



PanaxGDB: A Comprehensive Platform for *Panax*

Yuan Lin^{1,2,3†}, Bing Hao^{1,2†}, Ying Chun Lu^{1,2}, Yang Dong^{1,2,4}, Ying Li^{1,2}, Guang Hui Zhang^{1,2}, Zi Jiang Yang^{1,2,3}, Gui Sheng Xiang^{1,2}, Guan Ze Liu^{1,2}, Xue Jiao Li^{1,2}, Qin Zhu^{1,2,3}, Qing Hui Yang³, Xu Zhen Li^{1,4*} and Sheng Chao Yang^{1,2*}

¹ State Key Laboratory of Conservation and Utilization of Bio-Resources in Yunnan, The Key Laboratory of Medicinal Plant Biology of Yunnan Province, National & Local Joint Engineering Research Center on Germplasm Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University, Kunming, China, ² Key Laboratory of Medicinal Plant Biology, Yunnan Agricultural University, Kunming, China, ³ College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China, ⁴ College of Biological Big Data, Yunnan Agriculture University, Kunming, China

OPEN ACCESS

Edited by:

Kioumars Ghamkhar,
AgResearch Ltd., New Zealand

Reviewed by:

Zihua Liao,
Southwest University, China
Shihai Xing,
Anhui University of Chinese
Medicine, China

*Correspondence:

Sheng Chao Yang
shengchaoyang@163.com
Xu Zhen Li
manlixuzhen@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: 25 February 2022

Accepted: 07 April 2022

Published: 25 May 2022

Citation:

Lin Y, Hao B, Lu YC, Dong Y, Li Y,
Zhang GH, Yang ZJ, Xiang GS,
Liu GZ, Li XJ, Zhu Q, Yang QH, Li XZ
and Yang SC (2022) PanaxGDB: A
Comprehensive Platform for *Panax*.
Front. Plant Sci. 13:883818.
doi: 10.3389/fpls.2022.883818

The genus *Panax* is a valuable natural medicinal source used worldwide that contains high levels of triterpenoid saponins with extensive pharmacological activities. In past decades, molecular biotechnology and breeding techniques have been respectively used to generate omics data and information on cultivars primarily from *Panax ginseng* (ginseng), *Panax quinquefolium* (American ginseng), and *Panax notoginseng* (Sanqi) to biosynthesize valuable saponins, improve product quality, and conduct cost-controlled cultivation. Although much data have been produced, there are concerns that redundant data might be generated and that relatively scattered data might be overlooked. Therefore, many scientists desire a reliable, comprehensive omics database of the *Panax* genus that could save time and promote integrated analysis. Therefore, to provide all-inclusive, reliable, and valuable information on the *Panax* genus, PanaxGDB, an open comprehensive database that integrates data on omics and information on varieties, was established. The database contains information on nearly 600 compounds from 12 *Panax* species, draft genomic sequences with annotations and gene expression levels, single nucleotide polymorphisms, genome-wide association analysis based on agronomic traits, globally collected germplasm information, summaries, omics data of the *Panax* genus, and online versatile analytic tools. The *Panax* genus database will be updated when new data are released to continue serving as a central portal to boost research on the biology and functions of *Panax*. PanaxGDB is available at: <http://panaxGDB.yнау.edu.cn>.

Keywords: *Panax*, database, cultivars, genome, saponins

INTRODUCTION

The genus *Panax* (Araliaceae) consists of 16 species primarily distributed in southwestern and northeastern Asia and North America. The herb has been used as a traditional medicine for treating cardiovascular and cerebrovascular diseases and improving immune systems for thousands of years (Zuo et al., 2017). Three high-value species, *Panax quinquefolius*, *Panax ginseng*, and *Panax notoginseng*, have been successfully domesticated and are widely cultivated in Asia and North America to meet increasing market demands (Beedata, 2020; Zhang et al., 2020; Ibaogao, 2021). In 2019, the total global market for *P. notoginseng*, *P. ginseng*, and *P. quinquefolius* exceeded

\$9.34 billion (Beedata, 2020; Ibaogao, 2021). With the market increasing tremendously, several excellent cultivars are cultivated to fulfill high-quality requirements (Zhang et al., 2020). Evaluation standards and certification platforms used to select plants for breeding vary in different regions and countries, resulting in poor exchange of cultivars information and popularity of cultivars, which slows research on breeding.

Saponins are responsible for various pharmacological activities in the *Panax* genus (Xu et al., 2019). In 12 species of *Panax*, 492 damarane-type and oleanane-type saponins have been identified. Damarane-type saponins are classified based on different carbon skeletons as protopanaxadiol-type, protopanaxatriol-type, and ocotillo-type. Different chemical configurations and ligand modifications of those saponins generate many differences in physiological and pharmacological functions (Chen et al., 2009; Nag et al., 2012). There are large differences in the types of saponins among different cultivars and species of *Panax*. Damarane-type saponins are found in most species of *Panax*, except *Panax stipuleanatus*, and ocotillo-type saponins have not been detected in *P. notoginseng* (Zhu et al., 2004). Oleanane-type saponins are primarily distributed in *P. stipuleanatus*, *Panax japonicus*, *Panax major*, and *Panax zingiberensis* (Yang et al., 2014). Other compounds with pharmacological activity include dencichine for hemostasis and flavonoids for antibacterial, antidiabetic, and anti-inflammatory effects (Ding et al., 2018; Chen et al., 2019; Liu et al., 2020; Wang et al., 2020b).

