz, R., 2014. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci*, 19: 592–601.
- Duitama, J., Quintero, J.C., Cruz, D.F., Quintero, C., Hubmann, G., Foulque-Moreno, M.R., Verstrepen, K.J., Thevelein, J.M., Tohme, J., 2014. An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. *Nucleic Acids Res*, 42: e44.
- Endelman, J., 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, 4: 250–255.
- Fodor, A., Segura, V., Denis, M., Neuenschwander, S., Fournier-Level, A., Chatelet, P., Homa, F.A., Lacombe, T., This, P., Le Cunff, L., 2014. Genome-wide prediction methods in highly diverse and heterozygous species: proof-of-concept through simulation in grapevine. *PLoS One*, 9: e110436.
- Hardner, C.M., Hayes, B.J., Kumar, S., Vanderzande, S., Cai, L., Piaskowski, J., Quero-Garcia, J., Campoy, J.A., Barreneche, T., Giovannini, D., Liverani, A., Charlot, G., Villamil-Castro, M., Oraguzie, N., Peace, C.P., 2019. Prediction of genetic value for sweet cherry fruit maturity among environments using a 6k SNP array. *Hortic Res*, 6: 6.
- He, M., Li, L.F., Xu, Y., Mu, J.X., Xie, Z.H., Gu, C., Zhang, S.L., 2022. Identification of S-genotypes and a novel S-RNase in 84 native Chinese pear accessions. *Hortic Plant*, 8: 713–726.
- Heffner, E.L., Jannink, J.L., Sorrells, M.E., 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome*, 4: 65.
- Heffner, E.L., Lorenz, A.J., Jannink, J.L., Sorrells, M.E., 2010. Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci*, 50: 1681–1690.
- Heslot, N., Akdemir, D., Sorrells, M.E., Jannink, J.L., 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet*, 127: 463–480.
- Heslot, N., Yang, H., Sorrells, M.E., Jannink, J.L., 2012. Genomic selection in plant breeding: a comparison of models. *Crop Sci*, 52: 146–160.
- Howard, R., Carriquiry, A.L., Beavis, W.D., 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3-Genes Genom Genet*, 4: 1027–1046.
- Jiang, W., Li, N., Zhang, D., Meinhardt, L., Cao, B., Li, Y., Song, L., 2020. Elevated temperature and drought stress significantly affect fruit quality and activity of anthocyanin-related enzymes in jujube (*Ziziphus jujuba* mill. Cv. 'Lingwuchangzao'). *PLoS One*, 15: e0241491.
- Kumar, S., Chagne, D., Bink, M.C., Volz, R.K., Whitworth, C., Carlisle, C., 2012. Genomic selection for fruit quality traits in apple (*Malus × domestica* Borkh.). *PLoS One*, 7: e36674.
- Kumar, S., Kirk, C., Deng, C.H., Shirliff, A., Wiedow, C., Qin, M., Wu, J., Brewer, L., 2019. Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. *Sci Rep*, 9: 9072.
- Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27: 2987–2993.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25: 1754–1760.
- Li, X., Singh, J., Qi, M., Li, S., Zhang, X., Zhang, M., Khan, A., Zhang, S., Wu, J., 2019. Development of an integrated 200k SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol J*, 17: 1582–1594.
- Lorenz, A.J., Smith, K.P., Jannink, J.L., 2012. Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. *Crop Sci*, 52: 1609–1621.
- Lozada, D.N., Mason, R.E., Sarinelli, J.M., Brown-Guedira, G., 2019. Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genet*, 20: 82.
- Meuwissen, T., 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet Sel Evol*, 41: 35.
- Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819–1829.
- Millet, E.J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F., Tardieu, F., 2019. Genomic prediction of maize yield across European environmental conditions. *Nat Genet*, 51: 952–956.
- Minamikawa, M.F., Nonaka, K., Kaminuma, E., Kajiya-Kanegae, H., Onogi, A., Goto, S., Yoshioka, T., Imai, A., Hamada, H., Hayashi, T., Matsumoto, S., Katayose, Y., Toyoda, A.,



- Fujiyama, A., Nakamura, Y., Shimizu, T., Iwata, H., 2017. Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. *Sci Rep*, 7: 4721.
- Minamikawa, M.F., Takada, N., Terakami, S., Saito, T., Onogi, A., Kajiya-Kanegae, H., Hayashi, T., Yamamoto, T., Iwata, H., 2018. Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (*Pyrus pyrifolia* nakai). *Sci Rep*, 8: 11994.
- Montanari, S., Bianco, L., Allen, B.J., Martínez-García, P.J., Bassil, N.V., Postman, J., Knäbel, M., Kitson, B., Deng, C.H., Chagné, D., Crepeau, M.W., Langley, C.H., Evans, K., Dhingra, A., Troggio, M., Neale, D.B., 2019. Development of a highly efficient axiom™ 70 k SNP array for *Pyrus* and evaluation for high-density mapping and germplasm characterization. *BMC Genom*, 20: 331.
- Muir, W.M., 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet*, 124: 342–355.
- Muranty, H., Troggio, M., Sadok, I.B., Rifai, M.A., Auwerkerken, A., Banchi, E., Velasco, R., Stevanato, P., van de Weg, W.E., Di Guardo, M., Kumar, S., Laurens, F., Bink, M.C., 2015. Accuracy and responses of genomic selection on key traits in apple breeding. *Hortic Res*, 2: 15060.
- Nakaya, A., Isobe, S.N., 2012. Will genomic selection be a practical method for plant breeding? *Ann Bot*, 110: 1303–1316.
- Norman, A., Taylor, J., Edwards, J., Kuchel, H., 2018. Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3-Genes Genom Genet*, 8: 2889–2899.
- Perez, P., de los Campos, G., 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198: 483–495.
- Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J.E.E., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M., Dickinson, T.A., Campbell, C.S., 2007. Phylogeny and classification of Rosaceae. *Plant Systemat Evol*, 266: 5–43.
- Pszczola, M., Strabel, T., Mulder, H.A., Calus, M.P., 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci*, 95: 389–400.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81: 559–575.
- Riedelsheimer, C., Endelman, J.B., Stange, M., Sorrells, M.E., Jannink, J.L., Melchinger, A.E., 2013. Genomic predictability of interconnected biparental maize populations. *Genetics*, 194: 493–503.
- Saito, T., 2016. Advances in Japanese pear breeding in Japan. *Breed Sci*, 66: 46–59.
- San Jose, R., Sanchez-Mata, M.C., Camara, M., Prohens, J., 2014. Eggplant fruit composition as affected by the cultivation environment and genetic constitution. *J Sci Food Agric*, 94: 2774–2784.
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., Atlin, G., Jannink, J.L., McCouch, S.R., 2015. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet*, 11: e1004982.
- Stewart-Brown, B.B., Song, Q., Vaughn, J.N., Li, Z., 2019. Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3-Genes Genom Genet*, 9: 2253–2265.
- Stich, B., Van Inghelandt, D., 2018. Prospects and potential uses of genomic prediction of key performance traits in tetraploid potato. *Front Plant Sci*, 9: 159.
- Van Nocker, S., Gardiner, S.E., 2014. Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. *Hortic Res*, 1: 14022.
- VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. *J Dairy Sci*, 91: 4414–4423.
- Wang, Y., Mi, X., Rosa, G.J.M., Chen, Z., Lin, P., Wang, S., Bao, Z., 2018a. Technical note: an R package for fitting sparse neural networks with application in animal breeding. *J Anim Sci*, 96: 2016–2026.
- Wang, Y., Sun, G., Zeng, Q., Chen, Z., Hu, X., Li, H., Wang, S., Bao, Z., 2018b. Predicting growth traits with genomic selection methods in zhikong scallop (*Chlamys farreri*). *Mar Biotechnol*, 20: 769–779.
- Wiggans, G.R., Cole, J.B., Hubbard, S.M., Sonstegard, T.S., 2017. Genomic selection in dairy cattle: the USDA experience. *Annu Rev Anim Biosci*, 5: 309–327.
- Zhang, P.J., Zhang, H., Du, J.K., Qiao, Y.S., 2022. Genome-wide identification and co-expression analysis of GDSL genes related to suberin formation during fruit russetting in pear. *Hortic Plant J*, 8: 153–170.
- Zhao, Y., Gowda, M., Liu, W., Wurschum, T., Maurer, H.P., Longin, F.H., Ranc, N., Reif, J.C., 2012. Accuracy of genomic selection in European maize elite breeding populations. *Theor Appl Genet*, 124: 769–776.
- Zhou, Z., Zhang, C., Zhou, Y., Hao, Z., Wang, Z., Zeng, X., Di, H., Li, M., Zhang, D., Yong, H., Zhang, S., Weng, J., Li, X., 2016. Genetic dissection of maize plant architecture with an ultra-high density bin map based on recombinant inbred lines. *BMC Genom*, 17: 178.

# Horticulture Research

ISSN 2052-7276 (online)  
ISSN 2662-6810 (print)  
CN 32-1888/S6

[academic.oup.com/hr](http://academic.oup.com/hr)  
[www.hortres.com](http://www.hortres.com)



**Molecular basis of ornamental  
traits in *Clivia miniata***



 **OXFORD**  
UNIVERSITY PRESS

## Review Article

## Pear genetics: Recent advances, new prospects, and a roadmap for the future

Jiaming Li<sup>1,†</sup>, Mingyue Zhang<sup>2,†</sup>, Xiaolong Li<sup>1,†</sup>, Awais Khan<sup>3</sup>, Satish Kumar<sup>4</sup>, Andrew Charles Allan<sup>5</sup>, Kui Lin-Wang<sup>5</sup>, Richard Victor Espley<sup>5</sup>, Caihong Wang<sup>6</sup>, Runze Wang<sup>1</sup>, Cheng Xue<sup>2</sup>, Gaifang Yao<sup>7</sup>, Mengfan Qin<sup>1</sup>, Manyi Sun<sup>1</sup>, Richard Tegtmeier<sup>3</sup>, Hainan Liu<sup>1</sup>, Weilin Wei<sup>1</sup>, Meiling Ming<sup>1</sup>, Shaoling Zhang<sup>1</sup>, Kejiao Zhao<sup>1</sup>, Bobo Song<sup>1</sup>, Jiangping Ni<sup>1</sup>, Jianping An<sup>2</sup>, Schuyler S. Korban<sup>8,\*</sup> and Jun Wu<sup>1,\*</sup>

<sup>1</sup>Center of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

<sup>2</sup>State Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai-An, Shandong 271018, China

<sup>3</sup>Plant Pathology & Plant-Microbe Biology Section, Cornell University, Geneva, NY 14456, USA

<sup>4</sup>Hawke's Bay Research Centre, The New Zealand Institute for Plant and Food Research Limited, Havelock North 4157, New Zealand

<sup>5</sup>The New Zealand Institute for Plant and Food Research Limited, Auckland 1142, New Zealand

<sup>6</sup>College of Horticulture, Qingdao Agricultural University, Qingdao, 266109, China

<sup>7</sup>School of Food and Biological Engineering, Hefei University of Technology, 230009 Hefei, China

<sup>8</sup>Department of Natural Resources & Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

\*Corresponding authors. Email: wujun@njau.edu.cn, korban@illinois.edu

<sup>†</sup>Jiaming Li, Mingyue Zhang and Xiaolong Li contributed equally to this work.

## Abstract

Pear, belonging to the genus *Pyrus*, is one of the most economically important temperate fruit crops. *Pyrus* is an important genus of the Rosaceae family, subfamily Maloideae, and has at least 22 different species with over 5000 accessions maintained or identified worldwide. With the release of draft whole-genome sequences for *Pyrus*, opportunities for pursuing studies on the evolution, domestication, and molecular breeding of pear, as well as for conducting comparative genomics analyses within the Rosaceae family, have been greatly expanded. In this review, we highlight key advances in pear genetics, genomics, and breeding driven by the availability of whole-genome sequences, including whole-genome resequencing efforts, pear domestication, and evolution. We cover updates on new resources for undertaking gene identification and molecular breeding, as well as for pursuing functional validation of genes associated with desirable economic traits. We also explore future directions for “pear-omics”.

## Introduction

As a member of the Rosaceae family and the subfamily Maloideae [1], pear has a wide range of germplasm resources and an ancient cultivation history. There are at least 22 known *Pyrus* species, with over 5000 accessions that are either cataloged or maintained around the world. These accessions have wide morphological and physiological variability and different ecological adaptations. Among these species, *P. communis* is mostly cultivated in Western countries, while *P. pyrifolia*, *P. bretschneideri*, *P. ussuriensis*, and *P. × sinkiangensis* are primarily cultivated in Asian countries. China is a major producer of pear, accounting for 71.40% of world pear production, and supplies approximately 17.60% of the export pear market (<http://www.fao.org/faostat/en/#home>, 2019). However, the current export price of Chinese pears is lower than that of the international average market price. This is because fruits of major Asian pear cultivars have higher stone cell contents, often lack an overall attractive appearance, particularly compared with those of red-colored European pears,

and have relatively bland flavors [2–4]. Traditional pear breeding efforts are mainly dependent on sexual crosses between parental types and the selection of promising or opportunistic seedlings, mutation breeding and selection of improved mutants, and the selection of desirable spontaneous bud mutants. Overall, controlled sexual crosses are most often used in pear breeding.

Pears are often self-incompatible and have long juvenile periods (5–7 years), often leading to the independent segregation of traits of interest in the resultant hybrids from sexual crosses. Overall, traditional pear breeding is difficult, time-consuming, and costly, as it requires the long-term commitment of labor, materials, and land-space resources. Therefore, the availability and use of molecular markers and genomic selection can significantly contribute to overcoming some of these constraints of traditional breeding.

Prior to the release of the pear genome sequence, only a few genes and markers associated with important economic traits were identified. However, once genomic tools and new genetic approaches became available, a

Received: 29 April 2021; Accepted: 25 August 2021; Published: 5 January 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Table 1.** Comparisons among different sequenced pear genomes

| Contigs                              | 'Dangshansuli' | Bartlett1.0 | BartlettDHv2.0 | 'Shanxi Duli' | 'Zhongai 1' |
|--------------------------------------|----------------|-------------|----------------|---------------|-------------|
| Number of contigs                    | 25,312         | 182,196     | 620            | 595           | 1241        |
| Total size of contigs (Mb)           | 501.3          | 507.7       | 501            | 497           | 510.6       |
| N50 contig length (kb)               | 35.7           | 6.6         | 5300           | 1571.5        | 1277.3      |
| Longest contig (Mb)                  | 0.3            | 0.1         | /              | /             | 6.5         |
| <b>Scaffolds</b>                     |                |             |                |               |             |
| Number of scaffolds                  | 2,103          | 142083      | 592            | 139           | 784         |
| Total size of scaffolds (Mb)         | 512            | 577.3       | 496.9          | 532.7         | 510.6       |
| N50 scaffold length (kb)             | 540.8          | 88.1        | 6500           | 28122.4       | 23450       |
| Longest scaffold (Mb)                | 4.1            | 1.2         | /              | 45.5          | 31.9        |
| Anchored size to the chromosome (Mb) | 386.7          | 171.3       | 445.1          | 500           | 506.3       |
| Anchored rate to the chromosome (%)  | 75.5           | 29.7        | 84.2           | 94            | 99.2        |

wide variety of valuable technologies and outcomes have been pursued and achieved in pear, including genetic transformation, genome sequencing, molecular markers, genetic and physical mapping, and comparative genomic analyses. Furthermore, the availability of tools and data resources has offered new opportunities for the efficient and robust discovery of genes controlling desirable fruit quality traits, fruit productivity traits, and postharvest storage life, as well as those that greatly shorten the breeding cycle in pear.

In 2013, the first genome sequence of Asian pear was released [5], thereby allowing delineation of chromosome evolution, genomic structure, and patterns of genetic variations. With the availability of a high-quality pear genome sequence, several gene families involved in controlling desirable and economic traits have now been identified [6–13]; the development of reliable and robust molecular markers [14], new genetic mapping initiatives [15], and analysis of genome evolution [5] have also occurred. This new knowledge and these resources will continue to have significant impacts on efforts for the genetic improvement of pears.

In this review, we will provide a summary of pear genome sequences and highlight the applications of high-throughput technologies that have contributed new knowledge about the biology of pear. These advances have provided new knowledge on the origins of pear and new insights into our understanding of the independent domestication history of Asian and European pears. Furthermore, the construction of high-density genetic linkage maps and the pursuit of genome-wide association studies (GWAS) have contributed to the identification of genomic loci for key genes regulating various economic and desirable agronomic traits. In addition, this review covers developments of multiple omics resources useful for the identification of candidate genes and evaluates the impacts of these findings on future efforts for pear breeding. This roadmap for the pear genome will serve as a useful guide for pursuing genetic improvement efforts toward developing new high-quality and well-adapted pear cultivars. This roadmap is also critical for addressing issues of the pear

response to climate change, as well as of ever-changing consumer demands and preferences.

## Overview of Asian and European pear genome sequences

Pear has a basic chromosome number of 17 ( $2n = 34$ ). Different pear species have distinctly different genome sizes, ranging from 500 to 650 Mb, and possess high numbers of repeats and transposable elements (TEs), as well as high levels of heterozygosity.

The first draft genome sequence of the pear cv. 'Dangshansuli' (*P. bretschneideri* Rehd.), also known as Chinese white pear (Asian pear), was assembled based on a bacterial artificial chromosome (BAC)-by-BAC strategy, thereby alleviating issues of high rates of heterozygosity and the complexity of repeated reads [5]. A total of 2,103 scaffolds were assembled for this pear genome with an N50 of ~0.54 Mb, representing 97.1% (512.0 Mb) of the estimated genome size (527.0 Mb) with 194× genome coverage (Table 1). Based on a high-density genetic map, a total of 386.7 Mb sequences, corresponding to ~75.5% of the assembled genome, were anchored to all 17 chromosomes of the pear genome. Repetitive sequences accounted for 53.1% (271.9 Mb) of the assembled genome. A high long-terminal repeat (LTR) expansion rate suggested that the pear genome was in continuous expansion. Compared with the apple (*Malus × domestica* Borkh.) genome, another member of the subfamily Maloideae, it was proposed that the presence of large numbers of repeat sequences primarily contributed to the size differences between the pear and apple genomes [16]. Importantly, genes involved in the synthesis of stone cells, sugars, and volatile compounds, as well as those involved in disease resistance and self-incompatibility, were identified in the pear genome. However, haplotype features and allele-specific expression were difficult to determine owing to the high rates of heterozygosity, as well as the lack of available accessions with haplotype-derived homologous chromosomes in Asian pears. Subsequently, a haplotype-resolved genome for pear (*P. bretschneideri*) was developed

using a new approach for protoplast isolation from pollen combined with single-cell DNA sequencing of 12 pollen cells and 'barcode' phasing of 38,304 BAC sequences [17]. Thus, the assembled genome sizes of the haploid genomes A and B were 546 Mb and 536 Mb, respectively. The haploid genome assembly also revealed that 8.12% of the genes identified in the first pear reference genome featured mosaic assemblies [17].

European pear, *P. communis* L., also an economically important species that is widely cultivated in Western countries, has distinct phenotypic and fruit quality characteristics that differ from those of Asian pear, including fruit shape, taste, lignin content, and aroma [2–4, 18, 19]. The whole genome sequence and annotation of European pear have aided in pursuing comparative genomic studies with Asian pear [20].

A draft genome sequence of the European pear 'Bartlett' version 1.0 was assembled and released using next-generation sequencing (NGS) technology (Roche 454) [20]. A total of 142,083 scaffolds were assembled, corresponding to 577.3 Mb, and represented 96.2% of the expected 600 Mb of the European pear genome (Table 1) [20]. The number of predicted genes in European pear was higher than that reported for most other plant species, but this number was similar to that identified in Asian pear. This result might be expected due to the incidence of whole genome duplication (WGD) events in members of the Maloideae subfamily [16]. In addition, the predicted coding region length (1,209 bp), exon length, and gene density in the European pear genome were found to be similar to those detected in Asian pear.

Recently, an updated version of the European pear genome using a double-haploid 'Bartlett' cultivar was published [21]. The quality and completeness of the European pear draft genome, designated BartlettDHv2.0, were greatly enhanced by integrating multiple technologies [21]. In this updated draft genome, a total of 496.9 Mb sequences were assembled, and 445.1 Mb were anchored and oriented across all 17 chromosomes of the pear genome using both Hi-C data and a high-density genetic map (Table 1). A total of 50% of the sequences (~247 Mb) in the European pear genome were found to be repetitive sequences, and 37,445 protein-coding genes were annotated, corresponding to a 13% reduction in predicted protein-coding genes from the previous two draft genomes of both *P. communis* and *P. bretschneideri*.

In addition to these three genome sequences of cultivated pears, genomes of the wild pear *P. betulaefolia* Bunge [22] and 'Zhongai 1' (*P. ussuriensis* × *P. communis*), a dwarfing hybrid rootstock [23], were recently sequenced and assembled in ongoing efforts to expand the pool of sequenced pear genomes. Most of the pear genome sequences and annotation-related datasets have been stored in the Genome Database of Rosaceae (GDR: <https://www.rosaceae.org/tools/jbrowse>) [24].

With the development of NGS, many resequencing studies are now underway. These expanded sequencing datasets have been widely used to explore the origin,

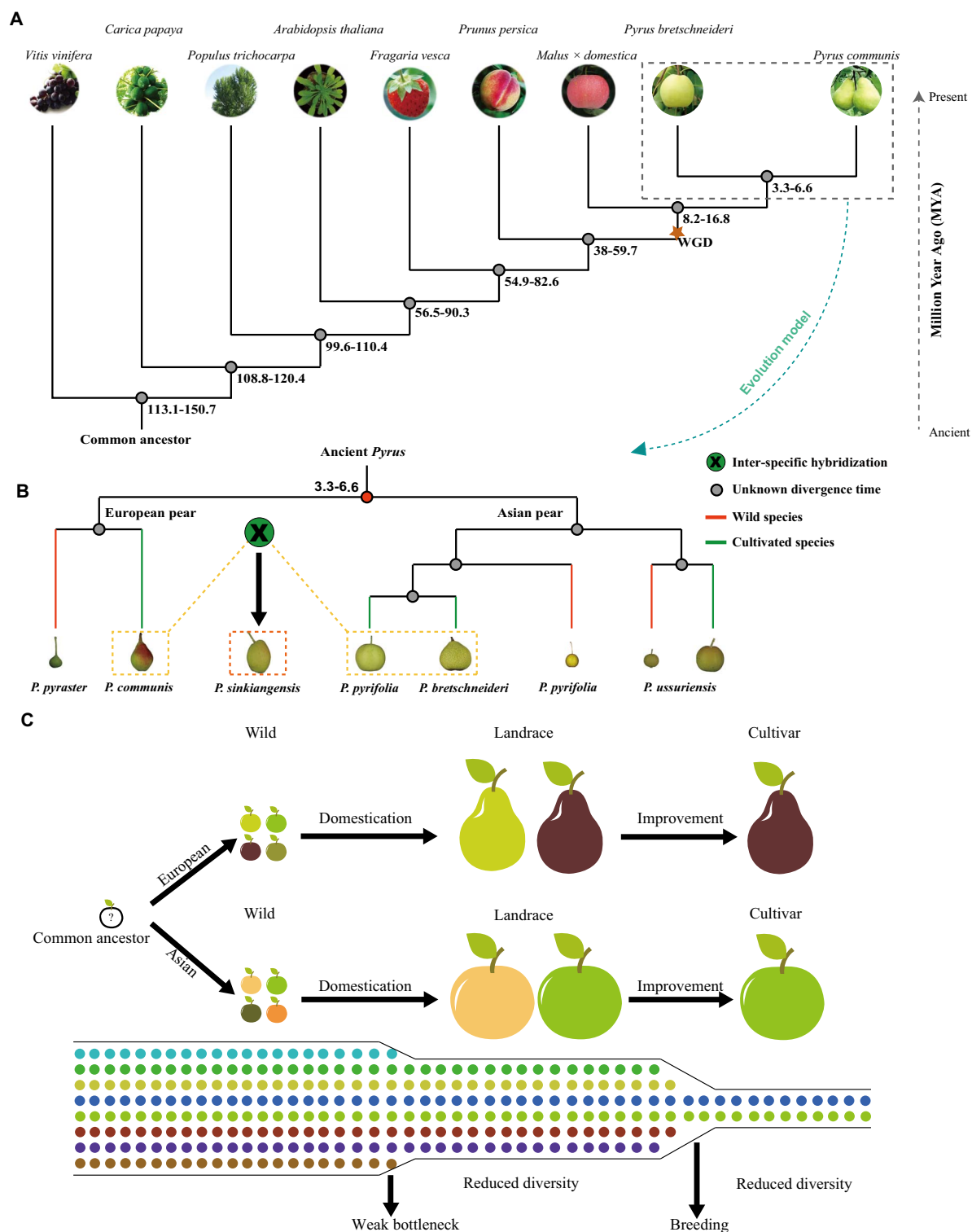
domestication history, and evolutionary processes of various fruit crops, including those for peach [25], grape [26], and apple [27].

Recently, 113 representative *Pyrus* accessions worldwide were resequenced [28]. This study led to the proposal of a new paradigm for pear germplasm dispersion and domestication. In this paradigm, it was suggested that current Asian and European pears originated from the southeast region of China, spread over to central Asia, and were then subsequently dispersed across Asia and Europe. Thus, each of the cultivated Asian and European pears independently underwent domestication in local regions [28]. Due to such independent domestication events, Asian and European pears were subjected to different selection pressures to meet the fruit appearance and taste preferences of different human populations. Generally, Asian pear has a round-shaped fruit with notable features, including crisp flesh, high stone cell content, low acidity, minimal aroma, high sugar, and mild flavor, whereas European pear is characterized by a typical pyriform-shaped fruit with smooth flesh, few stone cells, soft flesh, strong aroma, and strong flavor [2–4, 29]. At the genomic level, a total of 9.29 Mb of the genome, containing 857 genes, was identified to carry selective signatures in Asian pears, whereas a total of 5.35 Mb of genomic regions, containing 248 genes, was identified to carry selective signatures regions in European pears. However, only 47 genes were found to be common between the two pear types, suggesting that these genes were selected upon during the domestication of both Asian and European pears. These findings supported the hypothesis that Asian pears and European pears were independently domesticated in different regions.

## Pear domestication and improvement

Understanding the domestication and genetic improvement of a crop at the whole genome level can aid in future efforts to improve crop yield and other traits of interest.

By using resequencing and population genetic analysis of four different populations of pear (Asian wild, Asian cultivated, European wild, and European cultivated accessions), it was revealed that a weak domestication event occurred in pear [28]. The divergence time of Asian pears from European pears appeared to occur from 3.3 to 6.6 million years ago (MYA) compared to other plant species, including grape (*Vitis vinifera*), apple (*M. × domestica*), peach (*Prunus persica*), woodland strawberry (*Fragaria vesca*), poplar (*Populus trichocarpa*), papaya (*Carica papaya*), and the model plant *Arabidopsis* (*Arabidopsis thaliana*) (Fig. 1A). Furthermore, relationships between five domesticated pear populations and their corresponding wild relatives were delineated, with findings supporting an evolutionary model based on independent domestication of European and Asian pears (Fig. 1B). Specifically, it was demonstrated that S-RNase genes underwent rapid



**Figure 1.** A schematic diagram of divergence time and an evolutionary model of multiple plant species, including *Pyrus*, as well as a model of the domestication of Asian and European pears. (A) Divergence time of nine species, including *Vitis vinifera*, *Malus × domestica*, *Pyrus communis*, *Pyrus bretschneideri*, *Prunus persica*, *Fragaria vesca*, *Populus trichocarpa*, *Carica papaya*, and *Arabidopsis thaliana*. The estimated divergence times (MYA) are inferred based on single-copy orthologous groups and shown at each node. WGD, whole genome duplication. (B) An evolutionary model of wild and cultivated pear species [28]. (C) A schematic representation of the independent domestication process of Asian and European pears along with subsequent modern breeding efforts for cultivar improvement. The domestication process of pear has experienced a weak bottleneck resulting in slightly reduced diversity, while the improvement process shows significantly reduced diversity; thus, new efforts for restoring diversity are extremely urgent in the genetic improvement of pear cultivars.

evolution and balancing selection. Furthermore, the relatively low level of overlap in signatures of selection between Asian and European pears suggested that the

targets of selection differed between the two pear types. Moreover, the independent domestication model of Asian and European pears was also supported by the incidence



of separate groups in the phylogenetic tree, which further demonstrated that the divergence time of Asian and European pears occurred much earlier and prior to any possible human intervention. Notably, genes associated with fruit size, sugars, organic acids, stone cells, and volatile compounds were present in regions with selective sweep signatures (Fig. 1C). To further explore genetic changes at the RNA level that occurred during domestication and improvement, Li et al. [30] used a transcriptome dataset of 41 pear (*P. pyrifolia*) genotypes, consisting of 14 wild, 12 landrace, and 15 improved genotypes, to explore the genetic changes related to the domestication and improvement of pear. It was found that 11.13 Mb of genome sequence carried selective signatures of pear domestication, while 4.04 Mb of genomic regions carried selective signatures of pear improvement (Fig. 1C). Of particular note, several genes related to sugar content, stone cell content and fruit size were located in these selected regions, with some of these genes mapping to previously reported QTLs [30].

## Genome-wide variations

A high degree of diversity can be expected at the genomic level in self-incompatible perennial plants, such as that reported for *Pyrus* [28]. Often, NGS and long-read sequencing analyses detect numerous types of DNA sequence variants, including SNPs, InDels, and structural variations (SVs).

Genome resequencing efforts offer new knowledge pertaining to taxonomic classifications, phylogenetic relationships, evolutionary history, domestication, and genetic resources for pursuing innovative molecular-based breeding efforts. A recent genome-wide variation study involving 113 pear accessions representing all known *Pyrus* species (cultivated and wild, collected from 26 countries) confirmed the delineation of Asian and European pears [28]. This study revealed that *P. × sinkiangensis* was derived from a hybridization event between a cultivated Asian pear and a cultivated European pear.

Although RNA-seq is often used to assess levels of genome-wide gene expression, it also allows for the identification of genomic variants within transcribed coding gene regions. RNA-seq was used for SNP mining of the fruit of five different pear cultivars, including ‘Hosui’, ‘Yali’, ‘Nanguoli’, ‘Kuerlexiangli’, and ‘Starkrimson’ [31]. Subsequently, fruits at the enlarged fruit stage were collected from a group of 41 accessions of *P. pyrifolia* consisting of 15 enhanced genotypes, 12 landraces, and 14 wild accessions and subjected to RNA-seq. This transcriptome analysis led to the identification of 875,319 high-quality SNPs [30]. Based on these SNP data, landrace and wild pears were found to be closely related to each other, while a low level of genetic diversity was observed in the improved cultivar group [30]. Moreover, using nucleotide diversity ( $\pi$ ) and  $F_{ST}$  values, it was found that different selective sweeps occurred during both domestication and improvement. Moreover, the differential expression of

selected genes showed a 20.89% decrease from the wild group to the landrace group, while a 23.13% increase was observed from the landrace group to the enhanced genotype group. This result indicated that diversifying selection might play an important role during the improvement of pear [30].

A reduced-representation genotyping-by-sequencing (GBS) strategy targeting regions flanking restriction enzyme (RE) sites was developed by Elshire et al. [32]. This genotyping approach focuses on DNA sequence polymorphisms around methylation-sensitive RE sites and yields genome-wide polymorphism data across numerous samples. Unlike SNP arrays, GBS involves the simultaneous detection and scoring of SNPs in a population of interest; thus, this approach is free from ascertainment bias. The availability of reference genome sequences for both Asian [5] and European [20] pears has facilitated the implementation of GBS in pear genetic studies. Kumar et al. [33] conducted a GBS study using a single RE (*Bam*HI) to assess genetic diversity in Asian and European pears. Moreover, GBS was found to be useful for constructing a high-density linkage map, fine-mapping QTLs for red skin color [34], and pursuing genomic selection (GS) in interspecific populations of pear [35]. The restriction enzyme *Ape*KI was used to construct high-density linkage maps for Asian [36] and European pears [37]. Recently, a two-enzyme (*Eco*RI and *Nla*III)-based GBS was used to construct a linkage map for Asian pear [38]. Regardless of the enzyme system, GBS poses technical uncertainties that can result in the uneven sequencing of samples, the nonuniform distribution of SNPs, and missing genotype data [39].

With the release of pear reference genome sequences along with the availability of large numbers of SNPs, it has become feasible to design high-density SNP arrays for pear. Recently, a high-density 200K SNP genotyping array was constructed [40]. This array was used to improve the genome assembly, genetic mapping, and GWAS in pear [40]. A 70K Axiom array was also released [41]. Owing to its efficiency, flexibility, high throughput, and low cost, such a SNP array will serve as an important reference tool for GWAS and will be highly useful in pursuing further germplasm improvement and breeding efforts.

## Genetic linkage maps: Construction and mapping of trait-linked loci

Genetic linkage maps are powerful tools for use in investigating and understanding how agronomic traits are inherited from their parents. Prior to the assembly of the first Asian pear genome, studies focused on developing frameworks and encryption of linkage mapping [42–45]. It was not until Yamamoto et al. [46] used an F1 population of two European pear cultivars, ‘Bartlett’ and ‘La France’, that two independent maps for pear consisting of 17 linkage groups and corresponding to the basic chromosome number ( $n=17$ ) were constructed. Subsequently,

with the rapid development of pear genome sequences [5, 20], genome-wide DNA markers, including SNPs and simple sequence repeats (SSRs), were identified and developed. Wu et al. [15] used 3,143 SNPs from restriction-associated DNA sequencing (RAD-seq) to construct the first SNP-based high-density genetic map for pear and then used 98 SSR markers to anchor the corresponding linkage groups. The resulting map consisted of 3,241 markers, spanning 2,243.4 cM with an average distance of 0.70 cM between markers. This map has enhanced the development and analysis of pear genetic maps, along with the identification of QTLs for traits of interest that are useful in marker-assisted breeding (MAB) efforts [15].

Subsequently, linkage maps of different genetic backgrounds were constructed. However, these linkage maps lacked common markers, rendering it difficult to conduct comparative studies and to pursue analysis across different pear populations (Table 2). Li et al. [14] collected all genetic linkage maps of European pears and used common SSR markers across these different maps to merge them into a single integrated consensus map. This map allowed for anchoring a total of 291.5 Mb of the 'Bartlett' v1.0 sequence, exploring genetic structure patterns, conducting comparative studies among different maps, and identifying QTLs [14].

Genetic linkage maps can also be used to localize positions or genetic regions controlling target agronomic traits. Several important pear fruit traits have been located on particular chromosomes. As most traits assessed in linkage mapping populations to date are quantitative in nature, QTLs related to these traits are distributed along several linkage groups (LGs) (Table 2) [15, 47]. For example, QTLs for single fruit weight were identified on LGs 2, 3, 7, 8, 10, 11, 13, and 17. Moreover, a QTL for fruit firmness was identified on LG4, while those for harvest time were identified on LG3 and LG15 in Japanese pear. Another important economic trait is fruit skin color, in particular red-color pigmentation, a QTL for which was first located on LG4 using an F1 population of 'Max Red Bartlett', a red color mutant of 'Bartlett' (Table 2) [48]. Subsequently, map-based cloning was used to identify a red skin-related QTL on the tail end of LG5 associated with the *PyMYB114* gene, which regulates fruit anthocyanin biosynthesis [49]. Furthermore, QTLs for various disease resistance-related traits, such as fire blight [45], pear scab [50], and black spot [42], were identified along different linkage groups (Table 2) [51–65].

## Genome-wide association studies (GWAS) in pear

Genome-wide association studies (GWAS) serve as effective approaches for exploring genome-level genetic architecture(s) and have been widely used to identify genetic variants associated with human-related diseases. These approaches are adopted for use in plant population studies to identify candidate loci associated with complex

traits. Thus, GWAS provides opportunities for identifying candidate genetic loci associated with complex traits in pear and for developing robust molecular markers useful in pursuing MAB to enhance the rapid and accurate development of new cultivars.

The development of GWAS models varies with demands for degrees of accuracy, speed, and population size. A generalized linear model (GLM) can handle multiple variables and takes into account the population structure in the form of a covariance structure defined by the index *Q* [66]. Population structure can result in false-positive genotype-phenotype associations, and PCA is commonly used to quantify and correct for population structure in GWAS. Moreover, a mixed linear model (MLM) is also available for use in GWAS. An efficient mixed-model association (EMMA) [67] and a genome-wide efficient mixed-model association (GEMMA) [68] can serve as variance components and are exact methods, both belonging to MLM. EMMA eXpedited (EMMAX) [69, 70] and genome-wide rapid association using mixed model and regression (GRAMMAR) are approximate methods that allow for fast calculation of GWAS. However, GRAMMAR tends to underestimate associations [71]. Thus, a comprehensive overview of various factors must be taken into consideration, such as target species, quantitative and qualitative traits, population size, accuracy and speed, when different GWAS models are to be used.

For GWAS in perennial plants, such as fruit trees, natural germplasm populations are more readily available than controlled-cross (or hybridized) populations. GWAS has been conducted on various fruit trees, including apricots [72], peaches [73], and apples [27]. Recently, GWAS was conducted using 312 sand pear (*P. pyrifolia*) accessions and identified five loci associated with three fruit phenological traits, as well as 37 loci associated with eight fruit quality traits. Furthermore, a new gene, *PbrSTONE*, controlling stone cell conformation, was identified within these association loci [74].

Therefore, GWAS is a viable and useful approach for identifying QTLs/genes for various critical fruit quality traits and other desirable traits, as well as for contributing to the development of robust tools and resources for pursuing MAB in pear and other long-lived woody perennial trees, including self-incompatible fruit trees.

## Marker-assisted selection (MAS), marker-assisted breeding (MAB), and genomic selection (GS) in pear

As pear has a long generation time (5 to 7 years), traditional pear breeding programs are expensive and time-consuming [75]. Marker-assisted selection (MAS) and marker-assisted breeding (MAB) are deemed viable approaches for the molecular mapping of genes, as well as for pursuing genetic improvement efforts, particularly of long-lived perennial fruit trees, as most economic traits are complex and controlled by QTLs [76].

Table 2. QTL mapping of different agronomic traits in pear

| Population                        | Population size | Marker type                    | Number of markers |      | Map length (cM) |        | Number of linkage groups (LGs) |      | Interval (cM) |      | Traits                 | LGs                    | Reference |
|-----------------------------------|-----------------|--------------------------------|-------------------|------|-----------------|--------|--------------------------------|------|---------------|------|------------------------|------------------------|-----------|
|                                   |                 |                                | Female            | Male | Female          | Male   | Female                         | Male | Female        | Male |                        |                        |           |
| 'Kinchaku' × 'Kosui'              | 82              | RAPD                           | 120               | 78   | 768             | 508    | 18                             | 22   | 4.2           |      | Black spot             | 11                     | 42        |
| 'Bartlett' × 'Hosui'              | 63              | AFLP, SSR                      | 226               | 54   | 949             | 926    | 18                             | 17   | 4.9           |      | Self-incompatibility   | 17                     | 43        |
| 'Passe Crassane' × 'Harrow Sweet' | 99              | SSR, AFLP, MFLP, AFLP-RGA, RGA | 155               | 156  | 912             | 930    | 18                             | 19   | 5.8           |      | Fire blight            | 2a, 2b, 4, 9           | 45        |
| 'Bartlett' × 'Hosui'              | 63              | AFLP, SSR                      | 256               | 180  | 1020            | 995    | 19                             | 20   | 4             | 5.5  | Self-incompatibility   | 17                     | 44        |
| 'Bartlett' × 'Hosui'              | 63              | AFLP, SSR                      | 447               |      | 1000            |        | 17                             |      | 2.3           |      | -                      | -                      | 46        |
| 'Hosui' × 'La France'             | 55              | AFLP, SSR                      | 414               |      | 1156            |        | 17                             |      | 2.8           |      | -                      | -                      | 46        |
| 'Abbé Fétel' × 'Max Red Bartlett' | 95              | MFLP, SSR                      | 123               | 110  | 908.1           | 879.8  | 18                             | 19   | 7.4           | 8    | Pear scab              | 3, 7                   | 50        |
| 'Abbé Fétel' × 'Max Red Bartlett' | 95              | MFLP, SSR                      | 123               | 110  | 908.1           | 879.8  | 18                             | 19   | 7.4           | 8    | Skin color             | 4                      | 48        |
| 'Hosui' × 'La France'             | -               | AFLP, SSR                      | 484               |      | 1204            |        | 17                             |      | 2.49          |      | Pear scab              | 2,11,14                | 52        |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Self-incompatibility   | 17                     |           |
| 'Bartlett' × 'Hosui'              | -               | AFLP, SSR                      | 504               | 356  | 1080            | 1180   | 17                             | 16   | 2.1           | 3.31 | -                      | -                      | 53        |
| 'Bartlett' × 'Hosui'              | 63              | AFLP, SSR                      | 335               |      | 1174            |        | 17                             |      | 3.5           |      | -                      | -                      | 53        |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Leaf width             | 10, 15                 |           |
| 'Yali' × 'Jingbaili'              | 145             | AFLP, SSR                      | 402               |      | 18              |        | 1395.9                         |      | 3.8           |      | Leaf length/width      | 5                      | 54        |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Leaf length            | 8, 15, 16              |           |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Petiole length of leaf | 4, 15                  |           |
| 'Abbé Fétel' × 'Max Red Bartlett' | -               | SSR                            | -                 |      | -               |        | -                              |      | -             |      | Red skin color         | 9                      | 55        |
| 'Nittaka' × 'Suyhangri'           | 94              | RAPD, AFLP, SSR                | 106               | 122  | 1006            | 1168   | 19                             | 19   | 9.5           | 9.6  | -                      | -                      | 56        |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Fruit weight           | 2, 7, 8, 10            |           |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Fruit diameter         | 10, 15                 |           |
| 'Bayuehong' × 'Dangshansuli'      | 97              | AFLP, SRAP, and SSR            | 214               | 122  | 1352.7          | 1044.3 | 17                             | 17   | 6.3           | 8.6  | Fruit length           | 7 (two years), 8       | 47        |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | Fruit shape index      | 1, 2 (two years), 7, 8 |           |
|                                   |                 |                                |                   |      |                 |        |                                |      |               |      | SSC                    | 2, 5, 6                |           |
| 'Red Bartlett' × 'Nanguo pear'    | 74              | SRAP                           | 103               | 105  | 602.2           | 650    | 20                             | 20   | 4.9           | 5.2  | Fruit maturity date    | 8 (two years)          | 57        |
| 'Bartlett' × 'Hosui'              | 63              | SSR, SNP                       | 485               |      | 965             |        | 17                             |      | 2             |      | -                      | -                      | 58        |
| 'Hosui' × 'La France'             | 55              | SSR, SNP                       | 370               | 415  | 1160            | 1177   | 17                             | 20   | 3.1           | 2.8  | -                      | -                      | 58        |
| PEAR1 × PEAR2                     | 143             | SSR, SNP                       | 250               | 314  | 1132.3          | 1136.8 | 17                             | 17   | 4.53          | 3.62 | Pear scab              | 2, 5, 7, 10, 17        | 59        |

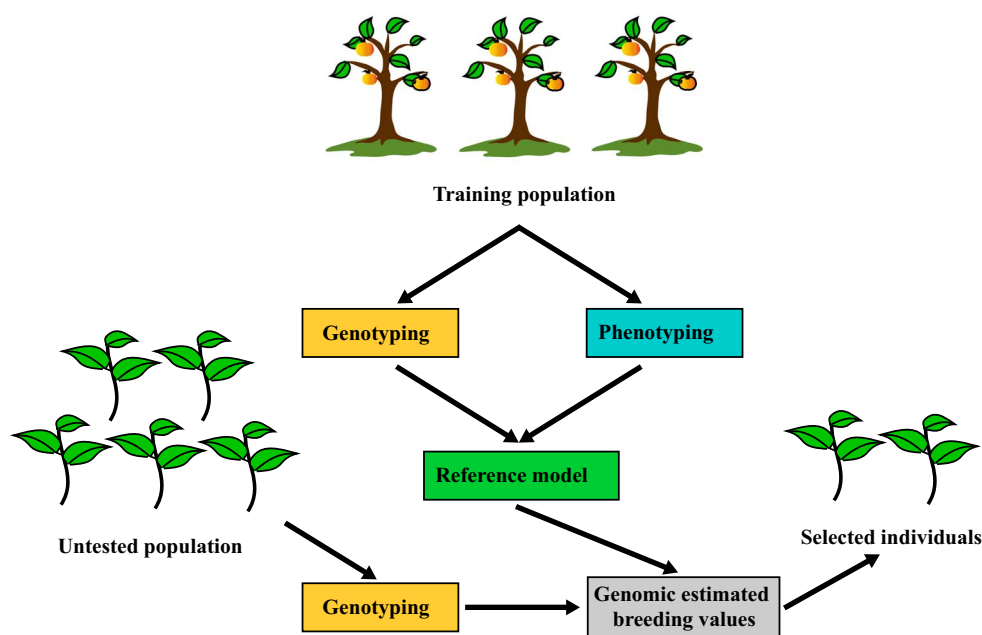
(Continued)



Table 2. Continued

| Population                        | Population size | Marker type  | Number of markers |      | Map length (cM) |        | Number of linkage groups (LGs) |      | Interval (cM) |      | Traits  | LGs   | Reference |
|-----------------------------------|-----------------|--------------|-------------------|------|-----------------|--------|--------------------------------|------|---------------|------|---|---|-----------|
|                                   |                 |              | Female            | Male | Female          | Male   | Female                         | Male | Female        | Male |   |   |           |
| 'Akiakari' × 'Taihaku'            | 93              | SSR, EST-SSR | 208               | 275  | 799.1           | 1039.1 | 17                             | 17   | 3.84          | 3.78 | Fruit skin color<br>Fruit weight<br>Total soluble solids content<br>Firmness<br>Preharvest fruit drop<br>Harvest time | 8 (two years)<br>3, 11<br>4, 8<br>4 (two years)<br>1, 15 (two years)<br>3 (two years), 15 (two years) | 60        |
| 'Bayuehong' × 'Dangshansuli'      | 56              | SSR          | 734               |      | 1661.4          |        | 17                             |      | 2.3           |      | -   | -   | 61        |
| 'Aihuali' × 'Chili'               | 215             | SSR, SLAF    | 12                |      | -               |        | 1                              |      | -             |      | Dwarf growth habit  | 16  | 62        |
| 'Red Clapp's Favorite' × 'Mansoo' | 161             | SSR, SLAF    | 4797              |      | 2703.6          |        | 17                             |      | 0.56          |      | -   | -   | 63        |
| Nine published maps               |                 | SSR, SNP     | 5085              |      | 3266            |        | 17                             |      | 0.6           |      | -   | -   | 14        |
| 'Spadona' × 'Harrow Sweet'        | 162             | SNP          | 2036              |      | 1433            |        | 17                             |      | 0.70          |      | Vegetative budbreak time  | 5, 8, 9, 13, 15, 17   | 37        |
| 'Mantianhong' × 'Hongxiangsu'     | 345             | SNP          | 2606              |      | 1847            |        | 17                             |      | 0.71          |      | -   | -   | 64        |
| 'Bayuehong' × 'Dangshansuli'      | 102             | SSR, SNP     | 3241              |      | 2243.4          |        | 17                             |      | 0.7           |      | Skin color<br>Single fruit weight<br>Transverse diameter<br>Vertical diameter<br>Soluble solid content<br>Flesh color | 4, 13, 16 (two years)<br>13, 17<br>3, 11, 17<br>11, 17 (two years)<br>5, 10, 14<br>9 (two years)      | 15        |
| 'Yuluxiang' × 'Mantianhong'       | 162             | SNP          | 2489              |      | 1668            |        | 17                             |      | 0.67          |      | Skin smooth<br>Length of pedicel<br>Calyx status<br>Juice content<br>Number of seeds                                  | 2, 17<br>2, 14, 17<br>6 (two years)<br>1, 5<br>5 (two years), 9, 14, 17 (two years)                   | 64        |
| 'E1 Dorado' × 'Potomac'           | 81              | SNP          | 27052             |      | 1034.97         |        | 17                             |      | 0.82          |      | Fire blight   | 2   | 65        |
| 'Old Home' × 'Bartlett'           | 100             | SNP          | 29703             |      | 1147.22         |        | 17                             |      | 0.63          |      | Fire blight   | 2   | 65        |
| 'NJA2R59T6' × 'Bartlett'          | 82              | SNP          | 37182             |      | 1080.5          |        | 17                             |      | 0.59          |      | Fire blight   | 2   | 65        |

Note: AFLP, amplified fragment length polymorphism; cM, centi-Morgan; EST, expressed sequenced tag; LGs, linkage group; MFLP, microsatellite-anchored fragment length polymorphism; RAPD, random amplified polymorphic DNA; RGA, resistance gene analog; SLAF, specific locus amplified fragment; SNP, single nucleotide polymorphism; SRAP, sequence-related amplified polymorphism; and SSR, simple sequence repeat.



**Figure 2.** A pipeline for genomic selection

Early use of molecular marker systems included SSRs, amplified-fragment length polymorphisms (AFLPs), sequence-characterized amplified regions (SCARs), random amplified polymorphic DNAs (RAPDs), and cleaved amplified polymorphic sequences (CAPS) [15, 77–79]. All these molecular markers relied on the use of gel electrophoresis methods for screening individuals, seedlings, selections, and germplasm accessions. The release of whole genome sequences and the availability of robust SNP markers have contributed to accelerated progress in breeding programs. For example, high-quality and high-density linkage maps combined with molecular markers have allowed for fine-mapping of the region around a gene for susceptibility to black spot disease, as well as for identifying QTLs for 11 pear fruit-related traits [80].

High-throughput and accurate SNP-based markers, SNP arrays, and GBS can assist in pursuing early trait-predictive assays [37, 81]. DNA markers flanking major causal loci, such as those for red pear fruit skin color, can be used for selection, but a traditional MAS scheme is not well suited for complex traits that are controlled by several loci. GS is a form of MAS that utilizes thousands of genome-wide markers simultaneously to calculate their associations with trait phenotypes in a training population and to predict the genomic estimated breeding values (GEBVs) of individuals in a tested population [82]. The GEBVs of selection candidates are estimated solely based on SNP genotypes and estimated SNP effects. Hence, outstanding candidates can be identified at very early stages of development prior to phenotyping, thereby reducing generation intervals and increasing breeding efficiency (Fig 2). GS is best suited for polygenic traits, and high-density genotyping is essential for the application of GS to ensure that all QTLs are in population-wide linkage disequilibrium with SNP markers.

In a preliminary study to assess the utility of GS in a sand pear (*P. pyrifolia*) breeding program, Iwata et al. [83] screened 76 cultivars using 162 genome-wide markers and found that genome-wide predictions for GS were highly accurate (0.75) for harvest time and moderately accurate (0.38–0.61) for resistance to black spot, firmness of flesh, fruit shape (longitudinal section), fruit size, acid content, and number of spurs. Minamikawa et al. [84] used an Illumina Golden Gate genotyping assay comprising 1536 SNPs to conduct genomic predictions of fruit phenotypes in a *P. pyrifolia* population consisting of 86 cultivars and 765 seedlings. They found that the accuracy of genomic prediction was further enhanced when full-sib family data of a target family were available and suggested that phenotypic data collected in a breeding program were useful for pursuing GWAS and GS when these methods were combined with genome-wide marker data. Kumar et al. [35] evaluated the genetic architecture of 10 fruit phenotypes (including sensory traits) and used GBS to assess the potential of GS in 550 hybrid seedlings.

The above studies have demonstrated that the accuracy of genomic predictions in pear is moderate, particularly as most fruit quality traits are complex and polygenically controlled. Therefore, additional studies must be conducted to develop larger genotype-phenotype datasets to further improve the prediction accuracy of the fruit phenotypes of untested seedlings.

## Multiple omics: Identifying genes related to important traits

### Candidate gene mining at the whole-genome level

As pear genomes have become available, the ability to identify gene families associated with desirable traits or

traits of interest has improved. Recently, studies have focused on identifying genes associated with fruit quality traits, such as those controlling skin color, sugar content, stone cells, and aromatic compounds. Members of the MADS-box gene family, MYB gene family, and lateral organ boundary domain (LBD) gene family were identified as candidate genes associated with anthocyanin biosynthesis in pear [8, 10, 85]. Sugar transporters, a SWEET gene family, the hexokinase gene family, and the phosphofructokinase gene family were identified as candidate genes associated with sugar biosynthesis [6, 9, 86, 87]. For the stone cell trait, MYB gene family, BZR gene family, and KNOX gene family were reported to be involved in stone cell conformation and related lignin biosynthesis in pear [8, 88, 89]. Flavor is a more complicated trait, and some candidate pear genes were identified as candidate genes contributing to aromatic compounds present in pear fruit [90, 91].

### Transcriptomes

A transcriptome, usually generated by RNA-seq, corresponds to all transcripts within a single cell or a population of cells at either a particular developmental stage or a physiological state [92].

The release of large-scale transcriptomic data for pear, including those for different cultivated species, developmental stages, tissues, and treatments for specific trait investigations, has facilitated the exploration of functional genes. Transcriptomes of 'Dangshansuli', 'Nanguoli', 'Yali', 'Hosui', 'Kuerlexiangli', 'Starkrimson', and *P. pyrifolia* accessions at key developmental stages were released, and these provided insights into biological processes underlying fruit quality traits, including stone cell contents, sugars, and volatiles [31]. In addition, gene expression during four developmental stages of pear pollen provided opportunities for studying the growth and cessation of pear pollen tubes [93]. Furthermore, transcriptome data from seven tissues were exploited to track the expression patterns of gene pairs between two pear subgenomes, thus providing critical evidence for unbiased subgenome evolution following palaeopolyploidization in pear and serving as valuable resources for investigating tissue-specific gene expression in pear [94].

In pursuit of specific pear traits, transcriptome analysis studies have focused on color pigment development, dormancy, and biotic and abiotic stresses. Transcriptomes of 'Starkrimson' pear versus its green color mutant, bagged versus unbagged 'Pingguoli', color-fading 'Red Bartlett' versus non-color-fading 'Starkrimson', russet- versus green-pericarp individuals of the 'Qingxiang' × 'Cuiguan' F1 group, and the color mutant 'Red Zaosu' versus 'Zaosu' were compared. As a result, several key candidate genes, including *LAR*, *ANR*, *Myb4-like1*, *myb4-like2*, *CCR*, *CAD*, and *PpBBX24*, were identified as candidates for color pigmentation [95–99]. Bud dormancy is a key developmental process for perennial plants to survive under adverse environmental conditions. When the transcriptomes of endodormant and ecodormant

Japanese pear ('Kosui') flower buds were subjected to RNA-seq, it was found that phytohormones, such as ethylene, were involved in endodormancy release [100]. Several studies were conducted to investigate cold, drought, salt stress, and black spot disease in pear, yielding critical datasets that would facilitate the identification and functional analysis of candidate genes [101–106].

### Proteomics

Proteomics is a tool for investigating proteins associated with gene expression and for delineating biochemical networks. In pear, a total of 1810 proteins were identified to be involved during the three stages of pear fruit development [107]. Moreover, 35 differentially expressed proteins related to fruit quality were identified, including three proteins related to sugar formation, seven proteins related to aroma synthesis, and 16 proteins related to lignin formation [107]. In another study, a total of 2841 proteins were identified during seven different pear fruit development stages, and it was suggested that invertases do not play a major role in sugar conversion in developing pear fruit; rather, it was likely that sucrose was broken down by sucrose synthases [108]. Furthermore, several putative sugar transporters from diverse gene families demonstrated developmental regulation [108].

To delineate the mechanism of fruit maturity of an early-maturing bud sport of 'Zaosu' pear, 75 differentially expressed protein spots were identified between an early-maturing bud sport and its original cultivar 'Zaosu' using a combination of 2-DE and MALDI-TOF MS technology [109]. Most of these differentially expressed proteins were closely associated with maturation, thus elucidating the maturation process in these pear genotypes [109].

### Metabolomics

Metabolomics deals with the analysis of large numbers of metabolites within a single cell, tissue, or organ in response to various conditions or treatments. In pear fruit, small molecule metabolites play important roles, such as those involved in basic metabolism; however, there are currently limited metabolomics studies on pear.

In early studies, the levels of glucose, fructose, sorbitol, and sucrose were determined in 10 pear accessions, while the levels of organic acids such as citric acid and malic acid were analyzed in 98 pear accessions using HPLC [3]. More than 100 volatile compounds could be identified in pear fruit [18, 19]. Recently, profiles of volatile compounds related to aroma were analyzed and compared in both 'Dangshansuli' and 'Nanguo' [110]. It was reported that fatty acid-related aroma volatiles were largely derived from metabolic precursors [110]. As wax is another important postharvest trait for pear, the metabolic profiles of wax compounds of 35 pear accessions from five pear species were evaluated. A total of 146 wax compounds were detected, including those



composed of fatty acids, esters, primary alcohols, and terpenoids [111].

### Small RNAs

MicroRNAs (miRNAs) are a class of 21- to 24-nucleotide noncoding RNAs that play important roles in growth and development [112]. As a result of NGS and high-throughput sequencing (HTS), many miRNAs associated with economic traits were detected in pear [113–115]. A particular focus was on pear fruit development and its regulation by miRNAs. Wu et al. [113] conducted deep sequencing of small RNAs from different developmental stages of ‘Dangshansuli’ fruit and found that miR160 acted as a regulator of an auxin response factor during fruit development [113]. Ma et al. [114] conducted an integrated analysis of RNA-seq and small RNA sequencing dataset from tissues of calyx abscission zones of ‘Korla fragrant pear’ and found that miRNAs miR858b and miR160a-3p were involved in calyx abscission [114]. In addition, miRNAs are also likely linked to pear fruit quality development. Using an integrated analysis of small RNAs and degradome sequencing of fruit skin tissues from ‘Meirensu’ pear combined with biochemical analysis, miR156 and its targeted *SQUAMOSA-promoter binding like* (SPL) gene family were found to participate in pear anthocyanin biosynthesis [115]. In another study, Wu et al. [113] identified nine sugar and acid metabolism-associated miRNAs (including miR1132, miR5077, and miR396b) and 11 lignin biosynthesis-related miRNAs (including miR397a, miR395a, and miR408b\*) following small RNA sequencing of pear fruit collected at different developmental stages [113]. In a subsequent study, it was demonstrated that *PbrmiR397a* reduced the lignin content by silencing three *PbrLAC* genes [116]. Other miRNAs (*pyr-miR1809* and *pyr-miR144-3p*) targeted *LAC* genes [117]. Several studies identified miRNAs associated with flavor and aroma in both ‘Suli’ and ‘Nanguoli’ pears, such as miR172, miR395, miR1132, and miR5077 [113, 118, 119]. Furthermore, biotic/abiotic stress-associated miRNAs were identified in pear through HTS, such as those involved in Apple Stem Grooving Virus (ASGV) defense, among other stress responses, including *pbr-miR164*, *pbr-miR399*, *pbr-miR168*, and *pbr-miR171* [118, 120, 121]. Interestingly, a pear-specific miR6390 that might promote dormancy release by degrading DAM (dormancy-associated MADS-box) was identified using small RNA and degradome sequencing [122].

### DNA methylation

DNA methylation is an important epigenetic modification that participates in diverse biological processes. Whole-genome bisulfite sequencing (WGBS or BS-Seq) or integrated analyses with RNA-seq are exploited in Rosaceae plants, and a Methylation Database for Rosaceae (MDR, <http://mdr.xieslab.org/>) has been established. As a result, various desirable traits associated with differentially methylated regions (DMRs) and DMR-regulated genes have been identified [123, 124].

DNA methylation plays essential roles in fruit size development and fruit ripening [125, 126]. Additionally, methylation plays an important role in regulating anthocyanin accumulation. It is well known that MYB TFs play important roles in anthocyanin biosynthesis in Rosaceae plants [127, 128] and that methylation levels of MYB promoters also influence anthocyanin biosynthesis. For example, methylation levels of the *PcMYB10* promoter were associated with the green-skin sport of the pear cv. Max Red Bartlett, where it was observed that higher methylation of the *PcMYB10* promoter resulted in lower expression levels of *PcMYB10* and *PcUGT* [128]. On the other hand, it was observed that the promoter *PyMYB10* had lower levels of methylation in anthocyanin-rich tissues in the red sport of ‘Zaosu’ pear and its red-striped pigmentation patterns [127].

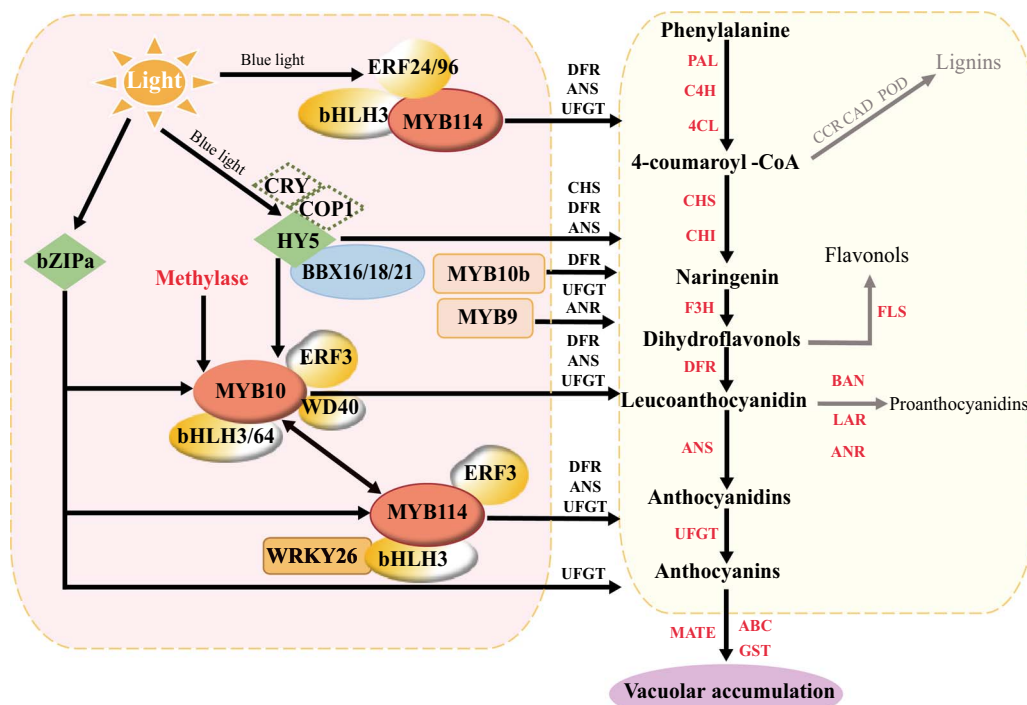
### Stone cells

Stone cells are sclerenchyma cells that develop following cessation of cell expansion in pear fruit, wherein secondary cell walls are deposited along primary walls. These stone cells can result in a gritty texture and poor fruit taste and have a negative impact on consumer satisfaction. Moreover, lignin and cellulose are major components of the secondary cell walls of stone cells [129]. Lignin biosynthesis and deposition are highly correlated with the formation of stone cells within the flesh of pear fruit [130, 131]. The component of lignin in pear stone cells consists primarily of guaiacyl-lignin (G-lignin), along with small amounts of syringyl-lignin (S-lignin) and p-hydroxyphenyl lignin (H-lignin) [132, 133]. It is important to point out that the dominant monolignols are mainly bonded with  $\beta$ -O-4 linkages [133].

Currently, stone cell contents in pears are most often evaluated using a frozen-HCl protocol. This protocol was used to determine the stone cell contents of 236 accessions of sand pear (*P. pyrifolia*) at 50 days after bloom [134]. However, this method is limited due to the incomplete separation of stone cells from the flesh of pear fruits and it is both awkward and time-consuming to perform. Recently, a PearProcess software program was developed based on an imaging protocol using computer vision to digitize images, and image processing algorithms were applied [135].

In the lignin pathway (Fig. 3), a series of enzymes are involved in the synthesis of three monolignols (G-, S-, and H-type lignin) from phenylalanine, such as phenylalanine ammonia-lyase (PAL) and cinnamate-4-hydroxylase (C4H) [136, 137]. These monolignols are polymerized into lignin, which is catalyzed by either peroxidase (PRX) or laccase (LAC) enzymes [138]. Among these enzymes, members of the 4CL, CCR, CAD, OMT, PRX, and LAC gene families have been identified [5, 139–143]. The expression levels of *Pb4CL1*, *PbCCR1*, *PbCCR2*, *PbCCR3*, *PbCAD2*, *PbCCOMT1*, *PbCCOMT3*, five *PbPRXs*, and five *PbLACs* were found to be consistent with changes in the lignin content during pear fruit development [139–143]. Several transcription factors are predicted as





**Figure 4.** A proposed mechanism for fruit skin color development in pear. ANS: anthocyanidin synthase; bZIP: basic region/leucine zipper; 4CL: 4-coumarate coenzyme A ligase; ABC: ATP-binding cassette transporter; ANR: anthocyanidin reductase; AP2/ERF: ethylene response factor/pathogenesis-related transcription factor; BAN: BANYULS; BBX: B-Box; bHLH: basic-helix-loop-helix; C4H: cinnamate 4-hydroxylase; CAD: cinnamyl alcohol dehydrogenase; CCR: cinnamoyl Co A reductase; CHI: chalcone isomerase; CHS: chalcone synthase; COP1: CONSTITUTIVE PHOTOMORPHOGENIC 1; CRY: cryptochrome; DFR: dihydroflavonol 4-reductase; F3H: flavanone 3-hydroxylase; FLS: flavonol synthase; GST: glutathione S- transferase; HY5: ELONGATED HYPOCOTYL 5; LAR: leucoanthocyanidin reductase; MATE: multidrug and toxic compound extrusion; MYB: v-myb avian myeloblastosis viral oncogene homolog; PAL: phenylalanine ammonialyase; POD: peroxidase; UFGT: UDP-glucose; flavonoid-3-O-glucosyltransferase; WD40: WD repeat protein; and WRKY: WRKYGQK domain.

PyMYB114 regulated anthocyanin biosynthesis by forming a complex with PybHLH3 and PyERF3 in Chinese red pears (Fig. 4) [49]. In 'Hongzaosu' pear, both PpERF24 and PpERF96 interacted with PpMYB114 to regulate blue light-modulated anthocyanin biosynthesis (Fig. 4) [157]. In addition, PyWRKY26 interacted with PybHLH3 to cotarget the promoter of PyMYB114 to regulate anthocyanin biosynthesis (Fig. 4) [156]. As light is essential for anthocyanin biosynthesis, Tao et al. [158] found that a blue-light signal transduction module, CRY-COP1-HY5, regulated anthocyanin biosynthesis in red pear [158]. Later, it was found that PybZIPa responded to light and promoted anthocyanin accumulation by activating PyUFGT in Asian pears [154]. Moreover, PyBBX16, PpBBX18 and PpBBX21 antagonistically regulated anthocyanin accumulation through competitive interactions with PpHY5 (Fig. 4) [159, 160].

### Sugar and acid contents

Among the various factors influencing fruit quality, sugar is key in determining flavor. The major soluble sugars in ripe pear fruit are glucose, fructose, sucrose, and sorbitol [3]. Sorbitol, a major photosynthetic product in pear, is the predominant carbohydrate that is translocated in the phloem [161] and unloaded into the fruit via the apoplasm [162]. Monosaccharide transporters mediate the transport of a variety of

sugars. Thus far, a total of 18 hexose transporters (STPs), six tonoplast monosaccharide transporters (TMTs or TSTs), six plastidic glucose transporters (pGlcTs), and six inositol transporters (INTs) have been detected in pear [6]. However, only PbTMT4 was experimentally demonstrated to participate in sugar accumulation in vacuoles (Fig. 5) [163].

Six sucrose transporters (SUTs) were identified in the pear genome [6]. Among these, PbSUT2 was found to influence the sucrose content in sinks, as well as the overall sugar content during flowering (Fig. 5) [164]. Oura et al. [165] reported that the rate of fruit growth depended on both the capability of the fruit to accumulate sorbitol and the rate of sorbitol conversion [165]. Sorbitol metabolic enzymes include NAD-sorbitol dehydrogenase (NAD-SDH), sorbitol-6-phosphate dehydrogenase (S6PDH), and sorbitol oxidase (SOX). Proteins and full-length cDNAs of both SD6PH and NAD-SDH in pears were purified and determined, respectively (Fig. 5) [165, 166].

Sucrose metabolic enzymes belong to the following two classes: sucrose phosphate synthase (SPS) and sucrose synthase (SS). It was previously reported that SPS and SS were involved in sucrose synthesis in pear, and with sucrose accumulation, the activities of both SPS and SS were found to increase [167]. Moreover, hexokinase (HXK) is a rate-limiting enzyme in glycolysis and controls





IAA amide during pear fruit ripening; moreover, GH3.1 expression is activated by ERFs, thereby reducing the IAA content. Additionally, an actin-related protein, ARP4, is involved in ethylene-mediated fruit ripening, and it may work with pectin methylesterase 1 (PME1) to regulate the pear fruit ripening process [195]. Fruit softening serves as an important index for pear fruit ripening, with polygalacturonase (PG)1 and PG2 playing critical roles in the rapid softening of 'Starkrimson' pear fruits [196].

In 'La France' pear, PcPG3, PcPME1, PcPME2, PcPME3, PcGAL1, and PcCel2 were proposed to be involved in fruit softening during storage [197, 198]. Moreover, after treatment of 'Nanguo' pear fruit with the chemical compound 1-methylcyclopropene (1-MCP), an ethylene inhibitor, ethylene biosynthesis (ACS1, ACS4, and ACO), reception (ERS1a), and response (PG1 and PG2) genes were inhibited by 1-MCP treatment. Furthermore, the -expansin-encoding genes PcExp2, PcExp3, PcExp5, and PcExp6 were also upregulated during 'La France' pear fruit ripening, and their expression patterns were consistent with the rates of softening [199].

### Biotic and abiotic stress resistance

Biotic stress is one of the critical factors responsible for decreased fruit quality, yield gaps, tree losses, and increased production costs in pear. The major modes of mitigation for these biological stressors include chemical treatments, pruning, planting disease/pest-resistant cultivars, and physical barriers, such as netting. Often, pests and diseases that attack pear have a wide range of infection strategies and virulence mechanisms that are constantly evolving to adapt to their host. This renders effective and long-term control measures rather challenging to develop. The introduction of R genes, major resistance genes, into commercial cultivars is an effective strategy, but these genes can be overcome by the rapid adaption of a pathogenic strain due to mutations in individual effector genes [200]. Nevertheless, polygenic disease resistance consists of many additive quantitative loci that can confer durable resistance against rapidly evolving pathogen populations. Moreover, pyramiding of such quantitatively controlled disease and pest resistance loci into susceptible cultivated backgrounds can lead to the development of cultivars with wide-spectrum and durable resistance [201]. QTLs linked to disease and pest resistance have been identified in pear, and molecular markers are becoming widely available for use in breeding programs [202].

As mentioned above, genome sequences of both Asian and European pears provided opportunities to develop high-throughput genotyping platforms to construct dense genetic maps for QTL mapping, identify genes/QTLs for resistance genes, and identify robust molecular markers useful for early screening and genotyping of young seedlings. Due to their economic importance, the genetic basis for fire blight [45, 203, 204], pear scab [80, 205, 206], black spot [80], and pear psylla [201] resistance was of particular interest, and

molecular markers linked to genetic resistance to some of these different biotic stressors were identified and became available for use in pear breeding programs.

Fire blight, caused by *Erwinia amylovora*, is a major threat to production in most of the countries where pears are grown. Different levels of susceptibility to fire blight were reported in European and Asian pear cultivars. Although no monogenic sources of fire blight resistance were reported, approximately 13 resistance loci of major and minor contributions to trait expression were identified for fire blight on LG2, LG4, LG9, LG10, LG11, and LG15 [45, 203, 204]. The genetic sources of these loci were derived from both European and Asian species, mainly *P. communis* and *P. ussuriensis*, respectively. Four QTLs linked to fire blight resistance were mapped onto LGs 2, 4, and 9 in the resistant European pear cultivar 'Harrow Sweet', and these QTLs had moderate/minor effects. A suite of AFLP-resistant gene analog (RGA) markers tightly linked to these QTLs were developed for use in MAB programs [45].

Abiotic stress is another important factor in the growth, development, productivity, and geographic adaptation of pear, including tolerance to drought, cold, and salt stresses. In recent years, a few key transcription factors associated with abiotic stress have been identified in pear, such as *PbrWRKY53* [207], *PbrMYB21* [208], *PbrMYB5* [209], *PbrbHLH1* [210], *PbeNAC1* [211], and *PbrNHX2* [212]. Among these, *PbrWRKY53* was found to play a positive role in drought tolerance by binding to the W-box element in the promoter region of *PbrNCED1* [207]. Moreover, transgenic tobacco expressing *PbrMYB21* exhibited higher levels of arginine decarboxylase expression along with a higher accumulation of polyamine, thus contributing to enhanced tolerance to both dehydration and drought stresses [208]. Furthermore, a novel R2R3-type MYB transcription factor, designated *PbrMYB5*, could bind to the promoter of *PbrDNAR2* and contribute to enhanced tolerance to chilling stress in pear [209].

### Dwarf growth habit

The dwarf growth habit allows for the establishment of high-density plantings for pear production. Utilizing suitable dwarf rootstocks or dwarf growth habit cultivars is an important approach to achieve this goal. The French cultivar 'Nain Vert', a seedling mutant of *P. communis* exhibiting a dwarf growth habit architecture with distinctive features of a compact crown, short internodes, and short stature, is observed to carry a single dominant gene *PcDw* [62]. However, the molecular mechanism of this dwarf architecture trait remains unknown.

Based on available pear and apple genetic maps along with draft genome sequences [16, 20], several DNA molecular markers, including SSRs and SNPs, tightly linked to the dwarf growth habit trait of 'Nain Vert' were developed, and these could serve as useful tools for MAS [213]. To uncover the sequence information, structure, and functional regulation of *PcDw*, the region containing this locus was narrowed down by fine-mapping using

these tightly linked DNA markers, and several candidate genes were inferred [213]. Recently, through comparative transcriptome analysis, the two most likely candidates, an arabinogalactan protein 7-like gene and a protein WVD2-like 7 gene, for the *PcDw* locus were identified, and a systemic overview of the complex regulatory network of the dwarf phenotype was provided [214].

As dwarf rootstocks are used for grafting scion cultivars to reduce the overall vigor of grafted trees, pear breeding efforts are underway to develop dwarf rootstocks that are graft-compatible with various scion cultivars, easily propagated, and highly resistant to both biotic and abiotic stresses. To accelerate the breeding scheme, it is necessary to elucidate the molecular mechanism and identify genetic determinants involved in the control of vigor in pear trees induced by rootstocks.

Two major QTLs, *Dw1* and *Dw2*, were identified from the apple rootstock 'M9' ('Malling 9'), accounting for most of the dwarfing effect conferred to the scion [215]. Recently, using an  $F_1$  segregating population of 'Old Home'  $\times$  'Louise Bonne de Jersey' (OH $\times$ LB), a QTL influencing scion vigor in pear was mapped to the syntenic position in the apple 'M.9' *Dw1* locus, which is located at the upper end of LG5 [216]. Based on the high degree of synteny between apple and pear [217], it will be worthwhile to conduct comparative genome analysis to explore and delineate the genetic and molecular mechanisms of rootstock-induced dwarfing in pear.

## Genome editing

Sequencing of the pear genome is important for identifying genes and desired traits using genome editing. Recently, CRISPR–Cas9 has been used for genome editing in diverse plant systems [218]. Owing to its high efficiency, simplicity, and versatility of multiplexing, Cas9, a Class 2 type II CRISPR system, was used extensively to manipulate different plant traits, such as disease resistance, crop yield, and quality [219–221]. Subsequently, Cas12a (formerly known as Cpf1), a Class 2 type V endonuclease, was used for editing several plant genomes, such as those for rice [222], tobacco [223], maize [224], and *Arabidopsis* [224], at high efficiencies.

Recently, genome editing studies have been undertaken in a number of horticultural crops. Among these, certain vegetable crops were subjected to more genome editing efforts (72% of all genome editing reports in horticultural crops) than other crops, particularly tomatoes (42% of all genome editing reports in horticultural crops) [225]. Fruit crops are very important commercial horticultural crops. Due to their time-consuming breeding cycle and complex genetic backgrounds, CRISPR–Cas systems have been successfully used in only a few fruit crops. Early efforts focused on using a mutation of the phytoene desaturase (*PDS*) gene that yielded an albino phenotype. This was demonstrated in banana [226], kiwifruit [227], grape [228], citrus [229], apple [230], and strawberry [231, 232]. Strawberry was used as a model fruit crop system for pursuing functional studies of

particular genes of interest, such as the anthocyanin biosynthesis TF R2R3 MYB10, whereby CRISPR–Cas9-mediated gene editing was found to reduce fruit color [231]. In addition, the CRISPR–Cas9 system was designed to target the *DIPM1*, 2, and 4 genes encoding disease-specific (Dsp)A/E-interacting proteins and pathogenicity effectors of *E. amylovora* in apple to enhance genetic resistance against fire blight disease [233].

Efforts have been undertaken to generate compact plants that develop rapid terminal flowers and fruits to accelerate the breeding cycle in tree fruit crops by gene editing of floral repressor genes, such as *CENTRO-RADIALIS* (*CEN*)-like genes [234]. Recently, the CRISPR–Cas9 system was used for the first time in pear by targeting the *TERMINAL FLOWER 1* (*TFL1*) gene, and early flowering phenotypes were observed in 9% of transgenic pear plants targeted for the *PcTFL1.1* gene and 93% of transgenic apple lines targeted for the *MdTFL1.1* gene [235].

Ongoing gene editing efforts are being pursued using the CRISPR–Cas9 system along with other novel gene editing systems, as these tools will be useful for the functional analysis of target genes and enhancing various desirable economic traits in pear.

## Future research prospects

The first pear genome, reported in 2013, has served as a highly valuable resource for gaining sequence information, gene discovery, fine mapping, tool development, and variant detection. Over the last decade, WGS efforts for pear have had significant impacts on our knowledge of pear biology, gene function, inheritance, and trait physiology, leading to the availability of valuable resources and tools that will contribute to the genetic improvement of this highly valuable fruit crop. These resources will also support efforts to shorten the breeding cycle and increase breeding efficiency. For future pear research and breeding strategies, it is highly advisable to integrate different omics technologies into the discovery process of genes controlling traits of interest, the elucidation of gene function mechanisms, support for genotyping and phenotyping, and the acceleration of the breeding cycle. Additionally, advanced NGS methods and development, approaches to shorten the juvenility period, gene editing technology, and gene transfer technologies will all contribute to efforts to breed new pear cultivars with highly desirable economic characteristics.

Below are some specific areas of research interest that should be taken into consideration in future efforts.

## Genome development

Although whole genome sequences of pear are available, each single reference genome does not represent the wide diversity within a species due to the presence of high levels of genomic variation. Recently, a pangenome concept has been proposed that involves comparing the genomes of multiple related individuals,



wherein core genome segments are present in all individuals, while dispensable genome sequences are absent in some individuals. As pear has experienced independent domestication across Asian and European pears, major differences have been observed in different groups of wild and cultivated Asian and European pear accessions. Thus, pangenome sequencing will help us better characterize those subspecies of pear genomes and identify any unique genes that might be present in these genomes. Furthermore, it is necessary to construct a pear pangenome assembly to capture the wide genetic diversity present in *Pyrus*, as this will further support evolutionary studies and aid in breeding efforts in pear. Due to the high level of heterozygosity and large numbers of pear cultivars, short-read sequences and the uncertainty of the physical locations of genes will undoubtedly raise new challenges in constructing such a pear pangenome sequence assembly and sequence mapping.

It is also important to point out that the obtaining the haplotype genome of pear serves as another bottleneck due to the self-incompatibility and long generation time of pear fruit trees. However, long-read sequencing technology could help identify strand-specific sequences and overcome some of these problems.

### Resequencing and phylogenetic studies

As NGS has become less expensive, long-read sequencing has become more popular. Earlier pear resequencing data reached tenfold coverage for NGS [28]; however, this is not sufficient for identifying structural variations (SVs), other than for SNP calling. Such SVs include deletions, insertions, duplications, inversions, and translocations, which can then be identified using NGS data and long-read sequencing combined with multiple algorithms, such as splitting reads, read pairs, and assembly methods, to discover and characterize SVs in the pear genome.

Whole-genome resequencing will also allow for a better understanding of the phylogenetics, domestication, improvement, and evolutionary history of a crop at the whole-genome level. This will provide unprecedented amounts of genomic datasets that will almost certainly allow for advances in modern pear molecular breeding programs.

### Whole-genome association studies

For pear trait association studies, GWAS is an effective approach in natural populations, overcoming the long juvenile period and self-incompatibility issues in *Pyrus*. However, it is important to standardize parameters and units of measurements of various phenotypic data from multiple years and sites to uncover reliable trait associations in such GWAS. With the enhanced reliability and accuracy of trait associations, generated datasets will allow for investigating SVs across multiple genomes associated with target traits. Thus, in addition to SNPs and small InDels, these expanded datasets will allow for

investigating genomes for copy number variants (CNVs), as these SVs confer larger effects on plant phenotypes. In recent years, several SV-GWAS have been undertaken in various plant species [236–238]; however, such studies have not yet been conducted in pear.

### Integrated omics technologies

Currently, omics, which includes genetics, transcriptomics, epigenomics, proteomics, metabolomics, phenomics, and bioinformatics, offers various promising approaches that will accelerate breeding efforts. Thus, pursuing integrated omics efforts will allow for generating new knowledge of yet unknown genes, proteins, and metabolites. Such new knowledge will aid in genetic improvement and breeding efforts in pear. For future pear research, it is critical to exploit integrated omics approaches to enhance our knowledge not only of fruit development and quality traits but also of biotic and abiotic stresses and subsequently translate this knowledge into pear breeding efforts.

### Gene editing

At present, gene editing offers a promising approach to improve new cultivars for traits of interest. Compared with other approaches, gene editing offers advantages in versatility, efficiency, and specificity. Currently, CRISPR-based gene editing can result in many types of mutations in target sequences, including precise base substitutions, small deletions/insertions, large fragment deletions, and gene replacement. Moreover, gene-edited plants can be free of exogenous DNA sequences either through genetic segregation or if CRISPR reagents are delivered as ribonucleoproteins to yield nontransgenic plants. This method may be a quick approach to developing new cultivars or improving existing cultivars.

As genetic transformation and genome editing technologies are limited for pear, it is important to undertake efforts to develop new protocols. This will be critical for pursuing the functional analysis of genes and for moving ahead with robust and efficient breeding of new and improved cultivars.

### Acknowledgments

We thank the editor-in-chief, Prof. Zong-Ming (Max) Cheng, for initiating these review papers. We apologize to our colleagues whose work was not included here owing to space limitations. The authors are grateful to various funding agencies and projects that contributed to the present work, including the National Natural Science Foundation of China (31725024, 31820103012, 31801835, 31901978), National Key Research and Development Program (2018YFD1000200), National Natural Science Foundation of Jiangsu Province For Young Scholars (BK20180516, BK20180529), and China Agriculture Research System of MOF and MARA.

## Author Contributions

J.W. and S.S.K. contributed to the manuscript outline and led editing of all sections of the manuscript; J.L. M.Z. and X.L. finalized the organization of the outline. The authors' contributions to the different sections of the manuscript are as follows: "Overview of Asian and European pear genome sequences" and "Pear domestication and improvement" (X.L. S.Z.); "Genome-wide variations" (M.Z., R.W.); "Genetic linkage maps: Construction and trait linked loci mapping" (M.Q., and M.S.); "Marker-assisted selection, marker-assisted breeding, and genomic selection in pear" (A.C.A., K.L.W., and R.V.E.); "Multiple omics: identifying genes related to important traits" (J.L., M.Z., X.L., A.K., C.W., R.W., C.X., G.Y., R.T., H.L., W.W., M.M., K.Z., B.B., J.N., and J.A.); and "Future research prospects" (J.L., M.Z. and X.L.). All authors reviewed the manuscript in the draft and final forms.

