nature portfolio

Peer Review File

The genome of Pseudotaxus chienii provides insights into the origin and evolution of taxane biosynthesis

Corresponding Author: Professor Chenjia Shen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
As a sister genus to Taxus, Pseudotaxus holds some potential for elucidating the evolutionary trajectory of the taxane biosynthesis pathway. The authors presented a chromosome-level genome assembly of Pseudotaxus and employed comparative genomics to investigate the evolution of its genome structure and gene families. They compared the metabolite profiles of Pseudotaxus, Torreya grandis, and Taxus wallichiana. Additionally, transcriptome analysis revealed that treatment with MeJA influences the accumulation of taxusin in Pseudotaxus. This study provides preliminary insights into the evolution of taxane metabolic pathways in Pseudotaxus. Some points need to be addressed and revised:

1. In figure 1, The authors showed the tissue-specific distribution of active ingredients such as flavonoids, prenol lipids, coumarins, and steroids in the stems and leaves of Pseudotaxus through spatial metabolomics and analyzed the accumulation of these ingredients in different tissues. However, this part lacks a coherent logical connection with the full text. It is advisable to either omit this portion or enhance the logical linkage between this section and the main body of the text.

2. In lines 171-173, the authors claim that "the LAI value was estimated to be 10.1, which is larger than the proposed standard, indicating that the chromosome-level assembly of P. chienii is of high quality." However, a LAI value of 10.1 does not necessarily reflect a high quality of a genome assembly, as a LAI value of 15 is generally considered a relatively good result, and LAI > 20 is regarded as the gold standard. Please explain this and provide evidence to support why the value of 10.1 is still considered larger than the proposed standard.

3. In Supplementary Table 5, the percentages of LTR-Gypsy and LTR-Copia are 37.93% and 8.88%, respectively; however, in Line186, the percentages of the two become 40.87% and 9.04%, respectively. Please explain why there is a contradiction between the two.

4. In Supplementary Table 5, the percentages of LTR-Gypsy and LTR-Copia are 37.93% and 8.88%, respectively. However, in Line 186, the percentages increase to 40.87% for LTR-Gypsy and 9.04% for LTR-Copia. This discrepancy may warrant further investigation. In addition, LTR elements make up a total of 79.87%. Among these, LTR-Gypsy accounts for 37.93%, and LTR-Copia accounts for 8.88%. Typically, LTR-Gypsy and LTR-Copia are expected to dominate the composition of LTR elements. However, this is not the case in this instance. Could you please explain what the remaining portion of the LTR elements comprises?

5. Line 169-170, authors stated “BUSCO evaluation revealed that 90.5% (1,460 out of 1,614) of the core eukaryotic genes were identified in the P. chienii genome”. However, in Supplementary Table 4, the BUSCO of genome and protein were 99.6% and 98.4%, respectively. Please explain why the two are different. In addition, in Supplementary Table 4, The database used by BUSCO seems to be eukaryota_odb10, which has only 255 gene sets. Obviously, this data is not enough to explain anything. It is recommended to use embryophyta_odb10 to evaluate the genome and annotate proteins. Finally, regarding BUSCO evaluation, please add detailed experimental methods to the manuscript.

6. In Line 228, 35404 protein-coding genes are annotated, but in Supplementary Table 7, there are 53407 annotated genes. What accounts for the significant difference between these two numbers?

7. The lack of genomic data and direct access to raw sequencing data and analysis codes hinders the verification of

conclusions and the reproducibility of analyses using this data.

8. Line 286, authors claim that “Taxusin is the end product of taxane biosynthesis pathway in P. chienii”, more evidence should be added to support this conclusion. It is known that T2OH enzymes can catalyze the formation of taxusin to the 2α-hydroxytaxusin. Since T2OH is present in the Pseudotaxus genome, it is necessary to test the catalytic activity of Pseudotaxus-derived T2OH on taxusin.

9. Line 296-298: “Due to the absence of TBT, which catalyzes the conversion of 2-debenzoyl-7,13-diacetylbaccatin III to 7,13-diacetylbaccatin III, P. chienii is unable to synthesize 10-deacetylbaccatin III and its subsequent intermediates.” Based on this point, the authors should conduct a rescue experiment to support this conclusion, i.e. the authors should integrate the reported functional TBT from Taxus mairei into the tissue of P. chienii and detect the 10-deacetylbaccatin III product by LCMS.

10. The corresponding author Chenjia Shen of this manuscript has reported that 10-deacetylbaccatin III and paclitaxel were able to be detected in Pseudotaxus chienii in 2021 (BMC Plant Biology, 2021, 21: 104). However, the author claimed that both metabolites were not detected in current manuscript. Please explain the conflicting results.

11. In Figure 5A, PcTS1 and PcTS2 cluster together with the bifunctional terpene synthases (TPS) into one large clade, but they were classified as class I TPSs in this manuscript. It is necessary to clarify why PcTS1 and PcTS2 are classified as class I TPSs rather than bifunctional TPSs. The structures of PcTS1 and PcTS2 can be predicted using AlphaFold 3 and their structural similarities (e.g., TM scores) can be compared with those of known class I and bifunctional TPSs to classify them correctly.

12. Line 581, Line 599, “Trans_Decoder” and “Repeat_Masker” should be correctly written as “TransDecoder” and “RepeatMasker”. These are not the only instances; there are many places in the manuscript where software names are incorrectly formatted. It is essential to review the entire text carefully for these errors. Additionally, all software referenced in this manuscript must be properly cited.

13. Line 213, The legends need to be correctly labeled, Fig 2f should be labeled on line 213. And the following Fig 2e should be adjusted to line 223. Authors are requested to carefully check the legend and its location in the full text.

14. The paclitaxel biosynthesis pathway was incomplete and some steps and new enzymes were missing in Fig3, the authors should add more detailed information based on several recent reports (Jiang et al., Science, 2024; Yang et al., Nature Communications, 2024; Li et al., Angewandte Chemie International Edition, 2024; Zhang et al., Molecular Plant, 2023, etc.).

15. Line 658-659, although the author claims that "Homologous genes in P. chienii were obtained by performing BLASTp searches with identity > 60 and protein length > 700 aa.", many of these proteins are smaller than 700 aa, and at least some genes of the CYP family in the paclitaxel biosynthesis pathway are not. Please explain why this experimental method is inconsistent with the actual situation. It is essential to validate this calculation method to ensure the reliability of the research findings.

16. Supplementary Figure S7 and Figure S10 are at low resolution. Please replace it with a high-resolution version.

Reviewer #2

(Remarks to the Author)
The manuscript by Wang et al presents the genome of Pseudotaxus chienii, a relict conifer endemic to China and purports to provide insights into the origin and evolution of taxane biosynthesis. Whilst I am impressed by the first part of this study namely the genome sequence I am less so regarding the origin and evolution of taxane biosynthesis.

My concerns here stem from the fact that they exclusively rely on the biosynthetic pathway published in Science last year - which contains massively fewer proteins than either the pathway proposed by the Fernie group in Molecular Plant in 2023 or that of the Sattely group in Nature Communications in 2024. A proper analyses would require all of the enzymes/genes identified in all three papers to be used in the evaluation. Only then, in my opinion would the study be complete. As a further explanation a recent review by Fernie et al. published in Nature Plants revealed that the heterologous overexpression of the pathway proposed by Fernie resulted in 100 fold higher levels of Baccatin III than that reported in the Science paper. I thus feel that it is compulsory that analyses of a much broader set of candidate genes is performed.

Reviewer #3

(Remarks to the Author)
This study conducts a chromosome-level genomic analysis of Pseudotaxus chienii to deeply explore the origin and evolution of the taxane biosynthesis pathway in the Taxaceae family. It reveals part of the taxane biosynthesis pathway in P. chienii and offers key insights into functional differences between the Taxus and Pseudotaxus genera during evolution, using comparative genomics, metabolomics, and functional validation.
Major comments:
1. In the section “Visualization of metabolite locations in the stems and leaves of P. chienii”, the MSI results of leaves and stems of P. chienii have little connection to the following experiments, at least the taxane biosynthesis-related metabolites

