**Taxus yunnanensis** genome offers insights into gymnosperm phylogeny and taxol production

Chi Song1,12,13, Fangfang Fu2,13, Lulu Yang3,4,13, Yan Niu4,13, Zhaoyang Tian4,13, Xiangxiang He4, Xiaoming Yang2, Jie Chen4, Wei Sun1, Tao Wan5,14, Han Zhang6, Yicheng Yang4, Tian Xiao4,13, Komivi Dossa4,7, Xiangxiao Meng1, Fuliang Cao2,14, Yves Van de Peer8,9,10,11,14✉, Guibin Wang2,14✉ & Shilin Chen1,14✉

Taxol, a natural product derived from *Taxus*, is one of the most effective natural anticancer drugs and the biosynthetic pathway of Taxol is the basis of heterologous bio-production. Here, we report a high-quality genome assembly and annotation of *Taxus yunnanensis* based on 10.7 Gb sequences assembled into 12 chromosomes with contig N50 and scaffold N50 of 2.89 Mb and 966.80 Mb, respectively. Phylogenomic analyses show that *T. yunnanensis* is most closely related to *Sequoiadendron giganteum* among the sampled taxa, with an estimated divergence time of 133.4–213.0 MYA. As with most gymnosperms, and unlike most angiosperms, there is no evidence of a recent whole-genome duplication in *T. yunnanensis*. Repetitive sequences, especially long terminal repeat retrotransposons, are prevalent in the *T. yunnanensis* genome, contributing to its large genome size. We further integrated genomic and transcriptomic data to unveil clusters of genes involved in Taxol synthesis, located on the chromosome 12, while gene families encoding hydroxylase in the Taxol pathway exhibited significant expansion. Our study contributes to the further elucidation of gymnosperm relationships and the Taxol biosynthetic pathway.
Taxis species, belonging to the Taxaceae (yews, gymnosperms), are slow-growing, long-lived coniferous trees or shrubs, that have been regarded as endangered Tertiary relict species. *Taxus* are well-known for their cancer-inhibitory alkaloid paclitaxel (Taxol), which is a trace natural product. Taxol is a polyoxynated cyclic diterpenoid, mainly used to treat numerous cancers, including ovarian, breast, lung, cervical, and pancreatic cancer. However, limited content (0.01–0.05 %) and localization of paclitaxel in specific organs (the bark of yew) renders production from natural sources low. Taxol biosynthesis begins with the universal diterpenoid precursor geranylatediphosphate (GGPP), which is then decorated with a series of cytochrome-P450 hydroxylases (CYP450s), acetyltransferases and other enzymes, leading to the end product paclitaxel. The complexity of Taxol biosynthesis has greatly hindered the mass production of Taxol.

To further elucidate the Taxol biosynthesis pathway, we here report a high-quality genome assembly of *T. yunnanensis*. In addition, the genome of *Taxus*, as a first representative of the Taxaceae, might help in unraveling the phylogenetic relationships within gymnosperms. There is much controversy about the evolutionary relationship between different gymnosperms (such as Cycads, Ginkgo, Gnetophytes and Conifers) within gymnosperms. The 1 KP transcriptome dataset provides strong support for Cycads and Gnetophytes sister to the rest of gymnosperms and Gnetophytes sister to, or within, the conifers. Whole-genome sequences, such as the one of *Taxus* presented here, provide an additional dataset to shed light on the elusive evolutionary relationships within gymnosperms.

**Results and discussion**

Based on the k-mer distribution analysis, we estimated the genome size of *T. yunnanensis* to be 10.49 Gb, with a high level of repetition (77.74%) and heterozygosity (0.54%) (Supplementary Fig. 1 and Supplementary Table 1). The genome sequence of *T. yunnanensis* was obtained using Oxford Nanopore high-throughput sequencing systems (85×), Illumina (50×) and high-throughput chromosome conformation capture (Hi-C, 60×) throughput chromosome conformation capture (Hi-C, 60×), Illumina (50×) and high-throughput sequencing systems (85×). The total length of the final assembly was 10.73 Gb with a contig N50 of 2.89 Mb and a scaffold N50 of 966.80 Mb (Table 1 and Supplementary Table 3). A total of 10.63 Gb of the assembly and 98.95% of the genes were distributed across 12 chromosome-level pseudomolecules (Supplementary Table 3 and Supplementary Fig. 2). The completeness of the genome assembly and gene set of *T. yunnanensis* was estimated at 72.6% and 73.7% using BUSCO, which is similar to the available gymnosperm genomes (Supplementary Table 4).