## Conflict of interest statement

The authors declare that they have no conflicts of interest.

## Supplementary data

Supplementary data is available at Horticulture Research online.

## References

- Potter D, Li J, Sun J et al. Phylogeny and classification of Rosaceae. *Plant Syst Evol.* 2007;**266**:5–43. <https://doi.org/10.1007/s00606-007-0539-9>.
- Cao YF, Tian LM, Liu-Lin LI et al. Comparison studies on the stone cell content in flesh of pear cultivars. *Acta Hort Sinica.* 2010;**22**:417–33. <https://doi.org/10.16420/j.issn.0513-353x.2010.08.016>.
- Yao GF, Zhang SL, Cao YF et al. Characteristics of components and contents of soluble sugars in pear fruits from different species. *Scient Agric Sinica.* 2010;**43**:4229–37. <https://doi.org/10.3864/j.issn.0578-1752.2010.20.014>.
- Quinet M, Wesel JP. Botany and taxonomy of pear. In: Korban SS, ed. *The Pear Genome*. Cham, Switzerland: Springer International Publishing, 2019,1–33.
- Wu J, Wang Z, Shi Z et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 2013;**23**:396–408. <https://doi.org/10.1101/gr.144311.112>.
- Li JM, Zheng DM, Li LT et al. Genome-wide function, evolutionary characterization and expression analysis of sugar transporter family genes in pear (*Pyrus bretschneideri* Rehd.). *Plant Cell Physiol.* 2015;**56**:1721–37. <https://doi.org/10.1093/pcp/pcv090>.
- Qiao X, Li M, Li et al. Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (*Pyrus bretschneideri*) and five other Rosaceae species. *BMC Plant Biol.* 2015;**15**:12. <https://doi.org/10.1186/s12870-014-0401-5>.
- Li X, Xue C, Li J et al. Genome-wide identification, evolution and functional divergence of MYB transcription factors in Chinese white pear (*Pyrus bretschneideri*). *Plant Cell Physiol.* 2016;**57**: 824–47. <https://doi.org/10.1093/pcp/pcw029>.
- Lü H, Li J, Huang Y et al. Genome-wide identification, expression and functional analysis of the phosphofructokinase gene family in Chinese white pear (*Pyrus bretschneideri*). *Gene.* 2019;**702**: 133–42. <https://doi.org/10.1016/j.gene.2019.03.005>.
- Song B et al. Mining and evolution analysis of lateral organ boundaries domain (LBD) genes in Chinese white pear (*Pyrus bretschneideri*). *BMC Genomics.* 2020;**21**:644. <https://doi.org/10.1186/s12864-020-06999-9>.
- Sun M, Zhang M, Singh J et al. Contrasting genetic variation and positive selection followed the divergence of NBS-encoding genes in Asian and European pears. *BMC Genomics.* 2020;**21**:809. <https://doi.org/10.1186/s12864-020-07226-1>.
- Hussain S, Niu Q, Qian M et al. Genome-wide identification, characterization, and expression analysis of the dehydrin gene family in Asian pear (*Pyrus pyrifolia*). *Tree Genet Genomes.* 2015;**11**:110. <https://doi.org/10.1007/s11295-015-0938-y>.
- Cao Y, Han Y, Meng D et al. B-BOX genes: genome-wide identification, evolution and their contribution to pollen growth in pear (*Pyrus bretschneideri* Rehd.). *BMC Plant Biol.* 2017;**17**:156. <https://doi.org/10.1186/s12870-017-1105-4>.
- Li L, Deng CH, Knabel M et al. Integrated high-density consensus genetic map of *Pyrus* and anchoring of the 'Bartlett' v1.0 (*Pyrus communis*) genome. *DNA Res.* 2017;**24**:dsw063–301. <https://doi.org/10.1093/dnares/dsw063>.
- Wu J, Li LT, Li M et al. High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *J Exp Bot.* 2014;**65**:5771–81. <https://doi.org/10.1093/jxb/eru311>.
- Velasco R, Zharkikh A, Affourtit J et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet.* 2010;**42**:833–9. <https://doi.org/10.1038/ng.654>.
- Shi D et al. Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants. *Genome Res.* 2019;**29**:1889–99. <https://doi.org/10.1101/gr.251033.119>.
- Qin G, Tao S, Cao Y et al. Evaluation of the volatile profile of 33 *Pyrus ussuriensis* cultivars by HS-SPME with GC–MS. *Food Chem.* 2012;**134**:2367–82. <https://doi.org/10.1016/j.foodchem.2012.04.053>.
- Suwanagul A, Richardson DG. Identification of headspace volatile compounds from different pear (*Pyrus communis* L.) varieties. *Acta Hort.* 1998;**475**:605–24. <https://doi.org/10.17660/ActaHortic.1998.475.73>.
- Chagne D, Crowhurst RN, Pindo M et al. The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS One.* 2014;**9**:e92644. <https://doi.org/10.1371/journal.pone.0092644>.
- Linsmith G, Rombauts S, Montanari S et al. Pseudo-chromosome-length genome assembly of a double haploid "Bartlett" pear (*Pyrus communis* L.). *Giga Science.* 2019;**8**:giz138. <https://doi.org/10.1093/gigascience/giz138>.
- Dong X, Wang Z, Tian L et al. De novo assembly of a wild pear (*Pyrus betuleafolia*) genome. *Plant Biotechnol J.* 2020;**18**:581–95. <https://doi.org/10.1111/pbi.13226>.
- Ou C, Wang F, Wang J et al. A de novo genome assembly of the dwarfing pear rootstock Zhongai 1. *Scientific data.* 2019;**6**:281. <https://doi.org/10.1038/s41597-019-0291-3>.
- Jung S, Lee T, Cheng CH et al. 15 years of GDR: new data and functionality in the genome database for Rosaceae. *Nucleic Acids Res.* 2019;**47**:D1137–45. <https://doi.org/10.1093/nar/gky1000>.
- Cao K, Zheng Z, Wang L et al. Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops.

- Genome Biol. 2014;**15**:415. <https://doi.org/10.1186/s13059-014-0415-1>.
26. Zhou Y, Massonnet M, Sanjak JS et al. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc Natl Acad Sci U S A*. 2017;**114**:11715–20. <https://doi.org/10.1073/pnas.1709257114>.
  27. Duan N, Bai Y, Sun H et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun*. 2017;**8**:249. <https://doi.org/10.1038/s41467-017-00336-7>.
  28. Wu J, Wang Y, Xu J et al. Diversification and independent domestication of Asian and European pears. *Genome Biol*. 2018;**19**:77. <https://doi.org/10.1186/s13059-018-1452-y>.
  29. Silva GJ, Souza TM, Barbieri RL et al. Origin, domestication, and dispersing of pear (*Pyrus* spp.). 82d Cong, 2d sess House of Representatives Report ; no 2356. 2014;**2014**:1. <https://doi.org/10.1155/2014/541097>.
  30. Li X, Li L, Ming M et al. Comparative transcriptomic analysis provides insight into the domestication and improvement of pear (*P. pyrifolia*) fruit. *Plant Physiol*. 2019;**180**:435–52. <https://doi.org/10.1104/pp.18.01322>.
  31. Zhang, M.Y, Xue, C, Xu, L. et al. Distinct transcriptome profiles reveal gene expression patterns during fruit development and maturation in five main cultivated species of pear (*Pyrus* L.). *Sci Rep* **6**, ???, <https://doi.org/10.1038/srep28130> (2016).
  32. Elshire RJ, Glaubitz JC, Sun Q et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;**6**:e19379. <https://doi.org/10.1371/journal.pone.0019379>.
  33. Kumar S, Kirk C, Deng C et al. Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. *Hortic. Res*. 2017;**4**:17015. <https://doi.org/10.1038/hortres.2017.15>.
  34. Kumar S, Kirk C, Deng CH et al. Fine-mapping and validation of the genomic region underpinning pear red skin colour. *Hortic Res*. 2019;**6**:29. <https://doi.org/10.1038/s41438-018-0112-4>.
  35. Kumar S, Kirk C, Deng CH et al. Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. *Sci Rep*. 2019;**9**:9072. <https://doi.org/10.1038/s41598-019-45618-w>.
  36. Han ML, Liu YL, Zheng XY et al. Construction of a genetic linkage map and QTL analysis for some fruit traits in pear. *J Fruit Sci*. 2010;**27**:496–503.
  37. Gabay G, Dahan Y, Izhaki Y et al. High-resolution genetic linkage map of European pear (*Pyrus communis*) and QTL fine-mapping of vegetative budbreak time. *BMC Plant Biol*. 2018;**18**:175. <https://doi.org/10.1186/s12870-018-1386-2>.
  38. Xue L, Liu Q, Hu H et al. The southwestern origin and eastward dispersal of pear (*Pyrus pyrifolia*) in East Asia revealed by comprehensive genetic structure analysis with SSR markers. *Tree Genet Genomes*. 2018;**14**:1–12. <https://doi.org/10.1007/s11295-018-1255-z>.
  39. Beissinger TM, Hirsch CN, Sekhon RS et al. Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics*. 2013;**193**:1073–81. <https://doi.org/10.1534/genetics.112.147710>.
  40. Li X, Singh J, Qin M et al. Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol J*. 2019;**17**:1582–94. <https://doi.org/10.1111/pbi.13085>.
  41. Montanari S, Bianco L, Allen BJ et al. Development of a highly efficient axion 70 K SNP array for *Pyrus* and evaluation for high-density mapping and germplasm characterization. *BMC Genomics*. 2019;**20**:331. <https://doi.org/10.1186/s12864-019-5712-3>.
  42. Iketani H, Abe K, Yamamoto T et al. Mapping of disease-related genes in Japanese pear using a molecular linkage map with RAPD markers. *Breed Sci*. 2001;**51**:179–84. <https://doi.org/10.1270/jsbbs.51.179>.
  43. Yamamoto T, Kimura T, Shoda M et al. Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. *Theor Appl Genet*. 2002;**106**:9–18. <https://doi.org/10.1007/s00122-002-0966-5>.
  44. Yamamoto T, Kimura T, Saito T et al. Genetic linkage maps of Japanese and European pears aligned to the apple consensus map. *Acta Hortic*. 2004;**663**:51–6. <https://doi.org/10.17660/ActaHortic.2004.663.2>.
  45. Dondini L, Pierantoni L, Gaiotti F et al. Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. *Mol Breed*. 2005;**14**:407–18. <https://doi.org/10.1007/s11032-005-0505-6>.
  46. Yamamoto T, Kimura T, Tekakami S et al. Integrated reference genetic linkage maps of pear based on SSR and AFLP markers. *Breed Sci*. 2007;**57**:321–9. <https://doi.org/10.1270/jsbbs.57.321>.
  47. Zhang R, Wu J, Li XG et al. An AFLP, SRAP, and SSR genetic linkage map and identification of QTLs for fruit traits in pear (*Pyrus* L.). *plant Mol. Plant Mol Biol Report*. 2013;**31**:678–87. <https://doi.org/10.1007/s11105-012-0544-1>.
  48. Dondini L, Pierantoni L, Sansavini S et al. The inheritance of the red colour character in European pear (*Pyrus communis*) and its map position in the mutated cultivar 'max red Bartlett'. *Plant Breed*. 2008;**127**:524–6. <https://doi.org/10.1111/j.1439-0523.2008.01500.x>.
  49. Yao G, Ming M, Allan AC et al. Map-based cloning of the pear gene MYB114 identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant J*. 2017;**92**:437–51. <https://doi.org/10.1111/tpj.13666>.
  50. Pierantoni L, Dondini L, Cho KH et al. Pear scab resistance QTLs via a European pear (*Pyrus communis*) linkage map. *Tree Genet Genomes*. 2007;**3**:311–7. <https://doi.org/10.1007/s11295-006-0070-0>.
  51. Wu J, Qin M. Linkage mapping in pear. In: Korban SS, ed. *The Pear Genome*. Cham, Switzerland: Springer International Publishing, 2019,103–12.
  52. Yamamoto T, Terakami S, Kimura T et al. Reference genetic linkage maps of European and Japanese pears. *Acta Hortic*. 2009;**814**:599–602. <https://doi.org/10.17660/actahortic.2009.814.101>.
  53. Terakami S, Kimura T, Nishitani C et al. Genetic linkage map of the Japanese pear 'Housui' identifying three homozygous genomic regions. *Horticulture journal*. 2009;**78**: 417–24. <https://doi.org/10.2503/jjshs.1.78.417>.
  54. Sun W, Zhang Y, Le W et al. Construction of a genetic linkage map and QTL analysis for some leaf traits in pear (*Pyrus* L.). *Front Agric China*. 2009;**3**:67–74. <https://doi.org/10.1007/s11703-009-0013-2>.
  55. Pierantoni L, Dondini L, De Franceschi P et al. Mapping of an anthocyanin-regulating MYB transcription factor and its expression in red and green pear, *Pyrus communis*. *Plant Physiol Biochem*. 2010;**48**:1020–6. <https://doi.org/10.1016/j.plaphy.2010.09.002>.
  56. Choi JK, Na DY, Kim D et al. Genetic linkage mapping using interspecific hybrid population between Korean wild pear (*Pyrus ussuriensis*) and Japanese pear (*P. pyrifolia*). *Hortic Environ Biotech*. 2010;**51**:319–25.



57. Hao Y, Zhao Y, Guo Y et al. Genetic linkage maps of pear based on SRAP markers. *Pak J Bot.* 2013;**45**:1265–71.
58. Yamamoto T, Terakami S, Moriya S et al. DNA markers developed from genome sequencing analysis in Japanese pear (*Pyrus pyrifolia*). *Acta Hortic.* 2013;**976**:477–83. <https://doi.org/10.17660/ActaHortic.2013.976.67>.
59. Won K, Bastiaanse H, Song JH et al. Genetic mapping of polygenic scab (*Venturia pirina*) resistance in an interspecific pear family. *Mol Breed.* 2014;**34**:2179–89. <https://doi.org/10.1007/s11032-014-0172-6>.
60. Yamamoto T, Terakami S, Takada N et al. Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed Sci.* 2014;**64**:351–61. <https://doi.org/10.1270/jsbbs.64.351>.
61. Chen H, Song Y, Li LT et al. Construction of a high-density simple sequence repeat consensus genetic map for pear (*Pyrus* spp.). *Plant Mol. Biol. Rep.* 2015;**33**:316–25. <https://doi.org/10.1007/s11105-014-0745-x>.
62. Wang C, Tian Y, Buck EJ et al. Genetic mapping of PcDw determining pear dwarf trait. *J Am Soc Hortic Sci.* 2011;**136**:48–53.
63. Wang L, Li X, Wang L et al. Construction of a high-density genetic linkage map in pear (*Pyrus communis* × *Pyrus pyrifolia* Nakai) using SSRs and SNPs developed by SLAF-seq. *Sci Hortic.* 2017;**218**:198–204. <https://doi.org/10.1016/j.scienta.2017.02.015>.
64. Xue H, Wang S, Yao JL et al. Chromosome level high-density integrated genetic maps improve the *Pyrus bretschneideri* 'DangshanSuli' v1.0 genome. *BMC Genomics.* 2018;**19**:833. <https://doi.org/10.1186/s12864-018-5224-6>.
65. Zurn JD, Norelli JL, Montanari S et al. Dissecting genetic resistance to fire blight in three pear populations. *Phytopathology.* 2020;**110**:1305–11. <https://doi.org/10.1094/PHYTO-02-20-0051-R>.
66. Pritchard JK, Stephens MJ, Donnelly PJ. Inference of population structure using multilocus genotype data. *Genetics.* 2000;**155**:945–59. <https://doi.org/10.1093/genetics/155.2.945>.
67. Kang HM, Zaitlen NA, Wade CM et al. Efficient control of population structure in model organism association mapping. *Genetics.* 2008;**178**:1709–23. <https://doi.org/10.1534/genetics.107.080101>.
68. Aulchenko YS, Ripke S, Isaacs A et al. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* 2007;**23**:1294–6. <https://doi.org/10.1093/bioinformatics/btm108>.
69. Abney M, Ober C, McPeck MS. Quantitative-trait homozygosity and association mapping and empirical genome-wide significance in large, complex pedigrees: fasting serum-insulin level in the hutterites. *Am J Hum Genet.* 2002;**70**:920–34. <https://doi.org/10.1086/339705>.
70. Aulchenko YS, de Koning DJ, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis. *Genetics.* 2007;**177**:577–85. <https://doi.org/10.1534/genetics.107.075614>.
71. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;**44**:821–4. <https://doi.org/10.1038/ng.2310>.
72. Mariette S, Tai FWJ, Roch G et al. Genome-wide association links candidate genes to resistance to plum pox virus in apricot (*Prunus armeniaca*). *New Phytol.* 2016;**209**:773–84. <https://doi.org/10.1111/nph.13627>.
73. Cao K, Zhou Z, Wang Q et al. Genome-wide association study of 12 agronomic traits in peach. *Nat Commun.* 2016;**7**:13246. <https://doi.org/10.1038/ncomms13246>.
74. Zhang MY, Xue C, Hu H et al. Genome-wide association studies provide insights into the genetic determination of fruit traits of pear. *Nat Commun.* 2021;**12**:1144. <https://doi.org/10.1038/s41467-021-21378-y>.
75. Peace CP. DNA-informed breeding of rosaceous crops: promises, progress and prospects. *Hortic. Res.* 2017;**4**:17006. <https://doi.org/10.1038/hortres.2017.6>.
76. Collard BC, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond Ser B Biol Sci.* 2008;**363**:557–72. <https://doi.org/10.1098/rstb.2007.2170>.
77. Itai A, Kotaki T, Tanabe K et al. Rapid identification of 1-aminocyclopropane-1-carboxylate (ACC) synthase genotypes in cultivars of Japanese pear (*Pyrus pyrifolia* Nakai) using CAPS markers. *Theor Appl Genet.* 2003;**106**:1266–72. <https://doi.org/10.1007/s00122-002-1186-8>.
78. Xu M, Huaracha E, Korban SS. Development of sequence-characterized amplified regions (SCARs) from amplified fragment length polymorphism (AFLP) markers tightly linked to the Vf gene in apple. *Genome.* 2001;**44**:63–70. <https://doi.org/10.1139/gen-44-1-63>.
79. Moriya Y, Yamamoto K, Okada K et al. Development of a CAPS marker system for genotyping European pear cultivars harboring 17 S alleles. *Plant Cell Rep.* 2007;**26**:345–54. <https://doi.org/10.1007/s00299-006-0254-y>.
80. Terakami S, Moriya S, Adachi Y et al. Fine mapping of the gene for susceptibility to black spot disease in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed Sci.* 2016;**66**:271–80. <https://doi.org/10.1270/jsbbs.66.271>.
81. Khan MA, Zhao YF, Korban SS. Identification of genetic loci associated with fire blight resistance in *Malus* through combined use of QTL and association mapping. *Physiol Plant.* 2013;**148**:344–53. <https://doi.org/10.1111/ppl.12068>.
82. Meuwissen TH, Goddard ME. Prediction of identity by descent probabilities from marker-haplotypes. *Genet Sel Evol.* 2001;**33**:605–34. <https://doi.org/10.1186/1297-9686-33-6-605>.
83. Iwata H, Hayashi T, Terakami S et al. Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breed Sci.* 2013;**63**:125–40. <https://doi.org/10.1270/jsbbs.63.125>.
84. Minamikawa MF, Takada N, Terakami S et al. Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (*Pyrus pyrifolia* Nakai). *Sci Rep.* 2018;**8**:11994. <https://doi.org/10.1038/s41598-018-30154-w>.
85. Wang R, Ming M, Li J et al. Genome-wide identification of the MADS-box transcription factor family in pear (*Pyrus bretschneideri*) reveals evolution and functional divergence. *PeerJ.* 2017;**5**:e3776. <https://doi.org/10.7717/peerj.3776>.
86. Li J, Qin M, Qiao X et al. A new insight into the evolution and functional divergence of SWEET transporters in Chinese white pear (*Pyrus bretschneideri*). *Plant Cell Physiol.* 2017;**58**:839–50. <https://doi.org/10.1093/pcp/pcx025>.
87. Yu L, Li J, Li L et al. Characterisation of the whole-genome wide hexokinase gene family unravels the functional divergence in pear (*Pyrus bretschneideri* Rehd.). *J Hortic Sci Biotechnol.* 2018;**93**:244–54. <https://doi.org/10.1080/14620316.2017.1362961>.
88. Cao Y, Meng D, Li X et al. A Chinese white pear (*Pyrus bretschneideri*) BZR gene PbBZR1 act as a transcriptional repressor of lignin biosynthetic genes in fruits. *Front Plant Sci.* 2020;**11**:1087. <https://doi.org/10.3389/fpls.2020.01087>.
89. Cheng X, Li M, Abdullah M et al. In Silico genome-wide analysis of the pear (*Pyrus bretschneideri*) KNOX family and

- the functional characterization of PbKNOX1, an Arabidopsis BREVIPEDICELLUS orthologue gene, involved in cell wall and lignin biosynthesis. *Front Genet.* 2019;**10**:632. <https://doi.org/10.3389/fgene.2019.00632>.
90. Liu C, Qiao X, Li Q et al. Genome-wide comparative analysis of the BAH1 superfamily in seven Rosaceae species and expression analysis in pear (*Pyrus bretschneideri*). *BMC Plant Biol.* 2020;**20**:14. <https://doi.org/10.1186/s12870-019-2230-z>.
  91. Zeng W, Qiao X, Li Q et al. Genome-wide identification and comparative analysis of the ADH gene family in Chinese white pear (*Pyrus bretschneideri*) and other Rosaceae species. *Genomics.* 2020;**112**:3484–96. <https://doi.org/10.1016/j.ygeno.2020.06.031>.
  92. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;**10**:57–63. <https://doi.org/10.1038/nrg2484>.
  93. Zhou H, Yin H, Chen J et al. Gene-expression profile of developing pollen tube of *Pyrus bretschneideri*. *Gene Expr Patterns.* 2016;**20**:11–21. <https://doi.org/10.1016/j.gep.2015.10.004>.
  94. Li QH, Qiao X, Yin H et al. Unbiased subgenome evolution following a recent whole-genome duplication in pear (*Pyrus bretschneideri* Rehd.). *Hortic. Res.* 2019;**6**:34. <https://doi.org/10.1038/s41438-018-0110-6>.
  95. Yang YN, Yao GF, Yue WQ et al. Transcriptome profiling reveals differential gene expression in proanthocyanidin biosynthesis associated with red/green skin color mutant of pear (*Pyrus communis* L.). *Front Plant Sci.* 2015;**6**:795. <https://doi.org/10.3389/fpls.2015.00795>.
  96. Liu BY, Wang L, Wang S et al. Transcriptomic analysis of bagging-treated 'Pingguo' pear shows that MYB4-like1, MYB4-like2, MYB1R1 and WDR involved in anthocyanin biosynthesis are up-regulated in fruit peels in response to light. *Sci Hortic.* 2019;**244**:428–34. <https://doi.org/10.1016/j.scienta.2018.09.040>.
  97. Wang ZG, Du H, Zhai R et al. Transcriptome analysis reveals candidate genes related to color fading of 'Red Bartlett' (*Pyrus communis* L.). *Front Plant Sci.* 2017;**8**:455. <https://doi.org/10.3389/fpls.2017.00455>.
  98. Wang YZ, Dai MS, Zhang SJ et al. Exploring candidate genes for pericarp russet pigmentation of sand pear (*Pyrus pyrifolia*) via RNA-seq data in two genotypes contrasting for pericarp color. *PLoS One.* 2014;**9**:e83675. <https://doi.org/10.1371/journal.pone.0083675>.
  99. Ou C, Zhang X, Wang F et al. A 14 nucleotide deletion mutation in the coding region of the PpBBX24 gene is associated with the red skin of "Zaosu red" pear (*Pyrus pyrifolia* white pear group): a deletion in the PpBBX24 gene is associated with the red skin of pear. *Hortic. Res.* 2020;**7**:39. <https://doi.org/10.1038/s41438-020-0259-7>.
  100. Bai SL, Saito T, Sakamoto D et al. Transcriptome analysis of Japanese pear (*Pyrus pyrifolia* Nakai) flower buds transitioning through endodormancy. *Plant Cell Physiol.* 2013;**54**:1132–51. <https://doi.org/10.1093/pcp/pct067>.
  101. Nham NT, Macnish AJ, Zakharov F et al. 'Bartlett' pear fruit (*Pyrus communis* L.) ripening regulation by low temperatures involves genes associated with jasmonic acid, cold response, and transcription factors. *Plant Sci.* 2017;**260**:8–18. <https://doi.org/10.1016/j.plantsci.2017.03.008>.
  102. Yang TY, Huang XS. Deep sequencing-based characterization of transcriptome of *Pyrus ussuriensis* in response to cold stress. *Gene.* 2018;**661**:109–18. <https://doi.org/10.1016/j.gene.2018.03.067>.
  103. Li KQ, Xu XY, Huang XS. Identification of differentially expressed genes related to dehydration resistance in a highly drought-tolerant pear, *Pyrus betulaeifolia*, as through RNA-Seq. *PLoS One.* 2016;**11**:e0149352. <https://doi.org/10.1371/journal.pone.0149352>.
  104. Li H, Lin J, Yang QS et al. Comprehensive analysis of differentially expressed genes under salt stress in pear (*Pyrus betulaeifolia*) using RNA-Seq. *Plant Growth Regul.* 2017;**82**:409–20. <https://doi.org/10.1007/s10725-017-0266-3>.
  105. Kan JL, Liu T, Ma N et al. Transcriptome analysis of Callery pear (*Pyrus calleryana*) reveals a comprehensive signalling network in response to *Alternaria alternata*. *PLoS One.* 2017;**12**:e0184988. <https://doi.org/10.1371/journal.pone.0184988>.
  106. Yang XP, Hu H, Yu D et al. Candidate resistant genes of sand pear (*Pyrus pyrifolia* Nakai) to *Alternaria alternata* revealed by transcriptome sequencing. *PLoS One.* 2015;**10**:e0135046. <https://doi.org/10.1371/journal.pone.0135046>.
  107. Li JM, Huang X, S, Ting, L. T. et al. Proteome analysis of pear reveals key genes associated with fruit development and quality. *Planta.* 2015;**241**:1363–79. <https://doi.org/10.1007/s00425-015-2263-y>.
  108. Reuscher S, Fukao Y, Morimoto R et al. Quantitative proteomics-based reconstruction and identification of metabolic pathways and membrane transport proteins related to sugar accumulation in developing fruits of pear (*Pyrus communis*). *Plant Cell Physiol.* 2016;**57**:505–18. <https://doi.org/10.1093/pcp/pcw004>.
  109. Liu X, Zhai R, Feng W et al. Proteomic analysis of 'Zaosu' pear (*Pyrus bretschneideri* Rehd.) and its early-maturing bud sport. *Plant Sci.* 2014;**224**:120–35. <https://doi.org/10.1016/j.plantsci.2014.04.012>.
  110. Qin G, Tao S, Zhang H et al. Evolution of the aroma volatiles of pear fruits supplemented with fatty acid metabolic precursors. *Molecules.* 2014;**19**:20183–96. <https://doi.org/10.3390/molecules191220183>.
  111. Wu X, Yin H, Shi Z et al. Chemical composition and crystal morphology of epicuticular wax in mature fruits of 35 pear (*Pyrus* spp.) cultivars. *Front Plant Sci.* 2018;**9**:679. <https://doi.org/10.3389/fpls.2018.00679>.
  112. Kidner CA, Martienssen RA. The developmental role of microRNA in plants. *Curr Opin Plant Biol.* 2005;**8**:38–44. <https://doi.org/10.1016/j.pbi.2004.11.008>.
  113. Wu J, Wang D, Liu Y et al. Identification of miRNAs involved in pear fruit development and quality. *BMC Genomics.* 2014;**15**:953. <https://doi.org/10.1186/1471-2164-15-953>.
  114. Ma L, Zhou L, Quan S et al. Integrated analysis of mRNA-seq and miRNA-seq in calyx abscission zone of Korla fragrant pear involved in calyx persistence. *BMC Plant Biol.* 2019;**19**:192. <https://doi.org/10.1186/s12870-019-1792-0>.
  115. Qian M, Ni J, Niu Q et al. Response of miR156-SPL module during the red peel coloration of bagging-treated Chinese sand pear (*Pyrus pyrifolia* Nakai). *Front Physiol.* 2017;**8**:500. <https://doi.org/10.3389/fphys.2017.00550>.
  116. Xue C, Yao JL, Qin MF et al. PbrmiR397a regulates lignification during stone cell development in pear fruit. *Plant Biotechnol J.* 2019;**17**:103–17. <https://doi.org/10.1111/pbi.12950>.
  117. Cheng X, Yan C, Zhang J et al. The effect of different pollination on the expression of Dangshan Su pear microRNA. *Biomed Res Int.* 2017;**2017**:1–18. <https://doi.org/10.1155/2017/2794040>.
  118. Niu Q, Qian M, Liu G et al. A genome-wide identification and characterization of microRNAs and their targets in 'Suli' pear (*Pyrus pyrifolia* white pear group). *Planta.* 2013;**238**:1095–112. <https://doi.org/10.1007/s00425-013-1954-5>.
  119. Shi F, Zhou X, Yao MM et al. miRNAs play important roles in aroma weakening during the shelf life of 'Nanguo' pear after cold storage. *Food Res Int.* 2019;**116**:942–52. <https://doi.org/10.1016/j.foodres.2018.09.031>.

120. Liu J, Zhang X, Zhang FP et al. Identification and characterization of microRNAs from in vitro-grown pear shoots infected with apple stem grooving virus in response to high temperature using small RNA sequencing. *BMC Genomics*. 2015;**16**:945. <https://doi.org/10.1186/s12864-015-2126-8>.
121. Zhang Q, Zhang Y, Wang S et al. Characterization of genome-wide microRNAs and their roles in development and biotic stress in pear. *Planta*. 2019;**249**:693–707. <https://doi.org/10.1007/s00425-018-3027-2>.
122. Niu Q, Li J, Cai D et al. Dormancy-associated MADS-box genes and microRNAs jointly control dormancy transition in pear (*Pyrus pyrifolia* white pear group) flower bud. *J Exp Bot*. 2016;**67**: 239–57. <https://doi.org/10.1093/jxb/erv454>.
123. Jiang SH, Sun QG, Chen M et al. Methylome and transcriptome analyses of apple fruit somatic mutations reveal the difference of red phenotype. *BMC Genomics*. 2019;**20**:117. <https://doi.org/10.1186/s12864-019-5499-2>.
124. Zhu YC, Zhang B, Allan AC et al. DNA demethylation is involved in the regulation of temperature-dependent anthocyanin accumulation in peach. *Plant J*. 2020;**102**:965–76. <https://doi.org/10.1111/tpj.14680>.
125. Daccord N, Celton JM, Linsmith G et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet*. 2017;**49**:1099–106. <https://doi.org/10.1038/ng.3886>.
126. Cheng J, Niu Q, Zhang B et al. Downregulation of RdDM during strawberry fruit ripening. *Genome Biol*. 2018;**19**:212. <https://doi.org/10.1186/s13059-018-1587-x>.
127. Qian M, Sun Y, Allan AC et al. The red sport of 'Zaosu' pear and its red-striped pigmentation pattern are associated with demethylation of the PyMYB10 promoter. *Phytochemistry*. 2014;**107**:16–23. <https://doi.org/10.1016/j.phytochem.2014.08.001>.
128. Wang Z, Meng D, Wang A et al. The methylation of the PcMYB10 promoter is associated with green-skinned sport in max red Bartlett pear. *Plant Physiol*. 2013;**162**:885–96. <https://doi.org/10.1104/pp.113.214700>.
129. Smith WW. The course of stone cell formation in pear fruits. *Plant Physiol*. 1935;**10**:587–611. <https://doi.org/10.1104/pp.10.4.587>.
130. Martin-Cabrejas MA, Waldron KW, Selvendran RR et al. Ripening-related changes in the cell walls of Spanish pear (*Pyrus communis*). *Physiol Plant*. 1994;**91**:671–9. <https://doi.org/10.1111/j.1399-3054.1994.tb03004.x>.
131. Tao S, Khanizadeh S, Zhang H et al. Anatomy, ultrastructure and lignin distribution of stone cells in two *Pyrus* species. *Plant Sci*. 2009;**176**:413–9. <https://doi.org/10.1016/j.plantsci.2008.12.011>.
132. Cai Y, Li M, Li D et al. Study of the structure and biosynthetic pathway of lignin in stone cells of pear. *Sci Hortic*. 2010;**125**: 374–9. <https://doi.org/10.1016/j.scienta.2010.04.029>.
133. Jin Q, Yan C, Qiu J et al. Structural characterization and deposition of stone cell lignin in Dangshan Su pear. *Sci Hortic*. 2013;**155**:123–30. <https://doi.org/10.1016/j.scienta.2013.03.020>.
134. Zhang J, Li J, Xue C et al. The variation of stone cell content in 236 germplasms of sand pear (*Pyrus pyrifolia*) and identification of related candidate genes. *Hortic Plant J*. 2021;**7**:108–16. <https://doi.org/10.1016/j.hpj.2020.09.003>.
135. Xue Y-S, Xu S, Z, Xue, C. et al. Pearprocess: a new phenotypic tool for stone cell trait evaluation in pear fruit. *J Integr Agric*. 2020;**19**:1625–34. [https://doi.org/10.1016/s2095-3119\(20\)63193-8](https://doi.org/10.1016/s2095-3119(20)63193-8).
136. Vanholme R, Cesarino I, Rataj K et al. Caffeoyl Shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science*. 2013;**341**:1103–6. <https://doi.org/10.1126/science.1241602>.
137. Zhao Q. Lignification: flexibility, biosynthesis and regulation. *Trends Plant Sci*. 2016;**21**:713–21. <https://doi.org/10.1016/j.tplants.2016.04.006>.
138. Dean JF, Eriksson KEL. Laccase and the deposition of lignin in vascular plants. *Holzforschung*. 1994;**48**:21–33. <https://doi.org/10.1515/hfsg.1994.48.s1.21>.
139. Cao Y, Han Y, Li D et al. Systematic analysis of the 4-coumarate: coenzyme a ligase (4CL) related genes and expression profiling during fruit development in the Chinese pear. *Genes*. 2016;**7**:89. <https://doi.org/10.3390/genes7100089>.
140. Cao Y, Han Y, Meng D et al. Structural, evolutionary, and functional analysis of the class III peroxidase gene family in Chinese pear (*Pyrus bretschneideri*). *Front Plant Sci*. 2016;**7**:1874. <https://doi.org/10.3389/fpls.2016.01874>.
141. Cheng X, Xiong Y, Li DH et al. Bioinformatic and expression analysis of the OMT gene family in *Pyrus bretschneideri* cv. GMR. 2016;**15**:15038664. <https://doi.org/10.4238/gmr.15038664>.
142. Cheng X et al. Characterization and analysis of CCR and CAD gene families at the whole-genome level for lignin synthesis of stone cells in pear (*Pyrus bretschneideri*) fruit. *Biol Open*. 2017;**6**: 1602–13. <https://doi.org/10.1242/bio.026997>.
143. Cheng X, Li G, Ma C et al. Comprehensive genome-wide analysis of the pear (*Pyrus bretschneideri*) laccase gene (PbLAC) family and functional identification of PbLAC1 involved in lignin biosynthesis. *PLoS One*. 2019;**14**:e0210892. <https://doi.org/10.1371/journal.pone.0210892>.
144. Xue C, Yao JL, Xue YS et al. PbrMYB169 positively regulates lignification of stone cells in pear fruit. *J Exp Bot*. 2019;**70**: 1801–14. <https://doi.org/10.1093/jxb/erz039>.
145. Cheng X, Li G, Muhammad A et al. Molecular identification, phylogenomic characterization and expression patterns analysis of the LIM (LIN-11, Isl1 and MEC-3 domains) gene family in pear (*Pyrus bretschneideri*) reveal its potential role in lignin metabolism. *Gene*. 2019;**686**:237–49. <https://doi.org/10.1016/j.gene.2018.11.064>.
146. Su X, Meng T, Zhao Y et al. Comparative genomic analysis of the IDD genes in five Rosaceae species and expression analysis in Chinese white pear (*Pyrus bretschneideri*). *PeerJ*. 2019;**7**:e6628. <https://doi.org/10.7717/peerj.6628>.
147. Cao Y, Han Y, Li D et al. MYB transcription factors in Chinese pear (*Pyrus bretschneideri* Rehd.): genome-wide identification, classification, and expression profiling during fruit development. *Front Plant Sci*. 2016;**7**:577. <https://doi.org/10.3389/fpls.2016.00577>.
148. Gong X, Xie Z, Qi K et al. PbMC1a/1b regulates lignification during stone cell development in pear (*Pyrus bretschneideri*) fruit. *Hortic. Res*. 2020;**7**:59. <https://doi.org/10.1038/s41438-020-0280-x>.
149. Yang YN et al. Expression differences of anthocyanin biosynthesis genes reveal regulation patterns for red pear coloration. *Plant Cell Rep*. 2015;**34**:189–98. <https://doi.org/10.1007/s00299-014-1698-0>.
150. Feng S, Wang Y, Yang S et al. Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. *Planta*. 2010;**232**:245–55. <https://doi.org/10.1007/s00425-010-1170-5>.
151. Zhang X, Allan AC, Yi Q et al. Differential gene expression analysis of Yunnan red pear, *Pyrus pyrifolia*, during fruit skin coloration. *Plant Mol. Biol. Rep*. 2011;**29**:305–14. <https://doi.org/10.1007/s11105-010-0231-z>.
152. Yu B, Zhang D, Huang C et al. Isolation of anthocyanin biosynthetic genes in red Chinese sand pear (*Pyrus pyrifolia* Nakai) and their expression as affected by organ/tissue,



- cultivar, bagging and fruit side. *Sci Hortic.* 2012;**136**:29–37. <https://doi.org/10.1016/j.scienta.2011.12.026>.
153. Zhai R, Wang Z, Zhang S et al. Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (*Pyrus bretschneideri* Rehd.). *J Exp Bot.* 2016;**67**:1275–84. <https://doi.org/10.1093/jxb/erv524>.
  154. Liu H, Su J, Zhu Y et al. The involvement of PybZIPa in light-induced anthocyanin accumulation via the activation of PyUGFT through binding to tandem G-boxes in its promoter. *Hortic. Res.* 2019;**6**:134. <https://doi.org/10.1038/s41438-019-0217-4>.
  155. Wu J, Zhao G, Yang Y et al. Identification of differentially expressed genes related to coloration in red/green mutant pear (*Pyrus communis* L.). *Tree Genet Genomes.* 2013;**9**:75–83. <https://doi.org/10.1007/s11295-012-0534-3>.
  156. Li C, Wu J, Hu KD et al. PyWRKY26 and PybHLH3 cotargeted the PyMYB114 promoter to regulate anthocyanin biosynthesis and transport in red-skinned pears. *Hortic. Res.* 2020;**7**:37. <https://doi.org/10.1038/s41438-020-0254-z>.
  157. Ni J, Bai S, Zhao Y et al. Ethylene response factors Pp4ERF24 and Pp12ERF96 regulate blue light-induced anthocyanin biosynthesis in 'Red Zaosu' pear fruits by interacting with MYB114. *Plant Mol Biol.* 2019;**99**:67–78. <https://doi.org/10.1007/s11103-018-0802-1>.
  158. Tao R, Bai S, Ni J et al. The blue light signal transduction pathway is involved in anthocyanin accumulation in 'Red Zaosu' pear. *Planta.* 2018;**248**:37–48. <https://doi.org/10.1007/s00425-018-2877-y>.
  159. Bai S, Tao R, Tang Y et al. BBX16, a B-box protein, positively regulates light-induced anthocyanin accumulation by activating MYB10 in red pear. *Plant Biotechnol J.* 2019;**17**:1985–97. <https://doi.org/10.1111/pbi.13114>.
  160. Bai S, Tao R, Yin L et al. Two B-box proteins, PpBBX18 and PpBBX21, antagonistically regulate anthocyanin biosynthesis via competitive association with *Pyrus pyrifolia* ELONGATED HYPOCOTYL 5 in the peel of pear fruit. *Plant J.* 2019;**100**:1208–23. <https://doi.org/10.1111/tpj.14510>.
  161. Bialeski RL, Redgwell RJ. Sorbitol metabolism in nectaries from flowers of Rosaceae. *Funct Plant Biol.* 1980;**7**:15–25. <https://doi.org/10.1071/PP9800015>.
  162. Zhang HP, Wu JY, Tao ST et al. Evidence for apoplasmic phloem unloading in pear fruit. *Plant Mol Biol Rep.* 2014;**32**:931–9. <https://doi.org/10.1007/s11105-013-0696-7>.
  163. Cheng R, Cheng Y, Lu J et al. The gene PbTMT4 from pear (*Pyrus bretschneideri*) mediates vacuolar sugar transport and strongly affects sugar accumulation in fruit. *Physiol Plant.* 2018;**164**:307–19. <https://doi.org/10.1111/ppl.12742>.
  164. Wang LF, Qi XX, Huang XS et al. Overexpression of sucrose transporter gene PbSUT2 from *Pyrus bretschneideri*, enhances sucrose content in *Solanum lycopersicum* fruit. *Plant Physiol Biochem.* 2016;**105**:150–61. <https://doi.org/10.1016/j.plaphy.2016.04.019>.
  165. Oura Y, Yamada K, Shiratake K et al. Purification and characterization of a NAD(+) dependent sorbitol dehydrogenase from Japanese pear fruit. *Phytochemistry.* 2000;**54**:567–72. [https://doi.org/10.1016/S0031-9422\(00\)00158-8](https://doi.org/10.1016/S0031-9422(00)00158-8).
  166. Kim HY, Ahn JC, Choi JH et al. Expression and cloning of the full-length cDNA for sorbitol-6-phosphate dehydrogenase and NAD-dependent sorbitol dehydrogenase from pear (*Pyrus pyrifolia* N.). *Sci Hortic.* 2007;**112**:406–12. <https://doi.org/10.1016/j.scienta.2007.01.015>.
  167. Moriguchi T, Abe K, Sanada T et al. Levels and role of sucrose synthase, sucrose-phosphate synthase, and acid invertase in sucrose accumulation in fruit of Asian pear. *J Am Soc Hortic Sci.* 1992;**117**:274–8. <https://doi.org/10.21273/JASHS.117.2.274>.
  168. Moore B, Zhou L, Rolland F et al. Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science.* 2003;**300**:332–6. <https://doi.org/10.1126/science.1080585>.
  169. Kim YM, Heinzel N, Giese JO et al. A dual role of tobacco hexokinase 1 in primary metabolism and sugar sensing. *Plant Cell Environ.* 2013;**36**:1311–27. <https://doi.org/10.1111/pce.12060>.
  170. Zhao B, Qi K, Yi X et al. Identification of hexokinase family members in pear (*Pyrus × bretschneideri*) and functional exploration of PbHXK1 in modulating sugar content and plant growth. *Gene.* 2019;**711**:143932. <https://doi.org/10.1016/j.gene.2019.06.022>.
  171. Chen J, Wang Z, Wu J et al. Chemical compositional characterization of eight pear cultivars grown in China. *Food Chem.* 2007;**104**:268–75. <https://doi.org/10.1016/j.foodchem.2006.11.038>.
  172. Popova TN, Pinheiro de Carvalho MA. Citrate and isocitrate in plant metabolism. *Biochim Biophys Acta.* 1998;**1364**:307–25. [https://doi.org/10.1016/S0005-2728\(98\)00008-5](https://doi.org/10.1016/S0005-2728(98)00008-5).
  173. Sadka A, Dahan E, Or E et al. NADP+-isocitrate dehydrogenase gene expression and isozyme activity during citrus fruit development. *Plant Sci.* 2000;**158**:173–81. [https://doi.org/10.1016/S0168-9452\(00\)00328-9](https://doi.org/10.1016/S0168-9452(00)00328-9).
  174. Sadka A, Dahan E, Cohen L et al. Aconitase activity and expression during the development of lemon fruit. *Physiol Plant.* 2000;**108**:255–62. <https://doi.org/10.1034/j.1399-3054.2000.108003255.x>.
  175. Canel C, Bailey-Serres JN, Roose ML. Molecular characterization of the mitochondrial citrate synthase gene of an acidless pummelo (*Citrus maxima*). *Plant Mol Biol.* 1996;**31**:143–7. <https://doi.org/10.1007/BF00020613>.
  176. Suzuki Y, Kanayama Y, Shiratake K et al. Vacuolar H(+)-pyrophosphatase purified from pear fruit. *Phytochemistry.* 1999;**50**:535–9. [https://doi.org/10.1016/S0031-9422\(98\)00554-8](https://doi.org/10.1016/S0031-9422(98)00554-8).
  177. Hosaka M, Kanayama Y, Shiratake K et al. Tonoplast H+-ATPase of mature pear fruit. *Phytochemistry.* 1994;**36**:565–7. [https://doi.org/10.1016/S0031-9422\(00\)89775-7](https://doi.org/10.1016/S0031-9422(00)89775-7).
  178. Suzuki Y, Shiratake K, Yamaki S. Seasonal changes in the activities of vacuolar H+-pumps and their gene expression in the developing Japanese pear fruit. *Engei Gakkai zasshi.* 2000;**69**:15–21. <https://doi.org/10.2503/jjshs.69.15>.
  179. Hiratake K, Kanayama Y, Maeshima M et al. Changes in H(+)-pumps and a tonoplast intrinsic protein of vacuolar membranes during the development of pear fruit. *Plant Cell Physiol.* 1997;**38**:1039–45. <https://doi.org/10.1093/oxfordjournals.pcp.a029269>.
  180. Yasuo S, Masayoshi M, Shohei Y. Molecular cloning of vacuolar H+-pyrophosphatase and its expression during the development of pear fruit. *Plant Cell Physiol.* 1999;**40**:900–4. <https://doi.org/10.1093/oxfordjournals.pcp.a029620>.
  181. Lu XP, Liu YZ, Zhou GF et al. Identification of organic acid-related genes and their expression profiles in two pear (*Pyrus pyrifolia*) cultivars with difference in predominant acid type at fruit ripening stage. *Sci Hortic.* 2011;**129**:680–7. <https://doi.org/10.1016/j.scienta.2011.05.014>.
  182. Willner B, Granvogl M, Schieberle P. Characterization of the key aroma compounds in Bartlett pear brandies by means of the sensomics concept. *J Agric Food Chem.* 2013;**61**:9583–93. <https://doi.org/10.1021/jf403024t>.

183. Wang C, Zhang W, Li H et al. Analysis of volatile compounds in pears by HS-SPME-GC×GC-TOFMS. *Molecules*. 2019;**24**:1795. <https://doi.org/10.3390/molecules24091795>.
184. El Hadi MA, Zhang FJ, Wu FF et al. Advances in fruit aroma volatile research. *Molecules*. 2013;**18**:8200–29. <https://doi.org/10.3390/molecules18078200>.
185. Shi F, Zhou X, Yao MM et al. Low-temperature stress-induced aroma loss by regulating fatty acid metabolism pathway in 'Nanguo' pear. *Food Chem*. 2019;**297**:124927. <https://doi.org/10.1016/j.foodchem.2019.05.201>.
186. Shi F, Zhou X, Zhou Q et al. Transcriptome analyses provide new possible mechanisms of aroma ester weakening of 'Nanguo' pear after cold storage. *Sci Hortic*. 2018;**237**:247–56. <https://doi.org/10.1016/j.scienta.2018.04.013>.
187. Li M, Dunwell JM, Zhang H et al. Network analysis reveals the co-expression of sugar and aroma genes in the Chinese white pear (*Pyrus bretschneideri*). *Gene*. 2018;**677**:370–7. <https://doi.org/10.1016/j.gene.2018.08.042>.
188. Wei S, Tao S, Qin S et al. Transcriptome profiling reveals the candidate genes associated with aroma metabolites and emission of pear (*Pyrus ussuriensis* cv.). *Sci Hortic*. 2016;**206**:33–42. <https://doi.org/10.1016/j.scienta.2016.04.019>.
189. Luo M, Zhou X, Sun H et al. Insights into profiling of volatile ester and LOX-pathway related gene families accompanying post-harvest ripening of 'Nanguo' pears. *Food Chem*. 2021;**335**:127665. <https://doi.org/10.1016/j.foodchem.2020.127665>.
190. Gonzalez M, Gaete-Eastman C, Valdenegro M et al. Aroma development during ripening of *Fragaria chiloensis* fruit and participation of an alcohol acyltransferase (F<sub>6</sub>AT1) gene. *J Agric Food Chem*. 2009;**57**:9123–32. <https://doi.org/10.1021/jf901693j>.
191. Zhou D, Sun Y, Li M et al. Postharvest hot air and UV-C treatments enhance aroma-related volatiles by simulating the lipoxygenase pathway in peaches during cold storage. *Food Chem*. 2019;**292**:294–303. <https://doi.org/10.1016/j.foodchem.2019.04.049>.
192. Singh RK, Srivastava S, Chidley HG et al. Overexpression of mango alcohol dehydrogenase (MiADH1) mimics hypoxia in transgenic tomato and alters fruit flavor components. *Agri gene*. 2018;**7**:23–33. <https://doi.org/10.1016/j.aggene.2017.10.003>.
193. Bapat VA, Trivedi PK, Ghosh A et al. Ripening of fleshy fruit: molecular insight and the role of ethylene. *Biotechnol Adv*. 2010;**28**:94–107. <https://doi.org/10.1016/j.biotechadv.2009.10.002>.
194. Lelievre J-M, Tichit L, Dao P et al. Effects of chilling on the expression of ethylene biosynthetic genes in Passe-Crassane pear (*Pyrus communis* L.) fruits. *Plant Mol Biol*. 1997;**33**:847–55. <https://doi.org/10.1023/A:1005750324531>.
195. Yuan H, Zhang L, Jiang Z et al. Characterization of ripening-related PuARP4 in pear (*Pyrus ussuriensis*). *J Plant Growth Regul*. 2017;**36**:766–72. <https://doi.org/10.1007/s00344-017-9680-z>.
196. Song L, Wang Z, Wang Z et al. Screening of cell wall-related genes that are expressed differentially during ripening of pears with different softening characteristics. *Postharvest Biol Technol*. 2016;**115**:1–8. <https://doi.org/10.1016/j.postharvbio.2015.12.012>.
197. Mercy WM, Mathooko FM, Matsuzaki M et al. Expression characteristics of seven members of the  $\beta$ -galactosidase gene family in 'La France' pear (*Pyrus communis* L.) fruit during growth and their regulation by 1-methylcyclopropene during postharvest ripening. *Postharvest Biol Technol*. 2005;**36**:253–63. <https://doi.org/10.1016/j.postharvbio.2005.02.002>.
198. Sekine D, Munemura I, Gao M et al. Cloning of cDNAs encoding cell-wall hydrolases from pear (*Pyrus communis*) fruit and their involvement in fruit softening and development of melting texture. *Physiol Plant*. 2006;**126**:163–74. <https://doi.org/10.1111/j.1399-3054.2006.00583.x>.
199. Hiwasa K, Rose JKC, Nakano R et al. Differential expression of seven  $\alpha$ -expansin genes during growth and ripening of pear fruit. *Physiol Plant*. 2003;**117**:564–72. <https://doi.org/10.1034/j.1399-3054.2003.00064.x>.
200. Peil A. Inoculation of *Malus × robusta* 5 progeny with a strain breaking resistance to fire blight reveals a minor QTL on LG5. *Acta Hortic*. 2011;**896**:357–62. <https://doi.org/10.17660/actahortic.2011.896.49>.
201. Percheplid L, Guerif P, Ravon E et al. Polygenic inheritance of resistance to *Cacopsylla pyri* in a *Pyrus communis* × *P. ussuriensis* progeny is explained by three QTLs involving an epistatic interaction. *Tree Genet Genomes*. 2016;**12**:108. <https://doi.org/10.1007/s11295-016-1072-1>.
202. De Franceschi, P. & Dondini, L. Molecular mapping of major genes and QTLs in pear. In: Korban, S.S. (ed.) *The Pear Genome*, pp 113–131, Cham, Switzerland. [https://doi.org/10.1007/978-3-030-11048-2\\_6](https://doi.org/10.1007/978-3-030-11048-2_6) (2019).
203. Bokszczanin K, Dondini L, Przybyla AA. First report on the presence of fire blight resistance in linkage group 11 of *Pyrus ussuriensis* maxim. *J Appl Genet*. 2009;**50**:99–104. <https://doi.org/10.1007/bf03195660>.
204. Montanari S, Percheplid L, Renault D et al. A QTL detected in an interspecific pear population confers stable fire blight resistance across different environments and genetic backgrounds. *Mol Breed*. 2016;**36**:47. <https://doi.org/10.1007/s11032-016-0473-z>.
205. Abe K, Saito T, Terai O et al. Genotypic difference for the susceptibility of Japanese, Chinese and European pears to *Venturia nashicola*, the cause of scab on Asian pears. *Plant Breed*. 2008;**127**:407–12. <https://doi.org/10.1111/j.1439-0523.2007.01482.x>.
206. Bouvier L, Bourcy M, Boulay M et al. A new pear scab resistance gene *Rvp1* from the European pear cultivar 'Navara' maps in a genomic region syntenic to an apple scab resistance gene cluster on linkage group 2. *Tree Genet Genomes*. 2012;**8**:53–60. <https://doi.org/10.1007/s11295-011-0419-x>.
207. Liu Y, Yang T, Lin Z et al. A WRKY transcription factor PbrWRKY53 from *Pyrus betulaefolia* is involved in drought tolerance and AsA accumulation. *Plant Biotechnol J*. 2019;**17**:1770–87. <https://doi.org/10.1111/pbi.13099>.
208. Li K, Xing C, Yao Z et al. PbrMYB21, a novel MYB protein of *Pyrus betulaefolia*, functions in drought tolerance and modulates polyamine levels by regulating arginine decarboxylase gene. *Plant Biotechnol J*. 2017;**15**:1186–203. <https://doi.org/10.1111/pbi.12708>.
209. Xing C, Liu Y, Zhao L et al. A novel MYB transcription factor regulates AsA synthesis and effects cold tolerance. *Plant Cell Environ*. 2019;**42**:832–45. <https://doi.org/10.1111/pce.13387>.
210. Jin C, Huang XS, Li KQ et al. Overexpression of a bHLH1 transcription factor of *Pyrus ussuriensis* confers enhanced cold tolerance and increases expression of stress-responsive genes. *Front Plant Sci*. 2016;**7**:441. <https://doi.org/10.3389/fpls.2016.00441>.
211. Jin C, Li KQ, Xu X et al. A novel NAC transcription factor, PbeNAC1, of *Pyrus betulaefolia* confers cold and drought tolerance via interacting with PbeDREBs and activating the expression of stress-responsive genes. *Front Plant Sci*. 2017;**8**:1049. <https://doi.org/10.3389/fpls.2017.01049>.

212. Dong H, Wang C, Xing C et al. Overexpression of PbrNHX2 gene, a Na<sup>+</sup>/H<sup>+</sup> antiporter gene isolated from *Pyrus betulaefolia*, confers enhanced tolerance to salt stress via modulating ROS levels. *Plant Sci.* 2019;**285**:14–25. <https://doi.org/10.1016/j.plantsci.2019.04.021>.
213. Wang CH, Li W, Tian YK et al. Development of molecular markers for genetic and physical mapping of the PcDw locus in pear (*Pyrus communis* L.). *J Hortic Sci Biotechnol.* 2016;**91**: 299–307. <https://doi.org/10.1080/14620316.2016.1155319>.
214. Xiao Y, Wang C, Tian Y et al. Candidates responsible for dwarf pear phenotype as revealed by comparative transcriptome analysis. *Mol Breed.* 2019;**39**:1. <https://doi.org/10.1007/s11032-018-0907-x>.
215. Foster TM, Celton JM, Chagne D et al. Two quantitative trait loci, Dw1 and Dw2, are primarily responsible for rootstock-induced dwarfing in apple. *Hortic. Res.* 2015;**2**:15001. <https://doi.org/10.1038/hortres.2015.1>.
216. Knäbel M, Friend AP, Palmer JW et al. Genetic control of pear rootstock-induced dwarfing and precocity is linked to a chromosomal region syntenic to the apple Dw1 loci. *BMC Plant Biol.* 2015;**15**:230. <https://doi.org/10.1186/s12870-015-0620-4>.
217. Celton J-M, Chagne D, Tustin SD et al. Update on comparative genome mapping between *Malus* and *Pyrus*. *BMC Res Notes.* 2009;**2**:182–2. <https://doi.org/10.1186/1756-0500-2-182>.
218. Nekrasov V, Staskawicz B, Weigel D et al. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol.* 2013;**31**:691–3. <https://doi.org/10.1038/nbt.2655>.
219. Chandrasekaran J, Brumin M, Wolf D et al. Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol.* 2016;**17**:1140–53. <https://doi.org/10.1111/mpp.12375>.
220. Peng A, Chen S, Lei T et al. Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnol J.* 2017;**15**: 1509–19. <https://doi.org/10.1111/pbi.12733>.
221. Tashkandi M, Ali Z, Aljedaani F et al. Engineering resistance against tomato yellow leaf curl virus via the CRISPR/Cas9 system in tomato. *Plant Signal Behav.* 2018;**13**:e1525996. <https://doi.org/10.1080/15592324.2018.1525996>.
222. Tang X, Ren Q, Yang L et al. Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnol J.* 2019;**17**:1431–45. <https://doi.org/10.1111/pbi.13068>.
223. Endo A, Masafumi M, Kaya H et al. Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Sci Rep.* 2016;**6**:38169. <https://doi.org/10.1038/srep38169>.
224. Malzahn AA, Tang X, Lee K et al. Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and *Arabidopsis*. *BMC Biol.* 2019;**17**:9. <https://doi.org/10.1186/s12915-019-0629-5>.
225. Xu J, Hua K, Lang Z. Genome editing for horticultural crop improvement. *Hortic. Res.* 2019;**6**:113. <https://doi.org/10.1038/s41438-019-0196-5>.
226. Kaur N, Alok A, Kaur N et al. CRISPR/Cas9-mediated efficient editing in phytoene desaturase (PDS) demonstrates precise manipulation in banana cv. Rasthali genome. *Funct. Integr. Genomics.* 2018;**18**:89–99. <https://doi.org/10.1007/s10142-017-0577-5>.
227. Wang Z, Wang S, Li D et al. Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. *Plant Biotechnol J.* 2018;**16**: 1424–33. <https://doi.org/10.1111/pbi.12884>.
228. Wang X, Tu M, Wang D et al. CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. *Plant Biotechnol J.* 2018;**16**:844–55. <https://doi.org/10.1111/pbi.12832>.
229. Zhang F, LeBlanc C, Irish VF et al. Rapid and efficient CRISPR/Cas9 gene editing in citrus using the YAO promoter. *Plant Cell Rep.* 2017;**36**:1883–7. <https://doi.org/10.1007/s00299-017-2202-4>.
230. Nishitani C, Hirai N, Komori S et al. Efficient genome editing in apple using a CRISPR/Cas9 system. *Sci Rep.* 2016;**6**:31481. <https://doi.org/10.1038/srep31481>.
231. Xing S, Jia M, Wei L et al. CRISPR/Cas9-introduced single and multiple mutagenesis in strawberry. *J Genet Genomics.* 2018;**45**: 685–7. <https://doi.org/10.1016/j.jgg.2018.04.006>.
232. Zhou J, Wang G, Liu Z. Efficient genome editing of wild strawberry genes, vector development and validation. *Plant Biotechnol J.* 2018;**16**:1868–77. <https://doi.org/10.1111/pbi.12922>.
233. Malnoy M, Viola R, Jung MH et al. DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Front Plant Sci.* 2016;**7**:1904. <https://doi.org/10.3389/fpls.2016.01904>.
234. Arkonyi-Gasic E, Wang T, Voogd C et al. Mutagenesis of kiwifruit CENTRORADIALIS-like genes transforms a climbing woody perennial with long juvenility and axillary flowering into a compact plant with rapid terminal flowering. *Plant Biotechnol J.* 2019;**17**:869–80. <https://doi.org/10.1111/pbi.13021>.
235. Charrier A, Vergne E, Dousset N et al. Efficient targeted mutagenesis in apple and first time edition of pear using the CRISPR-Cas9 system. *Front Plant Sci.* 2019;**10**:40. <https://doi.org/10.3389/fpls.2019.00040>.
236. Wang X, Gao L, Jiao C et al. Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat Commun.* 2020;**11**:5817. <https://doi.org/10.1038/s41467-020-19682-0>.
237. Guo J, Cao K, Deng C et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol.* 2020;**21**:258. <https://doi.org/10.1186/s13059-020-02169-y>.
238. Zhou Y, Minio A, Massonnet M et al. The population genetics of structural variants in grapevine domestication. *Nat Plants.* 2019;**5**:965–79. <https://doi.org/10.1038/s41477-019-0507-8>.



RESEARCH ARTICLE

Open Access



# Rearrangement and domestication as drivers of Rosaceae mitogenome plasticity

Manyi Sun<sup>1†</sup>, Mingyue Zhang<sup>2†</sup>, Xuening Chen<sup>1</sup>, Yueyuan Liu<sup>1</sup>, Binbin Liu<sup>3,4</sup>, Jiaming Li<sup>1</sup>, Runze Wang<sup>1</sup>, Kejiao Zhao<sup>1</sup> and Jun Wu<sup>1\*</sup>

## Abstract

**Background:** The mitochondrion is an important cellular component in plants and that functions in producing vital energy for the cell. However, the evolution and structure of mitochondrial genomes (mitogenomes) remain unclear in the Rosaceae family. In this study, we assembled 34 Rosaceae mitogenomes and characterized genome variation, rearrangement rate, and selection signal variation within these mitogenomes.

**Results:** Comparative analysis of six genera from the Amygdaloideae and five from the Rosoideae subfamilies of Rosaceae revealed that three protein-coding genes were absent from the mitogenomes of five Rosoideae genera. Positive correlations between genome size and repeat content were identified in 38 Rosaceae mitogenomes. Twenty repeats with high recombination frequency (> 50%) provided evidence for predominant substoichiometric conformation of the mitogenomes. Variations in rearrangement rates were identified between eleven genera, and within the *Pyrus*, *Malus*, *Prunus*, and *Fragaria* genera. Based on population data, phylogenetic inferences from *Pyrus* mitogenomes supported two distinct maternal lineages of Asian cultivated pears. A *Pyrus*-specific deletion (DEL-D) in selective sweeps was identified based on the assembled genomes and population data. After the DEL-D sequence fragments originally arose, they may have experienced a subsequent doubling event via homologous recombination and sequence transfer in the Amygdaloideae; afterwards, this variant sequence may have significantly expanded to cultivated groups, thereby improving adaptation during the domestication process.

**Conclusions:** This study characterizes the variations in gene content, genome size, rearrangement rate, and the impact of domestication in Rosaceae mitogenomes and provides insights into their structural variation patterns and phylogenetic relationships.

**Keywords:** Mitogenome, Rosaceae, Rearrangement rate, Domestication

## Background

As the cell's energy factory, the mitochondrion is an organelle essential in angiosperm development, growth, programmed cell death, and male sterility [1]. Each

mitochondrion has its own genome, which is usually uniparentally inherited [2]. Compared with plastid genomes, angiosperm mitogenomes vary in size and gene content [3, 4]. Currently known angiosperm mitochondrial genome (mitogenome) sizes range from 66 kb to 11.3 Mb, and the number of protein-coding genes ranges from 19 to 41 (excluding duplicated genes and open reading frames (ORFs)) [4–6]. Most genome size and structure variations occur in non-coding sequences, and these variations are primarily caused by foreign sequence importation, which increases the occurrence of repetitive sequences and recombination events [7–9].

<sup>†</sup>Manyi Sun and Mingyue Zhang contributed equally to this work.

\*Correspondence: wujun@njau.edu.cn

<sup>1</sup> College of Horticulture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Numerous inverted and direct repeats play a pivotal role in plant mitogenome size and structural evolution by participating in genome rearrangement, repeat-mediated recombination, insertion, and deletion events [8, 10, 11]. Repeat-mediated homologous recombination in mitogenomes has been investigated in angiosperm plants such as *Picea abies* [12] and *Nymphaea colorata* [13], and positive correlations between repeat length and recombination rate were detected in *Viscum scurruloideum* [4]. Minor to moderate recombination activity was detected among short (< 100 bp) and medium length repeats (100–1000 bp), while larger repeats (> 1000 bp) experienced more frequent recombination activity and isomerization in the genome [14, 15]. Recently, third-generation long-read sequencing technologies have been used to overcome the complexity of short read-based genome assembly, and this technology has proven sensitive at detecting the repeat-mediated recombination activity of large repeats [12, 13].

Mitogenome rearrangement is primarily caused by frequent repeat-mediated recombination [11], supported by the presence of rearrangement breakpoints close to repeats [8]. In plants, mitogenome rearrangements can influence ATP availability, plant growth, cytoplasmic male sterility (CMS), and overall fitness [16, 17]. Aside from the low substitution rate, mitogenomes in angiosperms have obviously different rearrangement rates. Within the genus *Monsonia*, the mitogenome of *M. ciliata* has a tenfold higher rearrangement rate than its sister species; overall, an over 600-fold variance in mitogenome rearrangement rates has been observed among seed plants [8].

Rosaceae has ca. 3000 species in 90 genera and includes herbs, shrubs, and trees adapted to a wide variety of environments [18]. Research on Rosaceae mitogenomes has remained limited despite recent progress in nuclear and chloroplast genomic sequencing of Rosaceae species [19–21]. Only 14 Rosaceae mitogenomes were available in the National Center for Biotechnology Information (NCBI) database (last access date: 20 January 2022) (Additional file 1). The evolution and divergence of Rosaceae mitogenomes remain unclear, and limited genetic information regarding Rosaceae mitogenome analysis exists. Third-generation long-read sequencing technologies and a series of assembly software like GetOrganelle [22], SOAPdenovo [23], and Canu [24] have provided the ability to assemble complete mitogenomes. In addition, many Rosaceae nuclear genomes can only employ different parental inheritance modes, such as in the *Malus*, *Pyrus*, and *Sorbus* genera; information on mitogenomes inherited from the maternal parent provides a chance to determine additional population and domestication information [25, 26].

Here, 38 complete Rosaceae mitogenomes were assembled and annotated (four of which were previously released). Variations in genes and repeat sequences were identified, and recombination and rearrangement events were investigated to explore the expansion and evolution patterns of Rosaceae mitogenomes. Subsequently, short-read sequencing data from 139 pear and 116 apple accessions was used to explore the genetic variations, phylogenetic relationships, and domestication processes in the *Pyrus* and *Malus* genera. We found that domestication and selection contributed to the variations in the mitogenomes of members of the Rosaceae family and resulted in the spread of structurally varied gene sequences.

## Results

### Profile of the mitogenomes of Rosaceae species

In this study, each of the 34 Rosaceae mitogenomes was de novo assembled into a single completely gapless contig with an average coverage depth of 323.52–6550.87× (Additional file 2). Coupled with the four previously released genomes, the sizes of the 38 Rosaceae mitogenomes ranged from 277.76 kb (*Rosa chinensis*: RoChi) to 535.73 kb (*Prunus mume*: Pmum) (Table 1). Genome sizes within the Amygdaloideae varied by up to 150.75 kb (*Sorbus aucuparia*: Sauc (384.98 kb) vs Pmum (535.73 kb)), and this variation increased up to 194.38 kb in Rosoideae species (Table 1). Twenty-four genes appeared in all 38 mitogenomes. Compared with six genera in Amygdaloideae, three protein-coding genes (*rpl5*, *rpl16*, and *sdh3*) were completely lost from the mitogenomes of five genera of the Rosoideae (Fig. 1). Within Rosoideae, *rps14* was lost in *Rosa*, *Geum*, *Potentilla*, and *Fragaria*, and *rps12* was lost in *Geum*, *Potentilla*, and *Fragaria*. *Rpl10* was lost in *Fragaria*, and *rps1* was lost in *Rosa* and *Potentilla*. In six Amygdaloideae genera, *rps7* was lost in *Sorbus*, *Photinia*, *Malus*, *Eriobotrya*, and *Pyrus*. Two varieties of *Pyrus bretschneideri* (“Yali”: Pbre-Y and “Dangshansuli” Pbre-D) contained copies of the *atp9* and *ccmB* genes. *Malus sylvestris* (Msyl) and *Malus domestica* (“Gala”: Mdom-G; “Yantai fuji 8”: Mdom-Y) contained copies of the *26rrn* and *rps12* genes. The GC content was relatively stable, averaging at about 45% in the 38 Rosaceae mitogenomes, except for the *Rubus chingii* (Ruchi) mitogenome (43.31%), which had a higher percent of chloroplast sequence imports relative to other species (Table 1).

### Repeat sequence variation and correlations between genome size and repeat sequences

In the 38 mitogenomes, total repeat number changes might be caused by short (< 100 bp) repeat sequences (Fig. 2a; Additional file 3). The number of repeat

**Table 1** Summary of 38 Rosaceae mitogenomes

| ID     | Species name                                       | Subfamily     | GenBank accession number | Genome size (bp) | Number of protein-coding genes | Number of rRNA | Number of tRNA | GC content (%) | Length of plastid-derived sequences (bp/%) |
|--------|--|---------------|--------------------------|------------------|--------------------------------|----------------|----------------|----------------|--|
| Ejap   | <i>Eriobotrya japonica</i>                         | Amygdaloideae | NC_045228                | 434,980          | 34                             | 3              | 23             | 45.42          | 2039 (0.47)                                |
| Fana-C | <i>Fragaria ananassa</i> ("Camarosa") <sup>a</sup> | Rosoideae     | OM763767                 | 285,543          | 30                             | 3              | 17             | 44.96          | 19,176 (6.72)                              |
| Fana-R | <i>Fragaria ananassa</i> ("Royal Royce")           | Rosoideae     | OM763768                 | 285,046          | 29                             | 3              | 18             | 44.98          | 18,507 (6.49)                              |
| Fiin   | <i>Fragaria iinumae</i>                            | Rosoideae     | OM763769                 | 329,263          | 30                             | 3              | 20             | 45.26          | 26,297 (7.99)                              |
| Fman   | <i>Fragaria mand-schurica</i>                      | Rosoideae     | ON478154                 | 313,022          | 30                             | 3              | 21             | 45.20          | 14,598 (4.66)                              |
| Fnil   | <i>Fragaria nilgerrensis</i>                       | Rosoideae     | OM763770                 | 315,210          | 29                             | 3              | 19             | 45.26          | 11,742 (3.73)                              |
| Fpen   | <i>Fragaria pentaphylla</i>                        | Rosoideae     | ON478155                 | 290,970          | 30                             | 3              | 19             | 45.07          | 12,360 (4.25)                              |
| Fves   | <i>Fragaria vesca</i>                              | Rosoideae     | ON478179                 | 312,993          | 30                             | 3              | 20             | 45.20          | 14,340 (4.58)                              |
| Fvir   | <i>Fragaria viridis</i>                            | Rosoideae     | ON478156                 | 289,146          | 30                             | 3              | 18             | 45.23          | 12,354 (4.27)                              |
| Gurb   | <i>Geum urbanum</i>                                | Rosoideae     | ON556624                 | 335,549          | 31                             | 4              | 23             | 44.42          | 30,222 (9.01)                              |
| Mbac   | <i>Malus baccata</i>                               | Amygdaloideae | ON478159                 | 400,769          | 33                             | 3              | 20             | 45.40          | 4049 (1.01)                                |
| Mdom-G | <i>Malus domestica</i> ("Gala")                    | Amygdaloideae | ON478160                 | 396,946          | 35                             | 4              | 21             | 45.41          | 2187 (0.55)                                |
| Mdom-Y | <i>Malus domestica</i> ("Yantai fuji 8")           | Amygdaloideae | MN964891                 | 396,947          | 32                             | 4              | 20             | 45.40          | 1968 (0.50)                                |
| Msie   | <i>Malus sieversii</i>                             | Amygdaloideae | ON478161                 | 385,869          | 34                             | 3              | 21             | 45.38          | 2140 (0.55)                                |
| Msyl   | <i>Malus sylvestris</i>                            | Amygdaloideae | ON478162                 | 396,940          | 35                             | 4              | 21             | 45.41          | 1968 (0.50)                                |
| Pans   | <i>Potentilla anserina</i>                         | Rosoideae     | ON478170                 | 294,682          | 28                             | 3              | 21             | 44.46          | 14,320 (4.86)                              |
| Parm   | <i>Prunus armeniaca</i>                            | Amygdaloideae | ON478164                 | 510,346          | 36                             | 3              | 24             | 45.43          | 5483 (1.07)                                |
| Pavi-G | <i>Prunus avium</i> ("Glory")                      | Amygdaloideae | ON478157                 | 444,576          | 36                             | 3              | 23             | 45.62          | 3887 (0.87)                                |
| Pavi-S | <i>Prunus avium</i> ("Staccato")                   | Amygdaloideae | ON478178                 | 444,576          | 36                             | 3              | 23             | 45.62          | 3887 (0.87)                                |
| Pbet   | <i>Pyrus betulifolia</i>                           | Amygdaloideae | ON478165                 | 432,493          | 35                             | 3              | 20             | 45.21          | 2039 (0.47)                                |
| Pbre-D | <i>Pyrus bretschneideri</i> ("Dangshansuli")       | Amygdaloideae | OM763766                 | 458,897          | 37                             | 3              | 21             | 45.21          | 3901 (0.85)                                |
| Pbre-Y | <i>Pyrus bretschneideri</i> ("Yali")               | Amygdaloideae | ON478180                 | 458,895          | 37                             | 3              | 21             | 45.21          | 3900 (0.85)                                |
| Pcom   | <i>Pyrus communis</i>                              | Amygdaloideae | ON478166                 | 443,525          | 35                             | 3              | 20             | 45.24          | 4195 (0.95)                                |
| Pkan   | <i>Prunus kanzakura</i>                            | Amygdaloideae | ON478167                 | 422,215          | 35                             | 3              | 27             | 45.54          | 3975 (0.94)                                |
| Pmir   | <i>Prunus mira</i>                                 | Amygdaloideae | ON478168                 | 429,732          | 35                             | 3              | 28             | 45.59          | 2320 (0.54)                                |
| Pmum   | <i>Prunus mume</i>                                 | Amygdaloideae | ON478169                 | 535,727          | 35                             | 3              | 28             | 45.45          | 5768 (1.08)                                |



**Table 1** (continued)

| ID    | Species name   | Subfamily     | GenBank accession number | Genome size (bp) | Number of protein-coding genes | Number of rRNA | Number of tRNA | GC content (%) | Length of plastid-derived sequences (bp/%) |
|-------|--|---------------|--------------------------|------------------|--------------------------------|----------------|----------------|----------------|--|
| Psal  | <i>Prunus salicina</i>   | Amygdaloideae | ON478171                 | 508,005          | 39                             | 3              | 23             | 45.43          | 3057 (0.60)                                |
| Pser  | <i>Photinia ser-ratifolia</i>                                      | Amygdaloideae | ON556623                 | 473,561          | 34                             | 4              | 21             | 45.25          | 2062 (0.44)                                |
| Psib  | <i>Prunus sibirica</i>   | Amygdaloideae | ON478172                 | 510,187          | 35                             | 3              | 26             | 45.42          | 5905 (1.16)                                |
| Pyed  | <i>Prunus yedoensis</i>  | Amygdaloideae | ON478173                 | 456,900          | 35                             | 3              | 26             | 45.56          | 3883 (0.85)                                |
| Pysb  | <i>Pyrus sinkiangensis</i> x <i>bretschneideri</i> ("Hongxiangsu") | Amygdaloideae | ON478158                 | 441,852          | 35                             | 3              | 20             | 45.21          | 2039 (0.46)                                |
| Pyuc  | <i>Pyrus ussuriensis</i> x <i>communis</i> ("No.1 Zhonghai")       | Amygdaloideae | ON478163                 | 441,853          | 35                             | 3              | 20             | 45.21          | 2039 (0.46)                                |
| Rochi | <i>Rosa chinensis</i>  | Rosoideae     | ON478174                 | 277,763          | 31                             | 3              | 18             | 45.35          | 6000 (2.16)                                |
| Rorug | <i>Rosa rugosa</i>   | Rosoideae     | ON478175                 | 302,947          | 31                             | 3              | 20             | 45.24          | 8863 (2.93)                                |
| Ruchi | <i>Rubus chingii</i>   | Rosoideae     | ON478176                 | 472,138          | 32                             | 3              | 35             | 43.31          | 77,163 (16.34)                             |
| Sauc  | <i>Sorbus aucuparia</i>  | Amygdaloideae | NC_052880                | 384,977          | 33                             | 2              | 23             | 45.39          | 3763 (0.98)                                |
| Spoh  | <i>Sorbus pohuashanensis</i>                                       | Amygdaloideae | ON478177                 | 396,857          | 32                             | 3              | 20             | 45.36          | 3044 (0.77)                                |
| Stor  | <i>Sorbus torminalis</i>   | Amygdaloideae | NC_052879                | 386,758          | 31                             | 3              | 18             | 45.31          | 3837 (0.99)                                |

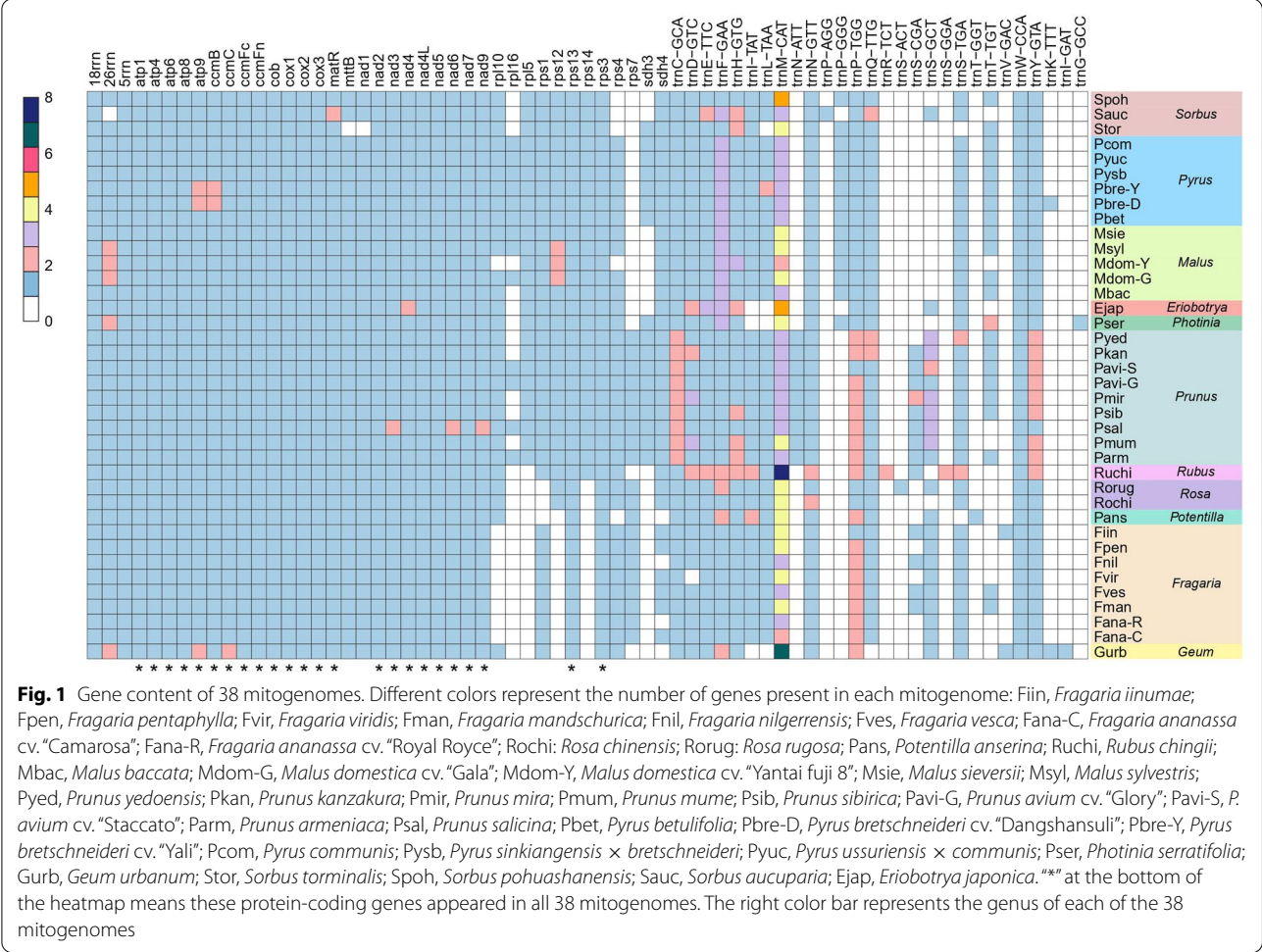
<sup>a</sup> The accession name is shown in parentheses

sequences ranged from 112 (*Fragaria ananassa* cv. "Camarosa": Fana-C) to 457 (Pmum), and 73.33–90.98% of repeats were less than 100 bp in length (Fig. 2a; Additional file 3). Among species of the Amygdaloideae, the number of short (< 100 bp) repeats in *Prunus* samples was significantly higher than samples from five other genera (*Photinia*, *Malus*, *Pyrus*, *Sorbus*, and *Prunus*) (*t*-test, *P*-value = 5.64e−8), while the number of repeats longer than 100 bp was not significantly increased (*t*-test, *P*-value = 0.11) (Additional file 3). In Rosoideae, the total repeat number (296) of Ruchi was higher than that of samples from four genera (total repeat number: 112–150) (Fig. 2a), but the total repeat length in Ruchi was lower than in *Geum urbanum* (Gurb) (Fig. 2b).

For all of the 38 Rosaceae samples, genome size showed a significantly high correlation with repeat number (phylogenetic generalized least squares: PGLS,  $R^2_{\text{adj}} = 0.35$ , *P*-value = 5.27e−5) (Fig. 2c). In addition, mitogenome size showed significantly high (*P*-value < 0.01) correlations with total repeat number and length (repeat length ≤ 500 bp) (Fig. 2e, f; Additional file 4: Fig. S1 a, b, e, f). However, negligible correlations ( $R^2_{\text{adj}} = -0.02$ ,

*P*-value = 0.51) appeared between total repeat length and genome size (Fig. 2d), and repeats longer than 1000 bp also showed low correlation with genome size ( $R^2_{\text{adj}} = -0.02$  and  $-0.03$ ) (Additional file 4: Fig. S1d, h). In 14 Fabaceae mitogenomes, genome size also showed high correlation with repeat sequences (total repeat number:  $R^2_{\text{adj}} = 0.73$ , *P*-value < 1.00e−3; total repeat length:  $R^2_{\text{adj}} = 0.68$ , *P*-value < 1.00e−3) (Additional file 4: Fig. S2a, g). All three repeat categories (length < 100 bp, 100 bp ≤ repeat length ≤ 500 bp, and length > 500 bp) showed significant correlation with variations in genome size ( $R^2_{\text{adj}}$  ranged from 0.67 to 0.90, *P*-value < 1.00e−3) (Additional file 4: Fig. S2b, c, f, h, i, l).

