

PPGR: a comprehensive perennial plant genomes and regulation database

Sen Yang^{1,2,3,†}, Wenting Zong^{4,5,6,†}, Lingling Shi^{1,2,3}, Ruisi Li^{1,2,3}, Zhenshu Ma^{1,2,3}, Shubao Ma^{1,2,3}, Jingna Si^{1,2,3}, Zhijing Wu^{1,2,3}, Jinglan Zhai^{1,2,3}, Yingke Ma^{4,5}, Zhuojing Fan^{4,5}, Sisi Chen^{1,2,3}, Huahong Huang⁷, Deqiang Zhang^{1,2,3}, Yiming Bao^{4,5,6,*}, Rujiao Li^{4,5,6,*} and Jianbo Xie^{1,2,3,*}

¹State Key Laboratory of Tree Genetics and Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

²National Engineering Research Center of Tree Breeding and Ecological Restoration, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

³The Tree and Ornamental Plant Breeding and Biotechnology Laboratory of National Forestry and Grassland Administration, Beijing Forestry University, Beijing 100083, China

⁴National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformatics, Beijing 100101, China

⁵China National Center for Bioinformatics, Beijing 100101, China

⁶University of Chinese Academy of Sciences, Beijing 100049, China

⁷State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Lin'an, Hangzhou 311300, China

*To whom correspondence should be addressed. Tel: +86 10 62336007; Fax: +86 10 62336164; Email: jbxie@bjfu.edu.cn

Correspondence may also be addressed to Rujiao Li. Tel: +86 10 84097638; Email: lirj@big.ac.cn

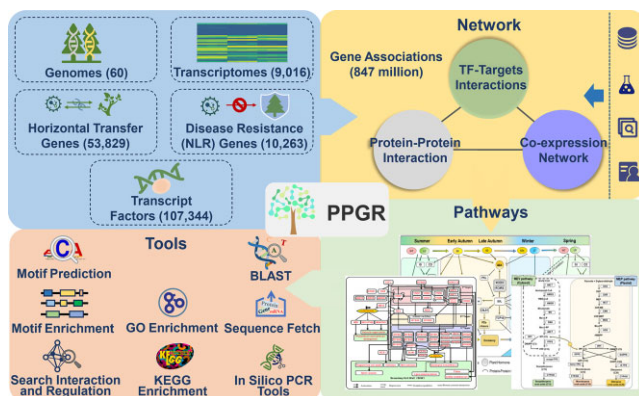
Correspondence may also be addressed to Yiming Bao. Tel: +86 10 84097858; Email: baoyim@big.ac.cn

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Perennial woody plants hold vital ecological significance, distinguished by their unique traits. While significant progress has been made in their genomic and functional studies, a major challenge persists: the absence of a comprehensive reference platform for collection, integration and in-depth analysis of the vast amount of data. Here, we present PPGR (Resource for Perennial Plant Genomes and Regulation; <https://ngdc.cnpcb.ac.cn/ppgr/>) to address this critical gap, by collecting, integrating, analyzing and visualizing genomic, gene regulation and functional data of perennial plants. PPGR currently includes 60 species, 847 million protein–protein/TF (transcription factor)–target interactions, 9016 transcriptome samples under various environmental conditions and genetic backgrounds. Noteworthy is the focus on genes that regulate wood production, seasonal dormancy, terpene biosynthesis and leaf senescence representing a wealth of information derived from experimental data, literature mining, public databases and genomic predictions. Furthermore, PPGR incorporates a range of multi-omics search and analysis tools to facilitate browsing and application of these extensive datasets. PPGR represents a comprehensive and high-quality resource for perennial plants, substantiated by an illustrative case study that demonstrates its capacity in unraveling gene functions and shedding light on potential regulatory processes.

Graphical abstract



Received: August 15, 2023. Revised: September 21, 2023. Editorial Decision: October 12, 2023. Accepted: October 13, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Perennial woody plants belong to an ecologically significant plant group with great value as a source of essential oils, clade-specific secondary products, wood biomass and benefits to ecosystems, industry and medicine (1–3). Perennial woody plants have developed strategies to regulate their growth and dormancy cycles, allowing them to survive harsh winters or dry seasons. This resilience makes them more resistant to environmental stresses compared to annual plants (4). Genomic sequence data provided evidence suggesting that the regulatory mechanisms in perennial woody plants are extremely complicated and involve distinct regulation processes that are not similar to those in annual plants (5–7). These distinct regulation processes could be deciphered by the reconstruction of the regulatory network (protein–protein/TF–target interactions). Over the past decade, genomic and functional studies have produced massive amount of data of perennial woody plants (8,9), deposited in public databases and documents. For instance, the *Populus trichocarpa* ‘Nisqually-1’ genome was released in 2006, which was the first woody perennial genome to be assembled and annotated (10). Nevertheless, the integration of its experimental findings from ChIP-seq, DAP-seq, Yeast one-hybrid and Dual luciferase reporter assays has remained unrealized.

Several databases like Phytozome (11), Plantgenie (12), EnsemblPlants (13), treegenesdb (14) and TRANSNAP (15) gather genomic and transcriptomic data, and others focus on gene function and regulation, such as LSD (16), MetaCyc (17), IntAct (18), STRING (19), GFDP (4) and ChIP-Hub (20). These, however, only cover a small portion of perennial woody plants and overlook the unique biological processes that cannot be addressed in annual plants. These datasets frequently differ in formats and features, complicating the integration and comparison of data from different research groups. Consequently, this situation hinders the thorough understanding of perennial woody plants and in-depth exploration and comparison with other plant species.