Although market demand for saponins is increasing, supplies are unable to meet the demand because of the long planting period of *Panax*. Therefore, there have been attempts to biosynthesize saponins using yeast in the past years, which requires large amounts of omics information on saponin biosynthesis (Wang et al., 2020a). Morphological differences among the related species are slight, and as a result, it is difficult to identify species such as *Panax bipinnatifidus* and its wild relatives (Zhou et al., 2020). To date, there are seven versions of genome assemblies and abundant RNA-seq data for *Panax* species, which have been used to uncover mechanisms underlying the biosynthesis of ginsenosides or to identify genes associated with agronomic traits (Jayakodi et al., 2018; Yang et al., 2021). The large omics data sets have been used to identify several UDP-glycosyltransferases involved in the saponin biosynthesis pathway in *P. ginseng*, *P. notoginseng*, and *P. zingiberensis* (Wang et al., 2015, 2016; Tang et al., 2019; Jiang et al., 2021). In addition, 67 chloroplast genomes of *Panax* species and cultivars were identified in phylogenetic studies (Kim et al., 2015a,b). However, further research is still needed on the relations among *Panax* species and phylogenetic relations resulting from complex population differentiation. Although much of the *Panax* omics data have been reported in dispersive publications and databases, the information utilization is inefficient.

A *Panax* comprehensive database is needed to provide better support and for research to progress more effectively on cultivation techniques and mechanisms affecting the quality of pharmacological compounds. In this study, PanaxGDB was constructed by using MySQL, ThinkPHP, and FastAdmin (<http://panaxGDB.ynau.edu.cn>; **Figure 1**). The database includes

geographical distribution, morphology, and cultivar information on 16 *Panax* species. A section on the compounds of those species focuses on the distributions and characteristics of nearly 600 compounds as well as their metabolic pathways. For the published *P. notoginseng* genomes, the database also integrates the draft genomic sequence with structural and functional annotations and calculates the gene expression levels based on well-organized transcriptomes sampled from different tissues, growth stages, and treatments. The resequencing results of *P. notoginseng* were also uploaded on the Population page to perform genome-wide association analysis (GWAS) on agronomic traits associated with the yield of *P. notoginseng*. In addition, the powerful tools Blast and JBrowse are used to perform functional analysis.

MATERIALS AND METHODS

Omics, Metabolome, and Germplasm Data Sets

The omics data set includes transcriptome sequencing, genomes, metagenomes, and chloroplast genomes. The bioinformatics of the data sets was summarized, and a list was formed for easy searching. Germplasm data include species and cultivar information. Information was also collected on more than 500 compounds from 13 *Panax* species by searching the literature on phytochemistry between 1963 and 2021. The above information and that on chemical compositions were manually collected from publications, online resources, and other literature.

Assembly and Gene Models of Genome Data

Genomic data sets primarily contain genome sequences, gene sequences, protein sequences, gff, gene functional annotations, and transcription factor (TF) annotations. Gene sequence and gff files of *P. ginseng* in the database were obtained from the ginseng genomic database (<http://ginsengdb.snu.ac.kr/data.php>). The 5th *P. notoginseng* genome assembly composed of 2.41 Gb from a previous study was included (Yang et al., 2021). Genome assemblies of *P. ginseng* and *P. notoginseng* included 59,352 and 47,870 genes, respectively. Genes were functionally annotated in InterPro (protein family; <http://www.ebi.ac.uk/interpro/>), pfam (protein family; <http://pfam.xfam.org/>), GO (Gene Ontology; <http://geneontology.org/>), NCBI non-redundant protein (Nr) database (<http://www.ncbi.nlm.nih.gov/>), and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg>) by using BLASTX (e-value < 10⁻⁵; <http://www.ncbi.nlm.nih.gov/BLAST/>). The transcription factors were identified with PlantTFDB serve (<http://planttfdb.gao-lab.org/>).

Digital Gene Expression Profiles

RNA-Seq data sets of *P. ginseng* and *P. notoginseng* with biological duplication were selected. Data sets included different tissues, development times, and treatments. Filtered RNA-seq was mapped to the latest reference genome by HISAT2 (Zhang et al., 2021). Matched reads were subsequently presented to

