WILEY-BLACKWELL

Cladistics (2018) 1–8

Cladistics

10.1111/cla.12362

RNAseq-based phylogenetic reconstruction of Taxaceae and Cephalotaxaceae

Aasim Majeed, Amandeep Singh, Shruti Choudhary and Pankaj Bhardwaj*

Molecular Genetics Laboratory, Department of Plant Sciences, Central University of Punjab, Mansa Road, Bathinda, 151001, India

Accepted 15 October 2018

Abstract

Taxaceae and Cephalotaxaceae are the two economically important conifer families. Over the years there has been much controversy over the issue of merging these families. The position of *Amentotaxus* and *Torreya* is also ambiguous. Some authors consider them closer to Taxaceae while others deemed them to fit within Cephalotaxaceae. Still, others prefer to raise them to their own tribe. Different morphological, anatomical, embryological and phylogenetic evidence supports one or the other view, making the precise delineation between them unresolved. Here we used an RNAseq-based approach to obtain orthologous genes across the selected species to reconstruct a more robust phylogeny of these families. A total of 233.123 million raw reads were de novo assembled to generate nine different transcript assemblies for the corresponding species. Of the 940 191 assembled transcripts across nine species, we generated 409 734 unigenes, which were clustered into orthologous groups. A total of 331 single-copy complete orthologous groups were selected for phylogenetic analysis. Maximum-likelihood, maximum-parsimony and Bayesian phylogenetic trees showed a sister relationship between Taxaceae and Cephalotaxaceae. Our analysis supports their distinctiveness at the family level and also shows that *Amentotaxus* and *Torreya* fit within Cephalotaxaceae.

Introduction

Gymnosperms are vascular plants that bear naked ovules. Among the largest gymnosperm groups, pines stand at the forefront while *Ginkgo* is the smallest group represented by just a single species. The Plant List, a database collaboratively produced by the Royal Botanic Gardens, Kew, and Missouri Botanical Garden, lists 10 40 426 species of plants belonging to 16 167 genera and 620 families. In this list, the Gymnosperms have been divided into 4588 species, 14 families and 88 genera (http://www.theplantlist.org). However, the Gymnosperm database (http://www.conifers.org) records only 13 families of gymnosperms. In a broad sense all vascular seed-bearing plants belong to either of the five lineages, namely angiosperms, coniferales, cycadales, ginkgoales and gnetales. Among other important families in the coniferales lineage, Taxaceae and Cephalotaxaceae are of note with regard to their economic importance and taxonomic confusion. Both Taxaceae and Cephalotaxaceae are closely related to each other. Taxaceae includes the well-known genera *Taxus*, *Pseudotaxus* and *Austrotaxus* while Cephalotaxaceae constitutes a single genus *Cephalotaxus*. The genera *Amentotaxus* and *Torreya* are viewed differently by different authors. They bear resemblances to as well as differences from both Taxaceae and Cephalotaxaceae and also with each other. This led to varied opinions regarding their taxonomic position. Due to contradicting proposals by different studies, a clear-cut delineation between Taxaceae and Cephalotaxaceae is not fully resolved and the current debate is whether they should be merged or treated as separate families.

Cephalotaxus was earlier considered as a member of Taxaceae (Eichler, 1889; Van Tieghem, 1891). Pilger (1903) also included the genus *Cephalotaxus* in the family Taxaceae. Although given the status of a family by Neger (1907), *Cephalotaxus* was merged back