Here, we describe PPGR (Resource for Perennial Plant Genomes and Regulation), the first database focusing on the genomes and regulatory network in hierarchical regulatory processes of perennial woody plants. PPGR integrates genomic data of 60 species, 847 million gene associations including co-expressed associations, protein–protein interactions (PPI) and transcription factor–targets interactions (TTI), 9016 independent transcriptomic datasets, 107 344 TFs, 10 263 nucleotide binding leucine-rich repeat receptor (NLR) genes, 53 829 horizontal transfer genes and 87 372 candidate genes involved in perennial woody plants characteristic pathways such as wood formation, seasonal dormancy, terpene biosynthesis and leaf senescence. By providing a centralized database and tools, PPGR facilitates research and promotes collaboration with extended benefits.

Construction

Organization of features

PPGR includes six main modules: Genomes, Network, Resources, Pathways, Tools and Document, each with distinct functionalities and advantages. ‘Quick Search’ and ‘Regulation Search’ tool for gene function and regulation networks are featured on the homepage. One helpful feature is the dynamic data overview page which allows users to browse the

current data types and numbers in PPGR. The ‘Genomes’ module allows users to easily view and analyze genomic data. The ‘Network’ module allows users to access to all pre-computed information both at high-level network view and at the individual gene association record (PPI, TTI and co-expressed). The ‘Resources’ module is organized into horizontal transfer genes, NLR genes, TFs and transcriptome section, which are essential for studying perennial plants evolution. The ‘Pathways’ module is organized into wood formation, seasonal dormancy, terpene biosynthesis and leaf senescence section. The ‘Tools’ module provides eight useful online analysis tools to analyze the regulatory mechanisms of candidate genes potential. The ‘Document’ module provides detailed instructions and guidelines on how to use the database effectively. It includes brief introduction of PPGR database, along with information on the different sections including Search, Genomes, Network, Resources and Pathway of the database. Below, we describe the currently available data and interfaces.

For a comprehensive understanding of the adaptation and evolution of perennial woody plants, we first collected and combined genomic, transcriptomic and experimental functional data from public documents and databases. Genome assembly completeness was evaluated and detailed sequencing information was manually curated. Co-expression analyses were performed by calculating Pearson correlation coefficient between gene pairs, using their gene expression levels. Second, the key regulators and genes controlling specific biological processes were mined by processing and analyzing these datasets. Third, the gene regulatory network was constructed based on multifaceted approaches. In addition, users are also provided with several practical and powerful tools that contribute to the deep mining of gene function. Finally, PPGR, a comprehensive and high-quality resource for perennial woody plants, was constructed (Figure 1).

Genomes

Currently, this module contains whole genome assemblies and annotations from 60 species, many of which are model species that have been extensively studied. For example, the genome assembly of black cottonwood *Populus trichocarpa* is available in PPGR, which serves as the reference genome for the *Populus* genus. The species page in the Genomes module provides users with a comprehensive summary of each species, a list of genes and even JBrowse (21) for genome visualization. This module allows users to view each genome summary, including common name, lineage, reference assembly version, assembly completeness, genome size, transcript number, gene number, gene position and related references. Users can freely view the adjacent genome areas in the same genome track by dragging or zooming in or out. Additionally, users can search the interested genomic regions by using the gene ID, products or genomic positions.

We reannotated a total of 2.36 million protein-coding genes across all genomes. These annotations included 1.34 million functional elements, such as InterPro domains (22) and Pfam domains (23). At the gene level, the annotations encompass essential information such as protein product, sequence structure, domains, regulatory interactions, expression profiles, cross-references to multiple databases, Gene ontology (GO) annotation (24,25) and ortholog groups. These comprehensive annotations aid users in understanding a gene’s function, characteristics and potential implications in various