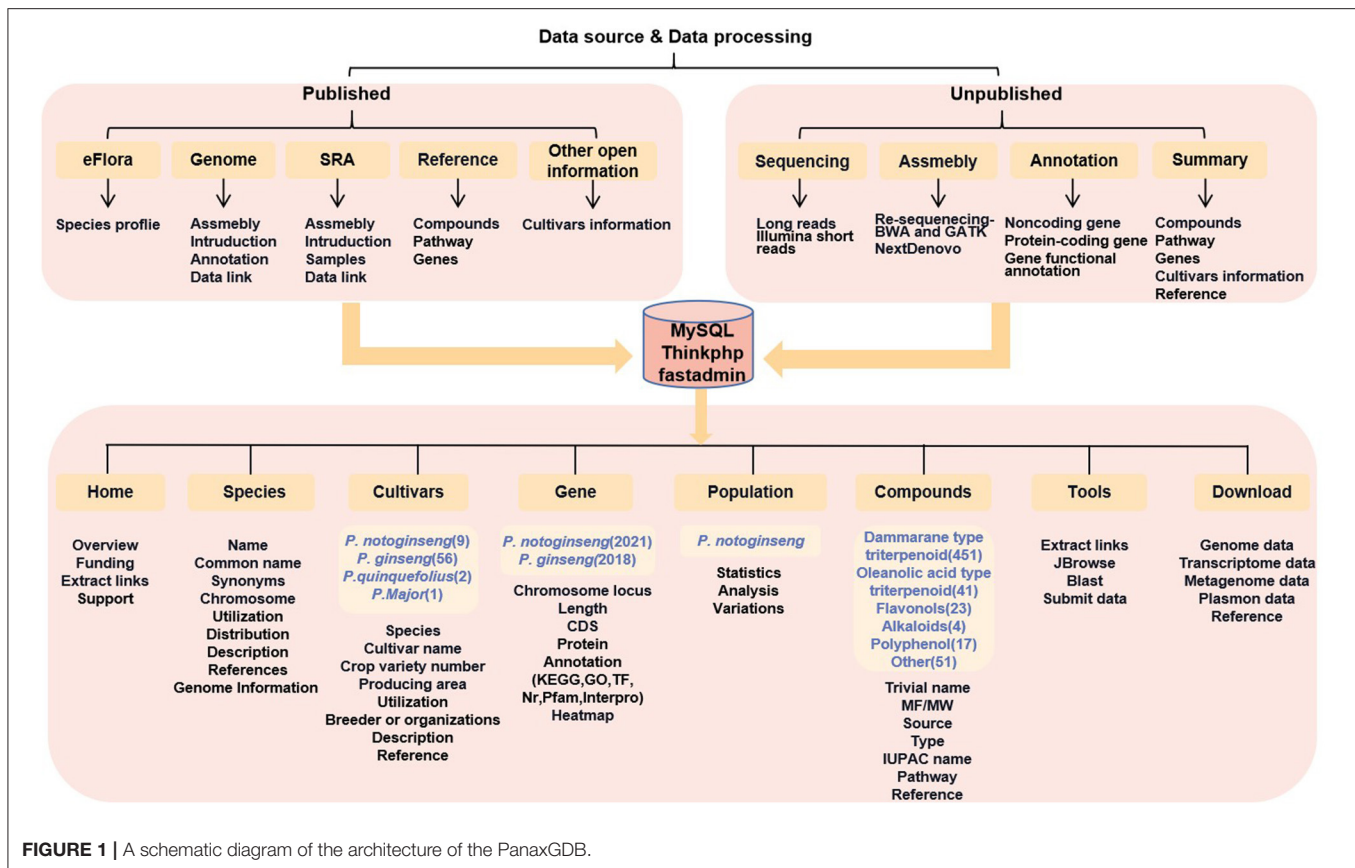


FIGURE 1 | A schematic diagram of the architecture of the PanaxGDB.

StringTie for *de novo* assembly, during which the expression level of each gene and the isoform was estimated (Pertea et al., 2015). Gene expression maps were prepared by manual selection on the Gene page.

Resequencing of *Panax notoginseng*

FastQC software (v0.11.5) was used for quality control of raw reads from Illumina HiSeq™2000 (WuHang, HuNan, China) sequencing, and clean reads were mapped to the latest reference genome using the BWA software (v0.7.15-R1140; Li and Durbin, 2009; De Sena Brandine and Smith, 2019). The GATK software (v4.0.9.0) was used for single nucleotide polymorphisms (SNPs) identification (Tajima, 1989). The ANNOVAR software was used to calculate the number of mutation types of SNP sites in each sample (Wang et al., 2010). The software LUMPY (v0.2.13) was used to detect structural variation (SV) in this project (Layer et al., 2014). PCA of the sample group structure was performed by using PLINK (v1.9b6) and EIGENSOFT (EIG-6.1.4) software (Patterson et al., 2006; Price et al., 2006; Purcell et al., 2007). The software ADMIXTURE (v1.3.0; Alexander et al., 2009) was used to analyze population structure. The software SNPhylo 2 (Lee et al., 2014) was used to construct evolutionary relations between different samples, and Phylogeny.IO (<https://github.com/oist/phylogeny-io>) was used to construct the phylogenetic tree. A graph of linkage disequilibrium was obtained by PopLDdecay (v3.40) analysis (Zhang et al., 2019). GCTA (v1.93.2) software