We annotated 34,931 high-quality protein-coding genes, which is slightly lower than for the *S. giganteum* genome (38,000) (Fig. 1 and Table 1). On average, the predicted gene sequence length was 305.46 bp (Table 1). Numerous long introns are a notable characteristic of the *T. yunnanensis* genome. The length distribution for the 10% longest introns in *T. yunnanensis* is from 14,790 bp to 462,177 bp, and average at 35,282 bp. A comparison of gene models for the 14 land plants revealed that the average length of the longest 10% of introns in most of the gymnosperms was longer than that in angiosperms (Supplementary Table 5 and Supplementary Fig. 3).

A total of 7.96 Gb of repetitive elements occupying 74.11% of the *T. yunnanensis* genome were annotated (Supplementary Table 6). Repetitive sequences, especially the long terminal repeat retrotransposons (LTR-RTs), have been deemed to be the major component of all gymnosperm genomes and the main cause of gymnosperm genome expansion. Consistent with other gymnosperm genomes, the majority of the repeats in the *T. yunnanensis* genome are LTR (40.95% of all assembled sequences), of which two super-families, 2,138,065 Ty3/Gypsy and 453,398 Try/Copia (the number of repeats sequences) were identified, accounting for 35.95% and 4.77% of all assembled sequences, respectively (Supplementary Table 6). Based on a mutation rate of 7.34573 × 10^{-10} substitutions per base per year, we found that the insertion for Gypsy and Copia occurred largely between 8−24 and 8−44 million years ago (MYA), respectively (Supplementary Fig. 4a). Since the Gypsy accounted for 87.78% of the total LTR sequences, the insertion of large amounts of Gypsy in 8−24 MYA resulted in genome expansion of *T. yunnanensis*. We identified and characterized full-length LTR in four gymnosperms and three angiosperms (*T. yunnanensis*, *Gnetum montanum*, *Ginkgo biloba*, *S. giganteum*, *Ampelopsis trifchopoda*, *Oryza sativa* and Arabidopsis thaliana), the number of LTRs contained in gymnosperms was higher than those in angiosperms (Supplementary Table 7). Phylogenetic reconstructions revealed that conifers displayed substantially higher diversity and abundance than *G. montanum* and *G. biloba*, possibly indicating gradual and/or rapid diversification in conifers.

The *T. yunnanensis* genome, as a second member belonging to the so-called conifers II clade for which the genome sequence has been determined, provides an opportunity to revisit the relationships of gymnosperms. Using six gymnosperms, five angiosperms and two pteridophytes, and *Anthoceros punctatus* as an outgroup, we identified 588 single-copy gene families (5951 genes in all of the 14 species) to construct a phylogenetic tree, using ASTRAL and ‘superbase’ based on amino acid alignments, DNA alignments, codon alignments and codon alignments with third-positions removed (Fig. 2a, Supplementary Table 8 and Supplementary Data 1). All of the phylogenomic analyses showed that *T. yunnanensis* was most related to *S. giganteum*, with an estimated divergence time of 133.4−213.0 MYA, representing the conifers II clade. The split between conifers I and conifers II was estimated at 219.1−257.2 MYA. All but one of the ASTRAL analyses (DNA alignment) placed *G. montanum* as sister to all other extant gymnosperm lineages, further supporting the Gnetales–other gymnosperms hypothesis of gymnosperm phylogeny (Supplementary Fig. 5).

However, this relationship is at odds with a general phylogeny proposed by the 1KP consortium, which finds Cyad and Ginkgo as sisters to the rest of gymnosperms based on transcriptome data. This difference may require further study, such as the use of genome data for additional gymnosperms.