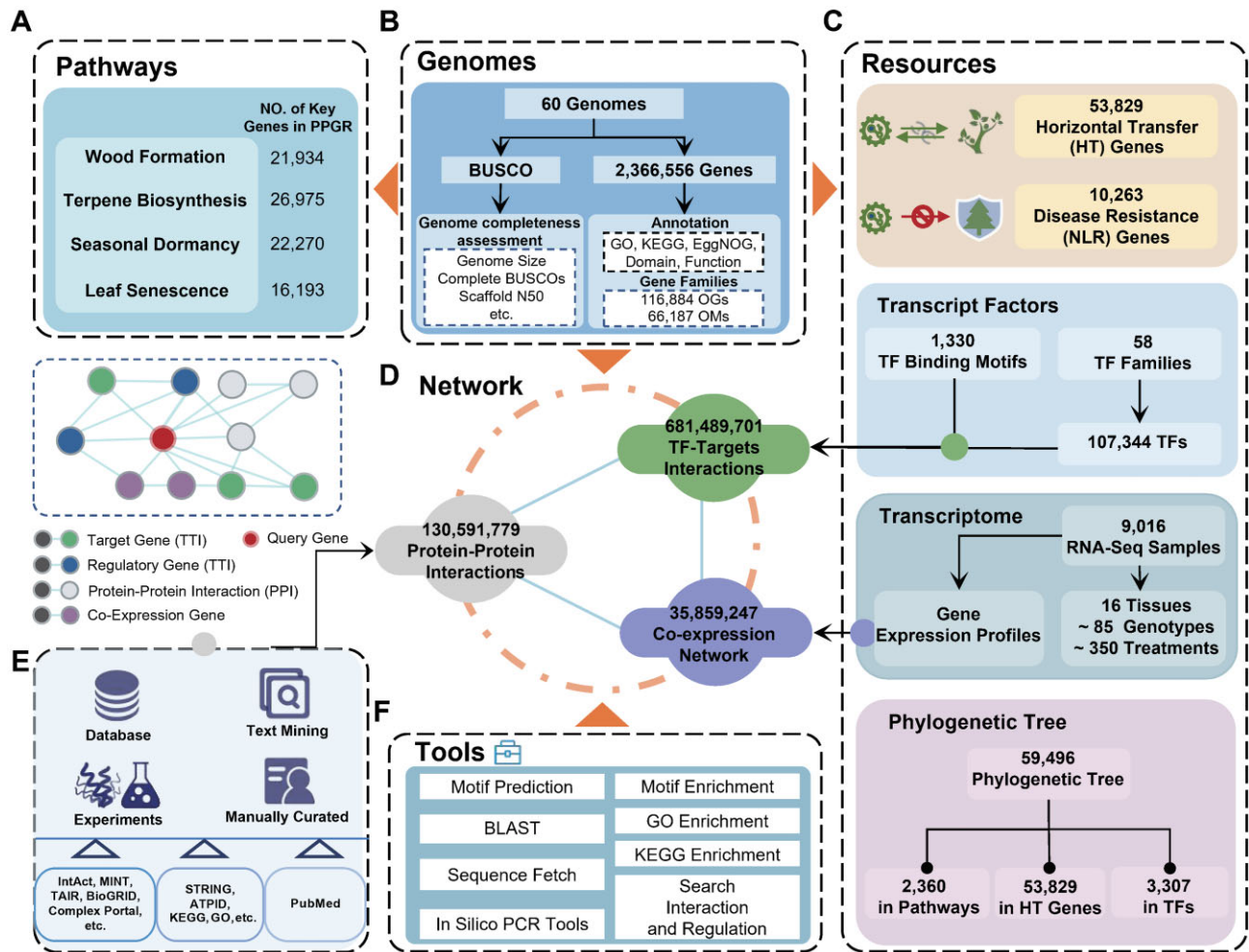


Figure 1. Overview of the PPGR database. **(A)** The ‘Pathways’ module is organized into wood formation, seasonal dormancy, terpene biosynthesis and leaf senescence sections. **(B)** Genomic resources included 60 species, and 2.36 million protein coding genes. **(C)** The ‘Resources’ module is organized into horizontal transfer genes, NLR genes, transcription factors and transcriptome sections. **(D)** Summary of the network construction methods. The network includes PPI, TTI and co-expressed subnetwork. **(E)** The approaches used in the network construction. **(F)** Display of web tools. NLR: nucleotide binding leucine-rich repeat receptor. PPI: protein–protein interaction. TTI: TF–targets interaction.

treatments or conditions. Additionally, we developed a gene search function that allows users to browse and compare gene structures, transcriptional profiles and regulatory networks. This module helps determining whether homologous genes have undergone functional differentiation.

Network

The sophisticated processes of perennial woody plants rely on complex gene regulation networks, which are carefully organized in space and time. This network consists of various subnetworks, including PPI, TTI and co-expressed associations that are seemingly unique but closely related. To fully understand gene function, PPGR database provides a comprehensive collection of quality-controlled PPI, TTI and co-expressed associations (Figure 2). Currently, the database contains 847 million pre-computed functional associations, which are constructed using multifaceted approaches.

The collection of regulatory gene associations in PPGR was obtained through several methods: (i) We mined transcriptional regulatory associations from literature and manually

curated each association, resulting in 3956 high-confidence PPI and 22 763 TTI based on 205 public datasets, (ii) We integrated 25 691 877 PPI and 326 431 826 TTI with experimental evidence imported from primary database repositories such as IntAct (18), Complex Portal (26), BioGRID (27), TAIR (28) and MINT (29), (iii) We integrated high-score association evidence from databases like STRING (19), KEGG (30) and Meta-Cyc (17), which identified 449 576 854 associations, (vi) We performed co-occurrence analysis across scientific literature, identifying 11 931 082 associations, (v) We identified 35 859 247 associations based on co-expression evidence from pre-computed transcriptome data and (vi) We predicted 681 489 701 TTI in silico through searching for potential binding sites of TFs.

The module provides user-friendly interfaces for searching and browsing gene regulatory data (PPI, TTI and co-expressed associations), as well as inspecting their underlying evidence. Users can query for genes of interest or a set of genes using different identifier name spaces. The resulting network can be interactively inspected, rearranged and filtered with variable stringency. Detailed information for each node in the net-