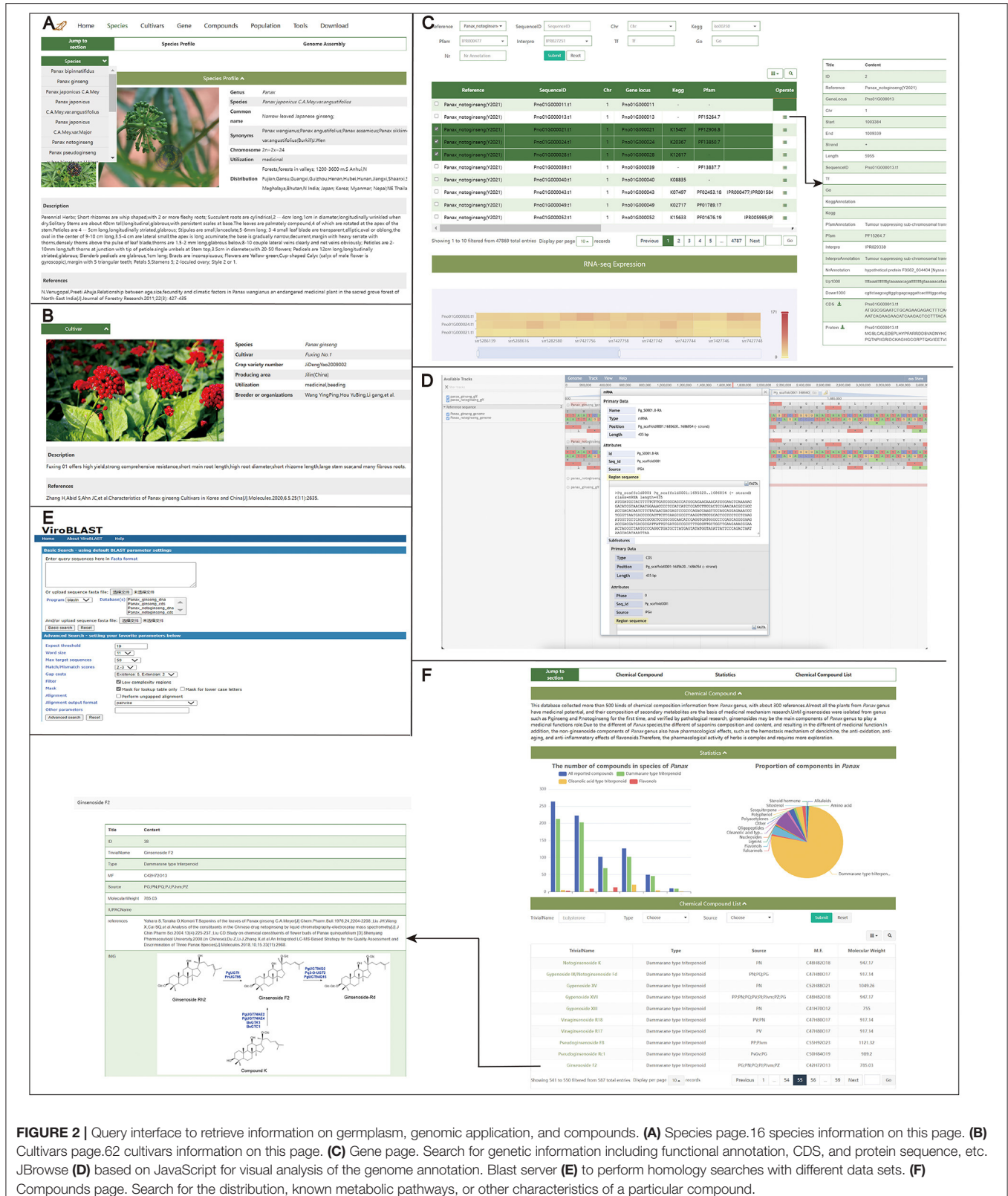
was performed for GWAS analysis, and a mixed linear model (MLM) was adopted as a modeling method (Yang et al., 2011).

Transcriptome of *Panax notoginseng*

A total of 20.3 G of PacBio Iso-Seq data and 92.6 G of raw RNA-seq data of three representative tissues (roots, stems, leaves) were acquired from *P. notoginseng* sequencing. The RNA from the three tissues was mixed to establish two SMRT cells for SMRT sequencing. Smrtlink 5.0 software was used to process the raw data, and the obtained redundant sequences were clustered by an iterative clustering for error correction algorithm to obtain full-length consensus sequences (FL; Pertea et al., 2015). Quiver was used to align FL sequences and non-FL reads. The high-quality sequence with accuracy >0.99 obtained in the previous step was further aligned with the assembled data from Illumina RNA-seq by using Lordec (v0.9) and mapped to the reference genome by using GMAP software (Wu and Watanabe, 2005).

Database Implementation

PanaxGDB is based on the Apache web server (<http://www.apache.org>), adopted ThinkPHP5.1 (<http://www.thinkphp.cn>)-based FastAdmin templates, included frameworks of CodeIgniter (<https://www.codeigniter.com/>) and Bootstrap (<https://getbootstrap.com>), and applied programming languages including CSS, PHP-HTML5, and JavaScript. MySQL (<https://>



www.mysql.com) was used for data sorting, storage, and management, and the AJAX asynchronous loading scheme was used for quick data loading and function implementation.

To provide an interactive user experience, Echarts (https://echarts.apache.org/zh/index.html), JBrowse (http://jbrowse.org), Phylogeny.IO (https://github.com/oist/phylogeny-io),

and JQuery (<https://jquery.com>) were applied. The File Transfer Protocol-based download function was also provided with a transfer speed of up to 50 Mbps. Additionally, all functions in PanaxGDB were tested for web browsing on Safari, Chrome, Firefox, IE, and Edge by mobile phone, pad, and computer.

RESULTS

Database Structure

PanaxGDB with a user-friendly web interface includes nine pages: Home, Species, Cultivars, Gene, Compounds, Population, Tools, and Download (Figure 1). The Home page presents an overview of PanaxGDB as well as links to popular ginseng sites and biological tools. Other pages represent the five main topics of PanaxGDB.

Germplasm of *Panax*

Two modules of the germplasm resources of *Panax* are included in PanaxGDB: Species (Figure 2A) and Cultivars (Figure 2B). These two pages can help users quickly understand the species structure of *Panax* and the breeding status of *Panax* cultivars, respectively. To determine species, authoritative plant taxonomic keys were used. Species not authenticated were identified according to morphological and molecular data in various public references, and species of uncertain status were indicated. *Panax assamicus* found in northeast India, which was not accepted by the Board of Trustees of the Royal Botanic Gardens, was cited as *P. bipinnatifidus* var. *angustifolius*. Similarly, *Panax shangianus* and *Panax omeiensis* were not included in the authoritative source for plant nomenclature. Although *Panax vietnamensis* var. *fuscidiscus* and *P. vietnamensis* var. *langbianensis* were not included in authoritative plant taxonomies, morphological and molecular evidence indicate they are new varieties (Van Duy et al., 2016).