*Corresponding author:
E-mail address: pankajihbt@gmail.com; pankajbhardwaj@cup.edu.in

within Taxaceae by Christenhusz et al. (2011) between *Taxus* and *Amentotaxus*. Both Cephalotaxaceae and Taxaceae were considered to have no alliance with each other (Florin, 1948, 1954; Singh, 1961). This opinion was refuted by Takhtajan (1953), Hart (1987), Price (1990) and Stefanoviac et al. (1998), who opined that both are closely related. Phylogenetic analysis based on nuclear genes, *NFY* and *NLY*, revealed *Cephalotaxus* as sister to Taxaceae but seemed to be nested in Taxaceae when only *NLY* was used (Lu et al., 2014). A sister relationship was also inferred between the two by studies utilizing chloroplast *matk* and the nuclear *rDNA ITS* region (Cheng et al., 2000), chloroplast *matk* (Wang and Shu, 2000), and chloroplast *matK*, *rbcL*, *trnL-trnF* spacer, *psbA-trnH* spacer and nuclear *ITS* (Xiao et al., 2008). However, based on *rbcL* and *matK*, Quinn et al. (2002) suggested that the two should be merged. Cladistic analysis based on morphology, leaf and wood anatomy revealed that Cephalotaxaceae should be merged within Taxaceae (Ghimire and Heo, 2014). While revisiting the taxonomic position of Cephalotaxaceae, Lang et al. (2013) agreed with Christenhusz et al. (2011) to retain Cephalotaxaceae within Taxaceae and followed the same classification. Thus, different studies seem to propose different taxonomic classifications of these families, and thus their precise delineation remains unresolved. *Amentotaxus* was considered as a member of Cephalotaxaceae (Pilger, 1926; Page, 1990). Based on their anatomical characteristics, *Amentotaxus* and *Torreya* were found closest to Cephalotaxaceae (Hu and Wang, 1989). However, phylogenetic analysis by Chaw et al. (1995, 1997) asserted that they are better resolved within Taxaceae. Based on fleshy structure in their reproductive parts, *Amentotaxus* has been considered as a link between Taxaceae and Cephalotaxaceae (Keng, 1969; Wilde, 1976). Whether *Amentotaxus* is closer to Cephalotaxaceae or links Taxaceae and Cephalotaxaceae, Cheng et al. (2000) regarded both Cephalotaxaceae and Taxaceae as sister groups; *Amentotaxus* and *Torreya* are closer to Taxaceae than to Cephalotaxaceae based on chloroplast *matk* and nuclear *rDNA ITS* region. *Amentotaxus* has also been suggested to be raised to its own family, Amentotaxaceae, due to sufficient distinction from the genera of Taxaceae (Kudo and Yamamoto, 1931; Chuang and Hu, 1963; Xi, 1986).

RNA sequencing (RNAseq) has recently emerged as a revolutionary technology. Besides its immense potential in diverse applications, its utility in the field of phylogenetics has recently been recognized, although the approach is still in its infancy. The constantly decreasing cost of next-generation sequencing and evolution of sophisticated tools for phylogenetic analysis have favoured the use of genomic-scale data in phylogenetics. Recently, robust phylogenies have been reconstructed using RNAseq data in mosquitoes (Hittinger et al., 2010), lice (Johnson et al., 2013) and ray-finned fish (Zou et al., 2012). The validity of using RNAseq data to recapitulate previous phylogenies has been shown for *Flaveria* (Lyu et al., 2015) and Orchidaceae (Deng et al., 2015). Thus, RNAseq data are becoming a new resource for phylogenetic analysis. Given the lack of clear taxonomic position between the genera of Taxaceae and Cephalotaxaceae, here we used an RNAseq-based approach to recover orthologous genes between the selected species and subsequently utilize them for phylogenetic analysis.