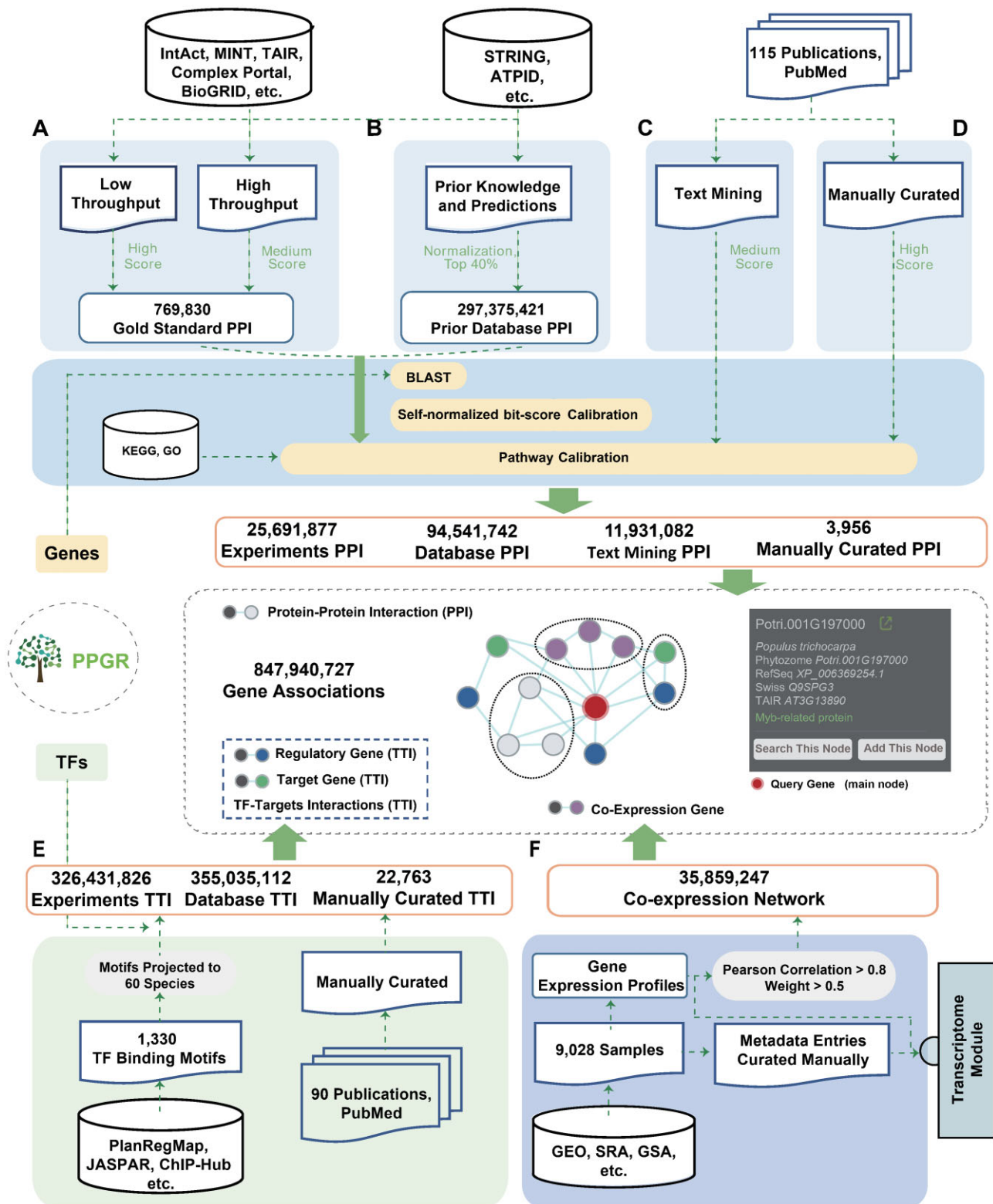


Figure 2. The approaches PPGR uses for the integration of gene associations. (A) Integrated experimental evidence imported from primary database repositories. (B) Integrated the high-score association evidence from curated databases. (C) Co-occurrence analysis across the scientific literature and protein–protein interactions from literature. (D) Mining and curation of the transcriptional regulatory associations from literature. (E) transcriptional regulatory associations from literature and *in silico* prediction through searching potential binding sites of TFs. (F) Identified associations based on the co-expression evidence from the pre-computed transcriptome samples.

work, including annotation, gene structure, expression profiles, PANTHER (31), MobiDBLite (32) and Pfam (33) functional domains, can be viewed in a pop-up window by clicking on the node. Each edge in the network represents a known or predicted interaction, and additional details about the underlying evidence are provided.

Resources

Horizontal transfer genes

Horizontal gene transfer (HGT) refers to the movement of new genes between distinct species and has long been considered as a crucial process in plants evolution (34,35). PPGR performed systematic analyses for the acquired genes derived from any foreign source in all perennial plant species. A summary of the horizontally acquired genes, including the donor, gene ID and the top hit, is provided. PPGR identified a total of 53 829 quality-controlled acquired genes transferred from foreign sources to perennial plants. Users can easily access detailed descriptions that include information on gene structure, domains, regulatory networks, annotations, orthologous groups and expression profiles. This resource can assist users in identifying new trait-associated genes and specific regulatory pathways across perennial plants.

Disease resistance (NLR) genes

Plants have evolved various defense mechanisms to protect themselves from microbial pathogens in their natural habitats. One important way plants defend against disease is through a cell-based surveillance system involving immune receptors called Nucleotide-binding domain and Leucine-rich repeat receptor genes (NLRs). NLRs play crucial role in detecting and responding to pathogen invasion (36). PPGR has identified 10 263 high-confidence NLR genes in perennial plants. PPGR offers a summary of NLR genes that includes the source, classification, gene ID and description. PPGR provides plant researchers and breeders with valuable information on NLR genes. This includes detailed descriptions of their functional domains, gene structures, orthologous groups and regulatory networks. By equipping them with this information, PPGR enhances their knowledge and strategies in combating plant pathogens, allowing them to develop a consensus plan.