Sixteen *Panax* species are summarized in the database, and eight categories of information are provided for each species on the Species page, including pictures, common name, synonyms, chromosome number, utilization, distribution, description, and references (Figure 2A). In China, local provincial authorities register and identify cultivars of medicinal plants, and as result, information on various cultivars is not concentrated. Therefore, enterprises and producers involved in *Panax* cultivation are limited in the collection of cultivar information, which is difficult to meet market demand. On the Cultivars page, those participating in the Panax industry can easily obtain the current breeding status of Panax cultivars to screen optimal cultivars for different products under local environmental conditions. The Cultivars page contains eight categories of information on cultivars, including crop variety number, uses, area of production, description, pictures, breeders or organizations, and references (Figure 2B). A total of fifty *P. ginseng* cultivars are recorded in PanaxGDB, including 19 from China, 30 from Korea, and one from Japan, as well as two cultivars of *P. quinquefolius*, one of *P. japonicus*, and nine of *P. notoginseng* from China.

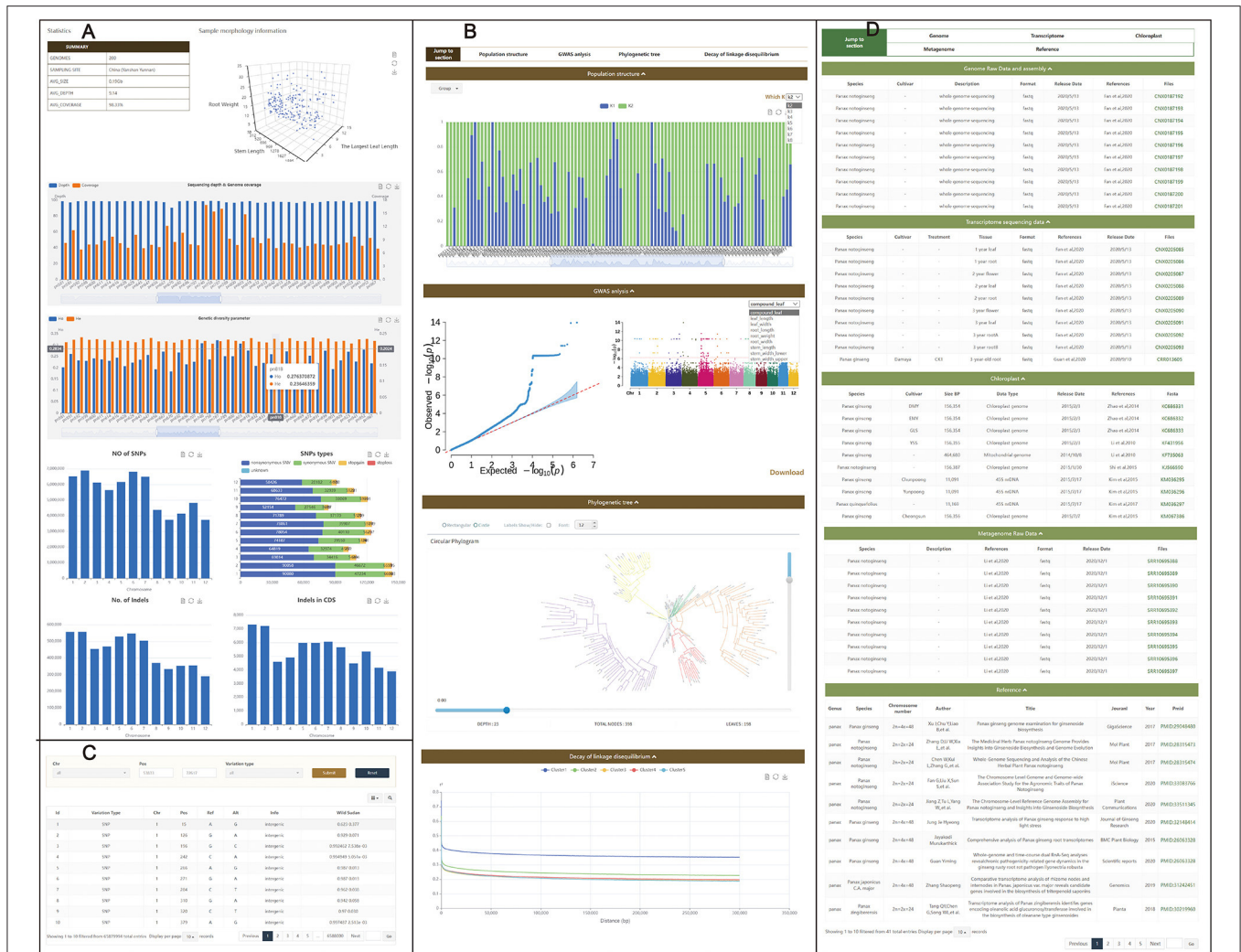
Genetic Information

PanaxGDB provides viewing of the latest assemblies of *Panax* genomes (*P. ginseng* and *P. notoginseng*). Genome assembly files of *P. notoginseng* are open to the public for research for the first time. For convenience and practicality, genome assemblies were re-annotated and uploaded. On the Gene page, users can enter a TF, GO, and KEGG orthology identifiers (IDs), Pfam domain, InterPro domain, or gene ID in different search boxes to obtain detailed corresponding information, such as structural and functional annotation of genes of interest. For each gene, 22 categories of information are provided, such as gene locus, length, and downloadable sequence, among others, which can be displayed using the “Operate” option in the upper-right of the display list (Figure 2C). Users can also access the ginseng genome databases from the Home page. In the gene expression module, one or more genes can be selected according to specific needs, and gene expression levels of different groups or treatments can be visualized from a heat map of ticked, related genes. Heat maps provide information on data accession numbers and associated reference (Figure 2C).