Materials and methods

We used in-house sequenced Illumina-based paired-end RNAseq raw reads of *Taxus wallichiana* together with the raw reads of eight species (Table 1) retrieved from the NCBI Sequence Read Archive (SRA) database for analysis. A schematic representation of the pipeline used for the orthology detection and phylogenetic analysis is shown in Fig. 1. The raw reads were cleaned by removing the adaptor sequences and low-quality bases through Trim Galore v0.4.1. The quality-trimmed reads were then de novo assembled using Trinity v1.6 (Grabherr et al., 2011). The completeness of the assembly was analysed using BUSCO v2 (Simão et al., 2015). The raw reads were then mapped back to the assembly using Bowtie2 v2.3.0 (Langmead and Salzberg, 2012) to assess the read representation for quality assessment. Quality was also checked by counting full-length transcripts through BLAST against the SwissProt database. Removal of the redundant sequences and generation of unigenes was done through CD-HIT-EST v4.6 at 95% sequence identity threshold (Li and Godzik, 2006). Non-redundant assemblies were then compared with each other by executing reciprocal one-to-one BLAST at an e-value of 1×10^{-5} for the identification of orthologous groups across the species through an orthology detection tool, Proteinortho v5.16b (Lechner et al., 2011). All degenerate or incomplete orthologous groups were filtered out to obtain only complete groups which contain no more than one gene from each species. The sequences of these single copy orthologous groups were processed to prepare the input files for alignment using in-house python scripts and then aligned individually by MAFFT v7.312 (Katoh and Standley, 2013). The individual aligned sequences were concatenated by AMAS v0.98 (Borowiec, 2016) to generate a super-alignment file. Trimming of spurious sequences and poorly aligned regions from the super-aligned sequences was performed using TrimAl v1.2 (Capella-Gutiérrez et al., 2009). We used a multiple phylogenetic approach in the reconstruction of phylogenetic

trees. Maximum-parsimony (MP) analysis was run on MEGA v7.0.26 (Kumar et al., 2008) using the SPR search method. Support for individual clades was achieved through bootstrapping with 500 replicates. A maximum-likelihood (ML) phylogenetic tree was reconstructed in IQ-TREE v1.5.5 (Nguyen et al., 2014) with Ultrafast Bootstrap support (Minh et al., 2013; Hoang et al., 2017) and SH-aLRT branch support for 10 000 replicates. For model testing, we used ModelFinder (Kalyaanamoorthy et al., 2017) implemented in the IQ-TREE. Bayesian Markov chain Monte Carlo (MCMC) analysis, which consisted of two runs for 1 000 000 generations, was performed by using Mr. Bayes v3.2.6 (Ronquist et al., 2012). A default 25% burn-in was used to reconstruct the tree under the 50% majority-rule. Comparison and calculation of differences between the trees obtained from the different methods was done by using TOPD/FMTS v3.3 (Puigbò et al., 2007).

Results

Assembly and orthology detection

Retrieval of raw reads belonging to eight different species (Table 1) from the NCBI SRA database together with in-house sequenced reads of *T. wallichiana* resulted in 233.123 million raw reads for processing. After adaptor removal and cleaning of low-quality bases, the generation of nine de novo assemblies yielded a total of 940 191 assembled transcripts. Clustering of these assembled transcripts resulted in 409 734 unigenes across nine species. Quality of the assemblies was assessed through different approaches. Read representation using Bowtie2 showed an average alignment rate of 83.1%. Completeness assessment of the assemblies showed that of 1440 core genes queried, the average number of core genes detected was 982 (68.19) with highest number of 1214 (84.31%) for *T. wallichiana* and lowest of 824 (57.22%) for *Torreya taxifolia* (Table 1). Initially, we also used *Cephalotaxus sinensis* in our analysis, but due to the low quality of the assembly, showing only 56.33% of reads mapping back to the assembly, we excluded it from further analysis. Using Proteinortho, 409 734 unigenes were clustered into 47 752 different orthologous groups. Following filtration of all the degenerate groups, we obtained a set of 331 complete groups which contain no more than one gene for each species.