Transcription factors

Currently, PPGR has identified 107 344 TFs from 60 species using a strict criterion and made extensive annotation at both family and gene levels. PPGR provides free online access to both the sequences and annotation information for each identified TF. Additionally, PPGR performed phylogenetic analysis for each TF family, which can help in identifying the common ancestors, determining the sequence conservation and inferring functional similarities or differences among TF members.

Transcriptome

To investigate gene function and expression, we collected 9016 RNA-seq data from 430 transcriptome studies conducted under various treatments or genetic backgrounds. We then reanalyzed the raw data using a quality control method (see Supplementary data sources and methods). The descriptions of the data type, genotype, tissues and samples were manually curated and combined. This combined information is displayed in the Transcriptome module, providing users with valuable insights into the function and ex-

pression features of the genes of interest. Additionally, we have developed and integrated visualization tools to assist users in gene mining and understanding their regulatory mechanisms.

Pathways

Several regulation pathways in woody plants are very complex and involve distinct physiological and anatomic processes, such as perennial secondary growth (37), long life-span (38) and cambium dormancy (39). PPGR represents an ongoing effort to provide insights into the evolution and resources for advancing research on perennial woody plants adaptation and development. Wood formation genes are involved in the synthesis of cell wall components, such as lignin and cellulose, which give plants strength and structure. Seasonal dormancy genes regulate plant's response to changing environmental conditions, allowing it to enter a dormant state during unfavorable seasons. Terpene biosynthesis genes are responsible for terpenes production, which are important for plant defense and communication. Leaf senescence genes control leaf aging and eventual death processes. The pathways currently include 87 372 key genes for controlling wood formation, seasonal dormancy, terpene biosynthesis and leaf senescence across 60 species. The annotation of these genes was done using domains, evolutionary analyses and manually curated experimental evidences, which help users identify their functions and evolutionary relationships with genes in other species. These data contribute to a better understanding of specific processes of perennial woody plants and consequently to precise manipulation of economic and horticultural traits, such as wood formation and flowering time. Detailed annotations and their functional associations were provided for each key gene.

Tools

PPGR offers a suite of eight practical tools. Among these, 'BLAST' tool is used to search for homologous sequences (either protein or nucleotide sequences) from the genome assembly of PPGR or input sequences. The 'GO enrichment' and 'KEGG enrichment' tools are used to perform the functional enrichment analyses using the R language clusterProfiler package (40). 'Motif Prediction' tool is employed to identify the potential binding motifs for TFs, which helps to determine of TFs likely involved in regulating the genes or genomic regions of interest. The 'Motif Enrichment' tool identifies statistically significant motifs that are enriched in the input sequences. Both motif prediction and motif enrichment analyses are important for understanding the regulatory mechanisms of gene expression and identifying potential TFs involved in specific biological processes. 'In silico PCR' helps users confirm the specificity of the designed primers. The 'Sequence Fetch' tool is designed to assist users in extracting specific nucleotide sequences, such as promoters from genome sequence. By providing the desired positions, users can easily retrieve the corresponding sequence. Specifically, 'Search Interaction and Regulation' tool could help users to search candidate interactions. Users can retrieve information about its interacting partners by inputting the name or identifier of a specific gene. This can be particularly useful in understanding the regulatory networks and pathways in which the candidate gene is involved. 'Search Interaction and Regulation' tool provides a comprehensive overview of the gene's interactions, allowing users to