However, in regard to the total number of repeats, repeats shorter than 100 bp or 500 bp showed negligible correlations ( $R^2_{\text{adj}} = 0.10$ , 0.08, and 0.11) with genome size in 88 (one sample per species) seed plant mitogenomes (Additional file 4: Fig. S3a–c), and repeats > 500 bp showed high correlations with genome size (Additional file 4: Fig. S3d, e). Total repeat length (except for length > 1000 bp) showed significant correlation ( $R^2_{\text{adj}}$  ranged from 0.45 to 0.51, *P*-value < 1.00e−3) with genome size

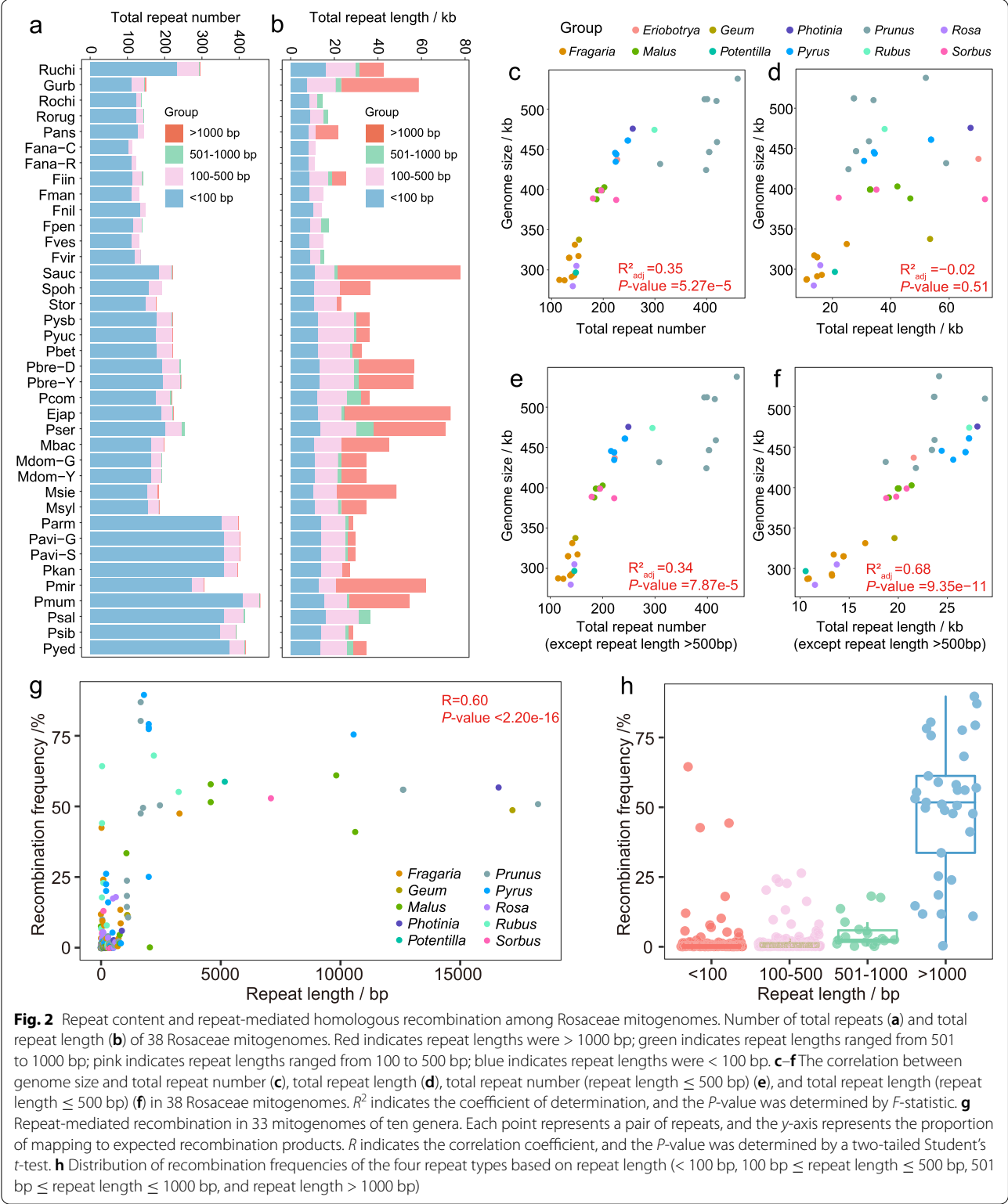


(Additional file 4: Fig. S3g-j, l). In addition, several mitogenomes may have an overrepresentation of repeat content (Additional file 5). For example, *Hyoscyamus niger* (total repeat length: 133.17 kb) had a similar genome size to *Prunus salicina* (Psal) (501.40 vs. 508.00 kb), but the total repeat length was about 3.97-fold of Psal (33.56 kb). Additionally, 50 mitogenomes, with genome sizes ranging from 271.60 to 525.67 kb, were selected, and 38.49- and 80.47-fold changes in total repeat length and numbers of repeat sequences were identified (Additional file 4: Fig. S4; Additional file 5).

**Recombination of repetitive sequences**

Based on long sequencing reads of 33 samples in ten genera, repeat length showed a relatively high correlation with recombination frequency (Pearson correlation coefficient:  $R = 0.60$ ,  $P$ -value  $< 2.2e-16$ ) (Fig. 2g). Higher recombination frequencies were observed for long ( $> 1000$  bp) repeats than for medium (100–500 bp and 501–1000 bp) or short ( $< 100$  bp) repeats (Fig. 2h), and percentages of long repeats associated

with homologous recombination were higher than that of short and medium repeats (Table 2). A total of 341 recombination events were identified, and 1–35 recombination events appeared in each of the 33 mitogenomes (Additional file 6). Among short repeats ( $< 100$  bp), only 2.45% (164/6707), underwent homologous recombination, but this percentage increased to 86.84% (33/38) for the long repeats ( $> 1000$  bp) (Table 2). Among the 341 repeats exhibiting recombination activity, 81.82% (27/33) of the long repeats recombined with a frequency greater than 20%, and 83.54% (137/164) of the short repeats had recombination frequencies lower than 1%. Twenty repeats had over 50% recombination frequency (Additional file 6). In *Pyrus*, a repeat of 2040 bp length and 25.31% recombination frequency in Pbre-Y exhibited 77.61–79.37% recombination frequency in "Hongxiang-suli" (*Pyrus sinkiangensis* × *bretschneideri*: Pysb), *Pyrus betulifolia* (Pbet), and "No.1 Zhong'ai" (*Pyrus ussuriensis* × *communis*: Pyuc). In *Pyrus communis* (Pcom), this repeat was shortened to 1841 bp, and the recombination frequency reached 89.69% (Additional file 6).



**Rearrangement rates of the 38 mitogenomes**  
Repeat-mediated recombination may further contribute to the rearrangement of mitogenomes [8]. In this study,

eleven mitogenomes (*Prunus mira*: Pmir, *Fragaria vesca*: Fves, *Eriobotrya japonica*: Ejap, Pbet, Ruchi, *Rosa rugosa*: Rorug, *Potentilla anserina*: Pans, *Photinia serratifolia*:



**Table 2** Recombination statistics on four types of repeats among 33 mitogenomes

| Sample | Number (%) of repeats <sup>b</sup> with recombination activity |                                 |                                  |                         |
|--------|--|---------------------------------|----------------------------------|-------------------------|
|        | Repeat length < 100 bp   | 100 bp ≤ repeat length ≤ 500 bp | 500 bp < repeat length ≤ 1000 bp | Repeat length > 1000 bp |
| Fana-C | 2 (1.98)   | 1 (9.09)                        | NA                               | NA                      |
| Fana-R | 11 (9.91)  | 6 (54.55)                       | NA                               | NA                      |
| Fiin   | 3 (2.68)   | 1 (3.70)                        | 0 (0.00)                         | 1 (100.00)              |
| Fman   | 3 (2.70)   | 1 (5.00)                        | NA                               | NA                      |
| Fnil   | 8 (5.97)   | 1 (6.67)                        | NA                               | NA                      |
| Fpen   | 2 (1.75)   | 3 (12.50)                       | 3 (100.00)                       | NA                      |
| Fves   | 0 (0.00)   | 1 (5.00)                        | NA                               | NA                      |
| Fvir   | 2 (1.68)   | 1 (6.25)                        | 1 (100.00)                       | NA                      |
| Mbac   | 6 (3.68)   | 11 (32.35)                      | NA                               | 1 (50.00)               |
| Mdom-G | 4 (2.47)   | 3 (10.34)                       | 1 (100.00)                       | 2 (100.00)              |
| Msie   | 1 (0.66)   | 1 (3.45)                        | NA                               | 2 (66.67)               |
| Msyl   | 2 (1.30)   | 5 (16.67)                       | 1 (100.00)                       | 1 (50.00)               |
| Pans   | 1 (0.78)   | 2 (13.33)                       | NA                               | 1 (100.00)              |
| Parm   | 7 (1.99)   | 7 (15.91)                       | 0 (0.00)                         | 1 (100.00)              |
| Pavi-G | 4 (1.11)   | 0 (0.00)                        | 0 (0.00)                         | 1 (100.00)              |
| Pavi-S | 4 (1.11)   | 1 (2.44)                        | 0 (0.00)                         | 1 (100.00)              |
| Pbet   | 5 (2.81)   | 5 (12.20)                       | 0 (0.00)                         | 1 (100.00)              |
| Pbre-Y | 5 (2.56)   | 9 (20.00)                       | 1 (33.33)                        | 2 (100.00)              |
| Pcom   | 2 (1.14)   | 0 (0.00)                        | 1 (16.67)                        | 1 (100.00)              |
| Pkan   | 7 (1.96)   | 0 (0.00)                        | NA                               | 1 (100.00)              |
| Pmir   | 11 (4.04)  | 6 (18.18)                       | NA                               | 2 (100.00)              |
| Pmum   | 6 (1.47)   | 3 (6.82)                        | 0 (0.00)                         | 3 (100.00)              |
| Psal   | 0 (0.00)   | 0 (0.00)                        | 3 (60.00)                        | NA                      |
| Psib   | 4 (1.15)   | 0 (0.00)                        | 0 (0.00)                         | 1 (100.00)              |
| Pyed   | 13 (3.49)  | 19 (47.50)                      | 1 (50.00)                        | 2 (100.00)              |
| Pysb   | 3 (1.69)   | 6 (14.29)                       | 0 (0.00)                         | 1 (50.00)               |
| Pyuc   | 3 (1.70)   | 2 (4.65)                        | 1 (100.00)                       | 1 (50.00)               |
| Rochi  | 21 (17.07)   | 6 (46.15)                       | 2 (100.00)                       | NA                      |
| Rorug  | 3 (2.44)   | 0 (0.00)                        | 2 (100.00)                       | NA                      |
| Ruchi  | 6 (2.58)   | 6 (10.17)                       | 0 (0.00)                         | 2 (100.00)              |
| Spoh   | 5 (3.18)   | 10 (28.57)                      | NA                               | 1 (100.00)              |
| Gurb   | 7 (6.36)   | 4 (11.43)                       | 0 (0.00)                         | 3 (100.00)              |
| Pser   | 3 (1.50)   | 4 (8.70)                        | 2 (28.57)                        | 1 (100.00)              |
| Total  | 164 (2.45)   | 125 (11.64)                     | 19 (40.43)                       | 33 (86.84)              |

<sup>a</sup>“NA” means repeat type was in the sample

<sup>a</sup> Percent = number of repeats with recombination/number of total repeats × 100%

<sup>b</sup> Repeats were divided into four types including short repeats (length < 100 bp), two types of medium repeats (100 bp ≤ length ≤ 500 bp and 500 bp < length ≤ 1000 bp), and long repeats (length > 1000 bp)

Pser, *Geum urbanum*: Gurb, *Sorbus pohuashanensis*: Spoh and *Malus sieversii*: Msie) were chosen to represent eleven genera. About 13.97–22.87% of the mitochondrial sequences were shared among all eleven genera, and more than 29 rearrangements were identified between

the Amygdaloideae and Rosoideae subfamilies (Fig. 3a). Twenty-one to 33 rearrangement events were identified in five Rosoideae genera, and 5–20 rearrangement events were identified in six Amygdaloideae genera. The rearrangements were then evaluated within each genus to avoid complications resulting from the high sequence divergence between genera [12]. In *Malus*, 91.76–95.31% of sequences were shared (Fig. 3c), and seven to nine rearrangement events were identified between *Malus baccata* (Mbac) and the other four apples. One or two rearrangements were detected between the remaining four accessions (Msie, Msyl, Mdom-Y, and Mdom-G). About 88.97–94.40% of sequences appeared within the six pears (Fig. 3d), and four to seven rearrangements were identified between Pcom and the other five Asian pears. Among the four Asian cultivated pears studied, six rearrangement events were identified between the Pysb and Pyuc and the Pbre-D and Pbre-Y species. Unexpectedly, no rearrangement events were detected between Pysb and Pyuc, despite their maternal parents coming from two pear systems (*Pyrus sinkiangensis* and *Pyrus ussuriensis*) (Fig. 3d). In *Prunus*, only 50.97–65.11% of sequences were shared, and 0–26 rearrangements were identified (Fig. 3e). In *Fragaria*, 66.89–77.20% homologous sequences and 0–17 rearrangements were identified (Fig. 3f).

Furthermore, obvious variations in rearrangement rate were identified at both the inter-genus (Fig. 3a, b) and intra-genus (Fig. 3c–f, Additional file 4: Fig. S5) levels. An over 10-fold variation in rearrangement rate (0.16–2.80) occurred among eleven genera, seven of which were lower than one (from 0.16 to 0.84) (Fig. 3b). The estimated common ancestor of Ejap, Msie, Pser, Pbet, and Spoh had a rearrangement rate as low as 0.18 after divergence with *Prunus*, which increased to 0.24 in Ejap, 1.14 in Pser, 2.08 in Msie, 2.45 in Pbet, and 2.80 in Spoh. Within *Malus*, the highest rearrangement rate (7.69 rearrangement events per million years ago (Mya)) was identified in the divergence between *Malus baccata* (Mbac) and the other three species (Additional file 4: Fig. S5a), and the pair-wise rearrangement rates (4.32 to 5.56) between Mbac and the other four samples (Fig. 3c) were higher than the others (from 0 to 2.00). In *Pyrus*, 2.88 rearrangement events/Mya were identified in Pcom (Additional file 4: Fig. S5b), and six rearrangement events which occurred at 0.05 Mya resulted in an extremely high rearrangement rate (120 rearrangement events/Mya) which experienced an increase in the pair-wise rearrangement rate between Pysb and Pbre-D (Fig. 3d). Variations in the rearrangement rate were also identified in *Prunus* and *Fragaria* (Fig. 3e, f; Additional file 4: Fig. S5c, d). Nine rearrangement events which occurred about 1.73 Mya in *Prunus avium* (“Glory”: Pavi-G and

“Staccato”: Pavi-S) resulted in a higher rearrangement rate than other species in *Prunus* (Additional file 4: Fig. S5c). Two *Fragaria* wild species (*Fragaria mandschurica*: Fman and Fves) experienced a threefold greater increase in rearrangement rate (10.52) than the other wild *Fragaria* species (0–3.35), and the rearrangement rate of *Fragaria ananassa* (“Royal Royce”: Fana-R and Fana-C) was 6.42 (Additional file 4: Fig. S5d).

### Mitogenomes reveal pear maternal phylogeny

The nuclear genome of pears is composed of biparental genetic background due to its self-incompatibility [21]. Compared with the nuclear genome phylogeny, the mitogenome phylogeny reveals the maternal relationship between different pear species. DNA re-sequencing data from 139 pear accessions were mapped to the “Dangshansuli” mitogenome (Additional file 7) to generate a SNP-based matrix, which included 85 Asian (52 cultivated and 33 wild) and 54 European (29 cultivated and 25 wild) pears. Our phylogenetic analysis of the associated mitogenomes revealed two groups, Asian and European pears (Fig. 4a). Among the Asian clade, three subclades were further subdivided, namely clades 1 and 3, which consisted of most of the Asian cultivated pear accessions, while clade 2 contained the wild Asian pear accessions. Cultivars of *Pyrus pyrifolia*, *P. ussuriensis*, and *P. bretschneideri* were mixed in clades 1 and 3. Four *P. sinkiangensis* cultivars clustered in the European group and one in the Asian group. Consistently, PCA (Fig. 4b) and structural analysis (Fig. 4c) also showed that Asian cultivated pears were divided into two groups.

### Identification of selective sweeps and divergent deletion types in mitogenomes

In 139 pear accessions, 1046 SNPs and 118 INDELs were identified (Fig. 5, Table 3), with only 95 SNPs (9.08%, Additional file 8) and two INDELs (1.69%, Additional file 9) being located in genes. To identify the specific regions under selection, selective sweeps were identified based on the diversity of the pear mitogenomes (Fig. 6a, b). For Asian pears, 5.88% (27.00 kb/458.90 kb) of the regions showed selective sweep signatures containing four protein-coding genes and one tRNA (Additional file 10). For European pears, there were selective sweep signatures for 2.18% (10.00 kb/458.90 kb) of sequences,

which contained three protein-coding genes. No overlapping selective sweeps were detected between Asian and European pears based on the mitogenomes.

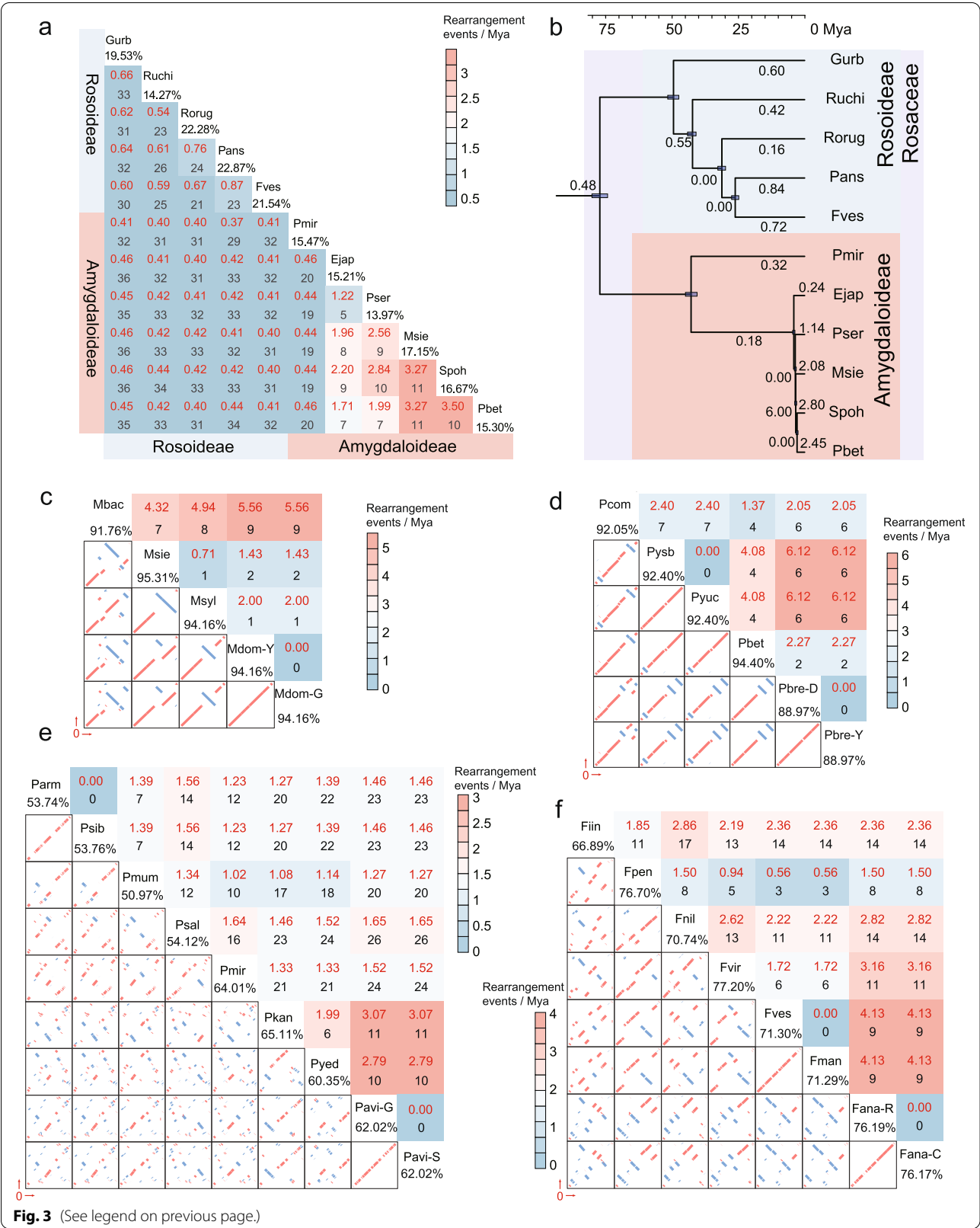
One continuous region from 185 to 190 kb showed a selective sweep signature in Asian pears, and *P. betulifolia* had deletions in this region (DEL-D, Pbre-D: 183.74–199.80 kb) (Fig. 6c). DEL-D was divided into three parts (Del1, Del2, and Del3); Del1 and Del3 were mitochondrial-specific sequences, and Del2 was similar to the chloroplast genome sequence (100% BLASTN identity). Therefore, we only analyzed the frequency of Del1 and Del3 in the four pears groups. Sixty-six percent (22/33) of Asian wild pears contained Del1, and the frequency was significantly (chi-square test,  $P$ -value =  $9.39 \times 10^{-15}$ ) higher than Asian cultivars (1.92%) (Fig. 6d). However, this divergence did not appear in European pears, and 92% of European wild pears and 100% of European cultivated pears did not contain Del1. This phenomenon also appeared in Del3, for which a significantly different frequency (chi-square test,  $P$ -value =  $3.11 \times 10^{-16}$ ) was observed between Asian wild and cultivated pears (Fig. 6e). As pears spread to the Middle East and Europe, most European wild and cultivated pears did not contain the Del1 (Fig. 6f).

A deletion (DEL-M) (*Malus domestica* cv. “Gala”: 180,287–186,952 bp), in a part of Del1, was also identified in *M. sieversii* (Fig. 6g), and DEL-M showed significantly (chi-square test,  $P$ -value < 0.01) different frequencies between wild and cultivated apples (Fig. 6h, Fig. S6a), based on the re-sequencing data of 116 apple accessions. Eighty percent of apples in the European wild (EW) group and 71.43% in the *M. domestica* (Dom) group did not contain DEL-M. Based on the apple distribution (Additional file 4: Fig. S6b), the *M. sieversii* group was divided into the Sie\_X (cultivated in the east of TianShan) and Sie\_K (cultivated in the west of TianShan) groups. One hundred percent of Sie\_X and 97.22% of Asian wild (AW) apples contained DEL-M.

Among the Rosaceae mitogenomes, large sequence fragments of DEL-D firstly appeared in Amygdaloideae and then expanded into *Malus* and *Pyrus* (Fig. 6i; Additional file 11). Compared with *Rosoideae*, large fragments (> 1 kb) of Del1 were identified in Amygdaloideae mitogenomes. Lengths of 1589–4201 bp of *Prunus* mitogenomes were mapped to Del1. A total of 8519 bp

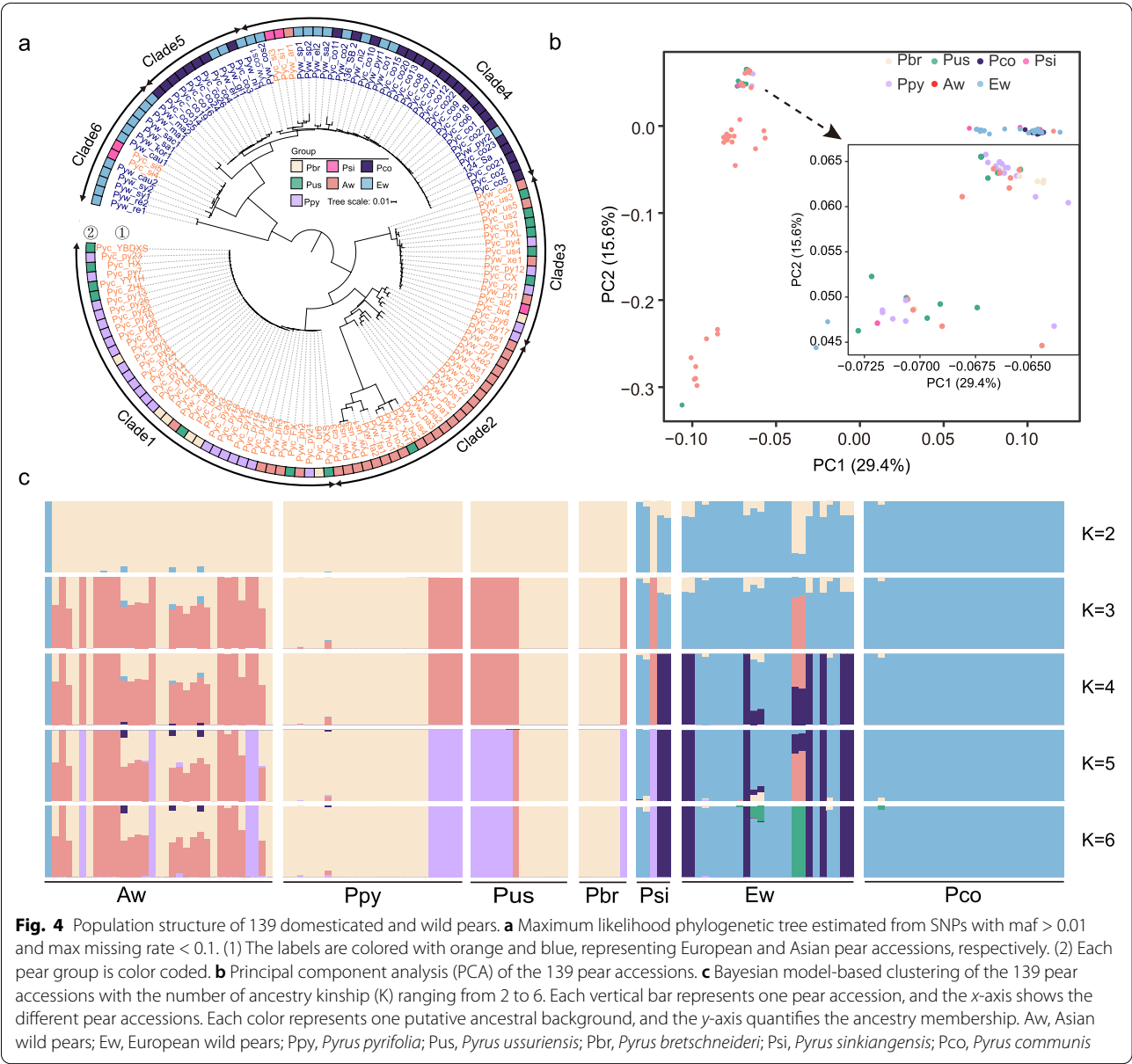
(See figure on next page.)

**Fig. 3** Rearrangement event and rate analysis. **a** Number and rate of pair-wise rearrangement events between eleven genera. Black numbers represent the rearrangement events (pair-wise rearrangement events), and red numbers represent the pair-wise rearrangement rates. **b** Rearrangement rates in eleven genera. The numbers of rearrangement events per million years are displayed on branches of the phylogeny. **c–f** The pair-wise analysis of rearrangement events and rate within *Malus* (**c**), *Pyrus* (**d**), *Prunus* (**e**), and *Fragaria* (**f**). Upper-right heatmaps represent the pair-wise rearrangement rates. Red numbers represent the rearrangement rates, and black numbers represent the rearrangement events. Bottom-left figures display the synteny analysis between two samples. Red represents direct, and blue represents inverted. Numbers under the sample ID represent the percent of shared sequences



**Fig. 3** (See legend on previous page.)



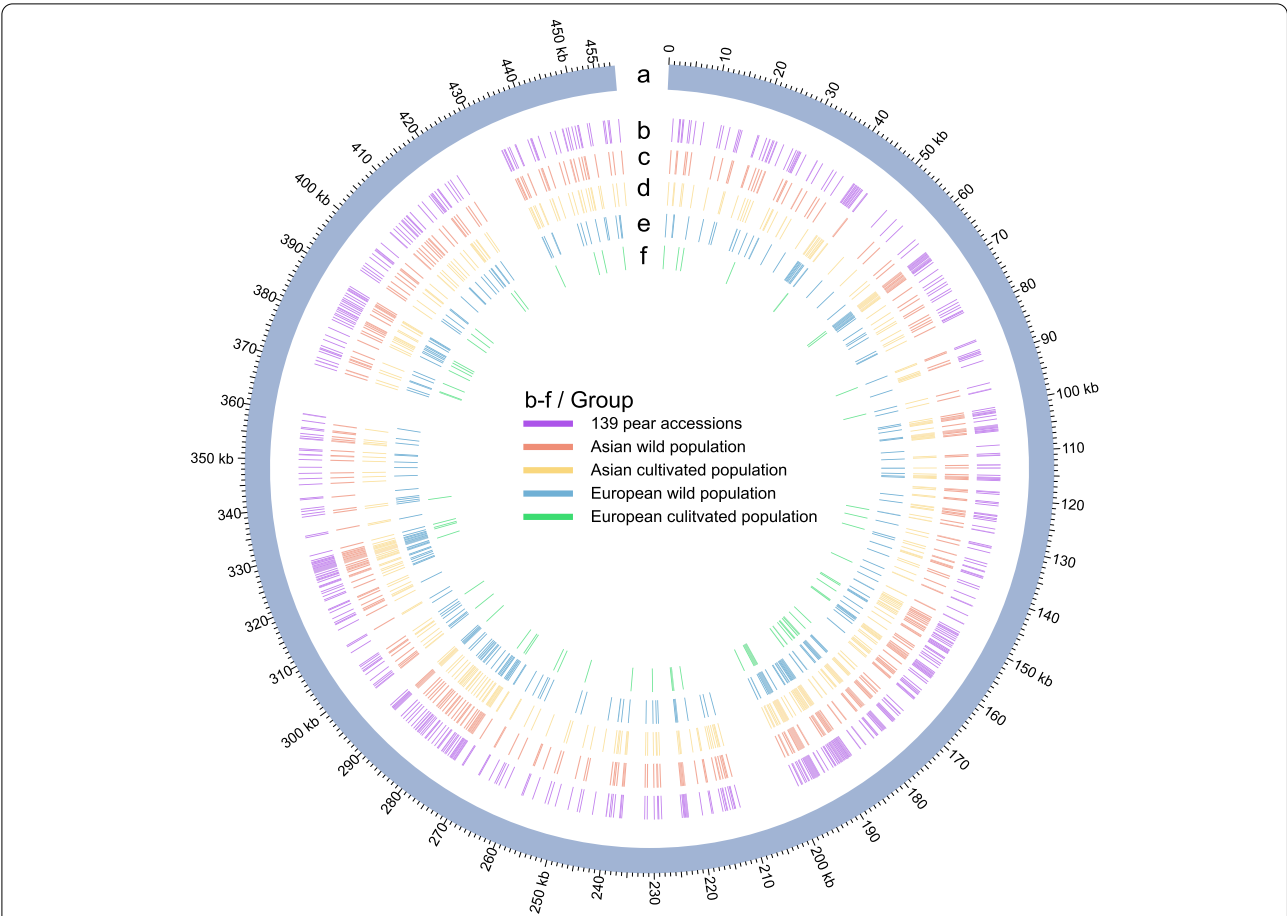


of sequence in the Pser mitogenome could be mapped to Del1. In *Malus*, a total of 6951 bp of sequence in the Mdom-G, Mdom-Y, and Msyl could be mapped to Del1. Spoh also contained 6133 bp of sequence mapping to Del1. More than 4000 bp of Del3 sequence was identified from the Pbre-D, Pbre-Y, Pcom, Mbac, Sauc, and Stor mitogenomes. Only 681 bp of Del3 was identified from Pysb and Pyuc, and 599–602 bp of mitogenome sequences of *Prunus* was mapped to Del3. Furthermore, the nuclear sequence mapping results showed that less than 10% of DEL-D sequences were shared with the nuclear sequences in *Fragaria*, *Rubus*, and *Rosa*, but this percentage increased in *Prunus* (5.69–44.06%), *Malus* (22.99–55.60%), and *Pyrus* (21.04–33.15%) (Additional file 11).

**Discussion**

**Gene loss and genome variation in 38 mitogenomes of Rosaceae**

Mitogenomes have variable gene content [27] and genome structure [11]. Thirty-eight mitogenomes from members of the Rosaceae were assembled and annotated to characterize the variations. Consistent with Fabaceae [9] and Poaceae [28], gene loss appeared before Rosaceae speciation, and the loss of *rpl2*, *rps10*, *rps11*, *rps19*, and *rps2* may have occurred in an ancestor



**Fig. 5** The distribution of SNPs and INDELs across mitogenomes in different groups. **a** The whole mitogenome of “Dangshansuli.” Total variants detected among **b** 139 pear accessions, **c** Asian wild pear population, **d** Asian cultivated pear population, **e** European wild pear population, and **f** European cultivated pear population

of Rosaceae. In addition, *rps2* and *rps11* were lost in all eudicots [27], indicating the more ancient losses of genes *rpl2*, *rps10*, and *rps19*. Within Rosaceae, *rpl16*, *sdh3*, and *rpl5* were absent in five genera of Rosoideae, which represent shrub and herb species, and *rps12* was lost in three herb genera (*Geum*, *Fragaria*, and *Potentilla*) (Fig. 1). These gene losses might affect the translocation and splicing of mitochondrial genes [29] and further influence plant development, reproduction, and other morphological and physiological traits, such as

stunting in maize [30], distorted leaves in *Arabidopsis* [31], stress responses in *Oryza sativa* [32], and the parasitic lifestyle of *V. scurruloideum* [4].

In this study, the genome sizes of the 38 mitogenomes were highly correlated with short (< 100 bp) and medium (100 bp ≤ repeat length ≤ 500 bp) repeat lengths (Fig. 2e, f; Additional file 4: Fig. S2a, b, e, f), indicating that repeat sequences may be related to the divergence of mitogenome sizes in Rosaceae. The DNA repair hypothesis suggests that repeat sequences are formed by non-homologous end joining and break-induced replication (BIR) and further drive genome expansion at evolutionary time scales [33]. However, this phenomenon was not consistently observed in 88 seed plant mitogenomes, and several mitogenomes had a burst in repeat sequences, which indicated that other mechanisms may drive genome size variation such as gains or losses of entire chromosomes [34], abundant rearrangements, or loss of non-coding sequences [35].

**Table 3** Summary of the SNPs and INDELs in 139 pear accessions

| Population                | Total variants | INDEL | SNP  |
|---------------------------|----------------|-------|------|
| All accessions            | 1164           | 118   | 1046 |
| Asian wild group          | 815            | 82    | 733  |
| Asian cultivated group    | 806            | 90    | 716  |
| European wild group       | 613            | 60    | 553  |
| European cultivated group | 194            | 20    | 174  |

Although repeats with lengths longer than 1000 bp showed a low correlation with genome size (Additional file 4: Fig. S2) in Rosaceae mitogenomes, they contained higher recombination frequencies than repeats with lengths shorter than 500 bp. Large mitochondrial repeats (> 1000 bp) undergo high-frequency reciprocal recombination to subdivide the genome in other plant species [36]. In addition, twenty repeats had recombination frequencies greater than 50%, indicating that a “master circle” was not the main conformation. High sub-genomic conformations have been observed in vivo, as exemplified in *Silene* [5], *Cucumis* [37], and *Selaginellaceae* [38], and no master conformation appeared in *Saccharum officinarum* [39]. Moreover, more than one repeat containing such recombination frequencies indicated that many conformations may appear at the same time (Additional file 6).

In this study, an over tenfold variation in rearrangement rate occurred between eleven genera of Rosaceae (Fig. 3b), and this variation also occurred within genera (Additional file 4: Fig. S5). In addition, at least 600-fold variation in rearrangement rate was identified in seed plants [8], and some studies found that environmental stress [40–43] and nuclear gene variation (like *MSH1* and *RECA*) [44] might contribute to mitogenome rearrangement. In *Malus*, Mbac originates from Siberia, Msie is distributed in Central Asia, and Msyl is distributed in Western Europe [45]. *Pyrus* spreads from southwest China to Europe [21]. *Fragaria* is widespread in Asia, Europe, and North America [46, 47]. *Prunus* spreads from Asia to Europe [48, 49]. These different geographical distributions and environmental changes might be one reason for the variation in rearrangement rate among Rosaceae species.

#### Domestication may have been involved in the evolution and expansion of mitogenomes

Human selection has modified many crop traits, and cultivated crops are divergent from their wild progenitors

[50]. DEL-D in selective sweep regions supports that selection drives mitogenome variation in pears. Formed by multi-step processes, DEL-D finally became fixed in Asian cultivated pears during domestication (Fig. 6j). Functional mitochondrial gene formation includes multiple steps and can cause phenotypic changes, biological diversity, and further benefits for natural adaptation [51]. DEL-D was formed by multi-recombination events, sequence imports, and new ORF formations (Fig. 6i), which may become a new resource conferring phenotypic or metabolic changes and contributing to adaptations to environmental stress. Afterwards, selection may quickly drive the allele frequency changes to improve the adaptive ability of the population [52]. DEL-D frequency is very different between Asian cultivated and wild pears and between Asian and European wild pears (Fig. 6d, e). DEL-M also had a significantly different frequency between Asian and European apples and between *M. sieversii* and cultivated apples (Fig. 6h). The selection sweeps and deletion frequency changes might aid in adaptation to environmental changes or be fit for human needs [53, 54].

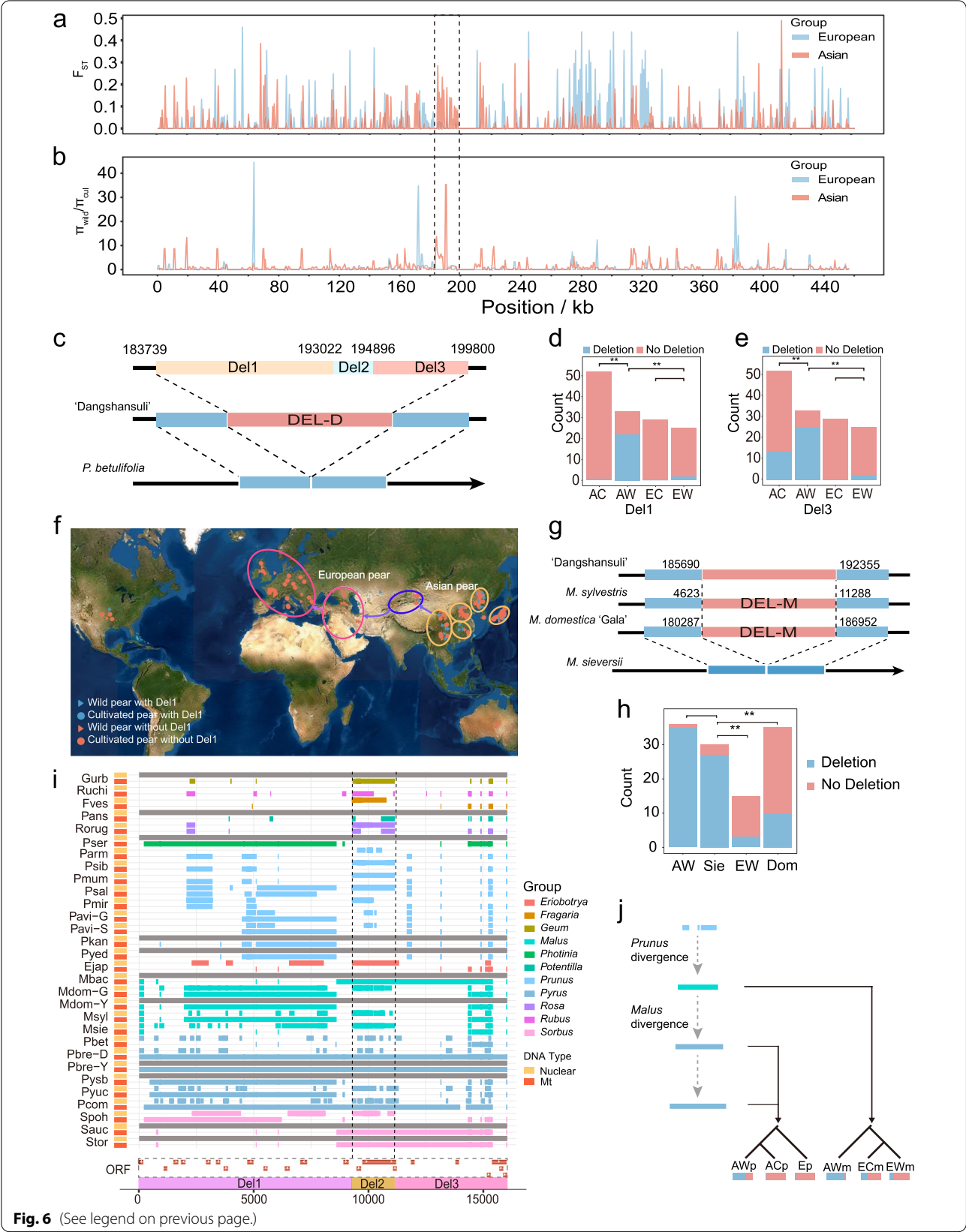
#### Mitochondrial variants shed new insights on the maternal relationships between *Pyrus* species

The topology based on the newly assembled mitogenomes provides insights into the maternal phylogenetic relationships of *Pyrus* species, and it presents an alternative framework to that based on nuclear sequences [21]. Compared with nuclear-based phylogenetic analysis, Asian cultivated pears were divided into clade 1 and clade 3, and three main cultivated pear species (*P. pyrifolia*, *P. bretschneideri*, and *P. ussuriensis*) were mixed in clades 1 and 3, suggesting the mitogenome divergence process produced two main maternal lines in Asian cultivated pears. What is more, the divergence occurred in the maternal parents of *M. domestica*, and the selection of fruit size, flavor, or unilateral compatibility in crosses may be responsible for this divergence [55]. Five *P.*

(See figure on next page.)

**Fig. 6** Deletion analysis of the mitogenomes of pear and apple. **a** Distribution of  $F_{ST}$  values across the whole Asian and European pear populations. **b** Distribution of  $\pi_{wild}/\pi_{cul}$  ratios; cul means cultivated pear population; dotted regions represent the deletions (Pbre-D: from 183,739 to 199,800 bp) in *P. betulifolia*. **c** A deletion (DEL-D) was identified in *P. betulifolia*. The deletion sequence is further divided into three parts (Del1, Del2, and Del3). **d**, **e** Appearance times of Del1 (**d**) and Del3 (**e**) in the four pear groups. Red represents the samples not containing the deletion, and blue represents the samples containing the deletion. AC, Asian cultivated group; AW, Asian wild group; EC, European cultivated group; EW, European wild group. **f** Geographical distribution of 139 pear accessions. The main distribution areas are marked by circles. Blue indicates pears with the deletion, and red indicates pears without deletion. Triangles represent wild pears, and circles represent cultivated pears. **g** A deletion (DEL-M) is identified in *M. sieversii* which did not appear in *M. sylvestris* and *M. domestica* cv. “Gala.” This deletion is homologous with a sequence (185,690–192,355 bp) of “Dangshansuli.” **h** Appearance times of DEL-M in the four apple groups. Red represents the samples not containing the deletion, and blue represents the samples containing the deletion. Sie, *M. sieversii*; AW, Asian wild group; Dom, cultivated group; EW, European wild group. **i** Homologous sequences of Del1, Del2, and Del3 in 30 mitogenomes and nuclear genomes. The gray bar indicates no information was available. ORFs were identified using ORFfinder with a minimum length of 150 bp. **j** Putative dynamics of evolution and expansion of the deletion sequence during the divergence of Rosaceae species. AWp, Asian wild pear; ACp, Asian cultivated pears; Ep, European pears; AWm, Asian wild apples; ECm, European cultivated apples; EWm, European wild apples. *P*-values were determined by a two-tailed Student’s *t*-test (\*\**P*-value < 0.01)





*sinkiangensis* cultivars were divided into Asian and European groups indicating that the maternal parents of *P. sinkiangensis* came from both Asian and European pears. Most Asian wild pear accessions were divergent from the cultivated species, and *Pyrus calleryana* (Pyw\_ca), *Pyrus xerophila* (Pyw\_xe), *Pyrus phaeocarpa* (Pyw\_ph), and *Pyrus serrulata* (Pyw\_se) showed a close relationship with cultivated pears indicating that introgression of maternal parents might happen because of cross-hybridization and adjacent distribution.

## Conclusions

In this study, in-depth comparisons showed the evolutionary patterns of 38 mitogenomes in Rosaceae. Apparent gene losses and shrinkage of the mitogenome size occurred in the Amygdaloideae and Rosoideae subfamilies. Repeat content may lead to genome size variations and primarily drive the dynamics of genome structure by homologous recombination and genomic rearrangements. We estimated the absolute rearrangement rate of Rosaceae mitogenomes, and variations in rearrangement rates were also identified in *Prunus*, *Malus*, *Pyrus*, and *Fragaria* genera. Two divergent maternal lineages were identified in Asian cultivated pears, and free hybridization might explain the mixed maternal lines of cultivated *P. pyrifolia*, *P. ussuriensis*, and *P. bretschneideri*. *Pyrus*-specific sequence variation (DEL-D) was determined, based on the complete mitogenome and population data, to have originated from Amygdaloideae, and this sequence quickly expanded from Asian wild species to Asian cultivated species and European populations. This comparative genomic study provides new insights into the evolutionary and selection patterns of Rosaceae mitogenomes.

## Method and plant materials

### Assembly and annotation of the mitochondrial and chloroplast genomes

Thirty-four of the 38 mitogenomes were assembled using NGS and long-read sequencing data. The Illumina HiSeq 2000 data generated from the whole genome of “Dangshansuli” were used for mitogenome assembly, and the series of the “Dangshansuli” (*P. bretschneideri* Rehd.) published genome BAC libraries were selected (library insertion sizes of 180 bp, 488 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb, and 20 kb) [56]. Fastq files were first filtered using Trimmomatic [57] with default parameters, using 800 bp library insertion size data for mitogenome assembly. Reads were assembled using SOAPdenovo2 [23], and the scaffolds were polished using Pilon v1.23 [58]. Furthermore, to ensure

that scaffolds were indeed mitogenomes, we chose the first ten longest assembled scaffolds to do the alignment in the NCBI assembly database, using a cutoff for the BLASTN *e*-value of  $1e-5$  for the scaffolds [59]. Lastly, scaffolds were selected from the mitogenomes. Insertion sizes of 2 kb, 5 kb, 10 kb, and 20 kb reads helped to concatenate the scaffolds into one, and then the mitogenome was polished by Pilon to fill the gaps. A similar method of organelle genome assembly based on whole genomes has been performed on plants [60].

The raw reads of 33 Rosaceae species were downloaded from NCBI (Additional file 12). We chose the “Dangshansuli,” *Rosa chinensis* (CM009589.1), and *Prunus avium* (MK816392.2) mitogenomes as the reference genomes. The mitochondrial long reads were identified by BLASR [61] to obtain candidate reads from the references and then assembled into contigs using the program Canu v1.8 [24]. Overlaps of mitochondrial candidate contigs were identified using the BLASTN program [59] and finally formed circular molecules. The circular molecules were polished by Pilon v1.23 [58]. Long and short reads were remapped to the polished genome sequences to check the completeness of all newly assembled mitogenomes (Additional file 2). The high-quality complete mitogenomes were annotated by Geseq [62] and Mitofy [63]. The final annotations were checked manually to correct the position of the start and stop codons. A strategy similar to the mitogenome assembly strategy was used for chloroplast genome assembly, and the genome sequences were annotated by Geseq [62]. The annotation files were further checked manually.

### Identification of plastid-derived and repeat sequences

To identify plastid-derived sequences, the 38 mitogenomes were searched against the corresponding plastid genomes in the BLASTN program using an *e*-value cutoff of  $1e-6$  and a word size of 7, simultaneously. Repeats identified in the 38 mitogenomes were carried out using similar methods [64], and the BLASTN program was used to search each mitogenome against itself, using an *e*-value limit lower than  $1e-6$  and a word size of 7. The Caper/R package was used to perform the phylogenetic generalized least squares (PGLS) analysis to identify correlations between genome sizes and repeat sequences in the 38 Rosaceae mitogenomes. For the analysis of Fabaceae and seed plants (including 14 and 88 genera), only one accession per genus was chosen.

### Identification of repeat-mediated homologous recombination events

To detect active, repeat-mediated, homologous recombination events within the long sequencing reads, we first

built up mitochondrial read databases of 33 mitogenomes (the five other samples were excluded due to lack of long sequencing data). We used the 33 mitogenome assemblies as a reference to obtain candidate mitochondrial sequences from whole DNA long sequencing reads by BLASTN, using an *e*-value cutoff of  $1e-100$ . Candidate mitochondrial reads were further searched against chloroplast genome sequences (Additional file 13) to remove putative plastid reads with overall alignment coverage of  $> 85\%$  of the read length, and the clean reads were self-corrected using Canu v1.8 [24]. Finally, we obtained 33 mitochondrial read databases (Additional file 14) and used similar methods [13] to identify repeat-mediated homologous recombination events. Briefly, each repeat pair with 200 bp of up- and down-stream sequence was extracted as reference sequences and used to build two recombinant sequences (repeat pairs with 100% BLASTN identity) or six recombinant sequences (repeat pairs were lower than 100% BLASTN identity) (Additional file 4: Fig. S7). Then, the mitochondrial reads were blasted against the reference and recombinant sequences, and reads having identities above 99% and hit coverages of 200 bp in two flanking regions were selected.

#### Species tree construction and divergence time estimation

A total of 38 Rosaceae chloroplast genomes (Additional file 13) were used for phylogenetic analysis and divergence time estimation. The coding sequences of 76 chloroplast protein-encoding genes of the 38 Rosaceae samples (Additional file 15) and an outgroup, *Vitis vinifera* (NC\_007957), were aligned. Phylogenetic trees were constructed using IQ-TREE [65]. Divergence time estimation was conducted by MCMCtree of PAML 4.9 [66] with the following parameters: burn-in of 5,000,000 iterations, sample frequency of 5000, and the MCMC process was performed 20,000 times. Three calibration points were used: one fossil of *Prunus* found in Shandong ( $> 44.3$  Mya) [67], one fossil of *Rubus* (47.8 to 41.3 Mya) [68], and the estimated divergence time (130 to 123 Mya) between *V. vinifera* and Rosaceae [69].

#### Rearrangement event identification in Rosaceae mitogenomes

To infer the rearrangement rate between eleven genera, multiple alignments of all pairwise combinations of the mitogenomes of the eleven genera (Pmir, Pser, Gurb, Fves, Ejap, Pbet, Ruchi, Rorug, Pans, Spoh, and Msie) were conducted using Mauve v2.0 [70] to analyze locally collinear blocks (LCBs) in each mitogenome with default parameters, and pairwise rearrangement distances in terms of a minimum number of rearrangements were inferred using GRIMM with the circle chromosome option [71]. To explore the rearrangement rate of

different branches of the tree, eleven samples were used in MLGO to infer the ancestral genome arrangement [72]. The rearrangement events between each node and neighboring nodes were calculated by GRIMM [71]. The rearrangement rate was calculated using the rearrangement events by dividing the absolute time of each branch. In addition, the number of pair-wise rearrangements was divided by double divergence time between the two samples to calculate the mean pair-wise rearrangement rate. *Pyru* (for *Pyrus*), *Fragaria viridis* (for *Fragaria*), *Msyl* (for *Malus*), and *Prunus armeniaca* (for *Prunus*) were chosen as the reference genomes for their respective genera to adjust the direction of other mitogenomes for rearrangement analysis, and the rearrangement rate within the genera *Pyrus*, *Malus*, *Prunus*, and *Fragaria* were calculated using the same calculation methods used for inter-genera analysis.

#### SNP and INDEL calling of 139 pear accessions

Together, with the published re-sequencing data of 113 pears [21], we also selected another 26 pear accessions to perform next-generation sequencing using the same method on the HiSeq 2000 platform (Additional file 7). We used the “Dangshansuli” mitogenome as a reference for SNP and INDEL calling. Raw data of 139 pear accessions were trimmed by Trimmomatic v0.39 [57]. Clean data was mapped to the reference genome using Burrows-Wheeler Alignment v0.7.16 (BWA) [73]. SAMtools [74] was used to convert the sequence alignment mapping file (SAM) into a binary SAM (BAM) file. Then, the removal of duplicated reads was performed using the Picard software (<http://broadinstitute.github.io/picard/>). Variant identification and filtering were performed using GATK v4.1.4 [75]. Finally, all SNPs and INDELs with minor allele frequencies (MAF) of  $> 0.01$  and max-missing rate of  $< 0.1$  were extracted for subsequent analysis. SNPeff v4.3t [76] was used for SNP and INDEL annotation.

#### Phylogenetic tree construction, PCA, and population structure analysis

All SNPs for each sample were connected one by one as a single locus to make fasta files using an in-house python script, and then IQ-TREE [65] was used to generate the phylogenetic tree with the maximum likelihood method, and the best model was detected using the “MF” function. We set the ultrafast bootstrap replication number as 1000. To evaluate the relationships, PCA and population structure analysis were performed using plink v1.90b [77] and admixture v1.3 [78].

#### Diversity analysis and selection sweep identification

$\pi$  and  $F_{ST}$  were calculated by VCFtools v0.1.16 [79] with a 1000-bp sliding window and 500-bp steps in pear.



To further identify the regions with signals of selection sweeps in cultivated pears, regions (1000-bp window) with signals for selective sweeps were identified with reference to previous criteria: the top  $F_{ST} > 0.1$ ,  $\pi_{wild}/\pi_{cul}$  ratio  $> 2$  based on common SNPs in the pear mitogenomes [80].

### Frequency of deletion analysis

To further evaluate the frequency of the deletion (DEL-D) in 139 pear accessions, BEDTools v 2.18 [81] was used to calculate the mapping coverage of DEL-D in the 139 pear accessions. First, DEL-D was divided into three parts (Del1, Del2, and Del3), and the read depths of each part (Idep) were calculated respectively. Furthermore, the whole-genome depth of each accession (Wdep) was calculated. To avoid the differences in sequencing depth in the 139 accessions, we used a ratio of Idep divided by Wdep to evaluate the presence and absence of the deletion. Fortunately, the ratio of Del1 was divided into two levels, namely low (0.24–0.72) and high (6.94–142.98), with a high ratio representing Del1 being present in the mitogenome of this accession and a low ratio representing absence. This phenomenon also appeared in Del3. Due to Del2 sharing homology with chloroplast sequences, we excluded Del2 from further analyses. The frequency of Del1 and Del3 in different pear populations were calculated, and the two-tailed Student's *t*-test was used to identify the significant differences. The same strategy was used to detect the frequency of the deletion (DEL-M: 6666 bp) in 116 apple accessions.

To detect the origin of the deletion sequence, we used a BLASTN search to detect the homologous sequence in 30 Rosaceae mitochondrial and nuclear genomes. The inferred putative origin of the intracellular transfer and nuclear-shared sequences were identified by performing BLASTN searches of mitogenomes against nuclear DNA, with an *e*-value cutoff lower than  $1e-100$  and hit length more than 100 bp, and the ggplot2 package (<https://cran.r-project.org/web/packages/ggplot2/index.html>) was used for visualization. An in-house python script was used to calculate the total length of homologous sequences from each mitochondrial and nuclear genome. ORFs with a minimum length of 150 bp were identified in DEL-D using ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>).

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01383-3>.

**Additional file 1.** Released Rosaceae mitogenomes (Last access date: 20-Jan-2022).

**Additional file 2.** Mapping depth of 34 newly assembled Rosaceae mitogenomes.

**Additional file 3.** Repeat statistics of 38 Rosaceae mitogenomes.

**Additional file 4: Figure S1.** The relationship between genome size and total repeat length and number in 38 Rosaceae mitogenomes. The repeats were divided into four types: length  $< 100$  bp (a, e);  $100 \text{ bp} \leq \text{repeat length} \leq 500$  bp (b, f);  $500 \text{ bp} < \text{repeat length} \leq 1,000$  bp (c, g); and repeat length  $> 1,000$  bp (d, h). The linear regression equation is displayed with adjusted R-square and *P*-values. **Figure S2.** The relationship between mitogenome size and total repeat length and count in 14 Fabaceae mitogenomes. The repeats were divided into six types: all repeats (a, g); repeat length  $< 100$  bp (b, h);  $100 \text{ bp} \leq \text{repeat length} \leq 500$  bp (c, i);  $500 < \text{repeat length} \leq 1,000$  bp (d, j); length  $> 1,000$  bp (e, k); and repeat length  $\leq 500$  bp (f, l). The linear regression equation is displayed with adjusted R-square and *P*-value. **Figure S3.** The relationship between mitogenome size and total repeat length and count in 88 seed plants. The repeats were divided into six types: all repeats (a, g); repeat length  $< 100$  bp (b, h);  $100 \text{ bp} \leq \text{repeat length} \leq 500$  bp (c, i);  $500 < \text{repeat length} \leq 1,000$  bp (d, j); repeat length  $> 1,000$  bp (e, k); and repeat length  $\leq 500$  bp (f, l). The linear regression equation is displayed with adjusted R-square and *P*-value.

**Figure S4.** The distribution of repeat count (a) and total repeat length (b) of 50 seed plant mitogenomes with genome sizes ranging from 271.60 to 525.67 kb. **Figure S5.** The rearrangement rate estimated using tree-based methods in *Malus* (a), *Pyrus* (b), *Prunus* (c) and *Fragaria* (d). Red numbers on the branches represent rearrangement events and rates (rearrangement events per million years), respectively. Yellow triangles represent the varieties within specie, and the rearrangement events and rates are calculated between species and neighboring nodes. The blue bar indicates the 95% highest posterior densities. **Figure S6.** The mapping depth and distribution analysis of 116 apple accessions. (a) The mapping depth of 116 apple accessions. The NGS reads are mapped to the *Malus domestica* cv. 'Gala' (Mdom-G) mitogenome. A ratio of Idep divided by Wdep was used to evaluate the mapping results, and the ratio was further normalized using the z-score method. Orange: high mt read mapping depth, blue: low mt read mapping depth. AW: Asian wild apples; EW: European wild apples; Sie: *Malus sieversii*; Dom: *Malus domestica*. (b) Distribution analysis of apple mitogenomes. Main distribution areas are marked by circles. Blue: apples containing the deletion (Del), red: apple not containing this deletion. Triangles represent wild apple and circles represents cultivated apple. Dom: *Malus domestica*; Syl: *Malus sylvestris*; Sie\_K: *Malus sieversii* in west of TianShan; Sie\_X: *Malus sieversii* in east of TianShan; Bac: *Malus baccata*; Asi: *Malus asiatica*; Hup: *Malus hupehensis*. **Figure S7.** Flow chart for repeat recombination analysis. (a) Recombinant sequence construction. 'b' and 'e' indicate repeat sequences; 'a' and 'd' indicate the upstream 200 bp sequences; 'c' and 'f' indicate the downstream 200 bp sequences. (b) Mitochondrial reads mapping to reference and recombinant sequences using BLASTN. (c) Recombination frequency calculation.

**Additional file 5.** Repeat statistics of 88 seed plant mitogenomes.

**Additional file 6.** Information on 341 repeats containing recombination activities.

**Additional file 7.** 139 wild and cultivated pear accessions and mapping profile.

**Additional file 8.** SNPs annotation of mitogenome.

**Additional file 9.** INDELs annotation of mitogenome.

**Additional file 10.** Mitochondrial genes in selective sweep regions.

**Additional file 11.** Homologous sequence of DEL-D (Pbre-D: 183,739–199,800 bp) in 30 mitogenomes and nuclear genomes.

**Additional file 12.** Project information for the raw sequences of 34 Rosaceae samples.

**Additional file 13.** Summary of 38 Rosaceae chloroplast genomes.

**Additional file 14.** Information for the mitochondrial read databases.

**Additional file 15.** Genes used in the chloroplast genome phylogeny analysis.

## Acknowledgements

We thank the high-performance computing platforms of the Bioinformatics Center of Nanjing Agricultural University and the State Key Laboratory of Crop Biology of Shandong Agricultural University for supporting this project. This project was also supported by the platform of the Center of Pear Engineering Technology Research of Nanjing Agricultural University. Honghe Sun and Leitong Li gave constructive suggestions for this project. The authors would like to thank everyone who contributed to this article. All authors read and approved the final manuscript.

## Authors' contributions

MYS and MYZ drafted the manuscript. MYS, MYZ, and XNC performed the bioinformatics analysis. YYL, BBL, JML, RZW, and KJZ reviewed the manuscript. JW conceived this study and prepared the manuscript. The authors read and approved the final manuscript.

## Funding

This work was supported by the National Key Research and Development Program (2018YFD1000200), the National Science Foundation of China (31820103012, 31901978), the Earmarked fund for China Agriculture Research System (CARS-28), the Earmarked Fund for Jiangsu Agricultural Industry Technology System (JATS [2021]453), and Natural Science Foundation of Jiangsu Province for Young Scholar (BK20221010).

## Availability of data and materials

Raw WGS data of pear and apple accessions were downloaded from the NCBI BioProject (PRJNA381668, PRJNA675194, PRJNA844501, and PRJNA322175). The NGS and Pacbio data used for mitogenome assembly were downloaded from NCBI, and the BioProject ID was supplied in Additional file 12. The 34 new assembly mitogenome sequences were all submitted to the NCBI database, and accession numbers are listed in Table 1.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>College of Horticulture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China. <sup>2</sup>State Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an 271018, China. <sup>3</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>4</sup>Department of Botany, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA.

Received: 16 February 2022 Accepted: 4 August 2022

Published online: 19 August 2022

## References

- Van Aken O, Van Breusegem F. Licensed to kill: mitochondria, chloroplasts, and cell death. *Trends Plant Sci.* 2015;20(11):754–66.
- Birky CW Jr. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci U S A.* 1995;92(25):11331–8.
- Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytol.* 2013;200(4):978–85.
- Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A.* 2015;112(27):E3515–24.
- Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 2012;10(1):e1001241.
- Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 2013;11:29.
- Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michellmore RW, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. *PLoS Genet.* 2019;15(8):e1008373.
- Cole LW, Guo W, Mower JP, Palmer JD. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Mol Biol Evol.* 2018;35(11):2773–85.
- Choi IS, Schwarz EN, Ruhlman TA, Khiyami MA, Sabir JSM, Hajarrah NH, et al. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. *BMC Plant Biol.* 2019;19(1):448.
- Mower JP, Sloan DB, Alverson AJ. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. *Plant genome diversity volume 1: plant genomes, their residents, and their evolutionary dynamics*. Vienna: Springer Vienna; 2012. p. 123–44.
- Davila JJ, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, et al. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. *BMC Biol.* 2011;9:64.
- Sullivan AR, Eldfjell Y, Schiffthaler B, Delhomme N, Asp T, Hebelstrup KH, et al. The mitogenome of Norway spruce and a reappraisal of mitochondrial recombination in plants. *Genome Biol Evol.* 2020;12(1):3586–98.
- Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, et al. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics.* 2018;19(1):614.
- Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the stability of the plant mitochondrial genome. *Int J Mol Sci.* 2020;21(1):328.
- Guo WH, Grewe F, Fan WS, Young GJ, Knoop V, Palmer JD, et al. Ginkgo and Welwitschia mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol Biol Evol.* 2016;33(6):1448–60.
- Juszczuk IM, Flexas J, Szal B, Dabrowska Z, Ribas-Carbo M, Rychter AM. Effect of mitochondrial genome rearrangement on respiratory activity, photosynthesis, photorespiration and energy status of MSC16 cucumber (*Cucumis sativus*) mutant. *Physiol Plant.* 2007;131(4):527–41.
- Bentolilla S, Stefanov S. A reevaluation of rice mitochondrial evolution based on the complete sequence of male-fertile and male-sterile mitochondrial genomes. *Plant Physiol.* 2012;158(2):996–1017.
- Shi S, Li J, Sun J, Yu J, Zhou S. Phylogeny and classification of *Prunus sensu lato* (Rosaceae). *J Integr Plant Biol.* 2013;55(11):1069–79.
- Rono PC, Dong X, Yang JX, Mutie FM, Oulo MA, Malombe I, et al. Initial complete chloroplast genomes of *Alchemilla* (Rosaceae): comparative analysis and phylogenetic relationships. *Front Genet.* 2020;11:560368.
- Sun X, Jiao C, Schwanninger H, Chao CT, Ma Y, Duan N, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet.* 2020;52(12):1423–32.
- Wu J, Wang Y, Xu J, Korban SS, Fei Z, Tao S, et al. Diversification and independent domestication of Asian and European pears. *Genome Biol.* 2018;19(1):77.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21(1):241.
- Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience.* 2012;1:18.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36.
- Yue X, Zheng X, Zong Y, Jiang S, Hu C, Yu P, et al. Combined analyses of chloroplast DNA haplotypes and microsatellite markers reveal new insights into the origin and dissemination route of cultivated pears native to East Asia. *Front Plant Sci.* 2018;9:591.

26. Barr CM, Neiman M, Taylor DR. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol.* 2005;168(1):39–50.
27. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci U S A.* 2002;99(15):9905–12.
28. Hall ND, Zhang H, Mower JP, McElroy JS, Goertzen LR. The mitochondrial genome of *Eleusine indica* and characterization of gene content within Poaceae. *Genome Biol Evol.* 2020;12(1):3684–97.
29. Kwasniak-Owczarek M, Kazmierczak U, Tomal A, Mackiewicz P, Janska H. Deficiency of mitochondrial S10 protein affects translation and splicing in *Arabidopsis* mitochondria. *Nucleic Acids Res.* 2019;47(22):11790–806.
30. Hunt MD, Newton KJ. The NCS3 mutation: genetic evidence for the expression of ribosomal protein genes in *Zea mays* mitochondria. *EMBO J.* 1991;10(5):1045–52.
31. Sakamoto W, Kondo H, Murata M, Motoyoshi F. Altered mitochondrial gene expression in a maternal distorted leaf mutant of *Arabidopsis* induced by chloroplast mutator. *Plant Cell.* 1996;8(8):1377–90.
32. Zhang X, Takano T, Liu S. Identification of a mitochondrial ATP synthase small subunit gene (RmtATP6) expressed in response to salts and osmotic stresses in rice (*Oryza sativa* L.). *J Exp Bot.* 2006;57(1):193–200.
33. Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol Evol.* 2013;5(6):1079–86.
34. Wu ZQ, Cuthbert JM, Taylor DR, Sloan DB. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc Natl Acad Sci U S A.* 2015;112(33):10185–91.
35. Wang S, Li D, Yao X, Song Q, Wang Z, Zhang Q, et al. Evolution and diversification of kiwifruit mitogenomes through extensive whole-genome rearrangement and mosaic loss of intergenic sequences in a highly variable region. *Genome Biol Evol.* 2019;11(4):1192–206.
36. Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, et al. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol Gen Genomics.* 2005;272(6):603–15.
37. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell.* 2011;23(7):2499–513.
38. Kang JS, Zhang HR, Wang YR, Liang SQ, Mao ZY, Zhang XC, et al. Distinctive evolutionary pattern of organelle genomes linked to the nuclear genome in Selaginellaceae. *Plant J.* 2020;104(6):1657–72.
39. Shearman JR, Sonthirod C, Naktang C, Pootakham W, Yoocha T, Sangsrakru D, et al. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. *Sci Rep.* 2016;6:31533.
40. Shedge V, Davila J, Arrieta-Montiel MP, Mohammed S, Mackenzie SA. Extensive rearrangement of the *Arabidopsis* mitochondrial genome elicits cellular conditions for thermotolerance. *Plant Physiol.* 2010;152(4):1960–70.
41. Xu YZ, Arrieta-Montiel MP, Viridi KS, de Paula WBM, Widhalm JR, Basset GJ, et al. MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. *Plant Cell.* 2011;23(9):3428–41.
42. Cheng L, Wang W, Yao Y, Sun Q. Mitochondrial RNase H1 activity regulates R-loop homeostasis to maintain genome integrity and enable early embryogenesis in *Arabidopsis*. *PLoS Biol.* 2021;19(8):e3001357.
43. Viridi KS, Wamboldt Y, Kundariya H, Laurie JD, Keren I, Kumar KRS, et al. MSH1 is a plant organellar DNA binding and thylakoid protein under precise spatial regulation to alter development. *Mol Plant.* 2016;9(2):245–60.
44. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol.* 2017;68:225–52.
45. Chen XL, Li SM, Zhang D, Han MY, Jin X, Zhao CP, et al. Sequencing of a wild apple (*Malus baccata*) genome unravels the differences between cultivated and wild apple species regarding disease resistance and cold tolerance. *G3 (Bethesda).* 2019;9(7):2051–60.
46. Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, et al. Origin and evolution of the octoploid strawberry genome. *Nat Genet.* 2019;51(3):541–7.
47. Qiao Q, Edger PP, Xue L, Qiong LJ, Zhang Y, Cao Q, et al. Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proc Natl Acad Sci U S A.* 2021;118(45):e2105431118.
48. Groppi A, Liu S, Cornille A, Decroocq S, Bui QT, Tricon D, et al. Population genomics of apricots unravels domestication history and adaptive events. *Nat Commun.* 2021;12(1):3956.
49. Li Y, Cao K, Li N, Zhu GR, Fang WC, Chen CW, et al. Genomic analyses provide insights into peach local adaptation and responses to climate change. *Genome Res.* 2021;31(4):592–606.
50. Gross BL, Olsen KM. Genetic perspectives on crop domestication. *Trends Plant Sci.* 2010;15(9):529–37.
51. Tang H, Zheng X, Li C, Xie X, Chen Y, Chen L, et al. Multi-step formation, evolution, and functionalization of new cytoplasmic male sterility genes in the plant mitochondrial genomes. *Cell Res.* 2017;27(1):130–46.
52. Lajbner Z, Pini R, Camus MF, Miller J, Dowling DK. Experimental evidence that thermal selection shapes mitochondrial genome evolution. *Sci Rep.* 2018;8(1):9500.
53. Li X, Liu L, Ming M, Hu H, Zhang M, Fan J, et al. Comparative transcriptomic analysis provides insight into the domestication and improvement of pear (*Pyrus pyrifolia*) fruit. *Plant Physiol.* 2019;180(1):435–52.
54. Duan NB, Bai Y, Sun HH, Wang N, Ma YM, Li MJ, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun.* 2017;8:249.
55. Nikiforova SV, Cavalieri D, Velasco R, Goremykin V. Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Mol Biol Evol.* 2013;30(8):1751–60.
56. Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 2013;23(2):396–408.
57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
58. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014;9(11):e112963.
59. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinf.* 2009;10:421.
60. Zhang T, Zhang X, Hu S, Yu J. An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. *Plant Methods.* 2011;7:38.
61. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinf.* 2012;13:238.
62. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–W11.
63. Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol.* 2010;27(6):1436–48.
64. Guo W, Zhu A, Fan W, Mower JP. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytol.* 2017;213(1):391–403.
65. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
66. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
67. Li Y, Smith T, Liu CJ, Awasthi N, Yang J, Wang YF, et al. Endocarps of *Prunus* (Rosaceae: Prunoideae) from the early Eocene of Wutu, Shandong Province, China. *Taxon.* 2011;60(2):555–64.
68. Gray J. The lower tertiary floras of Southern England. *Science.* 1964;144(3619):719–20.
69. Hohmann N, Wolf EM, Lysak MA, Koch MA. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell.* 2015;27(10):2770–84.
70. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010;5(6):e11147.
71. Tesler G. GRIMM: genome rearrangements web server. *Bioinformatics.* 2002;18(3):492–3.



72. Hu F, Lin Y, Tang J. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinformatics*. 2014;15:354.
73. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
74. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Proc GPD: the Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
75. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernyt-sky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
76. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92.
77. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
78. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64.
79. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
80. Li MR, Shi FX, Li YL, Jiang P, Jiao L, Liu B, et al. Genome-wide variation patterns uncover the origin and selection in cultivated ginseng (*Panax ginseng* Meyer). *Genome Biol Evol*. 2017;9(9):2159–69.
81. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



RESEARCH ARTICLE

Open Access



# Comparison of multiple algorithms to reliably detect structural variants in pears

Yueyuan Liu<sup>†</sup>, Mingyue Zhang<sup>†</sup>, Jieying Sun, Wenjing Chang, Manyi Sun, Shaoling Zhang and Jun Wu\*

## Abstract

**Background:** Structural variations (SVs) have been reported to play an important role in genetic diversity and trait regulation. Many computer algorithms detecting SVs have recently been developed, but the use of multiple algorithms to detect high-confidence SVs has not been studied. The most suitable sequencing depth for detecting SVs in pear is also not known.

**Results:** In this study, a pipeline to detect SVs using next-generation and long-read sequencing data was constructed. The performances of seven types of SV detection software using next-generation sequencing (NGS) data and two types of software using long-read sequencing data (SVIM and Sniffles), which are based on different algorithms, were compared. Of the nine software packages evaluated, SVIM identified the most SVs, and Sniffles detected SVs with the highest accuracy (> 90%). When the results from multiple SV detection tools were combined, the SVs identified by both MetaSV and IMR/DENOM, which use NGS data, were more accurate than those identified by both SVIM and Sniffles, with mean accuracies of 98.7 and 96.5%, respectively. The software packages using long-read sequencing data required fewer CPU cores and less memory and ran faster than those using NGS data. In addition, according to the performances of assembly-based algorithms using NGS data, we found that a sequencing depth of 50× is appropriate for detecting SVs in the pear genome.

**Conclusion:** This study provides strong evidence that more than one SV detection software package, each based on a different algorithm, should be used to detect SVs with higher confidence, and that long-read sequencing data are better than NGS data for SV detection. The SV detection pipeline that we have established will facilitate the study of diversity in other crops.

**Keywords:** SV detection, NGS, Long-read sequencing, Sequencing depth, Accuracy of SVs, SV calling pipeline

## Background

Structural variants (SVs), which include deletions, insertions, inversions, duplications and translocations, are defined as rearrangements in chromosomes larger than 50 nucleotides [1]. Translocations can also be classified as intra-chromosomal translocations (ITXs) and inter-chromosomal translocations (CTXs), based on whether the chromosome of the source locus is the same as that of the target locus [2]. Deletions, insertions and duplications are called unbalanced SVs because they give rise to copy number variants (CNVs), while inversions and translocations are called balanced SVs [2]. It is clear that

SVs play an important role in biological processes, and the identification of SVs is crucial for studying human genetic diversity, gene and genome variants, evolution and disease [3, 4]. SVs have been shown to be related to human diseases, such as immune escape of tumor cells [5], chronic hepatitis B virus infection [6] and heart failure [7]. SVs such as insertions and deletions and CNVs have been shown to contribute to natural variation of plants and have played a significant role in the differentiation of complex traits, domestication, evolution and adaptation [8, 9]. For example, a CNV involving four genes that define the *Female* locus in cucumber, which arose from a recent 30.2-kb duplication in a meiotically unstable region, gave rise to gynoeious plants [10]. The study of single nucleotide polymorphisms (SNPs), InDels and CNVs in tomato revealed introgressions from wild species and the mosaic structure of the genomes of

\* Correspondence: [wujun@njau.edu.cn](mailto:wujun@njau.edu.cn)

<sup>†</sup>Yueyuan Liu and Mingyue Zhang contributed equally to this work. Center of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China



cherry tomato accessions [11]. In ‘Su Shuai’ apple, SVs in 17 genes associated with disease resistance, 10 genes relevant to gibberellin and 19 genes related to fruit flavor were identified [12].

Pear is the third most important fruit species of the Rosaceae and is widely cultivated all over the world. The *Pyrus* genus is genetically diverse with thousands of cultivars, and studying SVs in *Pyrus* can lead to a better understanding of genetic diversity among cultivars and the genetic basis for complex traits. Previous studies have shown that SVs can influence crop traits, domestication, and evolution [8–12], but little is known about the SVs in *Pyrus*. Moreover, SV detection software was originally developed and tested using the human genome or the genome of the model plant *Arabidopsis thaliana*, so this software may not efficiently detect SVs in pear. Sequencing of the genome of *Pyrus bretschneideri* cv. ‘Dangshansuli’ pear, a variety that originated in China, in 2013 [13], revealed that it shows large differences from the *A. thaliana* genome. For example, the *A. thaliana* genome is smaller (only 125 Mb) and has fewer repetitive sequences than the genomes of pear and most fruit crops [13]. Thus, the development of a pipeline to detect SVs in *Pyrus* is of great significance for facilitating studies of genome complexity in the Rosaceae.

Recently, the availability of next-generation sequencing (NGS) and long-read sequencing data has greatly facilitated the characterization of SVs because variants of different sizes and types can be detected and breakpoints can accurately be identified at base-pair resolution [14–16]. NGS generates short reads ranging from 35 bp to 700 bp in length, while the long reads generated by third generation sequencing technology are over 10 kb in length [17]. A sufficient sequencing depth is required to detect SVs. For the human genome, 35-bp paired-end reads with an average depth of > 30× were used to build an accurate consensus sequence and characterize a million SNPs and 400,000 SVs [18]. A lower sequencing depth, > 10×, was found to be sufficient for detecting SVs when using reads over 10 kb in length [16]. However, the most suitable sequencing depth for detecting SVs in pear has not been determined.

To date, many approaches have been developed to detect SVs using NGS data. These algorithms are classified into four distinct categories based on the method used to detect SVs: read depth, read pairs, split reads, and assembly [19]. Algorithms based on read-depth signals can detect duplications and deletions using all mapped reads, but only at coarse resolution [20]. Read-depth algorithms are more effective for detecting larger (> 1 kb) CNVs. However, they cannot detect inversions. Read-pair algorithms are more popular for detecting SVs because of their relative simplicity and their ability to

detect all SV types [21–23]. Split read-based callers can work with low-coverage NGS data and identify SVs with base resolution. However, the disadvantages of split-read callers are that they cannot detect larger SVs such as duplications, inversions, translocations, and more complex variants because some short reads may map to many locations in the reference genome [24, 25]. When using assembly-based callers (de novo and reference-based assembly callers), short reads need to be assembled into longer sequence stretches called contigs before detection [26]. Because the contigs are longer than individual reads, SVs are called with high confidence. Many software packages have been developed for detecting more types of SVs with higher accuracy by integrating multiple algorithms (such as DELLY [27] and Lumpy [28]) or merging the outputs of multiple software (such as FusorSV [29], MetaSV [30] and Parliament2 (<https://github.com/dnanexus/parliament2>)). Callers using NGS data have a high rate of SV miscalling due to errors in alignment or de novo assembly, especially in repetitive regions that cannot be spanned with short reads [31]. To overcome these issues, software using long-reads such as SVIM [32] and Sniffles [16] have been developed; these algorithms are mostly based on split reads. The functions and features of each type of SV-calling software are known, but the reliability of using different combinations of software for detecting SVs has not been studied.

In this paper, we evaluated the effectiveness of several types of SV detection software in *Pyrus*. The pear cultivar chosen was ‘Yali’ (*P. bretschneideri*), which is genetically closely related to ‘Dangshansuli’ (*P. bretschneideri*) and is one of the primary pear cultivars grown in China. This cultivar is also exported to other countries where it is known as Asian pear. We have conducted a systematic analysis using ‘Yali’ genome NGS and long-read sequencing data to compare the performances of several commonly used SV-calling software packages using short reads, namely Pindel [25], BreakDancer [33], IMR/DENOM [34], Platypus [35], DELLY [27], Lumpy [28], and MetaSV [30], and software packages using long reads, namely SVIM [32] and Sniffles [16]. The effects of different sequencing depths on SV detection were investigated, and the most appropriate sequencing depth for detecting SVs in *Pyrus* was determined by comparing the number of SVs detected and the computational resources required for different sequencing depths. Moreover, we investigated the overlap in SVs identified by all possible combinations of two or three software packages to obtain high-confidence SVs. Then, the reliability of selected ‘Yali’ pear SVs was verified using visualization tools. Our findings lay the foundation for subsequent studies of SVs, and the pipeline we



constructed can be used to reliably detect SVs in other crops.

## Results

### Sequencing and mapping of the ‘Yali’ genome

Short read sequencing of the pear ‘Yali’ genome was conducted using the Illumina HiSeq™ 2000 platform for pair-end sequencing, and the sequencing depth was 60×. A total of 103,584,796,150-bp reads were obtained, and the GC content was 39%. The quality of the raw resequencing data was determined using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) software. After using Trimmomatic [36] to filter the low quality sequencing data, 97.84% of the reads were kept. Of the clean reads, 97.15% were mapped to the ‘Dangshansuli’ pear genome using Burrows-Wheeler-Aligner (BWA) software [37]. Seven SV detection software packages using NGS data (Table 1) were then used to identify SVs in ‘Yali’.

Long-read sequencing data for ‘Yali’ were generated using the PacBio platform, and the sequencing depth was 30×. A total of 2,977,899 subreads were obtained. The average subread length was 6 kb and the N50 was 8 kb. Two SV detection software packages (Sniffles and SVIM) using long read sequencing data (Table 1) were selected to identify SVs in ‘Yali’.

### SVs between ‘Yali’ and the reference genome detected using different algorithms and sequencing data

Depending on the performances of the nine SV callers, which are based on different algorithms (Table 1), up to eight types of SVs in the ‘Yali’ genome were detected: insertions, deletions, inversions, duplications, translocations, MNPs (multiple nucleotide polymorphisms), CTXs and ITXs (Table 1). Deletions were the only SVs detected by all nine callers. The number of SVs detected by the nine callers, categorized based on type and length,

is shown Fig. 1. Of the nine SV callers, SVIM detected the highest number of SVs. The software with assembly-based algorithms called fewer SVs than the other types of software, and Platypus called the fewest SVs. Although both DELLY and Lumpy use split-read and read-pair algorithms, DELLY called a higher number of SVs and more types of SVs than Lumpy. Detailed information about the number of SVs called by each software package is shown in Fig. 1.

For Pindel, which uses an split-read algorithm, short reads need to be broken into smaller fragments and mapped separately to the reference genome [25]. A total of 22,548 SVs were found using Pindel: 1178 insertions, 11,445 deletions, 9791 inversions and 134 duplications (Fig. 1). Deletions accounted for the largest proportion (50.76%) of the SVs and inversions accounted for the second largest proportion (43.42%). Compared with deletions and inversions, the numbers of insertions and duplications were very small, accounting for 5.22 and 0.59% of the SVs, respectively. In addition, Pindel could not detect insertions greater than 200 bp in length in the ‘Yali’ pear genome. Therefore, Pindel performed better in detecting small insertions and deletions and only detected a limited number of large SVs (>1 kb) (Fig. 1).