A total of 575 gene families were expanded, 55 of which exhibited significant expansion (*P* < 0.05), relative to the ancestor.
Some of these genes were annotated as a cellular component, such as apoplast (GO:0048046) and nucleosome (GO:0000786); the biological process chromosome stability such as DNA integration (GO:0015074), telomere capping (GO:0016233) and mitotic cell cycle (GO:0000278); molecular functions related to the synthesis of the primary metabolite, such as aspartic-type endopeptidase activity (GO:0004190), cysteine-type peptidase activity (GO:0008234), polysaccharide binding (GO:0030247) and protein heterodimerization activity (GO:0046982) (Supplementary Fig. 6 and Supplementary Data 2). Seventy-two genes related to the apoplast, of which 57 genes were annotated as Dirigent protein in the UniProt database, were discovered in coniferous trees, and participating in lignan biosynthesis for defense purposes (Supplementary Data 3). A total of 907 gene families, many involved in ATPase activity, coupled to transmembrane movement of substances (GO:0042626), ATPase activity (GO:0016887) and transmembrane transport (GO:0055085), showed contraction (Supplementary Fig. 7 and Supplementary Data 4).

Among T. yunnanensis, S. giganteum, G. montanum, and G. biloba gene families, a total of 2328 gene families appeared unique to T. yunnanensis (Fig. 2b and Supplementary Data 5), and were particularly enriched in isoquinoline alkaloid biosynthesis (ko00950), flavone and flavonol biosynthesis (ko00944), and ubiquinone and other terpenoid-quinone biosynthesis (ko00130) (Supplementary Fig. 8 and Supplementary Data 6).

Most angiosperms have undergone whole-genome duplication (WGD) somewhere during their evolutionary past. Although it has been reported that all seed plants shared an ancient WGD, WGDs in gymnosperms seem to be much rarer. WGDs are
usually identified from Ks (a measure of the number of substitutions per synonymous site) age distributions of paralogs, or from gene collinearity data. Since Ks age distributions showed no clear peaks and no widespread intragenomic colinear or syntenic segments could be detected, we assume no recent WGD event has occurred in the evolutionary past of *T. yunnanensis*, although older WGDs cannot be excluded (Fig. 2c, Supplementary Fig. 9 and Supplementary Fig. 10). Evidence for small-scale gene duplication events is more evident and general analysis of gene duplication in *T. yunnanensis* shows that dispersed duplicates (60.07%) from the dominant type compared to three other types: WGD/segmental duplication (0.75%), proximal (11.66%) and tandem (13.09%) (see ‘Methods’).

All of the hydroxylases involved in Taxol biosynthesis belong to CYP450s. The CYP450s responsible for hydroxylation at the C-2, C-5, C-7, C-10, C-13 and C-2’ positions have been characterized in *T. yunnanensis*, *G. biloba*, *P. abies*, *S. giganteum* and *A. trichopoda* (Fig. 3a and Supplementary Data 7). The genome assembly allowed us to locate all of the functionally characterized genes of Taxol metabolism in *Taxus*, as well as *S. giganteum* and *G. biloba*. So far, Taxol is found only in *Taxus* species, indicating that some of the CYPs involved in the biosynthesis of Taxol may be specific to *Taxus*. In order to identify such species-specific CYPs, we compared a total of 3368 CYP genes in *T. yunnanensis*, *S. giganteum*, *G. biloba*, *P. abies*, *S. giganteum* and *A. thaliana* (Supplementary Table 9). In total, 624 CYP450s genes were identified in the *T. yunnanensis* genome, and the number of CYP725 sub-family genes was substantially higher than that in other species (Supplementary Table 9). The CYP450s genes involved in the biosynthesis of Taxol belong to CYP725A sub-family. We constructed a phylogenetic tree from CYP725 genes obtained from a genome sequence alignment of five species (*T. yunnanensis*, *S. giganteum*, *G. biloba*, *P. abies*, *P. menziesii*) (Supplementary Fig. 11). Sixty-eight specific CYP725 genes were found in the *T. yunnanensis* genome, of which 62 genes belong to the CYP725A sub-family. We constructed a phylogenetic tree from CYP725 genes obtained from a genome sequence alignment of five species (*T. yunnanensis*, *S. giganteum*, *G. biloba*, *P. abies*, *P. menziesii*) (Supplementary Table 9). In total, 624 CYP450s genes were identified in *T. yunnanensis* genome and the number of CYP725 sub-family genes was substantially higher than that in other species (Supplementary Table 9). The CYP450s genes involved in the biosynthesis of Taxol belong to CYP725A sub-family. We constructed a phylogenetic tree from CYP725 genes obtained from a genome sequence alignment of five species (*T. yunnanensis*, *S. giganteum*, *G. biloba*, *P. abies*, *P. menziesii*) (Supplementary Data 8). The genome assembly allowed us to locate all of the functionally characterized genes of Taxol metabolism in *Taxus*, as well as *S. giganteum* and *G. biloba*.
as their closely related homologs, on either the chromosome or the unplaced scaffold positions. Forty CYP725A genes were found distributed on chromosome 12, including hydroxylation-like genes responsible for C-2, C-5, C-7, C-10, C-13 and C-14 hydroxylation. Moreover, taxadiene synthase (TXS) and 10-deacetylbaccatin III-10-O-acetyltransferase (DBAT) like genes involved in the Taxol biosynthetic pathway were also located on chromosome 12 (Fig. 3c and Supplementary Data 9). These genes were grouped on a 76.2 Mb region to form a taxol synthesis gene cluster which was artificially divided into three sub-clusters (sub-cluster I, II, III) (Fig. 3c). We detected 12 functionally uncharacterized CYP725 genes in the sub-cluster I, which exhibited a similar expression pattern (Pearson correlation coefficient > 0.8, P < 0.05) with T5αOH, T10βOH, T2αOH, T7βOH and T13αOH, and were highly expressed in bark (Fig. 3c, d). We suspect that these genes may participate in the production of Taxol. TXS and T5αOH are encoded by co-localized gene copies in sub-cluster II; T10βOH, T2αOH, T7βOH and T13αOH are encoded by co-localized gene copies in sub-cluster III, which were all highly expressed in the bark of *T. yunnanensis* (Fig. 3c, d; Supplementary Data 9, 10). Moreover, 15 functionally uncharacterized CYP725 genes localized in sub-cluster II and III, which have low homology with known hydroxylation-like genes in Taxol pathway, while exhibiting high and similar expression patterns (Pearson correlation coefficient > 0.8, P < 0.05) as the known genes functioning in Taxol metabolism. These genes might be interesting as potential candidates genes in Taxol biosynthesis pathway (Supplementary Data 9, 10).