Alignment and phylogenetic analysis

These single-copy orthologous groups were aligned individually. Concatenation of the individual alignment files generated a super-alignment file having an

Table 1
Statistics of raw reads, assembled transcripts and unigenes and quality parameters of the assemblies

	Taxus wallichiana	*Taxus cuspidata*	*Austrotaxus spicata*	*Pseudotaxus chienii*	*Torreya taxifolia*	*Amentotaxus argotaenia*	*Cephalotaxus hainanensis*	*Cephalotaxus harringtonia*	*Ginkgo biloba*
Assembly statistics									
Raw reads	101 071 384	8 984 233	13 333 334	13 333 334	16 506 045	17 070 241	25 721 203	14 757 991	22 346 199
Transcripts	209 860	88 580	72 876	81 530	77 796	127 205	91 025	61 407	129 912
Unigenes	129 869	75 411	21 802	20 724	22 785	30 747	35 032	24 899	48 465
MSL*	1244	571	1199	1122	1035	1086	1151	1498	1537
N50†	1606	740	1400	1278	1155	1237	1330	1840	1955
L50‡	29 069	17 041	7395	7049	7061	8995	10 518	6884	12 544
GC content	39.22	41.39	42.42	42.80	41.97	41.30	42.36	41.50	41.93
Quality									
A§	92.09%	87.20%	81.80%	72.94%	84.39%	80.29%	76.35%	91.50%	81.94%
B¶	6854	2783	4136	3735	3506	3821	4488	5355	6582
C**	1214	828	907	857	824	920	977	1109	1202

*Mean sequence length. †N50 is the assembly statistic which depicts that at least half of the nucleotides in the assembly belong to the transcripts having length equal to or greater than N50. ‡L50 is the number of sequences for which the length sum exceeds 50% of the size of the assembly. Quality parameters: §A represents the percentage of reads mapping back to assembly, ¶B is the number of proteins representing nearly full-length transcripts and having an alignment coverage of ≥ 80% using BLAST searches against the SwissProt database, and **C is the number of core genes detected against the query of 1440 seudotaxus genes in BUSCO completeness analysis.

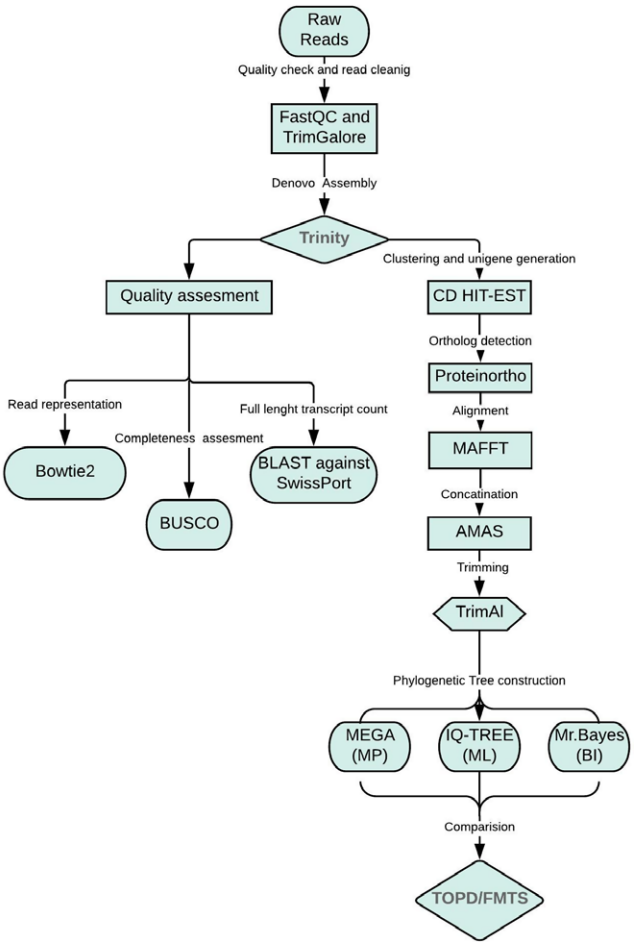

Fig. 1. Schematic representation of the pipeline used for detection of orthologues and phylogenetic analysis. [Colour figure can be viewed at wileyonlinelibrary.com]