On the Tool page, JBrowse and Blast can help users visualize *Panax* genomes and sequence alignments. Currently, *P. ginseng* and *P. notoginseng* genomes, gene annotations, and information on *Panax* variants can be imported into JBrowse. Sets of genomes, gene and density information on SNP/300 kb, and allele frequency can be explored in PanaxGDB in JBrowse (Figure 2D). The Blast tool is provided for sequence alignment with Sequence coding for amino acids in protein (CDS) and protein sequences of *P. ginseng* and *P. notoginseng* and uses the input of a similar sequence (Figure 2E). The final Blast result can be downloaded in HTML format.

Metabolites of *Panax*

Secondary metabolites in *Panax* are pharmacologically active. Therefore, a metabolites page was provided with an interface (Compounds page) to retrieve, visualize, and investigate metabolite distribution in 12 *Panax* species (Figure 2F). The page includes 587 compounds of *Panax*, of which 492 are saponins. Damarane-type saponins accounted for 76.8% (451) of the known compounds. More than 200 saponins have been detected in *P. ginseng* and *P. notoginseng* according to the available phytochemical and pharmacological literature. In the Statistics modules, the proportion, quantity, and distribution of different types of compounds across different species can be visualized. In the list of chemical compounds, users can find ginsenoside Rb1, Rb2, Rc, Rd, Rb3, Re, Rg1, 20(S)-Rg2, Rh1, Rf, and Ro, notoginsenoside R1, and gypenoside XVII in at least eight *Panax* species. More than two *Panax* species contain 148 types of saponins. However, some characteristic saponins, such as quinquenoside I, II, III, IV, and V, are found only in *P. quinquefolium* (Yoshikawa et al., 1998). According to statistics, characteristic saponins are found in *P. ginseng*, *P. notoginseng*, *P. quinquefolius*, *P. japonicus*, *P. japonicus* var. *major*, *P. vietnamensis*, *P. stipuleanatus*, and *P. bipinnatifidus*. However, the list might be incomplete because of insufficient detection depth. To search for additional information, users can search by entering the name of a specific compound, compound



type, or species origin. Users can click on a metabolite to link the metabolite information subpage and complete metabolite information, including biochemistry, biosynthesis pathway, and its species origin. Another hidden effect of this page is to help understand validated biosynthetic pathways of compounds, especially saponins, and validated genes in those pathways.

Population Structure

The resequencing database was set for *Panax* breeding and germplasm resource management. Data from *P. notoginseng* were collected on the Population page, and 200 3-year-old *P. notoginseng* samples from Yunnan, China, with nine important agricultural traits were randomly selected for resequencing. The average sequencing size of each sample was 28.83 Gb. Results were visualized and uploaded to the Population page with three modules which included Statistics (Figure 3A), Analysis (Figure 3B), and Variations (Figure 3C). Statistical results of the resequencing can be seen on the Statistics subpage. The Analysis

subpage includes population structure, phylogenetic tree, decay of linkage disequilibrium, and Manhattan figure. All images and corresponding data on this page can be downloaded. Users can also search for SNPs on the Variations subpage based on results of phenotype evaluation and GWAS.

A total of 1,615,492,994 SNPs were obtained in 200 samples. There were 1,462,055,399 SNPs of 200 samples within intergenic regions, 114,939,464 within intronic regions, and 38,498,131 within exons. These data can be visualized on the Statistics subpage. On the Analysis subpage, the phylogenetic tree was derived from the SNP data, and the tree shows *P. notoginseng* of Wenshan and can be divided into five subgroups. In the GWAS module, there were 192 highly significant peaks on 12 pseudochromosomes associated with traits (root weight and leaf width) related to saponin yield. Those SNPs can be found on the Variations pages, and 22 occurred in CDSs (Supplementary Table 1), which can be found on the Gene page. Those genes were

annotated as proteins associated with cell metabolism and growth, including peptidyl-prolyl cis-trans isomerase, aminotransferase, oxoglutarate dioxygenase, 2-isopropylmalate synthase, WD40-repeat protein, AMP-dependent synthetase, and phosphatidylinositol N-acetylglucosaminyltransferase, protein kinases, or as plant stress-related proteins, including heat shock protein (**Supplementary Table 1**).

Omics Data Acquisition and Upload

To make it more convenient to use *Panax* omics data, all publicly available data were gathered and classified. The Download page contains information and links to all available public *Panax* omics data, and long-read sequencing and resequencing data were published for the first time in PanaxGDB (**Figure 3D**). Downloadable content in this module includes 405 transcriptome sequencings, two genomes, 21 metagenomes, and 67 chloroplast genomes. The visual data can help users to summarize information. Because the database is an open platform, prospective ginseng omics data are welcomed and can be included in the database by contacting the authors through the Data Submission page.