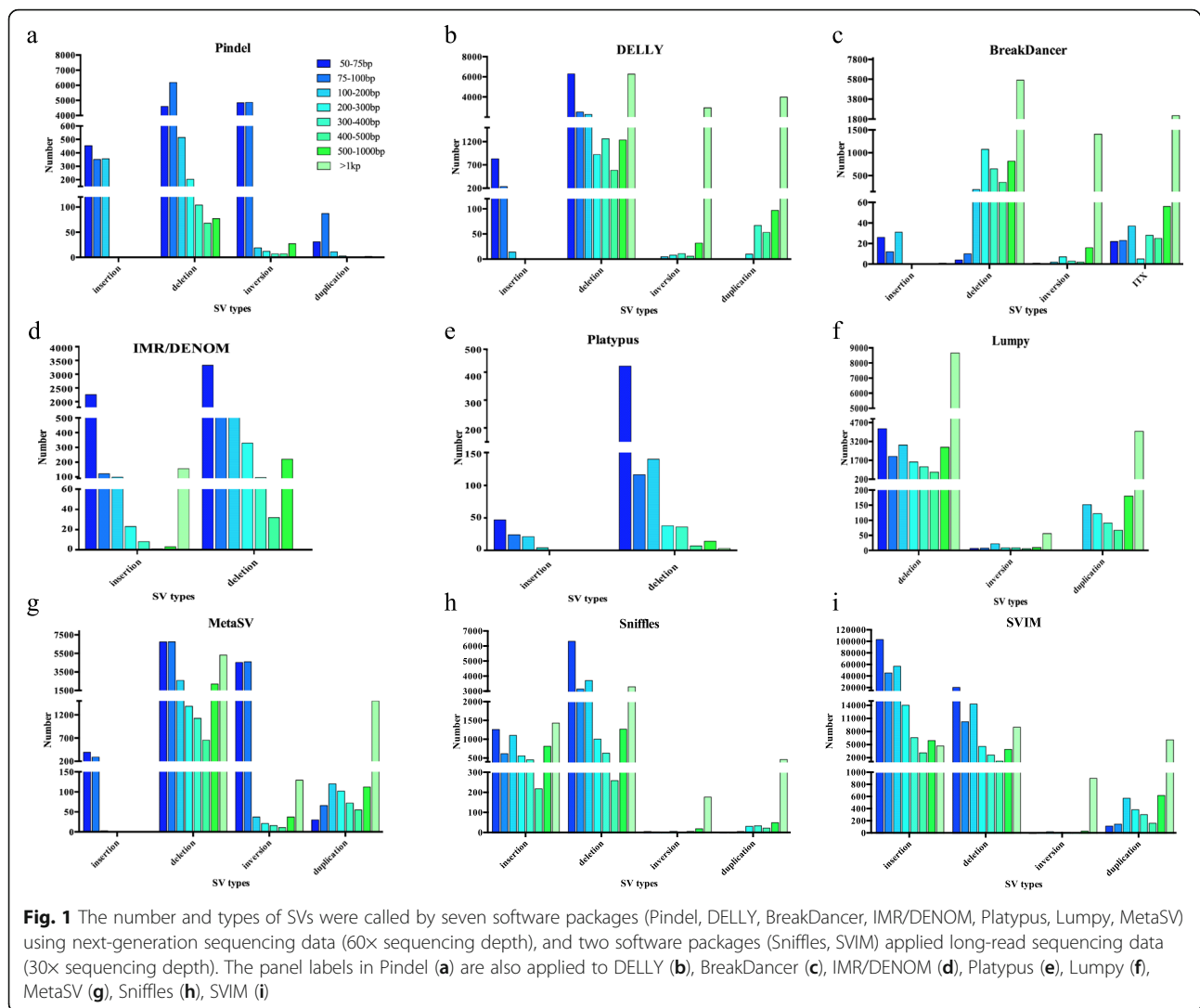
BreakDancer detects SVs using a read-pair algorithm; reads that map with an abnormal insert size or orientation are collected and then classified as insertions, deletions, inversions, or translocations [33]. Using BreakDancer, a total of 8682 SVs were detected: 90 insertions, 6900 deletions, 1398 inversions, and 294 ITXs. Of the SVs 79.47% were deletions, and no insertions longer than 400 bp were identified (Fig. 1). Therefore, BreakDancer is not suitable for detecting small variants or large insertions in pear.

IMR/DENOM [34] utilizes local de novo assemblies and iterative read mapping to the reference sequence to identify SVs [38]. IMR/DENOM called a total of 8398

**Table 1** Comparison of the nine types of SV detection software

| Data type     | Detection tools | Detectable SV types |     |     |     |     |     |     |      | Algorithms |
|---------------|-----------------|---------------------|-----|-----|-----|-----|-----|-----|------|------------|
|               |                 | INS                 | DEL | INV | DUP | ITX | CTX | TRA | MNPs |            |
| Illumina data | Pindel          | Yes                 | Yes | Yes | Yes | No  | No  | No  | No   | SR         |
|               | BreakDancer     | Yes                 | Yes | Yes | No  | Yes | Yes | No  | No   | RP         |
|               | DELLY           | Yes                 | Yes | Yes | Yes | No  | No  | Yes | No   | RP + SR    |
|               | IMR/DENOM       | Yes                 | Yes | No  | No  | No  | No  | No  | No   | AS         |
|               | LUMPY           | No                  | Yes | Yes | Yes | No  | No  | Yes | No   | RP + SR    |
|               | Platypus        | Yes                 | Yes | No  | No  | No  | No  | No  | Yes  | AS         |
|               | MetaSV          | Yes                 | Yes | Yes | Yes | Yes | Yes | Yes | No   | –          |
| PacBio data   | Sniffles        | Yes                 | Yes | Yes | Yes | No  | No  | Yes | No   | SR         |
|               | SVIM            | Yes                 | Yes | Yes | Yes | No  | No  | No  | No   | SR         |

Notes. An overview of the nine SV callers, including the types of SVs detected (INS: insertion, DEL: deletion, INV: inversion, DUP: duplication, TRA: Translocation, ITX: intra-chromosomal translocation, CTX: inter-chromosomal translocation) and the mutation signals used (SR: split reads, RP: read pairs, AS: assembly). The symbol ‘–’ indicates that the algorithm is chosen by the user



SVs (2514 insertions, 5884 deletions). IMR/DENOM could detect large insertions (> 1 kb) but it could not detect large deletions in 'Yali' (> 1 kb) (Fig. 1).

Platypus [35] detects deletions and insertions when using the assembly option, but this caller detected fewer and smaller SVs than the other callers; only 92 insertions, 776 deletions and 886 other complex SVs were detected. Moreover, Platypus could not call insertions longer than 300 bp, and over 50% of the SVs identified ranged from 50 bp to 75 bp in length. Therefore, this software performed better in detecting small insertions and deletions (Fig. 1).

DELLY has the ability to integrate pair-end data from libraries with different insert sizes with split-read data, making it a versatile tool for analyzing SVs using deep whole-genome sequencing data [27]. Using DELLY, 1054 insertions, 20,991 deletions, 2976 inversions and 4217 duplications were identified (Fig. 1). About 30% of deletions were longer than 1 kb. Similar to Pindel, DELLY could not

detect insertions longer than 200 bp. However, unlike Pindel, DELLY was not capable of detecting inversions and duplications less than 100 bp in length. Moreover, more than 97% of the inversions and more than 94% of the duplications called by DELLY were greater than 1 kb in length.

Lumpy [28] integrates multiple algorithms including those using read pairs, split reads and read depth. It detected 24,072 deletions, 127 inversions, and 4620 duplications. Over 35% of deletions, 44% of inversions and 87% of duplications were longer than 1 kb (Fig. 1). Therefore, Lumpy has superior sensitivity in detecting SVs longer than 1 kb.

MetaSV [30] detects SVs by merging the outputs of other SV detectors, such as Pindel, BreakDancer and Lumpy. It can also detect insertions by analyzing soft-clipped reads from alignments and improve the break-points of SVs using local assembly. To further compare the accuracy of SVs called by Pindel, BreakDancer and Lumpy, we only used the merge option without soft-clip-based analysis or local assembly. According to the

merged results, 689 insertions, 26,770 deletions, 9381 inversions and 2057 duplications were detected (Fig. 1). Almost all insertions and inversions ranged from 50 bp to 100 bp in size, and over 50% of deletions were between 50 bp and 100 bp in length. More than 50% of duplications were longer than 1 kb.

Sniffles, which uses long-read sequencing data [16], detects SVs from long-read alignments using a split-read algorithm with the NGMLR aligner. It detected 6556 insertions, 19,774 deletions, 242 inversions and 633 duplications (Fig. 1). The other software package using long-read sequencing data, SVIM [32], detects SVs in a process consisting of three steps: collection, clustering and combining of SVs from read alignments. SVIM detected 242,429 insertions, 67,950 deletions, 1019 inversions and 8609 duplications. SVIM detected more SVs than Sniffles, suggesting that SVIM detects SVs with higher sensitivity (Fig. 1).

### The SVs identified by multiple software are more accurate

We next investigated the overlap between SVs detected by multiple SV callers that use NGS data (each based on a different algorithm). The Integrative Genomics Viewer (IGV) browser was first used to confirm the presence of the SVs called by each caller. We randomly selected 660 deletions ranging from 50 bp to 500 bp in length from the output of single callers using NGS data. The accuracies of each type of software are shown in Additional file 8. The accuracies of Pindel (58%) and BreakDancer (58%) were lower than those of the other callers. For Pindel, the accuracy in calling SVs ranging from 50 bp to 75 bp in size was 75% while the accuracy in calling SVs ranging from 400 bp to 500 bp in size was 33%. Therefore, Pindel detected small SVs with high

sensitivity and confidence, with accuracy decreasing as SV length increased. The DELLY and Lumpy algorithms performed similarly, and the accuracy of SVs called by DELLY (63%) was a little better than that of Lumpy (60%). For the IMR/DENOM and Platypus software packages, which are based on assembly, the average accuracies of SV detection (81 and 66%, respectively) were higher than those of the other types of software, demonstrating that callers based on assembly algorithms detect SVs with higher confidence. The accuracy of the SVs called by MetaSV (70%), which were merged from the results of Pindel, BreakDancer and Lumpy, was higher than that of each caller alone. Therefore, the SVs called by merging outputs from multiple callers are more accurate than single SV caller.

According to the performances of the seven software packages using NGS data, Pindel, BreakDancer, IMR/DENOM and DELLY were selected for finding overlapping SVs (Table 2). Because the SVs called by MetaSV were merged from the outputs of Pindel, BreakDancer and Lumpy, we simply combined the outputs of MetaSV and IMR/DENOM to identify overlapping SVs and determine whether they were more accurate. We found the number of overlapping SVs from random combinations of Pindel, BreakDancer, IMR/DENOM and DELLY (Table 2). Based on the percentages of overlapping insertions, deletions, inversions and duplications identified by each software, DELLY performed better than the other three software packages (Table 2).

When focusing on Pindel and DELLY, we found very little overlap in the insertions identified by the two programs, with only 0.25% of Pindel insertions and 0.28% of DELLY insertions overlapping. However, greater than 80% of inversions were predicted by both software. A

**Table 2** The number of structural variations detected by individual algorithms and combinations of algorithms

| Combination                  | Insertion | Deletion | Inversion | Duplication |
|------------------------------|-----------|----------|-----------|-------------|
| Pindel                       | 1178      | 11,445   | 9791      | 134         |
| DELLY                        | 1054      | 20,991   | 2976      | 4217        |
| BreakDancer                  | 90        | 6900     | 1398      | 0           |
| IMR/DENOM                    | 2514      | 5884     | 0         | 0           |
| Pindel-DELLY                 | 3         | 8782     | 7997      | 89          |
| Pindel-BreakDancer           | 0         | 7616     | 6442      | 0           |
| Pindel-IMR/DENOM             | 1         | 502      | 0         | 0           |
| DELLY-BreakDancer            | 0         | 1192     | 129       | 0           |
| DELLY-IMR/DENOM              | 307       | 5152     | 0         | 0           |
| BreakDancer-IMR/DENOM        | 0         | 4729     | 0         | 0           |
| Pindel-DELLY-IMR/DENOM       | 1         | 443      | 0         | 0           |
| Pindel-DELLY-BreakDancer     | 0         | 7613     | 6441      | 0           |
| DELLY-BreakDancer-IMR/DENOM  | 0         | 4423     | 0         | 0           |
| Pindel-BreakDancer-IMR/DENOM | 0         | 361      | 0         | 0           |

high percentage, 66.42%, of the duplications identified by Pindel were also identified by DELLY, but only 2.11% of those identified by DELLY were also identified by Pindel. There was a higher number of overlapping deletions, with 76.73% of Pindel deletions also identified by DELLY, and 41.83% of DELLY deletions identified by Pindel.

The number of overlapping SVs between IMR/DENOM and Pindel and between IMR/DENOM and DELLY were shown in Table 2, respectively. Since IMR/DENOM can only detect insertions and deletions (Table 1), the number of inversions and duplications overlapping with those identified by the other three software packages was 0. Only one insertion and 502 deletions were detected by both Pindel and IMR/DENOM. Of the deletions identified by IMR/DENOM, 8.53% were also identified by Pindel, and 66.54% of the Pindel deletions overlapped with the IMR/DENOM deletions. For IMR/DENOM and DELLY, 307 insertions and 5152 deletions were discovered by both programs. Of the DELLY insertions, 26.06% were identified by IMR/DENOM, and 12.21% of IMR/DENOM insertions were identified by DELLY. However, 45.02% of the DELLY deletions overlapped with those identified by IMR/DENOM, while over 85% of IMR/DENOM deletions were identified by DELLY. IMR/DENOM and BreakDancer had no overlapping insertions, while the number of overlapping deletions was 4729.

There were few overlapping insertions between BreakDancer and DELLY and between BreakDancer and Pindel. However, a large number of deletions were called by both BreakDancer (100% overlapped with Pindel deletions) and Pindel (66.54% overlapped with BreakDancer deletions). Although 100% of the BreakDancer deletions also overlapped with those identified by DELLY, only 5.68% of DELLY deletions were identified by BreakDancer.

When comparing the combination of three software packages, few of the insertions called by Pindel, DELLY and IMR/DENOM overlapped, and no insertions called by these programs overlapped with those called by BreakDancer. However, there was better overlap in the deletions called by combinations of three software. Although Pindel, DELLY and IMR/DENOM shared fewer than 10% of deletions with each other, when comparing the output of Pindel, DELLY and BreakDancer, all of the deletions identified by BreakDancer, 66% of the deletions identified by Pindel and 36.27% of deletions identified by DELLY overlapped. A high number of overlapping inversions was also observed when combining DELLY (100%), BreakDancer (100%) and Pindel (65.78%). When comparing DELLY, BreakDancer and IMR/DENOM, 21.07% of deletions identified by DELLY, 75.17% of those identified by IMR/DENOM and 64.10% of those identified by

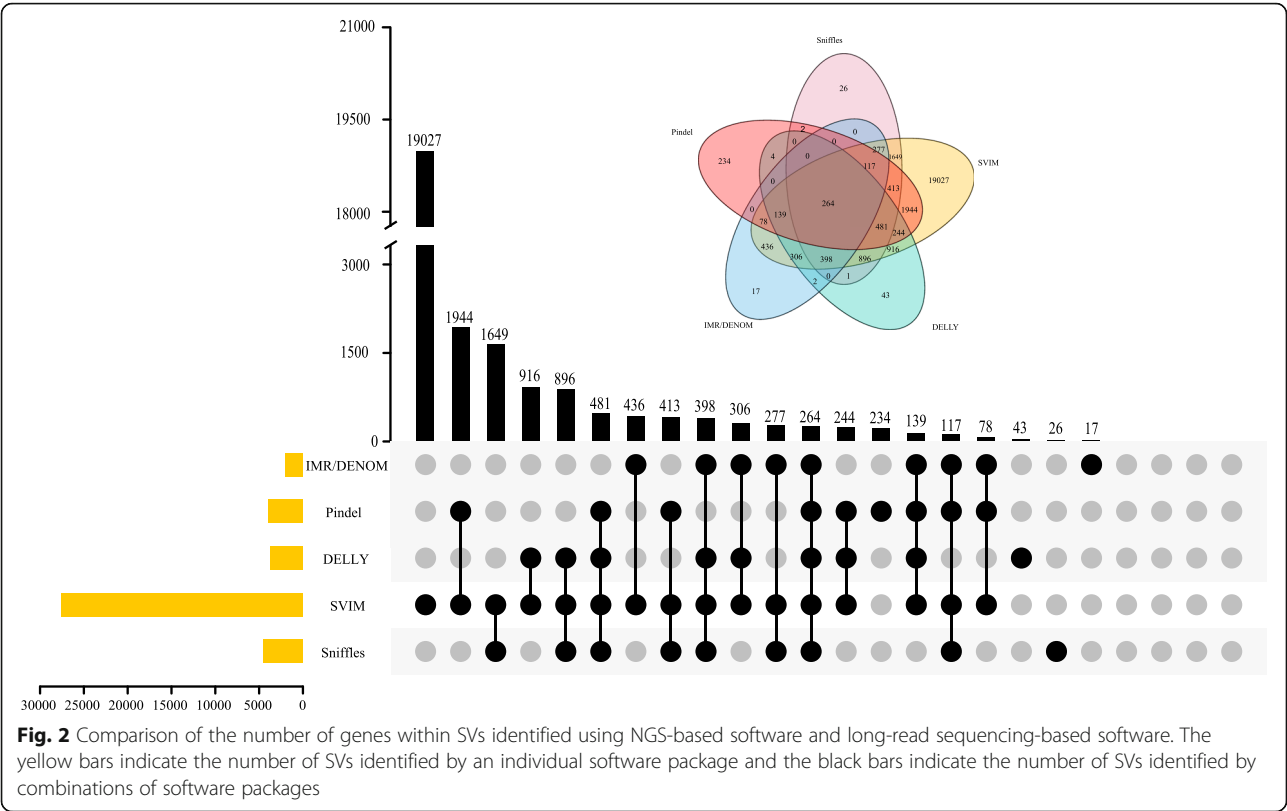
BreakDancer overlapped. When comparing Pindel, IMR/DENOM and BreakDancer, 3.16% of deletions identified by Pindel, 5.23% of those identified by BreakDancer and 6.14% of those identified by IMR/DENOM overlapped.

To confirm the accuracy of SVs from multiple software packages using NGS data, we randomly chose 940 overlapping SVs from the output of two software packages combined and three packages combined. The average accuracy of overlapping deletions was higher than the accuracy of deletions called by a single software package (Additional file 8). Moreover, the accuracies of SVs identified by the combinations Pindel and DELLY, Pindel and BreakDancer, and DELLY and BreakDancer were lower than those of SVs identified by the combinations Pindel and IMR/DENOM, DELLY and IMR/DENOM, and BreakDancer and IMR/DENOM. The average accuracy of overlapping SVs identified by Pindel, DELLY and BreakDancer was lower than that of overlapping SVs identified by Pindel, DELLY and IMR/DENOM; DELLY, BreakDancer and IMR/DENOM; and Pindel, BreakDancer and IMR/DENOM. In particular, the average accuracy of overlapping deletions from MetaSV, which included the merged results of Pindel, BreakDancer and Lumpy, and IMR/DENOM was greater than 90%. This indicates that the SVs detected by a combination of assembly-based software and multiple algorithm-based software were more accurate than those detected by the other combinations of software.

To further validate the accuracy by long-read resequencing data, we randomly selected 300 SVs identified by the software packages from Sniffles (100 SVs), SVIM (100 SVs) and Sniffles\_SVIM (SVs). The average accuracy of SVs detected by Sniffles was greater than 95%, while the accuracy of SVs detected by SVIM was less than 80%. The SVs overlapping between Sniffles and SVIM were high confidence SVs with an accuracy greater than 96%. Compared with algorithms using NGS data, the algorithms using long-read sequencing data detected SVs with higher accuracy, and large SVs with more confidence. However, the SVs overlapping between MetaSV and IMR/DENOM were more accurate than those overlapping between Sniffles and SVIM, which suggests that SVs detected by a combination of assembly-based software and multiple algorithm-based software are the most accurate.

We then annotated the SVs detected by five individual callers, three using NGS data, each based on a different algorithm (Pindel, DELLY, and IMR/DENOM), and two using long-read sequencing data (Sniffles, which detected more SVs, and SVIM, which detected higher-confidence SVs), and observed the number of genes within SVs commonly identified by these callers (Fig. 2). Among the callers based on paired-read algorithms, DELLY was chosen because it performed better than





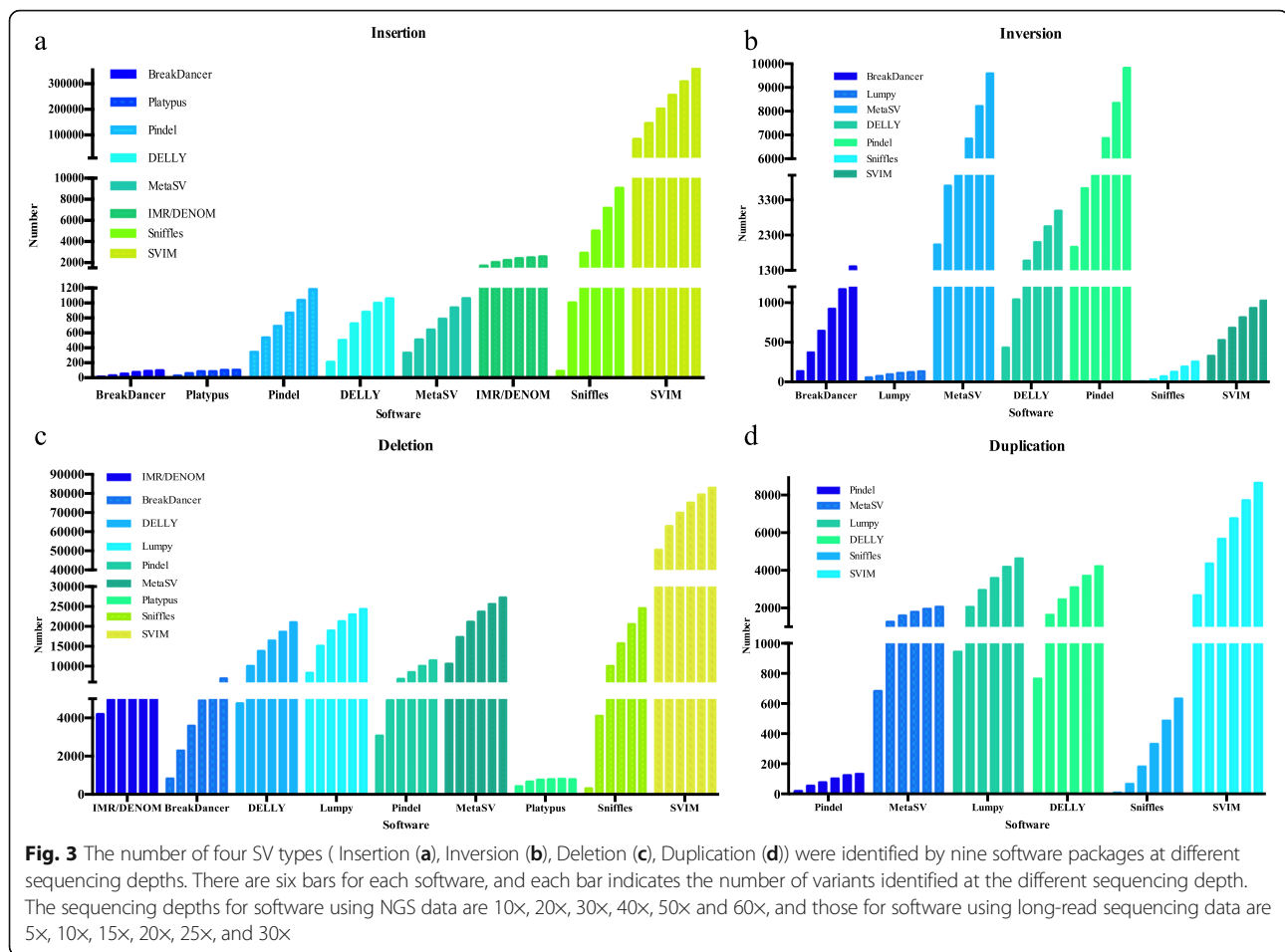
BreakDancer and Lumpy. The assembly-based caller IMR/DENOM was chosen because it detected more SVs than Platypus. The split-read-based caller, Pindel was chosen because it was better able to detect SVs less than 100 bp in length. A total of 264 genes within SVs were detected using the five software packages. These genes were subjected to functional enrichment analysis using both the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases (results are shown in Additional files 1 and 2). These 264 genes will be the main targets for future functional studies of the variants between ‘Yali’ and ‘Dangshansuli’ pear. A total of 403 genes within SVs were commonly detected by the callers using NGS data, and 4495 genes within SVs were commonly detected by the callers using long-read sequencing data (Additional file 3: Figure S1(a) and (b), respectively). The results of GO and KEGG analysis of these genes are shown in Additional files 4, 5, 6 and 7.

Effect of sequencing depth on SV detection

To determine the most appropriate sequencing depth for detecting SVs in pear, the performances of all software packages using NGS data (except MetaSV) and both software packages using long-read sequencing data at different sequencing depths were compared. *Seqtk* was used to obtain NGS (10×, 20×, 30×, 40×, 50×, 60×) and long-read sequencing (5×, 10×, 15×, 20×, 25×, 30×)

data at different sequencing depths (Fig. 3). For IMR/DENOM and Platypus, the number of SVs increased as sequencing depth increased to 50×. When the NGS depth increased to 60×, the number of variants called by IMR/DENOM and Platypus did not change too much, and even decreased. Based on this analysis, for assembly-based software an NGS depth of 50× is sufficient for detecting SVs in *Pyrus*. For Pindel, BreakDancer, DELLY, Lumpy, Sniffles and SVIM, the number of SVs called obviously increased as the sequencing depth increased. Therefore, for split read-based and read pair-based software, the higher the depth of sequencing, the higher the number of SVs detected in *Pyrus*.

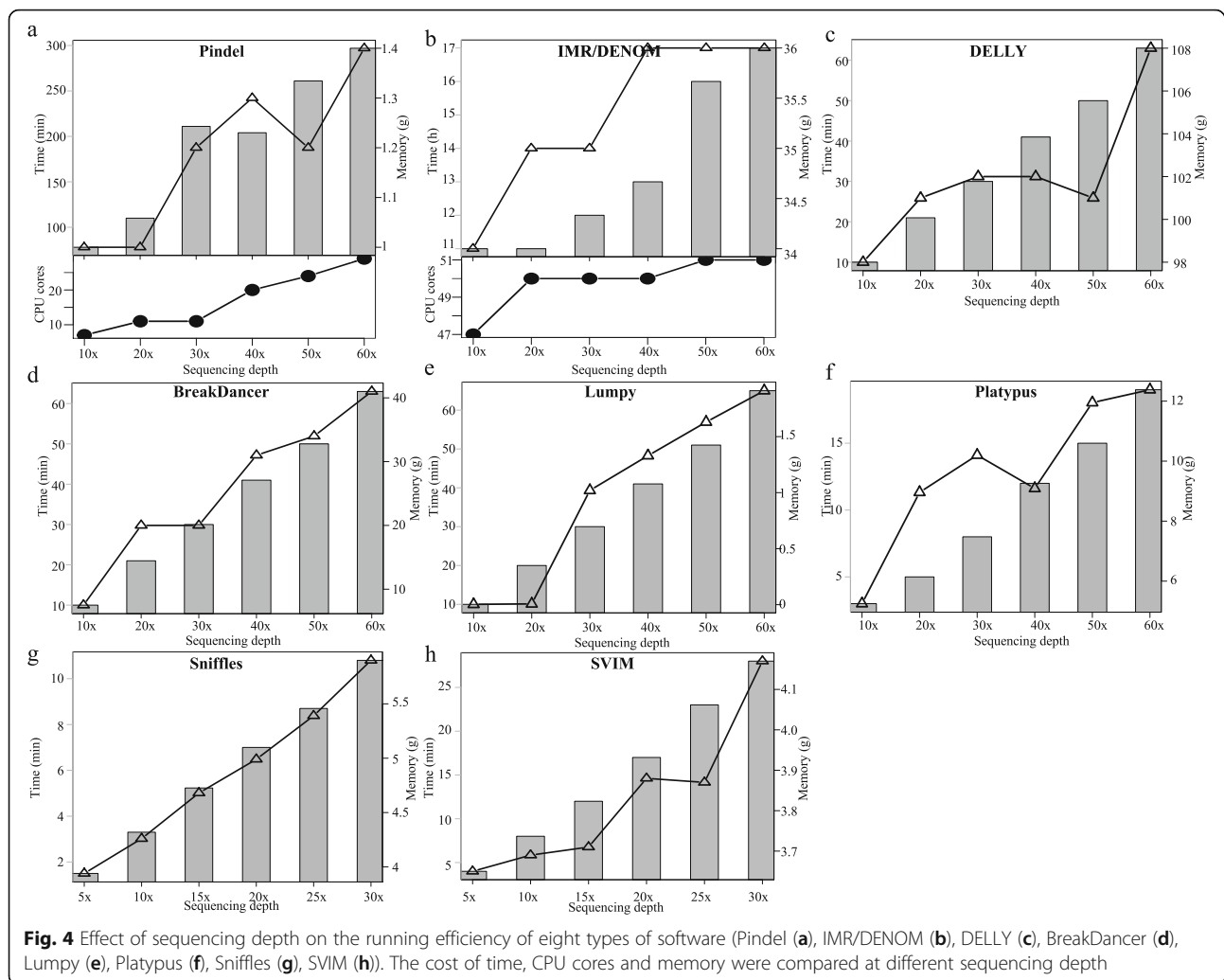
The computational time, the number of CPU cores required, and memory cost also need to be considered when determining the most suitable sequencing depth. Therefore, software performance at different sequencing depths was also evaluated. The performance of each SV caller was determined based on the mean computational time and computational memory cost with different parameters. The running time and maximum memory occupancies for the eight callers at different sequencing depths are shown in Fig. 4. When running DELLY, BreakDancer, Lumpy and SVIM, threads cannot be set, so the default CPU core was one. However, for Pindel, IMR/DENOM and Sniffles, different threads can be set to decrease the



computational time for SV detection depending on the running environment. Therefore, we set the thread to 50 for these programs to improve the detection efficiency. The number of CPU cores for Platypus can be specified, and we used 50 CPU cores. Platypus was able to detect SVs much faster than IMR/DENOM; Platypus required only about 3 min while IMR/DENOM required more than 10 h. As the depth of sequencing increased, so did the computational time, memory and the number of CPU cores required (Fig. 4). IMR/DENOM required more CPU cores and memory than the other programs. Sniffles was faster than SVIM, and both programs required the same amount memory. To sum up, for NGS data, DELLY is recommended because it requires less computational time and memory and because combinations of software that include DELLY identify more overlapping SVs than those that do not. If enough CPU cores and free memory on the server machine are available, IMR/DENOM is more suitable because of its high sensitivity and accuracy in detecting SVs. For long-read sequencing data, both Sniffles and SVIM are recommended, since SVIM can detect more SVs and Sniffles detects SVs with high confidence.

### Workflow for detecting accurate SVs

The goal of this study was to detect SVs with higher accuracy using the 'Yali' resequencing data. To facilitate the study of SVs in the future, we set up a workflow for SV detection based on the different algorithms evaluated in this study (Fig. 5). The workflows for SV detection using NGS data and long-read sequencing data were similar. Therefore, we describe the workflow for NGS data as an example. Firstly, quality control of the raw resequencing data was done, trimming the reads to obtain clean reads. Secondly, we mapped the clean reads to the 'Dangshansuli' reference genome. Thirdly, nine SV-calling software were used to detect SVs, and the overlapping SVs were identified using multiple types of software. The seven software packages using NGS data were mainly classified into two categories: software based on a single algorithm and software based on multiple algorithms. Pindel uses split reads and BreakDancer uses read pairs. IMR/DENOM and Platypus are based on assembly. The algorithms of Lumpy and DELLY are similar and both use read pairs and split reads. The algorithms of MetaSV merge outputs from multiple software. The overlapping SVs identified by multiple software packages were more

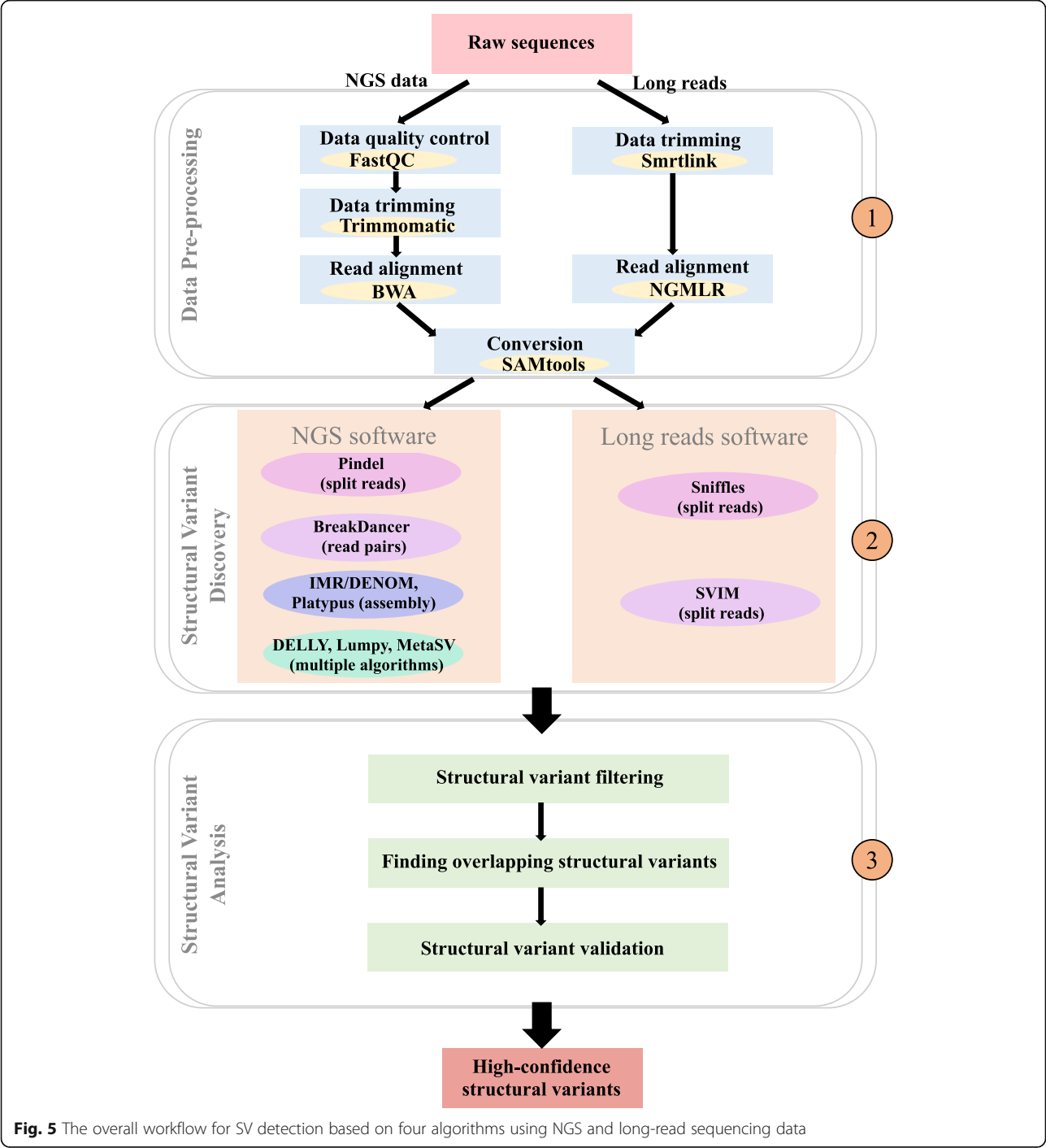


accurate. This high accuracy is essential when selecting SVs for further study.

## Discussion

In our study, we compared the sensitivities, accuracies and computational equipment requirements of seven common software packages using ‘Yali’ pear NGS data and two software packages using ‘Yali’ pear long-read sequencing data to provide insights for choosing the most appropriate SV-calling program. Detecting more SV types and decreasing the false discovery rate and increasing the sensitivity of SV detection have always been a concern for researchers studying SVs. Software developers have also focused on improving the sensitivity of SV detectors [27–29]. Here, we focused on the performance of seven software packages using NGS data (Fig. 1) and found that each software has its own advantages and disadvantages. For example, MetaSV detected many more SVs than the other packages tested; however, before running MetaSV, the VCF (variant call format) files of Pindel, BreakDancer,

Lumpy or other software need to be prepared, which is a cumbersome process. Not all software can detect insertions, and all programs except IMR/DENOM had limitations in the length of insertions detected. Using SAMtools, the mean sequence insert size of ‘Yali’ was found to be 320 bp. Only IMR/DENOM can detect insertions > 300 bp (Fig. 1), which explains why IMR/DENOM was the only caller able to detect insertions longer than the sequence inserts. Breakdancer, DELLY, Lumpy and MetaSV were more sensitive in detecting large deletions and Pindel was more sensitive in detecting small SVs (Fig. 1); this is because read-pair algorithms are less sensitive in detecting small SVs, which are below the standard deviation for insert size [14, 25, 27, 28]. Of the two software packages using long-read sequencing data, SVIM showed higher sensitivity, probably because SVIM collects, clusters and combines SV signatures from read alignments [32]. By contrast, Sniffles detects SVs from analysis of split-read alignments, high-mismatch regions and sequencing depth and coverage [16].



Validating the presence of SVs has been a challenge for researchers using the pear genome, but progress has been made in other model systems. For the human genome, a map of SVs was constructed based on whole-genome sequencing data from 185 human genomes, and most SVs were mapped to nucleotide resolution [39]. Many types of SV-calling software have been developed, and the sensitivities and false discovery rates of these callers have been tested using known SVs [28, 29]. In maize and rice, SVs were identified and mapped using the results of pan-genome analyses of population structure and diversity [40, 41]. However, the SVs in pear populations are unknown and characterizing high-confidence SVs is crucial for studying SVs using a single sample of pear. To evaluate the accuracy of different SV callers, we first compared the accuracy of SVs called by each software package, then



selected five software packages for finding overlapping SVs, and finally validated the accuracy of the overlapping SVs. The overlapping SVs from multiple packages had higher confidence than single software package.

Sequencing costs are often the biggest limitation for many laboratories [42]. For single nucleotide variant and small InDel detection, an average sequencing depth of 30× is the standard [43, 44]. But for SV detection, sensitivity and breakpoint detection can improve with increasing sequencing depth [42]. In our study, we found that the depth of NGS and long read sequencing absolutely affected the number of SVs called (Fig. 3), the cost of sequencing (Fig. 4) and the speed of variant calling (Fig. 4). Increasing sequencing depth within a certain range can increase the number of SVs identified and reduce the false discovery rate of variant calling when using NGS data. For IMR/DENOM and Platypus, the number of SVs called did not change with sequencing depths greater than 50×, and the computational time, number of CPU cores and memory required also showed no remarkable change. Therefore, when using IMR/DENOM and Platypus to detect SVs, the appropriate sequencing depth is 50×. By contrast, we did not identify an optimal sequencing depth for Pindel, DELLY, BreakDancer, Lumpy, MetaSV or software using long-read sequencing data (Sniffles and SVIM). For NGS-based callers with split-read, read-pair and read-depth algorithms and long-read sequencing-based callers with split-read algorithm, as the sequencing depth increased, the number of SVs called and the computational time, number of CPU cores and memory needed also increased.

## Conclusions

An SV detection pipeline using NGS and long-read sequencing data has been developed for application in pear, and this pipeline can be used for the study of SVs in other crops. Seven different types of SV-calling software packages using NGS data and two packages using long-read sequencing data were compared. SVIM detected SVs with the highest sensitivity, and Sniffles called SVs with the highest confidence. The SVs detected by software packages using long-read sequencing data showed higher accuracy than those using NGS data and also required less time, fewer CPU cores and less memory. A combination of multiple software packages is recommended for the detection of more types of SVs with higher accuracy. For NGS data, a sequencing depth of 50× is the most suitable for detecting SVs in pear based on the performance of assembly algorithm-based software. This information about the accuracy of SV detection, computational equipment requirements and suitable sequencing depth will benefit researchers engaged in the study of SVs. This study has provided important insights into methods for improving SV detection that can be applied in future studies of crop genomes.

## Methods

### Pear accession sequencing

‘Yali’ plants were grown in an experimental nursery at Changli Fruit Research Institute, Hebei Academy of Agricultural Sciences, China. Young leaves were collected 15 days after flowering, and extraction of DNA for NGS was performed using the Qiagen DNeasy 96 Plant Kit (Cat. no. 69181) following the manufacturer’s protocol. Paired-end sequencing libraries with an insert size of approximately 350 bp were sequenced on the Illumina HiSeq™ 2000 platform at the Biomarker Technologies Company (Beijing, China).

Long-read sequencing data have high error rates: ~ 15% with PacBio sequencing, and as high as ~ 40% with Oxford Nanopore sequencing [45]. We selected PacBio sequencing technology, and libraries for PacBio genome sequencing were constructed following the standard protocols from Pacific Biosciences. In brief, high molecular weight genomic DNA was sheared to a target size of 20 kb, followed by damage repair and end repair, blunt-end adaptor ligation, and size selection. Finally, the libraries were sequenced on the PacBio Sequel platform.

### Quality control of NGS sequencing data

*FastQC* was used to check raw sequencing data in FASTQ format during the first major step of sequence data preprocessing (`fastqc -o yali_fastq yali_1.fq yali_2.fq`). Based on the *FastQC* results, the overall quality values of raw sequence reads were calculated and reported in FASTQ format [46]. *Trimmomatic*, which is a fast, multithreaded command line tool for trimming paired-end and single reads produced by Illumina NGS technology [36], was used to trim reads using the following parameters: `java -jar trimmomatic-0.36.jar PE -phred33 -trimlog logfile yali_1.fq yali_2.fq yali.read_1.fq yali.trim.read_1.fq yali.read_2.fq yali.trim.read_2.fq ILLUMINACLIP: /Trimmomatic/adapters/TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDING-WINDOW:4:15 MINLEN:36`. The output of *Trimmomatic* was in the form of uncompressed filtered FASTQ files.

### Read mapping

Most SV detection software packages using NGS data detect SVs from alignments generated using the BWA aligner. However, the software packages using long-read sequencing data have higher requirements for aligners. Many new aligners have been developed for long-read alignment. NGMLR is recommended because of its better performance compared with other aligners [16]. For NGS data, once the sequences were quality checked and trimmed, the next step was to align the sequences to the ‘Dangshansuli’ pear genome (<http://peargenome.njau.edu.cn/>). Firstly, the reference genome fasta file was indexed using BWA (`bwa index dangshansuli.fasta`). The pair-end sequencing reads of ‘Yali’ were aligned to the

'Dangshansuli' reference genome using the 'align' step (bwa aln -t 20 dangshansuli.fasta yali.read\_1.fq > yali.read\_1.sai; bwa aln -t 20 reference yali.read\_2.fq > yali.read\_2.sai; bwa sampe dangshansuli.fasta yali.read\_1.sai yali.read\_2.sai yali.read\_1.fq yali.read\_2.fq > yali.sam). For PacBio sequencing data, Smrtlink v8.0 was used to filter low quality raw data and processed long reads with accuracies higher than 0.8 (<https://www.pacb.com/support/software-downloads/>). NGMLR was used to map long reads to the 'Dangshansuli' reference genome (ngmlr -t 50 -r dangshansuli.fasta -q yali\_pacbio.fastq -o yali.sam). Both the BWA and NGMLR results were in SAM format. SAMtools was used to convert SAM files into BAM files and to sort the BAM files and remove duplicates [47].

#### Randomly extracting data at different sequencing depths

*Seqtk* can help users to process sequences in FASTA or FASTQ format (<https://github.com/lh3/seqtk>). The command 'seqtk sample' was used to randomly extract a subsample of reads from the clean reads. The short reads were sampled at depths of 10×, 20×, 30×, 40×, 50× and 60×. The long reads were sampled at depths of 5×, 10×, 15×, 20×, 25× and 30×.

#### Description of the four types of SV software

The next step after mapping reads to the reference genome was to identify the SVs from the processed BAM format files. Seven types of SV-calling software using NGS data and two types of SV-calling software using long-read sequencing data, each based on different algorithms, were used to detect the SVs between 'Yali' and 'Dangshansuli' pear.

#### Pindel [25]

Pindel v0.2.5b9 is a C++ application based on the SR algorithm. Running Pindel requires two steps. The first step is 'bam2pindel.pl', the purpose of which is to extract read pairs for use by Pindel (bam2pindel\_bwa.pl -i yali-sortrmdup.bam -o output\_prefix -s yali -om). The second step is 'pindel'; the input files were the 'bam2pindel.pl' file and the reference fasta file (Pindel -f dangshansuli.fasta -p bam2pindel.txt -o output -T 50). By default, Pindel detects all chromosomes if the chromosome region is not specified. Pindel can identify the break points of large deletions (1 bp-10 kb) and medium-sized insertions (1-20 bp) from paired-end short reads. The output file contains the type and size of SV, the chromosome ID, the break point coordinates and the number of reads supporting each event.

#### BreakDancer [33]

BreakDancer-max v1.4.5 is a Perl application and is based on read pairs. The software includes two complementary programs, 'bam2cfg.pl' (bam2cfg.pl -g -h yalisortrmdup.bam >

config\_file) and 'breakdancer-max' (breakdancer-max -lh config\_file > output). The 'bam2cfg.pl' program is aimed at converting the BAM file into the specific file needed for 'breakdancer-max'. The option '-l' was chosen to analyze the Illumina long insert (mate-pair) library. The output contained important information such as the type and size of SV, the chromosome ID, and the SV length. BreakDancer-max had the ability to predict five types of SVs from 'Yali' sequencing data: insertions, deletions, inversions, and inter- and intra-chromosomal translocations.

#### IMR-DENOM [34]

IMR/DENOM v0.4.0 comprises three independent programs, namely 'IMR', 'DENOM', and 'MCMERGE', and is based on assembly. 'IMR' is designed to iterate realignment to the reference genome. In brief, in each iteration, reads are aligned to the reference genome and high-confidence SNPs and InDels are called and incorporated into a new consensus.

Before running the program, the config file was prepared, which contained the path of the output folder, reference, loaddata and settings for the iterations and threads. 'imr easyrun' (imr easyrun -e imr.bam -imrncall configfile) produces the sample BAM file, and the input file was configfile. The option '-e' was set to specify the bam file and '--imrncall' was chosen to only map reads and merge bam files without calling variants. Moreover, '-p' was set to 50 to increase the running speed. Then, 'imr imrcall' (imr imrcall -o imr.sdi -p 50 dangshansuli.fasta imr.bam) was run to detect variants from short reads. The input files were the indexed reference genome fasta file and the specific BAM file. The '-p' for this step was also set to 50.

The command 'denom soapinterface' was used to run DENOM by switching on SOAPdenovo, which creates soap4denom.contig, soap4denom.bam and soap4denom.sdi files (denom soapinterface -o denom.sdi -p 50 configfile). Before running, SOAPdenovo was installed on the server. The input file was configfile. The option '-p' was set to 50 to increase the running speed.

The final step was 'mcmerge dscmp' to merge the imr.sdi and denom.sdi files (mcmerge dscmp -o merge.sdi dangshansuli.fasta imr.sdi denom.sdi). The input files included the indexed reference genome fasta file, the imr.sdi file and the denom.sdi file. The output file was the merge.sdi file, which contained the chromosome ID, position, length, reference base, consensus base and the quality value.

#### Platypus [35]

Platypus v0.8.1 is a Python, Cython and C package, which has the option 'assemble' for detecting SVs based on assembly. It can detect SNPs, insertions, deletions and MNPs. However, it can only detect variants that are less than 10 kb in length. The option 'assembleBadReads'

was set to 1 for using filtered low quality reads for local assembly. The option ‘assembleBrokenPairs’ was set to 1 for using broken read pairs for local assembly (python Platypus.py --refFile dangshansuli.fasta --bamFile yali.bam --nCPU 50 --assemble 1 --assemblerKmerSize 85). The output file was a VCF file containing the chromosome ID, position, length, reference base, consensus base and the quality value.

#### **DELLY [27]**

DELLY v0.7.7 is a C++ application and an integrative program for SV discovery that combines short-range and long-range paired-end mapping and split-read analysis. The command ‘delly call’ was used to discover and genotype SVs. The input files consisted of the indexed reference fasta file and the ‘Yali’ sorted BAM file. The generic option ‘-t’ can be changed to detect other types of SVs. Here, ‘DEL’ (deletion), ‘DUP’ (duplication), ‘INV’ (inversion) and ‘INS’ (insertion) were selected to detect different types of SVs (delly call -t DEL -g dangshansuli.fasta yalisortrmdup.bam). The output was in BCF format. Then, Bcftools was used to convert BCF format into VCF format. The output file contained the SV type, the chromosome ID, the SV position, the reference sequence, the alteration, the quality, the filter and other SV information.

#### **Lumpy [28]**

Lumpy v0.2.13 is a C++ software package that integrates multiple SV signals, such as split reads, read pairs and read depth. ‘lumpyexpress’ in this package was used to detect SVs for standard analyses. Before detecting SVs, the BAM files \*.bam, \*.splitters.bam and \*.discordants.bam were obtained using the BWA-MEM alignment in speedseq [48] (speedseq align -R “@RG\tID:id\tSM:sample\tLB:lib” dangshansuli.fasta yali.read\_1.fq yali.read\_2.fq). The output contained important information such as the type and size of SV, the chromosome ID, and the SV length (lumpyexpress -B yali.bam -S yali\_splitters.bam -D yali.discordants.bam -o yali.output). Lumpy can predict four types of SVs: deletions, inversions, duplications and translocations.

#### **MetaSV [30]**

MetaSV v0.5.2 is a Python package that uses multiple algorithms: Pindel [25], BreakDancer [33], and Lumpy [28]. The reference, the BAM file, and the outputs of Pindel, BreakDancer and Lumpy were regarded as the input files for MetaSV (run metasv.py --reference dangshansuli.fasta --breakdancer\_native breakdancer.out --pindel\_native pindel\_D pindel\_SI pindel\_TD pindel\_INV --lumpy\_vcf yali.vcf --bam yali.bam --outdir. /out --disable\_assembly). We only merged the outputs of three SV detectors without performing further soft-clip based analysis or local assembly.

#### **Sniffles [16]**

Sniffles v1.0.11 is a C++ software package that detects SVs based on the split-read algorithm. Sniffles can detect SVs using PacBio or Oxford Nanopore sequencing data. The input file was the BAM file generated from the BWA (‘bwa mem -x’) or NGMLR aligner. Default parameters were used (sniffles -m yalisortrmdup.bam -v yali.vcf). The output file was in ‘vcf’ format and included chromosome, position, SV type, quality and other information.

#### **SVIM [32]**

SVIM v1.2.0 is a Python package that can detect five types of SVs (deletions, insertions, inversions, break ends, and duplications) using long reads from PacBio or Oxford Nanopore sequencing technology. The input file can be a fastq file or BAM file. To obtain the output file faster, we used the BAM file generated from the NGMLR aligner and ran the command (svim alignment yalisortrmdup.bam dangshansuli.fasta). The output directory ‘candidates’ contained BED format files of each SV type, and each BED file included the chromosome, start coordinate, end coordinate, SV type, score and other information.

#### **GO and KEGG analysis**

Based on the GO annotation information for all genes in the pear genome, WEGO (<http://wego.genomics.org.cn/>) was used to perform GO analysis. KEGG analysis of genes was done using the website <http://www.genome.jp/kegg/>.

#### **Computing resources**

Trimming, mapping and SV detection were performed on a server machine equipped with four 2.4GHz Intel® Xeon® 6 CPUs, with 18 cores within each CPU, and 2 TB of RAM. The operating system was CentOS release 6.8.

#### **Validation of SVs detected using NGS data and long-read sequencing data**

The SV calling tools are designed to maximize sensitivity. For example, to identify as many variants as possible in our data, we allowed many false positive calls due to artifacts from the sequencing process. Therefore, before validating the SVs, the SVs called by different software needed to be filtered. For Pindel, we selected the SVs with ‘Support > 5’. For DELLY, Lumpy and MetaSV, SVs with no ‘LowQual’ were selected, and for BreakDancer, SVs with scores over 60 were selected. For SVIM, SVs with a quality of ‘PASS’ were selected, and for IMR/DENOM, we removed the SNPs. Next, the IGV was used to visualize the SVs [49–51]. Specifically, the reference ‘Dangshansuli’ pear genome was loaded as the reference genome sequence, and the BAM file of ‘Yali’ was loaded to visually confirm the presence of the identified deletions and insertions.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-6455-x>.

**Additional file 1.** GO analysis of 264 genes within SVs commonly identified by five SV callers.

**Additional file 2.** KEGG analysis of 264 genes within SVs commonly identified by five SV callers.

**Additional file 3: Figure S1.** The number of genes within SVs detected by software packages using NGS data (a) and long-read sequencing data (b).

**Additional file 4.** GO analysis of 403 genes within SVs commonly identified by three SV callers using NGS data.

**Additional file 5.** KEGG analysis of 403 genes within SVs commonly identified by three SV callers using NGS data.

**Additional file 6.** GO analysis of 4495 genes within SVs commonly identified by two SV callers using long-read sequencing data.

**Additional file 7.** KEGG analysis of 4495 genes within SVs commonly identified by two SV callers using long-read sequencing data.

**Additional file 8.** Verification of SVs in 'Yali' through comparisons with the 'Dangshansuli' reference genome.

## Abbreviations

BAM file: Binary version of SAM file; BCF: Binary Counterpart Data; GO: Gene Ontology; InDel: Insertion and deletion; KEGG: Kyoto Encyclopedia of Genes and Genomes; MNP: Multiple nucleotide polymorphism; NGS: Next-generation sequencing; SAM: Sequence alignment/map format; SNP: Single nucleotide polymorphism; SV: Structural variant; VCF: Variant Call Format

## Acknowledgements

We gratefully thank the Changli Fruit Research Institute, Hebei Academy of Agricultural Sciences, China for providing the 'Yali' pear as experimental material.

## Authors' contributions

JW designed this project. YYL and MYZ wrote the manuscript. YYL, MYZ, WJC, and MYS performed data analysis. JYS provided the experimental materials. All authors read and approved the final manuscript.

## Funding

Funding from the National Science Foundation of China (31725024 and 31672111) supported the next-generation sequencing of 'Yali' pear, and funding from the Earmarked Fund for the China Agriculture Research System (CARS-28) supported the long-read sequencing of 'Yali' pear. DNA extraction and library preparation for 'Yali' were performed using funding from the "333 High Level Talents Project" of Jiangsu Province (BRA2016367).

## Availability of data and materials

All raw sequence data generated in this study are deposited in NCBI under BioProject accession number: PRJNA574796. All other supporting data are included as additional files.

## Ethics approval and consent to participate

The 'Yali' pear plant samples were obtained from the Changli Fruit Research Institute, Hebei Academy of Agricultural Sciences, China. Since these studies did not involve endangered or protected species, no specific permissions were required for this material. The authors declare that the experimental research on plants described in this paper complied with institutional and national guidelines.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 17 March 2019 Accepted: 7 January 2020

Published online: 20 January 2020

## References

- Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet.* 2011;12(5):363–76.
- Guan P, Sung WK. Structural variation detection using next-generation sequencing data a comparative technical review. *Methods.* 2016;102:36–49.
- Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature.* 2009;462(7276):1005–U1060.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MHY, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015;526(7571):75.
- Ogawa S. Novel mechanism of immune evasion in cancer via structural variations of the PD-L1 gene. *Rinsho Ketsueki.* 2017;58(8):957–65.
- Fujiwara K, Matsuura K, Matsunami K, Iio E, Nojiri S, Joh T. Novel non-canonical genetic rearrangements termed "complex structural variations" in HBV genome. *Hepatology.* 2017;66:805a.
- Haas J, Mester S, Lai A, Frese KS, Sedaghat-Hamedani F, Kayvanpour E, Rausch T, Nietsch R, Boeckel JN, Carstensen A, et al. Genomic structural variations lead to dysregulation of important coding and non-coding RNA species in dilated cardiomyopathy. *Embo Mol Med.* 2018;10(1):107–20.
- Zmienko A, Samelak A, Kozłowski P, Figlerowicz M. Copy number polymorphism in plant genomes. *Theor Appl Genet.* 2014;127(1):1–18.
- Marroni F, Pinosio S, Morgante M. Structural variation and genome complexity: is dispensable really dispensable? *Curr Opin Plant Biol.* 2014;18:31–6.
- Zhang ZH, Mao LY, Chen HM, Bu FJ, Li GC, Sun JJ, Li S, Sun HH, Jiao C, Blakely R, et al. Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell.* 2015;27(6):1595–604.
- Causse M, Desplat N, Pascual L, Le Paslier MC, Sauvage C, Bauchet G, Berard A, Bounon R, Tchoumakov M, Brunel D, et al. Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics.* 2013;14:791.
- Zhang SJ, Chen WP, Xin L, Gao ZH, Hou YJ, Yu XY, Zhang Z, Qu SC. Genomic variants of genes associated with three horticultural traits in apple revealed by genome re-sequencing. *Hortic Res.* 2014;1:14045.
- Wu J, Wang ZW, Shi ZB, Zhang S, Ming R, Zhu SL, Khan MA, Tao ST, Korban SS, Wang H, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 2013;23(2):396–408.
- Abel HJ, Duncavage EJ. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer Genet.* 2013;206(12):432–40.
- Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics.* 2012;28(3):311–7.
- Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods.* 2018;15(6):461.
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333–51.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008;456(7218):53–9.
- Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods.* 2009;6(11 Suppl):S13–20.
- Simpson JT, McIntyre RE, Adams DJ, Durbin R. Copy number variant detection in inbred strains from short read sequence data. *Bioinformatics.* 2010;26(4):565–7.
- Korbel JO, Abyzov A, Mu XJ, Carriero N, Cayting P, Zhang ZD, Snyder M, Gerstein MB. PEmr: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol.* 2009;10(2):R23.



22. Zeitouni B, Boeva V, Janoueix-Lerosey I, Loeillet S, Legoix-ne P, Nicolas A, Delattre O, Barillot E. SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics*. 2010;26(15):1895–6.
23. Marschall T, Costa IG, Canzar S, Bauer M, Klau GW, Schliep A, Schonhuth A. CLEVER: clique-enumerating variant finder. *Bioinformatics*. 2012;28(22):2875–82.
24. Emde AK, Schulz MH, Weese D, Sun RP, Vingron M, Kalscheuer VM, Haas SA, Reinert K. Detecting genomic indel variants with exact breakpoints in single- and paired-end sequencing data using SplazerS. *Bioinformatics*. 2012;28(5):619–27.
25. Ye K, Schulz MH, Long Q, Apweiler R, Ning ZM. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865–71.
26. Lin K, Smit S, Bonnema G, Sanchez-Perez G, de Ridder D. Making the difference: integrating structural variation detection tools. *Brief Bioinform*. 2015;16(5):852–64.
27. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):333–9.
28. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014;15(6):R84.
29. Becker T, Lee WP, Leone J, Zhu QH, Zhang CS, Liu S, Sargent J, Shanker K, Mil-Homens A, Cerveira E, et al. FusorSV: an algorithm for optimally combining data from multiple structural variation detection methods. *Genome Biol*. 2018;19:38.
30. Mohiyuddin M, Mu JC, Li J, Asadi NB, Gerstein MB, Abyzov A, Wong WH, Lam HYK. MetaSV: an accurate and integrative structural-variant caller for next generation sequencing. *Bioinformatics*. 2015;31(16):2741–4.
31. Kosugi S, Momozawa Y, Liu XX, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol*. 2019;20:117.
32. Heller D, Vingron M. SVM: structural variant identification using mapped Long reads. *Bioinformatics*. 2019;35:2907–15.
33. Fan X, Abbott TE, Larson D, Chen K. BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping. *Curr Protoc Bioinformatics*. 2014;45:15.16.11.
34. Gan XC, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*. 2011;477(7365):419–23.
35. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, McVean G, Lunter G, Consortium W. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet*. 2014;46(8):912–8.
36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
37. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
38. Gordon SP, Priest H, Marais DLD, Schackwitz W, Figueroa M, Martin J, Bragg JN, Tyler L, Lee CR, Bryant D, et al. Genome diversity in *Brachypodium distachyon*: deep sequencing of highly diverse inbred lines. *Plant J*. 2014;79(3):361–74.
39. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470(7332):59–65.
40. Wang WS, Mauleon R, Hu ZQ, Chebotarov D, Tai SS, Wu ZC, Li M, Zheng TQ, Fuentes RR, Zhang F, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 2018;557(7703):43.
41. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang TY, Li Y, Li YX, Semagn K, Zhang XC, et al. High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat Commun*. 2015;6:6914.
42. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*. 2014;15(2):121–32.
43. Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, et al. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res*. 2009;19(9):1622–9.
44. Wang J, Wang W, Li RQ, Li YR, Tian G, Goodman L, Fan W, Zhang JQ, Li J, Zhang JB, et al. The diploid genome sequence of an Asian individual. *Nature*. 2008;456(7218):60–U61.
45. Ye CX, Hill CM, Wu SG, Ruan J, Ma ZS. DBG2OLC: efficient assembly of large genomes using Long erroneous reads of the third generation sequencing technologies. *Sci Rep*. 2016;6:31900.
46. Kim T, Seo HD, Hennighausen L, Lee D, Kang K. Octopus-toolkit: a workflow to automate mining of public epigenomic and transcriptomic next-generation sequencing data. *Nucleic Acids Res*. 2018;46:e53.
47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
48. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan AR, Hall IM. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods*. 2015;12(10):966–8.
49. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–6.
50. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bahler J, Sedlazeck FJ. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun*. 2017;8:14061.
51. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



RESEARCH ARTICLE

Open Access



# Contrasting genetic variation and positive selection followed the divergence of NBS-encoding genes in Asian and European pears

Manyi Sun<sup>1†</sup>, Mingyue Zhang<sup>1†</sup>, Jugpreet Singh<sup>2</sup>, Bobo Song<sup>1</sup>, Zikai Tang<sup>1</sup>, Yueyuan Liu<sup>1</sup>, Runze Wang<sup>1</sup>, Mengfan Qin<sup>1</sup>, Jiaming Li<sup>1</sup>, Awais Khan<sup>2\*</sup> and Jun Wu<sup>1\*</sup>

## Abstract

**Background:** The NBS disease-related gene family coordinates the inherent immune system in plants in response to pathogen infections. Previous studies have identified NBS-encoding genes in *Pyrus bretschneideri* ('Dangshansuli', an Asian pear) and *Pyrus communis* ('Bartlett', a European pear) genomes, but the patterns of genetic variation and selection pressure on these genes during pear domestication have remained unsolved.

**Results:** In this study, 338 and 412 NBS-encoding genes were identified from Asian and European pear genomes. This difference between the two pear species was the result of proximal duplications. About 15.79% orthologous gene pairs had Ka/Ks ratio more than one, indicating two pear species undergo strong positive selection after the divergence of Asian and European pear. We identified 21 and 15 NBS-encoding genes under fire blight and black spot disease-related QTL, respectively, suggesting their importance in disease resistance. Domestication caused decreased nucleotide diversity across NBS genes in Asian cultivars (cultivated 6.23E-03; wild 6.47E-03), but opposite trend (cultivated 6.48E-03; wild 5.91E-03) appeared in European pears. Many NBS-encoding coding regions showed Ka/Ks ratio of greater than 1, indicating the role of positive selection in shaping diversity of NBS-encoding genes in pear. Furthermore, we detected 295 and 122 significantly different SNPs between wild and domesticated accessions in Asian and European pear populations. Two NBS genes (*Pbr025269.1* and *Pbr019876.1*) with significantly different SNPs showed >5x upregulation between wild and cultivated pear accessions, and > 2x upregulation in *Pyrus calleryana* after inoculation with *Alternaria alternata*. We propose that positively selected and significantly different SNPs of an NBS-encoding gene (*Pbr025269.1*) regulate gene expression differences in the wild and cultivated groups, which may affect resistance in pear against *A. alternata*.

(Continued on next page)

\* Correspondence: [awais.khan@cornell.edu](mailto:awais.khan@cornell.edu); [wujun@njau.edu.cn](mailto:wujun@njau.edu.cn)

<sup>†</sup>Manyi Sun and Mingyue Zhang contributed equally to this work.

<sup>2</sup>Plant Pathology & Plant-Microbe Biology Section, Cornell University, Geneva, NY 14456, USA

<sup>1</sup>College of Horticulture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusion:** Proximal duplication mainly led to the different number of NBS-encoding genes in *P. bretschneideri* and *P. communis* genomes. The patterns of genetic diversity and positive selection pressure differed between Asian and European pear populations, most likely due to their independent domestication events. This analysis helps us understand the evolution, diversity, and selection pressure in the NBS-encoding gene family in Asian and European populations, and provides opportunities to study mechanisms of disease resistance in pear.

**Keywords:** NBS, Pear, Expansion, Positive selection, Nucleotide diversity

## Background

Pear (*Pyrus*), the third most important fruit tree species in the world, has been cultivated for more than 3000 years and is one of the oldest fruit trees in the world [1]. Due to independent domestication under distinct geographical conditions, Asian and European pears display prominent differences in genetic and phenotypic diversity [2, 3]. The wild populations of these two pear groups have likely experienced unique disease pressures due to their completely different habitats. For example, brown spot (caused by fungus *Stemphylium vesicarium*) and fire blight (caused by bacteria *Erwinia amylovora*) diseases lead to death of pear trees and threaten the pear industry in Europe and west Asia [4–10]. Black spot (caused by fungus *A. alternata*), scab (caused by fungus *Venturia nashicola*), ring rot (caused by fungus *Botryosphaeria berengeriana*), and bitter rot (caused by bacteria *Colletotrichum fructicola*) diseases, in turn, cause huge losses to pear production in China and east Asia [11–14]. Furthermore, independent domestications of Asian and European pear [3] might have intensified the selection on a few genomic regions that have relevance for the specific diseases associated with their cultivation habitats. Hence, characterizing disease-related gene families can provide opportunities to understand the role of divergence and domestication in shaping host resistance responses in pear.

Studying disease resistance mechanisms in wild and cultivated germplasm remains a major focus of pear breeding and improvement [13, 15–20]. For example, Pierantoni L et al. [15] identified two major resistance QTL against scab disease caused by *Venturia pirina* in European pear (*P. communis*), and the causative scab disease gene 'Vnk' was found in Asian pear [16]. Similarly, QTL have also been identified for resistance to fire blight and black spot disease [21] in *Pyrus ussuriensis* [17] and *P. communis* [18]. Interestingly, the QTL region for black spot disease resistance contained two NBS-encoding genes [13], which were up-regulated after infection with *A. alternata*, the pathogen causing the black spot disease in *P. calleryana* [20]. These results highlight the importance of studying disease-related gene families, their genetic diversity, and the selection pressure on them in both Asian and European pear. Other pear species also show different levels of

disease resistance. For instance, *P. calleryana* is a wild species native to China with strong disease resistance that has been used for fire blight resistance in the US [20]. A genetic source of fire blight resistance was detected from *P. ussuriensis* and *P. pyrifolia*, which have been used to breed disease resistance pear cultivars like 'Harrow Sweet' and 'Moonglow' [22]. Furthermore, monogenic sources of scab resistance were detected in some Asian pear accessions including 'Kinchaku', 'Cangxili' and 'Hongli' [21]. These QTL, disease-resistance genes, and germplasm resources are potentially useful in understanding disease resistance mechanisms and breeding new disease-resistant cultivars in pear.

Plant disease resistance (R) genes play an important role in defense against pathogens and are classified based on their location in plant cells and their putative protein domain [23]. The nucleotide binding site plus leucine-rich repeat genes (NBS-LRR) represents the biggest class of R genes [24]. These genes mainly participate in the tracking and response of various pathogens including bacteria, viruses, fungi, nematodes and oomycetes [24, 25]. Recently, a range of NBS-encoding genes have been identified in different rosaceae species. For instance, the number of NBS-encoding genes varies from 346 to 1303 in strawberry and apple, respectively [26]. However, these numbers were different in other similar studies [27–29], probably due to different computational criteria used for their detection. It has been postulated that gene loss and expansion within species has caused the variation in number of NBS genes, and repeated duplication, divergence and eventual loss has facilitated their evolution in response to the rapidly changing pathogens [30]. In addition, NBS-encoding genes show a strong signature of selection, but the nature of selection differs across these genes [31]. Some NBS-encoding genes evolve rapidly through copy number variation and high ratio of non-synonymous to synonymous substitutions, but other genes evolve at a slower rate [31].

With the availability of new high-quality assembled genomes of double haploid *P. communis* [32] and *P. bretschneideri* [1], an accurate characterization of NBS-encoding genes and their comparative analysis with *P. bretschneideri* is feasible. In addition, whole-genome resequencing data from a large collection of pear accessions [3] also provides

a massive genomic resource to study diversification and selection effects in pear. Since the selection and evolution of NBS-encoding genes is yet unclear, the genome sequence information provides chances to explore their diversity and evolution in wild and cultivated groups in Asian and European pear.

In this study, we identified NBS-encoding genes and compared them across the genomes of *P. bretschneideri* and *P. communis*. We further used re-sequencing data from 131 pear accessions to characterize genetic variation and selection signatures on NBS-encoding genes in wild and domesticated Asian and European pear populations. In addition, NBS-encoding genes in disease resistance QTL were identified in Asian and European pears. These findings will provide additional resources for future research of NBS-encoding gene function and disease resistance in pear.

## Result

### Identification and phylogenetic analysis of NBS-encoding genes in *P. bretschneideri* and *P. communis*

In this study, a total of 338 and 412 NBS-encoding genes were identified in the ~510 Mb *P. bretschneideri* (Asian pear cultivar: 'Dangshansuli') genome, and the ~497 Mb *P. communis* genome (European pear cultivar: 'Bartlett'), respectively. The NBS-encoding genes were divided into six types including CC-NBS-LRR, TIR-NBS-LRR and four truncated NBS-LRR (Table 1), three of which (TIR-NBS-LRR, CC-NBS-LRR, NBS) determine the majority of the difference in the number of NBS-encoding genes between Asian and European pears. The NBS-LRR class was most frequent in both *P. bretschneideri* (36.4%) and *P. communis* (25.7%). However, two different classes,

CC-NBS-LRRs and NBS, represented the second largest classes in *P. bretschneideri* (26.6%) and *P. communis* (24.0%), respectively. The TIR-NBS (6.21%), CC-NBS (9.46%), NBS (10.36%), and TIR-NBS-LRR (10.95%) were least frequent in *P. bretschneideri*, while CC-NBS (7.04%), CC-NBS-LRR (9.22%), and TIR-NBS (13.35%) were smallest classes in *P. communis* (Table 1). The percentages of LRR domain-containing NBS genes were approximately 74.0 and 55.6% in *P. bretschneideri* and *P. communis*, respectively. However, the non-LRR containing genes showed a clear difference between Asian ( $n = 88$ ; 26.04%) and European ( $n = 183$ ; 44.41%) pear genomes (Table 1).

Phylogenetic analysis revealed that most non-TIR-NBS type genes and TIR-NBS genes were divided into two clear groups (Fig. 1), and this phenomenon remained consistent in the separate analysis of *P. bretschneideri* and *P. communis* (Additional file 1, Figure S1, S2). This analysis further showed that most NB-ARC domains were located at similar positions on the genes, with lengths ranging from 250 to 300 bp. The position, number, and length of LRR domains were comparatively more variable in NBS-encoding genes. In addition, a large number of clades specific to *P. bretschneideri* and *P. communis* were displayed on the phylogenetic tree. Many species-specific genes and the differences in the numbers of NBS-encoding genes might have appeared after the divergence of Asian and European pears [3].

Promoter region (~1500 bp) analysis identified 128 different kinds of cis-elements in the NBS-encoding gene families of *P. bretschneideri* and *P. communis*, with 16 cis-elements having frequency more than 40% among them (Table 2). The CAAT-box, TATA-box, and MYC cis-elements showed highest enrichment in the promoters of NBS-encoding genes of both *P. bretschneideri* and *P. communis*. These 16 cis-elements were associated with response to various environmental factors (light, temperature, water and anaerobic stresses) and plant hormones (jasmonic acid, methyl ester, ethylene and salicylic acid). Interestingly, five types of cis-elements only appeared in *P. communis*; two of them were H-box and JERE, which regulate defense and jasmonate-responsive genes.

### Genome-wide distribution and duplication of NBS-encoding genes

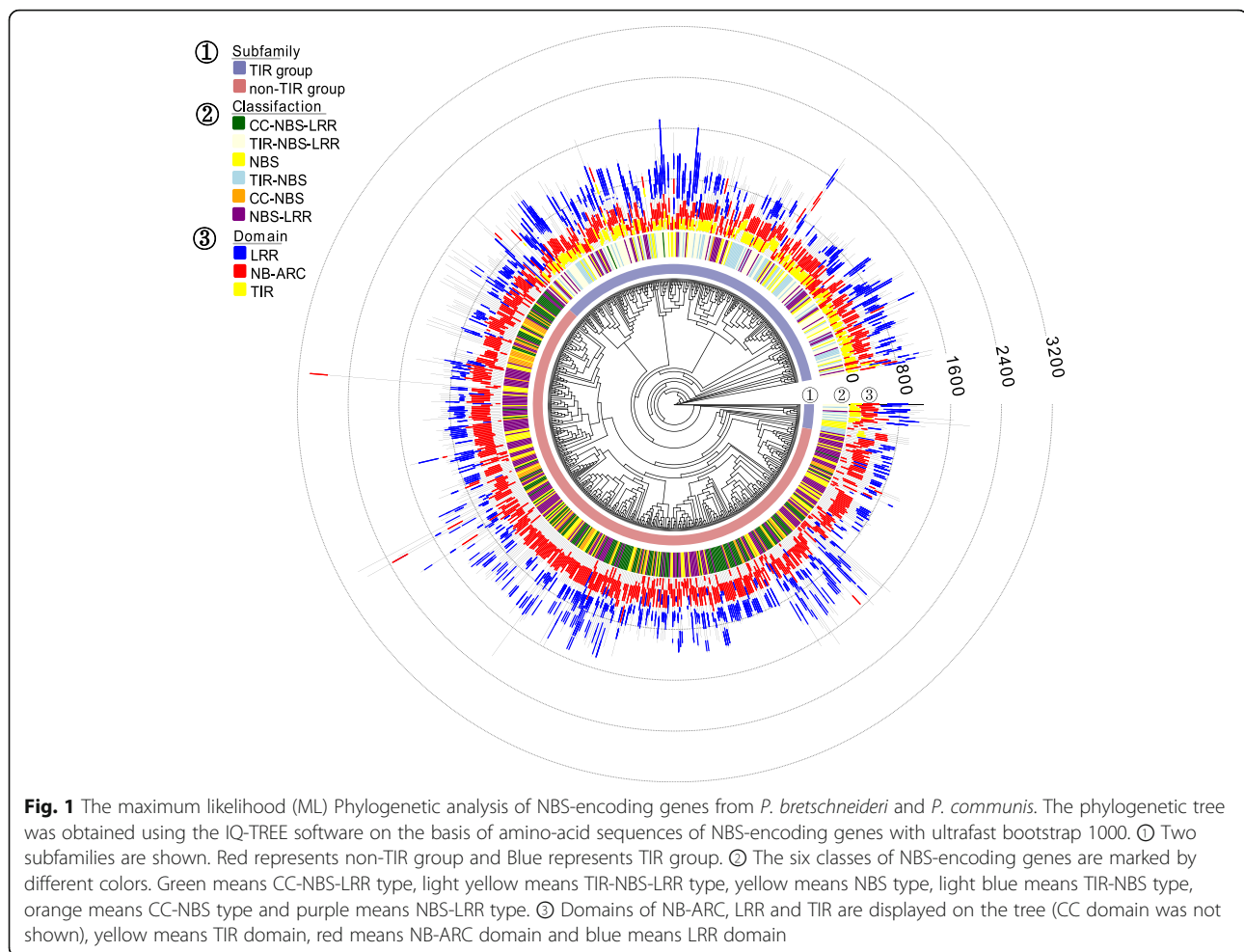
In *P. bretschneideri*, a total of 265 NBS-encoding genes were mapped to the 17 chromosomes and the remaining 73 genes were placed on scaffolds (Additional file 2). In the mapped genes, the highest number, 151 genes (56.98%), were located on four chromosomes; Chr2 (37), Chr5 (52), Chr7 (59), and Chr11 (50) (Fig. 2a). In *P. communis*, 408 NBS-encoding genes were mapped onto the 17 chromosomes, while 4 genes were on the

**Table 1** Classification of NBS-encoding genes in 'Dangshansuli' and 'Bartlett'

| Type <sup>a</sup>                    | Numbers        |            |
|--------------------------------------|----------------|------------|
|                                      | 'Dangshansuli' | 'Bartlett' |
| CC -NBS-LRR                          | 90             | 38         |
| TIR- NBS- LRR                        | 37             | 85         |
| <b>Truncated NBS-LRR<sup>b</sup></b> |                |            |
| NBS-LRR                              | 123            | 106        |
| TIR- NBS                             | 211            | 55         |
| CC-NBS                               | 32             | 29         |
| NBS                                  | 35             | 99         |
| <b>Total NBS</b>                     | <b>338</b>     | <b>412</b> |
| Total NBS with LRR                   | 250            | 229        |
| Total NBS without LRR                | 88             | 183        |

<sup>a</sup>Based on the presence or absence of CC, TIR, NBS and LRR domains, the NBS genes were classified into CC-NBS-LRR, TIR-NBS-LRR and four Truncated NBS-LRR types. <sup>b</sup>Truncated NBS-LRR means NBS genes lacking N-terminal domain (TIR or CC) or C-terminal domain (LRR), and classified into NBS-LRR, TIR-NBS, CC-NBS, NBS





unplaced scaffolds. As in *P. bretschneideri*, four chromosomes, Chr2 (52), Chr5 (42), Chr7 (58), Chr11 (41), had the highest proportion of the total NBS-encoding genes (Fig. 2b). A set of 40.38% (21/52) NBS-encoding genes on Chr2 co-localized with the previously detected fire blight disease resistance QTL of *P. communis* (Additional file 3) [13, 16, 17, 22, 33–35]. In addition, 30% (15/50) NBS-encoding genes on Chr11 were located within the black spot disease resistance QTL of *P. bretschneideri* (Additional file 3).

To detect the potential mechanism of expansion, we analyzed the duplication events of NBS-encoding genes in the *P. communis* and *P. bretschneideri* genomes. Each member of the NBS-encoding gene family was divided into five different categories: singleton, WGD or segmental, tandem, proximal, or dispersed. We found that 21,493 (50.20%) genes in the *P. bretschneideri* genome and 20,250 (54.08%) genes in the *P. communis* genome (data not shown) were primarily derived from WGD or segmental duplications. However, only 23.96% of NBS-encoding genes in *P. bretschneideri* and 19.90% in *P. communis* were duplicated and retained from WGD

events. These results indicated that the percentage of NBS-encoding genes derived from dispersed and proximal duplication were 27.81 and 28.11% in *P. bretschneideri*, and 20.15 and 40.05% in *P. communis*, respectively (Additional file 4).

#### Time of duplication events and selection pressure in NBS-encoding genes

By estimating evolutionary dates with the synonymous substitution rate per site (Ks), we identified that pear genome had undergone two WGD events [1]; an ancient WGD (Ks ~ 1.5–1.8) that occurred about 140 MYA ago [36], and a recent WGD (Ks ~ 0.15–0.3) event that may have occurred 30–45 MYA ago [1]. In *P. bretschneideri*, a total of 60 duplicated NBS-encoding gene pairs located on the collinear blocks were identified and, out of them, Ks values of 16 gene pairs (26.66%) ranged from 0.15–0.30 (Fig. 3a) [1]. As such, *P. communis* had 61 duplicated NBS-encoding gene pairs (Fig. 3a) and Ks values of 13 (21.31%) of these genes ranged from 0.15–0.30. The duplicated gene pairs derived from ancient WGD events have been lost in *P. bretschneideri*, but three

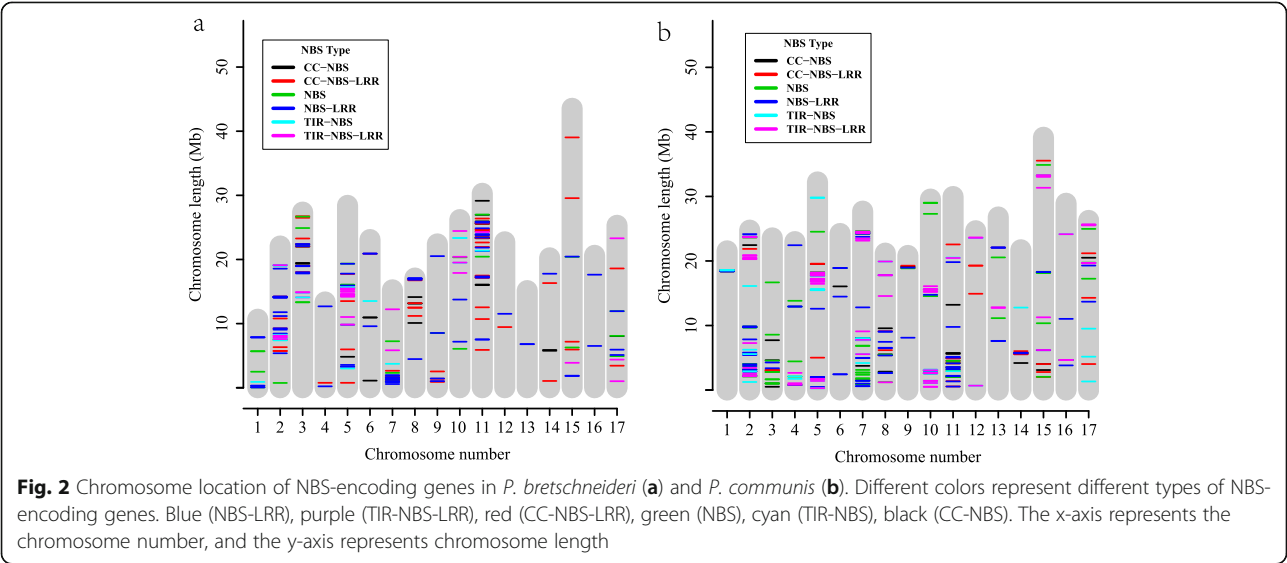
**Table 2** Frequency of cis-elements identified in the promoter regions of NBS-encoding genes in two pear species

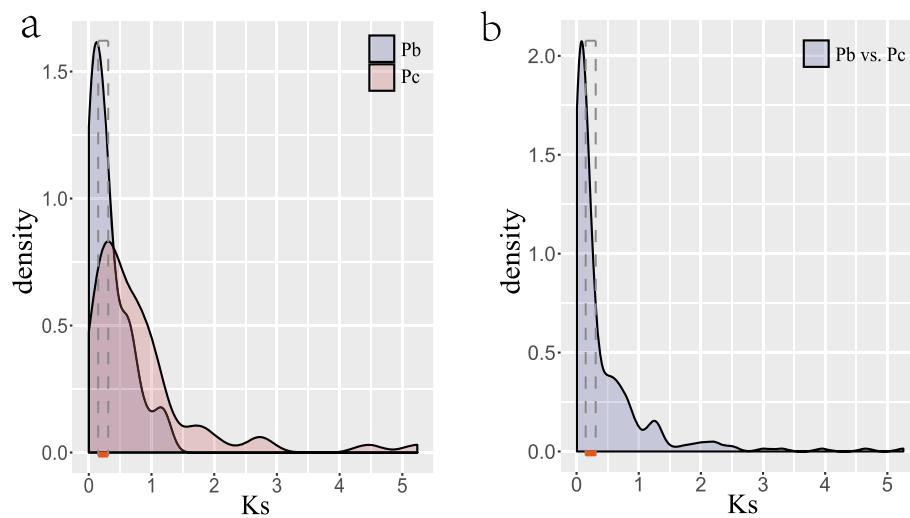
| Cis-element name | Sequence  | Putative response           | Frequency <sup>a</sup> |            |
|------------------|-----------|-----------------------------|------------------------|------------|
|                  |           |                             | 'Dangshansuli'         | 'Bartlett' |
| CAAT-box         | CCAAT     | Seed                        | 1.00                   | 1.00       |
| TATA-box         | TATAA     | core promoter element       | 1.00                   | 1.00       |
| MYC              | CATTTG    | Cold                        | 0.95                   | 0.96       |
| ARE              | AAACCA    | anaerobium                  | 0.75                   | 0.74       |
| STRE             | AGGGG     | Stress                      | 0.72                   | 0.73       |
| ABRE             | ACGTG     | ABA                         | 0.69                   | 0.74       |
| Box 4            | ATTAAT    | Light                       | 0.69                   | 0.72       |
| CGTCA-motif      | CGTCA     | MeJA                        | 0.67                   | 0.75       |
| TGACG-motif      | TGACG     | MeJA                        | 0.67                   | 0.75       |
| TCT-motif        | TCTTAC    | Light                       | 0.48                   | 0.51       |
| GT1-motif        | GGTTAA    | Light                       | 0.49                   | 0.46       |
| MBS              | CAACTG    | Drought                     | 0.49                   | 0.52       |
| ERE              | ATTTCATA  | Ethylene                    | 0.46                   | 0.49       |
| W box            | TTGACC    | phytochrome down-regulation | 0.43                   | 0.45       |
| LTR              | CCGAAA    | low-temperature             | 0.41                   | 0.46       |
| TCA-element      | CCATCTTTT | salicylic acid              | 0.39                   | 0.49       |

<sup>a</sup>Frequency = Genes (containing this cis-element)/All NBS-encoding genes in Dangshansuli (Bartlett). Cis-elements on upstream 1500 bp sequence of NBS genes were identified by Plant CARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) program. Genome sequences of Dangshansuli and Bartlett were downloaded from NJAU (<http://peargenome.njau.edu.cn/>) and phytozome database (<https://phytozome.jgi.doe.gov/>)

duplicated gene pairs ( $K_s \sim 1.7$ ) were retained in *P. communis*. We further estimated the ratio of non-synonymous ( $K_a$ ) and synonymous substitutions per site ( $K_s$ ) to investigate the role of positive and purifying selection in the evolution of NBS-encoding genes. For a total of 60 duplicated gene pairs in *P. bretschneideri*, the  $K_a/K_s$  ratio of 3 pairs was more than 1; the  $K_a/K_s$  ratio of the other 57 gene pairs were less than 1 (Additional file 5).