alignment length of 921 206, 463 145 variable sites and 277 597 parsimony-informative sites. Trimming of the super-aligned sequences resulted in the alignment matrix having 100% coverage with no gaps. MP analysis using MEGA showed that the trimmed sequences have an alignment length of 240 731, 135 756 parsimony-informative sites, 75 411 constant sites, 165 320 variable sites and 29 564 singleton sites. Model testing for ML analysis was performed under Bayesian information criterion (BIC), Akaike information criterion (AIC) and corrected Akaike information criterion (CAIC) for 286 models. The best fit model for construction of the ML phylogenetic tree was found to be TVM+R2 under BIC. In Bayesian analysis, MCMC searches at a 95% posterior probability resulted in a single tree with posterior probability (*P*) of 1.00 and cumulative posterior probability (*P*) of 1.00. The observation of a single tree after MCMC searching as the most probable and credible tree shows the topological certainty of our data. The MP analysis yielded a tree with strong statistical support which resolved the Taxaceae and Cephalotaxaceae as separate clades. *Amentotaxus* and *Torreya* were found closer to Cephalotaxaceae. The phylogenetic trees obtained from ML and Bayesian inference (BI) strongly corroborate the results obtained from MP analysis. All three methods showed two major clades with strong statistical support. *T. cuspidata*, *A. spicata*, *Pseudotaxus chienii* and *T. wallichiana* resolved into one clade while *A. argotaenia*, *T. taxifolia*, *C. hainanensis* and *C. harringtonia* segregated into another. *G. biloba* was used as the outgroup. Comparison of the ML, MP and BI trees (Table 2) showed that all had 100% taxon similarity. Moreover, the split distance, nodal distance and disagreement between taxa was found to be zero when calculating differences between the trees. Because all the trees are similar, for the sake of non-redundancy, a single tree representing the values from all methods is shown here (Fig. 2). Individual trees from the different methods are given in the Supporting Information (Figs S1–S3).

Discussion

In comparative genomics and molecular phylogenetics, orthology assessment forms a crucial part of data analysis. The ever increasing rise in generating genomic and transcriptomic sequences has led to rapid expansion of public databases. Studies on Gymnosperms at the whole genome level are rare. The transcriptomic data generated through RNAseq technology can provide an efficient alternative source for large-scale orthology detection. However, even at the transcriptomic level, the gymnosperms have received comparatively less attention and transcriptomic sequences of only limited taxa are available. In the absence of a reference genome, we developed de novo assemblies of the selected species for orthology analysis to address the taxonomic position of Taxaceae and Cephalotaxaceae. RNAseq read representation, BLAST searches against the SwissProt database and completeness assessment showed that the quality of our assemblies is good for orthology analysis. Phylogenetic analysis of a set of 331 orthologous groups showed separate clades for Taxaceae and Cephalotaxaceae with strong statistical support. As mentioned above, different studies give different opinions regarding the taxonomic position of the two families. Our analysis suggests a sister relationship between them and supports their distinctiveness at the family level. The validity of our results is strengthened by the observation that MP, ML and BI methods resolved the taxa in a similar manner. Our results are corroborated by evidence from morphological, anatomical, reproductive and phylogenetic studies. The sister relationship between them was also observed by Lu et al.

Table 2
Comparison of trees generated by different methods on the same taxa

S. no	Method	ML–BI	BI–MP	ML–MP
1	Proportion of taxa in common	100%	100%	100%
2	Nodal distance	0	0	0
3	Split distance	0	0	0
4	Disagreement	0/9	0/9	0/9

ML, maximum likelihood; MP, maximum parsimony; BI, Bayesian. Disagreement values 0/9 indicate that of 9 taxa there are 0 taxa disagreeing between the trees.

