

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  
Jun Wu  
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.

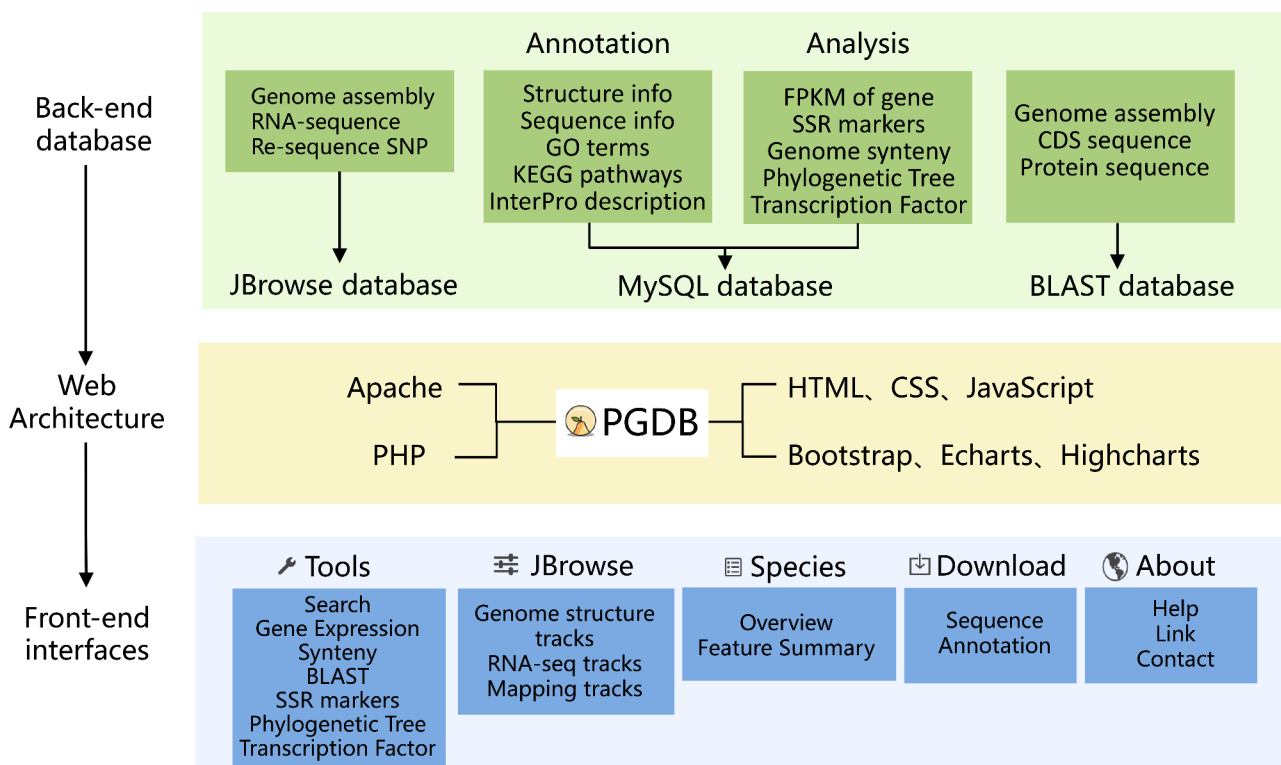
sequenced and assembled using the HiSeq Illumina technology combined with a BAC-by-BAC strategy [5]. After this, the western variety ‘Bartlett’ (*P. communis*) was sequenced with Roche’s 454 Sequencing Technology [6]. In recent years, several more pear reference genomes were published owing the rapid 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, syteny, 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), ‘Dangshansuli’ 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 MCSScanX [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