Likewise, out of total 61 duplicated pairs in *P. communis*, only 1 pair had  $K_a/K_s$  ratios > 1 while the remaining 60 pairs have  $K_a/K_s < 1$ . A total of 266 orthologous gene pairs were identified between *P. bretschneideri* and *P. communis* (Additional file 6), and most orthologous gene pairs were present on the homologous chromosomes of two pear species (Fig. 4).  $K_s$  values of these gene pairs ranged from 0.006 to 5.25 (Fig. 3b). In addition, the  $K_a/K_s$





**Fig. 3** Distribution of Ks values of NBS genes pairs. Orange lines at the x-axis and the gray dotted box represented the Ks value region from 0.15 to 0.30. **a** Distribution of Ks values in *P. bretschneideri* and *P. communis*, respectively **b** Distribution of Ks values of Orthologous gene pairs between two pear species. Letter 'Pb' means *P. bretschneideri*, and letter 'Pc' means *P. communis*. Ks represents synonymous substitutions per site

values of 266 orthologous gene pairs ranged from 0.15–2.51 (Additional file 7). A set of 42 orthologous gene pairs had Ka/Ks value > 1. These results indicated that about 35.21% NBS-encoding genes of *P. bretschneideri* had lost their orthologous genes in *P. communis*, and 17.79% orthologous gene pairs underwent positive selection.

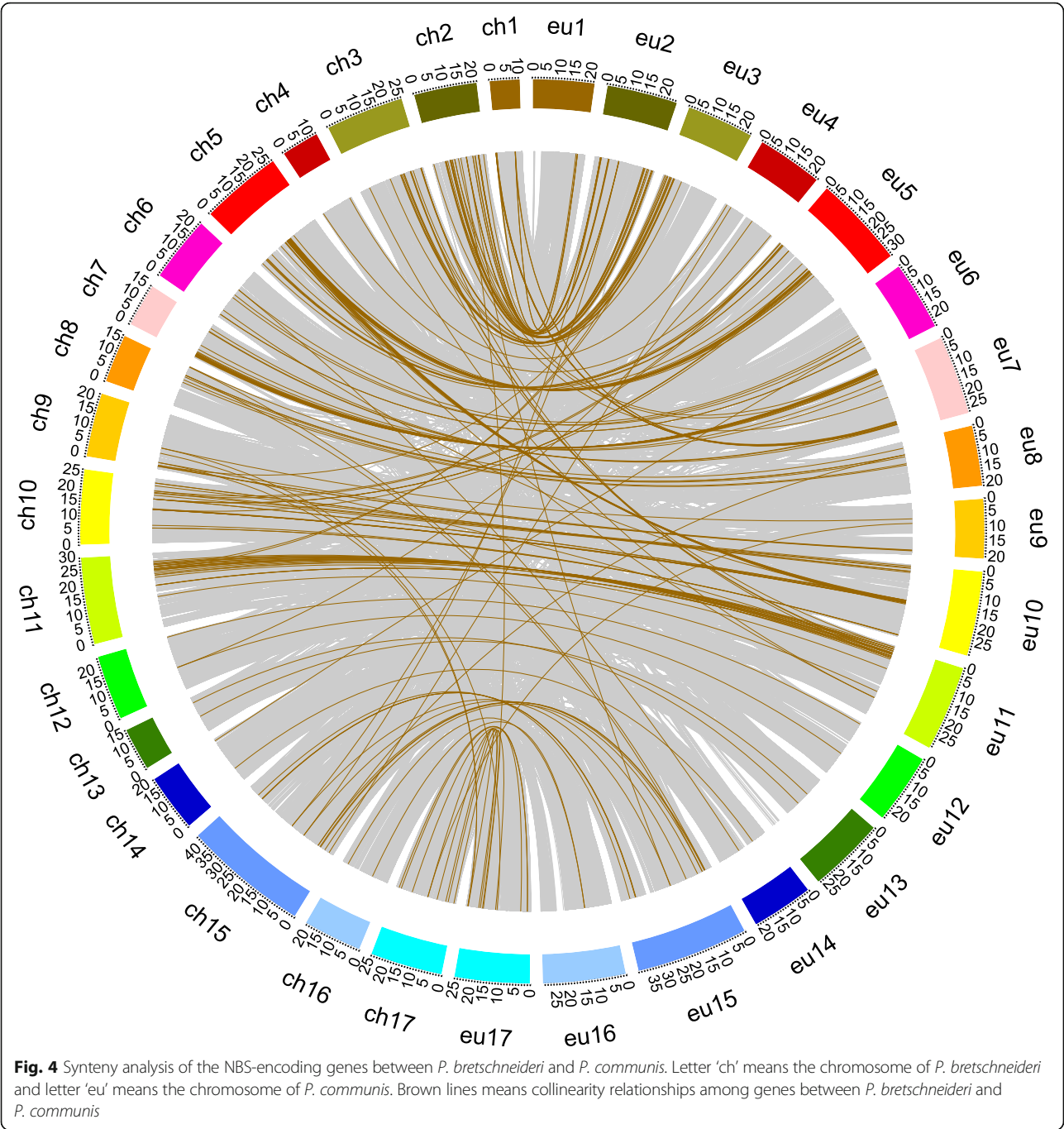
#### Genetic diversity and selection across NBS-encoding genes in Asian and European pear

An analysis of genetic diversity using genome sequences from 131 Asian and European pear accessions (including 61 wild and 70 cultivated accessions) revealed a total of 50,682 SNPs in NBS-encoding genomic regions, out of which 22,472 SNPs were present in the coding regions. After removing heterozygous SNPs, we obtained 7654 synonymous and 14,481 non-synonymous SNPs, and a non-synonymous/synonymous ratio of 1.89.

The average nucleotide diversity across NBS-encoding genes was 6.02E-03 and 5.65E-03 in the Asian and European populations, and 5.92E-03 and 6.88E-03 in the wild and cultivated groups (Additional file 8), respectively. A similar trend was observed when analysis was repeated using only the coding regions from NBS-encoding genes (Table 3). In general, nucleotide diversity had a wide distribution ranging from 0.02 to 2.19E-05 in Asian and 0.01 to 2.48E-05 in the European pears, respectively (Fig. 5a; Additional file 8). A comparison of cultivated and wild populations revealed that average nucleotide diversity across NBS-encoding genes reduced after domestication in Asian pears, but contrasting results were observed in the European population (Table 3). However, individual NBS gene coding regions showed varied

patterns of genetic diversity in wild and cultivated groups in both Asian and European pear populations. About 20% NBS gene coding regions showed a minimum 4-fold loss of nucleotide diversity in cultivated accessions than the wild germplasm in Asian pear (Additional file 8). Similarly, 17.2% of NBS-encoding regions had 4-fold genetic diversity loss during domestication in European pear. A total of 41 NBS-encoding genomic windows having reduction in genetic diversity were common in Asian and European populations (Additional file 8). Further analysis of fixation index across the whole NBS-encoding gene family revealed similar (T test,  $p$ -value = 0.31) average fixation index ( $F_{ST}$ ) values in Asian (5.77E-02; Fig. 5b) and European (5.89E-02; Fig. 5c) pear groups. However, similar to nucleotide diversity analysis, specific NBS gene coding regions showed high divergence between wild and cultivated groups in both the pear populations. For example, we found that 6 and 5 genes have  $F_{ST}$  values more than 0.30 in Asian and European populations, respectively.

A chi-square analysis revealed that a greater number of significantly ( $p$ -value < 0.01) different SNPs appeared in Asian (295) than European (122) population (Table 4; Additional file 9), and the number of non-synonymous SNPs also followed a similar trend (184 vs. 74) (Table 4; Additional file 10). We further calculated the Ka/Ks ratio to investigate positive selection in the 338 NBS-encoding genes. A set of 64 and 60 genes had signatures of positive selection ( $Ka/Ks > 1$ ) in Asian wild and cultivated groups (Table 4), and 51 genes were noticed in both groups (Additional file 11). A total of 57 and 63 genes had signatures of positive selection ( $Ka/Ks > 1$ ) in European wild and cultivated group (Table 4), and 51 genes



were in both groups. The trend in the number of genes under positive selection was very similar to the nucleotide diversity patterns in the four groups.

**Expression analysis of genes in wild and cultivated Asian pear**

We investigated the expression of a total of 17 genes expression. 16 of 17 genes had more than 2-fold mean Reads Per Kilobase per Million mapped reads (RPKM) difference between two wild pears and two cultivated pears, and another

gene (*Pbr003344.1*) was in black spot disease QTL (Additional file 12). These genes, *Pbr019876.1*, *Pbr039036.1* and *Pbr025269.1*, had significantly ( $p$ -value < 0.01) different SNPs between Asian wild and cultivated pear accessions and showed > 3-fold higher expression in 2 wild accessions than 2 cultivated pears at three different fruit developmental stages from RNA-Seq analysis (Fig. 6).

A quantitative real-time PCR (RT-qPCR) analysis of 17 genes also displayed consistent expression patterns as noticed in the RNA-seq analysis (Fig. 6). To investigate the



**Table 3** Nucleotide diversity ( $\pi$ ) of NBS gene family in different pear groups

| Pear accessions—groups<br>(number of samples) | $\pi$    | $\pi$<br>(Coding region) |
|---|----------|--------------------------|
| All (131)                                     | 5.88E-03 | 4.75E-03                 |
| Wild (61)                                     | 5.92E-03 | 4.82E-03                 |
| Cultivated (70)                               | 6.84E-03 | 5.46E-03                 |
| Asian (71)                                    | 6.02E-03 | 4.74E-03                 |
| European (60)                                 | 5.65E-03 | 4.76E-03                 |
| Asian wild (34)                               | 6.47E-03 | 5.15E-03                 |
| Asian cultivated (37)                         | 6.23E-03 | 4.88E-03                 |
| European wild (27)                            | 5.91E-03 | 5.02E-03                 |
| European cultivated (33)                      | 6.48E-03 | 5.40E-03                 |

The nucleotide diversity was calculated by VCFtools with a 1000 bp sliding windows and step size of 500 bp. The sequence data of 131 pear accessions were obtained from previous study [3]

role of these 17 genes in pathogen response, we compared RNA-seq data of control and *A. alternata* inoculated groups in accessions of a wild pear (*P. calleryana*) [20] and a cultivar of ‘Hongfen’ pear (*P. pyrifolia*) [19] (Additional file 13). Genes *Pbr019876.1*, *Pbr025269.1* and *Pbr018193.1* had upregulated (> 2-fold) expression in the inoculated than control samples of *P. calleryana*. The NBS-encoding gene, *Pbr003344.1*, under black spot disease resistance QTL also had up-regulated (~ 1.9-fold) expression after inoculation of *A. alternata* in *P. calleryana*.

#### A model to identify putative disease resistance in pear

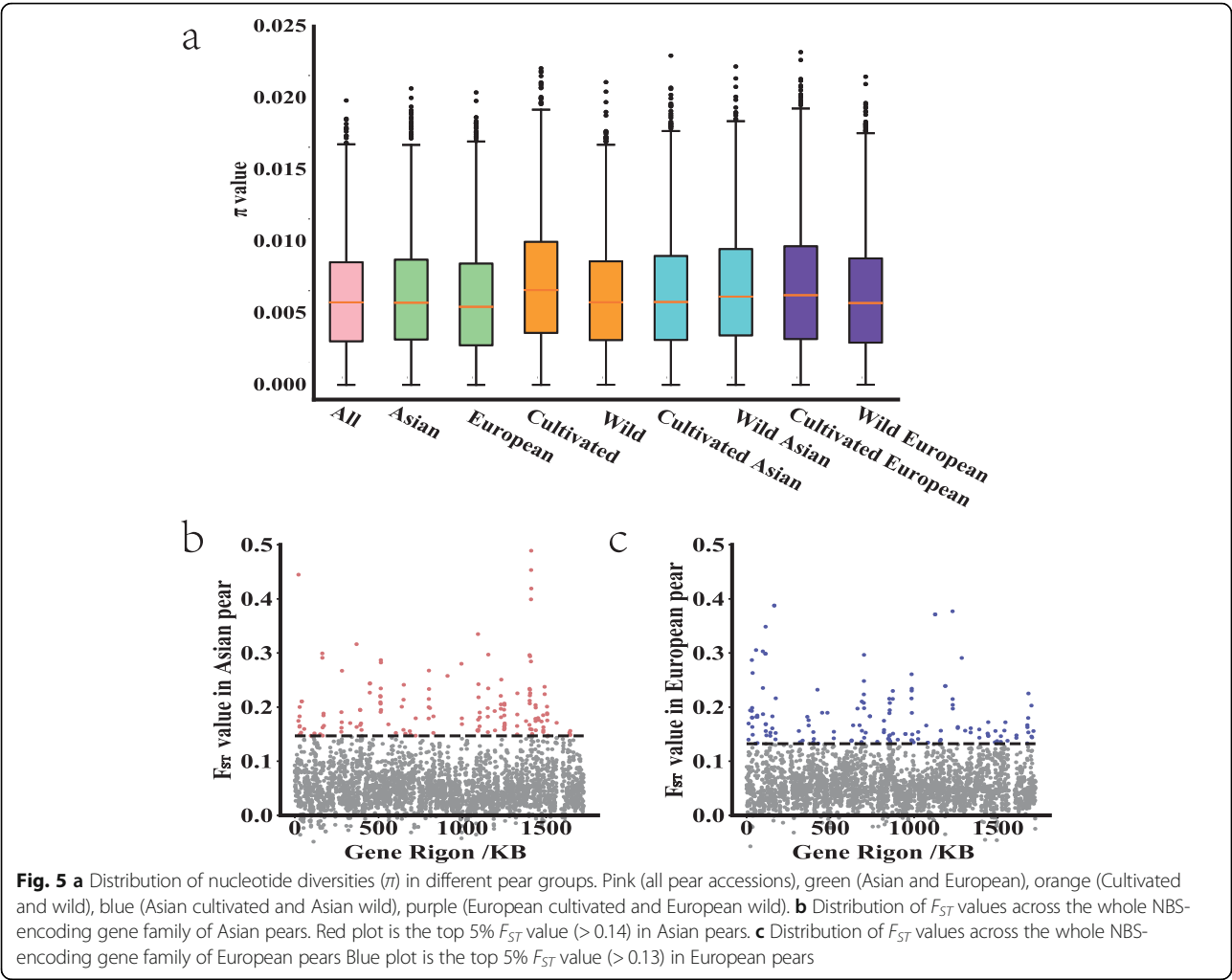
In order to detect the putative formation of resistance in wild and cultivated pear accessions, we combined the analysis of genetic variants, diversity, selection pressure and expression patterns in NBS-encoding genes. It was found that NBS-encoding gene of *Pbr025269.1* had 36 significantly (chi-square test  $p$ -value < 0.01) different SNPs between Asian wild and cultivated groups (Fig. 7a, b), and one of the SNP mutation caused amino acid change from leucine to tryptophan (Fig. 7a). This gene had higher nucleotide diversity in wild (1.22E-02) than cultivated (6.36E-03) group (Fig. 7c), had high divergence ( $F_{ST} = 0.32$ ) between Asian wild and cultivated groups, and showed a stronger signature of positive selection (Ka/Ks ratio = 2.04) in wild than cultivated (Ka/Ks ratio = 1.06) Asian pear accessions (Fig. 7c). Furthermore, *Pbr025269.1* also showed significantly ( $P$  value = 0.02) different expression between wild and cultivated accessions in Asian pears (Fig. 7d). Expression level of *Pbr025269.1* in the *A. alternata* inoculated *P. calleryana* (an Asian wild pear) accession was higher than observed in control groups (Fig. 7e), but this up-regulation did not appear in ‘Hongfen’ pear (Asian cultivated pear). Overall, positively selected *Pbr025269.1* had significantly

different SNPs and gene expression between the wild and domesticated Asian pear accessions.

## Discussion

### Difference in genetic variation patterns across the NBS-encoding genes between Asian and European pears

The NBS-encoding gene family has the most ancient and prevalent disease resistance genes, which play an important role in protecting plants from infection by diverse pathogens [37]. In this study, we observed a similarity in the distribution patterns and cis-elements types of NBS-encoding genes in *P. communis* and *P. bretschneideri*, indicating the conservation of core NBS genes before the divergence of these two pear species. The number of NBS-encoding genes in *P. bretschneideri* and *P. communis* were considerably different (338 vs 412), and only 64.79% orthologous NBS-encoding genes were present between them. Both Asian and European pear genomes had almost similar genome content, which did not define the differences in the number of NBS genes observed in this study [1, 32]. These differences might have originated from gene duplication and gene loss events associated with their adaptation to varying environments [38–40] during diversification of Asian and European pears approximately 6.6 and 3.3 MYA [1]. Due to the high heterozygosity and highly repetitive genome (53.1% repetitive sequences) of pear, 73 (21.60%) NBS genes of ‘Dangshansuli’ pear were not mapped to the chromosomes [1]. In addition to the gene gain/loss events, we also noticed higher nucleotide diversity in the cultivated pear group than the wild group from Europe (6.48E-03 vs. 5.91E-03), but the trend was opposite in the Asian population (wild vs. cultivated, 6.47E-03 vs. 6.23E-03). These findings suggest that a domestication-related loss of genetic diversity happened in Asian pear group. In contrast, the higher nucleotide diversity in NBS-encoding genes of cultivated groups in Europe might have occurred due to distinct selection pressure imposed by the cultivated habitats [41]. A recent study has shown that several regions with NBS-encoding genes had higher nucleotide diversity in landraces than the wild pear group [42]. The latter results, along with our findings, indicate that the genetic diversity of similar genes can be altered differently due to preferential selection of traits during independent domestication events as observed in pear [3]. This proposition was further supported by the identification of distinct NBS genes in Asian and European pears that showed reduction in genetic diversity after domestication bottleneck. Overall, the characterization and genetic diversity analysis suggest that NBS-encoding genes have undergone distinct selection pressures in pear due to its independent domestication in Asia and Europe.



The level of differentiation of NBS-encoding genes between wild and cultivated groups between two pear species also differs significantly from some other annual species. For instance, only 6 and 5 NBS-encoding genes show high divergence ( $F_{ST}$  value  $> 0.30$ ) between wild and cultivated groups from Asia and Europe, respectively, whereas 17 genes had high  $F_{ST}$  values ( $> 0.30$ ) between *Indica* and *nivara* rice species [43]. These differences reflect a much longer domestication history ( $> 10,000$  years) of rice [43] than the cultivated pear ( $\sim 3300$  years) [44]. The longer generation time and perennial nature might have also led to few sexual generations during pear domestication [3], which is a key factor causing lower divergence of NBS-encoding genes in cultivated and wild pear groups than annual crops.

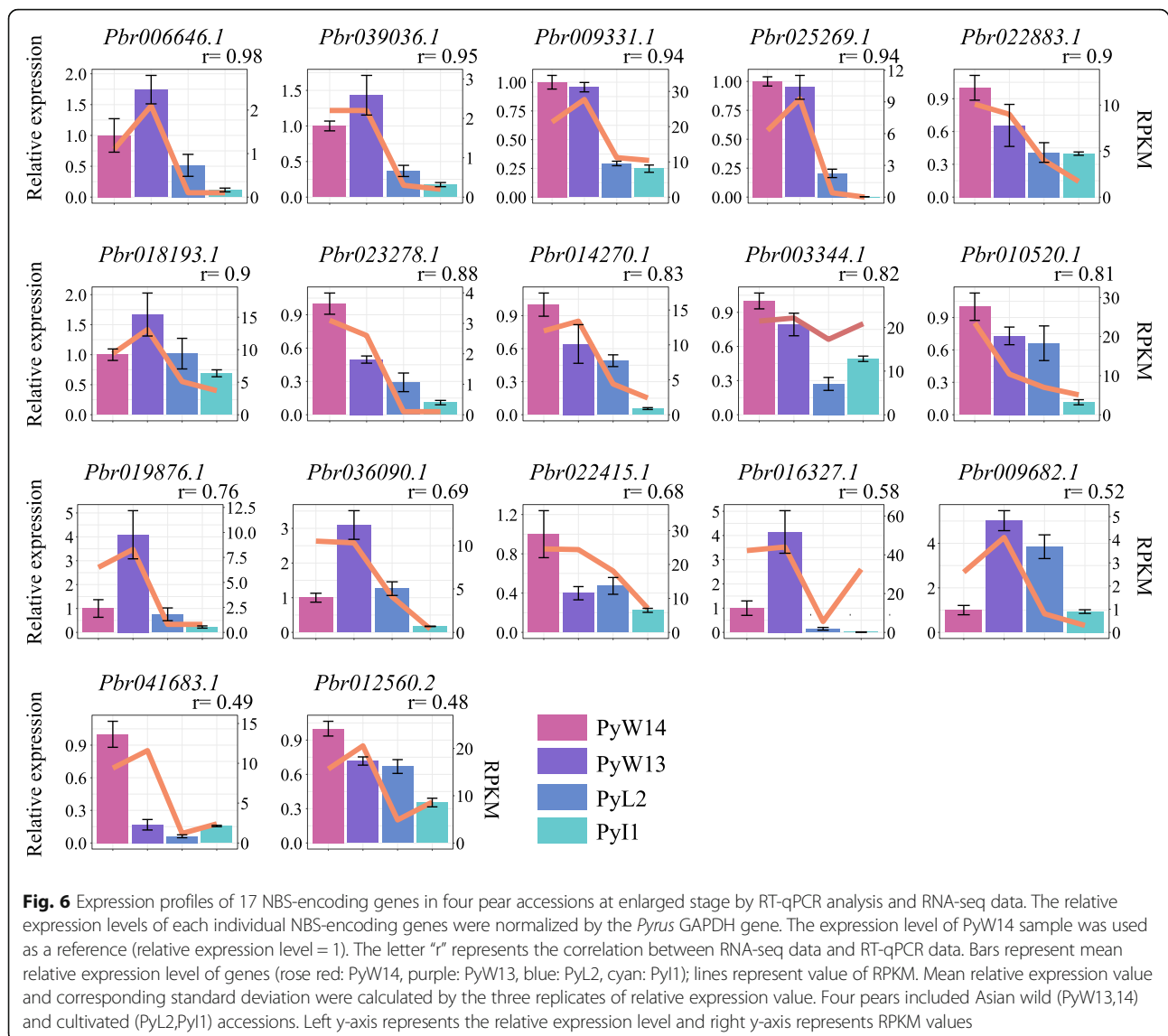
**Positive selection drives the evolution and selection of NBS-encoding genes in Asian and European pears**

Ka/Ks ratio of several paralogous gene pairs were greater than one, indicating a positive selection on NBS-

**Table 4** Summary of genetic variation and genes with signature of positive selection in different populations

| Population | Number of significantly different SNPs <sup>1</sup> | Number of non-synonymous SNPs <sup>a</sup> | Number of genes having non-synonymous SNPs | Subpopulation | Genes having signature of positive selection <sup>b</sup> |
|------------|---|--|--|---------------|---|
| Asian      | 295   | 184  | 61   | Wild          | 64  |
|            |   |  |  | Cultivated    | 60  |
| European   | 122   | 74   | 39   | Wild          | 57  |
|            |   |  |  | Cultivated    | 63  |

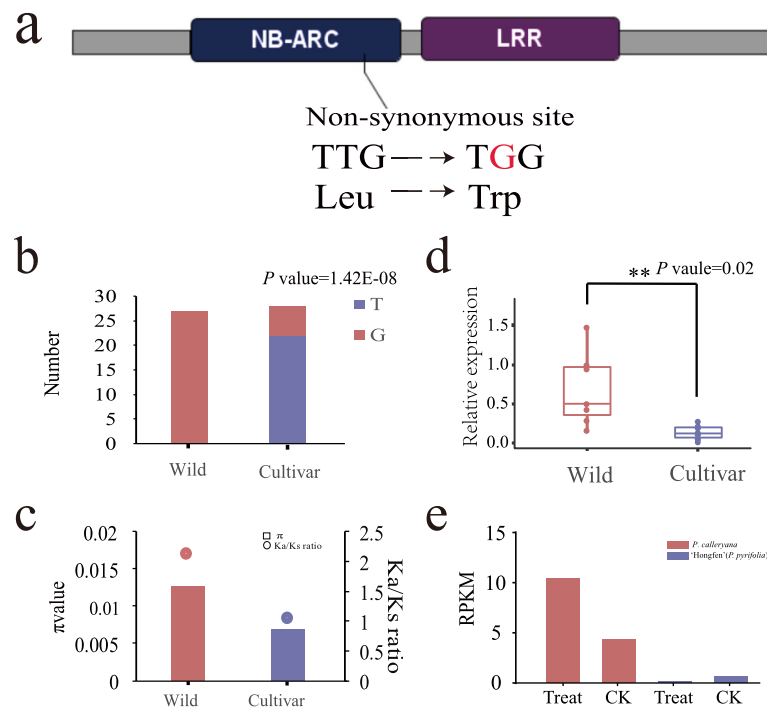
<sup>1</sup>Significantly different SNPs means this SNP has significant (chi-square test  $p$ -value  $< 0.01$ ) different alleles distribution between wild and cultivated groups. <sup>a</sup>Non-synonymous SNP means the significantly different SNP was also non-synonymous. <sup>b</sup>Genes with Ka/Ks ratio more than one were considered as having signature of positive selection in different populations



encoding gene family in pear. The NBS-encoding regions also had higher nonsynonymous / synonymous ratio (1.89) than the genome averages (1.20), a consistent pattern also seen in rice [41]. Moreover, the number of gene pairs showing signs of positive selection differed between *P. communis* and *P. bretschneideri*, with the latter showing comparatively stronger selection and more rapid evolution of NBS-encoding genes. The NBS-encoding genes also exhibited positive selection and rapid evolution in other species [26, 45]. Positive selection facilitates the prevalence of advantageous traits for the evolution of particular species [46–48], hence the signatures of positive selection indicate the increased fitness of NBS-encoding genes to overcome disease pressure by accumulating favorable alleles during evolution. These results also suggest that *P. communis* and *P. bretschneideri*, endured different disease pressures

during their divergence and domestication. For example, fire blight and brown spot damage was reported in pear production in Europe and west Asia [4–10], while ring rot, bitter rot, and scab infections occurred in China and east Asia [11–14]. The co-evolution with local pathogen populations and distinct disease pressure on pear production in different geographical areas can explain the difference in selection patterns across NBS-encoding genes in Asia and Europe.

We also found that the high percentage of these positively selected genes appeared in both wild and cultivated groups, indicating that these genes may play important roles in both wild and cultivated populations. Furthermore, a high percentage of non-synonymous SNPs showed significant (chi-square test  $p$ -value < 0.01) differences in Asian and European populations. These results were consistent with an existing study showing



**Fig. 7** Divergence of *Pbr025269.1* between Asian cultivated and wild groups. **a** One significantly different non-synonymous SNP located on NBS domain. **b** Genotyping distribution of SNP1 in Asian cultivated and wild groups (chi-square test,  $p$ -value = 1.42E-08). **c** Ka/Ks ratio and nucleotide diversity of *Pbr025269.1* in Asian wild and cultivated groups. Bars mean nucleotide diversities and spots mean Ka/Ks ratios. **d** Expression profiles of *Pbr025269.1* in ten Asian cultivated (PyL2, PyL7, PyL8, PyL9, PyL10, PyL11, PyL11, PyL11, PyL14) and seven Asian wild (PyW5, PyW6, PyW7, PyW9, PyW12, PyW13, PyW14) pear accessions at enlarged stage by RT-qPCR analysis. The relative expression levels of each individual NBS-encoding genes were normalized by the *Pyrus* GAPDH gene. The expression level of PyW14 sample was used as a reference (relative expression level = 1). Each point means relative expression level of one pear accession. \*\*\* means significant difference (T-test,  $p$ -value < 0.05) between wild and cultivated groups. **(e)** RPKM value of *Pbr025269.1* in *P. calleryana* and 'hongfen' pear (an Asian cultivate pear), 'Treat' means inoculated *A. alternata* groups and 'CK' means control groups; red represents *P. calleryana*, blue represents 'hongfen' (*P. pyrifolia*)

that significant divergent non-synonymous SNPs appeared among *M. x domestica* and other wild groups [49]. Host plants accumulate non-synonymous mutations to confer resistance against evolving pathogens [50], and higher frequency of these significantly different mutations in Asian and European pears suggest distinct pathogen pressure accompanying their independent domestication [37].

#### NBS-encoding genes and disease resistance in pear

The presence of 41 NBS-encoding genes in the previously identified disease-related QTL regions indicated their important roles in defense against these diseases. For example, two major QTL have been identified on chromosome 2 for fire blight resistance in 'Moonglow' (*P. communis*) and 'Harrow Sweet' (*P. communis*) pear cultivars [22, 34]. The distribution analysis has identified NBS gene clusters in these two QTL regions. Similarly, 15 NBS-encoding genes were located under the black spot resistance QTL on chromosome 11 of *P. bretschneideri*. Three NBS-encoding genes (*Pbr023226.1*, *Pbr023227.1*, *Pbr003344.1*) in the black spot QTL regions showed > 1.9-

fold expression after *A. alternata* inoculation in *P. calleryana* [20]. These gene expression differences were not significant after *A. alternata* inoculation in a susceptible 'Hongfen' pear (*P. pyrifolia*) [19], suggesting the specificity of gene expression based on the cultivar genetic background. Nonetheless, these expression differences might reflect the disease susceptibility or resistance depending on the genetic background of pear genotypes. Furthermore, the QTL-spanning NBS-encoding genes on chromosome 2 and 11 may serve as important functional candidates to study the genetic mechanisms of disease resistance against fire blight, pear scab and black spot in pear. For example, a detailed analysis identified positive selection and a high differential expression level of *Pbr025269.1*, a homolog of AT3G07040 to *Pseudomonas syringae* pv. *Maculicola* resistance [51, 52], in wild pears. This gene controls cell death at the site of infection [53, 54]. Transcriptome analysis also showed difference in expression of *Pbr025269.1* between *P. calleryana* and 'Hongfen' pear (*P. pyrifolia*) after infection with the *A. alternata* pathogen [55]. Thus, *Pbr025269.1*-like gene candidates can serve as functional candidates to fine map a



QTL-associated resistance and identifying their role in pathogen defense.

## Conclusion

The NBS-encoding gene family evolved through duplication events in *P. bretschneideri* and *P. communis*, and proximal duplication has resulted in different numbers of NBS-encoding genes in these two pear species. At a population level, the patterns of genetic diversity and positive selection were different across Asian and European pears. Moreover, several NBS-encoding genes were located under disease resistance QTL regions, indicating their potential role in disease resistance or susceptibility in pear. One specific NBS-encoding gene, *Pbr025269.1*, has shown signatures of selection, and exhibits high expression levels after inoculation with *A. alternata*.

## Methods

The pear genome was downloaded from Pear Center of Nanjing Agricultural University. The genome resequencing data from 113 pear accessions were obtained from previous study [3], and additional resequencing data from 18 pear accessions were obtained by using similar methods [3]. A total of 131 pear accessions were divided into two populations (Asian population and European population), and further divided into four groups (Asian cultivated group, Asian wild group, European cultivated group and European wild group) for genetic diversity and selection signature analysis.

### Identification and classification of NBS-encoding genes

For identification of NBS-encoding genes, the Chinese white pear (*P. bretschneideri* Rehd. cv. 'Dangshansuli') [1] and European pear (*P. communis* cv. 'Bartlett') [32] amino acid sequences were obtained from Pear Center of Nanjing Agricultural University (<http://peargenome.njau.edu.cn/>) and phytozome database (<https://phytozome.jgi.doe.gov/>), respectively. The sequence of NB-ARC domain (PF: 00931) was downloaded from Pfam database (<http://pfam.xfam.org/>). The software hmmer (version: 3.0) [50] was employed to search candidate NBS-encoding genes using a threshold expectation value of 1, and aiming to find the maximum number of candidate NBS-encoding genes. Also, we selected the candidate NBS-encoding genes (e-value <1E-20) as query sequence and used BLASTP [56] to search NBS-encoding genes separately in both the genomes. Subsequently, each of NBS gene in *Arabidopsis* was used as query sequence in BLASTP searching against the identified genes from above, and NCBI-Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/cdd/>) was used to identify the NBS domain identity in them. The same strategy was used to identify the NBS-encoding genes in *P. communis*.

The amino acid sequences of all 750 NBS-encoding genes in *P. bretschneideri* and *P. communis* genome were submitted to NCBI-CDD to identify the TIR, and LRR domain. The program COILS was used with a threshold of 0.9 [57] to specifically detect CC (coiled coils) domain in the identified genes.

### Alignment and phylogenetic tree analysis of NBS gene family

Clustalw2 [58] was employed to perform multiple alignments of the predicted amino acid sequences of gene coding region. Phylogenetic tree was constructed using the aligned amino acid sequences of the NBS proteins based on the ML (maximum likelihood) method by IQ-TREE (version 1.6.8) [59], and ultrafast bootstraps were estimated with 1000 replicates with default option.

### Cis-elements analysis of NBS-encoding genes

Promoters in the upstream 1500 bp regions of NBS-encoding genes were obtained from the pear annotation database [1] and phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). The Plant CARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) program was used to identify the cis-elements in promoter region of each gene. The same strategy was used to identify the cis-elements in NBS-encoding genes family of *P. communis*.

### Distribution and gene duplications of all NBS genes

Chromosomal locations of genes were displayed by using an in-house Rscript (<https://www.r-project.org/>). To analyze the NBS-encoding gene duplication events, the protein sequences of whole genomes were downloaded from pear genomic database and phytozome database. BLASTP was used to make an all-vs-all BLAST search (e-value of 1E-5 and top five matches). The BLAST output and the whole genome annotation file were imported to MCScanX [60] to identify homologous pairs, which were used to identify syntenic chains. The parameters were set to default as follows: Match score, 50; Overlap window, 5; Match size, 5; GAP penalty, -1; E-value, 1E-05; Max GAPS, 25 [61]. Duplicate gene classifier [60], the core function of MCScanX was used to classify all protein-coding genes into five types of duplications including tandem, singleton, WGD/ segmental, dispersed and proximal. BLASTP output and annotation file was used as input files. The assigning procedures were as following: (1) All genes were ranked based on their order along chromosomes and were classified as 'singletons', and all BLASTP hit results were evaluated for further classification; (2) If the genes had BLASTP hits to other genes, genes were labeled as dispersed duplicates; (3) If the two genes in a BLASTP hit, and their difference of gene rank was lower than 20 (less than 20 other genes insertion between two genes), then they

were labeled as proximal duplicates; (4) If the difference of two genes rank = 1 (no gene insertion between two genes), then they were labeled as tandem duplicates. (5) the anchor genes located in collinear blocks were labeled as WGD/segmental duplicates [60]. Finally, all duplicated genes were classified as a unique class based on the order of priority: WGD/segmental > tandem > proximal > dispersed [60]. We carried out the duplication analysis in both *P. bretschneideri* and *P. communis* using the same strategy.

#### Detection of positive/purifying selection

To detect positive and purifying selection in NBS-encoding gene family, the rate of non-synonymous (Ka) and synonymous (Ks) substitutions was calculated using the Ka/Ks\_Calculator2.0 [62] software with MYN method, and the ratio was obtained by dividing the Ka and Ks output from above.

#### Detection of NBS-encoding genes in disease resistance QTL

QTL information was retrieved from previous studies [13, 16, 17, 22, 33–35]. The marker sequences flanking the disease-related QTL peaks were extracted (Additional file 3), and bowtie2 (version 2.3.3.1) [63] was used to map primer sequences to the reference genomes (*P. bretschneideri* and *P. communis*) with the parameter: `-f -a -x` and max mismatches in seed is 3. The reference genome was selected based on the QTL backgrounds from Asian or European pears. The mapping coordinates from each primer were used to define the amplification regions on pear genomes. Finally, NBS-encoding genes in these QTL flanking regions were identified for further analysis.

#### SNP calling and screening SNP with significant difference

The software bwa (version 0.7.17-r1188) [64] was used to map the quality-filtered paired-end reads to the genomic sequences of 338 NBS-encoding genes. The potential PCR duplicates were removed from the alignment files using SAMtools (version 1.9) [65]. We used BCFtools (version 1.9) to perform SNP calls [66]. The low-quality SNPs were filtered out with the criteria: `QUAL < 20` or `DP < 4` or `QUAL/DP < 2`. The filtered SNP data set was used for further analysis.

After SNP calling, homozygous SNPs with missing rate < 0.1 were selected in 131 pear accessions. SNP genotypes of '0/1' were excepted from the analysis. Chi-square test was performed for the SNPs at each gene locus (excepted genotype of '0/1') in cultivated and wild populations, and the SNP having *p*-value < 0.01 were defined as significantly different SNP.

#### Diversity of NBS-encoding genes in four groups

To measure genetic diversity of the NBS-encoding gene family in pear, nucleotide diversity ( $\pi$ ) and fixation index ( $F_{ST}$ ) were used as population-differentiation statistics. The analysis was performed to compare the four pear groups (Asian wild, Asian cultivated, European wild, European cultivated). In brief, 338 NBS-encoding genes were combined to form a single sequence scaffold with the sorted order (Additional file 14), and SNPs were identified across them using the above defined methods. Then VCFtools (version 0.1.16) [67] was employed to further screen SNP data with the parameters: max-missing 0.50, maf 0.05. The resulting SNPs were used to calculate the  $\pi$  and  $F_{ST}$  with a 1000 bp sliding window and a step size of 500 bp across the linear sequence of all 338 NBS-encoding genes. The nucleotide diversity ( $\pi$ ) was also calculated separately for the coding regions.

#### Detecting signatures of selection in special genes

To identify signatures of selection, we calculated the Ka/Ks ratio of all orthologous NBS-encoding gene pairs to identify signatures of positive selection on them. Only SNPs from the coding regions were used form each reference gene. Furthermore, Ka/Ks\_calculator2.0 MYN method [62] was used to calculate the Ka and Ks of each orthologous pair in wild and cultivated groups, respectively. Afterwards, sequences with  $Ks > 0.1$  were discarded to avoid paralogs, and the average Ka, Ks and Ka/Ks ratio of each gene was calculated to identify the signatures of selection within specific group. Positive selection of genes was determined with a Ka/Ks ratio of more than one.

#### Genome-wide expression analysis of the NBS-encoding genes

To investigate the expression of the NBS-encoding genes that have signatures of selection, four pear genotypes representing two cultivated and two wild accessions at three different fruit developmental stages (small, enlarged, and mature) were used. The RNA-seq data of 12 samples were obtained from a previous study [42]. To investigate the expression changes of the NBS-encoding genes after inoculation with *A. alternata*, RNA-seq data of two pear species (*P. calleryana* and *P. pyrifolia*: 'Hongfen' pear) were used generated from previous study [19, 20]. RPKM value was used to measure the expression level of NBS-encoding genes. The R package (<https://cran.r-project.org/web/packages/pheatmap/>) was used to display the expression patterns of these genes.

#### Real-time PCR analysis

For gene expression validation, four pear accessions (PyW13, PyW14, PyI1 and PyL2) at enlarged stage were

chosen for quantitative real time PCR (RT-qPCR) analysis. For expression validation of *Pbr025269.1* with RT-qPCR, ten Asian cultivated (PyL2, PyL7, PyL8, PyL9, PyL10, PyL11, PyI1, PyI9, PyI11, PyI14) and seven Asian wild (PyW5, PyW6, PyW7, PyW9, PyW12, PyW13, PyW14) pear accessions were chosen from a previous study [42]. Plant Total RNA Isolation Kit plus was used to extract total RNA from pear fruits at the enlarged stage, and DNase I was used to remove the genomic DNA contamination. About 1 µg of total RNA was used for cDNA synthesis by using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech Co. Ltd.) according to the manufacturer's protocol. Primer5.0 was used to design primer sequences for gene amplification (Additional file 15). The SYBR®Green Premix kit (TaKaRa Biotechnology, Dalian, China) was used to carry out the analysis of RT-qPCR, and the PCR mixture was composed as follows: 150 ng of cDNA, 2.5 µl of each primer (10 µM), 10 µl of 2 × SYBR Premix ExTaq TM, and 5 µl of RNase-free water. The RT-qPCR started with 10 min at 95 °C, followed by 45 cycles of 95 °C for 15 s, 60 °C for 30 s and finally 72 °C for 30 s. The expression level of selected NBS-encoding genes was calculated by using  $2^{-\Delta\Delta C_t}$  method and normalized to the *Pyrus* GAPDH gene [42].

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-020-07226-1>.

**Additional file 1: Figure S1.** A phylogenetic tree of *P. bretschneideri* NBS proteins constructed by ML (Maximum likelihood) method using IQ-TREE. Fig. S2: A phylogenetic tree of *P. communis* NBS proteins constructed by ML method using IQ-TREE. ① Two subfamilies are shown. Red represents non-TIR group and Blue represents TIR group. ② The six classes of NBS-encoding genes are marked by different colors. Green means CC-NBS-LRR type, light yellow means TIR-NBS-LRR type, yellow means NBS type, light blue means TIR-NBS type, orange means CC-NBS type and purple means NBS-LRR type. ③ Domains of NB-ARC, LRR, and TIR are displayed on the tree (CC domain was not shown). Yellow means TIR domain, red means NB-ARC domain and blue means LRR domain.

**Additional file 2.** The information of all NBS-encoding genes in *P. bretschneideri* and *P. communis*, including type, location, and length of genomic, coding and protein sequences.

**Additional file 3.** NBS-encoding genes in pear disease resistance QTL regions. A total of 18 QTL were detected from previous study [13, 16, 17, 22, 33–35]. The primers of SSR markers in QTL were used to detect the marker location on the reference genome [1, 32]. Markers upstream and downstream 500 kbp were selected to identify the presence of NBS genes related to these QTL.

**Additional file 4.** Duplication categories of NBS-encoding genes in *P. bretschneideri* and *P. communis*. Five duplication categories (Singleton, Dispersed, Proximal, Tandem and WGD) were displayed.

**Additional file 5.** Ka/Ks ratios of paralogous gene pairs in *P. bretschneideri* and *P. communis*. Paralogous gene pairs were identified using MCScanX with default parameters. 'NA' means no values.

**Additional file 6.** Orthologous gene pairs between *P. bretschneideri* and *P. communis*. Orthologous gene pairs were identified using MCScanX with default parameters.

**Additional file 7.** Ka/Ks ratios of orthologous gene pairs between *P. bretschneideri* and *P. communis*. 'NA' means no values.

**Additional file 8.** The value of nucleotide diversity ( $\pi$ ) and  $F_{ST}$  within 1 kilobase sliding windows and step size of 500 bases across the whole NBS-encoding gene family from different populations.

**Additional file 9.** Significantly different SNPs between Asian (European) cultivated and wild groups. '0/0' means a homozygous reference site, '0/1' means a heterozygous site with two alleles (reference and variant) and '1/1' means a homozygous variant site. './.' means no data. Chi-square test was used to evaluate the significantly ( $p$ -value < 0.01) different SNPs between cultivated and wild groups.

**Additional file 10.** The list of non-synonymous or synonymous significantly different SNPs in Asian and European populations.

**Additional file 11.** Genes with Ka/Ks ratio > 1 in Asian and European cultivated and wild pear groups.

**Additional file 12.** The RPKM values of 338 NBS-encoding genes in four Asian pear accessions at three fruit stages. Letter 'NA' means no data.

**Additional file 13.** The expression patterns of 338 NBS-encoding genes' expression in *P. calleryana* and 'Hongfen' pear (*P. pyrifolia*). Letter 'Treat\_RPKM' means these genes' expression in treatment groups (inoculation of *A. alternata*). Letter 'CK\_RPKM' means these genes' expression in control groups. Letter 'NA' means no data.

**Additional file 14.** NBS-encoding genes in a sorted physical genome order. Genomic and coding sequences of 338 genes genomic were combined into a long sequence with the sorted order, respectively.

**Additional file 15.** Primer sequences of 17 genes used in RT-qPCR analysis.

## Acknowledgments

The authors would like to thank everyone who contributed to this article.

## Authors' contributions

MYS, MYZ and JS drafted the manuscript. YYS collected the public dataset and performed bioinformatics analysis. BBS contributed to data analysis. ZKT and JML worked for fruit samples collection, RNA extraction, RT-qPCR experiment and gene expression analysis. YYL, RZW and MFQ reviewed the manuscript. JW and AK conceived this study and prepared the manuscript. All authors read and approved the final manuscript.

## Funding

This work was funded by the Earmarked Fund for Jiangsu Agricultural Industry Technology System (JATS[2018]277), National Natural Science Foundation of China (31820103012, 31672111), the Earmarked Fund for China Agriculture Research System (CARS-28). This research was partially supported by the Federal Capacity Hatch Funds managed by the New York State Agricultural Experiment Station (NYSAES), Cornell University, Geneva, New York, USA.

## Availability of data and materials

The *P. bretschneideri* genome datasets used in this study are available in our pear center website (<http://peargenome.njau.edu.cn/>). The genome and amino acid sequences of *P. communis* were downloaded from the phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). Raw WGS data of all pear accessions were downloaded from NCBI BioProject (PRJNA381668, PRJNA675194). The RNA-seq data were obtained from the NCBI SRA database (PRJNA393405, PRJNA271833, PRJNA662252). The Rscript to draw the chromosome location was submitted to (<https://github.com/bioinformatic1996512/NBS>).

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 15 February 2020 Accepted: 11 November 2020

Published online: 19 November 2020

## References

- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 2013;23(2):396–408.
- Rubtsov GA. Geographical Distribution of the Genus *Pyrus* and Trends and Factors in Its Evolution. *Am Nat.* 1944;78(777):358–66.
- Wu J, Wang Y, Xu J, Korban SS, Fei Z, Tao S, Ming R, Tai S, Khan AM, Postman JD, et al. Diversification and independent domestication of Asian and European pears. *Genome Biol.* 2018;19(1):77.
- Llorente I, Montesinos E. Brown spot of pear: an emerging disease of economic importance in Europe. *Plant Dis.* 2006;90(11):1368–75.
- Llorente I, Moragrega C, Ruz L, Montesinos E. An update on control of brown spot of pear. *Trees (Berl West).* 2012;26(1):239–45.
- Shtienberg D, Manulis-Sasson S, Zilberstaine M, Oppenheim D, Shwartz H. The incessant battle against fire blight in pears: 30 years of challenges and successes in managing the disease in Israel. *Plant Dis.* 2015;99(8):1048–58.
- Shin DS, Heo GI, Son SH, Oh CS, Lee YK, Cha JS. Development of an improved loop-mediated isothermal amplification assay for on-site diagnosis of fire blight in apple and pear. *Plant Pathol J.* 2018;34(3):191–8.
- Bastas KK, Sahin F. First report of fire blight caused by *Erwinia amylovora* on meadowsweet (*Spiraea prunifolia*) in Turkey. *Plant Dis.* 2014;98(1):153.
- Jock S, Wensing A, Pulawska J, Drenova N, Dreo T, Geider K. Molecular analyses of *Erwinia amylovora* strains isolated in Russia, Poland, Slovenia and Austria describing further spread of fire blight in Europe. *Microbiol Res.* 2013;168(7):447–54.
- Bobev SG, Baeyen S, Crepel C, Maes M. First report of fire blight caused by *Erwinia amylovora* on *Pyracantha coccinea* in Bulgaria. *Plant Dis.* 2004;88(4):427.
- Li HN, Jiang JJ, Hong N, Wang GP, Xu WX. First report of Colletotrichum fructicola causing bitter rot of pear (*Pyrus bretschneideri*) in China. *Plant Dis.* 2013;97(7):1000.
- Gu X, Zhao J, Wang H, Lin FC, Guo Q, Shrivastava N, Jeewon R. ATMT transformation efficiencies with native promoters in Botryosphaeria kuwatsukai causing ring rot disease in pear. *World J Microbiol Biotechnol.* 2018;34(12):179.
- Terakami S, Moriya S, Adachi Y, Kunihisa M, Nishitani C, Saito T, Abe K, Yamamoto T. Fine mapping of the gene for susceptibility to black spot disease in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed Sci.* 2016;66(2):271–80.
- Choi ED, Kim GH, Park SY, Hoon Song J, Lee YS, Jung JS, Koh YJ. Genetic diversity of the pear scab fungus Venturia nashicola in Korea. *Mycobiology.* 2019;47(1):76–86.
- Pierantoni L, Dondini L, Cho KH, Shin IS, Gennari F, Chiodini R, Tartarini S, Kang SJ, Sansavini S. Pear scab resistance QTLs via a European pear (*Pyrus communis*) linkage map. *Tree Genet Genomes.* 2007;3(4):311–7.
- Terakami S, Shoda M, Adachi Y, Goni T, Kasumi M, Sawamura Y, Iketani H, Kotobuki K, Patocchi A, Gessler C, et al. Genetic mapping of the pear scab resistance gene Vnk of Japanese pear cultivar Kinchaku. *Theor Appl Genet.* 2006;113(4):743–52.
- Bokszczanin K, Dondini L, Przybyla AA. First report on the presence of fire blight resistance in linkage group 11 of *Pyrus ussuriensis* maxim. *J Appl Genet.* 2009;50(2):99–103.
- Dondini L, Pierantoni L, Gaiotti F, Chiodini R, Tartarini S, Bazzi C, Sansavini S. Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. *Mol Breed.* 2004;14(4):407–18.
- Yang X, Hu H, Yu D, Sun Z, He X, Zhang J, Chen Q, Tian R, Fan J. Candidate resistant genes of sand pear (*Pyrus pyrifolia* Nakai) to Alternaria alternata revealed by Transcriptome sequencing. *PLoS One.* 2015;10(8):e0135046.
- Kan J, Liu T, Ma N, Li H, Li X, Wang J, Zhang B, Chang Y, Lin J. Transcriptome analysis of Callery pear (*Pyrus calleryana*) reveals a comprehensive signalling network in response to Alternaria alternata. *PLoS One.* 2017;12(9):e0184988.
- Abe K, Saito T, Terai O, Sato Y, Kotobuki K. Genotypic difference for the susceptibility of Japanese, Chinese and European pears to Venturia nashicola, the cause of scab on Asian pears. *Plant Breed.* 2008;127(4):407–12.
- Montanari S, Perchev L, Renault D, Frijters L, Velasco R, Horner M, Gardiner SE, Chagne D, Bus VGM, Durel CE, et al. A QTL detected in an interspecific pear population confers stable fire blight resistance across different environments and genetic backgrounds. *Mol Breed.* 2016;36(4):1–16.
- van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL. Structure and function of resistance proteins in solanaceous plants. *Annu Rev Phytopathol.* 2007;45:43–72.
- McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006;7(4):212.
- DeYoung BJ, Innes RW. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol.* 2006;7(12):1243–9.
- Jia Y, Yuan Y, Zhang Y, Yang S, Zhang X. Extreme expansion of NBS-encoding genes in Rosaceae. *BMC Genet.* 2015;16(1):48.
- Zhong Y, Yin H, Sargent D, Malnoy M, Z-MJBG C. Species-specific duplications driving the recent expansion of NBS-LRR genes in five Rosaceae species. *BMC Genomics.* 2015;16(1):77.
- Arya P, Kumar G, Acharya V, AKJPO S. Genome-Wide Identification and Expression Analysis of NBS-Encoding Genes in *Malus x domestica* and Expansion of NBS Genes Family in Rosaceae. *PLoS One.* 2014;9(9):e107987.
- Perazzolli M, Malacarne G, Baldo A, Righetti L, Bailey A, Fontana P, Velasco R, Malnoy MJPO. Characterization of Resistance Gene Analogues (RGAs) in Apple (*Malus x domestica* Borkh.) and Their Evolutionary History of the Rosaceae Family. *PLoS One.* 2014;9:e83844.
- Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen JQ, Tian D, Yang S. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol Gen Genomics.* 2010;283(5):427–38.
- Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen JQ, Hartl DL, Tian D. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. *Proc Natl Acad Sci U S A.* 2013;110(46):18572–7.
- Linsmith G, Rombauts S, Montanari S, Deng CH, Celson J-M, Guérif P, Liu C, Lohaus R, Zurn JD, Cestaro A, et al. Pseudo-chromosome length genome assembly of a double haploid 'Bartlett' pear (*Pyrus communis* L.); 2019. p. 651778.
- Perchev L, Leforestier D, Ravon E, Guérif P, Denance C, Tellier M, Terakami S, Yamamoto T, Chevalier M, Lespinasse Y, et al. Genetic mapping and pyramiding of two new pear scab resistance QTLs. *Mol Breed.* 2015; 35(10):197.
- Le Roux PMF, Christen D, Duffy B, Tartarini S, Dondini L, Yamamoto T, Nishitani C, Terakami S, Lespinasse Y, Kellerhals M, et al. Redefinition of the map position and validation of a major quantitative trait locus for fire blight resistance of the pear cultivar 'Harrow Sweet' (*Pyrus communis* L.). *Plant Breed.* 2012;131(5):656–64.
- Cappai F, De Franceschi P, Ciriani A, Collina M, Dondini L. QTLs for susceptibility to *Stemphylium vesicarium* in pear. *Mol Breed.* 2018;38(3):24.
- Fawcett JA, Maere S, Van de Peer Y. Plants with double genomes might have had a better chance to survive the cretaceous-tertiary extinction event. *P Natl Acad Sci USA.* 2009;106(14):5737–42.
- McDowell JM, Simon SA. Recent insights into R gene evolution. *Mol Plant Pathol.* 2010;7(5):437–48.
- Xu M, Chen F, Qi S, Zhang L, Wu S. Loss or duplication of key regulatory genes coincides with environmental adaptation of the stomatal complex in *Nymphaea colorata* and *Kalanchoe laxiflora*. *Hortic Res.* 2018;5:42.
- Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CK, Severn-Ellis A, McCombie WR, Parkin IA, et al. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat Commun.* 2016;7: 13390.
- Zhong Y, Zhang XH, Cheng ZM. Lineage-specific duplications of NBS-LRR genes occurring before the divergence of six *Fragaria* species. *BMC Genomics.* 2018;19:128.
- Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol.* 2011;30(1):105–11.
- Li X, Liu L, Ming M, Hu H, Zhang M, Fan J, Song B, Zhang S, Wu J. Comparative Transcriptomic analysis provides insight into the domestication and improvement of pear (*P. pyrifolia*) fruit. *Plant Physiol.* 2019;180(1):435–52.
- Kovach MJ, Sweeney MT, McCouch SR. New insights into the history of rice domestication. *Trends Genet.* 2007;23(11):578–87.
- Yamamoto T, Terakami S. Genomics of pear and other Rosaceae fruit trees. *Breed Sci.* 2016;66(1):148–59.
- Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet.* 2010;42(12): 1053–9.



46. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 2002; 19(6):908–17.
47. Zheng F, Wu H, Zhang R, Li S, He W, Wong FL, Li G, Zhao S, Lam HM. Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. *BMC Genomics.* 2016;17:402.
48. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. *Science.* 2006;312(5780):1614–20.
49. Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, Wang X, Jiao C, Legall N, Mao L, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun.* 2017;8(1):249.
50. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 2009;23(1):205–11.
51. Keinath AP, Wechter WP, Smith JP. First Report of Bacterial Leaf Spot on Leafy *Brassica* Greens Caused by *Pseudomonas syringae* pv. *maculicola* in South Carolina. *Plant Dis.* 2006;90(5):683.
52. El Kasmi F, Chung EH, Anderson RG, Li J, Wan L, Eitas TK, Gao Z, Dangl JL. Signaling from the plasma-membrane localized plant immune receptor RPM1 requires self-association of the full-length protein. *Proc Natl Acad Sci U S A.* 2017;114(35):E7385–94.
53. Boyes DC, Nam J, Dangl JL. The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proc Natl Acad Sci U S A.* 1998;95(26):15849–54.
54. Grant M, Brown I, Adams S, Knight M, Ainslie A, Mansfield J. The RPM1 plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *Plant J.* 2000;23(4):441–50.
55. Buckley J, Kilbride E, Cevik V, Vicente JG, Holub EB, Mable BK. R-gene variation across *Arabidopsis lyrata* subspecies: effects of population structure, selection and mating system. *BMC Evol Biol.* 2016;16:93.
56. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
57. Mace E, Tai S, Innes D, Godwin I, Hu W, Campbell B, Gilding E, Cruickshank A, Prentis P, Wang J, et al. The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biol.* 2014;14(1):253.
58. Larkin MA, Blackshields G, Brown NP, Chenna R, PA MG, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947–8.
59. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
60. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40(7):e49.
61. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 2008;18(12):1944–54.
62. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinform.* 2010;8(1):77–80.
63. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9.
64. Li H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics.* 2012;28(14):1838–44.
65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
66. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27(21):2987–93.
67. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

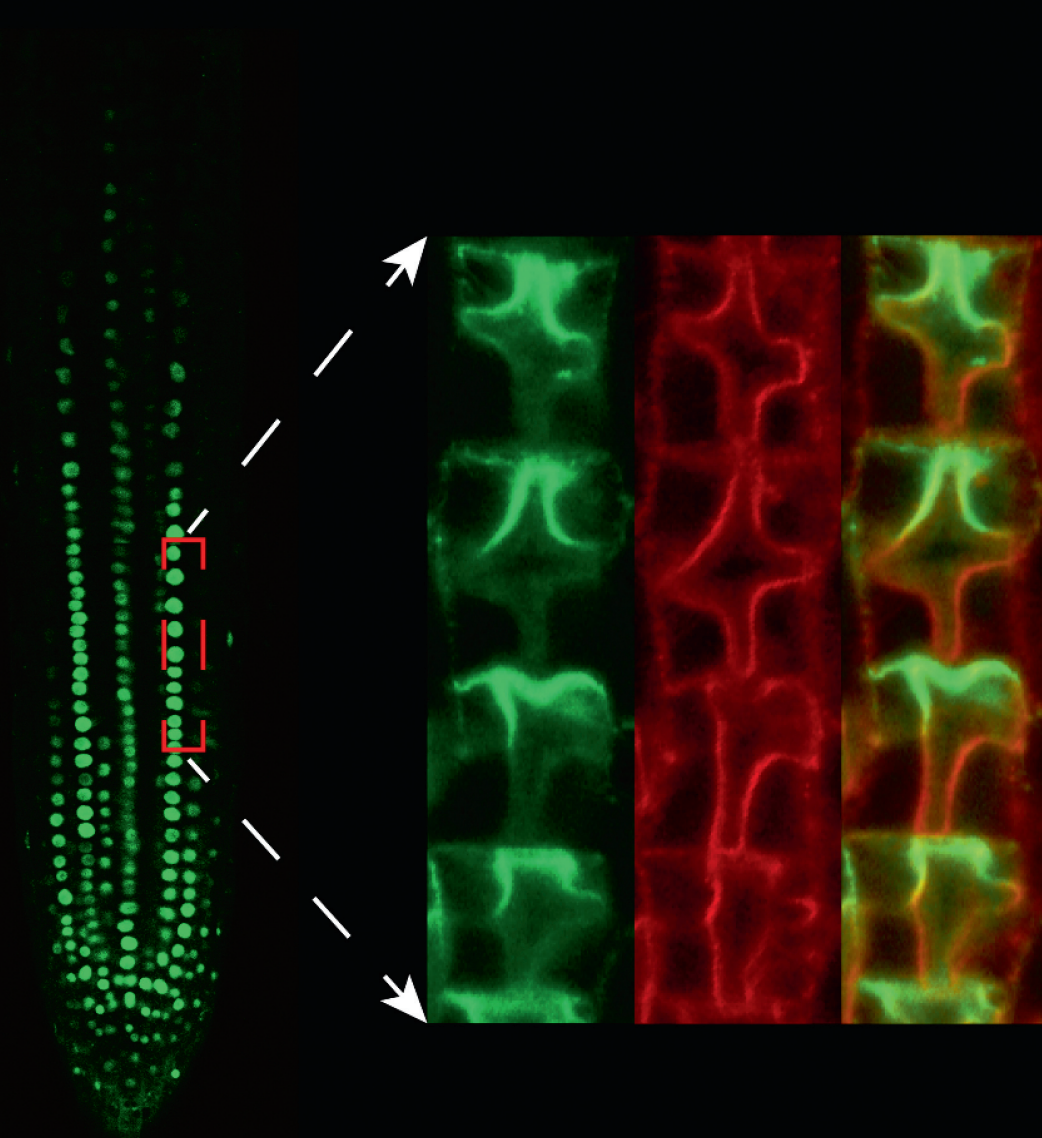
- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



# *the plant journal*



WILEY

VOLUME 114 | NUMBER 1 | APRIL 2023  
<http://www.theplantjournal.com> | ISSN 1365-313X

166

## Editorial information

*The Plant Journal* is edited by Professor Lee Sweetlove, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK. Professor Sweetlove is supported by an international team of Editors: <https://onlinelibrary.wiley.com/page/journal/1365313x/homepage/editorialboard.html>

Correspondence relating to editorial matters should be directed to *The Plant Journal* Editorial Office, John Wiley & Sons Ltd, 9600 Garsington Road, Oxford, OX4 2DQ, UK (Email: [tpj-editor@wiley.com](mailto:tpj-editor@wiley.com)).

## Subscription information

*The Plant Journal* is published semi-monthly (four volumes per annum), with two issues published each month. Subscription prices for 2023 are: Online only: US\$7735 (The Americas), US\$9026 (Rest of World), €5313 (Europe); £4186 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to [www.wileyonlinelibrary.com/tax-vat](http://www.wileyonlinelibrary.com/tax-vat). The price includes online access to the current and all online back files for previous 5 years, where available. For other pricing options, including access information and terms and conditions, please visit [www.wileyonlinelibrary.com/access](http://www.wileyonlinelibrary.com/access).

## Back issues

Single issues from current and recent volumes are available at the current single issue price from [cs-journals@wiley.com](mailto:cs-journals@wiley.com). Earlier issues may be obtained from Periodicals Service Company, 351 Fairview Avenue – Ste 300, Hudson, NY 12534, USA. Tel: +1 518 822-9300, Fax: +1 518 822-9305, Email: [psc@periodicals.com](mailto:psc@periodicals.com)

## Publisher

*The Plant Journal* is published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ. Tel: +44 (0)1865 776868; Fax: +44 (0)1865 714591.

## Journal Customer Services

For ordering information, claims and any enquiry concerning your journal subscription please go to <https://wolsupport.wiley.com/s/contactsupport?tabset-a7d10=2> or contact your nearest office.

**Americas:** Email: [cs-journals@wiley.com](mailto:cs-journals@wiley.com); Tel: +1 877 762 2974.

**Europe, Middle East and Africa:** Email: [cs-journals@wiley.com](mailto:cs-journals@wiley.com); Tel: +44 (0) 1865 778315; 0800 1800 536 (Germany)

**Germany, Austria, Switzerland, Luxembourg, Liechtenstein:** Email: [cs-germany@wiley.com](mailto:cs-germany@wiley.com); Tel: 0800 1800 536 (Germany)

**Asia Pacific:** E-mail: [cs-journals@wiley.com](mailto:cs-journals@wiley.com); Tel: +65 3165 0890

**Japan:** For Japanese speaking support, e-mail: [cs-japan@wiley.com](mailto:cs-japan@wiley.com)

**Visit our Online Customer Help** at <https://wolsupport.wiley.com/s/contactsupport?tabset-a7d10=2>

## Despatch

THE PLANT JOURNAL, (ISSN 0960-7412), is published semi-monthly (four volumes per annum).

Postmaster: Send all address changes to THE PLANT JOURNAL, Wiley Periodicals LLC, C/O The Sheridan Press, PO Box 465, Hanover, PA 17331 USA.

## Copyright and Copying

Copyright © 2023 Society for Experimental Biology and John Wiley & Sons Ltd. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission

in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organisation (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA ([www.copyright.com](http://www.copyright.com)), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. Special requests should be addressed to: [permissions@wiley.com](mailto:permissions@wiley.com).

## Online Open

OnlineOpen is available to authors of primary research articles who wish to make their article available to non-subscribers on publication, or whose funding agency requires grantees to archive the final version of their article. With OnlineOpen, the author, the author's funding agency, or the author's institution pays a fee to ensure that the article is made available to non-subscribers upon publication via Wiley Online Library, as well as deposited in the funding agency's preferred archive. For the full list of terms and conditions, see [http://wileyonlinelibrary.com/onlineopen#OnlineOpen\\_Terms](http://wileyonlinelibrary.com/onlineopen#OnlineOpen_Terms).

Any authors wishing to send their paper OnlineOpen will be required to complete the payment form available from our website at: [https://authorservices.wiley.com/bauthor/onlineopen\\_order.asp](https://authorservices.wiley.com/bauthor/onlineopen_order.asp).

Prior to acceptance there is no requirement to inform an Editorial Office that you intend to publish your paper OnlineOpen if you do not wish to. All OnlineOpen articles are treated in the same way as any other article. They go through the journal's standard peer-review process and will be accepted or rejected based on their own merit.

The Journal is indexed by Current Contents/Life Sciences, Science Citation Index, Index Medicus, MEDLINE and BIOBASE/Current Awareness in Biological Sciences.

Access to this journal is available free online within institutions in the developing world through the AGORA initiative with the FAO, the HINARI initiative with the WHO and the OARE initiative with UNEP. For information, visit [www.aginternet.org](http://www.aginternet.org), [www.healthinternetwork.org](http://www.healthinternetwork.org), [www.oarescience.org](http://www.oarescience.org).

Published by John Wiley & Sons Ltd, in association with the Society for Experimental Biology. Information on this journal can be accessed at <http://www.theplantjournal.com>. This journal is available online at Wiley Online Library. Visit [wileyonlinelibrary.com](http://wileyonlinelibrary.com) to search the articles and register for Table of Contents email alerts.

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. We have made a long-term commitment to standardize and improve our efforts around the world to reduce our carbon footprint. Follow our progress at [www.wiley.com/go/citizenship](http://www.wiley.com/go/citizenship).

## Disclaimer

The Publisher, the Society for Experimental Biology and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, the Society for Experimental Biology or editors, neither does the publication of advertisements constitute any endorsement by the Publisher, the Society for Experimental Biology or the Editors of the products advertised.

# the plant journal

CONTENTS OF VOL. 114, NO. 1, APRIL 2023

## RESEARCH HIGHLIGHT

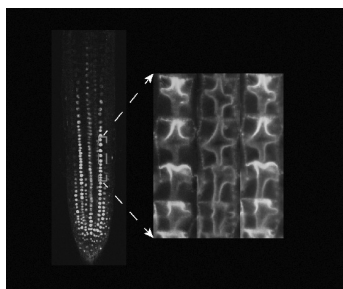
- 5 **2D or not 2D: how mRNA methylation regulates the transition to 3-dimensional growth in *Physcomitrium patens*.** G. Pierroz (Research Highlights Editor)

## ORIGINAL ARTICLES

- 7 **M<sup>6</sup>-Methyladenosine modification of mRNA contributes to the transition from 2D to 3D growth in the moss *Physcomitrium patens*.** D. Garcías-Morales, V. M. Palomar, F. Charlot, F. Nogué, A. A. Covarrubias and J. L. Reyes
- 23 **Genetic analysis of resistance to bean leaf crumple virus identifies a candidate *LRR-RLK* gene.** D. Ariza-Suarez, B. Keller, A. Spescha, J. S. Aparicio, V. Mayor, A. E. Portilla-Benavides, H. F. Buendía, J. M. Bueno, B. Studer and B. Raatz
- 39 **RAV1 family members function as transcriptional regulators and play a positive role in plant disease resistance.** R. K. Chandan, R. Kumar, D. M. Swain, S. Ghosh, P. K. Bhagat, S. Patel, G. Bagler, A. K. Sinha and G. Jha
- 55 **Raffinose positively regulates maize drought tolerance by reducing leaf transpiration.** Y. Liu, T. Li, C. Zhang, W. Zhang, N. Deng, L. M. A. Dirk, A. B. Downie and T. Zhao
- 68 **Chloroplast dehydroascorbate reductase and glutathione cooperatively determine the capacity for ascorbate accumulation under photooxidative stress conditions.** A. Hamada, Y. Tanaka, T. Ishikawa and T. Maruta
- 83 **Role of reactive oxygen species in the modulation of auxin flux and root development in *Arabidopsis thaliana*.** T. Pasternak, K. Palme and J. M. Pérez-Pérez
- 96 ***Arabidopsis* SMO2 modulates ribosome biogenesis by maintaining the RID2 abundance during organ growth.** Z. Guo, X. Wang, Y. Li, A. Xing, C. Wu, D. Li, C. Wang, A. de Bures, Y. Zhang, S. Guo, J. Sáez-Vasquez, Z. Shen and Z. Hu
- 110 **PHYTOCHROME-INTERACTING FACTORS are involved in starch degradation adjustment via inhibition of the carbon metabolic regulator *QUA-QUINE STARCH* in *Arabidopsis*.** S. W. Lee, D. Choi, H. Moon, S. Kim, H. Kang, I. Paik, E. Huq and D.-H. Kim
- 124 **Natural variations in the *PbCPK28* promoter regulate sugar content through interaction with *PbTST4* and *PbVHA-A1* in pear.** J. Li, R. Zhu, M. Zhang, B. Cao, X. Li, B. Song, Z. Liu and J. Wu
- 142 **Subgenome-dominant expression and alternative splicing in response to *Sclerotinia* infection in polyploid *Brassica napus* and progenitors.** G. W. de Jong and K. L. Adams
- 159 **Co-action of COP1, SPA and cryptochrome in light signal transduction and photomorphogenesis of the moss *Physcomitrium patens*.** M. Kreiss, F. B. Haas, M. Hansen, S. A. Rensing and U. Hoecker
- 176 **Boron supply restores aluminum-blocked auxin transport by the modulation of PIN2 trafficking in the root apical transition zone.** L. Tao, X. Xiao, Q. Huang, H. Zhu, Y. Feng, Y. Li, X. Li, Z. Guo, J. Liu, F. Wu, N. Pirayesh, S. Mahmud, R. F. Shen, S. Shabala, F. Baluška, L. Shi and M. Yu
- 193 **The transcription factor MYC1 interacts with FIT to negatively regulate iron homeostasis in *Arabidopsis thaliana*.** H. Song, Q. Geng, X. Wu, M. Hu, M. Ye, X. Yu, Y. Chen, J. Xu, L. Jiang and S. Cao

## RESOURCE

- 209 **The coordinated regulation of early meiotic stages is dominated by non-coding RNAs and stage-specific transcription in wheat.** Y. Jiang, A. N'Diaye, C. S. Koh, T. D. Quilichini, A. S. K. Shunmugam, M. W. Kirzinger, D. Konkin, Y. Bekkaoui, E. Sari, A. Pasha, E. Esteban, N. J. Provart, J. D. Higgins, K. Rozwadowski, A. G. Sharpe, C. J. Pozniak and S. Kagale



**Front cover:** Root apical auxin distribution visualized in the *DII-VENUS* reporter line showing a negative correlation with intracellular auxin signaling. Root basipetal polar auxin transport is ascribed to the PIN2 carrier (marked in green) in the cross-shaped plasma membrane (marked in red) of transgenic PIN2-GFP plants. The merged picture shows that PIN2 localizes to the upper half of the plasma membrane, directing auxin flux towards the elongation zone for root growth. Tao *et al.* (pp. 176–192)



# Natural variations in the *PbCPK28* promoter regulate sugar content through interaction with PbTST4 and PbVHA-A1 in pear

Jiaming Li<sup>1,†</sup>, Rongxiang Zhu<sup>1,†</sup>, Mingyue Zhang<sup>2,†</sup>, Beibei Cao<sup>1</sup>, Xiaolong Li<sup>3</sup>, Bobo Song<sup>1</sup>, Zhongchi Liu<sup>4</sup>  and Jun Wu<sup>1,5,\*</sup> 

<sup>1</sup>National Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, College of Horticulture, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China,

<sup>2</sup>State Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an, Shandong 271018, China,

<sup>3</sup>College of Horticulture Science, Zhejiang Agriculture and Forestry University, Hangzhou, Zhejiang 311200, China,

<sup>4</sup>Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742, USA, and

<sup>5</sup>Zhongshan Biological Breeding Laboratory, Nanjing, Jiangsu 210014, China

Received 31 May 2022; revised 12 January 2023; accepted 19 January 2023; published online 30 January 2023.

\*For correspondence (e-mail [wujun@njau.edu.cn](mailto:wujun@njau.edu.cn)).

<sup>†</sup>These authors contributed equally to this work.

## SUMMARY

Soluble sugars play an important role in plant growth, development and fruit quality. Pear fruits have demonstrated a considerable improvement in sugar quality during their long history of selection. However, little is known about the underlying molecular mechanisms accompanying the changes in fruit sugar content as a result of selection by horticulturists. Here, we identified a calcium-dependent protein kinase (PbCPK28), which is located on LG15 and is present within a selective sweep region, thus linked to the quantitative trait loci for soluble solids. Association analysis indicates that a single nucleotide polymorphism-13 variation (SNP13<sup>T/C</sup>) in the *PbCPK28* regulatory region led to fructose content diversity in pear. Elevated expression of *PbCPK28* resulted in significantly increased fructose levels in pear fruits. Furthermore, PbCPK28 interacts with and phosphorylates PbTST4, a proton antiporter, thereby coupling the sugar import into the vacuole with proton export. We demonstrated that residues S277 and S314 of PbTST4 are crucial for its function. Additionally, PbCPK28 interacts with and phosphorylates the vacuolar hydrogen proton pump PbVHA-A1, which could provide proton motive forces for PbTST4. We also found that the T11 and Y120 phosphorylation sites in PbVHA-A1 are essential for its function. Evolution analysis and yeast-two-hybrid results support that the CPK-TST/CPK-VHA-A regulatory network is highly conserved in plants, especially the corresponding phosphorylation sites. Together, our work identifies an agriculturally important natural variation and an important regulatory network, allowing genetic improvement of fruit sugar contents in pears through modulation of *PbCPK28* expression and phosphorylation of PbTST4 and PbVHA-A1.

**Keywords:** natural variations, calcium-dependent protein kinase, H<sup>+</sup>/sugar antiporter, phosphorylate, vacuolar hydrogen proton pump.

## INTRODUCTION

Pear (*Pyrus* L.) belongs to the subtribe Malinae of the Amygdaloideae subfamily within Rosaceae (Potter et al., 2007). As an ancient crop, pear has more than 3000 years of cultivation history (Lombard & Westwood, 1987), a cultivated area of approximately 1.29 million ha, and an annual production exceeding 23.11 million tons globally (FAOSTATA 2020, <http://www.fao.org/faostat/zh/#data/QCL/visualize>). Originating in southwest China and

disseminated throughout central Asia, pear eventually spread to western Asia and then to Europe (Wu et al., 2018). China was the first producer of pear in the world, possessing thousands of cultivars, dozens of wild species and five cultivated species, with an annual production of approximately 16.11 million tons, which accounts for 70% of the total global pear output. Recently, the genomes of 113 *Pyrus* accessions were reported, providing insights into the independent domestication of Asian and European

pears, as well as a sequence inventory for studies of the phenotypic differences between cultivated and wild pears, such as sugar content and stone cell content, which could be selected during domestication (Wu et al. 2018).

Domestication was the earliest form of plant breeding, and is a good example of evolution (Doebley et al., 2006). Crop domestication is a complex process of selection, and multiple traits of agronomical importance are involved. In the past decades, more and more crop genomes have been re-sequenced, making it possible to investigate the genes that underwent domestication (Cao et al., 2014; Du et al., 2018; Duan et al., 2017; Wang et al., 2017; Wu et al., 2018; Xu et al., 2012; Zhou et al., 2015). A number of domestication-related genes involved in seed shattering, seed size and dispersal, plant architecture and fruit size were identified and cloned from several crops (Chakrabarti et al., 2013; Dong et al., 2014; Li et al., 2011; Müller et al., 2018; Pourkheirandish et al., 2015; Sigmon & Vollbrecht, 2010; Soyk et al., 2017; Soyk et al., 2019; Studer et al., 2011; Zhang et al., 2019). A few genes that have undergone parallel selection within a single or across multiple plant families have been identified as the key factors controlling these important traits of domestication (Lin et al., 2012; Liu et al., 2015; Wang et al., 2018). In fruit crops, one of the most important motivations behind domestication is to improve fruit quality, as almost all corresponding wild species have low sugar content, high organic acid content, small fruit size and high flesh firmness, suggesting that these traits might be selected against during domestication (Chakrabarti et al., 2013; Frary et al., 2000; Soyk et al., 2017).

Sugar content is a quantitative trait controlled by multiple genes, a part of a complex metabolic network. In the plant kingdom, the vacuole occupies more than 90% of the plant cell volume (Winter et al., 1993), making it ideally suited for storage (Martinoia et al., 2006; Neuhaus, 2007). Sugars, upon entering cells, are directed towards the vacuole and stored for subsequent plant growth and development, biotic and abiotic stress tolerance, and energy source for the synthesis of other metabolites (Wormit et al., 2006). In fruits, the final storage level of sugars at the ripening stage affects fruit quality. Early evidence showed that sugars enter the vacuole through the tonoplast membrane, and the tonoplast monosaccharide sugar transporter (*TMT*) and vacuolar glucose transporter (*VG1*) were the first identified sugar transporters found to be localized on the tonoplast (Aluri & Buttner, 2007; Wormit et al., 2006). *TMT* (also referred to as *TST*) operates as a proton antiporter, coupling the import of sugars into the vacuole with the export of protons (Jung et al., 2015; Wingenter et al., 2010), this process is energized by Vacuole (V)-type ATPases (Jung et al., 2015; Schulz et al., 2011). Functional analysis showed that *TMT* activity in wild-type (WT) plants controls vacuolar sugar loading (Wingenter et al., 2010; Wormit et al., 2006). *AtTMT*

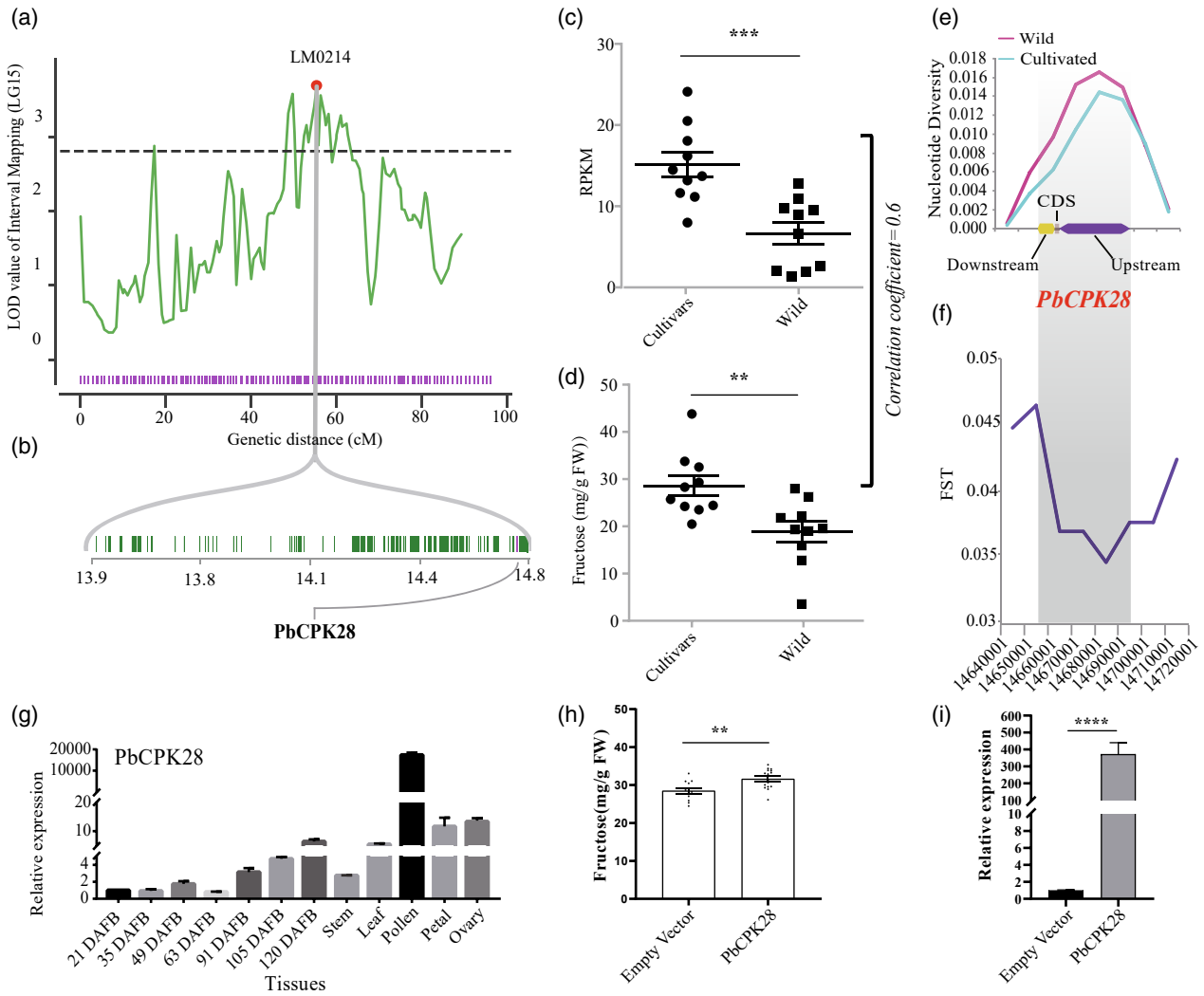
knockout lines accumulated less fructose and glucose during cold adaptation, suggesting that *AtTMTs* were required for cytosolic glucose homeostasis (Wormit et al., 2006). In watermelon, *CITST2* underlies a sugar accumulation quantitative trait locus (QTL) and contributes to the sugar content in fruit (Ren et al., 2018), and melon *CmTST2* is involved in increased sugar accumulation in melon fruit (Cheng et al., 2017). *BvTST2.1*, first identified as a sucrose-specific transporter, increases sugar yields in sugar beet (Jung et al., 2015). *TMT* is phosphorylated and activated by a mitogen-activated protein 3-kinase-like protein kinase (*VIK*) and *CIPK* (Deng et al., 2020; Wingenter et al., 2011), and *TMT* expression is regulated by sugar-induced transcription factor (*SUS1/WM1*; Ren et al., 2018). While the functional analysis of *TMT* genes has revealed a critical role in sugar transport, our knowledge of regulatory networks involving *TMT* expression is still limited in plants, especially concerning the functional relationship between *TMT* and the V-type ATPase.

