**Brief Communications**

**The chromosome-scale high-quality genome assembly of *Panax notoginseng* provides insight into dencichine biosynthesis**

Zijiang Yang1,†, Guanze Liu1,†, Guanghui Zhang1,†, Jing Yan1, Yang Dong1,2, Yingchun Lu1, Wei Fan1, Bing Hao1, Yuan Lin1, Ying Li1, Xuejiao Li1, Qingyan Tang1, Guisheng Xiang1, Simei He1, Junwen Chen1, Wei Chen1,2, Zhongping Xu2, Zichao Mao2, Shengchang Duan2,5, Shuangxia Jin3,† and Shengchao Yang1,†

1State Key Laboratory of Conservation and Utilization of Bio-Resources in Yunnan, The Key Laboratory of Medicinal Plant Biology of Yunnan Province, National and Local Joint Engineering Research Center on Germplasm Innovation and Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University, Kunming, China

2Yunnan Plateau Characteristic Agriculture Industry Research Institute, Kunming, China

3National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

4College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China

5NOWBIO Technology Co. Ltd, Kunming, China

The chromosome-scale high-quality genome assembly of *Panax notoginseng* (2n = 2x = 24), reputed as a valuable traditional Chinese medicine (Figure 1a). *P. notoginseng* has been used as a traditional Chinese medicine with obvious efficacy and favourable safety. Currently, *P. notoginseng* is an important material of many well-known Chinese patent medicines including Xuesaitong, Yunnan Baiyao and Compound Danshen Dripping Pills (Xu et al., 2019). Triterpenoid saponins (TSs) and dencichine are the main bioactive compounds in *P. notoginseng*. The therapeutic effects of TSs include anti-inflammatory, vasorelaxation and anticancer. Dencichine is used for treatment of injury-induced trauma, and its haemostatic function was proven by clinical practice (Ng, 2006).

Four versions of *P. notoginseng* genome assembly have been reported (Chen et al., 2017; Fan et al., 2020; Jiang et al., 2020; Zhang et al., 2017). The first two assemblies are highly fragmented, which limits genomic studies of *P. notoginseng*. The continuity was significantly improved in Fan’s and Jiang’s assembly (Figure 1b). The genome size of *P. notoginseng* has been estimated to be about 2.31 Gb using flow cytometry. While, the size of four assemblies varied between 1.85 to 2.66 Gb. Here, we present a chromosome-scale assembly for *P. notoginseng* created using PacBio long reads, Illumina short reads and Hi-C technology. The heterozygosity of *P. notoginseng* was estimated to be 0.23%, indicating its low genetic variability.

The size and contig N50 are 2.41 Gb and 1.45 Mb, respectively (Figure 1b). Approximately 87% of the contigs were anchored into 12 pseudo-chromosomes. The result showed that 2,219 out of 2,326 (95.4%) eudicots BUSCOs were found to be complete in the assembly with a decent redundancy (9.8% complete and duplicated BUSCOs) (Figure 1c). Our assembly is more complete compared with Fan’s assembly (89.3% complete BUSCOs). Since the data of Jiang’s assembly are not currently available (to be released in September, 2021), only genome BUSCO result with 2,121 BUSCOs data set was presented. The percentage of complete BUSCOs in Jiang’s assembly is 96.6%, yet the complete and duplicated BUSCOs are relatively high (26.1%) for a diploid species with low heterozygosity. Considering the assembled size of Jiang’s assembly is 15.2% percent larger than the estimated genome size, Jiang’s assembly might contain a large portion of differentiated haplotype sequences. We also aligned ~68X Illumina short reads and 266,984 RNA transcripts to our assembly and found 98.4% coverage rate and 96.0% mapping rate, respectively. Specifically, the depth distribution of mapped Illumina reads showed no anomaly (Figure 1d).