were not detected and visualized. The author should supplement the MSI results for taxane biosynthesis-related metabolites.
2. In Figure 1c, are whether MALDI-MS spectra attributed to leaves or stems? It is recommended that the mass spectra of both samples be provided and the active ingredients labeled.
3. The author should provide the optical images of the P. chienii leaf and stem sections in Figure 1d/e/f/g to better understand the segmentation of tissues. Besides, the author should supplement the information on the matrix used in MSI analysis, and the comparison between MSI results and the matrix background.
4. The clustering analyses of MSI results are confusing; the author should explain the differences between principal components (PCs) separation and the following clustering analyses of SAM. How to determine the attribution of identified metabolites to specific tissues? The clustering analyses identified “1,494 pith-, and xylem-SAMs”, but the pith and xylem are different tissues in PCs analysis; the statement should be “1,494 pith-, and 1531 xylem-SAMs”, corresponding to Figure S2c. Besides, the identified metabolites were clustered into six groups, and the author only explained clusters I to V.
5. According to Tables S1 and S2, the m/z errors of certain metabolites and active ingredients are notably high, including Taxine B (9.68), Cyanidin (8.40), ζ-carotene epoxide (9.71), Presqualene-PP (9.75), and Campesterol glucoside (9.64). These discrepancies could impact the accuracy of compound identification. The author should critically check the MS results.
6. In Figure 1h, the author calculated the proportion of different active ingredients in stems and leaves. The authors should explain the significance of the calculated proportion, as the results showed that leaves also contained more active ingredients (even higher than stems), while the twigs from a wild P. chienii tree were collected for the subsequent genome sequencing.
7. In the section “Genome assembly and annotation”, there appears to be no mention of genome annotation.
8. Please further explain "Most of the CYP725 genes are situated within a very restricted genomic region (Fig. 6), which may account for the consistent response of CYP725 to MeJA treatments" to better understand the underlying mechanism. Additionally, the author should supplement a reference to support the statement “MeJA is a significant elicitor that was commonly used to enhance the production of Taxol”.
9. In lines 347-349, "Further analysis of protein structures demonstrated that TPSs from the Taxus genus, along with PcTS1, exhibit a greater structural similarity to angiosperms than to other gymnosperms (Fig. 5g)." It is suggested to distinguish the structural similarities and differences between Taxus TPS and angiosperms in the figure, and make a more intuitive comparison in combination with protein models from other gymnosperms species. Additionally, the protein structures of the Taxus genus, TmTXS, and TbTDC1, show a significant similarity and seem to be the same image (Please double check these two figures). The author should discuss it or point out the differences between two protein structures.
10. The conclusion in the "Whole genome duplication (WGD)" section that P. chienii has not experienced recent WGD is reasonable. However, the evidence for ancient WGD is insufficient. Although a weak Ks peak was observed, its timing was not verified to be consistent with the evolutionary history of P. chienii; furthermore, Ks values are susceptible to heterogeneous evolutionary rates or sequencing errors.
11. The article used 184 single-copy genes for phylogenetic analysis and combined the MCMCTREE program to estimate divergence times. It inferred that P. chienii and Taxus diverged approximately 48.3 million years ago. However, the article did not provide detailed information on the calibration points or methods used, which may affect the accuracy of the divergence time estimation.
12. The author should provide the duplication levels of transgenic experiments.
13. In the section “Functional identification of the TS genes in P. chienii”, the new peak observed at an RT of 11.875 mins matched exactly with the analog of taxadiene, eicosane. However, the mass spectra lack sufficient information. The author should label the analog of taxadiene and eicosane in the MS profile and provide more information for mass spectrum annotation. Furthermore, the mass spectra of standard substances for the analog of taxadiene and eicosane should be included.
14. Annotating taxane biosynthesis-related genes based solely on sequence similarity is feasible but has limitations. To enhance the accuracy and reliability of the annotations, it is suggested to validate the functions of key enzymes involved in taxusin biosynthesis through in vitro activity assays. Additionally, metabolomic analyses should be conducted to detect other intermediate products accumulated in P. chienii, thereby further confirming whether taxusin is the end product of its taxane biosynthesis pathway.
15. It is recommended to italicize the Latin names of species in Figure S7 to conform to academic conventions; please enhance the resolution of Figure S10/S2b/S3b/S8a or optimize the image quality for clearer data presentation. There is an error in the figure numbering in Figure S12, and it is suggested to carefully check and correct it. There are also many traces in the image, including border residues that are recommended for modification, such as Figure S9. Besides, the format of labels in Figures 4c and 4d should be the same.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed some my questions in the previous round. However, there are still several points that need improvement.
1. The protein annotation results provided by the authors (https://figshare.com/s/1a98960b9985e2a0c198) were used to count the number of genes and evaluate the protein BUSCO (BUSCO v5.4.5, embryophyta_odb10). The results showed that the number of genes was 33789 and the BUSCO of proteins was 84.9%. This result is still inconsistent with the author‘s revision (BUSCO：90.5%; number of genes :35404), and the quality of the gene annotation results is much lower than the

results in the author's revised manuscript.

2. The genome annotation method provided by the author in the manuscript is to use EvidenceModeler to integrate the results of three annotation strategies: transcriptome-based prediction and homology-based gene prediction and ab initio prediction to obtain the final result. If this method is followed, the second column in the gene annotation result gff3 file should be a unified label, such as EVM. If this method is followed, the second column of the gene annotation gff3 file should be a unified label, such as EVM. However, in the gff3 file provided by the author, there are 5 labels in the second column (AUGUS, exonerate, GlimmerHMM, maker and transdecoder). It is more like artificial selection and manipulation.

3. In the author's response to comment 8 from “Reviewer #1”, the issue raised in my previous comment remains unresolved. The key concern is that the presence of T2OH in P. chienii indicates the potential for further conversion of taxusin into downstream metabolites such as 2α-hydroxytaxusin. Meanwhile, more data should to be provided to support the conclusion of “Considering the presence of both T2OH and T7OH, taxusin can be hydroxylated into 2α,7β-dihydroxytaxusin.” in P. chienii. Although, the acquirement of standard compound and the establishment of the genetic transform system are difficult, some works are able to be done. For examole:
1) Heterologous expressing in tobacco with the substrate feeding or purifying the protein and conducting a in vitro enzyme activity assays to prove that the T2OH or T7OH in P. chienii have the same or different function with that from other Taxus species.
2) showing the protein sequence identity or evolutionary relationship of T2OH or T7OH in P. chienii with that from other Taxus species.

4. In the author's response to comment 9 from “Reviewer #1”, the key point is to prove the directly relationship of the absence of TBT and the deficiency of 10-deacetylbaccatin III or baccatin III. Although, the absence of TBT maybe the reason of the undetected 10-deacetylbaccatin III or baccatin III, but other potential reasons should be pointed out. The absence of several taxanes can’t support the conclusion that the inability of P. chienii to synthesize 10-deacetylbaccatin III and its subsequent intermediates is solely due to the absence of TBT. Moreover, the sentence of the response “All these intermediate products could not be detected in P. chienii, thus we predicted that P. chienii is unable to synthesize 10-deacetylbaccatin III and its subsequent intermediates.” seemingly implies that the taxanes of 10-deacetyltaxol, 10-deacetyl-7-xylosyltaxol C, 13-acetyl-9-dihydrobaccatin III, and cephalomannine are the intermediate products of 10-deacetylbaccatin III and results in the misunderstands. Meanwhile, the 13-acetyl-9-dihydrobaccatin III hasn’t the side-chain and it is seemingly inappropriate to classify it with other three taxanes.

5. In the author's response to comment 10 from “Reviewer #1”, the contradiction between author's 2021 publication (BMC Plant Biology, t2021, 21: 104) and the current manuscript remains unresolved. The key issue is not whether P. chienii is a debated species, nor whether author's current sample has been authenticated, but why author own previous work reported the presence of 10-deacetylbaccatin III and paclitaxel while the current study reports their absence. What needs to be said more is that the author has negated the previous results by questioning the accuracy of the 2021 research materials. This approach is unconvincing, especially under the premise that both studies are led by the same corresponding author. If there are problems with the 2021 materials, please clearly state how the materials were identified in the study and any specific errors that may have existed, and explain why they were not explained or corrected at the time. If there are methodological biases or analytical misjudgments in the 2021 results, this should also be fully explained. Without a clear and scientifically grounded explanation, the conflicting results raise concerns about data reproducibility and the reliability of metabolite profiling in this species.

6. In the author's response to comment 11 by Reviewer #1, the authors state that TM-scores less than 0.18 indicate higher similarity with Class I TPSs. However, this interpretation is incorrect. TM-scores range from 0 to 1, with values above 0.5 generally indicating significant structural similarity, and scores below 0.2 typically reflecting random structural similarity. Therefore, claiming that a lower TM-score (e.g., <0.18) indicates greater structural similarity is inaccurate.

Reviewer #2

(Remarks to the Author)
I am fully satisfied that this resubmission addresses all the concerns I raised in my review of the previous submission.