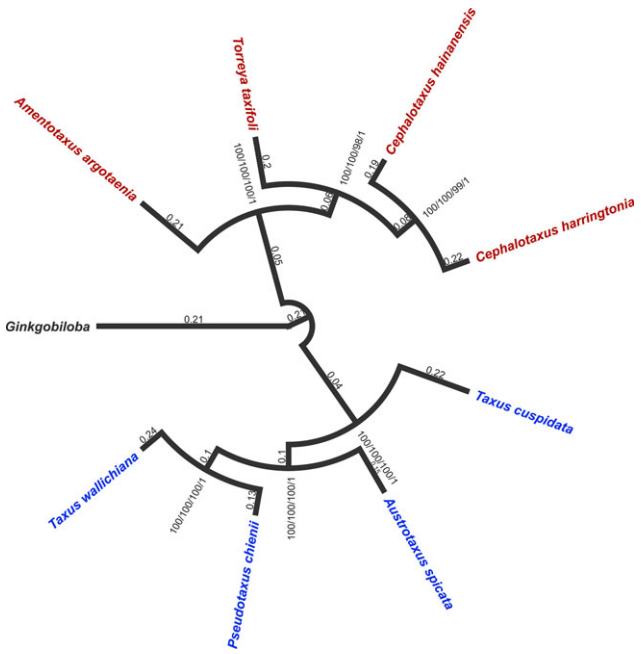

Fig. 2. Phylogenetic tree depicting the segregation of Taxaceae and Cephalotaxaceae with *G. biloba* as the outgroup. The tree was estimated with MP, ML and BI for a concatenated dataset of 331 unigenes. Support values are listed in the order: ML ultrafast bootstrap/ML SH-aLRT/MP bootstrap/Bayesian posterior probability. [Colour figure can be viewed at wileyonlinelibrary.com]

(2014), Cheng et al. (2000) and Xiao et al. (2008). Furthermore, morphological studies by Hart (1987) and Price (1990) support our results of a sister relationship between them. Our results did not support merging these families as proposed by Quinn et al. (2002), Ghimire and Heo (2014), Christenhusz et al. (2011) and Lang et al. (2013). Recently, a sister relationship has also been established between them using fluorescence microscopy of leaf anatomy (Elpe et al., 2018). Although Taxaceae and Cephalotaxaceae show similarity in several morphological features, they differ widely in others. In Taxaceae the aril of the cone partly encloses the seed while in Cephalotaxaceae it encloses the seed fully. Furthermore, maturation of the cone in Taxaceae takes 6–8 months compared with 18–20 months in Cephalotaxaceae. Moreover, mature seeds of Taxaceae have an average length of 5–8 mm compared with 12–40 mm in Cephalotaxaceae (Singh, 2006). They also differ in embryological characters (Xiao et al., 2008; Singh, 1961). The absence of foliar resin canals in Taxaceae also distinguishes them from Cephalotaxaceae. Cephalotaxaceae differ from Taxaceae in having several two-ovulate bracts in their seed cones in contrast to single-ovulate bracts of Taxaceae (Xiao et al., 2008). Embryological studies also show differences between Cephalotaxaceae and Taxaceae (Singh, 1961). These attributes support our analysis and highlight their distinctiveness at the family level.