DISCUSSION

An early ginseng genome database was limited to the genome and transcriptome of *P. ginseng*, and it was not possible to visually compare expression levels of more than two genes (Jayakodi et al., 2018). In PanaxGDB, additional transcriptome data from different tissues and experiments were re-assembled, and information on those data, including sources and references, can be intuitively visualized, including comparisons of expression levels of different genes. PanaxGDB provides the most comprehensive overview of the omics, metabolome, and germplasm resources of *Panax* and forms a user-friendly interface. In addition, the most complete analysis of resequencing and annotation results of the 5th version of the *P. notoginseng* genome and the full-length RNA-seq of *P. notoginseng*, as well as *Panax* cultivars, are fully displayed.

PanaxGDB is introduced as a comprehensive database with the first highly integrated multidirectional research results of the *Panax* genus to improve the practical use of data. PanaxGDB can be an important resource for those working with *Panax* in comparative studies of major cultivars, in the development of pharmacological value, and in heterogeneous biosynthetic research on saponins. As an open platform, advanced users are welcome to submit data related to *Panax* species to regularly update the database. Thus, PanaxGDB can continuously serve as

a central portal to boost research on the biology and chemistry of *Panax*.

DATA AVAILABILITY STATEMENT

All data of this study are available through the PanaxGDB database website (<http://panaxGDB.yнау.edu.cn>).

AUTHOR CONTRIBUTIONS

XuL and SY: conception and design of the study. YLin, YLi, ZY, GX, YLu, QZ, XueL, GL, and BH: acquisition of data or analysis and interpretation of data. YLin, XuL, YD, GZ, and QY: database construction and visualization. YLin, BH, XuL, and SY: drafting of the article or revising it critically for important intellectual content. YLin, BH, YLu, YLi, YD, ZY, GX, GL, XueL, QZ, QY, GZ, SY, and XuL: final approval of the version to be published.

FUNDING

This work was supported by National Key R & D Plan (2017YFC1702500), Major Science and Technology Projects in Yunnan Province (2019ZF011-1), Science and Technology Innovation team of Yunnan (202105AE160011), the First Projects of Science and Technology Plan in the Biomedical field in 2021 (202102AA310048), and the Digitalization of Biological Resource Project (202002AA1000051), Guangxi Innovation Driven Development Project (GuiKe AA18242040).

ACKNOWLEDGMENTS

We are thankful to the co-authors for their contributions to this project, the Key Laboratory of Medicinal Plant Biology, and the College of Biological Big Data for essential research facilities support. We also acknowledge the National Key R & D Plan, the Major Science and Technique Programs in Yunnan Province, the First Projects of Science and Technology Plan in the Biomedical field in 2021, and the Digitalization of Biological Resource Project for financial support for this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.883818/full#supplementary-material>

REFERENCES

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Beedata (2020). *Report on Data Analysis of China's Sanqi Industries in 2020*. Available online at: <https://finance.sina.com.cn/money/future/agri/2020-10-16/doc-iiznezxr6369582.shtml> (accessed December 24, 2021).
- Chen, R. J., Chung, T.-Y., Li, F.-Y., Lin, N.-H., and Tzen, J. T. (2009). Effect of sugar positions in ginsenosides and their inhibitory potency on Na⁺/K⁺-ATPase activity. *Acta Pharmacol. Sin.* 30, 61–69. doi: 10.1038/aps.2008.6
- Chen, W., Balan, P., and Popovich, D. G. (2019). Review of ginseng anti-diabetic studies. *Molecules* 24:4501. doi: 10.3390/molecules24244501
- De Sena Brandine, G., and Smith, A. D. (2019). Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res.* 8:1874. doi: 10.12688/f1000research.21142.1
- Ding, S., Wang, M., Fang, S., Xu, H., Fan, H., Tian, Y., et al. (2018). D-dencichine regulates thrombopoiesis by promoting megakaryocyte