Reviewer #3

(Remarks to the Author)
1. In the author’s reply to the 9th comment from Reviewer 1, Fig. 4g should be Fig. 4f.
2. In the author’s reply to the 5th comment from Reviewer 3, the author stated that high mass errors may be due to calibration bias or insufficient mass spectrometer resolution. They still need to provide fragment (MS/MS) data of the key compounds for compound identification.
3. The author should supplement the scale bar for Figure 1b/d. Besides, the author should check the size of the MALDI-2 chip. The 6.5*6.5 mm chip might be small for the leaf and stem materials.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
1. In the author's response to comment 5 from “Reviewer #1”, the response does not address the key scientific concerns raised in previous comments. The explanation is arbitrary and lacks the necessary scientific rigor.
The discrepancy between the two studies from the same research group regarding the detection or presence of paclitaxel is significant and cannot be dismissed as minor. Without a thorough and rigorous reconciliation of these conflicting findings, the credibility of both studies is compromised. The authors attempt to address the inconsistency between the 2021 and current results by suggesting that the material analyzed in the present study represents a new variety, Pseudotaxus chienii var. Nanjianyan. However, this claim lacks support from morphological, genetic, or chemotaxonomic evidence. In the absence of formal diagnostic criteria and designated voucher specimens, the proposed varietal name is scientifically unwarranted and should be retracted unless substantiated by appropriate data.

2. In the author's response to comment 3 of "Reviewer #1", the authors avoided using misleading statements and provided data on protein sequence identity, which did not help much with the key issues. The authors should conduct an experiment to heterologously express T2OH and T7OH in tobacco fed with taxusin and use T2OH and T7OH expressed in T. maieri as controls.

3. In the author's response to comment 6 from “Reviewer #1”, in the section "Protein structure analysis of Taxaceae TPSs", the authors identified PcTS1 and PcTS3 as ClassI based on Figure e, but there is no PcTS3 in Figure a, only PcTS1 and PcTS2, and the branch where PcTS1 and PcTS2 are located is closer to the branch of bifunctional enzymes than other ClassI, which means that PcTS1 and PcTS2 belong to bifunctional enzymes. Similarly, the structural analysis of Figure h shows that TS1 is a bifunctional enzyme. All these results are contradictory, and it is difficult to judge which one is correct.

4. In figure3a, 1-Dehydroxybaccatin IV cannot be directly synthesized from Baccatin I. In addition, the structural formulas of 2α,7β-Dihydroxytaxusin, Taxadiene hexa-acetate, 1-Dehydroxybaccatin IV, and Baccatin I are in opposite directions.

5. Line156-line157, it should be made clear that the object of this BUSCO assessment is the genome or the protein sequence.

Reviewer #3

(Remarks to the Author)
I agreed with Reviewer 3's comment that MS/MS data should be supplemented. In situ MS/MS is quite significant for compound identification in MS imaging. Due to the lack of LC separation, peak assignment is always ambiguous and may be overlapped with matrix or background ions. Therefore, only MS1 data cannot guarantee the accuracy of the results. Unlike the mouse brain (Nat Methods. 2025 May;22(5):1051-1058), a model tissue for MSI has been thoroughly interrogated and public databases are available; identification of plant secondary metabolites needs more efforts, particularly for key metabolites.

Version 3:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript has gone through several rounds of revision. While its quality has greatly improved, some issues remain as follows:
1.Several key inconsistencies need to be addressed and corrected throughout the paper. For example, the Abstract and Results sections state the genome size as "15.6 Gb assembled" (line 33) and "15.5 Gb (99.0%) anchored" (lines 148–149), respectively, while the Introduction states, "It contains 16.79 Gb of data, of which 15.45 Gb were assigned to 12 pseudochromosomes" (line 105).

2.The authors reanalyzed their previously published results and explained the conflict in results about whether paclitaxel is present in P. chienii. It is recommended to include the specimen IDs of the materials they had used to improve the reliability of the findings.

3.Pathway intermediates claimed in metabolite detection (e.g., taxusin) need to be verified by targeted LC-MS/MS. Supplementary materials should include retention times and MS/MS spectra of taxusin in P. chienii samples and standards, along with LOD/LOQ, ion ratios, and recoveries. Furthermore, Figure S9e shows similar accumulation levels of taxusin across different groups, but only the peak for the control group "pBWA + Taxusin" is detected in the figure, which is clearly unreasonable and needs to be explained regarding how to quantify the residual taxusin.

4.This paper contains several errors and contradictions regarding the classification of TPS sequences. Initially, it classifies TPSs with DDXXD/E motifs as class I and those with DXDD motifs as class II (lines 83-84). However, the paper then states that only Taxaceae TPSs lacking the "DDXD" motif belong to class I and further describes the core DXDD motif as characteristic of class I TPSs (lines 467-474), which contradicts the widely accepted classification approach, as DXDD is

characteristic of class II TPSs.

5.The manuscript still has several typographical and grammatical errors, including “signa” corrected to “signal” (Line 534), “tblasn” corrected to “blastn” (Line 590), “we h must acknowledge” (Line 400), and “should be conducted out” (Line 403).

Reviewer #3

(Remarks to the Author)
No additional comments.

Version 4:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I carefully reviewed the revised manuscript by Wang et al and especially the comments to the last round of reviews to which I did not contribute to. I was generally very happy with the responses provided. However, I must admit I found the comment that "we are not really experts in the use of the software rather worrying. However, looking at the Supplementary Figure it would appear that there annotation is likely correct and I am guessing that this was just unfortunate use of English since the annotation makes a lot of biological sense and also seems to be technically correct. Should the authors generally have meant that they were unsure with the software I would however suggest that they contact the vendor and ask for their help in confirming the annotation.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

As a sister genus to Taxus, Pseudotaxus holds some potential for elucidating the evolutionary trajectory of the taxane biosynthesis pathway. The authors presented a chromosome-level genome assembly of Pseudotaxus and employed comparative genomics to investigate the evolution of its genome structure and gene families. They compared the metabolite profiles of Pseudotaxus, Torreya grandis, and Taxus wallichiana. Additionally, transcriptome analysis revealed that treatment with MeJA influences the accumulation of taxusin in Pseudotaxus. This study provides preliminary insights into the evolution of taxane metabolic pathways in Pseudotaxus. Some points need to be addressed and revised:

1. In figure 1, The authors showed the tissue-specific distribution of active ingredients such as flavonoids, prenol lipids, coumarins, and steroids in the stems and leaves of Pseudotaxus through spatial metabolomics and analyzed the accumulation of these ingredients in different tissues. However, this part lacks a coherent logical connection with the full text. It is advisable to either omit this portion or enhance the logical linkage between this section and the main body of the text.
A: Thank you for your suggestion. *P. chienii* is a rare gymnosperm with limited research, and its economic value remains unclear. Before studying its genome, I would like to analyze its potential active ingredients through spatial metabolomics. However, the logical linkage between this section and the main body of the text is quite week. In the revised version, the data on flavonoids, coumarins, and steroids were removed from the manuscript. Considering that taxanes are also a type of diterpenoid compound, we only retained the analysis of some potential diterpenoids in *P. chienii* to enhance the logical coherence the text.
(Lines 113-1114, 116-119, and 136-142)

2. In lines 171-173, the authors claim that "the LAI value was estimated to be 10.1, which is larger than the proposed standard, indicating that the chromosome-level assembly of P. chienii is of high quality." However, a LAI value of 10.1 does not necessarily reflect a high quality of a genome assembly, as a LAI value of 15 is generally considered a relatively good result, and LAI > 20 is regarded as the gold standard. Please explain this and provide evidence to support why the value of 10.1 is still considered larger than the proposed standard.
A: In a previous published article, the authors proposed a genome classification system for the assembly of repetitive and intergenic sequence space using LAI: draft quality, with LAI score less than 10; reference quality, with LAI score ranges from 10 to 20; and gold quality, with LAI score greater than 20. According to their classification, our genome can reach to the reference-level.
Reference: Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. 46, e126-e126 (2018).
We have revised the text to make it clear. (Lines 159)

3. In Supplementary Table 5, the percentages of LTR-Gypsy and LTR-Copia are 37.93% and 8.88%, respectively; however, in Line186, the percentages of the two become 40.87% and 9.04%, respectively. Please explain why there is a contradiction between the two.
A: The Supplementary Table 5 is an old version of statistics of repetitive sequences, we have

replaced this table with the right one, sorry for our carelessness.