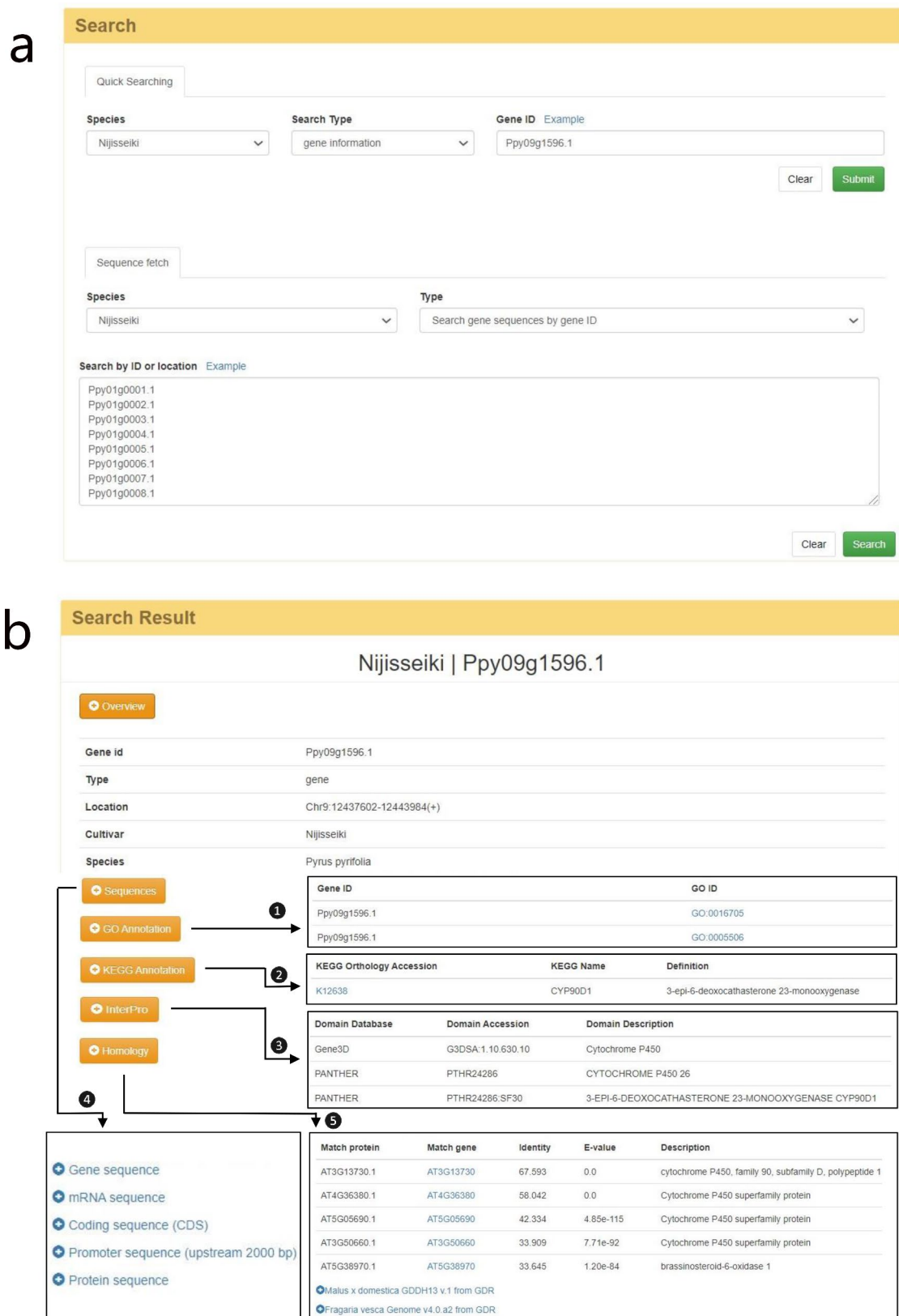
The ‘Search’ page provides two retrieval modules (Fig. 2a). In the ‘Quick searching’ module, users can first search for detailed annotations on genes by simply selecting the cultivar genome and inputting the gene ID. The results page includes information on the sequences (including gene, CDS, and protein), 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. 2** Search page of the PGDB. **(a)** Search for genetic information and sequences. **(b)** Genetic information search results, including gene details, GO <sup>Ⓞ</sup>, KEGG <sup>Ⓞ</sup> and InterPro <sup>Ⓞ</sup> functional annotations, sequences <sup>Ⓞ</sup>, and homologous genes <sup>Ⓞ</sup>



**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.





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. MCSScanX: 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, Elsik 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.