- adhesion, migration and proplatelet formation. *Front. Pharmacol.* 9:297. doi: 10.3389/fphar.2018.00297
- Ibaogao (2021). *Report of Produce Sale Demand and Investment Forecast on China Ginseng Industry (2001-2026) [Preprint]*. Available online at: <https://bg.qianzhan.com/report/detail/300/200731-1fb56d43.html> (accessed December 28, 2021).
- Jayakodi, M., Choi, B. S., Lee, S. C., Kim, N. H., Park, J. Y., Jang, W., et al. (2018). Ginseng Genome Database: an open-access platform for genomics of *Panax ginseng*. *BMC Plant Biol.* 18:62. doi: 10.1186/s12870-018-1282-9
- Jiang, Z., Tu, L., Yang, W., Zhang, Y., Hu, T., Ma, B., et al. (2021). The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. *Plant Commun.* 2:100113. doi: 10.1016/j.xplc.2020.100113
- Kim, K., Lee, S. C., Lee, J., Lee, H. O., Joh, H. J., Kim, N. H., et al. (2015a). Comprehensive survey of genetic diversity in chloroplast genomes and 45S rDNAs within *Panax ginseng* species. *PLoS ONE*. 10:e0117159. doi: 10.1371/journal.pone.0117159
- Kim, K., Lee, S. C., Lee, J., Yu, Y., Yang, K., Choi, B. S., et al. (2015b). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* 5:15655. doi: 10.1038/srep15655
- Layer, R. M., Chiang, C., Quinlan, A. R., and Hall, I. M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 15, 1–19. doi: 10.1186/gb-2014-15-6-r84
- Lee, T. H., Guo, H., Wang, X., Kim, C., and Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*. 15:162. doi: 10.1186/1471-2164-15-162
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Liu, L., Xu, F. R., and Wang, Y. Z. (2020). Traditional uses, chemical diversity and biological activities of *Panax* L. (Araliaceae): a review. *J. Ethnopharmacol.* 263:112792. doi: 10.1016/j.jep.2020.112792
- Nag, S., Qin, J., Wang, W., Wang, M.-H., Wang, H., and Zhang, R. (2012). Ginsenosides as anticancer agents: *in vitro* and *in vivo* activities, structure-activity relationships, and molecular mechanisms of action. *Front. Pharmacol.* 3:25. doi: 10.3389/fphar.2012.00025
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2:e190. doi: 10.1371/journal.pgen.0020190
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595. doi: 10.1093/genetics/123.3.585
- Tang, Q. Y., Chen, G., Song, W. L., Fan, W., Wei, K. H., He, S. M., et al. (2019). Transcriptome analysis of *Panax zingiberensis* identifies genes encoding oleanolic acid glucuronosyltransferase involved in the biosynthesis of oleanane-type ginsenosides. *Planta* 249, 393–406. doi: 10.1007/s00425-018-2995-6
- Van Duy, N., Le Ngoc, T., Chinh, N. D., and Van Tien, T. (2016). A new variety of *Panax* (Araliaceae) from Lam Vien Plateau, Vietnam and its molecular evidence. *Phytotaxa* 277, 47–58. doi: 10.11646/phytotaxa.277.1.4
- Wang, D., Wang, J., Shi, Y., Li, R., Fan, F., Huang, Y., et al. (2020a). Elucidation of the complete biosynthetic pathway of the main triterpene glycosylation products of *Panax notoginseng* using a synthetic biology platform. *Metab. Eng.* 61, 131–140. doi: 10.1016/j.ymben.2020.05.007
- Wang, H., Yang, Y., Lin, L., Zhou, W., Liu, M., Cheng, K., et al. (2016). Engineering *Saccharomyces cerevisiae* with the deletion of endogenous glucosidases for the production of flavonoid glucosides. *Microb. Cell Fact.* 15:134. doi: 10.1186/s12934-016-0535-2
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603
- Wang, L., Huang, Y., Yin, G., Wang, J., Wang, P., Chen, Z. Y., et al. (2020b). Antimicrobial activities of Asian ginseng, American ginseng, and notoginseng. *Phytother. Res.* 34, 1226–1236. doi: 10.1002/ptr.6605
- Wang, P., Wei, Y., Fan, Y., Liu, Q., Wei, W., Yang, C., et al. (2015). Production of bioactive ginsenosides Rh2 and Rg3 by metabolically engineered yeasts. *Metab. Eng.* 29, 97–105. doi: 10.1016/j.ymben.2015.03.003
- Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi: 10.1093/bioinformatics/bti310
- Xu, C., Wang, W., Wang, B., Zhang, T., Cui, X., Pu, Y., et al. (2019). Analytical methods and biological activities of *Panax notoginseng* saponins: recent trends. *J. Ethnopharmacol.* 236, 443–465. doi: 10.1016/j.jep.2019.02.035
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, W. Z., Hu, Y., Wu, W. Y., Ye, M., and Guo, D. A. (2014). Saponins in the genus *Panax* L. (Araliaceae): a systematic review of their chemical diversity. *Phytochemistry* 106, 7–24. doi: 10.1016/j.phytochem.2014.07.012
- Yang, Z., Liu, G., Zhang, G., Yan, J., Dong, Y., Lu, Y., et al. (2021). The chromosome-scale high-quality genome assembly of *Panax notoginseng* provides insight into dencichine biosynthesis. *Plant Biotechnol. J.* 19, 869–871. doi: 10.1111/pbi.13558
- Yoshikawa, M., Murakami, T., Yashiro, K., Yamahara, J., Matsuda, H., Saijoh, R., et al. (1998). Bioactive saponins and glycosides. XI. Structures of new dammarane-type triterpene oligoglycosides, quinquenosides I, II, III, IV, and V, from American ginseng, the roots of *Panax quinquefolium* L. *Chem. Pharm. Bull.* 46, 647–654. doi: 10.1248/cpb.46.647
- Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875
- Zhang, H., Abid, S., Ahn, J. C., Mathiyalagan, R., Kim, Y. J., Yang, D. C., et al. (2020). Characteristics of *Panax ginseng* Cultivars in Korea and China. *Molecules* 25:2635. doi: 10.3390/molecules25112635
- Zhang, Y., Park, C., Bennett, C., Thornton, M., and Kim, D. (2021). Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.* 31, 1290–1295. doi: 10.1101/gr.275193.120
- Zhou, M., Yang, G., Sun, G., Guo, Z., Gong, X., and Pan, Y. (2020). Resolving complicated relationships of the *Panax bipinnatifidus* complex in southwestern China by RAD-seq data. *Mol. Phylogenet. Evol.* 149:106851. doi: 10.1016/j.ympev.2020.106851
- Zhu, S., Zou, K., Fushimi, H., Cai, S., and Komatsu, K. (2004). Comparative study on triterpene saponins of Ginseng drugs. *Planta Med.* 70, 666–677. doi: 10.1055/s-2004-827192
- Zuo, Y. J., Wen, J., and Zhou, S. L. (2017). Intercontinental and intracontinental biogeography of the eastern Asian - Eastern North American disjunct *Panax* (the ginseng genus, Araliaceae), emphasizing its diversification processes in eastern Asia. *Mol. Phylogenet. Evol.* 117, 60–74. doi: 10.1016/j.ympev.2017.06.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lin, Hao, Lu, Dong, Li, Zhang, Yang, Xiang, Liu, Li, Zhu, Yang, Li and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.