4. In Supplementary Table 5, the percentages of LTR-Gypsy and LTR-Copia are 37.93% and 8.88%, respectively. However, in Line 186, the percentages increase to 40.87% for LTR-Gypsy and 9.04% for LTR-Copia. This discrepancy may warrant further investigation. In addition, LTR elements make up a total of 79.87%. Among these, LTR-Gypsy accounts for 37.93%, and LTR-Copia accounts for 8.88%. Typically, LTR-Gypsy and LTR-Copia are expected to dominate the composition of LTR elements. However, this is not the case in this instance. Could you please explain what the remaining portion of the LTR elements comprises?
A: The Supplementary Table 5 is an old version of statistics of repetitive sequences, we have replaced this table with the right one.
We surveyed many articles and found that LTR-Gypsy and LTR-Copia are indeed the most proportional LTRs, however, there still may exist a large proportion of LTRs with unknown classification. For example, the Giant lily genome contain 64.40% LTR retrotransposons with Copia and Gypsy accounting for 16.62% and 31.53%, respectively. In the *P. chienii* genome, the ratio of LTR-unknown seems to be much larger (~30%).
Reference: Xu, Sujuan et al., The evolutionary tale of lilies: Giant genomes derived from transposon insertions and polyploidization. The Innovation, 5 (6), 100726 (2024)

5. Line 169-170, authors stated “BUSCO evaluation revealed that 90.5% (1,460 out of 1,614) of the core eukaryotic genes were identified in the P. chienii genome”. However, in Supplementary Table 4, the BUSCO of genome and protein were 99.6% and 98.4%, respectively. Please explain why the two are different. In addition, in Supplementary Table 4, The database used by BUSCO seems to be eukaryota_odb10, which has only 255 gene sets. Obviously, this data is not enough to explain anything. It is recommended to use embryophyta_odb10 to evaluate the genome and annotate proteins. Finally, regarding BUSCO evaluation, please add detailed experimental methods to the manuscript.
A: In the main text, the BUSCO was calculated using the embryophyta_odb10 database and those numbers are right. But we put a wrong BUSCO table using the eukaryota_odb10 database in the Supplementary Table. We have replaced that table. Sorry for our carelessness.
The method of BUSCO evaluation was added to the M&M section (Lines 559-561).

6. In Line 228, 35404 protein-coding genes are annotated, but in Supplementary Table 7, there are 53407 annotated genes. What accounts for the significant difference between these two numbers?
A:The information in the main text is right, however, Supplementary Table 7 is a wrong copy. In the revised version, we have replaced it with the right one.

7. The lack of genomic data and direct access to raw sequencing data and analysis codes hinders the verification of conclusions and the reproducibility of analyses using this data.
A: We have provided the reviewer link for those data when the first time to submit the manuscript. Once the article is accepted, we will release the raw data as soon as possible.
The raw data can be accessed by the link https://dataview.ncbi.nlm.nih.gov/object/PRJNA1205499?reviewer=abkg7ki8dhq2iv9etv45bjk07r.
The genome assembly and genome annotation files can be accessed at https://figshare.com/s/1a98960b9985e2a0c198.

8. Line 286, authors claim that “Taxusin is the end product of taxane biosynthesis pathway in *P. chienii*”, more evidence should be added to support this conclusion. It is known that T2OH enzymes

can catalyze the formation of taxusin to the 2α-hydroxytaxusin. Since T2OH is present in the Pseudotaxus genome, it is necessary to test the catalytic activity of Pseudotaxus-derived T2OH on taxusin.

A: Thank you for your professional advice. In our lab, the standard of 2α-hydroxytaxusin was unavailable, thus quantitative detection of 2α-hydroxytaxusin cannot be completed. As a woody gymnosperm, the genetic transformation system of *P. chienii* has not been established, and it is very difficult for us to detect the catalytic activity of *Pseudotaxus* derived T2OH on taxusin.

In the revised version, we have modified the relevant description to avoid defining taxusin as the only end product of taxane biosynthesis pathway in *P. chienii*.

(Lines 259-260)

9. Line 296-298: "Due to the absence of TBT, which catalyzes the conversion of 2-debenzoyl-7,13-diacetylbaccatin III to 7,13-diacetylbaccatin III, P. chienii is unable to synthesize 10-deacetylbaccatin III and its subsequent intermediates." Based on this point, the authors should conduct a rescue experiment to support this conclusion, i.e. the authors should integrate the reported functional TBT from Taxus mairei into the tissue of P. chienii and detect the 10-deacetylbaccatin III product by LCMS.

A: Thank you for your professional advice. As a woody gymnosperm, the genetic transformation system of *P. chienii* has not been established, and it is very difficult for us to conduct a rescue experiment to support this conclusion. In the future, we will focused on the genetic transformation system of *P. chienii* and strive to achieve genetic transformation of *P. chienii* as soon as possible.

Here, the accumulation levels of several key intermediate products, including 10-deacetyltaxol, 10-deacetyl-7-xylosyltaxol C, 13-acetyl-9-dihydrobaccatin III, and cephalomannine, were analyzed and provided in Fig. 4g. All these intermediate products could not be detected in *P. chienii*, thus we predicted that *P. chienii* is unable to synthesize 10-deacetylbaccatin III and its subsequent intermediates.

(Lines 284-287)

10. The corresponding author Chenjia Shen of this manuscript has reported that 10-deacetylbaccatin III and paclitaxel were able to be detected in Pseudotaxus chienii in 2021 (BMC Plant Biology, t2021, 21: 104). However, the author claimed that both metabolites were not detected in current manuscript. Please explain the conflicting results.

A: The existence of paclitaxel in *P. chienii* has been a topic of debate for a long time. We have carefully reviewed the previous published papers and found that there are different opinions. Due to the sparse and scattered distribution of the natural population of *P. chienii*, the plant materials in different researchers are in inconsistent, which has led to contradictions between different research results. There is also some controversy over the plant materials used in the research published in 2021 (BMC Plant Biology, t2021, 21: 104). Therefore, the present research aims to provide a standard for molecular identification of *P. chienii* by completing whole genome sequencing.

In the present study, the research materials have been appraised by various experts and several pictures, videos, and plant specimens have been preserved, which confirmed that the plant used in our study is *Pseudotaxus chienii* (W. C. Cheng) W. C. Cheng.

11. In Figure 5A, PcTS1 and PcTS2 cluster together with the bifunctional terpene synthases (TPS) into one large clade, but they were classified as class I TPSs in this manuscript. It is necessary to clarify why PcTS1 and PcTS2 are classified as class I TPSs rather than bifunctional TPSs. The structures of PcTS1 and PcTS2 can be predicted using AlphaFold 3 and their structural similarities (e.g., TM scores) can be compared with those of known class I and bifunctional TPSs to classify

them correctly.

A: Thank you for your professional suggestion. Without considering sequence similarity, Class I TPSs contain motif DDXXD and Class II TPSs contain motif DXDD (Chin J Nat Med. 2021 Sep;19(9):666-674.) Yes, PcTS1 and PcTS2 cluster together with the bifunctional terpene synthases (TPS) into one large clade according to their sequence similarity. But all the TPS proteins were grouped into different classes according to their subtle variation in DXDD motif, which are only contain five amino acid residues. Furthermore, the 3D structure and TM-score were analyzed again, which was used to confirm the structural similarities. The TM-scores showed that PcTS1 displays a higher similarity with the two Angiosperm (Class I) TPSs (<0.18), thus PcTS1 should be grouped into Class I. The data has been provided in **Fig. 5h**.

(Lines 310-312, and 323-327)

12. Line 581, Line 599, “Trans_Decoder” and “Repeat_Masker” should be correctly written as “TransDecoder” and “RepeatMasker”. These are not the only instances; there are many places in the manuscript where software names are incorrectly formatted. It is essential to review the entire text carefully for these errors. Additionally, all software referenced in this manuscript must be properly cited.

A: Thank you for your reminder. Yes, several software names have not been correctly written. All the software referenced in our manuscript were properly cited.

13. Line 213, The legends need to be correctly labeled, Fig 2f should be labeled on line 213. And the following Fig 2e should be adjusted to line 223. Authors are requested to carefully check the legend and its location in the full text.

A: Yes, the legends also are very important. We have carefully checked the Figure legends to make them correct.

(Line 205 and 212)

14. The paclitaxel biosynthesis pathway was incomplete and some steps and new enzymes were missing in Fig3, the authors should add more detailed information based on several recent reports (Jiang et al., Science, 2024; Yang et al., Nature Communications, 2024; Li et al., Angewandte Chemie International Edition, 2024; Zhang et al., Molecular Plant, 2023, etc.).

A: Yes, the paclitaxel biosynthesis in Fig.3 is incomplete and some new enzymes are missing. In the revised version, more detailed information based on several recent reports has been added to the Fig. 3. For examples, several newly identified genes, such as TOT1 (Jiang et al., Science, 2024), CYP725A22-1, CYP725A23-1, T9a oxidase (Zhang et al., Molecular Plant, 2023), CYP725A55, CYP725A37, TAX19, and TCPR (Yang et al., Nature Communications, 2024), and TmCYP1 (Li et al., Angewandte Chemie International Edition, 2024), were added to the pathway in Fig. 3. After carefully sequence checking, we found that TmCYP1 and TOT1 may be the same gene.

In the revised version, a new Fig. 3 has been uploaded. (Lines 269-275)

15. Line 658-659, although the author claims that "Homologous genes in P. chienii were obtained by performing BLASTp searches with identity > 60 and protein length > 700 aa.", many of these proteins are smaller than 700 aa, and at least some genes of the CYP family in the paclitaxel biosynthesis pathway are not. Please explain why this experimental method is inconsistent with the actual situation. It is essential to validate this calculation method to ensure the reliability of the research findings.