In the field, sugar content is susceptible to environmental conditions and exhibits plasticity between plants. Therefore, it is more difficult to characterize sugar content than other agronomic features, such as fruit color. Recently, our group had constructed the high-density genetic map based on bin markers of pear, taking use of this map and two consecutive years of phenotypic data, we identified the QTLs for 18 important agronomic traits of pear fruits. Among them, QTLs for soluble solid content (SSC) were detected on several linkage groups (LGs), including LG1, LG5 and LG15, the QTLs on LG15 owed the highest logarithm of odds (LOD) value with the mapping analysis (Qin et al., 2022). However, which key genes regulate SSC in the QTL region remains unknown. Based on the QTL mapping and transcriptome association analysis, we identified a candidate gene *PbCPK28*, annotated as a calcium-dependent protein kinase, the function of this gene involving in sugar accumulation had never been reported. To determine the effects of candidate genes *PbCPK28* on sugar accumulation in pear fruits and uncover the regulatory mechanisms, the comprehensive genetic variation analysis, biochemical and physiological experiments were conducted. Our data demonstrated that *PbCPK28* could interact with and phosphorylate the vacuolar sugar transporter *PbTST4*, as well as vacuolar hydrogen proton pump *PbVHA-A1*, which could provide proton motive forces for *PbTST4* and promote transport of excess sugars into vacuoles for storage, eventually increasing sugar accumulation during pear fruit development.

## RESULTS

### QTL of SSC and map-based cloning of the sugar-related gene *PbCPK28*

To map the gene controlling sugar content of pear fruit, a preliminary genetic analysis of an F1 pear population



**Figure 1.** Map-based cloning and characterization of *PbCPK28*.

(a) Fine-mapping for soluble solids of pear fruit in the quantitative trait loci (QTL) validation population. Interval mapping (IM) methods were used to identify the QTL locus contributing to the soluble solids accumulation in pear fruit.  
 (b) The genes contained within LG15:13.9–14.8 Mb.  
 (c) *PbCPK28* gene expression of cultivar and wild Asian pear fruits.  
 (d) The fructose content of cultivar and wild Asian pear fruits.  
 (e, f) The comparison of nucleotide diversity ( $\pi$ ) and  $F_{ST}$  values between wild and cultivated Asian pears. The position is the physical position immediately upstream and downstream of the translation start and stop codon site of the *PbCPK28* gene.  
 (g) Expression of *PbCPK28* in different tissues of 'Dangshansuli' by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). Seven fruit stages of 'Dangshansuli' and five different pear tissues were sampled in 2018, including 21 DAFB (days after full blooming), 35 DAFB, 49 DAFB, 63 DAFB, 91 DAFB, 105 DAFB, 120 DAFB, Petal, Pollen, Ovary, Leaf, Stem.  
 (h) Transient transformation assays demonstrating the function of *PbCPK28* in pear fruit.  
 (i) qRT-PCR analysis after overexpressing *PbCPK28* in 'Dangshansuli' pear fruit at 135 DAFB.  
 (c, d, h, i) Error bars indicate standard error, and significant differences are indicated by asterisks, as determined by one-way ANOVA tests, with a significance level accepted at \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ .

consisting of 176 individuals derived from a cross between 'Niitaka' and 'Hongxiangsu' was performed by Qin et al. (2022). Interval mapping (IM) and a multiple QTL model were used to identify potential QTLs for fruit SSC of pear. As shown in Figure 1a, QTLs (Im0214) of SSC, which is an important index of the SSC of pear fruit, were significantly linked on LG15. Im0214 was located in the highest

LOD peak region, suggesting that a major genetic determinant of the SSC trait was located in this region. Then, candidate genes that surrounded (200 kb upstream and downstream) Im0214 were extracted for manual screening. This 0.87-Mb genomic interval contained 88 predicted protein-encoding genes (Figure 1b; Table S1). Previous reports showed that sugar transporter proteins possess a

high number of predicted phosphorylation sites, nearly all of which are located in the loop region (Endler et al., 2009), and one of these sites has been shown to be phosphorylated in native *AtTMT1* (Wingenter et al., 2010). Among those candidate genes, a calcium-dependent protein kinase gene located on plasma membrane (*Pbr034839.1*; *PbCPK28*; Figure S1) was selected for further analysis, which is a classical plant Ser/Thr protein kinase (Zhu et al., 2007).

Fruit sugar content is a trait of domestication (Wu et al., 2018). Fructose accounts for the greatest sugar content differential at maturity between cultivated and wild pears when comparing the contents of components of sucrose, glucose, fructose and sorbitol. Then, we compared the *PbCPK28* gene expression levels between cultivated and wild Asian pears. The expression levels of *PbCPK28* in cultivated pears were significantly higher than in wild pears (Figure 1c). The fructose content also exhibited significant differences between cultivated and wild Asian pear fruits (Figure 1d). Furthermore, there was a high correlation coefficient obtained when comparing gene expression and fructose content in pear fruits. To test whether the CDS of *PbCPK28* and its 2-kb promoter have been targeted during artificial selection during pear domestication, we further explored the evidence of selection on *PbCPK28* at the sequence level using nucleotide diversity ( $\pi$ ,  $F_{ST}$ ; Figure 1e,f) in Asian pears. Our results showed evidence of selection during domestication in a genomic study comparing populations of cultivated accessions and wild pear accessions (Figure 1e,f). The  $\pi$  and  $F_{ST}$  values all showed that *PbCPK28* was located in a selective sweep region, supporting that *PbCPK28* is probably a domestication-related gene.

Next, we performed quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) to analyze *PbCPK28* expression in different tissues. *PbCPK28* is highly expressed during pear fruit development, similar to fructose accumulation patterns (Figure 1g). Finally, we transformed the *35S::PbCPK28* construct into the pear fruits, and found that overexpression of *PbCPK28* led to significantly increased fruit fructose content in pear (Figure 1h,i).

#### A single-nucleotide polymorphism (SNP) in the promoter of *PbCPK28* is associated with fructose content variation in pear fruit

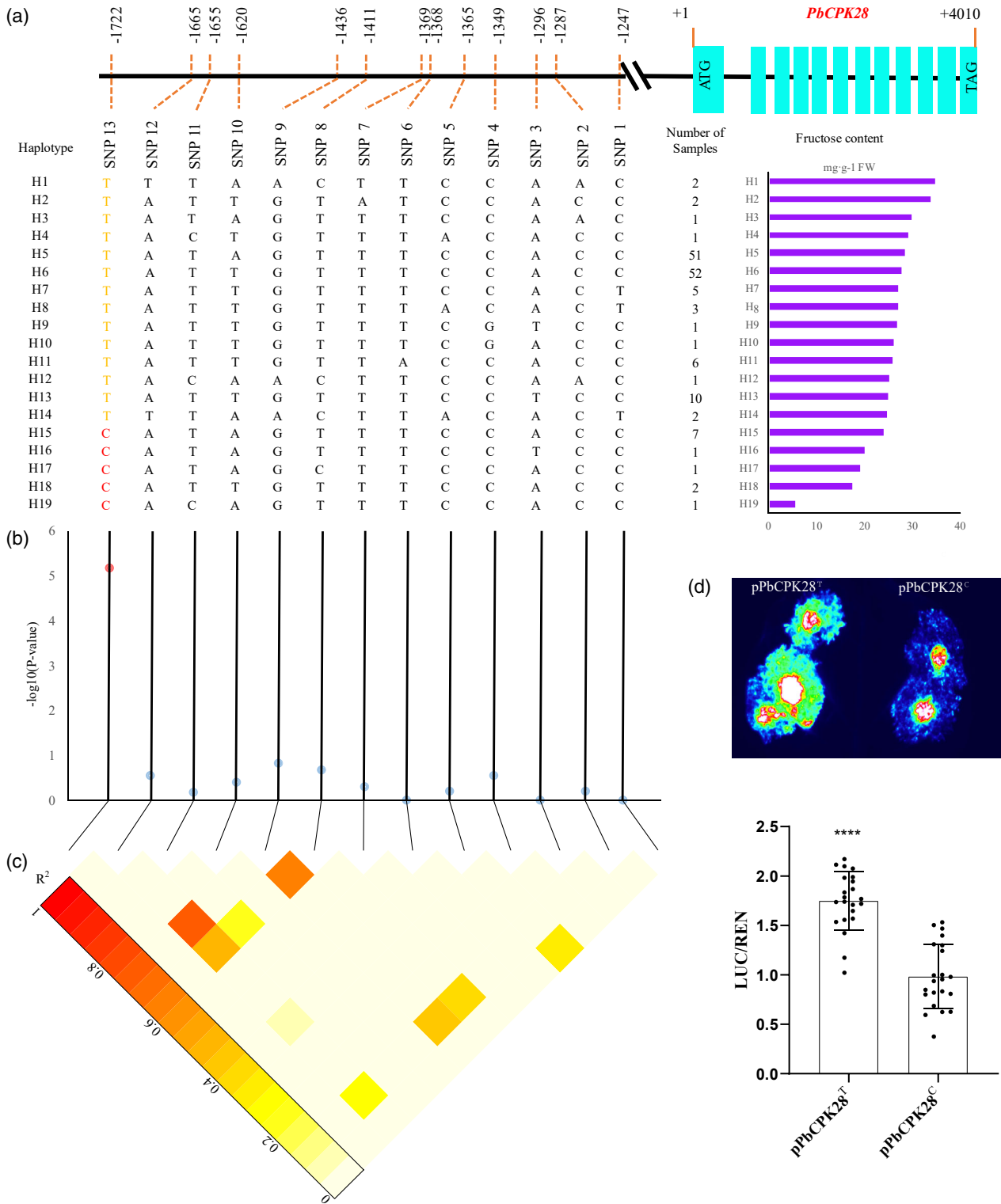
We found that the gene expression level of *PbCPK28* in different varieties was significantly different, suggesting that DNA variants in regulatory regions of *PbCPK28* may underlie their expression differences. To determine which mutation(s) in the promoter regulated gene expression and was associated with fructose content, we cloned and sequenced the *PbCPK28* promoter region of 150 pear cultivars, including nine wild species, and found that all the pear accessions or cultivars clustered into 19 haplotypes

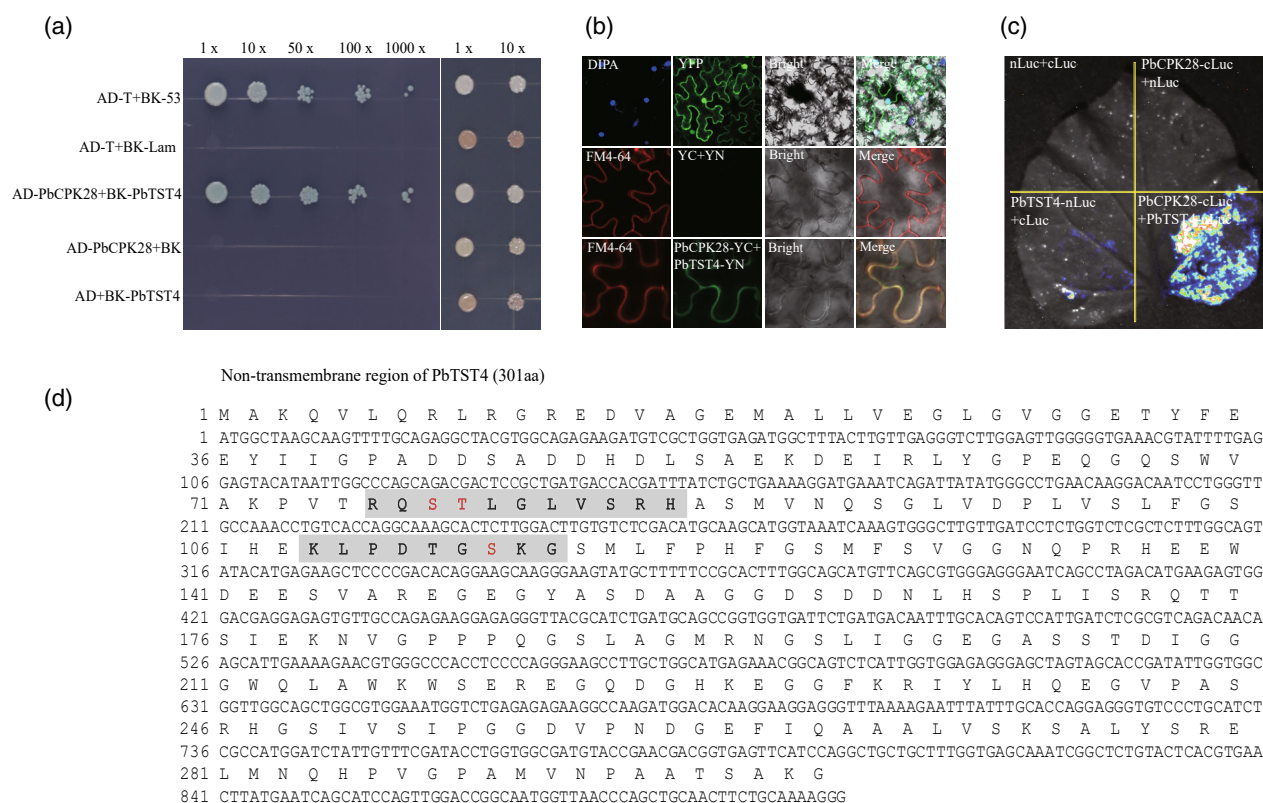
(H1–H19; Figure 2a; Table S2). The wild and cultivated haplotypes were indistinguishable, which might be due to the self-incompatibility of pears that encourages outcrosses. As a result, the genetic background of cultivated and wild accessions is highly complex due to frequent gene exchanges between wild and cultivated accessions. Subsequently, an association test with pear fruit fructose content and the 13 SNPs in the promoter of *PbCPK28* revealed the strongest signal at the SNP13 ( $P = 6.654 \times 10^{-6}$ ; Figure 2b,c), and medium signals at the SNP8, SNP9 and SNP12 (Figure 2b,c). Further sequence comparisons revealed that the nucleotide at SNP13 of low-fructose pears was C, and that of the high-fructose pears was T, suggesting that the genotypes at the SNP13 site affect the level of pear fruit fructose content. To probe whether SNP13 alleles contribute to differences in *PbCPK28* expression and subsequent differences in fructose content, we conducted transient expression assays with the two SNP13 variant-containing *PbCPK28* promoters in *Nicotiana benthamiana* leaves. The results showed that the pPbCPK28<sup>T</sup>-LUC reporter exhibited significantly higher expression levels than the pPbCPK28<sup>C</sup>-LUC reporter (Figure 2d), indicating that the T variant at the SNP13 site enhanced the *PbCPK28* promoter activity. As a whole, our findings indicate that the SNP13 variant in the *PbCPK28* promoter contributed to fruit fructose content variation in different pear accessions.

#### *PbCPK28* increases the sugar content through interaction with and phosphorylation of *PbTST4*

The above results indicated that *PbCPK28* could promote fructose content. However, the regulatory mechanism is unknown. To isolate proteins that interact with *PbCPK28* in pear fruits, the transcriptome database of 35 developmental fruit samples of different pear species (Zhang et al., 2016) was used to detect co-expression (high correlation coefficient) of genes with *PbCPK28*. Interestingly, one gene encoding a tonoplast membrane sugar transporter (*Pbr032130.1*) with high correlation with *PbCPK28* was identified, and here renamed as *PbTST4* (Table S3). Yeast-two-hybrid (Y2H) assays detected positive interactions between *PbTST4* and *PbCPK28* (Figure 3a). To determine which domain of *PbTST4* mediates its interaction with *PbCPK28*, a series of deletion constructs of *PbTST4* were made and then tested for interaction with *PbCPK28* using the Y2H assay, our data showed that *PbCPK28* interacted with the non-membrane region of *PbTST4* (Figure S2). Next, the split yellow fluorescent protein (YFP)-based bimolecular fluorescence complementation (BiFC) assay was used for confirming the interaction between *PbCPK28* and *PbTST4* in *N. benthamiana* leaves. A strong YFP fluorescence signal on the cell membrane indicated that *PbCPK28* interacted with *PbTST4* (Figure 3b). To validate the BiFC results, we performed luciferase complementation







**Figure 3.** PbCPK28 interacts with and phosphorylates the non-membrane region of PbTST4.

(a) PbCPK28 interacts with PbTST4 in yeast. Yeast colonies containing AD-PbCPK28 and BK-PbTST4 were grown on the -Trp, -Leu, -His, -Ade medium for 3 days.

(b) Bimolecular fluorescence (BiFC) assays showing the interaction between PbCPK28 and PbTST4. The indicated plasmid vectors were co-transformed into *Nicotiana benthamiana* leaves. Co-expression of PbCPK28-YCE and YNE, and PbTST4-YNE and YCE were used as negative controls. Scale bars: 120  $\mu$ m.

(c) Luciferase complementation imaging (LCI) assay showing the interaction of PbCPK28 and PbTST4 in *N. benthamiana* cells containing different pairs of constructs. Luminescent images were captured using a cooled charge-coupled device imaging apparatus.

(d) Identification of three phosphorylation sites of PbTST4 by PbCPK28 through liquid chromatographic tandem mass spectrometry (LC-MS/MS) analysis. The identified PbTST4 peptide sequences (in bold under the gray box), phosphorylation sites (in red).

imaging (LCI) assays. PbCPK28 was fused to the C-terminal fragment of the luciferase (cLUC), and PbTST4 was fused to the N-terminal end of the luciferase (nLUC). LUC activity was detected when *PbCPK28-cLUC* was co-expressed with *PbTST4-nLUC* in *N. benthamiana* leaves (Figure 3c). In contrast, the co-expression of *PbCPK28-cLUC* or *PbTST4-nLUC* with the negative controls, nLUC or cLUC, respectively, produced no detectable LUC activity.

To investigate the potential biochemical mechanism by which PbCPK28 phosphorylates PbTST4, we used a parallel *in vitro* kinase assay with PbCPK28 and PbTST4, and then used liquid chromatographic tandem mass spectrometry (LC-MS/MS) to determine the phosphorylation sites of the non-membrane region of PbTST4. Three amino acid residues including Ser and Thr were identified as target sites of PbTST4 by PbCPK28 with high confidence scores. The corresponding peptides are shown in Figure 3(d), and the three phosphorylation sites are highlighted in red and bolded (Ser277, Thr278 and Ser314), with the corresponding tandem spectra shown in Figure S3.

### PbTST4 uses the proton-motive force for loading sugars in pear fruit

In pear, *PbTST4* was previously identified as an important candidate gene that controls sugar content in pear fruit by genome-wide proteomic and transcriptomic analyses (Li, Huang, et al., 2015; Li, Zheng, et al., 2015). Combined with our study of *PbTST4* gene expression during pear fruit development, these data collectively support our hypothesis that *PbTST4* is an important gene that controls sugar content in pear fruit (Figure 4a).

First, to identify the transport function of *PbTST4*, the patch clamp technique was applied to intact human embryonic kidney 293T (HEK293T) cells transiently expressing *PbTST4*. First, both luminal (pipette) and cytosolic (bath) mediums were buffered to pH 7.5. When 50 mM glucose (Glc) and fructose (Fru) were applied to the bath solution (outside) of the HEK293T cell membrane, HEK293T cells harboring *PbTST4* responded with a strong current trace when compared with the background

current levels (Figure 4b,c). The current trace was more than 6 pA, indicating that Fru and Glc entered the cells at pH 7.5.

To mimic physiological proton gradients across the HEK293T cell membrane and to provide further evidence for proton coupling, we studied the pH dependency of the PbTST4-mediated sugar transport. The pipette medium was buffered to pH 5.5 and the bath medium was set to pH 7.5. when Fru and Glc were applied to the bath solution. HEK293T cells expressing *PbTST4* showed a higher current trace under asymmetrical pH conditions than under symmetrical pH conditions (Figure 4d,e). Accordingly, the sugar-induced current responses of PbTST4-decorated HEK293T cells were strongly promoted by a pH gradient oriented in the opposite direction to that of the sugar (Fru or Glc) gradient (Figure 4d,e). Our results indicated that the sugar-induced currents were carried by protons released from the vacuole. Furthermore, this demonstrates that both a sugar gradient and a pH gradient alone can drive the sugar/proton transport function of PbTST4. Our results thereby classify PbTST4 as a sugar-proton antiporter that couples the uphill import of sugars (from the cytosol into the vacuole) to the downhill export of protons (out of the vacuole into the cytosol). This process is energized by V-type ATPase activity (Jung et al., 2015), which is highly abundant in the vacuolar membrane during pear fruit development (Table S4). This leads to proton pump activity (Figure 4f) that is sufficient for the generation of the required proton motive force.

Finally, to assess the function of tonoplastic PbTST4 in living plants (Figure 4g), we stably transformed the 35S::*PbTST4* construct into strawberry cotyledons and generated several transgenic lines. Transgenic lines 3, 5 and 6 were selected for further analysis (Figure 4h,i). The mature stage fruit sugar content of three 35S::*PbTST4* transgenic strawberry lines were measured, and we found that the Fru and Glc contents in transgenic strawberry fruit were significantly higher than that of non-transgenic fruit (Figure 4j). Furthermore, 35S::*PbTST4* constructs were agroinfiltrated into 'Dangshansuli' pear fruit at 135 days after full blooming (DAFB) for transient expression. Five days post-infiltration, a significant increase in fructose content was observed at the infiltration sites of *PbTST4* overexpression (Figure 4k,l), but glucose content was not increased at these sites. This difference observed in strawberry might be due to the dynamics of sugar accumulation in fruit as a long-term process, where the gene functional efficiency in the transient expression system is lower than that of the stable transgenic plants. To characterize the biological function of phosphorylation sites in PbTST4, we overexpressed mutant variants of PbTST4 (with three phosphorylation-dead sites) in pear fruit at 135 DAFB. Notably, the functionality of the phosphorylation-dead variants of PbTST4 was decreased (Figure 4k,l). This

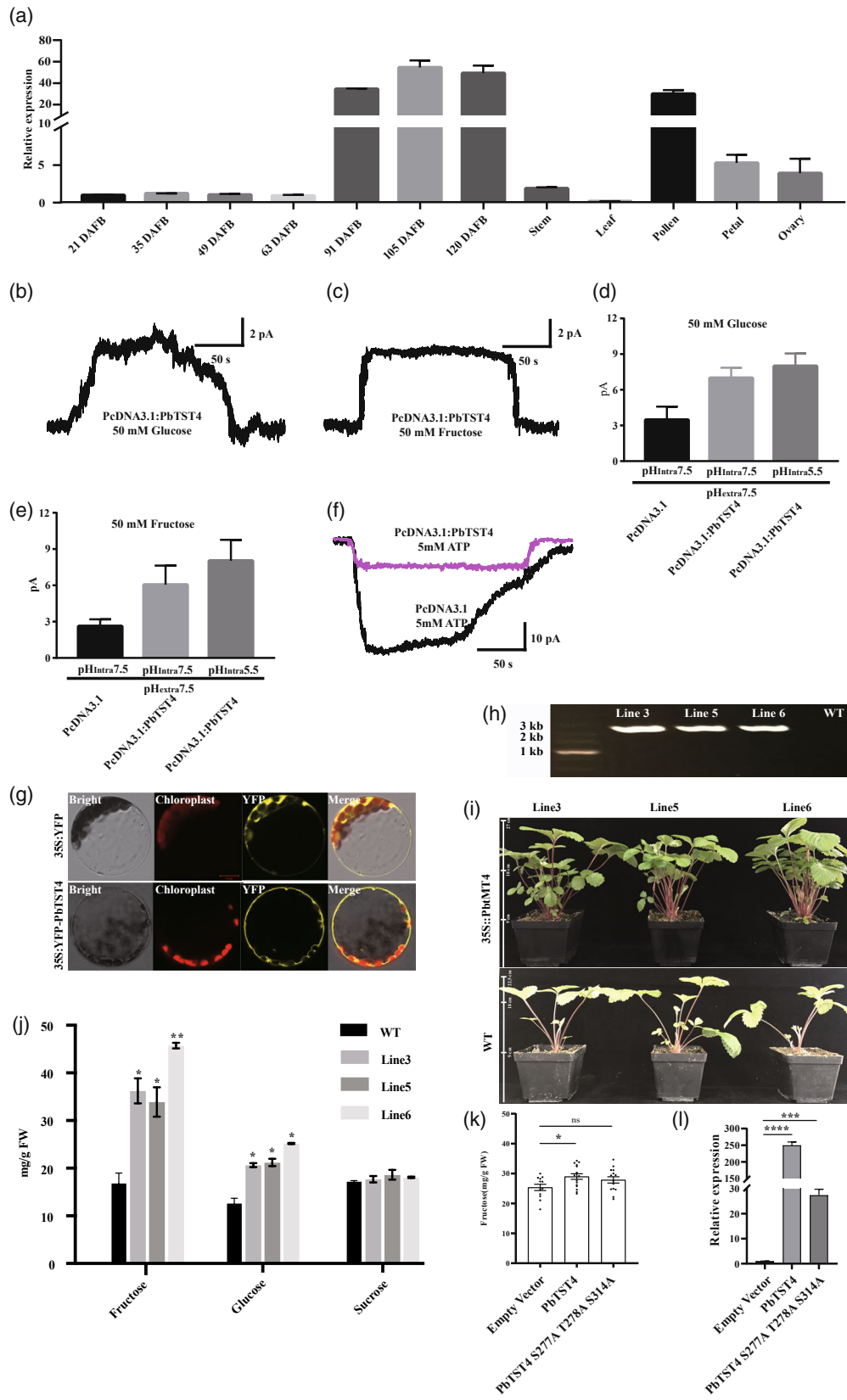
observation indicated that phosphorylation residues are critical for the transport capacity of PbTST4.

### PbCPK28 interacts with and phosphorylates PbVHA-A1 to consume adenosine triphosphate

Our previously described results suggest that PbTST4 is energized by the V-type ATPase enzyme. We revisited and mined the transcriptomic database of 35 development pear fruit samples from five pear cultivars, and a vacuolar ATP synthase subunit A gene (Figure 5a; *Pbr039217.1*, *PbVHA-A1*; Table S4), a component of V-type ATPase belonging to the V1 subfamily (Figure S4), was identified and selected for further analysis. qRT-PCR analysis of *PbVHA-A1* transcript levels at different fruit developmental stages and tissues showed that *PbVHA-A1* was upregulated, similar to *PbCPK28* and *PbTST4* (Figure 5b), suggesting that *PbVHA-A1* might provide energy for *PbTST4* during pear fruit development through interaction with and phosphorylation of PbCPK28 to participate in sugar accumulation. The interaction between PbCPK28 and PbVHA-A1 was demonstrated using the Y2H system (Figure 5c). Additionally, BiFC assays showed that PbCPK28-YCE, when co-expressed with PbVHA-A1-YNE into the tobacco leaves, caused a strong YFP fluorescence signal at the cell membrane, indicating that PbCPK28 interacted with the PbVHA-A1 (Figure 5d). The interaction was also detected by LCI assays (Figure 5e). A similar approach was used to determine whether PbCPK28 phosphorylates PbVHA-A1, which showed that PbCPK28 phosphorylates PbVHA-A1 (Figure 5f) at the three amino acid residues Thr11, Tyr120 and Thr622 with high confidence scores (Figure S5).

To determine the function of *PbVHA-A1* and *PbCPK28* in pH gradient generation, we over-expressed *PbVHA-A1* and *PbCPK28*, respectively, in apple calli to determine whether they could consume ATP to transport protons. The Ox-*PbVHA-A1* transgenic apple calli had significantly reduced ATP content when compared with that of WT; a similar result was also obtained in Ox-*PbCPK28* transgenic apple calli (Figure 5g). We also used the ATPase inhibitor (concanamycin A, ConcA) to treat the Ox-*PbCPK28* and Ox-*PbVHA-A1* transgenic apple calli, and found that the ATP content was significantly higher than WT and untreated Ox-*PbCPK28* and Ox-*PbVHA-A1* transgenic apple calli (Figure 5g). Taken together, our results indicate that *PbVHA-A1* could consume ATP and provide a pH gradient for both sides of the tonoplast, and that *PbCPK28* interacts with and phosphorylates *PbVHA-A1* to increase ATPase enzyme activity, leading to higher ATP consumption.

Next, the 35S::*PbVHA-A1* constructs were agroinfiltrated into 'Dangshansuli' pear fruit at 135 DAFB. Five days after infiltration, a significant increase in fructose content was observed at the infiltration sites where *PbVHA-A1* overexpression was prominent (Figure 5h). To characterize the biological function of phosphorylation residues in





**Figure 4.** Molecular characterization of PbTST4.

- (a) Expression profiles of *PbTST4* in different pear tissues and fruit developmental stages by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR).
- (b, c) Incubation with 50 mM glucose (b) and fructose (c) solutions induced inward current responses in HEK293T cells expressing *PbTST4*. HEK293T cells only did not exhibit a significant change in current in response to incubation of 50 mM glucose and fructose.
- (d, e) 50 mM glucose in (d), fructose in (e)-induced current responses determined under different HEK293T cell pH conditions. Both pH conditions were significant ( $P < 0.05$ ).
- (f) Incubation with 5 mM ATP solutions induced inward current responses in HEK293T cells expressing pcDNA3.1-PbTST4 construct and HEK293T cells transfected with the pcDNA3.1 vector.
- (g) Subcellular localization of the over-expressed YFP-PbTST4 fusion protein in protoplasts from *Arabidopsis* suspension culture.
- (h) RT-PCR analysis of three 35S::*PbTST4* transgenic lines and the wild-type (WT) control.
- (i) Strawberry plants were grown on soil in a growth chamber with 16 h light at 25°C and 8 h dark at 22°C.
- (j) Sugars content of transgenic and WT strawberry fruits.
- (k) Transient transformation assays demonstrating the function of PbTST4 and PbTST4 mutation in pear fruit.
- (l) The qRT-PCR analysis after overexpressing *PbTST4* in 'Dangshansuli' pear fruit at 135 days after full blooming (DAFB).
- (j–l) Error bars indicate standard error, and significant differences are indicated by asterisks, as determined by one-way ANOVA tests, with a significance level accepted at \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ .

*PbVHA-A1*, we expressed mutant variants of *PbVHA-A1* (three sites phosphorylation-dead) in pear fruit at 135 DAFB. Notably, the functionality of the phosphorylation-dead variants of *PbVHA-A1* was decreased (Figure 5i). This observation indicated that phosphorylation residues are critical for the function of *PbVHA-A1*.

#### The CPK-TST/CPK-VHA-A regulation network was conserved in plant

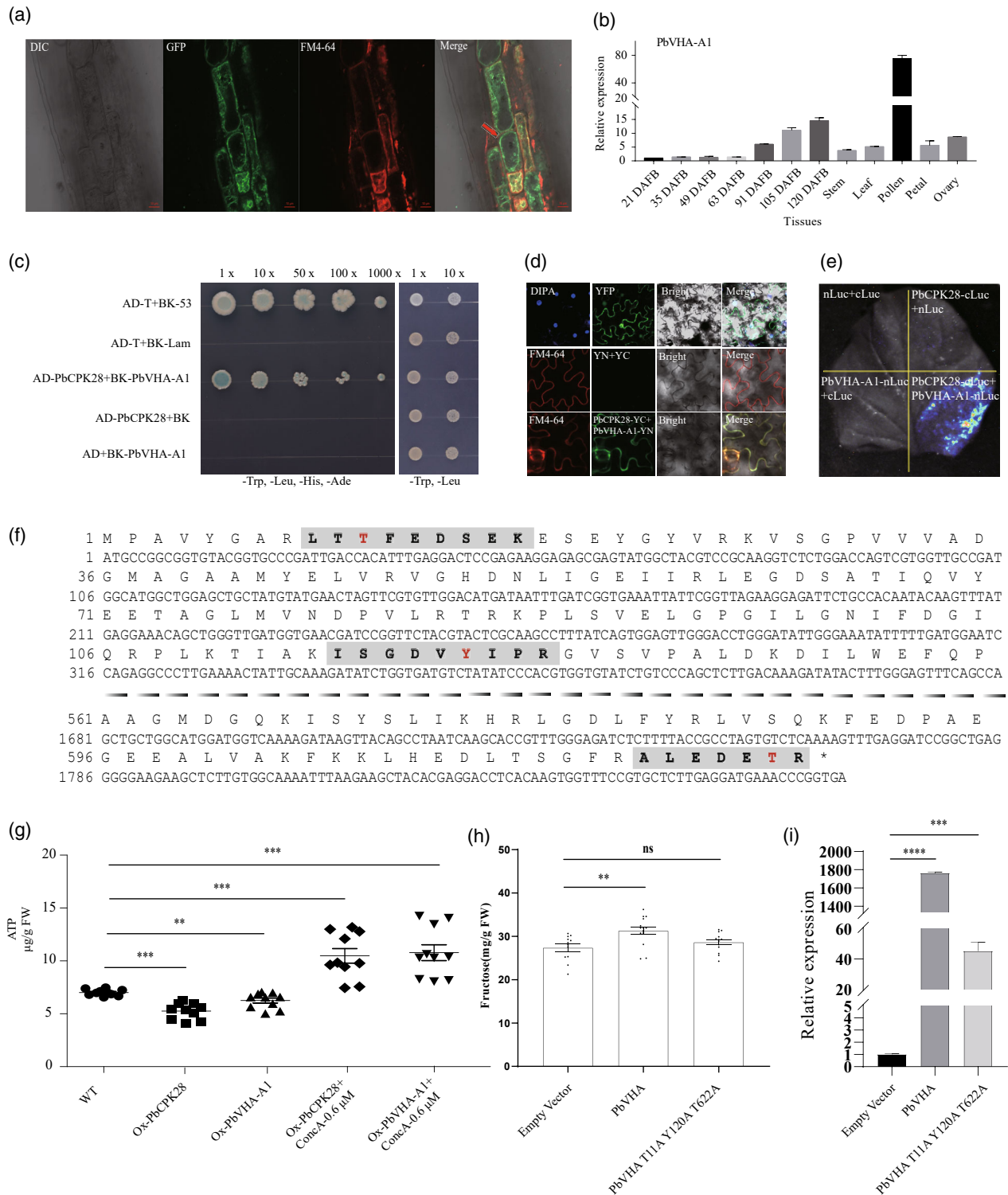
Here, we found that PbCPK28 interacted with and phosphorylated PbTST4 and PbVHA-A1. Then, we investigated whether regulation through three phosphorylation sites of PbTST4 (S277, T278 and S314) and PbVHA-A1 (T11, Y120 and T622) are conserved in other plant TST and VHA-A proteins. Firstly, we constructed a phylogenetic tree of CPKs from 65 sequenced plant genomes (<https://phytozome-next.jgi.doe.gov>). Plant CPKs form four clades, and PbCPK28 belongs to Clade IV (Figure 6a). Based on the protein sequence alignments, we found that the Serine/Threonine catalytic domain of CPKs was highly conserved in plants (Figure 6b). Furthermore, the amino acid sequences surrounding a phosphorylation site might be conserved for recognition of the site by compatible kinases (Miller & Turk, 2016). Accordingly, we aligned the TST and VHA-A orthologous genes, which belong to the same clade of plants, and found that the identification of phosphorylation sites corresponding to S277, T278, S314 in TST (Figures 6c and S6), and T11, Y120 T622 in VHA-A (Figures 6d and S7), thereby exhibiting conservation of amino acid motifs surrounding the phosphorylation site, respectively. Among them, S277 and S314 in TST or T11 and Y120 in VHA-A are highly conserved in all plants, implicating that those phosphorylation sites are essential for TST and VHA-A activity, and that the CPK-TST/CPK-VHA-A regulatory network might be conserved in higher plants. To test the hypothesis that the CPK-TST/CPK-VHA-A regulatory networks were conserved in other plants, the orthologs of CPK, TST and VHA-A in tomato were isolated, and their interactions between CPK-TST and

CPK-VHA-A were similarly observed and shown by the Y2H (Figure 6e,f).

#### DISCUSSION

Crop domestication is the genetic modification of wild species through human selection, a process enabled by years of selection of numerous genomic loci (Doebley et al., 2006). Identification of genes involved in domestication, especially the genes related to the fruit quality, yield and stress response, will help us accelerate the process of domesticating new crops. Genetic modification in the cis-regulatory region can alter transcription levels and patterns of gene expression. Such modifications underlie trait transitions. In rice, a SNP in the 5' regulatory region of the *qSH1* gene caused loss of seed shattering owing to the absence of abscission layer formation (Konishi et al., 2006). A SNP in the cis-regulatory region of the *OsLG1* gene led to a compact panicle architecture in cultivars during rice domestication (Zhu et al., 2013). Three SNPs in the *OsTIG1* promoter of indica cultivars of rice resulted in decreased expression of *OsTIG1* in the adaxial side of the tiller base and reduced cell length and tiller angle, leading to the transition from inclined tiller growth in wild rice to erect tiller growth during rice domestication (Zhang et al., 2019). We sequenced the *PbCPK28* promoter region of 150 pear cultivars, and they clustered into 19 haplotypes. The association results showed that a SNP (T/C) present in the cis-regulatory region of *PbCPK28* was associated with the fructose content in the natural pear population (Figure 2a). The C to T SNP at the 1722-bp site of the cis-regulatory region affected *PbCPK28* expression levels (Figure 2d), as well as the sugar content in pear fruit (Figure 1c). Our results provide another example of the selection of mutations in the cis-regulatory region during crop domestication.

In previous reports, CPK was a classical plant Ser/Thr protein kinase (Zhu et al., 2007), with roles in rapid abiotic stress and immune signaling responses (Delormel & Boudsocq, 2019). There were few reports of the involvement of CPK in long-term adaptive processes or plant



development. In *Arabidopsis*, 34 CPK genes were identified (Hrabak et al., 2003), and 12 of them are predominantly expressed in pollen (Myers et al., 2009). Among them, *AtCPK17*, *AtCPK28*, *AtCPK11* and *AtCPK24* play important roles in pollen tube fitness and growth (Myers et al., 2009;

Zhao et al., 2013). The shaker pollen K<sup>+</sup> in the channel may act as the CPK target protein (Zhao et al., 2013). Recently, a CPK32-cyclic nucleotide-gated channel 18 might regulate the pollen tube growth by increasing Ca<sup>2+</sup> accumulation (Zhou et al., 2014), and CPK4/5/11 phosphorylated the N-

**Figure 5.** PbCPK28 interacts with and phosphorylates PbVHA-A1.

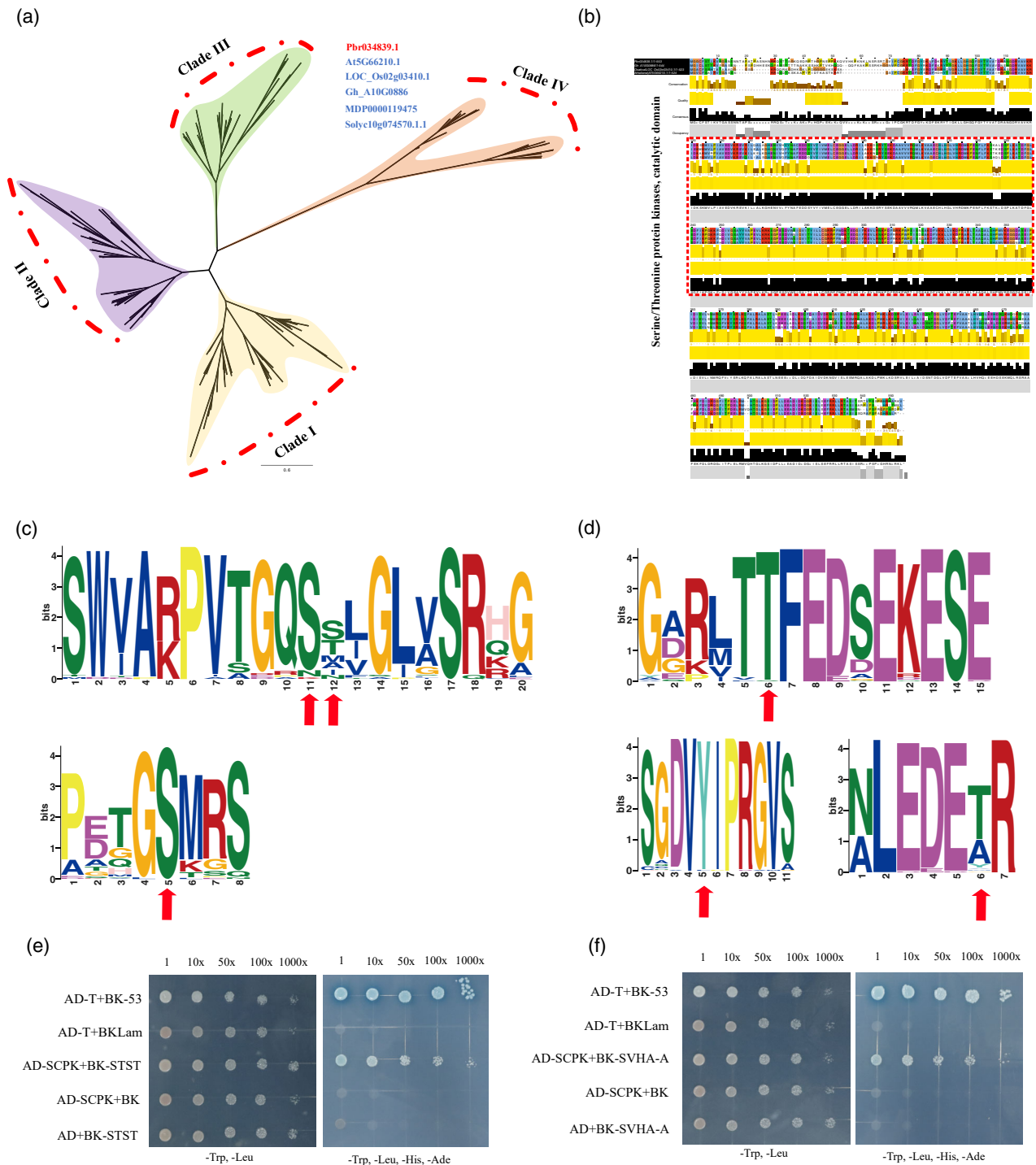
- (a) Subcellular localization of *PbVHA-A1*. The *PbVHA-A1*-GFP fusion gene under the control of *CaMV35S* promoter was expressed in transgenic *Arabidopsis* root cells.
- (b) Expression of *PbVHA-A1* in different tissues of 'Dangshansuli' fruit by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR).
- (c) PbCPK28 interacts with PbVHA-A1 in yeast. Yeast colonies containing AD-PbCPK28 and BK-PbVHA-A1 were grown on the -Trp, -Leu, -His, -Ade medium for 3 days.
- (d) Bimolecular fluorescence (BiFC) assays depicting the interaction between PbCPK28 and PbVHA-A1. The indicated plasmids were co-transformed into *Nicotiana benthamiana* leaves. Co-expression of PbCPK28-YCE and YNE, PbVHA-A1-YNE and YCE were used as negative controls. Scale bars: 120  $\mu$ m.
- (e) Luciferase complementation imaging (LCI) assay showing the interaction of PbCPK28 and PbVHA-A1. *Nicotiana benthamiana* cells containing different pairs of constructs. Luciferase images were captured using a cooled charge-coupled device imaging apparatus.
- (f) Identification of three phosphorylation sites of PbVHA-A1 by PbCPK28 through liquid chromatographic tandem mass spectrometry (LC-MS/MS) analysis. The identified PbVHA-A1 peptide sequences are in bold under the gray box and the phosphorylation sites are in red.
- (g) The ATP concentration in transgenic apple calli by high-performance liquid chromatography. Concanamycin A (ConcA)-0.6  $\mu$ M indicates 0.6  $\mu$ M ConcA treatment on Ox-PbVHA-A1 and Ox-PbCPK28 transgenic apple calli.
- (h) Transient transformation assays demonstrating the function of PbVHA-A1 and PbVHA-A1 mutation in pear fruit.
- (i) The qRT-PCR analysis after overexpressing *PbVHA-A1* in 'Dangshansuli' pear fruit at 135 days after full blooming (DAFB).
- (g-i) Error bars indicate standard error, and significant differences are indicated by asterisks, as determined by one-way ANOVA tests, with a significance level accepted at \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ .

terminal domain of MTP8 primarily at Ser31 and Ser32 residues to regulate Mn transporters (Zhang et al., 2021). However, there has been no report of CPK in sugar accumulation. In *Arabidopsis*, *AtCPK28* is a regulatory component that controls stem elongation and vascular development. Two independent loss-of-function mutants of *AtCPK28* show a reduction of stem elongation accompanied by shorter leaf petioles and enhanced anthocyanin levels, indicating that *AtCPK28* encodes a regulator for coordinated stem elongation and secondary growth (Matschi et al., 2013). Additionally, loss-of-function mutants of *AtCPK28* exhibited a NaCl- and mannitol-sensitive phenotype in green cotyledons (Gao et al., 2014). *AtCPK28*-overexpressing plants displayed stronger tolerance to NaCl and mannitol stresses than WT plants (Gao et al., 2014). Similar results were also found in rice, where *OsCPK4* (a homolog of *AtCPK28*) functions as a positive regulator of salt and drought stress responses in rice (Campo et al., 2014). Interestingly, *AtTMT1* and *AtTMT2* expression is induced by drought, salt and cold treatment. *TMT* loss-of-function mutant plants showed lower glucose and fructose accumulation when compared with WT plants, indicating that *TMTs* play a major role during stress responses (Wormit et al., 2006). The *TST* loss-of-function mutants also showed slower growth than WT (Wingenter et al., 2010) and had a similar phenotype to *CPK28* loss-of-function mutants (Gao et al., 2014). The phenotype of *AtCPK28* loss-of-function mutants could be recovered by gibberellic acid ( $GA_3$ ) treatment (Matschi et al., 2013), which might be due to  $GA_3$ -enhancement of sugar transporter expression and increased sugar accumulation (Murcia et al., 2016). Sugars are required for the synthesis of plant cell walls and carbohydrate metabolism, such that plants can grow stronger and faster in a high-sugar liquid medium (Wingenter et al., 2010). The loss-of-function mutants of *TST* and *CPK28* led to slower growth, hinting that the two genes might act in the same regulatory network. Here, we determined that *PbCPK28*, an ortholog of

*AtCPK28*, could interact with the non-transmembrane region of *TST* in the cytoplasm to increase *TST* protein activity and enhance sugar transport for sugar accumulation in pear fruit, providing insights into their molecular interaction and coordinated activity in other growth and developmental processes.

Plant *TST*, as a tonoplast sugar transporter, plays an important role in sugar accumulation (Cheng et al., 2017; Cho et al., 2010; Jung et al., 2015; Ren et al., 2018; Schulze et al., 2012; Wingenter et al., 2010, 2011; Wormit et al., 2006). In sugar beet taproots, *TST* acts as a sucrose-specific proton antiporter that contributes to increased sugar yields and could be energized by V-type ATPases that provide a pH gradient (Jung et al., 2015). Additionally, in watermelon and melon, *TST* expression can promote sugar accumulation in fruits (Cheng et al., 2017; Ren et al., 2018). Our result demonstrates that both the sugar gradient and pH gradient alone could drive the sugar/proton transport function of *PbTST4*. As an antiporter, *PbTST4* couples the uphill import of sugars into the vacuole with the downhill export of protons to the cytosol. The former property being the biophysical prerequisite for accumulating high amounts of sugars in pear fruit as the pH gradient was established by the V-type ATPase (Jung et al., 2015).

All plasma membrane SUC/SUT sucrose transporters catalyze a proton symport and then allow sugar loading from the apoplast. Similar results were also found in SUT4, which is a tonoplast sucrose exporter and a proton symport (Eom et al., 2011; Schneider et al., 2012; Schulz et al., 2011). Previous reports showed that sugar transporters (such as SUC and TST) were  $H^+$  dependent (Eom et al., 2011; Jung et al., 2015; Ren et al., 2018; Schulz et al., 2011). However, there has been no evidence indicating a direct relationship between TST and V-type ATPase. We showed that *PbCPK28* interacted with *PbVHA-A1* to enhance *PbVHA-A1* enzyme activity and provide a pH gradient across the tonoplast. This was the first indication of a relationship between sugar accumulation and pH gradient



**Figure 6.** Conservation of CPK-TST-VHA-A in plant species.

(a) Alignment of PbCPK28 clade IV, showing conserved domains of CPK from apple, sorghum, maize and tomato. The conserved domain in the red box indicated the Serine/Threonine protein kinase catalytic domain.

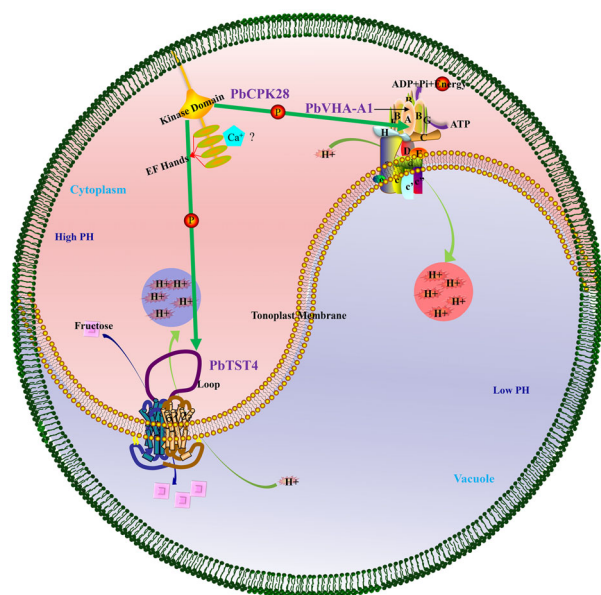
(b) Conservation of the amino acid sequence surrounding the phosphor-sites of TST in plants. Arrows indicate the position of the phosphor-sites.

(c) Conservation of the amino acid sequence surrounding the phosphor-sites of VHA-A in plants. Arrows indicate the position of the phosphor-sites.

(d) SICPK (SI means tomato) interacts with SITST in yeast. Yeast colonies containing AD-SICPK and BK-SITST were grown on the -Trp, -Leu, -His, -Ade medium for 3 days.

(f) SICPK interacts with SIVHA-A in yeast. Yeast colonies containing AD-SICPK and BK-SIVHA-A were grown on the -Trp, -Leu, -His, -Ade medium for 3 days.





**Figure 7.** A model of the *PbCPK28* regulatory network in pear fruit.

due to the function of *PbCPK28* in the phosphorylation of both *TST* and the *V*-type *ATPase* subunit.

We presented additional evidence indicating that *CPK*, *TST* and *VHA-A* genes are highly conserved in the plant kingdom. Therefore, similar regulatory mechanisms of sugar accumulation discovered here could operate in other plants. This is further supported by Y2H data investigating tomato homologs (Figure 6e,f) interaction with the *SITST1* and *SITST2* genes to regulate sugar content in tomato fruit (Zhu et al., 2021).

In summary, our identification and characterization of *PbCPK28*, *PbTST4* and *PbVHA-A1* genes revealed the molecular basis of fruit sugar accumulation in pear (Figure 7). The regulating mechanism of sugar accumulation not only provides genetic resources to develop pears with enhanced fruit qualities, but also provides a reference to improve sugar content for other fruit crops.

## EXPERIMENTAL PROCEDURES

### Cultivation of plants and seedlings

Strawberry [*Fragaria vesca* ('YW5AF7')] was grown in a growth chamber with 16 h light at 25°C and 8 h dark at 20°C, with illumination at approximately 100  $\mu\text{mol m}^{-2} \text{sec}^{-1}$  illumination. *Nicotiana benthamiana* plants were grown in a growth chamber with 16 h light at 23°C. Plants were grown until they had more than six leaves. *Arabidopsis* seeds were sterilized with 10% bleach before being sown on half-strength Murashige and Skoog basal medium (1/2 MS; Sigma-Aldrich, St. Louis, MO, USA) with 1% sucrose. *Arabidopsis* seeds were kept at 4°C for 3 days before being moved to white light illumination for 8 h at 22°C. After the seedlings developed two leaves, they were transferred to soil and grown in a growth chamber with 8 h light at 22°C.

### Plant materials for gene expression studies

The cultivar 'Dangshansuli' used for qRT-PCR (*Pyrus bretschneideri* Rehder) was collected from orchard of Fengxian in Jiangsu province. Six different tissues and seven fruit stages of 'Dangshansuli' were collected from the same trees, including petal, receptacle, style, anther, leaf, stem and fruit of 21, 35, 49, 63, 91, 105, 120 DAFB. The samples were frozen in liquid nitrogen immediately after collection and then stored at -80°C. The 135 DAFB of 'Dangshansuli' fruit were used for transient transformation. All materials were collected for three biological replicates.

### Identification of QTLs related to SSC in pear

The SSC of 176 progenies from the hybridization of 'Niitaka' (*Pyrus pyrifolia*) and 'Hongxiangsu' (a hybrid of *P. bretschneideri* × *Pyrus sinkiangensis*) were recorded. Genetic maps, phenotype data and methods for the identification of QTLs associated with SSC were described previously (Qin et al., 2022). QTL mapping was performed using MapQTL v6.0 with the Kruskal–Wallis (KW) test and IM methods. For IM mapping method, QTLs with the mean values higher than LOD = 2.5 were considered significant. For the KW test, 'Signif.' < 0.005 (\*\*\*\*) was considered significant. Candidate genes associated with SSC were identified by annotation 200 kb upstream and downstream of identified QTLs.

### Genetic diversity analysis

The SNP data were collected from our published database (Wu et al., 2018) and used for the genetic diversity analysis of *PbCPK28* in pears. The SNPs with missing data > 5% or minor allele frequency < 5% were filtered out.  $F_{ST}$  values were calculated with a 30-k to 20-k sliding window by VCFtools software to analyze the pairwise genomic differentiation for cultivated and wild populations of pear (Table S2).

### Strawberry germline transformation and pear fruit transient expression

A full-length CDS of *PbTST4* was inserted into the Gateway vector pCR8/GW/TOPO, and then via LR recombined into the Plant Gateway vector pMDC32 containing the cauliflower mosaic virus (CaMV) 35S promoter. The recombinant plasmid was introduced into *Agrobacterium tumefaciens* strain GV3101 and then used for plant transformation. Hygromycin-resistant plants derived from independent transformation events were identified through PCR using the primers 35S-F and *PbTST4*-R and by sequencing.

Strawberry transformation and regeneration were based on previously published protocols (Caruana et al., 2018; Kang et al., 2013). The cotyledons from 'YW5AF7' plants were co-cultivated with *Agrobacterium* GV3101 harboring the pMDC32::PbTST4 construct in MS salt agar medium, at pH 5.8, and supplemented with 2% sucrose, 3.4 mg L<sup>-1</sup> benzyladenine, 0.3 mg L<sup>-1</sup> indole-3-butyric acid and 0.7% agar. After 3 days of co-cultivation, cotyledon explants of 'YW5AF7' were washed to remove *Agrobacterium* and transferred to the same new agar media plus 250 mg L<sup>-1</sup> timentin and 250 mg L<sup>-1</sup> carbenicillin for 1 week in the dark without hygromycin. Then, the cotyledon explants of 'YW5AF7' were cultured in the same media with 2 mg L<sup>-1</sup> hygromycin and transferred to fresh medium containing the same ingredients every 2 weeks until shoot emergence. The shoots were transferred to rooting media (0.1 mg L<sup>-1</sup> indole-3-butyric acid, 2% glucose, 0.5 × MS salts, 0.7% agar, pH 5.8). The rooted plants were transferred to soil and genotyped by the primers 35S-F and *PbTST4*-R, respectively. Positive lines were confirmed by PCR.

To transiently overexpress *PbCPK28*, *PbTST4* and *PbVHA-A1*, the full-length coding sequences of *PbCPK28*, *PbTST4* and *PbVHA-A1* were constructed in pEarleyGate 104 to form the fusion vector. The overexpression constructs were transferred into GV3101 cells. Then, these cells were centrifuged, re-suspended in infiltration buffer (10 mM MgCl<sub>2</sub>; 10 mM MES, pH 5.7; 200 µM acetosyringone) to OD<sub>600</sub> 0.9–1.0 and maintained at 22°C for 4 h. The cells were agroinfiltrated into ‘Dangshansuli’ pear fruits at 135 DAFB using needleless syringes. Six fruits were injected for each construct in an experiment. The transformed fruits were placed in the dark at 22°C overnight and then incubated in a growth chamber at 22°C under a 16-h photoperiod for 5 days before being examined. Each experiment was performed with three biological replicates.

### Patch-clamp analyses of transformed HEK293T cells

The transient overexpression of *PbTST4* was under the control of the human cytomegalovirus immediate-early promoter in HEK293T cells. The technique previously used to identify proton-coupled sugar antiport activity catalyzed by *TST* in *Arabidopsis*, sugar beet and watermelon was used (Jung et al., 2015; Ren et al., 2018; Wingenter et al., 2010). Briefly, HEK293T cells were grown in DMEM medium (Gibco, Grand Island, NE, USA) with 10% fetal calf serum, 10% fetal bovine serum, 50 µg ml<sup>-1</sup> streptomycin and 50 µg ml<sup>-1</sup> penicillin (Invitrogen Life Technologies, Carlsbad, CA, USA) at 37°C. The cells were transiently transfected using XtremeGENE HP DNA Transfection Reagent (Sigma-Aldrich) according to the manufacturer’s protocol. After transient transfection, all cells were incubated for 48 h before electrophysiological measurement. The electrophysiological recordings were obtained under visual control with a microscope, and the amplifier was used to record the electrophysiological signal. The bath and pipette solutions were identical in solute composition (45 mM NaCl, 2.5 mM KCl, 1 mM MgCl<sub>2</sub>·6H<sub>2</sub>O, 2 mM CaCl<sub>2</sub>, 1 mM D-glucose, 10 mM HEPES), with the bath medium adjusted to pH 7.5 and the pipette solution to pH 5.5 or 7.5. ATPase-mediated proton pump currents were stimulated by ATP (5 mM) application to the outside (bath solution) of the HEK293T cell. The data were stored and analyzed with Patchmaster software. All of the experiments were performed at room temperature (25°C), and three biological replicates were conducted.

### Y2H assays

To analyze the interaction between *PbCPK28* and *PbTST4* or *PbVHA-A1*, the CDS of *PbCPK28* was cloned into the pGADT7 (AD) vector. Additionally, the fourth non-transmembrane region of *PbTST4* and the coding sequence of *PbTST4* were cloned into the pGBKT7 (BK) vector. The resulting vectors were co-transformed into yeast strain AH109 via the lithium acetate method. The transformed cells were grown on SD medium without Leu, Trp, His and Ura, but containing X-α-Gal (50 µg ml<sup>-1</sup>). All of the experiments were performed with three biological replicates.

### BiFC assay

cDNA of *PbCPK28* without its stop codon was amplified using PCR and cloned into the *KpnI* site of pSPYCE155 to construct *PbCPK28-YCE*. The cDNAs of *PbTST4* and *PbVHA-A1* without their stop codons were cloned into pSPYNE173 using *KpnI* and *XhoI* to construct *PbTST4-YNE* and *PbVHA-A1-YNE*, respectively. *Agrobacterium tumefaciens* strain GV3101 was transformed with the above vectors or the control vector, and then infiltrated into young leaves of *N. benthamiana*. Plants were grown in a growth

chamber with 16 h light at 25°C and 8 h dark at 23°C for 3 days. Fluorescence signals were observed with a confocal microscope (Carl Zeiss, Germany). All of the experiments were performed with three biological replicates.

### Luciferase fluorescence complementation assay

The LCI assays were carried out as follows. *Agrobacterium tumefaciens* strain GV3101 was transformed with the vectors or the control vector, and then infiltrated into young leaves of *N. benthamiana*. After infiltration, plants were cultured in a growth chamber at 23°C for 48 h at 16 h light–8 h dark photoperiod. The infiltrated leaves were sprayed with 1 mM luciferin (Sigma-Aldrich St. Louis, MO, USA) in darkness for 10 min before detection of luminescence. Images of luminescence were captured by a low-light cooled charge-coupled device imaging apparatus (Lumazone Pylon 2048B; Princeton). All of the experiments were performed with three biological replicates.

### Dual-luciferase transcriptional activity assay

To construct the *pPbCPK28<sup>T</sup>-Luc* and *pPbCPK28<sup>C</sup>-Luc* plasmids, we amplified the promoter sequences of *PbCPK28* from ‘Dangshansuli’ and then used the primer overlapping method to change the specific T to C to obtain the *pPbCPK28<sup>C</sup>* promoter. The resulting promoter was cloned into the pGreen II-0800-Luc vector and used as reporters. The Luc activities were determined using a dual-luciferase reporter assay system (Promega, USA). Relative luciferase activities of treatment samples were calculated by normalizing against the activity of control samples. Each experiment was performed with three biological replicates.

### Production of *PbTST4*, *PbCPK28* and *PbVHA-A1* recombinant proteins in *Escherichia coli*

The coding sequences of *PbTST4*, *PbCPK28* and *PbVHA-A1* were cloned into the pCold-TF vector to construct His-tagged recombinant vectors. The vectors were transformed into *E. coli* BL21, and the strain was cultivated at 37°C until OD<sub>600</sub> = 0.6–0.8. The culture was incubated at 4°C for 1 h, and then incubated at 15°C for 24 h after the addition of 0.5 mM isopropyl-β-D-thiogalactoside. Cells were collected in the tube by centrifuging at 2800 g, 4°C, and resuspended in phosphate-buffered saline (PBS; pH 7.4), and underwent ultra-sonication until the solution became transparent. Next, the solution was centrifuged at 4°C (18000 g, 10 min), and the supernatant was stored in tubes on ice. Recombinant proteins were assessed by 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). To purify the recombinant protein, the supernatant was loaded onto a 10-ml column containing 1 ml Ni–NTA His Bind (Merck & Millipore); the velocity of the constant flow pump linked to the column base was 100 µl min<sup>-1</sup>. Protein was washed with 20 ml of pH 7.9 buffer solution containing 5 mM imidazole, 200 mM NaCl and 20 mM Tris–HCl. Elution of 5 ml target protein was achieved using pH 7.9 elution buffer containing 200 mM imidazole, 200 mM NaCl and 20 mM Tris–HCl. The eluent was dialyzed in PBS at 4°C three times with centrifugation. Finally, the remaining 100–200 µl was stored at –80°C.

### Kinase assays for mapping phosphorylation sites

*In vitro* kinase assays were performed according to Rudd et al. (1996) description. Briefly, recombinant His-tagged proteins were incubated at 30°C with 1 µg *PbCPK28* and 1 µg *PbTST4* (or *PbVHA-A1*) in 100 µl of a kinase buffer containing 50 mM Tris–HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub> and 1 mM ATP. After 30 min,

the mixture was loaded on a 10% SDS-PAGE gel. Aliquots were digested with Filter-aided sample preparation and analyzed using LC-MS/MS.

### Identification of *in vitro* phosphorylation sites of PbTST4 and PbVHA-A1 by LC-MS/MS

The proteins samples were analyzed by a Q Exactive mass spectrometer (Thermo Fisher Scientific) connected to an Easy NanoLC 1000 (Thermo Fisher Scientific) for a total run time of 60 min. The peptide mixture was loaded onto a reverse-phase trap column (Thermo Scientific Acclaim PepMap100, 100  $\mu\text{m}^2$  cm, nanoViper C18) connected to the C18 reverse-phase analytical column (Thermo Scientific Easy Column, 10 cm long, 75  $\mu\text{m}$  inner diameter, 3  $\mu\text{m}$  resin) in buffer A (0.1% formic acid) and separated with a linear gradient of buffer B (84% acetonitrile and 0.1% formic acid) at a flow rate of 300  $\text{nl min}^{-1}$  controlled by IntelliFlow technology. One-hour gradient: 0–35% buffer B for 50 min, 35–100% buffer B for 5 min, hold in 100% buffer B for 5 min.

The mass spectrometer was operated in positive ion mode. MS data were acquired using a data-dependent top10 method dynamically choosing the most abundant precursor ions from the survey scan (300–1800  $m/z$ ) for higher-energy collisional dissociation (HCD) fragmentation. The automatic gain control target was set to 3–6, and the maximum injection time to 10 msec. The dynamic exclusion duration was 40.0 sec. Survey scans were acquired at a resolution of 70 000 at  $m/z$  200 and resolution for HCD spectra was set to 17 500 at  $m/z$  200, and the isolation width was 2  $m/z$ . The normalized collision energy was 30 eV and the underfill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.1%. The instrument was run with peptide recognition mode enabled. Raw data were queried for protein identification using the Proteome Discoverer and then connected to an in-house Mascot engine (Version 2.4; <https://www.thermofisher.com/ci/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html>). Phosphorylation site locations were validated using the phos-phoRS algorithm.

### RNA extraction for gene cloning and qRT-PCR analysis

Total RNA was isolated using the CTAB method and digested with RNase-free DNaseI (TaKaRa Biotechnology, Dalian, China) to remove genomic DNA. Then, 2  $\mu\text{g}$  of total RNA was used to synthesize the first-strand cDNA with the ReverTraAce-a First Strand cDNA Synthesis Kit (TOYOBO Biotech Co. Ltd., Osaka, Japan) according to the manufacturer's protocol. To investigate *PbTST4*, *PbCPK28* and *PbVHA-A1* gene expression, qRT-PCR analysis with a LightCycler 480 SYBR Green I Master kit (Roche, Indianapolis, IN, USA) was conducted according to the manufacturer's protocol. The PCR mixture contained 1  $\mu\text{l}$  cDNA and 8.2  $\mu\text{l}$  of nuclease-free water, 0.4  $\mu\text{l}$  of each primer, 10  $\mu\text{l}$  of LightCycler 480 SYBR GREEN I Master. qRT-PCR was performed on the LightCycler 480 (Roche), and all reactions were run as duplicates in 96-well plates. The reaction procedure entailed: pre-incubation at 95°C for 3 min, followed by 40 cycles of 95°C for 3 sec, and 60°C for 10 sec, 72°C for 30 sec, with a melt curve between 60 and 95°C. At the end, the average threshold cycle was calculated per sample using the  $2^{-\Delta\Delta\text{Ct}}$  method described by Livak and Schmittgen (2001). The pear *GAPDH* gene (*Pbr036263.1*, primer sequence shown in Table S5) was used as an internal control for pear. The qRT-PCR experiments reported in Figures 1, 4 and 5 are from one representative biological experiment (with three technical repeats). Three times biological replicates gave similar results.

### ATP extraction and measurement

The protocol for ATP extraction and measurement was described by Liu et al. (2006). A kit provided by Suzhou Keming Biotechnology was used to extract ATP from apple calli. In short, 0.3 g of the sample was added to 1 ml of reagent one, heated at 100°C for 5 min, centrifuged at 8000  $g$  for 15 min; the supernatant was filtered with a 0.45- $\mu\text{m}$  syringe filter, and then set aside. To determine ATP content by high-performance liquid chromatography, the injection volume was 10  $\mu\text{l}$ , and the flow rate was 1  $\text{ml min}^{-1}$ . With 20 min retention, the detection wavelength is 254 nm. The data reported in Figure 5 are from one representative biological experiment (with three technical repeats). Biological replicates (three times) gave similar results.

### Soluble sugar extraction from pear and strawberry fruit and quantification

Soluble sugars were extracted as described by Liu et al. (2016). Briefly, the pear fruit samples (2 g) were removed from  $-80^\circ\text{C}$  storage, and ground into a powder with a mortar and pestle in liquid nitrogen. The frozen powder was homogenized in 80% ethanol and incubated in a water bath at 37°C for 30 min to inactivate invertase enzymes. The solution with 2 g of sample was centrifuged at 12 000  $g$  at 4 °C for 30 min. The supernatant was collected in a 25 ml constant volume bottle. The residues (pellets) were mixed with 10 ml 80% ethanol again, incubated at 37°C for 30 min, and then centrifuged at 12 000  $g$  for 10 min. This entire process was repeated twice, after which the supernatants were combined, and then brought to a constant volume of 25 ml with 80% ethanol. Two-millilitre supernatants were processed in the rotary evaporator and dissolved in 1 ml sterile water. The soluble sugars extraction method of strawberry fruits was described previously (Jia et al., 2013).

The aliquots were passed through a 0.45- $\mu\text{m}$  membrane filter and the sugar was analyzed using a Sugar-pak 1 column (6.5 mm  $\times$  300 mm, Inner Diameter, 10  $\mu\text{m}$  particle size; Waters, Shanghai, China) with a guard column cartridge (Sugar-pak 1 Guard-Pak Holder and Insert; Waters). A refractive index detector (code 2414; Waters) with reference cell was maintained at 35°C. The column was maintained at 85°C. The mobile phase was deionized, distilled, degassed and filtered through 0.45- $\mu\text{m}$  Sep-Pak filter water at a flow rate of 0.6  $\text{ml min}^{-1}$ . The injection volume was 10  $\mu\text{l}$ .

### Phylogenetic analysis and sequences alignment

The plant CPK, TST and VHA-A gene sequences were extracted from public genomic databases (Phytozome 12, <https://phytozome.jgi.doe.gov/pz/portal.html>) and manually curated. All sequences were aligned using MAFFT software using iterative refinement. The CPK alignment was used to infer a phylogenetic maximum likelihood tree using RAxML software with JTT substitution models, the sequence alignment of plant TSTs and VHAs are shown in Figures S6 and S7 to determine the conserved phosphorylation sites of TST and VHA-A proteins.

### Statistical analysis

The data that refer to Figures 1, 2, 4 and 5 were statistically processed using the SPSS software (IBM, USA); plotting the mean with standard error; one-way ANOVA tests were analyzed, with a significance level accepted at  $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ;  $****P < 0.0001$ .



## Primers

All primers that were used for experiments could be observed in Table S5.

## ACKNOWLEDGEMENTS

The authors are grateful to various funding agencies and projects that contributed to the present work, including the National Natural Science Foundation of China (31820103012, 31801835, 31901983), the National Natural Science Foundation of Jiangsu Province (BK20180516, BK20180529), Earmarked Fund for China Agriculture Research System (Grant No. CARS-28), and the high-performance computing platform of Bioinformatics Center, Nanjing Agricultural University. In addition, the authors thank Ms Yuehua Ma for assistance in using the ultra-high-resolution confocal microscope. The authors thank Dr Wanpeng Wang and Junhui Zhou for help with subcellular localization analysis.

## AUTHOR CONTRIBUTIONS

JW designed the project. JL and RZ performed the experiments and wrote the manuscript. BC contributed to transient expression and sugar extraction of pear fruit. MZ, XL and BS contributed to bioinformatics analysis. ZL, JL and JW revised the manuscript.

## CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

## DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Subcellular localization of PbCPK28. The PbCPK28-GFP fusion gene under the control of the CaMV35S promoter was expressed in the *Arabidopsis* root cells.

**Figure S2.** PbCPK28 interacts with PbTST4 in yeast. Yeast colonies containing AD-PbCPK28 and BK-PbTST4 were grown on the -Trp, -Leu, -His, -Ade medium for 3 days.

**Figure S3.** The corresponding tandem spectra of PbTST4.

**Figure S4.** Structure and mechanism of the V-ATPase complex.

**Figure S5.** The corresponding tandem spectra of PbVHA-A1.

**Figure S6.** Homology analysis of putative orthologs for the PbTST4 clade III, showing the full identity of PbTST4 with genes from plants.

**Figure S7.** Homology analysis of putative orthologs for the PbVHA-A1 clade IV, showing the full identity of PbVHA-A1 with genes from plants.

**Table S1.** Genes localized within the QTL region.

**Table S2.** Major haplotypes of SNPs in the PbCPK28 promoter region.

**Table S3.** Correlation analysis of *PbCPK28* and *PbTST4* gene expression during fruit development.

**Table S4.** Full list of PbVHA-A1 proteins identified during pear fruit development.

**Table S5.** Primer sequences.

## REFERENCES

- Aluri, S. & Buttner, M. (2007) Identification and functional expression of the *Arabidopsis thaliana* vacuolar glucose transporter 1 and its role in seed germination and flowering. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2537–2542.
- Campo, S., Baldrich, P., Messegue, J., Lalanne, E., Coca, M. & San, S.B. (2014) Overexpression of a calcium-dependent protein kinase confers salt and drought tolerance in rice by preventing membrane lipid peroxidation. *Plant Physiology*, **165**, 688–704.
- Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W. *et al.* (2014) Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome Biology*, **15**, 415.
- Caruana, J.C., Sittmann, J.W., Wang, W. & Liu, Z. (2018) Suppressor of runnerless encodes a DELLA protein that controls runner formation for asexual reproduction in strawberry. *Molecular Plant*, **11**, 230–233.
- Chakrabarti, M., Zhang, N., Sauvage, C., Munos, S., Blanca, J., Canizares, J. *et al.* (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 17125–17130.
- Cheng, J., Wen, S., Xiao, S., Lu, B., Ma, M. & Bie, Z. (2017) Overexpression of the tonoplast sugar transporter CmtST2 in melon fruit increases sugar accumulation. *Journal of Experimental Botany*, **69**, 511–523.
- Cho, J.I., Burla, B., Lee, D.W., Ryoo, N., Hong, S.K., Kim, H.B. *et al.* (2010) Expression analysis and functional characterization of the monosaccharide transporters, OsTMTs, involving vacuolar sugar transport in rice (*Oryza sativa*). *The New Phytologist*, **186**, 657–668.
- Delormel, T.Y. & Boudsocq, M. (2019) Properties and functions of calcium-dependent protein kinases and their relatives in *Arabidopsis thaliana*. *The New Phytologist*, **224**, 585–604.
- Deng, J.W., Yang, X.Y., Sun, W.N., Miao, Y.H., He, L.G. & Zhang, X.L. (2020) The calcium sensor CBL2 and its interacting kinase CIPK6 are involved in plant sugar homeostasis via interacting with tonoplast sugar transporter TST2. *Plant Physiology*, **183**, 236–249.
- Doebley, J.F., Gaut, B.S. & Smith, B.D. (2006) The molecular genetics of crop domestication. *Cell*, **127**, 1309–1321.
- Dong, Y., Yang, X., Liu, J., Wang, B.H., Liu, B.L. & Wang, Y.Z. (2014) Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature Communications*, **5**, 3352.
- Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X. *et al.* (2018) Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. *Nature Genetics*, **50**, 796–802.
- Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M. *et al.* (2017) Genome resequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, **8**, 249.
- Endler, A., Reiland, S., Gerrits, B., Schmidt, U.G., Baginsky, S. & Martinoia, E.J.P. (2009) In vivo phosphorylation sites of barley tonoplast proteins identified by a phosphoproteomic approach. *Proteomics*, **9**, 310–321.
- Eom, J.S., Cho, J.I., Reinders, A., Lee, S.W., Yoo, Y., Tuan, P.O. *et al.* (2011) Impaired function of the tonoplast-localized sucrose transporter in rice, OsSUT2, limits the transport of vacuolar reserve sucrose and affects plant growth. *Plant Physiology*, **157**, 109–119.
- Frery, A., Nesbitt, T.C., Frery, A., Grandillo, S., Knaap, E., Cong, B. *et al.* (2000) Fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.
- Gao, A., Wu, Q., Zhang, Y., Miao, Y. & Song, C. (2014) Arabidopsis calcium-dependent protein kinase CPK28 is potentially involved in the response to osmotic stress. *Chinese Science Bulletin*, **59**, 1113–1122.
- Hrabak, E.M., Chan, C.W.M., Gribskov, M., Harper, J.F., Choi, J.H., Halford, N. *et al.* (2003) The arabidopsis CDPK-SnRK superfamily of protein kinases. *Plant Physiology*, **132**, 666–680.
- Jia, H., Wang, Y., Sun, M., Li, B., Han, Y., Zhao, Y. *et al.* (2013) Sucrose functions as a signal involved in the regulation of strawberry fruit development and ripening. *The New Phytologist*, **198**, 453–465.
- Jung, B., Ludewig, F., Schulz, A., Meissner, G., Wostefeld, N., Flugge, U.I. *et al.* (2015) Identification of the transporter responsible for sucrose accumulation in sugar beet taproots. *Nature Plants*, **1**, 14001.



- Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N. & Liu, Z. (2013) Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell*, **25**, 1960–1978.
- Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. *et al.* (2006) An SNP caused loss of seed shattering during rice domestication. *Science*, **312**, 1392–1396.
- Li, J., Zheng, D., Li, L., Qiao, X., Wei, S., Bai, B. *et al.* (2015) Genome-wide function, evolutionary characterization and expression analysis of sugar transporter family genes in pear (*Pyrus bretschneideri* Rehd). *Plant Cell Physiology*, **56**, 1721–1737.
- Li, J.M., Huang, X.S., Li, L.T., Zheng, D.M., Xue, C., Zhang, S.L. *et al.* (2015) Proteome analysis of pear reveals key genes associated with fruit development and quality. *Planta*, **241**, 1363–1379.
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L. *et al.* (2011) Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nature Genetics*, **43**, 1266–1269.
- Lin, Z., Li, X., Shannon, L.M., Yeh, C.T., Wang, M.L., Bai, G. *et al.* (2012) Parallel domestication of the Shattering1 genes in cereals. *Nature Genetics*, **44**, 720–724.
- Liu, H., Liu, H., Zhou, L., Zhang, Z., Zhang, X., Wang, M. *et al.* (2015) Parallel domestication of the heading date 1 gene in cereals. *Molecular Biology and Evolution*, **32**, 2726–2737.
- Liu, H.J., Jiang, Y., Luo, Y. & Jiang, W. (2006) A simple and rapid determination of ATP, ADP and AMP concentrations in pericarp tissue of litchi fruit by high performance liquid chromatography. *Food Technology and Biotechnology*, **44**, 531–534.
- Liu, L., Chen, C.X., Zhu, Y.F., Xue, L., Liu, Q.W., Qi, K.J. *et al.* (2016) Maternal inheritance has impact on organic acid content in progeny of pear (*Pyrus* spp.) fruit. *Euphytica*, **209**, 305–321.
- Livak, K. & Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $^{-\Delta\Delta CT}$  method. *Methods*, **25**, 402–408.
- Lombard, P. & Westwood, M. (1987) Pear rootstocks. In: Rom, R.C. & Carlson, R.F. (Eds.) *Rootstocks for fruit crops*. New York: John Wiley and Sons, pp. 145–183.
- Martinoia, E., Maeshima, M. & Neuhaus, H.E. (2006) Vacuolar transporters and their essential role in plant metabolism. *Journal of Experimental Botany*, **58**, 83–102.
- Matschi, S., Werner, S., Schulze, W.X., Legen, J., Hilger, H.H. & Romeis, T. (2013) Function of calcium-dependent protein kinase CPK28 of *Arabidopsis thaliana* in plant stem elongation and vascular development. *The Plant Journal*, **73**, 883–896.
- Miller, C.J. & Turk, B.E. (2016) Kinase screening and profiling: rapid identification of protein kinase phosphorylation site motifs using combinatorial peptide libraries. In: Zegzouti, H. & Goueli, S. (Eds.) *Methods in molecular biology*. New York: Humana Press, pp. 203–216.
- Müller, N.A., Zhang, L., Koornneef, M. & Jiménez-Gómez, J.M. (2018) Mutations in EID1 and LNK2 caused light-conditional clock deceleration during tomato domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 7135–7140.
- Murcia, G., Pontin, M., Reinoso, H., Baraldi, R., Bertazza, G., Gomez-Talquenca, S. *et al.* (2016) ABA and GA3 increase carbon allocation in different organs of grapevine plants by inducing accumulation of non-structural carbohydrates in leaves, enhancement of phloem area and expression of sugar transporters. *Physiologia Plantarum*, **156**, 323–337.
- Myers, C., Romanowsky, S.M., Barron, Y.D. & Garg, S. (2009) Calcium-dependent protein kinases regulate polarized tip growth in pollen tubes. *The Plant Journal*, **59**, 528–539.
- Neuhaus, H.E. (2007) Transport of primary metabolites across the plant vacuolar membrane. *FEBS Letters*, **581**, 2223–2226.
- Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J.E.E., Morgan, D.R. *et al.* (2007) Evolution phylogeny and classification of Rosaceae. *Plant Systematics and Evolution*, **266**, 5–43.
- Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M. *et al.* (2015) Evolution of the grain dispersal system in barley. *Cell*, **162**, 527–539.
- Qin, M.F., Li, L.T., Singh, J., Sun, M.Y., Bai, B., Li, S.W. *et al.* (2022) Construction of a high-density bin-map and identification of fruit quality-related quantitative trait loci and functional genes in pear. *Horticulture Research*, **9**, uhac141.
- Ren, Y., Guo, S., Zhang, J., He, H., Sun, H. & Tian, S. (2018) A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiology*, **176**, 836–850.
- Rudd, J.J., Franklin, F.C.H., Lord, J.M. & Franklin-Tong, V.E. (1996) Increased phosphorylation of a 26-kD pollen protein is induced by the self-incompatibility response in *Papaver rhoeas*. *Plant Cell*, **8**, 713–724.
- Schneider, S., Hulpke, S., Schulz, A., Yaron, I., Höll, J., Imlau, A. *et al.* (2012) Vacuoles release sucrose via tonoplast-localised SUC4-type transporters. *Plant Biology*, **14**, 325–336.
- Schulz, A., Beyhl, D., Marten, I., Wormit, A., Neuhaus, E., Poschet, G. *et al.* (2011) Proton-driven sucrose symport and antiport are provided by the vacuolar transporters SUC4 and TMT1/2. *The Plant Journal*, **68**, 129–136.
- Schulze, W.X., Schneider, T., Starck, S., Martinoia, E. & Trentmann, O. (2012) Cold acclimation induces changes in Arabidopsis tonoplast protein abundance and activity and alters phosphorylation of tonoplast monosaccharide transporters. *The Plant Journal*, **69**, 529–541.
- Sigmon, B. & Vollbrecht, E. (2010) Evidence of selection at the *ramosa1* locus during maize domestication. *Molecular Ecology*, **19**, 1296–1311.
- Soyk, S., Lemmon, Z.H., Oved, M., Fisher, J., Liberatore, K.L., Park, S.J. *et al.* (2017) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell*, **169**, 1142–1155.e12.
- Soyk, S., Lemmon, Z.H., Sedlazeck, F.J., Jiménez-Gómez, J.M., Alonge, M., Hutton, S.F. *et al.* (2019) Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. *Nature Plants*, **5**, 471–479.
- Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*, **43**, 1160–1163.
- Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z. *et al.* (2018) Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nature Genetics*, **50**, 1435–1441.
- Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q. *et al.* (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics*, **49**, 579–587.
- Wingenter, K., Schulz, A., Wormit, A., Wic, S., Trentmann, O., Hoermiller, I.I. *et al.* (2010) Increased activity of the vacuolar monosaccharide transporter TMT1 alters cellular sugar partitioning, sugar signaling, and seed yield in Arabidopsis. *Plant Physiology*, **154**, 665–677.
- Wingenter, K., Trentmann, O., Wanschuh, I., Hormiller, I.I., Heyer, A.G., Reinners, J. *et al.* (2011) A member of the mitogen-activated protein 3-kinase family is involved in the regulation of plant vacuolar glucose uptake. *The Plant Journal*, **68**, 890–900.
- Winter, H., Robinson, D.G. & Heldt, H.W. (1993) Subcellular volumes and metabolite concentrations in barley leaves. *Planta*, **191**, 180–190.
- Wormit, A., Trentmann, O., Feifer, I., Lohr, C., Tjaden, J., Meyer, S. *et al.* (2006) Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. *Plant Cell*, **18**, 3476–3490.
- Wu, J., Wang, Y., Xu, J., Korban, S.S., Fei, Z., Tao, S. *et al.* (2018) Diversification and independent domestication of Asian and European pears. *Genome Biology*, **19**, 77.
- Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X. *et al.* (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*, **30**, 105–111.
- Zhang, M.Y., Xue, C., Xu, L., Sun, H., Qin, M.F., Zhang, S. *et al.* (2016) Distinct transcriptome profiles reveal gene expression patterns during fruit development and maturation in five main cultivated species of pear (*Pyrus* L.). *Scientific Reports*, **6**, 28130.
- Zhang, W., Tan, L., Sun, H., Zhao, X., Liu, F., Cai, H. *et al.* (2019) Natural variations at TIG1 encoding a TCP transcription factor contribute to plant architecture domestication in rice. *Molecular Plant*, **12**, 1075–1089.
- Zhang, Z., Fu, D., Sun, Z., Ju, C., Miao, C., Wang, Z. *et al.* (2021) A tonoplast-associated calcium signaling regulates manganese homeostasis in Arabidopsis. *Molecular Plant*, **14**, 805–819.
- Zhao, L.N., Shen, L.K., Zhang, W.Z., Zhang, W., Wang, Y. & Wu, W. (2013)  $Ca^{2+}$ -dependent protein kinase11 and 24 modulate the activity of the inward rectifying  $K^{+}$  channels in Arabidopsis pollen tubes. *Plant Cell*, **25**, 649–661.
- Zhou, L., Lan, W., Jiang, Y., Fang, W. & Luan, S. (2014) A calcium-dependent protein kinase interacts with and activates a calcium channel to regulate pollen tube growth. *Molecular Plant*, **7**, 369–376.

- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W. *et al.* (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, **33**, 408–414.
- Zhu, L., Li, B., Wu, L., Li, H., Wang, Z., Wei, X. *et al.* (2021) MdERDL6-mediated glucose efflux to the cytosol promotes sugar accumulation in the vacuole through up-regulating TSTs in apple and tomato. *Proceedings of the National Academy of Sciences of the United States of America*, **5**, 118 (1).
- Zhu, S.Y., Yu, X.C., Wang, X.J., Zhao, R., Li, Y., Fan, R.C. *et al.* (2007) Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis. *Plant Cell*, **19**, 3019–3036.
- Zhu, Z., Tan, L., Fu, Y., Liu, F., Cai, H., Xie, D. *et al.* (2013) Genetic control of inflorescence architecture during rice domestication. *Nature Communications*, **4**, 2200.



# Genome-wide genetic diversity and IBD analysis reveals historic dissemination routes of pear in China

Xuening Chen<sup>1</sup> · Mingyue Zhang<sup>2</sup> · Manyi Sun<sup>1</sup> · Yueyuan Liu<sup>1</sup> · Shengnan Li<sup>1</sup> · Bobo Song<sup>1</sup> · Mengyan Li<sup>1</sup> · Shaoling Zhang<sup>1</sup> · Runze Wang<sup>1</sup> · Jiaming Li<sup>1</sup> · Kejiao Zhao<sup>1</sup> · Jun Wu<sup>1</sup>

Received: 21 September 2021 / Revised: 7 November 2021 / Accepted: 15 November 2021 / Published online: 4 December 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

Pear (*Pyrus*) is an important temperate fruit, which originates in the southwestern region of China and has more than 3000 year's cultivation history. However, the historic routes of pear dissemination in China have not been fully elucidated. In this study, a total of 2,412,930 single nucleotide polymorphisms (SNPs) at a density of 4.74 SNP/kb were identified by resequencing. The SNP-based phylogenetic analysis revealed that 102 pear samples from 23 provinces in China were divided into two major clades and eight geographic groups, and these divisions were supported by results of a population structure analysis and principal component analysis (PCA). Combined with the results of population diversity and identity-by-descent (IBD) analysis, it was revealed that the dissemination direction of pear was from southwest to southeast and from south to north. In the southern region, the dispersal pattern of pear spreading from west to east was generally in line with the course of the Yangtze River and Pearl River. The southern pear spread by multiple routes to its north neighboring areas, and regions in the middle and lower reaches of the Yellow River played important roles in the further dissemination of pear in the northern region of China. Moreover, we identified comparative higher genetic diversity of Ussurian pear than other populations, which might be due to low degree of domestication and closely resembled the high diversity of its wild counterpart. Our study provides new information to further our understanding of pear evolution in China, while laying a foundation of data for population genetic research, germplasm protection, and utilization for pear breeding in the future.

**Keywords** Pear · SNP · Genetic diversity · Phylogenetic relationships · Identity-by-descent (IBD) · Dissemination route

## Introduction

Pear is one of the most widespread fruit trees in temperate regions of the world (Bell 1984). The genus *Pyrus* is classified into the subfamily Amygdaloideae of the family Rosaceae (Shi et al. 2013). *Pyrus* originated from the

southwestern region of China (Challice and Westwood 1973; Rubtsov 1944; Wu et al. 2018) and has been cultivated for more than 3000 years (Pu and Wang 1963). China is one of the three diversity centers of cultivated pear (Vavilov 1951; Zhukovsky 1965) and has 13 native pear species (Pu and Wang 1963; Yu 1979). Among them, sand pear (*P. pyrifolia*), white pear (*P. bretschneideri*), Ussurian pear (*P. ussuriensis*), and Xinjiang pear (*P. sinkiangensis*) are the four major species cultivated in China.

In previous studies, various molecular markers were used to explore pear germplasm resources around the world (Chang et al. 2017; Dolatowski et al. 2004; Iketani et al. 1998; Song et al. 2014; Teng et al. 2002), including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), and noncoding regions of chloroplast DNA (cpDNA). However, these molecular markers often have a limited number of sites and thus cannot achieve the high sequence coverage needed to represent genetic information

Xuening Chen and Mingyue Zhang contributed equally to this work.

Xuening Chen and Mingyue Zhang are co-first authors.

Responsible Editor: D. Chagné

✉ Jun Wu  
wujun@njau.edu.cn

<sup>1</sup> College of Horticulture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Jiangsu 210095 Nanjing, China

<sup>2</sup> State Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an 271018, Shandong, China

at the whole-genome level. With the development of genome sequencing technology, genome-wide SNPs can provide more genetic information to analyze the genetic differences between individuals (Gupta, Rustgi, and Kulwal 2005; Ma et al. 2019), and they have been widely used to explore and understand genetic, evolutionary, and domestication processes of agricultural crops (Cao et al. 2019; Kumar et al. 2017; Park and Burke 2020; Tam et al. 2019; Wang et al. 2020). For example, the domestication of apples was studied using SNP data of 117 apple accessions, and these data provided evidence supporting a two-step domestication model for apple fruit enlargement (Duan et al. 2017). The SNP data of 158 mango materials from seven geographical populations illustrated a complex domestication background of mango (Warschefsky and von Wettberg 2019). Since the ‘Dangshansuli’ genome was released in 2013 (Wu et al. 2013), many SNPs have been developed from pear sample sequences with the help of the pear genome. SNP markers play an important role in genetic diversity analysis and investigations of genetic relationships of the pear germplasm (Li et al. 2019; Montanari et al. 2019; Wu et al. 2018).

Previous studies have improved our understanding of the dissemination of *Pyrus* in China. The southwestern region is considered the starting point of pear spread, and the Yangtze and Pearl Rivers, two of China’s major rivers, facilitated the spread of southern pear (Xue et al. 2018; Yue et al. 2018). Additionally, Shanxi Province was considered the starting point of evolution of routes for pear dispersal in northern China (Chang et al. 2017). However, there are still doubts and contradictions in the reported genetic relationships among Chinese pear populations in various regions and the specific routes of dissemination among different regions.

In this study, we selected 102 representative pear landraces from 23 provinces in China. SNPs obtained from resequencing data of all pear accessions in this study were used to explore the genetic diversity and dispersal routes of pears in China. The objectives of this study were to assess the genetic diversity and relationship of pear germplasms from different regions of China and determine the historic dissemination routes of pear. Furthermore, this investigation will contribute to the understanding of the evolution and domestication of pear and protection and utilization of pear germplasm, and it can also serve as a useful reference for germplasm studies on other perennial plants.

## Methods

### Sampling information

We used a total of 102 pear accessions from pear samples originally collected from 23 provinces in China, all of which were landraces (Table S1). Among them, genome

resequencing data from 78 pear accessions were obtained from our previous studies (Wu et al. 2018; Zhang et al. 2021), and additional resequencing data of 24 accessions were obtained using the methods of Wu et al. (2018).

### Mapping and SNP calling

Trimmomatic version 0.39 (Bolger, Lohse, and Usadel 2014) was used to remove adaptor and low-quality sequences of resequencing data of the 102 landraces using the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 MINLEN:36. The clean reads were aligned to the ‘Dangshansuli’ reference genome (<http://peargenome.njau.edu.cn>) using BWA version 0.7.17-r1188 (Li and Durbin 2009), and only uniquely mapped reads were retained. Picard-tools version 1.119 (<http://broadinstitute.github.io/picard/>) was used to remove potential PCR repeats. SAMtools version 1.9 (Li et al. 2009) was used to establish index (indices) of BAM format file. SNP calling was performed with the HaplotypeCaller tool in Genome Analysis Toolkit version 4.1.4.0 (GATK) (McKenna et al. 2010). All identified SNPs were processed through a two-step filter. First, all SNPs were filtered using the VariantFiltration tool in GATK with the follow parameters: QualByDepth (QD) < 2.0, FisherStrand (FS) > 60.0, RMSMappingQuality (MQ) < 40.0, MappingQualityRankSumTest (MQRankSum) < -12.5, and ReadPosRankSumTest (ReadPosRankSum) < -8.0 (Han et al. 2020). Second, the SNPs were filtered using VCFtools version 0.1.16 (Danecek et al. 2011) with the minor allele frequency set to > 0.05 and missing < 0.9. The final filtered SNP dataset was used for further analysis.

### Phylogenetic and population structure analysis

To infer phylogenetic relationships among pear landraces across the different areas of sample collection, we conducted a phylogenetic analysis. We selected a total of 126,398 SNPs at fourfold degenerate sites (4Dtv SNPs), and maximum likelihood phylogenetic trees were constructed using IQ-TREE version 2.0.3 (Nguyen et al. 2015). We used 1000 bootstrap replicates to assess support. The tree was visualized and annotated using ITOLv5 (<https://itol.embl.de/>). Population structure was investigated using ADMIXTURE version 1.3.0 (Alexander, Novembre, and Lange 2009) with default parameters. The population structure results were displayed using the barplot function in R. The principal component analysis (PCA) was performed using PLINK version 1.9 (Purcell et al. 2007), and the result was visualized using the R package ggplot2 (Ginestet 2011).



## Statistics of population genetics

Common measures of genetic diversity were calculated for the populations of pear accessions using the SNP dataset. Values of nucleotide diversity ( $\pi$ ) and pairwise fixation index ( $F_{ST}$ ) were computed using VCFtools (version 0.1.16) with a sliding window of 10 kb and a step length of 5 kb. Expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ), percentage of polymorphism (%Poly), and number of private alleles (Ap) were calculated using Stacks: populations version 2.52 (Catchen et al. 2013).

Detection of selective sweeps in *P. pyrifolia* and *P. ussuriensis*.

We used the SNP dataset of pear population from previous research (Wu et al. 2018). The wild and cultivated *P. pyrifolia* accessions and wild and cultivated *P. ussuriensis* accessions were used to detect selective sweep analysis, respectively. The SNP datasets for selective sweep analysis were filtered according to the methods of our previous research (Wu et al. 2018). To identify regions with selective signals, the  $\pi$  ratio ( $\pi_{\text{wild}}/\pi_{\text{cultivated}}$ ) and  $F_{ST}$  values were calculated in non-overlapping windows of 10-kb slide windows across the entire pear genome. Regions (10-kb slide windows) with significant selective signals were identified using the following criteria: the top 5% of  $F_{ST}$  and top 5% of the  $\pi$  ratio. The result was visualized using the R package ggplot2 (Ginestet 2011).

## Identification of identity-by-descent segments

To detect the similar identity-by-descent (IBD) segments of pear landraces originating from the various geographical regions, IBD segments were identified using the algorithm of BEAGLE version 4.1 (Browning and Browning 2007) with the default parameters. Subsequently, the IBD segments of pairwise samples were grouped according to phylogenetic tree result. In order to eliminate the influence of different sample sizes across groups, the average length of IBD segments was calculated by dividing the total length of IBD segments by the number of sample pairs between two groups (Campbell et al. 2012). The result was visualized using the R package ggplot2 (Ginestet 2011).

## Results

### Genome resequencing and variation calling

A total of 102 pear landraces were collected from 23 provinces, which were 67 *P. pyrifolia*, 15 *P. ussuriensis*, 14 *P. bretschneideri*, and 6 *P. sinkiangensis* (Table 1). Resequencing data of the 102 pear genomes generated a total of 669.5 Gb of cleaned sequences with an average of 6.86 Gb per accession (approximately 13× coverage of the

**Table 1** Measures of genetic diversity for eight pear groups

| Group  | $H_E$  | $H_O$  | %Poly | Ap     |
|--------|--------|--------|-------|--------|
| Group1 | 0.2529 | 0.2504 | 83.78 | 367    |
| Group2 | 0.2443 | 0.2797 | 76.75 | 84     |
| Group3 | 0.2323 | 0.2366 | 74.12 | 139    |
| Group4 | 0.2342 | 0.2578 | 75.49 | 177    |
| Group5 | 0.2428 | 0.2428 | 81.63 | 3543   |
| Group6 | 0.2468 | 0.2674 | 80.79 | 26     |
| Group7 | 0.2294 | 0.2614 | 65.89 | 327    |
| Group8 | 0.2623 | 0.2880 | 81.74 | 47,054 |

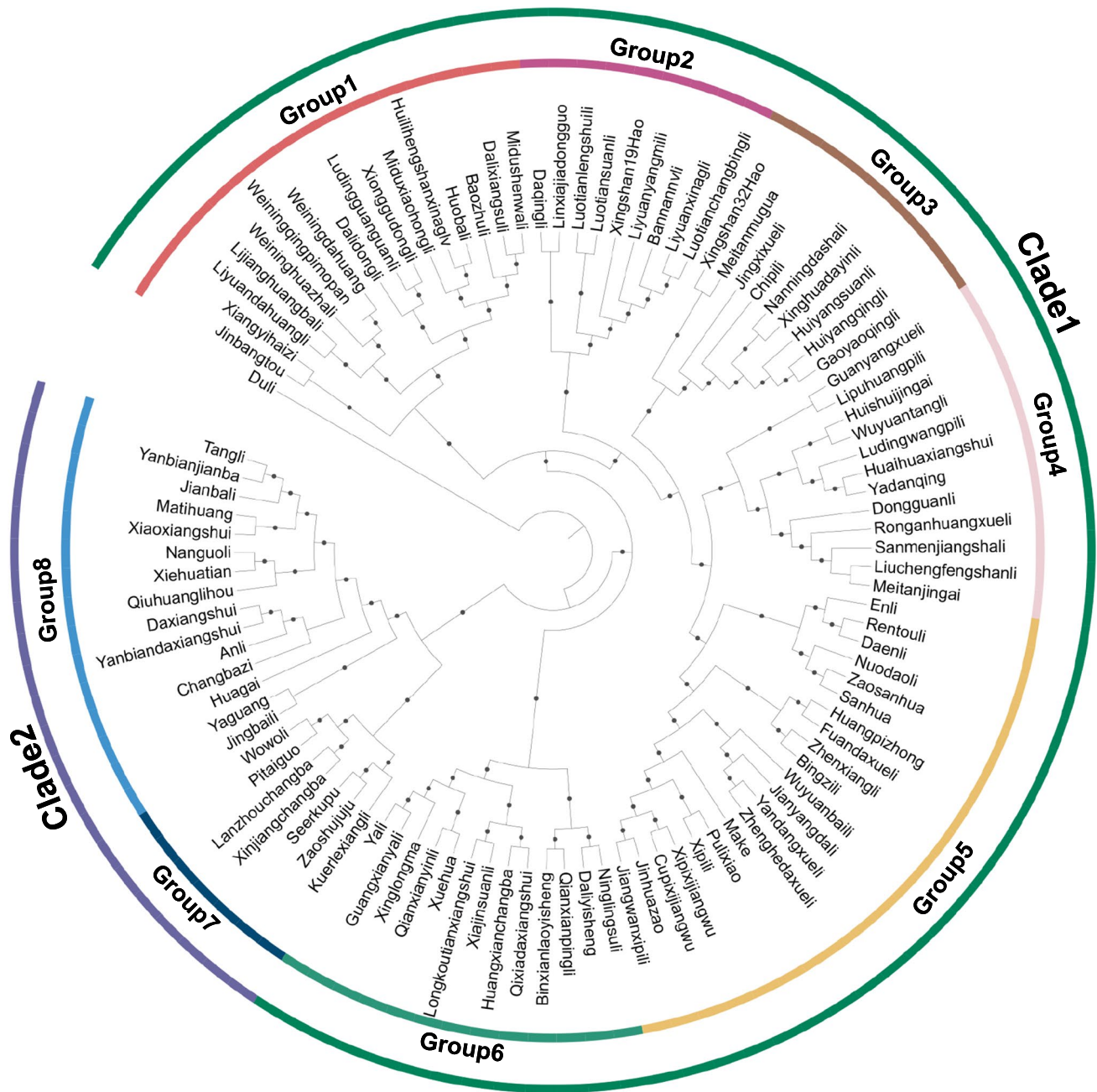
$H_E$ , heterozygosity within populations;  $H_O$ , observed heterozygosity; %Poly, percentage of polymorphism; Ap, private alleles

pear genome). These sequences were aligned to the reference genome ‘Dangshansuli’ (Wu et al. 2013). The average mapping rate was 87.85% (Table 1). After SNP calling and filtering with a missing rate < 0.9 and minor allele frequency > 0.05, we identified a final set of 2,412,930 SNPs. The number of SNPs identified in different individuals varied from 614,417 (‘Ninglingsuli’) to 1,119,646 (‘Xiaoxiangshui’). Among these SNPs, 16.3% were located in coding regions, 8.9% were synonymous and 7.4% were non-synonymous. This SNP dataset was used for downstream analyses.

### Accessions clustered into two major clades and eight geographic groups

The phylogenetic tree was constructed based on the maximum likelihood method using SNPs at fourfold degenerate sites (4Dtv SNPs) and the wild species ‘Duli’ (*Pyrus betulifolia* Bunge) as the outgroup. Nodes with a bootstrap value > 80% are indicated by a black circle in the result. The 102 accessions from different areas in China were divided into two major clades (Fig. 1). Clade1 consisted of all southern accessions and some northern accessions from the middle and lower reaches of the Yellow River region. Clade2 contained pear accessions from northwestern and northeastern China.

According to the geographical distribution of the samples, Clade1 was divided into six groups. Group1 contained 16 sand pear accessions, which mostly came from the southwestern region. Group2 was a cluster of nine sand pear accessions, the majority of which were from the Hubei (HB) and Sichuan (SC) provinces in the middle of the Yangtze River region. Group2 was located adjacent to Group1. Group3 contained nine sand pear accessions from three provinces in the southernmost part of the Chinese mainland coast and two accessions from Hubei and Guizhou (GZ). Group4 had 12 sand pear accessions from the Guangxi (GX) region and the Guizhou, Guangdong (GD), Hubei, Hunan (HN), Jiangxi (JX), and Sichuan provinces in the southern



central region of the country. Group5 consisted of 20 sand pear accessions and one white pear accession, and mainly from the southeastern region, indicating genetic relationships of pear samples from southeast China are relatively close. Twelve white pear accessions and one sand pear accession clustered into Group6 and they were from the Henan (HA), Hebei (HE), Shandong (SD), Shaanxi (SN), and Shanxi (SX) provinces in the middle and lower reaches of the Yellow River region.

geographical distribution. Group7 contained seven Xinjiang pear accessions, which are cultivated in provinces of the northwestern region, such as the Xinjiang (XJ) and Gansu (GS) provinces. Group8 consisted of 15 Ussurian pears, which were mostly from northeast China.

The results of the population structure analysis revealed the genetic backgrounds and relationships among pear

populations (Fig. 2a, b; Fig. S1a–e). The program ADMIXTURE (Alexander, Novembre, and Lange 2009) was used to establish the population structure of 102 pear accessions. When  $K=2$  (Fig. 2a), the resulting data showed that the eight geographical groups were divided into two populations. Pears from the southern region (Group1–Group5) and the middle and lower reaches of the Yellow River region (Group6) exhibited the same genetic background, while Group7 (composed of Xinjiang pear) and Group8 (composed of Ussurian pear) consisted of a high-level purple gene pool. Among the eight groups, we found that Ussurian pears in Group8 were all clustered together with a single color and showed no distinctive variation with increasing  $K$  values. The result of  $K=2$  was consistent with the result of having determined two major branches in the phylogenetic tree.

When  $K=3$ , a new gene pool (pink) mainly appeared in Group1–Group6 (Fig. S1a). In Group1, most accessions were composed of the high-level green gene pool, and these accessions were from the southwest region of China. The pears from the southeastern region (Group5) were grouped into the pink gene pool. Interestingly, Group2 to Group4, which geographically lie in between Group1 and Group5, showed an obvious mixture of green and pink gene pools. For  $K=4$  (Fig. S1b), a high level of the orange gene pool appeared among the samples of the middle reaches of the Yangtze River (Group2) and the middle and lower reaches of the Yellow River (Group6). When  $K=5$  (Fig. S1c), another new gene pool mainly appeared in the southern central region (Group4), and four pear accessions in Group4 showed a single genetic background. When  $K=6$  (Fig. S1d), the middle reaches of the Yangtze River region (Group2) and southernmost region (Group3) mostly showed light green and yellow genetic backgrounds and there was no change with the  $K=5$ . When  $K=7$  (Fig. S1e), four pear accessions in Group4 showed a single background, while other pear accessions in Group4 showed mixed backgrounds, indicating that the four accessions were different from other accessions of Group4. With the increase of  $K$  to 8, Group7, composed

of Xinjiang pears grown in the northwest region, formed a separate gene pool (gray), which shows the particularity of the genetic background of Xinjiang pear. In general, the results of the analysis of population structure at  $K=8$  (Fig. 2b) were consistent with the results of the phylogenetic tree, supporting the geographical distribution of samples.

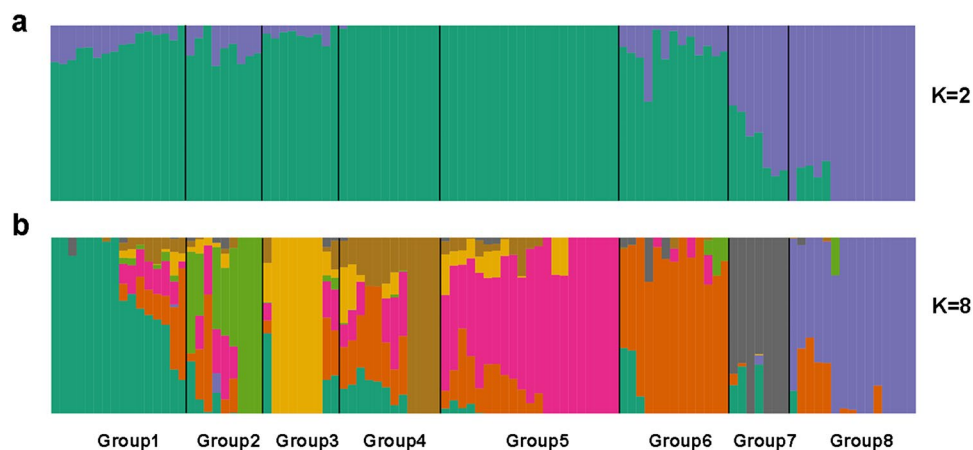
The PCA of the 102 pear accessions showed that the genetic relationships of pear accessions in different regions were similar to the results of the phylogenetic and population structure analyses. The first principal component explained 68.70% of the variance, and the second explained 31.30% of the variance (Fig. 3a). The PCA revealed that sand pear and white pear accessions from southern and northern regions clustered together. Xinjiang pear and Ussurian pear accessions (Group7 and Group8) were distantly separated from the accessions of the other groups. This also corresponded well with the results of phylogenetic tree and population structure analyses.

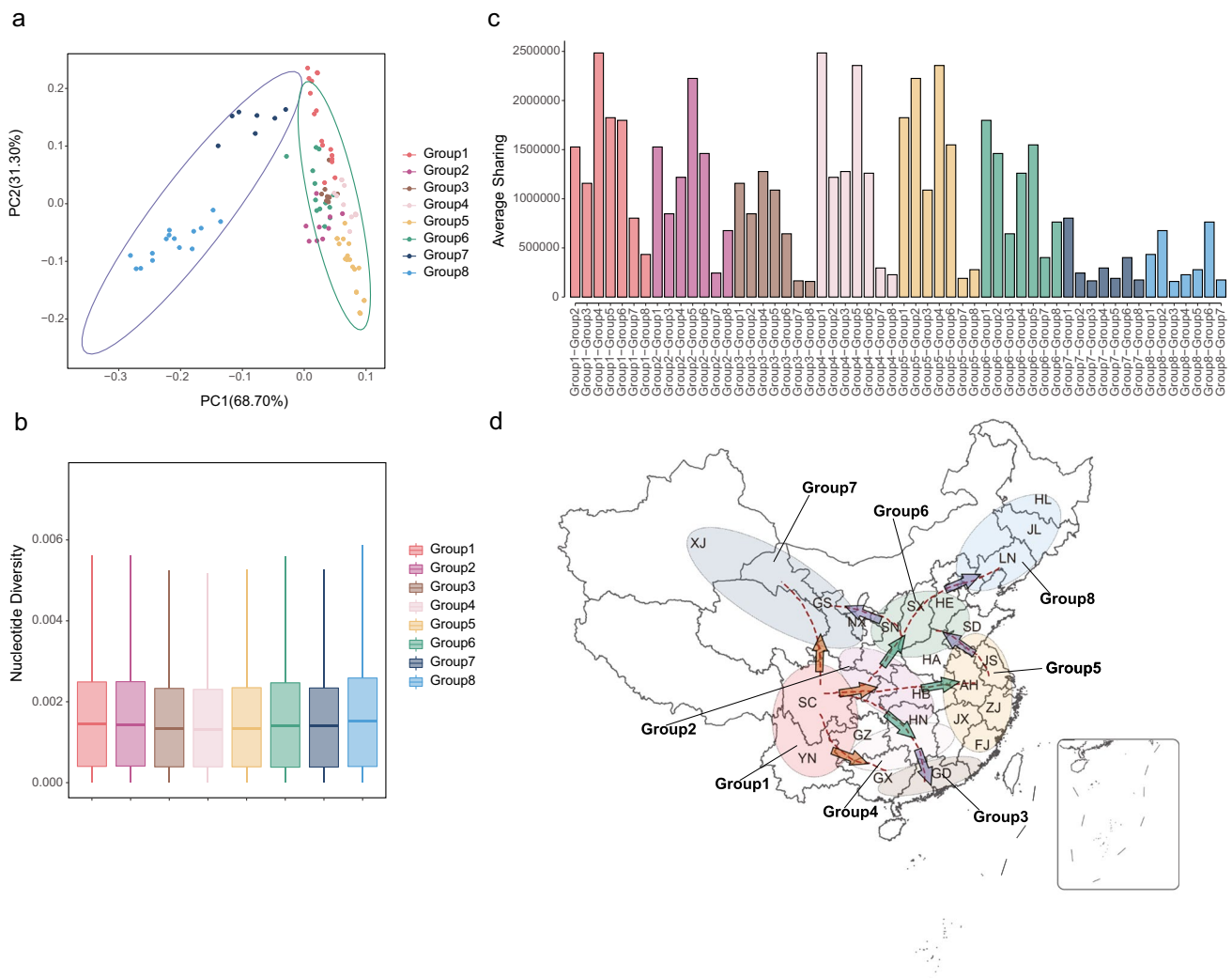
### Genetic diversity and population differentiation revealed the dissemination route and genetic relationships of pear groups

The population statistics of genetic diversity were calculated for the eight geographic pear groups, revealing the genetic and dispersal relationships among the populations. The calculated population indices of  $H_O$ ,  $H_E$ , %Poly, and  $A_p$  are listed in Table 1, whereas Fig. 3b illustrates the  $\pi$  data.

In southern China, the levels of genetic diversity indicated the pattern of pear dissemination generally moved from west to east. The  $\pi$  results showed that samples of two groups, Group1 and Group2, respectively, collected from southwest China and the middle reaches of the Yangtze River region, had higher  $\pi$  ( $1.620 \times 10^{-3}$ ,  $1.625 \times 10^{-3}$  respectively) than samples of three other groups, Group3–Group5. Similarly, the values of  $H_E$  were higher in the southwestern Group1 (0.2529) and the Yangtze River Group2 (0.2443), while comparatively lower in the farther south and southeastern

**Fig. 2** Population structure of 102 pear accessions. **a**, **b** Population structure based on the number of ancestry populations ( $K=2$  and  $K=8$ ). Each vertical bar represents a single accession that was assigned ancestry to one or more of the populations (different colors)





**Fig. 3** Genetic relationships of pears in different geographical regions. **a, b** Principal component analysis and nucleotide diversity of the 102 pear accessions that clustered into eight pear groups. **c**

Genome-wide IBD sharing for the average pair of accessions across eight pear groups. **d** Map of dissemination routes of pears in China

Group3–Group5. Levels of  $H_O$  were highest for Group2 (0.2797) and lowest for Group3 (0.2366). The percentage of polymorphic loci varied from 83.78% in Group1 to 74.12% in Group3. The  $A_p$  in the southeast coastal area (Group5) was the highest at 3543, which was 9.7 times higher than that in the next highest population, Group1 (367), and 20 times higher than that in the third highest population, Group4 (177). A general pattern in diversity appeared to correspond with geography: diversity decreased from pear groups in the southwestern mountainous area (Group1) to the adjacent central region (Group2, Group4), and then to the southeastern coastal area (Group5). This evidence also supports dispersal route of pear from west to east in southern China. Interestingly,  $A_p$  indicated that genetic diversity of pear samples in Group5 was higher than that in Group1 and Group2; this difference might be caused by the larger pear population in Group5 (Kalinowski 2004). Moreover, the results of  $H_O$ ,  $H_E$ , %Poly, and  $\pi$  showed that the genetic

diversity of Group2 was higher than that of Group4 in the southern central region, which revealed a pattern of dissemination from Group2 to Group4. The results of  $H_O$ ,  $H_E$ , %Poly, and  $A_p$  showed that the genetic diversity of Group4 was higher than that of Group3 in the southernmost region, which revealed a dissemination route from Group4 to Group3.

From south to north, the dissemination direction of pear was dispersive. The levels of  $H_O$ ,  $H_E$ , %Poly,  $A_p$ , and  $\pi$  determined in pear accessions from the northwest region (Group7) were 0.2614, 0.2294, 65.89%, 327, and  $1.537 \times 10^{-3}$ , respectively. The obvious reductions in all indices from the southwest mountain (Group1) to northwest regions (Group7) reflected the pattern of pear dispersal from the southwest mountain region to the north. The values for  $H_O$ ,  $H_E$ , %Poly,  $A_p$ , and  $\pi$  of pear accessions from the middle and lower reaches of the Yellow River (Group6) were 0.2674, 0.2468, 80.79%, 26, and  $1.594 \times 10^{-3}$ , respectively.



The  $H_O$ ,  $H_E$ , and  $\pi$  of pear accessions from the middle reaches of the Yangtze River (Group2) were higher than those of Group6, which revealed a dissemination route from Group2 to Group6. Most indices also showed that the genetic diversity of pear populations in Group6 was higher than that of the northwest region (Group7), indicating that there may have been a dissemination route from the middle and lower reaches of the Yellow River to the northwest. Notably, the  $H_O$ ,  $H_E$ ,  $A_p$ , and  $\pi$  of the Ussurian pears from the northeast region (Group8) were higher than those of pear populations in other regions.

The genetic diversity of sand pear in southern China was lower than that of Ussurian pear in northern China (Fig. 3c; Table 1), and the disparity may be caused by the higher similarity between the cultivated Ussurian pear and its wild counterpart. We included in our study the SNP datasets of wild and cultivated samples of Chinese sand pear and Ussurian pear used in a previous study (Wu et al. 2018). The top 5% of the  $\pi$  ratios ( $\pi_{\text{Wild}}/\pi_{\text{Cultivated}}$ ) and  $F_{ST}$  values of these samples were used as two thresholds to identify candidate selective sweeps that have occurred during sand pear and Ussurian pear domestication by comparing the wild against the cultivated samples. The number of candidate-selected slide windows in the domestication of sand pear was more than the number of candidate-selected slide windows in the domestication of Ussurian pear (Fig. S2), indicating the Ussurian pear was less domesticated than sand pear. Compared with cultivated sand pear, the genetic diversity of cultivated Ussurian pear more closely resembled the high diversity of its wild counterpart.

The value of  $F_{ST}$  indicates the measure of population differentiation between each group (Table 2), and thus was used to lend additional support to the potential dissemination routes of pear in China. The  $F_{ST}$  values of each Group7 and Group8 versus groups of other regions were high, indicating that the pear populations in other regions were highly differentiated from the Xinjiang pear in Group7 and the Ussurian pear in Group8. The  $F_{ST}$  values between the closest groups were lower than those between groups geographically farther away from each other (Table 2), implying extensive gene flow between pear populations in neighboring areas.

Unexpectedly, the  $F_{ST}$  values between Group6 and southern groups (Group1–Group5) were smaller than the  $F_{ST}$  values between Group6 and northern neighboring groups (Group7 and Group8). These values suggest a strong genetic relationship between Group6 and southern pear accessions, which is consistent with the results of the phylogenetic and PCA.

### Analysis of mean IBD provides evidence of relationships of historic dispersal among eight pear groups

The IBD analysis was conducted on accessions across populations to identify shared DNA fragments among populations, explain relatedness between different groups, and provide evidence of potential dissemination routes of pear accessions (Fig. 3c). In comparing average IBD between pear populations in the middle reaches of the Yangtze River (Group2) and pear groups in other regions, the degree of IBD with the southwest and southeast regions (Group1 and Group5, respectively) was both high. Similarly, the degree of IBD between accessions of the southern central region (Group4) and accessions from the southwest and southeast regions (Group1 and Group5, respectively) was also high. These results reflect the close relationship between the pear populations of the middle reaches of the Yangtze River or southern central region and the pear populations of the southwest and southeast regions of China. Furthermore, they support the dissemination route from west to east in southern China and were consistent with results from the genetic diversity analysis. When comparing the average amounts of IBD sharing between pears from the southernmost region (Group3) and pears from other regions, the amount for the southern central region (Group4) was the highest, supporting the strong gene flow and population dispersal patterns between pear accessions from the southern central region and southernmost region.

For pears from the middle and lower reaches of the Yellow River (Group6), the average amount of IBD sharing with most of the pear groups from the southern regions was very high (Fig. 3c), especially for accessions in the southwest region (Group1), middle reaches of the Yangtze River

**Table 2** Pairwise differentiation ( $F_{ST}$ ) between eight pear groups

|        | Group1 | Group2 | Group3 | Group4 | Group5 | Group6 | Group7 | Group8 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Group1 |        |        |        |        |        |        |        |        |
| Group2 | 0.060  |        |        |        |        |        |        |        |
| Group3 | 0.053  | 0.068  |        |        |        |        |        |        |
| Group4 | 0.055  | 0.072  | 0.061  |        |        |        |        |        |
| Group5 | 0.064  | 0.063  | 0.061  | 0.055  |        |        |        |        |
| Group6 | 0.044  | 0.053  | 0.063  | 0.060  | 0.056  |        |        |        |
| Group7 | 0.074  | 0.106  | 0.097  | 0.113  | 0.115  | 0.087  |        |        |
| Group8 | 0.099  | 0.104  | 0.106  | 0.122  | 0.122  | 0.089  | 0.084  |        |

(Group2), and the southeast region (Group5). This indicates that pear accessions in Group6 have close genetic connections with the southern region, suggesting that pears from the middle and lower reaches of the Yellow River (Group6) may have close communication with neighboring areas in the south (Group2 and Group5). Of the Xinjiang pears from northwest China (Group7), the average amount of IBD sharing with the southwest region (Group1) was the highest and with the middle and lower reaches of the Yellow River (Group6) was the second highest. These data support the genetic diversity result that pears in the northwest region may have spread from the southwest and the middle and lower reaches of the Yellow River. For the Ussurian pears from the northeast region (Group8), the mean IBD sharing with pears from the middle and lower reaches of the Yellow River (Group6) was the highest, signifying that there was relatively strong gene flow between the samples from northeast China and the middle and lower reaches of the Yellow River. Moreover, the average IBD sharing was also high between pear samples from Group8 in the northeast region and Group2 in the middle reaches of the Yangtze River, indicating the occurrence of dissemination between Group2 and the adjacent Group6 and implying a close but indirect relationship between Group2 and Group8.

## Discussion

### Genetic diversity of pear revealed by genome-wide SNPs

Previous studies on pear germplasm resources or genetic relationships mostly used RFLP, RAPD, AFLP, and SSR markers. Compared with these traditional molecular markers, the SNP marker is based on whole-genome resequencing, has greater genome-wide density, has higher accuracy, and has more genetic information (Andari et al. 2020; Ma et al. 2019). In this study, we used resequencing data to obtain a total of 2.4 M SNP sites with a density of 4.74 SNP/kb, and the coverage density was much higher than that of other markers in pear research (Chang et al. 2017; Liu et al. 2015; Zong et al. 2014). Several diversity statistics determined in this study demonstrated the high level of genetic information in the 2.4 M SNP dataset.

The average value of  $H_E$  of the 102 pear samples obtained through our analysis was 0.2431. This differs from the genetic diversity and  $F_{ST}$  results obtained by researchers using SSR markers. For example, Xue et al. (2018) used 17 SSR markers to evaluate the genetic diversity of 470 sand pears and determined an average  $H_E$  of 0.78. Song et al. (2014) used 134 SSR markers to calculate the genetic diversity of 99 cultivated sand pears and obtained an average  $H_E$  of 0.70. The differences may be due to the molecular markers

used, and other previous research have shown that the  $H_E$  based on SSR markers was higher than that based on SNP markers applied to the same germplasms (Emanuelli et al. 2013; Lijavetzky et al. 2007). Meanwhile, previous studies have shown that  $H_E$  calculated by SSR cannot fully reflect the genome-wide genetic diversity of the investigated population and SNPs can more accurately estimate population-level diversity than SSR markers (Fischer et al. 2017; Zimmerman, Aldridge, and Oyler-McCance 2020). Moreover, the average value of  $F_{ST}$  between different groups obtained through our analysis was 0.082 (range: 0.044–0.122), which was higher than the  $F_{ST}$  value between different sand pear groups calculated by Xue et al. (2018) using SSR markers (mean: 0.0023; range: 0.001–0.005). This difference might due to different genetic information identified by SNP and SSR. In terms of availability and cost-effectiveness, SNP may be more suitable for  $F_{ST}$  calculation than SSR (Fischer et al. 2017).

### Historic routes of pear dispersal in China

Studies on historic patterns of pear dispersal in China using different molecular markers are supported and expanded upon by our study. Xue et al. (2018) speculated on the dissemination route of sand pear in the south from west to east. Yue et al. (2018) reported the dissemination pattern of pear in the south that spread into the north. Chang et al. (2017) explored the evolution route of northern pears based on cpDNA divergence. However, there are still different views on whether the middle reaches of the Yangtze River region were the center of pear dissemination between the south and north, and whether there was a route of pear dispersal from the middle and lower reaches of the Yellow River region to the northeast region of China. After a series of analyses, we not only provide evidence of the middle reaches of the Yangtze River region as an important geographical point of pear dispersal but also support the significance of the southwest and southeast regions in the process of northward dissemination. We also report a route of pear dissemination from the middle to lower reaches of the Yellow River region to the northeast region.

Genetic diversity studies have contributed much evidence to speculate about the histories of crop domestication and dissemination, and the reduction in genetic diversity can be used to infer the direction of crop spread and domestication (Warschefsky and von Wettberg 2019; Wuyun et al. 2015; Xue et al. 2017). Analyses based on IBD can explain relatedness between different populations (Campbell et al. 2012; Gusev et al. 2012; Ralph and Coop 2013). Accordingly, our results of genetic diversity and IBD analyses indicated that pear spread from a point of origin in the southwest (Group1) to the adjacent regions of the middle reaches of the Yangtze and Pearl Rivers (Group2 and Group4) and the

northwest region (Group7), and then it spread farther to the other southern (Group3 and Group5) and northern regions (Group6 and Group8) (Fig. 3d). We inferred more detailed relationships of dispersal among pear populations in various regions of southern China, particularly of how the pattern of spread was generally in line with the flow direction of the Yangtze and Pearl Rivers. This pattern is also consistent with that reported in previous research (Xue et al. 2018). The Yangtze River and the Pearl River Basin were important transportation routes in ancient China. Frequent biological and human activities in the two major watersheds likely led to the spread of pears. Records in ancient literature describe the cultivation of pear in the Yangtze River Basin since the Western Han Dynasty (Yuan 1993), and a large number of fine varieties were gradually developed (Shen 1965; Zhao 1965).

In general, pear dispersal occurred from southern to northern China. The pear accessions of our study had dissemination routes from the southwest region (Group1) going toward the northwest region (Group7), and from the middle reaches of the Yangtze River (Group2) going toward the middle and lower reaches of the Yellow River region (Group6). Meanwhile, the southeast region (Group5) and the middle and lower reaches of the Yellow River region (Group6) had similar genetic diversity level and high level of average IBD sharing. And the historical evidence of excellent pear tree cultivation in the Yangtze River Basin during the Western Han Dynasty (founded in 202 BC) (Yuan 1993) further supports that there is a possible dispersal pattern between pear accessions from Group5 to Group6. Therefore, our data suggest that there was close connectivity between pear populations in the adjacent areas of the north and south. In the north, the path from the middle and lower reaches of the Yellow River region (Group6) to the northwest (Group7) coincides with the ancient Silk Road. Coincidentally, historical records show that the renowned diplomat Zhang Qian of the Han Dynasty (founded in 202 BC) brought Asian pear trees to Xinjiang (Zhao 1965), which corroborates the existence of this dissemination route. Moreover, based on IBD analysis, sweep selection, and geographical data, a strong relationship was identified between pear accessions from Group6 and Group8 (northeast region), indicating that there was a route of dispersal from the middle and lower reaches of the Yellow River region to the three northeast provinces.

In the northern region of China, the middle and lower reaches of the Yellow River region (Group6) may have played important roles in the dissemination of pears. Shanxi Province, which is located in Group6's region, was the starting point of evolution of routes for northern pear accessions as reported by Chang et al. (2017). In addition, a record in the Chinese literature explains that when the historic Shanglin Garden was built in the city of Xi'an during the Western Han Dynasty (founded in 202 BC), there were

excellent pear varieties collected from all over the country and planted in Shanglin Garden (He 2005; Liu 2006). In ancient times, many dynasties had their capital in this region; therefore, the significance of Group6's region as a central location of pear dispersal in northern China is reasonable. The extensive development of pear breeding and the political and commercial activities between different regions since ancient times have contributed to the dissemination patterns observed among pear populations (Xue et al. 2018).

### Genetic basis of Ussurian pear in northeast China

The Ussurian pear accessions of Group8 were derived from samples from the extremely cold northeast region of China. These accessions had relatively distant relationships with other pear populations, which is consistent with previous studies (Iketani et al. 2012; Wu et al. 2018). The population structure results identified a relatively single genetic background for the Ussurian pear population, indicating that Ussurian pear had diverged long ago from the other pear groups. At the same time, based on the results of population structure (Fig. 2; Fig. S1), Ussurian pear showed partial composition of genetic background similar with sand pear and white pear. With the  $K$  value increase from 2 to 8, Ussurian pear showed higher genetic background from white pear than sand pear. Secondly, combined with IBD analysis, we found that the gene flow between Ussurian pears and Group6 (mainly composed of white pears) was the highest, then followed by Group2 (mainly composed of sand pear). As a whole, we supposed that white pear contributed more than sand pear to the Ussurian pear.

Through selection sweep analysis of wild and cultivated samples of Ussurian pear, we found that the degree of domestication of this species was lower than that of sand pear. We suspect that a weak selection process may have contributed to the higher genetic diversity observed in cultivated Ussurian pear populations. This may also support the proposed hypothesis in previous studies that the cultivated Ussurian pear may be a result of hybridization between wild Ussurian pear and cultivated sand pear, instead of direct domestication from wild Ussurian pear (Shuang et al. 2016; Yu et al. 2016). Moreover, the study of genetic diversity also helps to protect and better utilize existing crop resources (Erfani et al. 2012; Ferradini et al. 2017; Zong et al. 2014), and it provides a basis for formulating conservation strategies (Millar and Westfall 1992; Moritz 1994; Newton et al. 1999). The genetic diversity of Group8, which was composed of Ussurian pear accessions, was higher than that of other pear populations, including the pear population in southwest China. The greater genetic diversity confers a higher conservation value to Ussurian pears in northeast China.

## Conclusion

In summary, a total of 102 pear samples from 23 provinces of China were divided into two major clades and eight geographic groups based on our genome-wide SNP data. The route of pear dispersal in southern China was consistent with the direction of flow of the Yangtze and Pearl Rivers, and pear accessions also spread from southern China to the neighboring areas in the north. The middle and lower reaches of the Yellow River region also played an important role in the further spread of pear in northern China. Our results provide insight into the diversity as well as the dispersal routine of pears in China. The data also provide a solid foundation for further study of pear genetic diversity and domestication, and this study may benefit germplasm resource protection and utilization in the future.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11295-021-01530-x>.

**Acknowledgements** This project was supported by the Bioinformatics Center of Nanjing Agricultural University.

**Author contribution** JW designed this project. XNC and MYZ drafted the manuscript. XNC collected the public dataset and performed bioinformatics analysis. MYZ and MYS contributed to results analysis. BBS contributed to data analysis. SNL worked with data collection. YYL and MYL reviewed the manuscript. SLZ provided help to the platform. RZW, JML, and KJZ provided the suggestion to do this study. All authors read and approved the final manuscript.

**Funding** This work was funded by the Earmarked Fund for Jiangsu Agricultural Industry Technology System (JATS [2020]401), the National Science Foundation of China (31820103012, 31725024, 31801835, 31901983), and China Agriculture Research System of MOF and MARA.

**Data availability** The datasets generated and analyzed during the current study are available from the corresponding author upon request. Raw genome resequencing reads have been deposited into the NCBI sequence read archive (SRA) under BioProject accession PRJNA782471.

**Code availability** The software application during the current study is downloaded from its official website, and custom codes are available from the corresponding author upon request.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

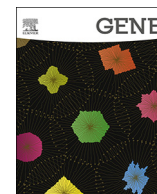
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664
- Andari MVC, Bussamra SLC, Tedesco TGD, Peixoto PAB, Pares P, Braga A, Araujo Júnior E, Aoki T (2020) Noninvasive prenatal testing: benefits and limitations of the available tests. *Ceska Gynekol* 85:41–48
- Bell RL (1984) Evaluation of *Pyrus* germplasm for resistance to pear Psylla in the orchard. *Acta Hort* 161:133–134
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Amer J Human Genet* 81:1084–1097
- Campbell CL, Palamara PF, Dubrovsky M, Botigue LR, Fellous M, Atzmon G, Oddoux C, Pearlman A, Hao L, Henn BM (2012) North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. *P Natl Acad Sci USA* 109:13865–13870
- Cao K, Li Y, Deng CH, Gardiner SE, Zhu GR, Fang WC, Chen CW, Wang XW, Wang LR (2019) Comparative population genomics identified genomic regions and candidate genes associated with fruit domestication traits in peach. *Plant Biotechnol J* 17:1954–1970
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
- Challice JS, Westwood MN (1973) Numerical taxonomic studies of the genus *Pyrus* using both chemical and botanical characters. *Bot J Linn Soc*: 121–48
- Chang YJ, Cao YF, Zhang JM, Tian LM, Dong XG, Zhang Y, Qi D, Zhang XS (2017) Study on chloroplast DNA diversity of cultivated and wild pears (*Pyrus* L.) in Northern China. *Tree Genet Genomes* 13: 44
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Dolotowski J, Zych M, Podyma W, Nowosielski J, Szymańska M (2004) Molecular studies on the variability of Polish semi-wild pears (*Pyrus*) using AFLP. *J Fruit Ornament Plant Res* 12:331–337
- Duan NB, Bai Y, Sun HH, Wang N, Ma YM, Li MJ, Wang X, Jiao C, Legall N, Mao LY, Wan SB, Wang K, He TM, Feng SQ, Zhang ZY, Mao ZQ, Shen X, Chen XL, Jiang YM, Wu SJ, Yin CM, Ge SF, Yang L, Jiang SH, Xu HF, Liu JX, Wang DY, Qu CZ, Wang YC, Zuo WF, Xiang L, Liu C, Zhang DY, Gao Y, Xu YM, Xu KN, Chao T, Fazio G, Shu HR, Zhong GY, Cheng LL, Fei ZJ, Chen XS (2017) Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun* 8:1–11
- Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troglio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grando MS (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol* 13:39
- Erfani J, Ebadi A, Abdollahi H, Fatahi R (2012) Genetic diversity of some pear cultivars and genotypes using simple sequence repeat (SSR) markers. *Plant Mol Biol Rep* 30:1065–1072
- Ferradini N, Lancioni H, Torricelli R, Russi L, Ragione ID, Cardinali I, Marconi G, Gramaccia M, Concezzi L, Achilli A, Veronesi F, Albertini E (2017) Characterization and phylogenetic analysis of ancient Italian landraces of pear. *Front Plant Sci* 8
- Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2017) Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics* 18:69
- Ginestet C (2011) ggplot2: elegant graphics for data analysis. *J Roy Stat Soc* 174:245–246



- Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461–485
- Gusev A, Palamara PF, Aponte G, Zhuang Z, Darvasi A, Gregersen P, Pe'er I (2012) The architecture of long-range haplotypes shared within and across populations. *Mol Biol Evol* 29:473–86
- Han Z, Hu Y, Tian Q, Cao Y, Si A, Si Z, Zang Y, Xu C, Shen W, Dai F (2020) Genomic signatures and candidate genes of lint yield and fibre quality improvement in Upland cotton in Xinjiang. *Plant Biotechnol J* 18:1–13
- He Q (2005) Collation and interpretation of panoramic maps of the three county level administrative districts in Qin and Han Dynasties (in Chinese). ZhongHua Book Company, Beijing
- Iketani H, Katayama H, Uematsu C, Mase N, Sato Y, Yamamoto T (2012) Genetic structure of East Asian cultivated pears (*Pyrus* spp.) and their reclassification in accordance with the nomenclature of cultivated plants. *Plant Syst Evol* 298:1689–1700
- Iketani H, Manabe T, Matsuta N, Akihami T, Hayashi T (1998) Incongruence between RFLPs of chloroplast DNA and morphological classification in east Asian pear (*Pyrus* spp.). *Genet Resour Crop Ev* 45:533–539
- Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conserv Genet* 5:539–543
- Kumar S, Kirk C, Deng C, Wiedow C, Knaebel M, Brewer L (2017) Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. *Hortic Res* 4
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPD (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
- Li XL, Singh J, Qi MF, Li SW, Zhang X, Zhang MY, Khan A, Zhang SL, Wu J (2019) Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol J* 17:1582–1594
- Lijavetzky D, Cabezas JA, Ibanez A, Rodriguez V, Martinez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8
- Liu QW, Song Y, Liu L, Zhang MY, Sun JM, Zhang SL, Wu J (2015) Genetic diversity and population structure of pear (*Pyrus* spp.) collections revealed by a set of core genome-wide SSR markers. *Tree Genet Genomes* 11:1–22
- Liu X (2006) Miscellany of the western capital (in Chinese). Xi'an: SANQIN Publishing House, Xi'an
- Ma H, Wang S, Zeng G, Guo J, Guo M, Dong X, Hua G, Liu Y, Wang M, Ling Y, Ding X, Zhao C, Wu C (2019) The origin of a coastal indigenous horse breed in China revealed by genome-wide SNP data. *Genes* 10
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Millar CI, Westfall RD (1992) Allozyme markers in forest genetic conservation. *Populat Genet Forest Trees* 42:347–371
- Montanari S, Bianco L, Allen BJ, Martinez-Garcia PJ, Bassil NV, Postman J, Knabel M, Kitson B, Deng CH, Chagne D, Crepeau MW, Langley CH, Evans K, Dhingra A, Troglio M, Neale DB (2019) Development of a highly efficient Axiom™ 70 K SNP array for *Pyrus* and evaluation for high-density mapping and germplasm characterization. *BMC Genomics* 20:1–18
- Moritz C (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. *Mol Ecol* 3:401–411
- Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends Ecol Evol* 14:140–145
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
- Park B, Burke JM (2020) Phylogeography and the evolutionary history of sunflower (*Helianthus annuus* L.): wild diversity and the dynamics of domestication. *Genes* 11:1–18
- Pu F, Wang Y (1963) Pomology of China pears (in Chinese). Shanghai Scientific and Technical Publishers, Shanghai
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Ralph P, Coop G (2013) The geography of recent genetic ancestry across Europe. *Plos Biol* 11: e1001555
- Rubtsov GA (1944) Geographical distribution of the genus *Pyrus* and trends and factors in its evolution. *Amer Naturalist* 78:358–366
- Shen Z (1965) Records of Kuaiji (in Chinese). Nanjing: manuscripts of China Agricultural Heritage Museum, Nanjing
- Shi S, Li J, Sun J, Yu J, Zhou S (2013) Phylogeny and classification of *Prunus sensu lato* (Rosaceae). *J Integr Plant Biol* 55:1069–1079
- Shuang J, Zheng X, Yu P, Yue X, Maqsood A, Cai D, Teng Y, Takaya MJPO (2016) Primitive genepools of Asian pears and their complex hybrid origins inferred from fluorescent sequence-specific amplification polymorphism (SSAP) markers based on LTR retrotransposons. *Plos One* 11: e0149192
- Song Y, Fan L, Chen H, Zhang M, Ma Q, Zhang S, Wu J (2014) Identifying genetic diversity and a preliminary core collection of *Pyrus pyrifolia* cultivars by a genome-wide set of SSR markers. *Sci Hortic* 167:5–16
- Tam NT, Dwiyantri MS, Koide Y, Nagano AJ, Ky H, Tin HQ, Hien NL, Dung LV, Kishima Y (2019) Profiling SNP and nucleotide diversity to characterize Mekong Delta rice landraces in Southeast Asian populations. *Plant Genome* 12:1–11
- Teng YW, Tanabe K, Tamura F, Itai A (2002) Genetic relationships of *Pyrus* species and cultivars native to East Asia revealed by randomly amplified polymorphic DNA markers. *J Am Soc Hortic Sci* 127:262–270
- Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Ronald Press, New York
- Wang X, Feng H, Chang Y, Ma C, Wang L, Hao X, Ai Li, Cheng H, Wang L, Cui P, Jin J, Wang X, Wei K, Ai C, Zhao S, Wu Z, Li Y, Liu B, Wang G-D, Chen L, Ruan J, Yang Y (2020) Population sequencing enhances understanding of tea plant evolution. *Nat Commun* 11:1–10
- Warschefsky EJ, von Wettberg EJB (2019) Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication. *New Phytol* 222:2023–2037
- Wu J, Wang YT, Xu JB, Korban SS, Fei ZJ, Tao ST, Ming R, Tai SS, Khan AM, Postman JD, Gu C, Yin H, Zheng DM, Qi KJ, Li Y, Wang RZ, Deng CH, Kumar S, Chagne D, Li XL, Wu JY, Huang XS, Zhang HP, Xie ZH, Li X, Zhang MY, Li YH, Yue Z, Fang XD, Li JM, Li LT, Jin C, Qin MF, Zhang JY, Wu X, Ke YQ, Wang J, Yang HM, Zhang SL (2018) Diversification and independent domestication of Asian and European pears. *Genome Biol* 19
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, Chen NJ, Nishio T, Xu X, Cong L, Qi K, Huang X, Wang Y, Zhao X, Wu J, Deng C, Gou C, Zhou W, Yin H, Qin G, Sha Y, Tao Y, Chen H, Yang Y, Song Y, Zhan D, Wang J, Li L, Dai M, Gu C, Wang Y, Shi D, Wang X, Zhang H, Zeng L, Zheng D, Wang C, Chen M, Wang G, Xie L, Sovero V, Sha S, Huang W, Zhang S, Zhang M, Sun J, Xu L, Li Y, Liu X, Li Q,

- Shen J, Wang J, Paull RE, Bennetzen JL, Wang J, Zhang S (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res* 23:396–408
- Wuyun T, Amo H, Xu JS, Ma T, Uematsu C, Katayama H (2015) Population structure of and conservation strategies for wild *Pyrus ussuriensis* Maxim. in China. *PLoS ONE* 10:1–20
- Xue L, Liu QW, Hu HJ, Song Y, Fan J, Bai B, Zhang MY, Wang RZ, Qin MF, Li XL, Wu J (2018) The southwestern origin and eastward dispersal of pear (*Pyrus pyrifolia*) in East Asia revealed by comprehensive genetic structure analysis with SSR markers. *Tree Genet Genomes* 14:1–12
- Xue L, Liu QW, Qin MF, Zhang MY, Wu X, Wu J (2017) Genetic variation and population structure of “Zangli” pear landraces in Tibet revealed by SSR markers. *Tree Genet Genomes* 13:1–11
- Yu D (1979) Taxonomy of the fruit tree in China. China Agr Press, Beijing
- Yu P, Shuang J, Wang X, Bai S, Teng Y (2016) Retrotransposon-based sequence-specific amplification polymorphism markers reveal that cultivated *Pyrus ussuriensis* originated from an interspecific hybridization. *Eur J Hort Sci* 81:264–272
- Yuan K (1993) Proofreading and annotation of Shanhaijing (in Chinese). Bashu Publishing House, Sichuan
- Yue XY, Zheng XY, Zong Y, Jiang S, Hu CY, Yu PY, Liu GQ, Cao YF, Hu HJ, Teng YW (2018) Combined analyses of chloroplast DNA haplotypes and microsatellite markers reveal new insights into the origin and dissemination route of cultivated pears native to East Asia. *Front Plant Sci* 9
- Zhang MY, Xue C, Hu H, Li J, Xue Y, Wang R, Fan J, Zou C, Tao S, Qin M (2021) Genome-wide association studies provide insights into the genetic determination of fruit traits of pear. *Nat Commun* 12:1–10
- Zhao H (1965) General annals of Jiangnan (in Chinese) Nanjing: manuscripts of China Agricultural Heritage Museum, Nanjing
- Zhukovsky PM (1965) Main gene centres of cultivated plants and their wild relatives within the territory of the U.S.S.R. *Euphytica* 14:177–188
- Zimmerman SJ, Aldridge CL, Oyler-McCance SJ (2020) An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics* 21:382
- Zong Y, Sun P, Liu J, Yue XY, Niu QF, Teng YW (2014) Chloroplast DNA-based genetic diversity and phylogeography of *Pyrus betulaefolia* (Rosaceae) in Northern China. *Tree Genet Genomes* 10:739–749

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



## Research paper

# Identification of key genes related to seedlessness by genome-wide detection of structural variation and transcriptome analysis in ‘Shijiwuhe’ pear



Yueyuan Liu<sup>a,1</sup>, Jieying Sun<sup>a,1</sup>, Mingyue Zhang<sup>a,1</sup>, Guangyan Yang<sup>a</sup>, Runze Wang<sup>a</sup>, Jintao Xu<sup>b</sup>, Qingyu Li<sup>c</sup>, Shaoling Zhang<sup>a</sup>, Wenquan Le<sup>b</sup>, Baofeng Hao<sup>b</sup>, Yuanjun Li<sup>c</sup>, Jun Wu<sup>a,\*</sup>

<sup>a</sup> Center of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

<sup>b</sup> Changli Institute of Pomology, Hebei Academy of Agricultural and Forestry Sciences, Changli, Hebei 066600, China

<sup>c</sup> Yantai Academy of Agricultural Sciences, Shandong 264000, China

## ARTICLE INFO

## Keywords:

‘Shijiwuhe’ pear

Seedlessness

Genomic structural variation

RNA-seq

## ABSTRACT

Seedless fruits are highly marketable because they are easier to eat than fruits with seeds. ‘Shijiwuhe’ is a seedless pear cultivar that is a mutant derived from an F1 hybridization population (‘Bartlett’ x ‘Yali’). Little is known about the key genes controlling seedless pear fruit. In this study, field experiments revealed that seedless ‘Shijiwuhe’ pear was not due to parthenocarpy, and that it was self-incompatible. Single nucleotide polymorphisms (SNPs), small insertions and deletions (InDels) and structural variations (SVs) were characterized using DNA sequencing data between ‘Shijiwuhe’ and parental cultivars. A total of 1498 genes were found to be affected by SV and over 50% of SVs were located in promoter regions. Transcriptome analysis was conducted at three time points (4, 8, and 12 days after cross-pollination) during early fruit development of ‘Shijiwuhe’, ‘Bartlett’, and ‘Yali’. In total, 1438 differentially expressed genes (DEGs) were found between ‘Shijiwuhe’ and parental cultivars ‘Bartlett’ and ‘Yali’. We found 1193 SVs that caused differential expression of genes at 4 DACP. Among them, over 100 genes were in pathways related to seed nutrition and energy storage and 41 candidate genes encoded several important transcription factors, such as *MYB*, *WRKY*, *NAC*, and *bHLH*, which might play important roles in seed development. The qRT-PCR results also confirmed that the candidate genes with SVs showed differential expression between ‘Shijiwuhe’ pear and ‘Bartlett’ or ‘Yali’. This study, which combined field experiments, SV detection, and transcriptome analysis might provide an effective way to predict the candidate genes regulating the seedless trait and important gene resources for genetic improvement of pear.

## 1. Introduction

Seedlessness is a major goal for fruit breeders around the world and consumers’ preference for seedless fruit has increased because seedless fruit is easier to eat (Zhang et al., 2017). Seedless fruits are thus desirable for markets in many countries and more seedless fruits have been selected and popularized in recent decades, including grape (Lo’ay and El-Boray, 2018), litchi (Chu et al., 2015), citrus (Ma et al., 2017), mango (Singh, 2005), and loquat (Blasco et al., 2015). Seedlessness can be obtained through phytohormones, female sterility and male sterility, pollination or fertilization failure, embryo abortion, parthenocarpy, and various other factors (Zhang et al., 2017).

A variety of agricultural species are made seedless by the exogenous application of auxins, cytokinins, or gibberellin (GA), indicating that a number of independent and possibly redundant hormone pathways can promote seed development (Chiwocha et al., 2005). Auxins play a significant role in the development of embryo, endosperm, seed, and fruit. In tomatoes, *Aux/IAA* genes were reported to give rise to parthenocarpy by triggering fruit development before fertilization (Wang et al., 2005). In fruit crops, spraying GA<sub>3</sub> can significantly induce seedless fruit (Koura et al., 2004). In addition to hormones causing seedlessness, female sterility includes abortion of female organs, and degeneration of nutritional organs and megasporogenesis (Pannell and Ojeda, 2000). Male sterility includes the failure of plants to produce

**Abbreviations:** SV, structural variation; DACP, days after cross-pollination; DEGs, differentially expressed genes; SNP, single nucleotide variation; InDels, small insertions and deletions; SR, split-read; RP, read-pair; SAM, Sequence alignment/map format; BAM file, Binary version of SAM file; BED, Browser Extensible Data; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ARF, auxin response factor

\* Corresponding author.

E-mail address: [wujun@njau.edu.cn](mailto:wujun@njau.edu.cn) (J. Wu).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.gene.2020.144480>

Received 31 August 2019; Received in revised form 14 February 2020; Accepted 14 February 2020

Available online 17 February 2020

0378-1119/© 2020 Elsevier B.V. All rights reserved.

functional anthers, pollen, or gametes (Budar and Pelletier, 2001). Pollination or fertilization failure are caused by self-incompatibility or cross incompatibility (Wilcock and Neiland, 2002). Embryo abortion varies in time and anatomical characteristics among species (Bawa et al., 1989); parthenocarpy is caused by a lack of fertilization or embryo abortion (Hazra and Dutta, 2011).

Lack of certain nutrient elements can also lead to seed abortion. For example, sugar, as a nutrient, is an example of a small molecule involved in the communication between seed structures. Seed development needs nutrients and forms a carbon sink with high sugar flow from vegetative tissues to seeds (Robert, 2019). In addition, amino acids play pivotal roles in the central metabolism of seeds. They are mainly used for synthesis of seed-storage proteins and are precursors for the biosynthesis of secondary metabolites as well as a source of energy (Amir et al., 2018). Changes in lipid accumulation and fatty acid composition have huge effects on seed growth (Chen et al., 2017). Hormone metabolism and response (auxin in particular), and metabolism of sugars and lipids were identified to induce parthenocarpy in tomato (Martinelli et al., 2009). However, little is known about the key genes and genetic mechanism controlling for seedless fruit.

DNA diversity between cultivars can be studied with a high quality reference genome by mapping millions of reads generated from next-generation sequencing. Based on the advantages of next-generation DNA-sequencing, some bioinformatics tools were developed to detect DNA variation, including SNPs (single nucleotide variations), InDels (small insertions and deletions), and SVs (structural variations) (Medvedev et al., 2009; Nielsen et al., 2011). SVs include deletion, insertion, inversion, and duplication, and are often longer than 50 nucleotides (Alkan et al., 2011). The occurrence of SVs in crops was determined to be associated with the number of flowers (Zhang et al., 2015), the size of leaves (Horiguchi et al., 2009), the shape of fruits (Xiao et al., 2008), and other important phenotypic traits. Whether the occurrence of SVs is associated with molecular mechanisms of seedless fruit is still unknown.

Next-generation RNA-sequencing can generate an enormous amount of data effectively. Transcriptome analysis can provide comprehensive molecular biology insights through differentially expressed genes (DEGs) into the traits of interest and identify a series of important candidate genes regulating target traits. Furthermore, RNA sequencing supports the accuracy of DEG analysis based on the reference genome. In previous studies, high-throughput RNA-sequencing was used to determine: the molecular mechanism of accumulation of vitamin C in jujube fruit (Liu et al., 2014); the genes involved in glucose metabolism on the mature development process of bayberry fruit (Feng et al., 2012); and the observations related to changes in aroma substances in peach (Zhang et al., 2011).

'Shijiwuhe' pear is a mutant cultivar with no seeds derived from 'Bartlett' × 'Yali' F1 progeny. To identify the basis of genetic variation and identify the candidate genes that may contribute to the seedlessness of 'Shijiwuhe' pear, we first observed the pollen tube growth of 'Shijiwuhe' pear in field experiments. Then, DNA and RNA datasets of the seedless 'Shijiwuhe' and the parental cultivars with seeds were analyzed. Transcriptome analysis and the genes from structural variation detection were combined to identify candidate genes associated with seedlessness. qRT-PCR was used to validate the accuracy of candidates. Ultimately, the findings of this study provide valuable information concerning the mechanism of seedlessness in pear and other fruit species.

## 2. Materials and methods

### 2.1. DNA sequencing analysis

#### 2.1.1. Plant materials and DNA sequencing

For DNA resequencing materials, young pear leaves after 15 days of flowering were collected from the orchard of the Changli Institute of

Pomology in Hebei Province and Yantai Academy of Agricultural Sciences, China. DNA extraction was performed using the Tiangen plant genomic DNA kit following the manufacturer's protocol. Paired-end sequencing libraries with an insert size of approximately 350 bp were sequenced using an Illumina HiSeq™ 2000 system at the Biomarker Technologies Company (Beijing, China).

#### 2.1.2. Detection and analysis of genomic variation of 'Shijiwuhe' – 'Bartlett' and 'Shijiwuhe' – 'Yali'

For the DNA re-sequencing data, FastQC was used to check the quality of raw sequencing data (Kim et al., 2018). Trimmomatic was used to remove reads containing adapters, reads containing poly-N, and low quality reads from the raw data, yielding clean data for further analysis (Bolger et al., 2014). The 'Dangshansuli' genome was used as a reference genome. SNP and InDel detection were performed using BWA (Li and Durbin, 2010) and Genome Analysis Toolkit (McKenna et al., 2010) (GATK, v4.1.3.0) with the following filtering options: Quality by depth (QD) < 2.0, Fisher strand (FS) > 60.0, RMS mapping quality (MQ) < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, Maximum depth (DP) < 20.

Three software packages (Pindel (Ye et al., 2009), DELLY (Rausch et al., 2012) and DENOM (Rimmer et al., 2014)) were chosen to detect SVs and then find the overlapped SVs. Clean reads were mapped to the 'Dangshansuli' genome with BWA (bwa aln). SAMtools (Li et al., 2009) was used to convert the obtained SAM file to a BAM file and remove duplicate reads. Pindel (using the -f -p -o parameters) and DELLY (using the -t -o parameters) were used to detect SVs from the BAM file. DENOM (Rimmer et al., 2014) was used to detect SVs based on assembly. For detecting SVs with assembly algorithms, the short reads were assembled using SOAPdenovo2 (Luo et al., 2015). Before assembly, kmergenie (Chikhi and Medvedev, 2014) was used to predict the best k-mer length for each *de novo* genome assembly. The obtained contigs were mapped to the 'Dangshansuli' genome using BWA (Li and Durbin, 2009b) and then the contig reads were aligned to the 'Dangshansuli' pear genome and the BAM file was generated using BWA and SAMtools. Finally, 'denom varcall' was used to detect SVs from the BAM file.

#### 2.1.3. Genome variation annotation and GO and KEGG analysis

The functional annotation related analysis of SNPs and InDels was performed using SnpEff (Cingolani et al., 2012) software. The same positions of genomic variants in 'Bartlett' – 'Shijiwuhe' and 'Yali' – 'Shijiwuhe' were filtered. The breakpoints of the SVs from the output of three algorithms were converted to BED format with in-house scripts. Bedtools (Quinlan and Hall, 2010) was used to find the overlapping SVs in the BED files. Overlapping SVs longer than 50 bp were retained. Then, the breakpoint of SVs in BED format were collected for functional annotation using InterProScan (Zdobnov and Apweiler, 2001) and Gene Ontology (GO) enrichment analysis, including biological process, molecular function, and cellular component.

Based on the Gene Ontology (GO) annotation information for all genes in the pear genome, WEGO (<http://wego.genomics.org.cn/>) was used to perform GO analysis. GO terms with corrected *P-value* < 0.05 were considered significantly enriched for differentially expressed genes (DEGs). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding high-level functions and utilities of biological systems, such as cells, organisms, and ecosystems, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (<http://www.genome.jp/kegg/>). We used KOBAS (Xie et al., 2011) software to test the statistical enrichment of DEGs in the KEGG pathway.



## 2.2. RNA sequencing analysis

### 2.2.1. Plant materials and RNA extraction

Young fruits of ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ were collected at 4, 8, and 12 days after cross-pollination (DACP) from the orchard of the Changli Institute of Pomology in Hebei Province and Yantai Academy of Agricultural Sciences during the 2016 growing season. All samples were screened randomly for uniform size and absence of mechanical damage. Part of the samples were used to observe the morphology and the seeds, and the rest of the samples were cut into pieces and immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

Total RNA was extracted using a Plant Total RNA Isolation Kit Plus (World’s Foregene) according to the manufacturer’s instructions. RNA degradation and contamination were monitored on 1% agarose gels. RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using a Qubit® RNA Assay Kit in a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). The integrity of RNA was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

A total of 18 independent libraries named Wz1-1, Wz1-2, B1-1, B1-2, Y1-1, Y1-2 for 4 DACP; Wz2-1, Wz2-2, B2-1, B2-2, Y2-1, Y2-2 for 8 DACP; and Wz3-1, Wz3-2; B3-1, B3-2, Y3-1, Y3-2 for 12 DACP were prepared following the instructions of the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA). A total of 3  $\mu\text{g}$  RNA per sample was used as input material for the RNA sample preparation. The libraries were sequenced on an Illumina HiSeq 2000 platform at the Beijing Genomics Institute (BGI, Shenzhen, China) and raw reads were generated in 125-bp paired-end format.

### 2.2.2. Quality control and read mapping to the reference genome

Because the female parent ‘Bartlett’ and the male parent ‘Yali’ belong to occidental pear (Chagne et al., 2014) and oriental pear (Wu et al., 2013), respectively, the hybrid progeny ‘Shijiwuhe’ inherits the genetic background of oriental pear and occidental pear. So, we intended to map two pear genomes (‘Dangshansuli’ and ‘Bartlett’) to allow us to mine the key genes regulating the molecular mechanism of seedlessness. For the RNA sequencing data, the raw data in fastq format were processed using Trimmomatic for removing low quality reads and obtaining clean reads. At the same time, Q20, Q30, and GC content of the clean data were calculated. The following analyses were based on the clean data with high quality. The index of the reference genome was built using Bowtie (Li and Durbin, 2009a) and clean reads were aligned to ‘Dangshansuli’ (*Pyrus bretschneideri*) and ‘Bartlett’ (*P. communis*) (Chagne et al., 2014) CDS sequences by TopHat (Trapnell et al., 2009).

### 2.2.3. Transcript annotation and gene expression analysis

Transcript expression was calculated using the FPKM method (expected number of Fragments Per Kilobase of transcript sequence per Millions of base pairs sequenced), which simultaneously considers the effect of sequencing depth and gene length for the read counts (Garber et al., 2011). For each sequenced library, prior to differential gene expression analysis, the read counts were adjusted using the edgeR program package with a one-scale normalizing factor (Robinson et al., 2010). The differential expression analysis of two groups was performed using the DESeq R package (1.18.0) (Anders and Huber, 2010). Genes with  $P\text{-value} < 0.05$  found by DESeq were set as the threshold for significantly differential expression.

### 2.2.4. Validation of candidate genes by real-time quantitative PCR

By merging the results of DNA sequencing and RNA sequencing, hundreds of candidate genes were generated. The genes selected for heatmap and further validation met two conditions: the  $|\log_2(\text{fold change})| \geq 2$  and the genes affected by SVs. Then, according to the performance of each gene in the heatmap, several candidate genes were chosen for validation by quantitative RT-PCR (qRT-PCR). Coding

sequences of the genes were acquired from the pear genome project (<http://peargenome.njau.edu.cn>) and the primers were designed (Supplemental Table S1) using the online software Primer 5 (<http://singene.com/Primer5>). qRT-PCR was performed using a LightCycler 480 SYBR GREEN Master (Roche, USA) according to the manufacturer’s instructions. The raw data were analyzed with LightCycler 480 Software release version 1.5.0 (Roche), and the gene expression levels were determined with the  $2^{-\Delta\Delta\text{Ct}}$  algorithm (Livak and Schmittgen, 2001). Reactions of selected genes and reference genes were performed in triplicate in a total volume of 20  $\mu\text{l}$ . The protocol of real-time PCR was as follows: initial incubation at  $95^{\circ}\text{C}$  for 10 min and at  $95^{\circ}\text{C}$  for 15 s, and then cycled at  $60^{\circ}\text{C}$  for 15 s, and  $72^{\circ}\text{C}$  for 20 s for 40 cycles, and final extension at  $72^{\circ}\text{C}$  for 30 min. A melting curve was performed from 60 to  $95^{\circ}\text{C}$  to check the specificity to the amplified product.

## 3. Results and discussion

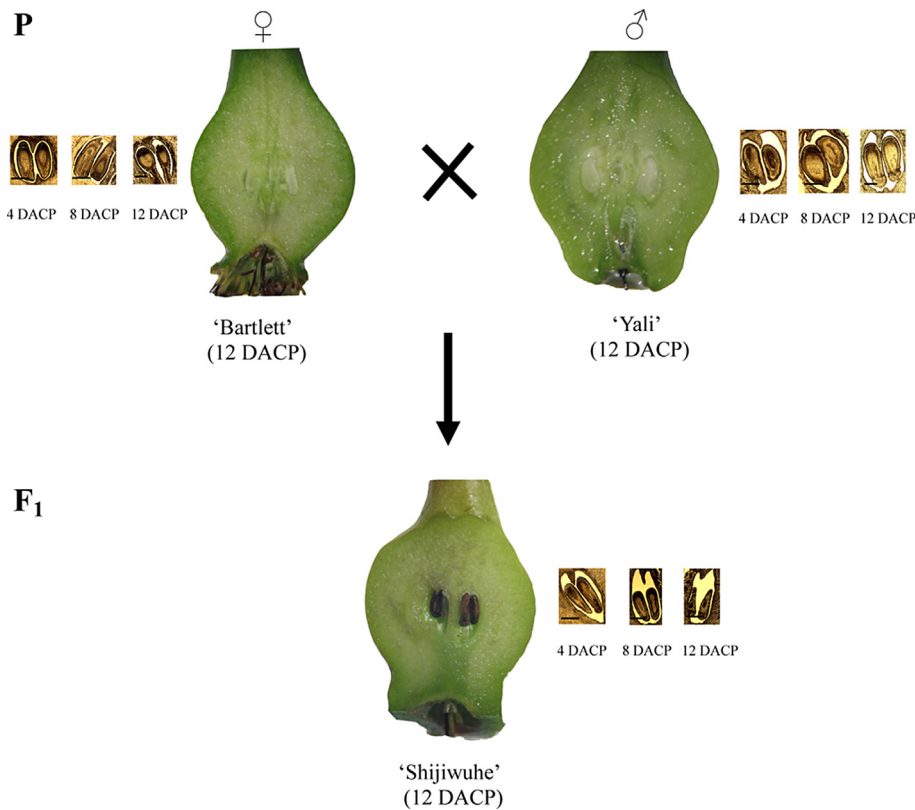
### 3.1. ‘Shijiwuhe’ pear showed seed abortion during fruit development

To verify whether the seedlessness of ‘Shijiwuhe’ pear was caused by parthenocarpy, we did a field experiment of artificial emasculum followed by bagging. The fruit setting rate was zero at 12 DACP (days after cross-pollination). The reason for seedlessness of ‘Shijiwuhe’ was therefore not parthenocarpy. To verify whether ‘Shijiwuhe’ was self-incompatible, we did field experiments of self-pollination with bagging (a) and the pollen of ‘Cuiguan’ was used to pollinate ‘Shijiwuhe’ after artificial emasculum and then bagging (b). After 96–120 h of pollination, we collected the pistils of experimental plants to observe their pollen tube growth referring to the method of Sogo and Tobe (2008) using fluorescence microscopy. The results showed that the pollen tube was elongated to about 1/3 of the pollen tube after self-pollination, while the pollen tube of treatment b could be extended to the bottom of the style when cross-pollinated. Therefore, ‘Shijiwuhe’ pear was self-incompatible. Thus, the fruit setting rate of self-pollination was zero, whereas it was 63.56% with cross-pollination, further indicating that the reason for seedlessness in ‘Shijiwuhe’ happened in the ovary, so the ovary was used for further RNA-sequencing.

We collected the young fruit of ‘Shijiwuhe’ at 4 DACP, 8 DACP, and 12 DACP as materials for studying seedlessness by observing seed abortion in early fruit development. The ovules of ‘Shijiwuhe’ changed in color and volume compared with its parents ‘Bartlett’ and ‘Yali’ over time. According to phenotypic observation, the ovary samples collected at 4 days, 8 days and 12 days post anthesis covered three key developmental stages of ‘Shijiwuhe’ as materials; that is, ovary at 4 DACP, ovary at 8 DACP with seed partially black and the size of seeds no longer increasing, and ovary at 12 DACP with completely black and obviously small seeds (Fig. 1). The parental seeds from Bartlett’ and ‘Yali’ had normal growth based on both volume and color (Fig. 1). In following studies, ‘Shijiwuhe’ and parental cultivars of ‘Bartlett’ and ‘Yali’ were used to scan the genetic variation and differentially expressed genes (DEGs).

### 3.2. Identification of genomic variation between seedless ‘Shijiwuhe’ and seeded parental cultivars

The raw data of the pear genomes from ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ was generated using the Illumina HiSeq™ 2000 platform for paired-end sequencing, and the sequencing depth was  $60 \times$ . The quality of the raw data was determined using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) software. The total number of reads in ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ genomes were 111,400,760, 103,584,796, and 108,689,525, respectively, with 150 paired-end sequencing, and the GC contents were 40.10%, 39.95%, and 40.05%, respectively (Supplemental Table S2). The percentages of clean reads were all over 95% after filtering the low quality reads by Trimmomatic. Clean reads of ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ were



**Fig. 1.** Seed development of three pear cultivars. The fruits of ‘Yali’, ‘Bartlett’, and ‘Shijiwuhe’ at 12 DACP (12 days after cross-pollination). Longitudinal sections of three cultivars’ seeds from three periods (4 DACP, 8 DACP, 12 DACP). Scale bar is 0.5 mm.

aligned to the reference genome using BWA and SAMtools (Li and Durbin, 2009b; Li et al., 2009). Percentages of the mapped rates were 95.58%, 97.15%, and 96.38%, respectively.

> 5 million SNPs were identified between ‘Bartlett’ and ‘Shijiwuhe’ and over 8 million SNPs were found between ‘Yali’ and ‘Shijiwuhe’ (Table 1). The same SNPs and InDels between ‘Bartlett’ and ‘Shijiwuhe’ and between ‘Yali’ and ‘Shijiwuhe’ were deleted. In total, the number of genes with SNPs between ‘Shijiwuhe’ and seeded parents was over 30,000, accounting for ~ 70% of the total 42,341 genes in the ‘Dangshansuli’ reference genome. A total of 1,173,080 InDels were identified between ‘Bartlett’ and ‘Shijiwuhe’. The number of genes with InDels in ‘Bartlett’ was 23,322 (~55%) and that in ‘Shijiwuhe’ was 21,593 (> 51%). In addition, 1,533,092 InDels were identified between ‘Yali’ and ‘Shijiwuhe’. The number of genes with InDels in ‘Yali’ was 22,065 (~52% of the whole genes), and that in ‘Shijiwuhe’ was 27,845 (> 66%).

Meanwhile, three SV software packages based on different

algorithms were chosen to carry out the analysis. The assembly algorithm has high sensitivity and specificity to detect structural variations (SVs) (Rimmer et al., 2014), and SOAPdenovo2 can obtain longer contigs with high accuracy (Luo et al., 2015). Before detecting SVs with assembly algorithms, the genomes of ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ pear were assembled using SOAPdenovo2 (Luo et al., 2015). The best k-mers of the three cultivars estimated by kmergenie (Chikhi and Medvedev, 2014) were 91, 85, and 91, respectively. The N50 of contigs assembled from ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ genome sequences were 583 bp, 603 bp, and 780 bp, respectively, and the total contig lengths were 905 Mb, 852 Mb, and 1001 Mb, respectively (Table 2). We aligned these contigs to the ‘Dangshansuli’ reference genome to produce the BAM file for SV detection using BWA and SAMtools (Li and Durbin, 2009b; Li et al., 2009).