Annotation revealed 88.2% of the genome as repetitive elements. Transposable elements (TEs) were identified as the most abundant, spanning 87.8% of the assembly. Among the subfamilies of TEs, long terminal repeats (LTRs) took up 81.2% of the assembly. Insertion time analysis of 24,861 intact LTRs revealed *P. notoginseng* experienced a burst increase of LTRs started 1 million years ago (Mya) (Figure 1e). The LTR assembly index (LAI) reached 12.44, validating the high quality of our assembly. A total of 47,870 genes were predicted in *P. notoginseng*. Compared with previous studies, more genes were predicted in *P. notoginseng* (Figure 1b). Comparison of gene set BUSCO results showed our gene set as the most complete (94% complete BUSCOs) (Figure 1c), which might explain the higher number of predicted genes in our assembly. Among the predicted genes, 98.51% were supported by RNA-seq data and 94.61% could be annotated in public database. The gene number was smaller than the closely related species *Panax*...
ginseng (59,352 genes), probably due to the tetraploid nature of *P. ginseng* (Kim et al., 2018).

Gene family analysis of *P. notoginseng* and 11 other angiosperms suggest *P. notoginseng* genes were clustered into 17,306 families and *P. notoginseng* had much less multiple-copy orthologs compared with *P. ginseng* (Figure 1f). Phylogenetic tree based on single-copy genes suggest Panax genus diverged from the Apiaceae species Daucus carota.
44.7–65.0 Mya, and divergence of *P. notoginseng* and *P. ginseng* is between 6.0 - 17.1 Mya (Figure 1g). Chromosome synteny analysis of Fan’s assembly with ours showed many discontinuities and segmental inversions (Figure 1h), where most of these anomalies fell into TE-rich regions (Figure 1d). This suggests limitations of current technologies in assembling highly repetitive plant genomes.

Three key enzyme families are involved in biosynthesis of TSs: oxidosqualene cyclases (OSC), cytochrome P450 (CYPs) and glycoltransferases (GTs). Dammarenediol-II synthase (DDS) from OSCs family catalyses the cyclization of 2,3-oxidosqualene, forming the triterpene scaffolds dammarenediol-II. Then, dammarenediol-II was modified by CYPs and GTs through hydroxylation and glycosylation of certain positions (mainly C-3, C-6 and C-20).

Depending on whether C-6 contains a hydroxyl group, TSs are divided into protopanaxadiol saponins (PDS) and protopanaxatriol saponins (PTS) (Figure 1i part1). Functional studies revealed that CYP716A47 and CYP716A53v2 are responsible for biosynthesis of PDS and PTS, respectively (Kim et al., 2015). DDS, CYP716A47 and CYP716A53v2 were all identified in *P. notoginseng* genome. Specifically, PnDDS1 and PnDDS2 were derived from proximal duplication (separated by two genes on chromosome 3). Unlike *P. ginseng*, PTS are abundant in roots while scarce in leaves in *P. notoginseng* (Figure 1i, part 3). RNA expression of key genes was investigated to unveil the mechanism of tissue-specific PTS distribution. No tissue-specific expression patterns were found for DDS and CYP716A47, whereas the expression level of CYP716A53v2 was significantly higher in roots than in leaves (Figure 1i part2). This suggests the differential expression of CYP716A53v2 in leaves and roots is responsible for the imbalance PTS distribution.

Dencichine is a non-protein amino acid present in *Panax*, *Lathurus* and several other species. In *Lathurus sativus*, the biosynthesis of dencichine involves L-serine, which is transformed into O-acetyl-L-serine via serine acetyltransferase (SAT). β-cyanoalanine synthase (β-CAS) catalyses the formation of β-(isoxazolin-5-on-2-yl)-L-alanine (BIA) using O-acetyl-L-serine and isoxazolin-5-one. BIA is proposed to be converted into L-2,3-Diaminopropanoic acid (L-Dap). Finally, dencichine is synthesized from L-Dap by enzymes from BAHD acyltransferase family. (Figure 1j, part1) (Yan et al., 2006). The intermediates isoxazolin-5-one and BIA were not detected in *P. notoginseng* indicating the mechanism of dencichine biosynthesis might be different from *L. sativus*. In *Staphylococcus aureus*, two enzymes from Class II PLP-dependent enzymes (PALP) and ornithine cyclodeaminase (OCD) family could produce L-Dap using O-phospho-L-serine (Kobylyarz et al., 2014). Based on these findings, we proposed a novel biosynthetic pathway for dencichine in *P. notoginseng* involving five different type of enzymes: 3-phosphoserine aminotransferase (PSAT), PALP, OCD, acyl activating enzyme 3 (AAE3) and BAHD. Candidate genes for these five enzymes were identified in our study. Notably, dencichine is found to be more abundant in rhizome, fibril and root of *P. notoginseng* whereas less in leaves (Figure 1j, part3), which is in accordance with expression profiles of candidate genes (Figure 1j, part2). Additionally, we measured the expression levels of several candidate genes and found no tissue-specific expression patterns using real-time quantitative PCR (Figure 1j, part4).