A: The of the method of “Identification of taxane biosynthesis-related genes” is incorrect. Homologous genes in *P. chienii* were obtained by using OrthoFinder ver.2.09 with default

parameters. We have corrected this error in the revised version. (Lines 648-649)

16. Supplementary Figure S7 and Figure S10 are at low resolution. Please replace it with a high-resolution version.
A: Yes, the two supplementary files are at low resolution. We have replaced them with high-resolution version.

Reviewer #2 (Remarks to the Author):

The manuscript by Wang et al presents the genome of Pseudotaxus chienii, a relict conifer endemic to China and purports to provide insights into the origin and evolution of taxane biosynthesis. Whilst I am impressed by the first part of this study namely the genome sequence I am less so regarding the origin and evolution of taxane biosynthesis.

My concerns here stem from the fact that they exclusively rely on the biosynthetic pathway published in Science last year - which contains massively fewer proteins than either the pathway proposed by the Fernie group in Molecular Plant in 2023 or that of the Sattely group in Nature Communications in 2024. A proper analyses would require all of the enzymes/genes identified in all three papers to be used in the evaluation. Only then, in my opinion would the study be complete. As a further explanation a recent review by Fernie et al. published in Nature Plants revealed that the heterologous overexpression of the pathway proposed by Fernie resulted in 100 fold higher levels of Baccatin III than that reported in the Science paper. I thus feel that it is compulsory that analyses of a much broader set of candidate genes is performed.
A: Thank you for your suggestion. Considering that the genome of *T. mairei* was published by Jianbin Yan's group, thus our work rely on the biosynthetic pathway published in Science last year, which is also published by the same group. However, Taxol metabolism is a complex pathway several new papers on Taxol pathway were published recently.

In the revised version, more detailed information based on several recent reports has been added to the Fig. 3. For examples, several newly identified genes, such as TOT1 (Jiang et al., Science, 2024), CYP725A22-1, CYP725A23-1, T9a oxidase (Zhang et al., Molecular Plant, 2023), CYP725A55, CYP725A37, TAX19, TCPR, AAE4 (Yang et al., Nature Communications, 2024), and TmCYP1 (Li et al., Angewandte Chemie International Edition, 2024), were added to the pathway in Fig. 3. More proteins from the pathway proposed by the Fernie group in Molecular Plant in 2023 and that of the Sattely group in Nature Communications in 2024 also have been added to our manuscript.
(Fig.3a and lines 269-275)

Reviewer #3 (Remarks to the Author):

This study conducts a chromosome-level genomic analysis of Pseudotaxus chienii to deeply explore the origin and evolution of the taxane biosynthesis pathway in the Taxaceae family. It reveals part of the taxane biosynthesis pathway in P. chienii and offers key insights into functional differences between the Taxus and Pseudotaxus genera during evolution, using comparative genomics, metabolomics, and functional validation.

Major comments:

1. In the section "Visualization of metabolite locations in the stems and leaves of P. chienii", the

MSI results of leaves and stems of P. chienii have little connection to the following experiments, at least the taxane biosynthesis-related metabolites were not detected and visualized. The author should supplement the MSI results for taxane biosynthesis-related metabolites.

A: *P. chienii* is rare gymnosperm with very little research, and its economic value is still unclear. Before studying its genome, I would like to analyze its potential active ingredients through spatial metabolomics. But the logical linkage between this section and the main body of the text is very week (which was pointed by the reviewer 1). Due to database limitations, taxanes was not detected.

In the revised version, the data on flavonoids, coumarins, and steroids were removed from the manuscript. Considering that taxanes are also a type of diterpenoid compound, we only retained the analysis of some potential diterpenoids in *P. chienii* to enhance the logical coherence. (Lines 116-119 and 136-142)

2. In Figure 1c, are whether MALDI-MS spectra attributed to leaves or stems? It is recommended that the mass spectra of both samples be provided and the active ingredients labeled.

A: In Figure 1c, the MALDI-MS spectra is for both stems and leaves. The chip size we use is 6.5*6.5 mm. We placed the leaf and stem materials on the same chip to ensure comparability between them. Thus, only one mass spectra is available. We have added this information to the M&M section to make it clear.

(Lines 515-517)

3. The author should provide the optical images of the P. chienii leaf and stem sections in Figure 1d/e/f/g to better understand the segmentation of tissues. Besides, the author should supplement the information on the matrix used in MSI analysis, and the comparison between MSI results and the matrix background.

A: The optical images of the *P. chienii* leaf and stem sections has been provided in Fig. 1b. No signal of the matrix background was detected. More information on the matrix used in MSI analysis has been added to the M&M section.

(Lines 116-119, 511-513 and 514-517)

4. The clustering analyses of MSI results are confusing; the author should explain the differences between principal components (PCs) separation and the following clustering analyses of SAM. How to determine the attribution of identified metabolites to specific tissues? The clustering analyses identified “1,494 pith-, and xylem-SAMs”, but the pith and xylem are different tissues in PCs analysis; the statement should be “1,494 pith-, and 1531 xylem-SAMs”, corresponding to Figure S2c. Besides, the identified metabolites were clustered into six groups, and the author only explained clusters I to V.

A: Yes, we should explain how to determine the attribution of identified metabolites to specific tissues in the M&M section.

(Lines 525-528)

Thank you for your reminder. There is a mistake in the statement on SAM analysis. We have carefully checked the text corresponding to Fig. S2c to make it correct. For the stem, the cluster VI contains 390 epidermis/xylem-SAMs. For the leaves, the cluster VI contains 336 epidermis/mesophyll-SAM. We have revised the text to make the description complete.

(Lines 126-127, 129-132, and 134-135)

5. According to Tables S1 and S2, the m/z errors of certain metabolites and active ingredients are notably high, including Taxine B (9.68), Cyanidin (8.40), ζ-carotene epoxide (9.71), Presqualene-PP (9.75), and Campesterol glucoside (9.64). These discrepancies could impact the accuracy of

compound identification. The author should critically check the MS results.
A: The newly developed visualization technology has certain shortcomings, and its database is also not large enough. High m/z error may be due to calibration bias or insufficient resolution of the mass spectrometer. In the future, it is recommended to re-evaluate the data preprocessing steps to improve the *m/z* errors.

6. In Figure 1h, the author calculated the proportion of different active ingredients in stems and leaves. The authors should explain the significance of the calculated proportion, as the results showed that leaves also contained more active ingredients (even higher than stems), while the twigs from a wild *P. chienii* tree were collected for the subsequent genome sequencing.
A: The method of quantitative analysis is missing. Receiver Operation Characteristic analysis and Student's t test analysis were applied to determine the significance of differences between the samples. Corrected *p* value < 0.01 were used to screen significant changed metabolites. In our study, *P. chienii* twigs, including tender stems and attached leaves, were collected for the genome sequencing. We have added this information to the M&M section.
(Lines 525-528)

7. In the section "Genome assembly and annotation", there appears to be no mention of genome annotation.
A: Thank you for your reminder. A paragraph of gene annotation was added to the section "Genome assembly and annotation".
(Lines 161-165)

8. Please further explain "Most of the CYP725 genes are situated within a very restricted genomic region (Fig. 6), which may account for the consistent response of CYP725 to MeJA treatments" to better understand the underlying mechanism. Additionally, the author should supplement a reference to support the statement "MeJA is a significant elicitor that was commonly used to enhance the production of Taxol".
A: Thank you for your suggestion. Gene expression is regulated by various *cis*-acting elements, and the distribution of these elements within the plant genome is uneven. Consequently, genes that are clustered together may be influenced by the same regulatory elements and display similar expression patterns. We have added this explanation to the text.
In *Taxus* cell suspension cultures, MeJA elicitation is an effective strategy to induce and enhance biosynthesis of Taxol. A reference supporting that MeJA is a significant elicitor has been added to the text.
Patil RA, Lenka SK, Normanly J, Walker EL, Roberts SC. Methyl jasmonate represses growth and affects cell cycle progression in cultured Taxus cells. Plant Cell Rep. 2014 Sep;33(9):1479-92.
(Lines 476-478 and 481-484)

9. In lines 347-349, "Further analysis of protein structures demonstrated that TPSs from the Taxus genus, along with PcTS1, exhibit a greater structural similarity to angiosperms than to other gymnosperms (Fig. 5g)." It is suggested to distinguish the structural similarities and differences between Taxus TPS and angiosperms in the figure, and make a more intuitive comparison in combination with protein models from other gymnosperms species. Additionally, the protein structures of the Taxus genus, TmTXS, and TbTDC1, show a significant similarity and seem to be the same image (Please double check these two figures). The author should discuss it or point out the differences between two protein structures.
A: Thank you for your professional suggestion. Class I TPSs contain motif DDXXD and Class II

TPSs contain motif DXDD (Chin J Nat Med. 2021 Sep;19(9):666-674.) Yes, PcTS1 and PcTS2 cluster together with the bifunctional terpene synthases (TPS) into one large clade according to the phylogenetic analysis. Furthermore, the 3D structure and TM-score were analyzed again, which was used to confirm the structural similarities. The TM-scores showed that PcTS1 displays a higher similarity with the two Angiosperm (Class I) TPSs (<0.18), thus PcTS1 should be grouped into Class I. The data has been provided in **Fig. 5h**. We have checked the sequences of TmTXS, and TbTDC1 again. TmTXS and TbTDC1 are the same protein, which are from *T. media* and *T. baccata*, respectively, thus they show a significant similarity.
(Lines 310-312 and 323-327)

10. The conclusion in the "Whole genome duplication (WGD)" section that P. chienii has not experienced recent WGD is reasonable. However, the evidence for ancient WGD is insufficient. Although a weak Ks peak was observed, its timing was not verified to be consistent with the evolutionary history of P. chienii; furthermore, Ks values are susceptible to heterogeneous evolutionary rates or sequencing errors.
A: We agree with your opinion. As we could not provide a valid evidence for ancient WGD, we deleted this conclusion.