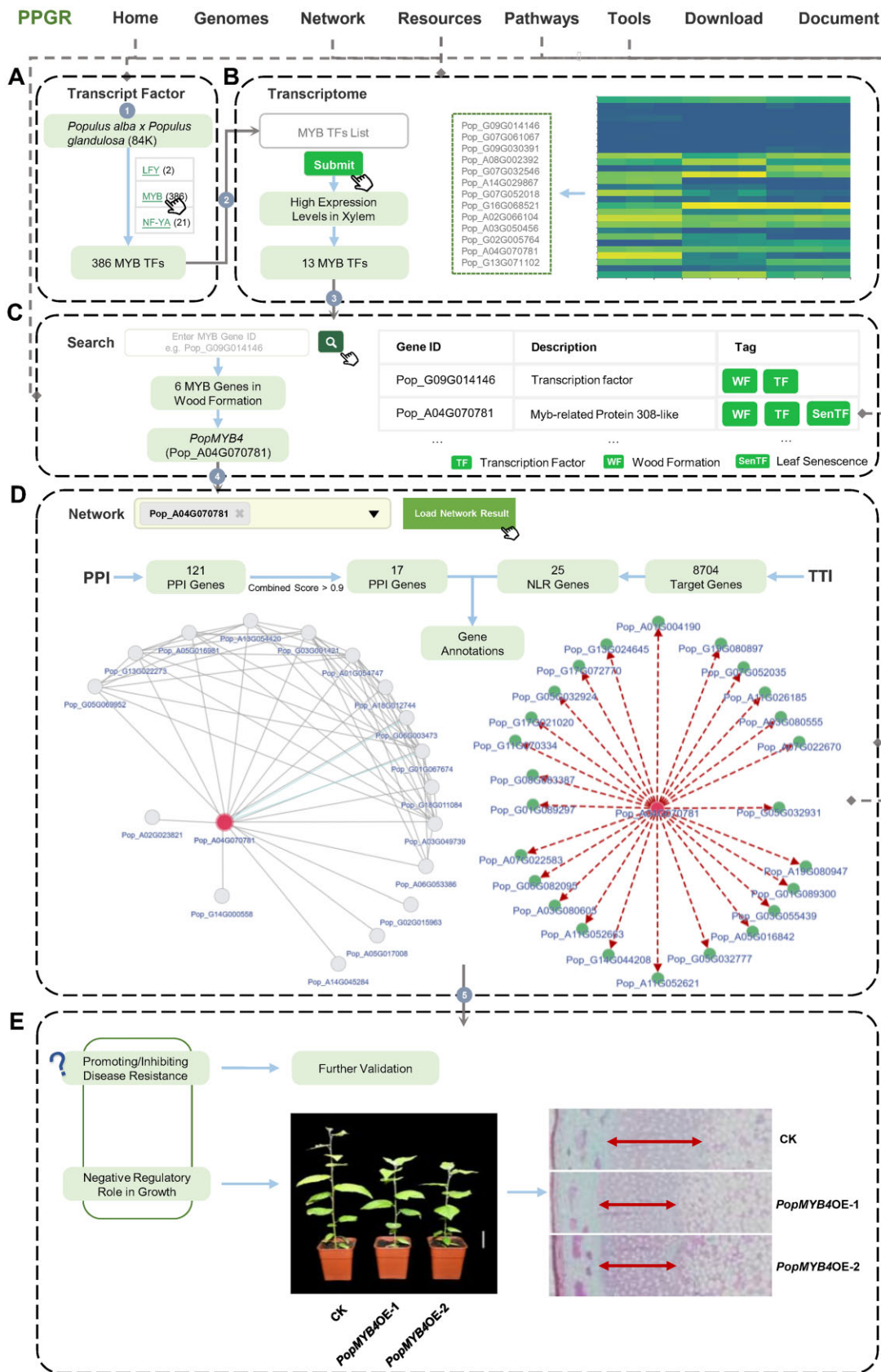


Figure 3. Case study of identifying *MYB* genes associated with wood formation. **(A)** Searching the gene list of the *MYB* TFs. **(B)** The transcription analyses of the *MYB* TFs. **(C)** The annotation of the candidate *MYB* genes. **(D)** The regulatory network analyses of *PopMYB4*. **(E)** The function prediction of *PopMYB4*.

explore potential functional relationships and identify biological processes key players. By leveraging these online analytical tools, users can flexibly reanalyze the public datasets and obtain new insights.

Case study: mining genes associated with wood formation

Wood formation represents a crucial biological process in perennial woody plants (41). To characterize MYB members that are involved in regulating wood formation, we performed a series of extensive analyses using PPGR. Leveraging the ‘Transcript Factor’ module, we first identified 386 MYB family members in *Populus alba* x *Populus glandulosa* ‘84K’ (Figure 3A). The thirteen members exhibiting high expression levels in xylem tissues were selected for further analysis (Figure 3B). Significant upregulation specifically in xylem tissue suggests their potential involvement in wood formation. With the aid of the global search tool provided by PPGR, six genes that may function in the wood formation process were identified (Figure 3C). Then *PopMYB4* (Pop_A04G070781) was further selected for subsequent analyses. Subsequent functional assessment of *PopMYB4* was facilitated by the ‘Network’ module (Figure 3D). Seventeen interaction proteins were detected, with half being annotated as transcriptional corepressors, suggesting their negative regulatory roles. These insights propose that *PopMYB4* may play a negative regulatory role in growth. Furthermore, 25 NLR genes were identified to be the targets of *PopMYB4*, suggesting *PopMYB4* was highly associated with disease resistance. This is consistent with previous studies that have reported *PopMYB4* ability to negatively regulate lignin synthesis and enhance disease resistance in plants (42,43). Notably, our recent experimental investigations also revealed inhibited growth and lignin synthesis in overexpressing *PopMYB4* transgenic poplars (Figure 3E).

Discussion and perspectives

PPGR offers a wide range of multi-omics data, including genome sequences, annotations and expression profiles. It also provides collections of PPI, TTI and co-expressed associations for various perennial woody plants. Additionally, PPGR provides online analysis and visualization tools to assist users in exploring the data and gaining insights into gene expression regulation and phenotypic formation. Overall, PPGR is a valuable resource for studying gene function and understanding biological processes.

PPGR has the following characteristics: (i) it integrates and optimizes the data resources of perennial woody plants. This suggests that PPGR brings together various data resources and functions related to perennial woody plants, potentially making it a valuable resource for researchers in this field, (ii) PPGR has a vast collection of multi-omics datasets, including 60 published genome assemblies, transcriptome samples, gene associations and key genes associated with the adaptation and evolution of perennial woody plants. It is beneficial for researchers studying perennial woody plants and (iii) PPGR offers a range of common online analysis tools. These tools are designed to support all published genomes and provide a convenient and efficient analysis platform.

In summary, we provide a valuable database for future functional genomics research on perennial plants. In the future, with the release of more multi-omics datasets, the database will be annually updated. Advanced multi-omics methods, such as the multilayered hierarchical gene regulatory network construction tool, will be integrated and applied. We aim to enhance our understanding of the genetic basis of perennial plant-specific traits and improve the identification of candidate loci and genes associated with these traits.