**Conclusions**

This study reports a high-quality chromosome-level genome assembly for *T. yunnanensis*. This provides crucial information for the study of the evolution of gymnosperms. We estimated that there is no evidence of a recent WGD in *T. yunnanensis* and LTR expansion is the main cause of its large genome size. Interestingly, the CYP725A gene families, encoding hydroxylase involved in Taxol synthesis, exhibited significant expansion, and most of
them clustered on chromosome 12 and exhibited co-expression, which contributes to the further elucidation of the Taxol bio-synthetic pathway.

Methods

Plant materials, DNA library construction and sequencing: Fresh leaves were collected from *T. yunnanensis* in Yunnan province. High-quality genomic DNA was isolated from five fresh leaves using the CTAB method and the DNA quality and concentration were tested by 0.75% agarose gel electrophoresis. NanoDrop One spectrophotometer (Thermo Fisher Scientific) and Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

After the DNA quality and integrity were tested, it was randomly sheared by Covaris ultrasonic disruptive. Illumina sequencing pair-end libraries with an insert size of 300 bp were prepared using Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). Sequencing was performed using the Illumina NovaSeq platform (Illumina, San Diego, CA, USA). Raw reads were cleaned to discard low-quality reads (reads with adaptors or unknown nucleotides (Ns) or reads with more than 20% low-quality bases) using the SOAPnuke (v2.1.4) tool (https://github.com/BGI-flexlab/SOAPnuke) and, after data filtering, clean data were used for subsequent analyses.

For Oxford Nanopore sequencing, the libraries were prepared using the SQK-LSK109 kit and using the protocol. The purified library was loaded onto primed R9.4 Spot On Flow Cells and sequenced using a PromethIon sequencer (Oxford Nanopore Technologies, Oxford, UK) with 48-h runs at Wuhan Benagen Tech Solutions Company Limited, Wuhan, China. Base-calling analysis of raw data was performed using the Oxford Nanopore GPPY software (v3.0.5).