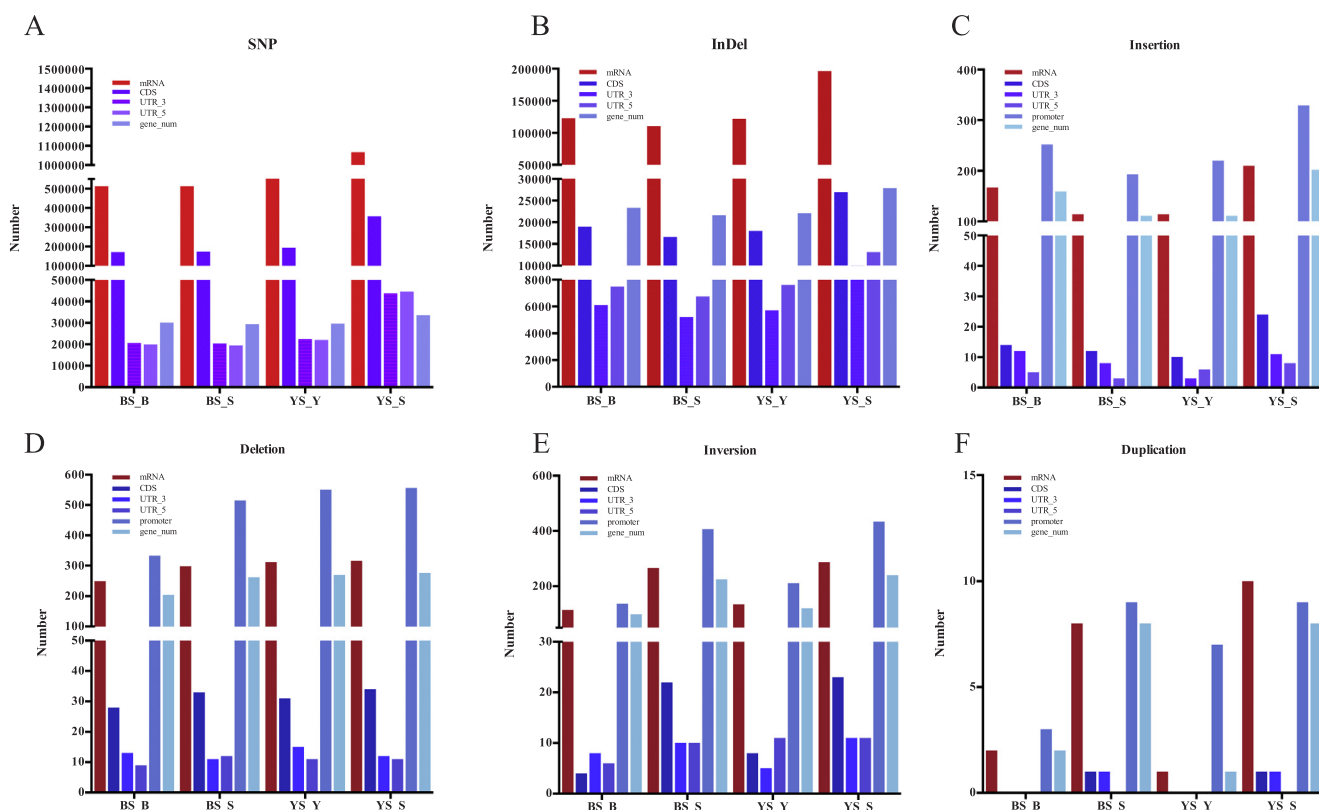
Accurately predicting the breakpoints of SVs remains a great challenge for heterozygous species, so we tried to use different algorithms for detecting SVs to potentially improve the sensitivity and decrease the false-positive rate. We used three SV programs to detect SVs, and find overlapped SVs. The numbers of overlapped SVs were much lower than the output of a single software program for the SVs, because lengths less

**Table 1**  
The number of genomic variants between ‘Shijiwuhe’ and its seeded parents (‘Bartlett’ and ‘Yali’).

|                            | BS        | YS        |
|----------------------------|-----------|-----------|
| SNPs                       | 5,570,500 | 8,241,627 |
| InDels                     | 1,173,080 | 1,533,092 |
| SVs                        | 4363      | 5320      |
| Gene with SNPs             | 33,709    | 34,900    |
| Gene with InDels           | 27,549    | 30,307    |
| Gene within SVs            | 2823      | 3296      |
| Gene with SNPs in 4 DACP   | 7915      | 6679      |
| Gene with InDels in 4 DACP | 7579      | 6914      |
| Gene within SVs in 4 DACP  | 413       | 328       |

Note: ‘BS’ represents the genomic variants between ‘Bartlett’ and ‘Shijiwuhe’. ‘YS’ represents the genomic variants between ‘Yali’ and ‘Shijiwuhe’.

| Table 2<br>Summary of genome assembly of three pears. |             |             |               |
|---|-------------|-------------|---------------|
| Item  | ‘Bartlett’  | ‘Yali’      | ‘Shijiwuhe’   |
| Size without N bases                                  | 905,273,376 | 852,436,650 | 1,000,820,457 |
| Total number of contigs                               | 2,297,305   | 2,119,628   | 2,256,593     |
| Total length of contigs (Mb)                          | 905         | 852         | 1001          |
| Mean size of contigs (bp)                             | 394         | 402         | 443           |
| Median size of contigs (bp)                           | 218         | 220         | 214           |
| Longest sequence of contigs                           | 39,881      | 46,490      | 27,131        |
| GC content  | 37.87%      | 37.66%      | 37.81%        |
| N50 length of contigs (> 150 bp)                      | 583         | 603         | 780           |
| N90 length of contigs (> 150 bp)                      | 175         | 171         | 182           |



**Fig. 2.** The number of SNPs, InDels and SVs with different regions between ‘Shijiwuhe’ and its seeded parental cultivars (‘Bartlett’ and ‘Yali’). (Note: ‘BS\_B’ represents the genomic variations in ‘Bartlett’ by comparing ‘Bartlett’ and ‘Shijiwuhe’. ‘BS\_S’ represents the genomic variations in ‘Shijiwuhe’ by comparing ‘Bartlett’ and ‘Shijiwuhe’. ‘YS\_Y’ represents the genomic variations in ‘Yali’ by comparing ‘Yali’ and ‘Shijiwuhe’. ‘YS\_S’ represents the genomic variations in ‘Shijiwuhe’ by comparing ‘Yali’ and ‘Shijiwuhe’).

than 50 bp were filtered. Through comparing between ‘Bartlett’ and ‘Shijiwuhe’, a total of 4363 SVs were identified, including 1080 insertions, 2058 deletions, 1188 inversions, and 37 duplications. Between ‘Yali’ and ‘Shijiwuhe’, 5320 SVs were identified, including 1346 insertions, 2560 deletions, 1357 inversions, and 39 duplications. The information of SVs is shown in Table S3 (Supplemental file) including the chromosome and the start and end positions.

To assess the functional impact of these genomic variations, we further investigated their relationship with annotated protein coding genes (Fig. 2). The number of SNPs and InDels were far more than that of SVs (Fig. 2A and 2B), it is difficult to narrow down the candidate genes. According to the locations of SVs in genes, the number of SVs occurring in promoter regions was more than that in messenger RNA regions and there were few SVs in CDS regions (Fig. 2 (C, D, E, F)). We speculated that SVs occurring in the promoter regions of genes could have a greater influence in differential expression of genes.

To understand the specific function of each gene within SVs, they were subjected to functional annotation using InterProScan (Zdobnov and Apweiler, 2001), Gene Ontology (GO) enrichment analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. The distribution of 1222 genes of ‘Bartlett’ in different GO categories is shown in Table S4. According to the results of KEGG analysis, three KEGG pathways had  $P$ -value  $\leq 0.05$  (Supplemental Table S5) and 12 genes were involved in ‘lipid metabolism’ pathways. The distribution of 1761 genes of ‘Shijiwuhe’ in different GO categories is shown in Table S6 and four KEGG pathways had  $P$ -value  $\leq 0.05$  (Supplemental Table S7). Comparing the SV genes of ‘Yali’ and ‘Shijiwuhe’, the distribution of 1542 genes of ‘Yali’ in different GO categories is shown in Table S8 and four KEGG pathways had  $P$ -value  $\leq 0.05$  (Supplemental Table S9). The distribution of 2079 genes of ‘Shijiwuhe’ in different GO categories is shown in Table S10 and nine KEGG pathways had  $P$ -

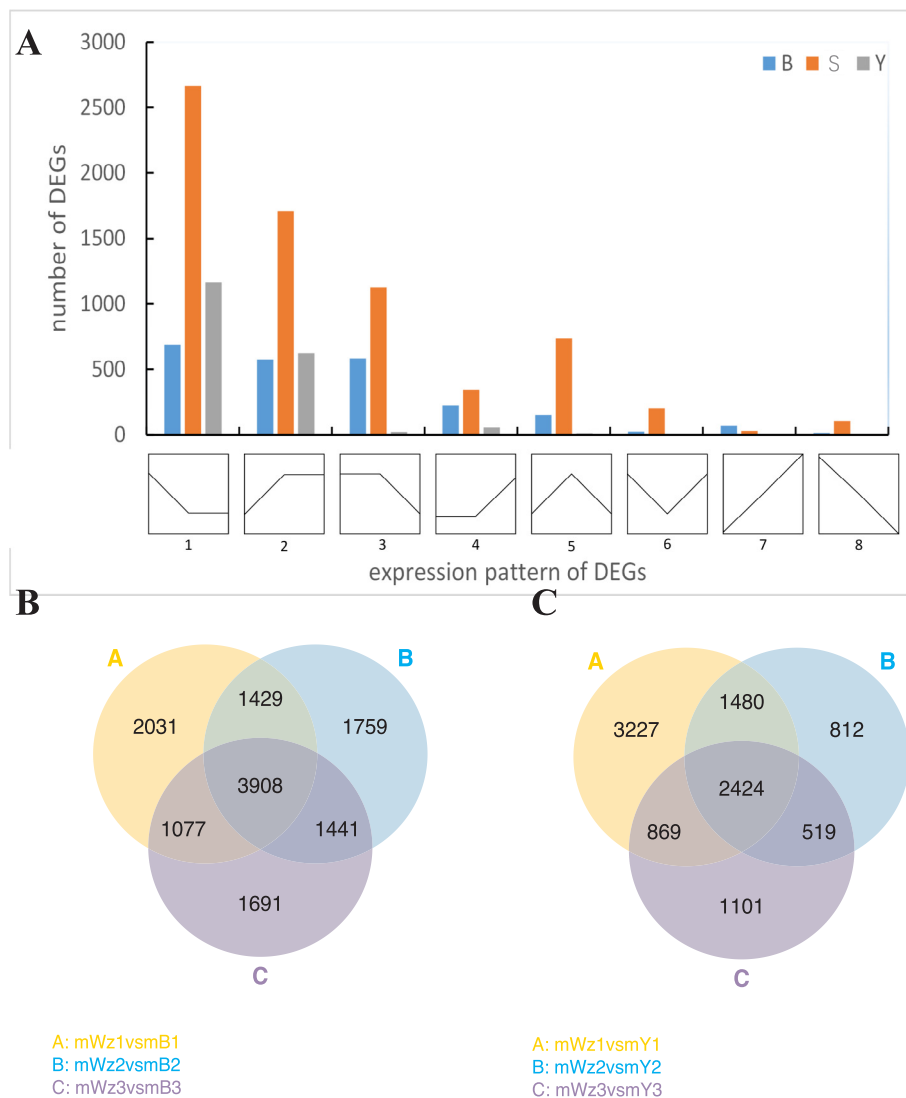
value  $\leq 0.05$  (Supplemental Table S11).

### 3.3. Identification of DEGs between seedless ‘Shijiwuhe’ and seeded parental cultivars

The transcriptomes of the selected 4 DACP, 8 DACP, and 12 DACP samples of ‘Bartlett’, ‘Yali’, and ‘Shijiwuhe’ were sequenced at three ovary developmental stages. The number of raw reads for each library ranged from 27 to 39 millions and the rates of clean reads/raw reads were all above 97.80% (Supplemental Table S12). The clean reads of all libraries were aligned to the pear genome reference sequence of ‘Dangshansuli’ (Wu et al., 2013) and ‘Bartlett’ (Chagne et al., 2014), with mapping rates from 79.81% to 93.35% (Supplemental Table S12). In total, 87,558 transcripts mapped to the genome were analyzed. Furthermore, the Pearson correlation coefficients (Sedgwick, 2012) were all above 90.0% (Supplemental Fig. S1) between biological repeats, which ensured that the sequencing and RNA quality were high and experimental repetition was reliable for further analysis of gene expression.

The analysis strategy of selecting DEGs was as depicted in Fig. S2 (Supplemental file). The upregulated and downregulated genes between every two developmental ovule stages of the same cultivar were obtained based on the relative read count data, an index to analyze the gene expression levels, and  $P$ -value  $\leq 0.05$ . Eight gene expression patterns were found among three periods in each cultivar (Fig. 3A). For the male parent ‘Yali’, gene expression patterns 1 and 2 ( $> 500$  genes for each pattern) were the main patterns. For the female parent ‘Bartlett’, patterns 1, 2, and 3 were the main patterns. Most gene expression patterns of seedless ‘Shijiwuhe’ were similar to female parent ‘Bartlett’.

The Venn diagrams revealed the number of DEGs of three ovule development periods (4 DACP, 8 DACP, and 12 DACP) between



**Fig. 3.** Profiles of genes in three seed development stages of pears. (A). Genes with different expression patterns in each cultivar during the three developmental stages (B: 'Bartlett'; W: 'Shijiwuhe'; Y: 'Yali'); (B). Venn diagram showing the overlap of DEGs between 'Bartlett' and 'Shijiwuhe' (including 4 DACP, 8 DACP, and 12 DACP); (C). Venn diagram showing the overlap of DEGs between 'Yali' and 'Shijiwuhe' (including 4 DACP, 8 DACP, 12 and DACP).

'Shijiwuhe' and 'Bartlett' and 'Shijiwuhe' and 'Yali' (Fig. 3B and 3C). Through comparisons, we found that the number of unique DEGs at 4 DACP was the highest. In 'Bartlett' and 'Shijiwuhe' pear, a total of 3,908 DEGs overlapped by comparing 4 DACP, 8 DACP, and 12 DACP. Between 'Yali' and 'Shijiwuhe' pear, 2,424 DEGs were shared at 4 DACP, 8 DACP, and 12 DACP. The overlapped genes from different comparisons were used to identify the genes with SVs.

The DEGs of 'Shijiwuhe' with its parents ('Bartlett' and 'Yali') were significantly enriched in pattern 1 and pattern 2 and were categorized into different functional groups. A total of 263 DEGs were identified. KEGG pathway enrichment revealed that these DEGs were mapped in seven KEGG sub-categories (Supplemental Table S13). Among them, most DEGs were enriched in lipid metabolism ( $P$ -value  $\leq 0.05$ ) and carbohydrate metabolism ( $P$ -value  $\leq 0.05$ , a sub-pathway of fatty acid metabolism). We also found six DEGs in the pathway of plant hormone signal transduction: AUX1, IAA, SAUR, ETR, and EBF1/2, which showed that auxin and ethylene might play important roles during abnormal/normal ovule development.

#### 3.4. Selection of candidate genes involved in seedless 'Shijiwuhe' based on SV and DEGs

The DEGs at 4 DACP are of concern. According to the phenotype observations in Fig. 1, the size of seeds was no longer increasing, and the seeds turned black by 8 DACP. Therefore, we thought the genes were differentially expressed before 8 DACP. The number of genes with non-synonymous SNPs that were differentially expressed at 4 DACP in 'Shijiwuhe' was over 7000 and those with InDels differentially expressed at 4 DACP in 'Shijiwuhe' was over 5800. GO annotations of differentially expressed genes with non-synonymous SNPs and InDels at 4 DACP in 'Shijiwuhe' pear were conducted (Supplemental Table S14–S17). In this study, we paid more attention to the DEGs that overlapped with SV at 4 DACP and they were subjected to GO and KEGG analysis.

The number of DEGs at 4 DACP in BS ('Bartlett' Vs 'Shijiwuhe') and YS ('Yali' Vs 'Shijiwuhe') were 8445 and 8000, respectively. Among the DEGs, there were 429 transcription factors potentially associated with seedlessness (78 MYB, 45 WRKY, 48 NAC and 52 bHLH transcription factors in 'Bartlett' and 'Shijiwuhe' pear, 55 MYB, 41 WRKY, 44 NAC and 452 bHLH transcription factors in 'Yali' and 'Shijiwuhe' pear).

Moreover, we identified that SNPs at transcription binding sites led



to differential expression of these transcription factors. PlantTFDB (<http://planttfdb.cbi.pku.edu.cn/>) was used to predict their transcription factor binding sites (TFBS). Each candidate transcription factor was predicted to have multiple TFBS. In ‘Bartlett’ and ‘Shijiwuhe’ pear, a total of 447 SNPs occurred in TFBS, including 198 SNPs in MYB TFBS, 77 SNPs in WRKY TFBS, 98 SNPs in NAC TFBS and 74 SNPs in bHLH TFBS (Supplemental Table S18). In ‘Yali’ and ‘Shijiwuhe’ pear, 651 SNPs occurred in TFBS, including 257 SNPs in MYB TFBS, 80 SNPs in WRKY TFBS, 215 SNPs in NAC TFBS and 99 SNPs in bHLH TFBS (Supplemental Table S19). Because there were many SNPs in the DEGs related to seedlessness and to narrow the scope of screening genomic variations in genes, we paid more attention to effectively detecting the DEGs within SVs of ‘Shijiwuhe’ pear, and predicting the potential functions of the candidate genes.

We explored the types and number of SVs enriched in the genes of seed development, seed energy storage, lipid metabolism, and genes encoding transcription factors and hormone transduction levels at 4 DACP in ‘Shijiwuhe’ pear (Table S20). Between ‘Bartlett’ and ‘Shijiwuhe’ pear, inversion was the most important SV inducing DEG-related lipid metabolism. Deletion was the second most important SV, affecting DEGs associated with auxin. Between ‘Yali’ and ‘Shijiwuhe’ pear, the most important SV type was deletion, which may affect the DEGs associated with ARF (auxin response factor). Number of inversions was the second most important, which potentially contributed the DEGs associated with lipid metabolism.

Next, the upregulated and downregulated genes with SV in ‘Shijiwuhe’ pear were detected (Supplemental Table S21). Some genes with SVs were selected for the heatmap (Fig. 4 (A, B, C, D)) from the comparison of three development periods. In Fig. 4 (A, B, C, D), we can see that some genes with SVs were expressed highly in ‘Shijiwuhe’ while the same genes with no SVs, such as *Pbr029844.1*, *Pbr030840.1*, *Pbr003092.1*, had low expression in ‘Bartlett’ and ‘Yali’ in the three development periods. Some genes affected by SVs had low expression in ‘Shijiwuhe’ while the same genes with no SVs, such as *Pbr031666.1*, *Pbr030926.1*, and *Pbr011494.1*, had high expression in ‘Bartlett’ and ‘Yali’ in the three development periods.

In ‘Bartlett’ and ‘Shijiwuhe’, 382 genes were identified as upregulated with SVs in ‘Shijiwuhe’ pear, and 348 genes with SVs were downregulated in ‘Shijiwuhe’ pear. All of these genes were subjected to functional enrichment analysis using both GO and KEGG. For upregulated genes, 12 KEGG pathways had  $P\text{-value} \leq 0.05$  (Supplemental Table S22). For downregulated genes, five KEGG pathways had  $P\text{-value} \leq 0.05$  (Supplemental Table S23). When comparing ‘Yali’ and ‘Shijiwuhe’, 431 upregulated genes were identified with SVs and 372 downregulated genes were identified with SVs in ‘Shijiwuhe’ pear. Among the results of KEGG in upregulated genes, ten KEGG pathways had  $P\text{-value} \leq 0.05$  (Supplemental Table S24). Eight KEGG pathways had  $P\text{-value} \leq 0.05$  in downregulated genes (Supplemental Table S25).

Among the mapped pathways at 4 DACP, we found 41 candidate genes that were involved in amino acid metabolism processes; 40 candidates were related to carbohydrate metabolism processes; and 28 candidate genes were associated with lipid metabolism processes. Amino acid metabolism had a major contribution to the seed energy status (Galili et al., 2014). Carbohydrates play multiple roles in cellular metabolism: they are a precursor and essential substrate for fatty acid biosynthesis and can convert into lipids, proteins, and secondary metabolites such as fiber in an elaborate regulation system during seed development, which can greatly affect seed oil content (Hua et al., 2014). Lipids are involved in hormone metabolism processes and phytohormones are essential in seed growth (Chapman et al., 2012). Lipids are interrelated with seed nutrition and energy storage.

Some genes downregulated in ‘Shijiwuhe’ pear may influence the development of seeds. For example, *Pbr006713.1* with SVs was downregulated in ‘Shijiwuhe’ compared with ‘Bartlett’. *Pbr006703.1* with SVs was also downregulated in ‘Shijiwuhe’ based on a comparison of ‘Bartlett’ and ‘Shijiwuhe’; this gene is a glycoside hydrolase.

*Pbr022627.1* with SVs was downregulated in ‘Shijiwuhe’ and upregulated in ‘Bartlett’ and ‘Yali’ and it was previously identified to be involved in the development of seed coats and pollen in *Arabidopsis thaliana* (Edstam and Edqvist, 2014). In addition, 12 SV genes were involved in the seed and embryogenesis processes. *Late embryogenesis abundant* (LEA) genes are highly expressed in seeds (Galau et al., 1986); the SVs in these genes could lead to changes of expression level in ‘Shijiwuhe’ pear.

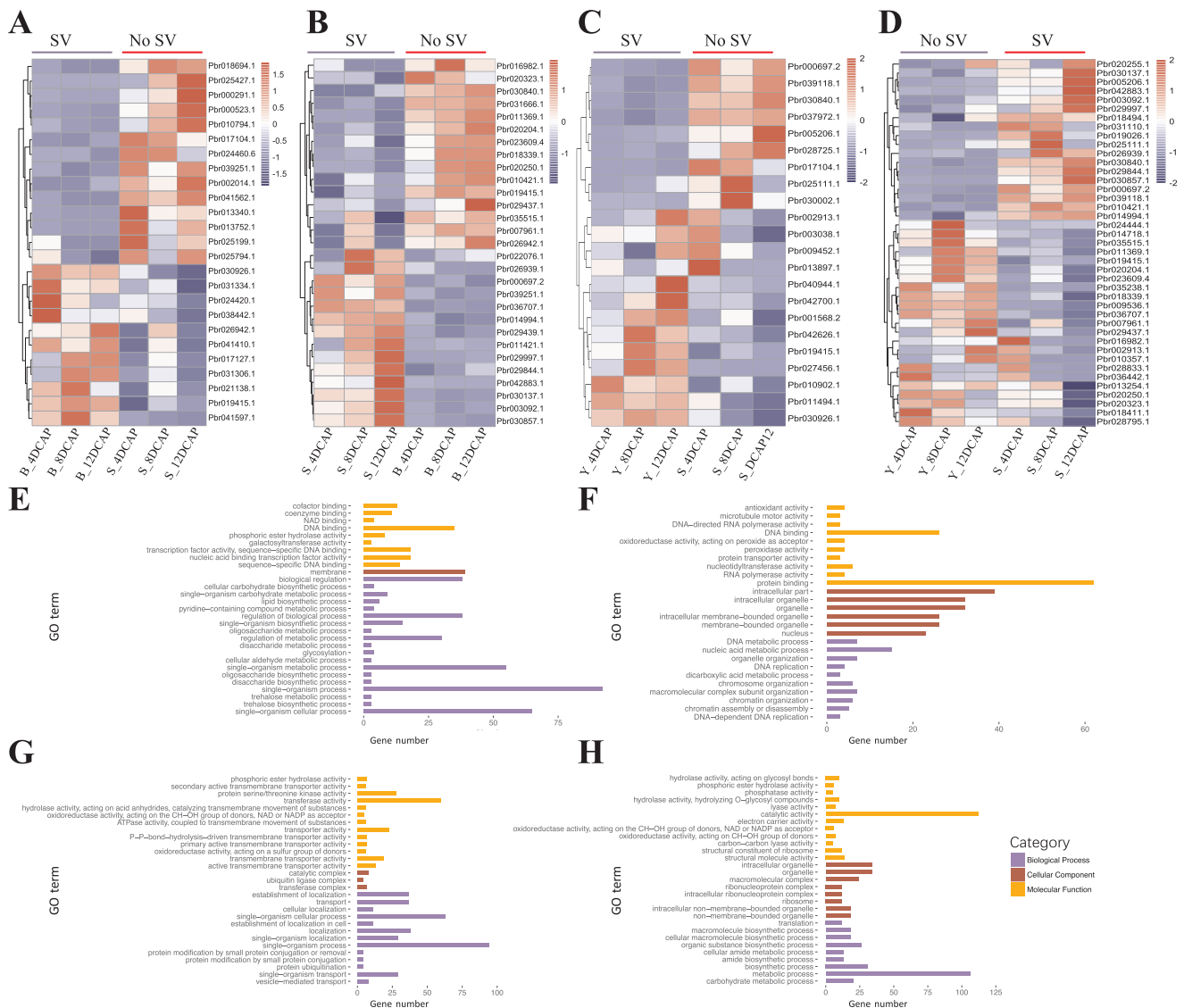
Candidate genes in KEGG pathways also included some important transcription factors (TFs) such as WRKY/MYB/NAC, which function in seed abortion. A total of 41 genes encoding crucial TFs were screened; these genes were identified as candidates associated with ovule abortion development. These genes were subdivided into four categories: MYB, WRKY, NAC, domains/NAC and other TF genes, such as bHLH and ICE1 (Supplemental Table S26).

WRKY TFs, one of the largest families of transcriptional regulators in plants, are implicated in comprehensive plant processes such as senescence, germination, and pathogen defense (Jing et al., 2009), and respond to abiotic stresses such as drought (Rushton et al., 2010) and salt (Qiu and Yu, 2009). The role of WRKY genes in seed development was explored in previous studies. In wild potato (*Solanum chacoense*), a Group-I WRKY transcription factor, ScWRKY1, was detected to be expressed transiently and strongly in fertilized ovules at the late torpedo stage (Lagace and Matton, 2004). The WRKY transcription factor DGE1 of orchard grass (*Dactylis glomerata*) was expressed during somatic embryogenesis (Alexandrova and Conger, 2002). AtWRKY10 is expressed in pollen and developing endosperms of *Arabidopsis*, as well as in the globular embryo. Plants homozygous for a knockout mutation in this gene produce small seeds. Reduced growth of the embryo and early cellularization of the endosperm were shown in the development of these small seeds, illustrating the pivotal role that WRKY genes play during this process (Luo et al., 2005). The list of WRKYs may be an important source for studying the molecular mechanism of seedlessness in pear.

Combining the DEGs from three DACP periods showed that some genes were in hormone signal transduction pathways. Candidate genes involved in plant hormone signal transduction may affect seed formation and abortion (Supplemental Table S24). Hormones including auxins, cytokinins, gibberellins, and abscisic acid are found at relatively high levels in seeds from different developmental stages and are essential in seed development (Wang and Irving, 2011). The plant signaling molecule auxin, acting through AUXIN RESPONSE FACTOR (ARF) transcription factors, is important for embryo development (Moller et al., 2017). *Pbr014718.1* (ARF11) was downregulated in ‘Shijiwuhe’ and upregulated in ‘Yali’ and it was reported to induce somatic embryogenesis in *Arabidopsis* (Wojcikowska and Gaj, 2017). Cytokinins could induce parthenocarp since they were shown to promote cell division in tomato (Matsuo et al., 2012). However, the seedlessness of ‘Shijiwuhe’ was not due to parthenocarp. Gibberellins control a range of growth and developmental processes in higher plants. MADS domain proteins have been shown to control ovule development (Alvarez-Buylla et al., 2000) and formation of the seed coat (Nesi et al., 2002). The MADS-domain transcriptional regulator AGAMOUS-LIKE15 (AGL15) has been reported to enhance somatic embryo development when constitutively expressed (Thakare et al., 2008).

### 3.5. Verification of DEGs within SV related to seedless ‘Shijiwuhe’ by qRT-PCR

A total of six candidate genes (*Pbr000291.1*, *Pbr000523.1*, *Pbr017104.1*, *Pbr014994.1*, *Pbr029844.1*, and *Pbr038442.1*) were selected for qRT-PCR analysis to verify differences in expression (Supplemental Table S1). The correlations of the FPKM of the six genes and the relative qPCR expression levels were evaluated. Pearson’s correlation coefficients showed that the qRT-PCR data and digital transcript abundance measurements were highly correlated (Fig. 5). R-

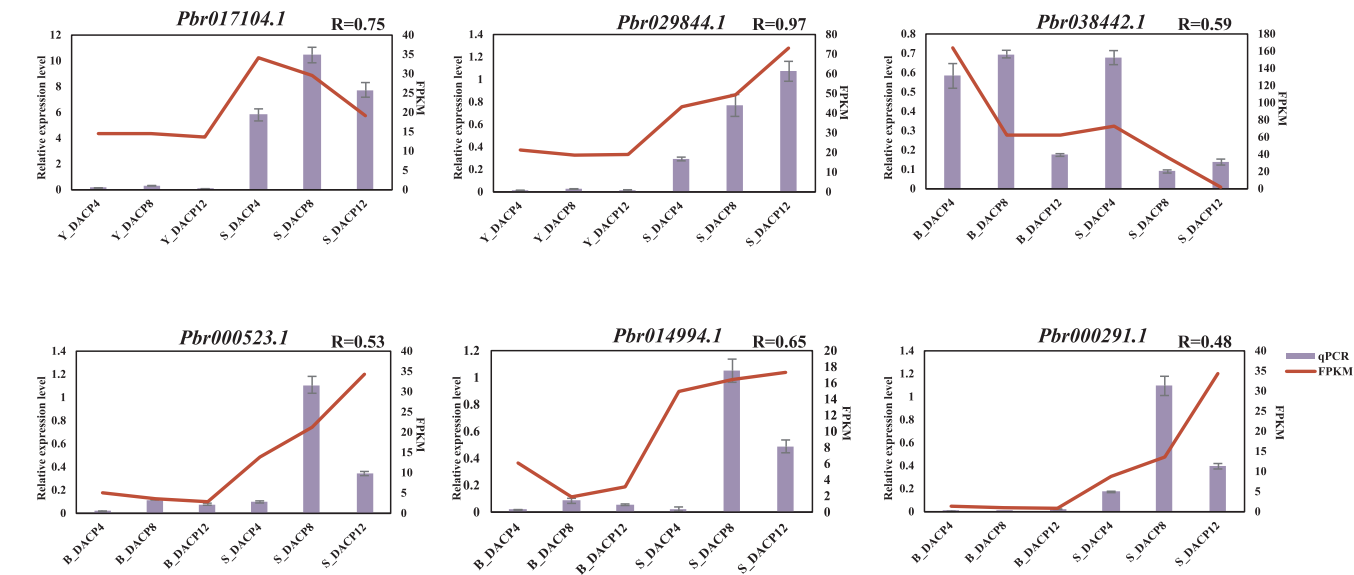


**Fig. 4.** Heatmaps of differentially expressed genes (DEGs) of each DACP and GO analysis of upregulated and downregulated genes with SVs of 4 DACP in 'Shijiwuhe' pear. (A, B). Heatmaps of DEGs with SVs in 'Bartlett' and 'Shijiwuhe', respectively, by comparing 'Bartlett' and 'Shijiwuhe' genomes. (C, D). Heatmaps of DEGs with SVs in 'Yali' and 'Shijiwuhe', respectively, by comparing 'Yali' and 'Shijiwuhe' genomes. The FPKM of each gene among the three DACP was normalized to draw the heatmap. Red indicates high relative gene expression level and blue indicates low relative gene expression level. (E, F). GO analysis of upregulated and downregulated genes with SVs of 4 DACP in 'Shijiwuhe' pear by comparing the 'Bartlett' and 'Shijiwuhe' genomes. (G, H). GO analysis of upregulated and downregulated genes with SVs of 4 DACP in 'Shijiwuhe' pear by comparing the 'Yali' and 'Shijiwuhe' genomes.

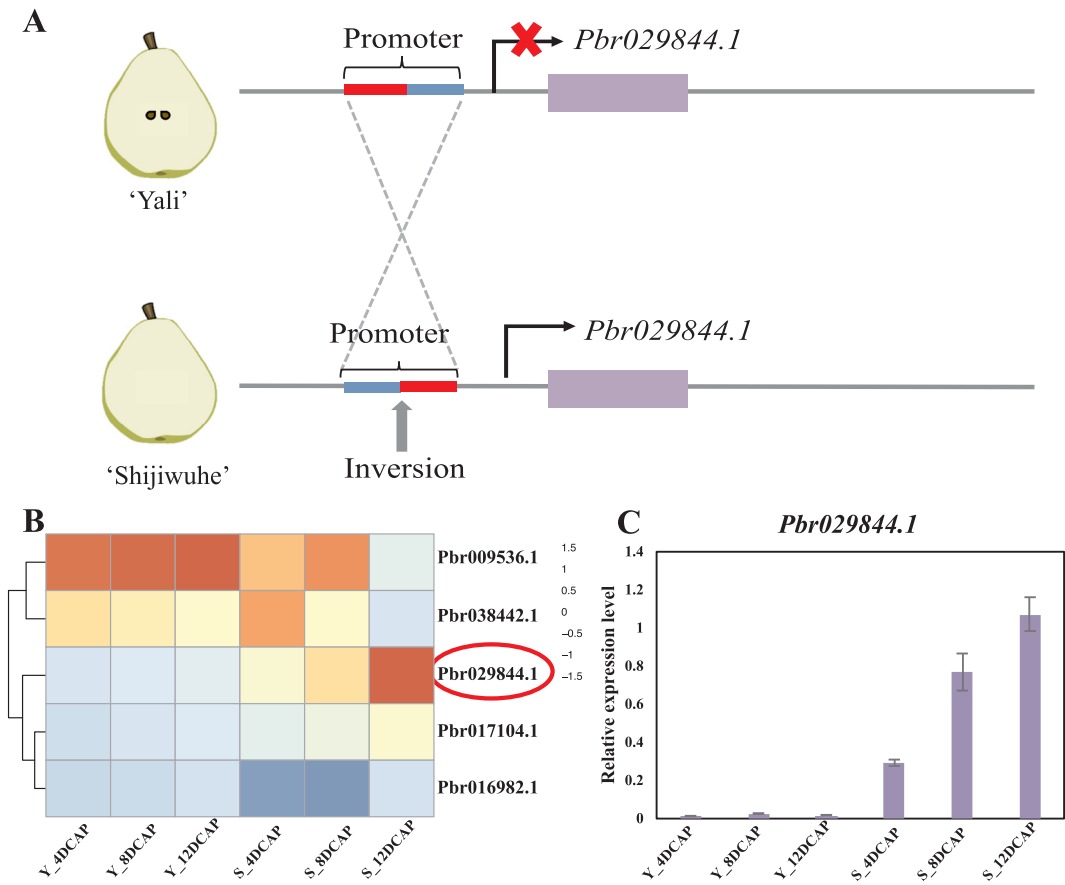
values of five of the six genes were over 0.5 and the highest R-value was 0.97. Most of them had obviously different expression from their parents. For example, *Pbr017104.1* was expressed highly in 'Shijiwuhe' during three development periods and almost had no expression in the parent 'Yali' (Fig. 5). This clearly demonstrated similar trends as those of RNA-Seq data, thus supporting the reliability of our analysis. We found an inversion in the promoter of *Pbr029844.1* in 'Shijiwuhe' by comparing 'Yali' and 'Shijiwuhe'. *Pbr029844.1* (*ML1*) had higher expression in seedless 'Shijiwuhe' than seeded 'Yali' and its expression increased from 4 DACP to 12 DACP (Fig. 6). *ML1* was reported to lead to aberrant embryo patterning along the apical-basal axis and to impair the chloroplast development resulting in embryo abortion in *Arabidopsis* (Chen et al., 2015). Based on this, we speculated that *Pbr029844.1* (*ML1*) may have no expression in seeded 'Yali' and the SV in the promoter may result in high expression of *Pbr029844.1* (*ML1*) inducing seedlessness in 'Shijiwuhe'. However, the seedless interactions and mechanisms of candidate genes still need further investigation.

#### 4. Conclusions

Seedlessness is a desirable character in pear and other Rosaceae crops. In this study, we conducted field experiments and confirmed that seedlessness in 'Shijiwuhe' pear was not due to parthenocarp and that it was self-incompatible. By combining the genes from SV detection using DNA sequencing data and transcriptome profiles at three time points (4 DACP, 8 DACP, 12 DACP) in early fruit development of seeded 'Bartlett' and 'Yali' with seedless 'Shijiwuhe' pear, we identified a list of SV affecting differentially expressed candidate genes associated with seedlessness between seeded and seedless pear cultivars. Six genes were selected randomly from the list of candidates to verify that the RNA-seq expression and Pearson's correlation coefficients were highly correlated. The candidates were involved in seed development, seed energy storage, and lipid metabolism, with some predicted to encode essential transcription factors and hormone transduction signals. The detected candidate genes might be the main target for characterization and functional studies, which could facilitate the development of new



**Fig. 5.** Expression levels of six candidate genes. The relative expression levels of six candidate genes selected randomly from the list of candidates are shown. X axes indicate two cultivars. The left Y axes indicate relative expression level of each selected gene. The right Y axes indicate the values of FPKM. qRT-PCR data are technical replicates with error bars, representing means  $\pm$  SE (n = 3).



**Fig. 6.** An example of SVs affecting differentially expressed genes in 'Shijiwuhe' pear. (A) A deletion in the promoter region of a candidate gene (*Pbr029844.1*) associated with seedlessness. (B) The heatmap of some candidate genes (including *Pbr029844.1*) by comparing 'Yali' and 'Shijiwuhe' pear. (C) The RT-qPCR expression of one candidate gene (*Pbr029844.1*).

seedless pear cultivars. This study lays a foundation for selecting genes associated with seedlessness and verifying the function of candidate genes.

**Author contribution**

J. Wu designed and supervised the whole experiments. J.Y. Sun, J.T. Xu, Q.Y. Li, W.Q. Le, B.F. Hao and Y.J. Li prepared material and G.Y. Yang performed RT-qPCR. Y.Y. Liu and R.Z. Wang performed

bioinformatics analysis. S.L. Zhang provided experimental support. Y.Y. Liu and J. Y. Sun drafted the manuscript, J. Wu and M.Y. Zhang revised the manuscript. All the authors have read and approved the final manuscript.

## Funding

This work was supported by the Earmarked fund for Jiangsu Agricultural Industry Technology System (JATS[2018]277); National Natural Science Foundation of China (31672111); the Earmarked Fund for China Agriculture Research System (CARS-28).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank the Changli Institute of Pomology, Hebei Academy of Agricultural and Forestry Sciences, China for providing the 'Yali' and 'Shijiwuhe' pears as experimental materials and the Yantai Academy of Agricultural Sciences, China for providing the 'Bartlett' pear as experimental materials.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gene.2020.144480>.

## References

- Alexandrova, K.S., Conger, B.V., 2002. Isolation of two somatic embryogenesis-related genes from orchardgrass (*Dactylis glomerata*). *Plant Sci.* 162, 301–307.
- Alkan, C., Coe, B.P., Eichler, E.E., 2011. Applications of next-generation sequencing genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12, 363–375.
- Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F., Yanofsky, M.F., 2000. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* 24, 457–466.
- Amir, R., Galili, G., Cohen, H., 2018. The metabolic roles of free amino acids during seed development. *Plant Sci.* 275, 11–18.
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11.
- Bawa, K.S., Hedge, S.G., Ganeshaiah, K.N., Shaanker, R.U., 1989. Embryo and seed abortion in plants. *Nature* 342, 625.
- Blasco, M., Badenes, M.L., Naval, M.D., 2015. Colchicine-induced polyploidy in loquat (*Eriobotrya japonica* (Thunb.) Lindl.). *Plant Cell, Tissue Organ Cult.* 120, 453–461.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Budar, F., Pelletier, G., 2001. Male sterility in plants: occurrence, determinism, significance and use. *Comptes Rend. Acad. Sci. Serie Iii-Sci. Vie-Life Sci.* 324, 543–550.
- Chagne, D., Crowhurst, R.N., Pindo, M., Thrimawithana, A., Deng, C., Ireland, H., Fiers, M., Dzierzon, H., Cestaro, A., Fontana, P., Bianco, L., Lu, A., Storey, R., Knabel, M., Saeed, M., Montanari, S., Kim, Y.K., Nicolini, D., Larger, S., Stefani, E., Allan, A.C., Bowen, J., Harvey, I., Johnston, J., Malnoy, M., Troggio, M., Perceped, L., Sawyer, G., Wiedow, C., Won, K., Viola, R., Hellens, R.P., Brewer, L., Bus, V.G.M., Schaffer, R.J., Gardiner, S.E., Velasco, R., 2014. The draft genome sequence of European pear (*Pyrus communis* L 'Bartlett'). *Plos One* 9.
- Chapman, K.D., Dyer, J.M., Mullen, R.T., 2012. Biogenesis and functions of lipid droplets in plants: thematic review series: lipid droplet synthesis and metabolism: from yeast to man. *J. Lipid Res.* 53, 215–226.
- Chen, G.Q., Riiff, T.J., Johnson, K., Morales, E., Kim, H.U., Lee, K.R., Lin, J.T., 2017. Seed development and hydroxy fatty acid biosynthesis in *Physaria lindheimeri*. *Ind. Crops Prod.* 108, 410–415.
- Chen, H.Y., Zou, W.X., Zhao, J., 2015. Ribonuclease J is required for chloroplast and embryo development in *Arabidopsis*. *J. Exp. Bot.* 66, 2079–2091.
- Chikhi, R., Medvedev, P., 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30, 31–37.
- Chiwocha, S.D.S., Cutler, A.J., Abrams, S.R., Ambrose, S.J., Yang, J., Ross, A.R.S., Kermode, A.R., 2005. The *etr1-2* mutation in *Arabidopsis thaliana* affects the abscisic acid, auxin, cytokinin and gibberellin metabolic pathways during maintenance of seed dormancy, moist-chilling and germination. *Plant J.* 42, 35–48.
- Chu, Y.C., Lin, T.S., Chang, J.C., 2015. Pollen effects on fruit set, seed weight, and shriveling of '73-S-20' litchi: with special reference to artificial induction of parthenocarpy. *HortScience* 50, 369–373.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.Y., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly* 6, 80–92.
- Edstam, M.M., Edqvist, J., 2014. Involvement of GPI-anchored lipid transfer proteins in the development of seed coats and pollen in *Arabidopsis thaliana*. *Physiol. Plant.* 152, 32–42.
- Feng, C., Chen, M., Xu, C.J., Bai, L., Yin, X.R., Li, X., Allan, A.C., Ferguson, I.B., Chen, K.S., 2012. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *Bmc. Genomics* 13.
- Galau, G.A., Hughes, D.W., Dure, L., 1986. Absciscic-acid induction of cloned cotton late embryogenesis-abundant (Lea) messenger-Rnas. *Plant Mol. Biol.* 7, 155–170.
- Galili, G., Avin-Wittenberg, T., Angelovici, R., Fernie, A.R., 2014. The role of photosynthesis and amino acid metabolism in the energy status during seed development. *Front. Plant Sci.* 5.
- Garber, M., Grabherr, M.G., Guttman, M., Trapnell, C., 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–477.
- Hazra, P., Dutta, A.K., 2011. Parthenocarpy in tomato (*Solanum lycopersicum*), its inheritance pattern and association with marker characters. *Indian J. Genetics Plant Breed.* 71, 248–253.
- Horiguchi, G., Gonzalez, N., Beemster, G.T.S., Inze, D., Tsukaya, H., 2009. Impact of segmental chromosomal duplications on leaf size in the grandifolia-D mutants of *Arabidopsis thaliana*. *Plant J.* 60, 122–133.
- Hua, S.J., Chen, Z.H., Zhang, Y.F., Yu, H.S., Lin, B.G., Zhang, D.Q., 2014. Chlorophyll and carbohydrate metabolism in developing silique and seed are prerequisite to seed oil content of *Brassica napus* L. *Botanical Studies* 55.
- Jing, S.J., Zhou, X., Song, Y., Yu, D.Q., 2009. Heterologous expression of OsWRKY23 gene enhances pathogen defense and dark-induced leaf senescence in *Arabidopsis*. *Plant Growth Regul.* 58, 181–190.
- Kim, T., Seo, H.D., Hennighausen, L., Lee, D., Kang, K., 2018. Octopus-toolkit: a workflow to automate mining of public epigenomic and transcriptomic next-generation sequencing data. *Nucleic Acids Res.*
- Koura, S., Hasegawa, K., Yamamoto, Y., Yonemoto, Y., 2004. Fruit set and fruit growth of seedless cherimoya (*Annona cherimola* Mill.) induced by GA(3) under greenhouse cultivation in Japan. In: Proceedings of the Ninth International Symposium on Plant Bioregulators in Fruit Production, pp. 63–66.
- Lagace, M., Matton, D.P., 2004. Characterization of a WRKY transcription factor expressed in late torpedo-stage embryos of *Solanum chacoense*. *Planta* 219, 185–189.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Proc, G.P.D., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liu, M.J., Zhao, J., Cai, Q.L., Liu, G.C., Wang, J.R., Zhao, Z.H., Liu, P., Dai, L., Yan, G.J., Wang, W.J., Li, X.S., Chen, Y., Sun, Y.D., Liu, Z.G., Lin, M.J., Xiao, J., Chen, Y.Y., Li, X.F., Wu, B., Ma, Y., Jian, J.B., Zhang, W., Yuan, Z., Sun, X.C., Wei, Y.L., Yu, L.L., Zhang, C., Liao, S.G., He, R.J., Guang, X.M., Wang, Z., Zhang, Y.Y., Luo, L.H., 2014. The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.* 5.
- Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 25, 402–408.
- Lo'ay, A.A., El-Boray, M.S., 2018. Improving fruit cluster quality attributes of 'Flame Seedless' grapes using preharvest application of ascorbic and salicylic acid. *Sci. Hortic.* 233, 339–348.
- Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., Chaudhury, A., 2005. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *PNAS* 102, 17531–17536.
- Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., Yuan, J.Y., He, G.Z., Chen, Y.X., Pan, Q., Liu, Y.J., Tang, J.B., Wu, G.X., Zhang, H., Shi, Y.J., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C.L., Cheung, D.W., Yiu, S.M., Peng, S.L., Zhu, X.Q., Liu, G.M., Liao, X.K., Li, Y.R., Yang, H.M., Wang, J., Lam, T.W., Wang, J., 2015. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 4 (vol. 1, 18, 2012).
- Ma, Y.W., Li, Q.L., Hu, G.B., Qin, Y.H., 2017. Comparative transcriptional survey between self-incompatibility and self-compatibility in *Citrus reticulata* Blanco. *Gene* 609, 52–61.
- Martinelli, F., Uratsu, S.L., Reagan, R.L., Chen, Y., Tricoli, D., Fiehn, O., Rocke, D.M., Gasser, C.S., Dandekar, A.M., 2009. Gene regulation in parthenocarpic tomato fruit. *J. Exp. Bot.* 60, 3873–3890.
- Matsuo, S., Kikuchi, K., Fukuda, M., Honda, I., Imanishi, S., 2012. Roles and regulation of cytokinins in tomato fruit development. *J. Exp. Bot.* 63, 5569–5579.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Medvedev, P., Stanciu, M., Brudno, M., 2009. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods* 6, S13–S20.
- Moller, B.K., ten Hove, C.A., Xiang, D.Q., Williams, N., Lopez, L.G., Yoshida, S., Smit, M., Datla, R., Weijers, D., 2017. Auxin response cell-autonomously controls ground tissue initiation in the early *Arabidopsis* embryo. *PNAS* 114, E2533–E2539.
- Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., Lepiniec, L., 2002. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS



- domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* 14, 2463–2479.
- Nielsen, R., Paul, J.S., Albrechtsen, A., Song, Y.S., 2011. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451.
- Pannell, J.R., Ojeda, F., 2000. Patterns of flowering and sex-ratio variation in the Mediterranean shrub *Phillyrea angustifolia* (Oleaceae): implications for the maintenance of males with hermaphrodites. *Ecol. Lett.* 3, 495–502.
- Qiu, Y.P., Yu, D.Q., 2009. Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in *Arabidopsis*. *Environ. Exp. Bot.* 65, 35–47.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., Korbel, J.O., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, I333–I339.
- Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., Wilkie, A.O.M., McVean, G., Lunter, G., Consortium, W., 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* 46, 912–918.
- Robert, H.S., 2019. Molecular communication for coordinated seed and fruit development: what can we learn from auxin and sugars? *Int. J. Mol. Sci.* 20.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rushton, P.J., Somssich, I.E., Ringler, P., Shen, Q.X.J., 2010. WRKY transcription factors. *Trends Plant Sci.* 15, 247–258.
- Sedgwick, P., 2012. Pearson's correlation coefficient. *British Med. J.* 344.
- Singh, Z., 2005. Embryo abortion in relation to fruit size, quality, and concentrations of nutrients in skin and pulp of mango. *J. Plant Nutr.* 28, 1723–1737.
- Sogo, A., Tobe, H., 2008. Mode of pollen tube growth in pistils of *Ticodendron incognitum* (Ticodendraceae, Fagales) and the evolution of chalazogamy. *Bot. J. Linn. Soc.* 157, 621–631.
- Thakare, D., Tang, W., Hill, K., Perry, S.E., 2008. The MADS-domain transcriptional regulator AGAMOUS-LIKE15 promotes somatic embryo development in *Arabidopsis* and soybean. *Plant Physiol.* 146, 1663–1672.
- Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Wang, H., Jones, B., Li, Z.G., Frasse, P., Delalande, C., Regad, F., Chaabouni, S., Latche, A., Pech, J.C., Bouzayen, M., 2005. The tomato Aux/IAA transcription factor IAA9 is involved in fruit development and leaf morphogenesis. *Plant Cell* 17, 2676–2692.
- Wang, Y.H., Irving, H.R., 2011. Developing a model of plant hormone interactions. *Plant Signal Behav.* 6, 494–500.
- Wilcock, C., Neiland, R., 2002. Pollination failure in plants: why it happens and when it matters. *Trends Plant Sci.* 7, 270–277.
- Wojcikowska, B., Gaj, M.D., 2017. Expression profiling of AUXIN RESPONSE FACTOR genes during somatic embryogenesis induction in *Arabidopsis*. *Plant Cell Rep.* 36, 843–858.
- Wu, J., Wang, Z.W., Shi, Z.B., Zhang, S., Ming, R., Zhu, S.L., Khan, M.A., Tao, S.T., Korban, S.S., Wang, H., Chen, N.J., Nishio, T., Xu, X., Cong, L., Qi, K.J., Huang, X.S., Wang, Y.T., Zhao, X., Wu, J.Y., Deng, C., Gou, C.Y., Zhou, W.L., Yin, H., Qin, G.H., Sha, Y.H., Tao, Y., Chen, H., Yang, Y.A., Song, Y., Zhan, D.L., Wang, J., Li, L.T., Dai, M.S., Gu, C., Wang, Y.Z., Shi, D.H., Wang, X.W., Zhang, H.P., Zeng, L., Zheng, D.M., Wang, C.L., Chen, M.S., Wang, G.B., Xie, L., Sovero, V., Sha, S.F., Huang, W.J., Zhang, S.J., Zhang, M.Y., Sun, J.M., Xu, L.L., Li, Y., Liu, X., Li, Q.S., Shen, J.H., Wang, J.Y., Paull, R.E., Bennetzen, J.L., Wang, J., Zhang, S.L., 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 23, 396–408.
- Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., van der Knaap, E., 2008. A retro-transposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319, 1527–1530.
- Xie, C., Mao, X.Z., Huang, J.J., Ding, Y., Wu, J.M., Dong, S., Kong, L., Gao, G., Li, C.Y., Wei, L.P., 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322.
- Ye, K., Schulz, M.H., Long, Q., Apweiler, R., Ning, Z.M., 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865–2871.
- Zdobnov, E.M., Apweiler, R., 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848.
- Zhang, B., Xi, W.P., Wei, W.W., Shen, J.Y., Ferguson, I., Chen, K.S., 2011. Changes in aroma-related volatiles and gene expression during low temperature storage and subsequent shelf-life of peach fruit. *Postharvest Biol. Technol.* 60, 7–16.
- Zhang, S.J., Shi, Q.C., Albrecht, U., Shatters, R.G., Stange, R., McCollum, G., Zhang, S., Fan, C.M., Stover, E., 2017. Comparative transcriptome analysis during early fruit development between three seedy citrus genotypes and their seedless mutants. *Hortic. Res.* 4.
- Zhang, Z.H., Mao, L.Y., Chen, H.M., Bu, F.J., Li, G.C., Sun, J.J., Li, S., Sun, H.H., Jiao, C., Blakely, R., Pan, J.S., Cai, R., Luo, R.B., Van de Peer, Y., Jacobsen, E., Fei, Z.J., Huang, S.W., 2015. Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell* 27, 1595–1604.

RESEARCH

Open Access



# The pear genomics database (PGDB): a comprehensive multi-omics research platform for *Pyrus* spp.

Shulin Chen<sup>1</sup>, Manyi Sun<sup>2</sup>, Shaozhuo Xu<sup>1</sup>, Cheng Xue<sup>1</sup>, Shuwei Wei<sup>3</sup>, Pengfei Zheng<sup>1</sup>, Kaidi Gu<sup>1</sup>, Zhiwen Qiao<sup>1</sup>, Zhiying Liu<sup>1</sup>, **Mingyue Zhang<sup>1\*</sup>** and Jun Wu<sup>2\*</sup>

## Abstract

**Background** Pears are among the most important temperate fruit trees in the world, with significant research efforts increasing over the last years. However, available omics data for pear cannot be easily and quickly retrieved to enable further studies using these biological data.

**Description** Here, we present a publicly accessible multi-omics *pear* resource platform, the Pear Genomics Database (PGDB). We collected and collated data on genomic sequences, genome structure, functional annotation, transcription factor predictions, comparative genomics, and transcriptomics. We provide user-friendly functional modules to facilitate querying, browsing and usage of these data. The platform also includes basic and useful tools, including JBrowse, BLAST, phylogenetic tree building, and additional resources providing the possibility for bulk data download and quick usage guide services.

**Conclusions** The Pear Genomics Database (PGDB, <http://pyrusgdb.sdau.edu.cn>) is an online data analysis and query resource that integrates comprehensive multi-omics data for pear. This database is equipped with user-friendly interactive functional modules and data visualization tools, and constitutes a convenient platform for integrated research on pear.

**Keywords** Pear, Database, Genome, Transcriptome, Re-sequencing

## Background

Pear is a member of the *Rosaceae* family and the *Amygdaloideae* subfamily [1], and one of the most important temperate fruit trees globally, with a cultivation history spanning more than 3,000 years [2, 3]. At present, 22 species and 5,000 accessions of pear have been described, including 5 major domesticate species cultivated for the production of fruit, specifically *P. communis*, *P. pyrifolia*, *P. bretschneideri*, *P. ussuriensis*, and *P. sinkiangensis* [4].

Most cultivated pears are diploid ( $2n=34$ ), and the genome is highly heterozygous and contains several repetitive sequences. The genome of an important oriental pear variety ‘Dangshansuli’ (*P. bretschneideri*) was

\*Correspondence:

Mingyue Zhang  
[myzhang2020@163.com](mailto:myzhang2020@163.com)

Jun Wu

[wujun@njau.edu.cn](mailto:wujun@njau.edu.cn)

<sup>1</sup>College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an 271018, Shandong, China

<sup>2</sup>College of Horticulture, National Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

<sup>3</sup>Shandong Institute of Pomology, Tai'an 271000, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

sequenced and assembled using the HiSeq Illumina technology combined with a BAC-by-BAC strategy [5]. After this, the western variety ‘Bartlett’ (*P. communis*) was sequenced with Roche’s 454 Sequencing Technology [6]. In recent years, several more pear reference genomes were published owing the rapid development of sequencing technologies [7–13]. These developments further led to the generation of a large number of transcriptome and population DNA re-sequencing data, allowing mining key genes responsible for important agronomic traits and studying the domestication history of pears [14, 15]. At present, pear genome and resequencing data have been collected in the Rosaceae Genome database GDR, but transcriptome data is lacking. Therefore, there is an urgent need for a database that can effectively integrate, analyze and disseminate pear multiomics data, and provide a platform for researchers to quickly access and utilize these resources. These resources are already available for a variety of plants, such as bayberry and pineapple [16, 17]. Therefore, we integrated the advantages of the above-mentioned databases and constructed the Pear Genomics Database (PGDB). In this study, a total of nine genome sequences, 35 transcription group datasets, and re-sequencing data from 30 pear accessions were collected. We also included commonly used tools, such as BLAST, JBrowse, phylogenetic tree building in the PGDB which will facilitate the future development of pear functional genomics and molecular biology approaches.

Database construction and content

The PGDB collected and processed data on genome sequences, annotation, expression, synteny, and resequencing, which are stored in the MySQL database server (5.7.34). The web interface mainly uses the front-end framework Twitter Bootstrap based on HTML5 (Hyper-Text Markup Language 5), CSS (Cascading Style Sheets) and JavaScript, and allows users to connect various levels of information, query the data and generate results. The data can be downloaded through a PHP protocol (7.4.21). The entire website was developed using the Web server software Apache (2.4.48), and implemented in the Linux (CentOS 7.6) operating system (Fig. 1).

Genome assemblies and functional annotations

The PGDB collected information on 9 pear genomes, including ‘Dangshansuli’ v.1.0 (*P. bretschneideri*), ‘Dangshansuli’ v.1.1 (*P. bretschneideri*), ‘Cuiguan’ (*P. pyrifolia*), ‘Zhongai 1’ [(*P. ussuriensis* × *communis*) × spp.], ‘Shanxi Duli’ (*P. betulifolia*), ‘Nijisseiki’ (*P. pyrifolia*), ‘Bartlett’ v.1.0 (*P. communis*), ‘Bartlett’ v.2.0 (*P. communis*) and ‘d’Anjou’ (*P. communis*) (Additional file 1: Table S1). We kept the ID of each gene in the database consistent with the gene ID available in the original GFF annotation file. Gene Ontology (GO) [18] and InterPro [19] annotations were performed using InterProScan (v5.53-87.0) [20]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were annotated with the bi-directional best hit

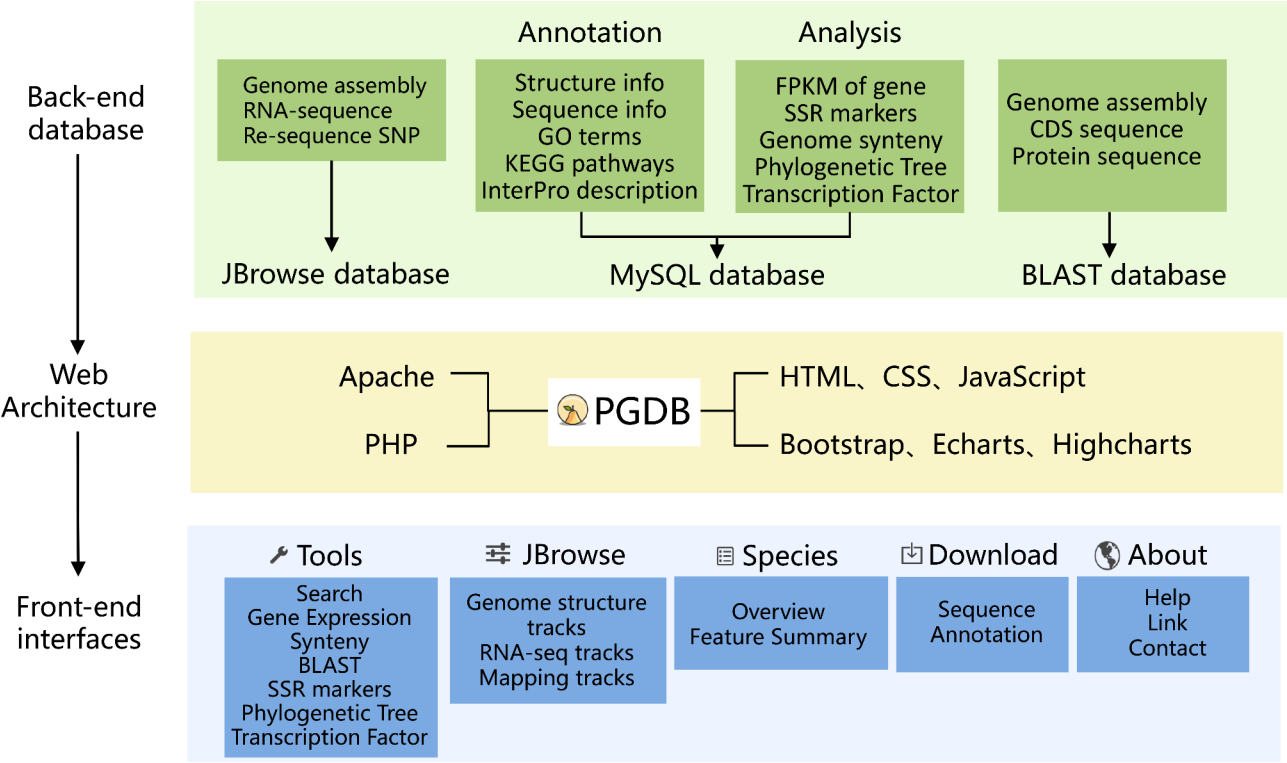


Fig. 1 Overview of the PGDB website architecture

(BBH) method from KAAS [21] using 40 plant species as reference (Additional file 2: Table S2).

### Transcription factors

Transcription factors (TFs) and transcriptional regulators (TRs) modulate the expression of target genes, which in turn are involved in several important life processes, such as growth and development, secondary metabolism and abiotic stress responses [22, 23]. We used the iTAK program for TF and TR predictions, a software based on a set of consensus domain assignment rules [24]. We detected TFs and TRs from the aforementioned 9 pear genomes, i.e., ‘Dangshansuli’ v1.0 (2433, 563), ‘Cuiguan’ (2705, 616), ‘Shanxi Duli’ (2670, 628), ‘Zhongai 1’ (2409, 570), ‘Bartlett’ v1.0 (2541, 624), ‘Bartlett’ v2.0 (1910, 481), ‘Nijisseiki’ (2544, 625), ‘d’Anjou’ (2666,703), ‘Dangshan-suli’ v1.1 (3517,475) (Additional file 3: Table S3), and found that the most numerous were the MYB and NAC families. These families include widely known key factors regulating development and stress responses [25, 26].

### Synteny data

We identified synteny blocks and homologous gene pairs from 9 pear genome data. The protein sequences were aligned against each other and themselves using BLASTP (E-value  $\leq 1e-10$ ). The MCScanX [27] software was then employed with default parameters to determine the synteny blocks and homologous gene pairs from the BLASTP results.

### Marker data

The Krait tool [28] was used to mine simple sequence repeat (SSR) resources in nine pear genome data. A total of 386,779 SSR markers were identified and divided into five categories, namely dinucleotides to hexanucleotides, with the minimum number of repeats of 6,5,4,4,4 for each SSR type. Primer3 software (58) [29] implemented in Krait tool was used to design SSR primers. The specific parameters are: the size range of polymerase chain reaction (PCR) product is 100–300 bp, the length of primer is 20–25 bases, the best is 22 bases, the best annealing temperature is 50–60 °C, the GC content is 40–60%, the best is 50%. Retain the default values for other parameters. In addition, 579 pairs of SSR markers were collected from reported literatures [30–35].

### Transcriptomic data

The PGDB included transcriptomic data from seven key stages of fruit development on the following 5 cultivars: ‘Hosui’ (*P. pyrifolia*), ‘Yali’ (*P. bretschneideri*), ‘Kuerlexiangli’ (*P. sinkiangensis*), ‘Nanguoli’ (*P. ussuriensis*) and ‘Starkrimson’ (*P. communis*) [36]. The RNA-seq reads were mapped to the reference genome using software of SOAPaligner [37]. Transcription abundance was

quantified by in-house perl scripts using the method of mapped sequence reads per million kilobytes per exon (RPKM). In addition, the results were changed to bigwig format using deepTools [38] software and placed in JBrowse.

### Utility

#### Database content

The homepage of the PGDB database is mainly composed of three parts. The top navigation bar is a fast link entry of each module, including: ‘Tools’, ‘JBrowse’, ‘Species’, ‘Download’, and ‘About’. The middle part contains a brief introduction to the database and the fast link to the ‘Tools’ and ‘Species’ modules. The bottom portion includes the website’s launch date and other information.

### Available tools

#### Search

The ‘Search’ page provides two retrieval modules (Fig. 2a). In the ‘Quick searching’ module, users can first search for detailed annotations on genes by simply selecting the cultivar genome and inputting the gene ID. The results page includes information on the sequences (including gene, CDS, and protein), functional annotations (GO, KEGG and InterPro) and the existence of homologous genes (Fig. 2b). Users can select which information should be displayed by clicking on different drop-down box options. In addition, users can also employ Bedtools [39] to retrieve genomic sequences by entering the reference genome coordinates. The results can be visualized online or downloaded for local storage. The ‘Sequence fetch’ module provides a batch search function for gene, CDS, and protein sequences.

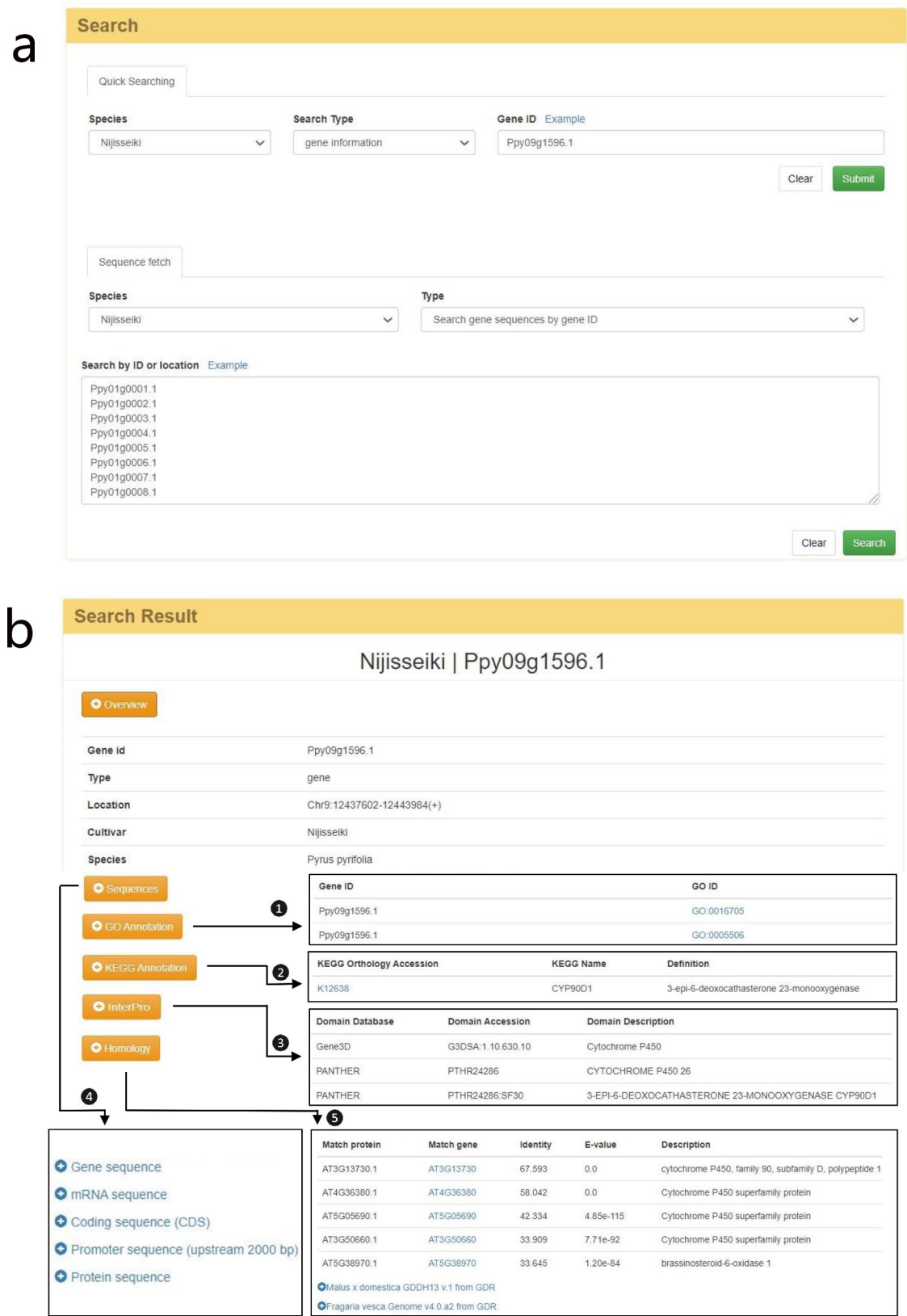
### Gene expression

The ‘Gene Expression’ page provides a search function for genes with annotated RPKM values. Users can find this function in the navigation bar or the ‘Tools’ module in the middle section of the home page. The results are presented as line or bar charts drawn by Echarts [40] to display RPKM values at different development stages in pear fruits. The query results support online browsing and downloading to facilitate researchers conducting in-depth analyses.

### Synteny

In this page, comparative genomic information between different pear varieties is provided to facilitate quick retrieval of genomic collinearity and homologous gene pairs (Fig. 3a). In the ‘Synteny Block’ module, users can obtain synteny blocks by selecting the pear genome and chromosome of choice. The top half of the results page contains an image showing the quantitative relationship between synteny blocks of the query and compared







**Fig. 3** Synteny page of the PGDB. **(a)** Querying synteny blocks between genomes and drawing synteny images. **(b)** The synteny blocks in the query and compared chromosomes. **(c)** The genes contained in each synteny block. **(d)** The collinear image drawn online

genomes. This is implemented by HighCharts. The bottom half of the page provides complete synteny block information (block ID, location, source, e-value) in the form of a list (Fig. 3b). By clicking on different synteny blocks, users will be linked to detailed information on homologous gene pairs within synteny blocks (Fig. 3c). In the 'Synteny Image' module, synteny images can be constructed between the chromosomes of any two genomes, and downloaded for further study (Fig. 3d).

### BLAST

This page provides a user-friendly BLAST tool for sequence alignment with ViroBlast [41]. Nucleotide and amino acid sequence similarity searches can be performed through a user-friendly input-output interface. We provide three types of query databases for genomic sequences, CDS sequences and protein sequences (Fig. 4a, b). Users can search the nucleotide sequence and the protein sequence databases by query sequences in BLASTN or BLASTX, and TBLASTN and BLASTP, respectively. In addition, users can choose TBLASTX to translate nucleotide sequences into protein sequences before comparison.

### SSR markers

PGDB provides a query page for two types of SSR markers based on genomic prediction and literature reports. Users search for molecular marker data by filling in SSR IDs or selecting special items. Users can submit the search criteria to obtain detailed information including variety, SSR ID, scaffold, motif, type, repeat, start, end, and length. In addition, for genomic SSR markers, detailed information related to primers can be obtained by clicking SSR ID, such as forward sequence, reverse sequence, Tm (temperature), GC content and product size, etc.

### Phylogenetic tree building

This page provides a simple and quick tool for constructing phylogenetic trees. Users can input FASTA formatted sequences, with alignment performed with MAFFT (V7.158) [42]. IQ-Tree, a stochastic algorithm to infer phylogenetic trees by maximum likelihood, is then used to assemble these sequences [43, 44]. Both the aligned sequence file and the NWK file containing the phylogenetic tree can be downloaded. Finally the Phylo.io [45] tool was used for the visual presentation of the phylogenetic trees.



**Fig. 4** The BLAST and the JBrowse tools available in the PGDB. **(a)** BLAST page to search for regions of similarity between sequences. **(b)** The BLAST results page. **(c)** Visualization of genomic regions using the Genome browser. **(d)** Detailed information about a single region

## Transcription factor

This page provides a search function for predicted TF and TR families in the 9 pear genomes. The search form allows users to retrieve additional TF families by entering a specific gene ID or, instead, the family name for a complete list of genes in specific families. We also provide a list of 94 families at the bottom of the search page to serve as reference.

## Genome browser

The genome browser is an important tool for visualization of high-throughput sequencing data. JBrowse [46] is a genome browser based on HTML5 and JavaScript, which contains a fully dynamic AJAX interface. We collected genome and annotation information for 9 pear varieties, as well as genome re-sequencing data for 30 pear cultivars, which were mapped to the ‘Dangshansuli’ v1.0 genome [1, 47, 48]. In addition, we mapped transcriptome data from five pear cultivars of seven stages to nine pear reference genomes. These data can all be viewed in JBrowse. On the left-hand side of the genome browser, the ‘Available Tracks’ option provides all displayable file options. After choosing which files to display, the information will appear on a window located in the right-hand side (Fig. 4c). Clicking on the different parts of the sequences will display detailed data information and allows users to browse gene sequences, structure and annotations (Fig. 4d).

## Other options

The ‘Species’ page contains a brief introduction to the 9 pear genomes available and provides links to the relevant literature. The ‘About’ module contains three parts: the ‘Download’ page allows users to download genomic information, including FASTA files of the genome assembly, gene, CDS, and protein sequences, and gene structure data in the GFF format. The ‘Link’ page provides quick links to other plant-related databases and resources. The ‘Contact’ option allows users to contact the administrators of the PGDB.

## Conclusion

PGDB currently includes genomic, transcriptomic and re-sequencing data for pear, which can be displayed through a user-friendly platform that is functionally practical. This can help researchers quickly retrieving, browsing and analyzing multi-omics data and promote in-depth studies and development of pear omics.

## Abbreviations

|            |   |
|------------|---|
| BAC-by-BAC | Bacterial artificial chromosome         |
| MySQL      | My's Structured Query Language          |
| HTML5      | HyperText Markup Language, version 5    |
| CSS        | Cascading Style Sheets                  |
| PHP        | PHP Hypertext Preprocessor              |
| AJAX       | Asynchronous JavaScript and XML         |
| GFF        | Generic Feature Format                  |
| GO         | Gene Ontology                           |
| KEGG       | Kyoto Encyclopedia of Genes and Genomes |
| GDR        | Genome Database for Rosaceae            |
| KAAS       | KEGG Automatic Annotation Server        |

|       |   |
|-------|---|
| CDS   | Coding DNA Sequence                           |
| TFs   | Transcription factors                         |
| TRs   | Transcriptional regulators                    |
| RPKM  | Reads Per Kilobase per Million mapped reads   |
| BLAST | Basic Local Alignment Search Tool             |
| NWK   | Newick  |
| AJAX  | Asynchronous JavaScript and XML               |
| NCBI  | National Center for Biotechnology Information |

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04406-5>.

Table S1: Profile of the 8 pear genomes available in the PGDB

Table S2: Functional annotations of pear genomes, including GO terms, KEGG pathways and InterPro terms

Table S3: Transcription factor and regulator gene families in 9 pear genomes

## Acknowledgements

We thank Che Lu for technical help provided in the construction of the website, and the Network Information Technology Center of Shandong Agricultural University.

## Authors' contributions

JW, YMZ designed and directed the project. LSC, YMZ collected and analyzed the data and constructed the database. LSC wrote the manuscript, which was revised by JW, YMZ, CX, ZSX, YMS contributed to the project conception. CX, ZSX, FPZ, DKG collected samples from the Shandong Academy of Agricultural Sciences. WSW, CX, ZSX, FPZ, DKG, WZQ, and YZL tested the database. All authors read and approved the final version of the manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (31820103012, 32172531), the Earmarked Fund for China Agriculture Research System (CARS-28), the High-level Talent Project of Shandong Agriculture University, and the Natural Science Foundation of Shandong Province (ZR2021MC177).

## Data availability

All original RNA-seq data (BioProject PRJNA309745) and re-sequencing data (BioProject PRJNA563813, PRJNA782471, PRJNA381668) are available at the National Center for Biotechnology Information (NCBI) database. This database can be accessed for free at <http://pyrusgdb.sdau.edu.cn>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare they have no competing interests.

Received: 10 March 2023 / Accepted: 9 August 2023

Published online: 15 September 2023

## References

- Wu J, Wang Y, Xu J, Korban SS, Fei Z, Tao S, Ming R, Tai S, Khan AM, Postman JD, et al. Diversification and independent domestication of asian and european pears. *Genome Biol.* 2018;19(1):77.
- Ferradini N, Lancioni H, Torricelli R, Russi L, Dalla Ragione I, Cardinali I, Marconi G, Gramaccia M, Concezzi L, Achilli A, et al. Characterization and phylogenetic analysis of ancient italian landraces of Pear. *Front Plant Sci.* 2017;8:751.
- Wu J, Li LT, Li M, Khan MA, Li XG, Chen H, Yin H, Zhang SL. High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *J Exp Bot.* 2014;65(20):5771–81.
- Li J, Zhang M, Li X, Khan A, Kumar S, Allan AC, Lin-Wang K, Espley RV, Wang C, Wang R et al. Pear genetics: recent advances, new prospects, and a roadmap for the future. *Hortic Res* 2022, 9.
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, et al. The genome of the pear (*Pyrus bretschneideri* Rehd). *Genome Res.* 2013;23(2):396–408.
- Chagne D, Crowhurst RN, Pindo M, Thrimawithana A, Deng C, Ireland H, Fiers M, Dzierzon H, Cestaro A, Fontana P et al. The Draft Genome Sequence of European Pear (*Pyrus communis* L. 'Bartlett'). *PLoS ONE* 2014, 9(4).
- Shirasawa K, Itai A, Isobe S. Chromosome-scale genome assembly of Japanese pear (*Pyrus pyrifolia*) variety 'Nijisseiki'. *DNA Res* 2021, 28(2).
- Ou C, Wang F, Wang J, Li S, Zhang Y, Fang M, Ma L, Zhao Y, Jiang S. A de novo genome assembly of the dwarfing pear rootstock Zhongai 1. *Sci Data.* 2019;6(1):281.
- Gao YH, Yang QS, Yan XH, Wu XY, Yang F, Li JZ, Wei J, Ni JB, Ahmad M, Bai SL et al. High-quality genome assembly of 'Cuiguan' pear (*Pyrus pyrifolia*) as a reference genome for identifying regulatory genes and epigenetic modifications responsible for bud dormancy. *Hortic Res-England* 2021, 8(1).
- Linsmith G, Rombauts S, Montanari S, Deng CH, Celton JM, Guerif P, Liu C, Lohaus R, Zurn JD, Cestaro A et al. Pseudo-chromosome-length genome assembly of a double haploid "Bartlett" pear (*Pyrus communis* L.). *Gigascience* 2019, 8(12).
- Dong X, Wang Z, Tian L, Zhang Y, Qi D, Huo H, Xu J, Li Z, Liao R, Shi M, et al. De novo assembly of a wild pear (*Pyrus betuleaefolia*) genome. *Plant Biotechnol J.* 2020;18(2):581–95.
- Zhang H, Wafula EK, Eilers J, Harkess AE, Ralph PE, Timilsena PR, dePamphilis CW, Waite JM, Honaas LA. Building a foundation for gene family analysis in Rosaceae genomes with a novel workflow: a case study in *Pyrus* architecture genes. *Front Plant Sci.* 2022;13:975942.
- Xue H, Wang S, Yao JL, Deng CH, Wang L, Su Y, Zhang H, Zhou H, Sun M, Li X, et al. Chromosome level high-density integrated genetic maps improve the *Pyrus bretschneideri* 'DangshanSuli' v1.0 genome. *BMC Genomics.* 2018;19(1):833.
- Li XL, Liu L, Ming ML, Hu HJ, Zhang MY, Fan J, Song BB, Zhang SL, Wu J. Comparative transcriptomic analysis provides insight into the domestication and improvement of Pear (*P. pyrifolia*) Fruit. *Plant Physiol.* 2019;180(1):435–52.
- Song B, Li X, Cao B, Zhang M, Korban SS, Yu L, Yang W, Zhao K, Li J, Wu J. An identical-by-descent segment harbors a 12-bp insertion determining fruit softening during domestication and speciation in *Pyrus*. *BMC Biol.* 2022;20(1):215.
- Ren HY, He YH, Qi XJ, Zheng XL, Zhang SW, Yu ZP, Hu FR. The bayberry database: a multiomic database for *Myrica rubra*, an important fruit tree with medicinal value. *BMC Plant Biol* 2021, 21(1).
- Xu HM, Yu QY, Shi Y, Hua XT, Tang HB, Yang L, Ming R, Zhang JS. PGD: Pine-apple Genomics database. *Hortic Res-England* 2018, 5.
- Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 2015;43(Database issue):D1049–1056.
- Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M, et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 2017;45(D1):D190–9.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9):1236–40.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007, 35(Web Server issue):W182–185.
- Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, et al. Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiol.* 2004;135(2):773–82.
- Manna M, Thakur T, Chirom O, Mandlik R, Deshmukh R, Salvi P. Transcription factors as key molecular target to strengthen the drought stress tolerance in plants. *Physiol Plant.* 2021;172(2):847–68.
- Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, et al. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant.* 2016;9(12):1667–70.



25. Jensen MK, Kjaersgaard T, Nielsen MM, Galberg P, Petersen K, O'Shea C, Skriver K. The Arabidopsis thaliana NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem J*. 2010;426(2):183–96.
26. Dubos C, Stracke R, Grotewold E, Weissshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. *Trends Plant Sci*. 2010;15(10):573–81.
27. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49.
28. Du L, Zhang C, Liu Q, Zhang X, Yue B, Hancock J. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. *Bioinformatics*. 2018;34(4):681–3.
29. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.
30. Chen H, Song Y, Li LT, Khan MA, Li XG, Korban SS, Wu J, Zhang SL. Construction of a high-density simple sequence repeat Consensus Genetic Map for Pear (*Pyrus* spp). *Plant Mol Biol Rep*. 2015;33(2):316–25.
31. Yamamoto T, Kimura T, Shoda M, Imai T, Saito T, Sawamura Y, Kotobuki K, Hayashi T, Matsuta N. Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. *Theor Appl Genet*. 2002;106(1):9–18.
32. Yamamoto T, Kimura T, Shoda M, Ban Y, Hayashi T, Matsuta N. Development of microsatellite markers in the Japanese pear (*Pyrus pyrifolia* Nakai). *Mol Ecol Notes*. 2002(1):2.
33. Inoue E, Matsuki Y, Anzai H, Evans K. Isolation and characterization of microsatellite markers in Japanese pear (*Pyrus pyrifolia* Nakai). *Molecular Ecology Notes*. 2007.
34. FERNÁNDEZ-FERNÁNDEZ F, Harvey NG, James CM. Isolation and characterization of polymorphic microsatellite markers from European pear (*Pyrus communis* L). *Mol Ecol Notes*. 2010;6(4):1039–41.
35. Lei WANG, Long WANG, Hua-bai XUE, Xiu-gen LI, Jiang LI. Construction of SSR genetic linkage map and comparison on Pears[J]. *Scientia Agricultura Sinica*. 2016;49(12):2353–67.
36. Zhang MY, Xue C, Xu L, Sun H, Qin MF, Zhang S, Wu J. Distinct transcriptome profiles reveal gene expression patterns during fruit development and maturation in five main cultivated species of pear (*Pyrus* L). *Sci Rep*. 2016;6:28130.
37. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25(15):1966–7.
38. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187–191.
39. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
40. Li DQ, Mei HH, Shen Y, Su S, Zhang WL, Wang JT, Zu M, Chen W. ECharts: A declarative framework for rapid construction of web-based visualization (vol 2, pg 136, 2018). *Visual Informatics*. 2021; 5(1):43–43.
41. Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*. 2007;23(17):2334–6.
42. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
43. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74.
44. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9.
45. Robinson O, Dylus D, Dessimoz C. Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol Biol Evol*. 2016;33(8):2163–6.
46. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elisk CG, Lewis SE, Stein L, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol*. 2016;17:66.
47. Zhang MY, Xue C, Hu H, Li J, Xue Y, Wang R, Fan J, Zou C, Tao S, Qin M, et al. Genome-wide association studies provide insights into the genetic determination of fruit traits of pear. *Nat Commun*. 2021;12(1):1144.
48. Chen XN, Zhang MY, Sun MY, Liu YY, Li SN, Song BB, Li MY, Zhang SL, Wang RZ, Li JM et al. Genome-wide genetic diversity and IBD analysis reveals historic dissemination routes of pear in China. *Tree Genet Genomes*. 2022; 18(1).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# 神农中华农业科技奖 证书

为表彰在我国农业科学技术进步工作中做出突出贡献的获奖者，特颁发此证书，以资鼓励。

成果名称：南京农业大学梨遗传与种质创新团队  
奖励等级：优秀创新团队奖  
获奖者：张明月（第14完成人）  
获奖者单位：南京农业大学



证书编号： 2019-TD10-1-R14





第二届  
优秀女青年奖

张明月

中国植物生理与植物分子生物学学会  
先正达生物科技(中国)有限公司  
二零二一年七月