11. The article used 184 single-copy genes for phylogenetic analysis and combined the MCMCTREE program to estimate divergence times. It inferred that P. chienii and Taxus diverged approximately 48.3 million years ago. However, the article did not provide detailed information on the calibration points or methods used, which may affect the accuracy of the divergence time estimation.
A: Divergence times between *O. sativa* and *G. biloba* (326.4 - 336.8 million years ago [Mya]), *G. montanum* and *C. panzhihuaensis* (275.3 - 316.4 Mya), and *A. trichopoda* and *N. colorata* (179.9 - 205.0 Mya) were used as fossil calibration points. We added this information into the Methods section.
(Lines 615-619).

12. The author should provide the duplication levels of transgenic experiments.
A: For the transgenic experiments, two independent repeats were performed. The data of the other one repeat was provided in **Supplementary Fig. 10a-b.**

13. In the section “Functional identification of the TS genes in P. chienii”, the new peak observed at an RT of 11.875 mins matched exactly with the analog of taxadiene, eicosane. However, the mass spectra lack sufficient information. The author should label the analog of taxadiene and eicosane in the MS profile and provide more information for mass spectrum annotation. Furthermore, the mass spectra of standard substances for the analog of taxadiene and eicosane should be included.
A: Yes, more information should be provided for the mass spectra. Although functional identification of the TS genes in *P. chienii* is very important, no standard of taxadiene is available. In our study, eicosane (an analog of taxadiene) was used as a standard to analyze taxadiene. The mass spectra of standard substances for the analog of taxadiene and eicosane has been provided in **Supplementary Fig. 10c**.

14. Annotating taxane biosynthesis-related genes based solely on sequence similarity is feasible but has limitations. To enhance the accuracy and reliability of the annotations, it is suggested to validate the functions of key enzymes involved in taxusin biosynthesis through in vitro activity assays. Additionally, metabolomic analyses should be conducted to detect other intermediate products

accumulated in P. chienii, thereby further confirming whether taxusin is the end product of its taxane biosynthesis pathway.

A: Thank you for your professional advice. In the present study, most of the taxusin biosynthesis-related genes have not been cloned yet. In the future, we will focus on the functions of key enzymes involved in taxusin biosynthesis through in vitro activity assays. In the revised version, untargeted metabolomic analyses were performed to detect other intermediate products accumulated in *P. chienii*.

The accumulation levels of several key intermediate products, including 10-deacetyltaxol, 10-deacetyl-7-xylosyltaxol C, 13-acetyl-9-dihydrobaccatin III, and cephalomannine, were analyzed and provided in Fig. 4g. All the intermediate products could not be detected in *P. chienii*, thus we predicted that taxusin is one of the end products of taxane biosynthesis pathway.

(Lines 285-287)

15. It is recommended to italicize the Latin names of species in Figure S7 to conform to academic conventions; please enhance the resolution of Figure S10/S2b/S3b/S8a or optimize the image quality for clearer data presentation. There is an error in the figure numbering in Figure S12, and it is suggested to carefully check and correct it. There are also many traces in the image, including border residues that are recommended for modification, such as Figure S9. Besides, the format of labels in Figures 4c and 4d should be the same.

A: Yes, there are many small errors in the Figures. In the revised version, we have carefully checked all the Figures and supplemental Figures to make them correct. All the traces in the images were carefully removed.

Reviewer #1 (Remarks to the Author):

The authors have addressed some my questions in the previous round. However, there are still several points that need improvement.

1. The protein annotation results provided by the authors (https://figshare.com/s/1a98960b9985e2a0c198) were used to count the number of genes and evaluate the protein BUSCO (BUSCO v5.4.5, embryophyta_odb10). The results showed that the number of genes was 33789 and the BUSCO of proteins was 84.9%. This result is still inconsistent with the author‘s revision (BUSCO:90.5%; number of genes :35404), and the quality of the gene annotation results is much lower than the results in the author's revised manuscript.

A: We are sorry for the inappropriate file and description of the BUSCO value. We have updated the latest data with a new link https://figshare.com/s/67260743782347f41dbf.

(Lines 750-752)

The BUSCO of proteins was 85.2% with the embryophyta_odb10 database (last update 2024-01-08). We also compared the BUSCO value of *P. chienii* with those of other gymnosperms and the results were recorded in Supplementary Table S4, which showed that the quality of our genome assembly ranks in the upper-middle range among gymnosperms.

(Lines 156-158 and 575-577)

2. The genome annotation method provided by the author in the manuscript is to use EvidenceModeler to integrate the results of three annotation strategies: transcriptome-based prediction and homology-based gene prediction and ab initio prediction to obtain the final result. If this method is followed, the second column in the gene annotation result gff3 file should be a unified label, such as EVM. If this method is followed, the second column of the gene annotation gff3 file should be a unified label, such as EVM. However, in the gff3 file provided by the author, there are 5 labels in the second column (AUGUS, exonerate, GlimmerHMM, maker and transdecoder). It is more like artificial selection and manipulation.

A: We used Maker to integrate data instead of EVM, we did not recognize the wrong description in the Methods. We are sorry for our carelessness and thank you for pointing out this mistake. The second column showed 5 labels because we traced the source of gene prediction, which means that the specific software provides the best gene model. The label “maker” indicates the integration of gene models from multiple software. Considering that this might lead to misunderstandings among reviewers or readers, we changed the labels to the unified label “maker” in the second column.

Also, in the Methods section, we cited a reference which described the gene prediction pipeline in more details.

Reference: Leng, L. et al. Cepharanthine analogs mining and genomes of Stephania accelerate anti-coronavirus drug discovery. Nature Communications 15, 1537 (2024)

(Line 591)

3. In the author's response to comment 8 from “Reviewer #1”, the issue raised in my previous comment remains unresolved. The key concern is that the presence of T2OH in P. chienii indicates the potential for further conversion of taxusin into downstream metabolites such as 2α-hydroxytaxusin. Meanwhile, more data should to be provided to support the conclusion of “Considering the presence of both T2OH and T7OH, taxusin can be hydroxylated into 2α,7β-dihydroxytaxusin.” in P. chienii. Although, the acquirement of standard compound and the establishment of the genetic transform system are difficult, some works are able to be done. For example:

1) Heterologous expressing in tobacco with the substrate feeding or purifying the protein and

conducting a in vitro enzyme activity assays to prove that the T2OH or T7OH in P. chienii have the same or different function with that from other Taxus species.

2) showing the protein sequence identity or evolutionary relationship of T2OH or T7OH in P. chienii with that from other Taxus species.

A: Yes, more data should be provided to support the conclusion. In our lab, the standard of taxusin was available, but the standard of 2α,7β-dihydroxytaxusin was unavailable. It is difficult for us to check the enzyme activities of T2OH or T7OH in *P. chienii*. After carefully sequence analysis, we confirmed that the T2OH or T7OH in P. chienii have high similarities to that from *T. mairei*.

In the revised version, the protein sequence identity was analyzed using Needleman-Wunsch alignment method. Furthermore, multiple Sequence Alignment of T2OH or T7OH in *P. chienii* with that from other Taxus species were analyzed. The data was provided in the supplementary Figure 10.