In conclusion, this high-quality genome assembly of *P. notoginseng* provides novel insights into unique saponins distribution pattern and reveals possible dencichine biosynthetic pathway.

**Acknowledgements**

This work was supported by the National Key Research and Development Plan (No. 2017YFC1702500), the National Natural Science Foundation of China (No. 81760691), the Major Science and Technique Programs in Yunnan Province (No. 2016ZFK01), and the Digitalization of Biological Resource Project of Yunnan Province (No. 2020002AA100007).

**Author contributions**

S. Y. and S. J. conceived the study. Z. Y., G. L., G. Z., J. Y., Y. D., Y. L., W. F., B. H., Y. L., Y. L., Q. T., G. X., S. H., J. C., W. C., Z. X., Z. M. and S. D. performed the experiments and carried out the analysis. Z. Y., G. L., G. Z., S. Y. and S. J. designed the experiments and wrote the manuscript. All the authors approved the manuscript.

**Conflict of interest**

No conflict of interest declared.

**Data availability**

Raw sequencing data have been deposited in NCBI under BioProject number PRJNA608068. Genome assembly and annotation have been deposited in Herbal Medicine Omics Database (http://herbalplant.ynau.edu.cn/).

**References**

Chen, W., Kui, L., Zhang, G.H., Zhu, S.S., Zhang, J., Wang, X. and Yang, M. et al. (2017) Whole-genome sequencing and analysis of the Chinese herbal plant *Panax notoginseng*. Mol. Plant, 10, 899–902.

Fan, G., Liu, X., Sun, S., Shi, C., Du, X., Han, K., Yang, B. et al. (2020) The chromosome level genome and Genome-Wide Association Study for the agronomic traits of *Panax notoginseng*. *Science*, 23, 101538.

Jiang, Z., Tu, L., Yang, W., Zhang, Y., Hu, T., Ma, B., Lu, Y. et al. (2020) The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. *Plant Commun.*, 2, 101113.

Kim, N.-H., Jayakodi, M., Lee, S.-C., Choi, B.-S., Jang, W., Lee, J. and Kim, H.H. et al. (2018) Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant Biotechnol. J.*, 16, 1904–1917.

Kim, Y.J., Zhang, D.B. and Yang, D.C. (2015) Biosynthesis and biotechnological production of ginsenosides. *Biotechnol. Adv.*, 33, 717–735.

Kobylyarz, M.J., Grigg, J.C., Takayama, S.-I., Rai, D.K., Heinrichs, D.E. and Murphy, M.E.P. (2014) Synthesis of L-2,3-diaminopropanoic acid, a siderophore and antibiotic precursor. *Cell Chem. Biol.*, 21, 379–388.

Ng, T.B. (2006) Pharmacological activity of sanchi ginseng (*Panax notoginseng*). *J. Pharm. Pharmacol.*, 58, 1007–1019.

Xu, C., Wang, W., Wang, B., Zhang, T., Cui, X., Pu, Y. and Li, N. (2019) Analytical methods and biological activities of *Panax notoginseng* saponins: recent trends. *J. Ethnopharmacol.*, 236, 443–465.

Yan, Z.-Y., Spencer, P.S., Li, Z.-X., Liang, Y.-M., Wang, Y.-F., Wang, C.-Y. and Li, F.-M. (2006) *Lathurus sativus* (grass pea) and its neurotoxin ODAP. *Phytochemistry*, 67, 107–121.

Zhang, D., Li, W., Xia, E.H., Zhang, Q.J., Liu, Y., Zhang, Y. and Tong, Y. et al. (2017) The medicinal herb *Panax notoginseng* genome provides insights into ginsenoside biosynthesis and genome evolution. *Mol. Plant*, 10, 903–907.