RNA library construction, sequencing and data processing: For gene prediction analysis, total RNA was extracted from young leaves of *T. yunnanensis* using the RNA prep Pure Plant Plus Kit according to the manufacturer’s instructions (NanoDrop One spectrophotometer (Thermo Fisher Scientific) and Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA)). Sequencing was performed using the Illumina NovaSeq platform (Illumina, San Diego, CA, USA). Raw reads were cleaned to discard low-quality reads (reads with adaptors or unknown nucleotides (Ns) or reads with more than 20% low-quality bases) using the SOAPnuke (v2.1.4) tool (https://github.com/BGI-flexlab/SOAPnuke) and, after data filtering, clean data were used for subsequent analyses.

Gene prediction. Evidence from transcript mapping, ab initio gene prediction, and homologous gene alignment was combined to predict protein-coding genes in the *T. yunnanensis* genome. CNT cDNA reads from *T. yunnanensis* were aligned against the *T. yunnanensis* genome using Minimap2 (v2.17). Transcriptome assemblies were each aligned using MUSCLE (v3.8.31) with default parameters. Phylogenetic trees were inferred based on multiple sequence alignment using FastTree (v2.1.9). The integration of genes (f) of intact LTRs were estimated using the equation *t* = K/r, where K is the number of nucleotide substitutions per site between each LTR pair and *r* is the substitution rate, which was set to 7.34573 × 10⁻¹ substitutions per site per year⁻¹.

Phylogenetic tree construction. All amino acid sequences of the 14 selected species were aligned using Blastp (v.2.6.0) with the parameters: -e value cut-off 1e⁻⁰⁵ searches against entries in both the NCBI and UniProt databases (http://www.uniprot.org/). Searches for gene motifs and domains were performed using InterProScan (v3.53) and HMMER (v3.1). The GO terms for genes were obtained from the corresponding InterPro (https://www.ebi.ac.uk/interpro) or UniProt entry (https://www.uniprot.org/). Pathway annotation was performed using KOBAS (v3.5) (https://github.com/samoa/kobas) against the KEGG database.
and WAG models for amino acid data. For each analysis, support was inferred for branches on the final tree from 100 bootstrap replicates.

Based on the phylogenetic tree result, the incmctree of PAML (v.4.9; parameter: nsample = 1000000; burnin = 200000; seqtype = 0; clock = 3; model = 4) was used to estimate the divergence time of the different species. Published divergence times45,46 for A. thaliana and V. vinifera: 90–120 MYA, N. colorata and A. trichopoda: 215–265 MYA, P. abies and P. taeda: 123–220 MYA, S. moellendorffii and P. abies: 410–440 MYA, P. abies and A. trichopoda: 250–390 MYA and O. sativa and V. vinifera: 125–150 MYA were used to calibrate the divergence time.

Gene family contraction and expansion analysis were performed using CAFE (v.2.1; parameter: –filter) software47 based on gene family clustering results.

Whole-gene duplication analysis. All T. yunnanensis amino acid sequences were self-aligned using BLASTp and the value cut-off 1e−05 and the best BLASTp result was retained. To obtain paralogous gene families, we performed gene cluster analyses based on the CDS alignment using OrthoMCL (v.2.0.9)48. Ks values were calculated from all paralogous families using y0n0 in the PAML package47. The Ks of a given family was represented by the median value, and the distribution of corrected Ks values was plotted by median values48. To distinguish whether this peak represents a whole-genome duplication event or background small-scale duplications, we identified paralogous gene pairs using Blastp methods and determined syntonic blocks using MCScanX30 (https://github.com/wyp1125/MCScanx). Although the synonymous substitution rate (Ks) was calculated for T. yunnanensis syntonic block gene pairs and Ks distribution clearly showed a major peak at around 0.1, there were no widespread and well-maintained one-versus-one syntonic blocks indicates that a recent whole-genome duplication (WGD) event has not occurred in the T. yunnanensis genome. Indeed, analysis of duplication types of the T. yunnanensis paralogs by Duplicate gene classifier tool of MCScanX30 indicates that there are four types: WGD/segmental duplication (match genes in the genome), pseudogene (match genes with insertions), family (match genes with insertions) and tandem (match genes with insertions). For T. yunnanensis and WAG models for amino acid data. For each analysis, support was inferred for branches on the final tree from 100 bootstrap replicates.