Needleman-Wunsch alignment of two sequences				
T2OH (ctg2120_gene.6) from *T. mairei*				
	NW Score	Identities	Positives	Gaps
PchiChr11G298350.1	2058	406/506(80%)	439/506(86%)	13/506(2%)
PchiChr11G298480.1	2161	420/506(83%)	458/506(90%)	12/506(2%)
PchiChr1G033150.1	2123	414/513(81%)	447/513(87%)	11/513(2%)
PchiChr1G033160.1	2101	407/505(81%)	442/505(87%)	10/505(1%)
T7OH (ctg2120_gene.1) from *T. mairei*				
	NW Score	Identities	Positives	Gaps
PchiChr11G298490.1	2166	425/502(85%)	456/502(90%)	9/502(1%)
PchiChr11G299570.1	2185	427/504(85%)	459/504(91%)	11/504(2%)
T7OH (ctg2120_gene.2) from *T. mairei*				
	NW Score	Identities	Positives	Gaps
PchiChr11G298490.1	2208	433/502(86%)	462/502(92%)	9/502(1%)
PchiChr11G299570.1	2227	435/504(86%)	465/504(92%)	11/504(2%)

(Lines 277-278)

4. In the author's response to comment 9 from “Reviewer #1”, the key point is to prove the directly relationship of the absence of TBT and the deficiency of 10-deacetylbaccatin III or baccatin III. Although, the absence of TBT maybe the reason of the undetected 10-deacetylbaccatin III or baccatin III, but other potential reasons should be pointed out. The absence of several taxanes can’t support the conclusion that the inability of P. chienii to synthesize 10-deacetylbaccatin III and its subsequent intermediates is solely due to the absence of TBT. Moreover, the sentence of the response “All these intermediate products could not be detected in P. chienii, thus we predicted that P. chienii is unable to synthesize 10-deacetylbaccatin III and its subsequent intermediates.” seemingly implies that the taxanes of 10-deacetyltaxol, 10-deacetyl-7-xylosyltaxol C, 13-acetyl-9-dihydrobaccatin III, and cephalomannine are the intermediate products of 10-deacetylbaccatin III and results in the misunderstands. Meanwhile, the 13-acetyl-9-dihydrobaccatin III hasn’t the side-chain and it is seemingly inappropriate to classify it with other three taxanes.

A: Thank you for your reminder. Yes, the absence of several taxanes cannot support the conclusion. More potential reasons have been provided to discuss this point. There may be another pathway that consumes an amount of 10-deacetylbaccatin III or baccatin III, making them undetectable. This potential reason has been added to the revised discussion section.

There is a problem in the sentence “All these intermediate products could not be detected in P. chienii, thus we predicted that *P. chienii* is unable to synthesize 10-deacetylbaccatin III and its subsequent intermediates.” To avoid the misunderstands, we have revised the sentence as follow:

“All of these intermediate products could not be detected in *P. chienii*. Therefore, we predict that *P.*

chienii is unable to synthesize 10-deacetylbaccatin III and the subsequent intermediates following 10-deacetylbaccatin III."
Yes, the 13-acetyl-9-dihydrobaccatin III hasn't the side-chain, and we deleted the data of 13-acetyl-9-dihydrobaccatin III.
(Lines 272-277 and 290-291)

5. In the author's response to comment 10 from "Reviewer #1", the contradiction between author's 2021 publication (BMC Plant Biology, t2021, 21: 104) and the current manuscript remains unresolved. The key issue is not whether P. chienii is a debated species, nor whether author's current sample has been authenticated, but why author own previous work reported the presence of 10-deacetylbaccatin III and paclitaxel while the current study reports their absence. What needs to be said more is that the author has negated the previous results by questioning the accuracy of the 2021 research materials. This approach is unconvincing, especially under the premise that both studies are led by the same corresponding author. If there are problems with the 2021 materials, please clearly state how the materials were identified in the study and any specific errors that may have existed, and explain why they were not explained or corrected at the time. If there are methodological biases or analytical misjudgments in the 2021 results, this should also be fully explained. Without a clear and scientifically grounded explanation, the conflicting results raise concerns about data reproducibility and the reliability of metabolite profiling in this species.
A: Yes, it is a serious problem of the contradiction between the 2021 publication and the current manuscript. The contradictory outcome may cause confusion for readers. The methods used in the two studies are consistent, so it must be due to differences in materials. We should clearly point this out instead of hastily denying the previous materials.

We have reviewed several papers on the study of *P. chienii* and found that there may be significant differences between different natural populations. The materials for the 2021 study were harvested from a wild population located in Mount Tianmu (30.0°18.0′30.0″N, 119.0°23.0′47.0″E), Hangzhou, and the materials for the present study were harvested from another wild population located Nan Jian Yan Scenic Area (28.0°20.0'39.19''N, 119.0°6.0'0.97''E), Lishui. Considering the difference in taxanes between the two materials, we designated the material used in the present study a variant of *Pseudotaxus chienii*, *Pseudotaxus chienii var. Nanjianyan*. The present study was conducted based on this variety of *P. chienii*.

To avoid confusing readers, we have explained this issue in the discussion and M&M sections, respectively.
(Lines 395-404 and 517-521)

6. In the author's response to comment 11 by Reviewer #1, the authors state that TM-scores less than 0.18 indicate higher similarity with Class I TPSs. However, this interpretation is incorrect. TM-scores range from 0 to 1, with values above 0.5 generally indicating significant structural similarity, and scores below 0.2 typically reflecting random structural similarity. Therefore, claiming that a lower TM-score (e.g., <0.18) indicates greater structural similarity is inaccurate.
A: I am sorry that I didn't understand the meaning of TM-value before. I have relearned the definition and calculation method of protein TM-values, and used the alphafold3 online tool to recalculate the TM values. The updated values was added to the Figure 5h.

	OsCPSsyn	ScCPS	GrTPS4	AtEKS	PsTPS-LAS	AgAS
PcTS1	0.731	0.739	0.765	0.772	0.838	0.795

(Lines 329-331)

Reviewer #2 (Remarks to the Author):

I am fully satisfied that this resubmission addresses all the concerns I raised in my review of the previous submission.

Reviewer #3 (Remarks to the Author):

1. In the author's reply to the 9th comment from Reviewer 1, Fig. 4g should be Fig. 4f.
A: Thank you for your reminder. We have corrected this error.
2. In the author's reply to the 5th comment from Reviewer 3, the author stated that high mass errors may be due to calibration bias or insufficient mass spectrometer resolution. They still need to provide fragment (MS/MS) data of the key compounds for compound identification.
A: We consulted the company providing technical services, confirmed the experimental parameters. We believed that the mass error of these substances is not significant (within 10 ppm), which belongs to the normal error range for qualitative substances. Space metabolomics generally do not perform secondary mass spectrometry, therefore there is no corresponding MS/MS data available
TEMI: tissue-expansion mass-spectrometry imaging. Nat Methods. 2025 May;22(5):1051-1058.
3. The author should supplement the scale bar for Figure 1b/d. Besides, the author should check the size of the MALDI-2 chip. The 6.5*6.5 mm chip might be small for the leaf and stem materials.
A: The scale bars for Figure 1b/d have added in the revised version. The correct size of the MALDI-2 chip is 2.5*7.5 cm. We have corrected this error in the revised version.
(Lines 531-532)

Reviewer #1 (Remarks to the Author):

A:Thank you for your expertise. Your modification suggestions have greatly improved the quality of our article.

1. In the author's response to comment 5 from “Reviewer #1”, the response does not address the key scientific concerns raised in previous comments. The explanation is arbitrary and lacks the necessary scientific rigor.

The discrepancy between the two studies from the same research group regarding the detection or presence of paclitaxel is significant and cannot be dismissed as minor. Without a thorough and rigorous reconciliation of these conflicting findings, the credibility of both studies is compromised. The authors attempt to address the inconsistency between the 2021 and current results by suggesting that the material analyzed in the present study represents a new variety, Pseudotaxus chienii var. Nanjianyan. However, this claim lacks support from morphological, genetic, or chemotaxonomic evidence. In the absence of formal diagnostic criteria and designated voucher specimens, the proposed varietal name is scientifically unwarranted and should be retracted unless substantiated by appropriate data.

A: Yes, your opinion is very reasonable. The discrepancy between the two studies from the same research group is unacceptable. Due to the lack of morphological, genetic, or chemotaxonomic evidence, the claim regarding a new variety, *Pseudotaxus chienii* var. Nanjianyan was retracted in the revised version.

Over the past two months, we have repeatedly checked the presence of paclitaxel in *P. chienii*. We are now convinced that paclitaxel is not present in *P. chienii*. Therefore, we must acknowledge an error in the sampling process in our previous article. We have clearly pointed out this error in the current manuscript to prevent any potential misunderstanding among readers. In 2021, lacking experience in identifying P. chienii, we conducted sampling during a season when the plants bore no fruit, which was a mistake. We may have inadvertently confused some Taxus specimens for P. chienii. Since *Taxus* and *Pseudotaxus* are very similar, sampling should be carried out in autumn to confirm the species based on the color of their fruits. We have included this important clarification in the revised manuscript. Furthermore, we will to contact the editor of BMC Plant Biology to discuss how to correct the errors in the previous article.