Data availability

PPGR is freely available online at <https://ngdc.cncb.ac.cn/ppgr/>. Users can directly download the data resources without registration or login.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

This work is supported by funding from the Fundamental Research Funds for the Central Universities [QNTD202305]; the Project of the National Natural Science Foundation of China [32371906, 32022057, 31972954]; the Fundamental Research Funds for the National Key R&D Program of China [2022YFD2201600, 2022YFD2200602]; Forestry and Grassland Science and Technology Innovation Youth Top Talent Project of China [2020132607]; Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38030200, XDB38030100]; Genomics Data Center Construction of Chinese Academy of Sciences [WX145XQ07-04]; Professional Association of the Alliance of International Science Organizations [ANSO-PA-2020-07]; and Open Biodiversity and Health Big Data Programme of IUBS.

Conflict of interest statement

None declared.

References

1. Tuskan, G.A., Groover, A.T., Schmutz, J., DiFazio, S.P., Myburg, A., Grattapaglia, D., Smart, L.B., Yin, T.M., Aury, J.M., Kremer, A., *et al.* (2018) Hardwood Tree Genomics: unlocking Woody Plant Biology. *Front. Plant Sci.*, **9**, 1799.
2. Li, P., Xiao, L., Du, Q.Z., Quan, M.Y., Song, Y.P., He, Y.L., Huang, W.X., Xie, J.B., Lv, C.F., Wang, D., *et al.* (2023) Genomic insights into selection for heterozygous alleles and woody traits in *Populus tomentosa*. *Plant Biotechnol. J.*, **17**, 2002–2018.
3. Taylor, A., Weigelt, P., Denelle, P., Cai, L.R. and Kreft, H. (2023) The contribution of plant life and growth forms to global gradients of vascular plant diversity. *New Phytol.*, **240**, 1548–1560.
4. Wang, H., Yan, H.W., Liu, H.L., Liu, R., Chen, J. and Xiang, Y. (2018) GFDP: the gene family database in poplar. *Database-Oxford*, **2018**, bay107.
5. Shefferson, R.P. (2009) The evolutionary ecology of vegetative dormancy in mature herbaceous perennial plants. *J. Ecol.*, **97**, 1000–1009.
6. Rohde, A. and Bhalerao, R.P. (2007) Plant dormancy in the perennial context. *Trends Plant Sci.*, **12**, 217–223.