Received: 21 June 2021; Accepted: 18 September 2021; Published online: 20 October 2021

References
1. Yu, C. et al. Comparative metabolomics reveals the metabolic variations between two endangered Taxus species (T. fujiana and T. yunnanensis) in the Himalayas. BMC Plant Biol. 18, 197 (2018).
2. Robin Foa, L. N. & Andrew, D. Seidman Taxol (paclitaxel): a novel anticancer microtubule agent with remarkable anti-neoplastic activity. Int. J. Clin. Lab. Res. 24, 6–14 (1994).
3. Li, Y. L. et al. A protocol of homozygous haploid cell induction from endosperm of Taxus chinensis Relh. var. mairei. SpringerPlus 5, 659 (2016).
4. Yuan, H. et al. Albumin nanoparticle of paclitaxel (Abraxane) decreases while taxol increases breast cancer stem cells in treatment of triple negative breast cancer. Mol. Pharm. 17, 2275–2286 (2020).
5. Zheng, L. L., Wen, G., Yao, Y. X., Li, X. H. & Gao, F. Design, synthesis, and anticancer activity of natural product hybrids with paclitaxel side chain inducing apoptosis in human colon cancer cells. Nat. Prod. Commun. 15, 1934575X20971299 (2020).
6. Xi, X. J. et al. Genetic diversity and taxol content variation in the Chinese yew Taxus mairei. Plant Syst. Evol. 300, 2191–2198 (2014).
7. Kuang, X., Sun, S., Wei, J., Li, Y. & Sun, C. Iso-Seq analysis of the Taxus cuspidata transcriptome reveals the complexity of Taxol biosynthesis. BMC Plant Biol. 19, 210 (2019).
8. Sanchez-Munoz, R. et al. A novel hydroxylation step in the taxane biosynthetic pathway: a new approach to paclitaxel production by synthetic biology. Front. Bioeng. Biotechnol. 8, 410 (2020).
9. Schneider, F., Samarin, K., Zanella, S. & Gaich, T. Total synthesis of the complex taxane diterpene canastaxpropellane. Science 367, 667–681 (2020).
10. Wang, X. Q. & Ran, J. H. Evolution and biogeography of gymnosperms. Mol. Phylogeny Evol. 75, 24–40 (2014).
11. De La Torre, A. R., Li, Z., Van de Peer, Y. & Ingvarsson, P. K. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol. Biol. Evol. 34, 1363–1377 (2017).
12. Wan, T. et al. A genome for gnophetes and early evolution of seed plants. Nat. Plants 4, 82–89 (2018).
13. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574, 679–685 (2019).
14. Wong, G. K. S. et al. Sequencing and analyzing the transcriptomes of a thousand species across the tree of life for green plants. Annu. Rev. Plant Biol. 71, 741–765 (2020).
15. Wickett, N. J. et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl Acad. Sci. USA 111, E4859–E4868 (2014).
16. Guan, R. et al. Draft genome of the living fossil Ginkgo biloba. Gigascience 5, e1002411 (2016).
17. Scott, A. D. et al. The giant sequoia genome and proliferation of disease resistance genes. https://doi.org/10.1101/2020.03.17.995944 (2020).
18. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015).
19. Nystedt, B. et al. The Norway spruce genome sequence and conifer genome evolution. Nature 497, 579–584 (2013).
20. Lee, K. et al. A functional phylogenomic view of the seed plants. PLoS Genet. 7, e1002411 (2011).
21. Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. Nature 473, 97–100 (2010).
22. Roodt, D. et al. Evidence for an ancient whole genome duplication in the cyad lineage. PLoS ONE 12, e0184454 (2017).
23. Van de Peer, Y., Mizrahi, E. & Marchal, K. The evolutionary significance of polyploidy. Nat. Rev. Genet. 18, 411–424 (2017).
24. Chau, M. & Croteau, R. Molecular cloning and characterization of a cytochrome P450 taxoid 2alpha-hydroxylase involved in Taxol biosynthesis. Arch. Biochem. Biophys. 427, 48–57 (2004).
25. Chau, M., Jennewein, S., Walker, K. & Croteau, R. Taxol biosynthesis: molecular cloning and characterization of a cytochrome P450 taxoid 2alpha-hydroxylase. Arch. Biochem. Biophys. 387, 379–387 (2004).
26. Jennewein, S., Rithner, C. D., Williams, R. M. & Croteau, R. B. Taxol biosynthesis: taxane 13- hydroxylase is a cytochrome P450-dependent monoxygenase. Proc. Natl Acad. Sci. USA 98, 13595–13600 (2001).