(Lines 399-404 and 522-524)

2. In the author's response to comment 3 of "Reviewer #1", the authors avoided using misleading statements and provided data on protein sequence identity, which did not help much with the key issues. The authors should conduct an experiment to heterologously express T2OH and T7OH in tobacco fed with taxusin and use T2OH and T7OH expressed in *T. mairei* as controls.

A: Yes, a heterologous expression experiment should be performed. In the revised version, we cloned the full-length sequences of *P. chienii* T2OH (PchiChr11G298480.1) and T7OH (PchiChr11G298490.1), and *T. mairei* T2OH (ctg2120_gene.6) and T7OH (ctg2120_gene.1). A feeding experiment using Taxusin as the substrate was conducted. Considering the unavailability of 2α,7β-Dihydroxytaxusin standard, we assessed the activities of T2OH and T7OH by measuring the decrease in Taxusin content. The target genes were individually transiently expressed in *N. benthamiana* leaves via *Agrobacterium*-mediated transformation. After four days of cultivation, Taxusin was fed to the *N. benthamiana* leaves, which were then collected after an additional 1 day of incubation. Our data showed that the T2OH and T7OH enzymes from *P. chienii* exhibit activities similar to those from *T. mairei*.

(Lines 276-278 and 705-717)

3. In the author's response to comment 6 from “Reviewer #1”, in the section "Protein structure analysis of Taxaceae TPSs", the authors identified PcTS1 and PcTS3 as ClassI based on Figure e, but there is no PcTS3 in Figure a, only PcTS1 and PcTS2, and the branch where PcTS1 and PcTS2 are located is closer to the branch of bifunctional enzymes than other ClassI, which means that PcTS1 and PcTS2 belong to bifunctional enzymes. Similarly, the structural analysis of Figure h shows that TS1 is a bifunctional enzyme. All these results are contradictory, and it is difficult to judge which one is correct.

A: I am sorry for this mistake in Figure 5e. The correct name is PcTS2, not PcTS3.

Yes, this is a very interesting phenomenon. My description may not be accurate enough, so it has caused misunderstanding. Previous study pointed out that TS protein in *T. brevifolia* is a Class I diTPS (Med Res Rev. 2021 Nov;41(6):2971-2997.) The classification of TPS is mainly based on the presence of two motifs (DXDD and DDXXD/E). PcTS1/2 does not contain the “DDXD” motif, so it can not be considered as bifunctional TPS, although they are closer to class I members in sequence. We have revised the text to make it clear.

(Lines 465-471)

4. In figure3a, 1-Dehydroxybaccatin IV cannot be directly synthesized from Baccatin I. In addition, the structural formulas of 2α,7β-Dihydroxytaxusin, Taxadiene hexa-acetate, 1-Dehydroxybaccatin IV, and Baccatin I are in opposite directions.

A: I am very sorry that I did not read the literature carefully enough. Thank you for your reminder. 1-Dehydroxybaccatin IV cannot be directly synthesized from Baccatin I. We have corrected this error in the revised version. The directions of the structural formulas of 2α,7β-Dihydroxytaxusin, Taxadiene hexa-acetate, 1-Dehydroxybaccatin IV, and Baccatin I have been corrected. In the revised version, a new Figure 3 has been updated.

5. Line156-line157, it should be made clear that the object of this BUSCO assessment is the genome or the protein sequence.

A: The object of thus BUSCO assessment is the protein sequence. In the revised version, we have revised the text to make it clear.

(Lines 155-156)

Reviewer #3 (Remarks to the Author):

I agreed with Reviewer 3's comment that MS/MS data should be supplemented. In situ MS/MS is quite significant for compound identification in MS imaging. Due to the lack of LC separation, peak assignment is always ambiguous and may be overlapped with matrix or background ions. Therefore, only MS1 data cannot guarantee the accuracy of the results. Unlike the mouse brain (Nat Methods. 2025 May;22(5):1051-1058), a model tissue for MSI has been thoroughly interrogated and public databases are available; identification of plant secondary metabolites needs more efforts, particularly for key metabolites.

A: Yes, the MS/MS raw data is very important for plant study. In the revised version, MS/MS data has been uploaded to the Metaspace database with project ID “m-2025”. Reviewers can view the MS/MS data by the reviewer link (https://metaspace2020.org/api_auth/review?prj=07a71fa8-5c9a-11f0-a049-b3ec65131efb&token=KzN_YV6IGpbf).

(Lines 770-773)

Reviewer #1 (Remarks to the Author):

This manuscript has gone through several rounds of revision. While its quality has greatly improved, some issues remain as follows:

1.Several key inconsistencies need to be addressed and corrected throughout the paper. For example, the Abstract and Results sections state the genome size as "15.6 Gb assembled" (line 33) and "15.5 Gb (99.0%) anchored" (lines 148–149), respectively, while the Introduction states, "It contains 16.79 Gb of data, of which 15.45 Gb were assigned to 12 pseudochromosomes" (line 105).

A: Yes, this inconsistency should be corrected. We have revised the manuscript to make them consistent.

(line 105)

2.The authors reanalyzed their previously published results and explained the conflict in results about whether paclitaxel is present in P. chienii. It is recommended to include the specimen IDs of the materials they had used to improve the reliability of the findings.

A: Thank you for your understanding. A specimen ID of our plant material should be provided in the revised version. We have added a specimen ID (ZQBX) to the Method section.

(Line 520)

3.Pathway intermediates claimed in metabolite detection (e.g., taxusin) need to be verified by targeted LC-MS/MS. Supplementary materials should include retention times and MS/MS spectra of taxusin in P. chienii samples and standards, along with LOD/LOQ, ion ratios, and recoveries. Furthermore, Figure S9e shows similar accumulation levels of taxusin across different groups, but only the peak for the control group "pBWA + Taxusin" is detected in the figure, which is clearly unreasonable and needs to be explained regarding how to quantify the residual taxusin.

A: The contents of taxusin has been verified by targeted LC-MS/MS analysis. Using MRM method, we measured the difference between *Taxus* and *P. chienii*, and the effect of MeJA on the content of taxusin in *P. chienii*. Recently, the instruments in our school have broken down. In the follow-up supplementary experiments on T2OH and T7OH activities were performed by the instrument of the Chinese Academy of traditional Chinese medicine. Therefore, the parameters of the two batches of experiments are slightly different.

The spectra and retention times of taxusin in *P. chienii* and *T. wallichiana* samples were shown in Fig. 4d and Fig. 6e. The spectra, ion ratios, and recoveries of *N. benthamiana* samples was provided in Fig. S10. The transition of *m/z* 527.3→467.2 was used for taxusin quantitative analysis and the transitions of *m/z* 527.3→407.2 were utilized for qualitative analysis. LOD is 2 ng/mL. This important information has been added to the M&M section.

(Lines 735-738)

The Figure S9e is relatively vague. Because we are not very familiar with the software of spectral analysis, there is a problem in the interception of spectra. In the revised version, we attached the original MS spectra in Figure S10 to make the data more accurate.

(See Figure S10)

4.This paper contains several errors and contradictions regarding the classification of TPS sequences. Initially, it classifies TPSs with DDXXD/E motifs as class I and those with DXDD motifs as class II (lines 83-84). However, the paper then states that only Taxaceae TPSs lacking the "DDXD" motif belong to class I and further describes the core DXDD motif as characteristic of class I TPSs (lines 467-474), which contradicts the widely accepted classification approach, as DXDD is characteristic of class II TPSs.

A: Thank you for your reminder. There is a mistake in our description in the discussion section (Lines 467-474). The core DXDD motif is the characteristic of class II TPSs, not class I TPS. We have revised this paragraph to make it correct.
(Lines 468 and 472)

5.The manuscript still has several typographical and grammatical errors, including “signa” corrected to “signal” (Line 534), “tblasn” corrected to “blastn” (Line 590), “we h must acknowledge” (Line 400), and “should be conducted out” (Line 403).
A: We have carefully checked the manuscript again to correct these typographical and grammatical errors.

Reviewer #2 (Remarks to the Author):

I carefully reviewed the revised manuscript by Wang et al and especially the comments to the last round of reviews to which I did not contribute to. I was generally very happy with the responses provided. However, I must admit I found the comment that "we are not really experts in the use of the software rather worrying. However, looking at the Supplementary Figure it would appear that there annotation is likely correct and I am guessing that this was just unfortunate use of English since the annotation makes a lot of biological sense and also seems to be technically correct. Should the authors generally have meant that they were unsure with the software I would however suggest that they contact the vendor and ask for their help in confirming the annotation.

A: Thank you for your kindness! The correctness of software use and MS analysis is very important. We asked experts from the Chinese Academy of traditional Chinese medicine (Jian Yang and Ping Su) to help check our MS data and software use, and they also confirmed that there was no problem with the analysis and identification of compounds. Furthermore, we contacted the vendor to help us confirm the MS data. They also conformed that our annotation is correct.