7. Spicer,R. and Groover,A. (2010) Evolution of development of vascular cambium and secondary growth. *New Phytol.*, **186**, 577–592.
8. Li,X., Cai,K.W., Han,Z.M., Zhang,S.K., Sun,A.R., Xie,Y., Han,R., Guo,R.X., Tigabu,M., Sederoff,R., *et al.* (2022) Chromosome-level genome assembly for *Acer pseudosieboldianum* and highlights to mechanisms for leaf color and shape change. *Front. Plant Sci.*, **13**, e850054.
9. Liu,H.L., Wang,X.B., Wang,G.B., Cui,P., Wu,S.G., Ai,C., Hu,N., Li,A.L., He,B., Shao,X.J., *et al.* (2021) The nearly complete genome of *Ginkgo biloba* illuminates gymnosperm evolution. *Nat. Plants*, **7**, 748–756.
10. Tuskan,G.A., DiFazio,S., Jansson,S., Bohlmann,J., Grigoriev,I., Hellsten,U., Putnam,N., Ralph,S., Rombauts,S., Salamov,A., *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.
11. Goodstein,D.M., Shu,S.Q., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W., Hellsten,U., Putnam,N., *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.
12. Sundell,D., Mannapperuma,C., Netotea,S., Delhomme,N., Lin,Y.C., Sjodin,A., Van de Peer,Y., Jansson,S., Hvidsten,T.R. and Street,N.R. (2015) The Plant Genome Integrative Explorer Resource: plantGenIE.org. *New Phytol.*, **208**, 1149–1156.
13. Yates,A.D., Allen,J., Amode,R.M., Azov,A.G., Barba,M., Becerra,A., Bhai,J., Campbell,L.I., Martinez,M.C., Chakiachvili,M., *et al.* (2022) Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res.*, **50**, D996–D1003.
14. Falk,T., Herndon,N., Grau,E., Buehler,S., Richter,P., Zaman,S., Baker,E.M., Ramnath,R., Ficklin,S., Staton,M., *et al.* (2018) Growing and cultivating the forest genomics database, TreeGenes. *Database-Oxford*, **2019**, baz043.
15. Koshimizu,S., Nakamura,Y., Nishitani,C., Kobayashi,M., Ohyanagi,H., Yamamoto,T. and Yano,K. (2019) TRANSNAP: a web database providing comprehensive information on Japanese pear transcriptome. *Sci. Rep.*, **9**, 18922.
16. Li,Z.H., Zhang,Y., Zou,D., Zhao,Y., Wang,H.L., Zhang,Y., Xia,X.L., Luo,J.C., Guo,H.W. and Zhang,Z. (2020) LSD 3.0: a comprehensive resource for the leaf senescence research community. *Nucleic Acids Res.*, **48**, D1069–D1075.
17. Caspi,R., Foerster,H., Fulcher,C.A., Hopkinson,R., Ingraham,J., Kaipa,P., Krummenacker,M., Paley,S., Pick,J., Rhee,S.Y., *et al.* (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.*, **34**, D511–D516.
18. Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R., *et al.* (2007) IntAct - open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
19. von Mering,C., Jensen,L.J., Snel,B., Hooper,S.D., Krupp,M., Foglierini,M., Jouffre,N., Huynen,M.A. and Bork,P. (2005) STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res.*, **33**, D433–D437.
20. Fu,L.Y., Zhu,T., Zhou,X.K., Yu,R.R., He,Z.H., Zhang,P.J., Wu,Z.G., Chen,M., Kaufmann,K. and Chen,D.J. (2022) ChIP-Hub provides an integrative platform for exploring plant regulome. *Nat. Commun.*, **13**, 3413.
21. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
22. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W.Z., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G., *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.
23. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J., *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
24. Ashburner,M., Ball,C., A.,B., Judith,A.,Botstein and,David. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–34.
25. Aleksander,S.A., Balhoff,J., Carbon,S., Cherry,J.M., Drabkin,H.J., Ebert,D., Feuermann,M., Gaudet,P., Harris,N.L., Hill,D.P., *et al.* (2023) The Gene Ontology knowledgebase in 2023. *Genetics*, **224**, iyad031.
26. Meldal,B.H.M., Perfetto,L., Combe,C., Lubiana,T., Cavalcante,J.V.F., Bye-A-Jee,H., Waagmeester,A., Del-Toro,N., Shrivastava,A., Barrera,E., *et al.* (2022) Complex Portal 2022: new curation frontiers. *Nucleic Acids Res.*, **50**, D578–D586.
27. Oughtred,R., Rust,J., Chang,C., Breitkreutz,B.J., Stark,C., Willems,A., Boucher,L., Leung,G., Kolas,N., Zhang,F., *et al.* (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.*, **30**, 187–200.
28. Lamesch,P., Berardini,T.Z., Li,D.H., Swarbreck,D., Wilks,C., Sasidharan,R., Muller,R., Dreher,K., Alexander,D.L., Garcia-Hernandez,M., *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
29. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardoza,A.P., Santonico,E., *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
30. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
31. Mi,H.Y., Muruganujan,A., Ebert,D., Huang,X.S. and Thomas,P.D. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.*, **47**, D419–D426.
32. Necci,M., Piovesan,D., Dosztanyi,Z. and Tosatto,S.C.E. (2017) MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*, **33**, 1402–1404.
33. El-Gebali,S., Mistry,J., Bateman,A., Eddy,S.R., Luciani,A., Potter,S.C., Qureshi,M., Richardson,L.J., Salazar,G.A., Smart,A., *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432.
34. Ma,J.C., Wang,S.H., Zhu,X.J., Sun,G.L., Chang,G.X., Li,L.H., Hu,X.Y., Zhang,S.Z., Zhou,Y., Song,C.P., *et al.* (2022) Major episodes of horizontal gene transfer drove the evolution of land plants. *Mol. Plant*, **15**, 857–871.
35. Wang,H., Li,Y.H., Zhang,Z.H. and Zhong,B.J. (2023) Horizontal gene transfer: driving the evolution and adaptation of plants. *J. Integr. Plant Biol.*, **65**, 613–616.
36. Baggs,E.L., Monroe,J.G., Thanki,A.S., O’Grady,R., Schudoma,C., Haerty,W. and Krasileva,K.V. (2020) Convergent loss of an EDS1/PAD4 signaling pathway in several plant lineages reveals coevolved components of plant immunity and drought response. *Plant Cell*, **32**, 2158–2177.
37. Groover,A. and Robischon,M. (2006) Developmental mechanisms regulating secondary growth in woody plants. *Curr. Opin. Plant Biol.*, **9**, 55–58.
38. Plomion,C., Aury,J.M., Amselem,J., Leroy,T., Murat,F., Duplessis,S., Faye,S., Francillonne,N., Labadie,K., Le Provost,G., *et al.* (2018) Oak genome reveals facets of long lifespan. *Nat. Plants*, **4**, 440–452.
39. Ruttink,T., Arend,M., Morreel,K., Storme,V., Rombauts,S., Fromm,J., Bhalerao,R.P., Boerjan,W. and Rohde,A. (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell*, **19**, 2370–2390.
40. Wu,T.Z., Hu,E.Q., Xu,S.B., Chen,M.J., Guo,P.F., Dai,Z.H., Feng,T.Z., Zhou,L., Tang,W.L., Zhan,L., *et al.* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*, **2**, 100141.
41. Mizrahi,E. and Myburg,A.A. (2016) Systems genetics of wood formation. *Curr. Opin. Plant Biol.*, **30**, 94–100.

42. Cho, J.S., Jeon, H.W., Kim, M.H., Vo, T.K., Kim, J., Park, E.J., Choi, Y.I., Lee, H., Han, K.H. and Ko, J.H. (2019) Wood forming tissue-specific bicistronic expression of PdGA20ox1 and PtrMYB221 improves both the quality and quantity of woody biomass production in a hybrid poplar. *Plant Biotechnol. J.*, **17**, 1048–1057.
43. Tang, X.F., Zhuang, Y.M., Qi, G., Wang, D., Liu, H.H., Wang, K.R., Chai, G.H. and Zhou, G.K. (2015) Poplar PdMYB221 is involved in the direct and indirect regulation of secondary wall biosynthesis during wood formation. *Sci. Rep.*, **5**, 12240.