Regarding the position of *Amentotaxus* and *Torreya*, it is clear that they resemble Taxaceae but show also strong likenesses to Cephalotaxaceae. Both show some similarities as well as differences with both Taxaceae and Cephalotaxaceae and also with each other morphologically, anatomically, embryologically and phylogenetically. Thus, different authors held varied opinions and placed them in one or the other group. Our results suggest that they are closer to Cephalotaxaceae than to Taxaceae. They are corroborated by the findings of Lu et al. (2014) who also observed *Amentotaxus* and *Torreya* segregating with the *Cephalotaxus* clade. Moreover, *Cephalotaxus*, *Amentotaxus* and *Torreya* were found to have foliar resin canals and discontinuous fibrous hypodermis, suggesting they are closely related to each other. Such resin canals are absent in extant taxad genera (Lu et al., 2014). *Torreya* and *Cephalotaxus* resemble each other in having bijugate phyllotaxis (Tomlinson and Zacharias, 2001). *Amentotaxus* was considered as a member of Cephalotaxaceae (Pilger, 1926; Page, 1990). Based on their anatomical characteristics, *Amentotaxus* and *Torreya* were found closest to Cephalotaxaceae (Hu and Wang, 1989). *Torreya* and *Amentotaxus* differ from Taxaceae in having abaxial microsporangia and an aril adnate to the seed coat, enclosing the seed tightly in contrast to radial microsporangia and cup-like aril free from the integument of the ovule (Price, 1990). While studying leaf anatomy, Ghimire et al. (2014) found that *Torreya*, *Amentotaxus* and *Cephalotaxus* have only one resin canal below the vascular bundle, in contrast to *Taxus*, *Austrotaxus* and *Pseudotaxus* which do not possess

any resin canals. Furthermore, *Amentotaxus* and *Torreya* were found to be closer to *Cephalotaxus* in having discontinuous fibrous hypodermis. The authors tentatively suggested Cephalotaxaceae neither as a separate family nor totally merged in Taxaceae. Although the authors agree on support for minor familial classification between them, they favoured Taxaceae containing Cephalotaxaceae (Ghimire et al., 2014). Our analysis rejects this and proposes to maintain their distinctiveness at the family level. Alliance between *Amentotaxus* and *Cephalotaxaceae* has also been suggested by Hart (1987) and Keng (1969). Except for the lack of bracts, the pollen strobilus of *Amentotaxus* resembles that of Cephalotaxaceae in being compound. The dorsiventral symmetry of microsporangia (rarely radial) also resembles that of Cephalotaxaceae (Wilde, 1976). *Torreya* shows cleavage polyembryony in contrast to simple polyembryony and formation of the cell wall at the four- to eight-celled stage, making it distinct from Taxaceae (Doyle and Brennan, 1971; Tang et al., 1986). The aril completely surrounding the seed, opposite or sub-opposite branchlets and absence of a prominent midrib differentiates it from *Taxus* (Cope, 1998). All these features bring *Amentotaxus* and *Torreya* closer to Cephalotaxaceae and thus corroborate our results. Our results did not favour the non-alliance of *Amentotaxus* and *Torreya* with Cephalotaxaceae as propounded by Cheng et al. (2000) or Janchen's (1949) two tribe classification of Taxaceae, Torreyeae (encompassing *Torreya* and *Amentotaxus*) and Taxeae (containing taxad genera) based on the presence of foliar resin canals. Furthermore, placing *Amentotaxus* and *Austrotaxus* within the same tribe, Amentotaxeae, on the basis of having spike-like male strobilus as proposed by Chen and Wang (1984) and Koidzumi (1932), is not supported by our results.

Thus, phylogenetic studies involving a different set of loci arrive at different results regarding the position of these families. This is because different loci tend to evolve at different rates so will resolve between taxa differently. Here we used a set of 331 unigenes that resolved Taxaceae and Cephalotaxaceae as sister groups and also resolved *Amentotaxus* and *Torreya* within Cephalotaxaceae. This analysis is based on very large number of loci and sufficiently long alignment length and its resolving power is greater than for earlier phylogenetic studies involving few loci and a comparatively shorter alignment length. Furthermore, the convergence of the MP, ML and BI methods strengthens this observation. The topological credibility of our data is shown by generation of a single tree as the most probable tree at 95% credibility in MCMC searches. The high bootstrap and posterior probability values obtained for the tree topologies add additional strength to the results.

Conclusion

The potential for RNAseq in phylogenetics has recently been recognized. Utilization of RNAseq data provides the opportunity to incorporate a large number of loci and sufficiently long alignment length for construction of a robust phylogeny. Based on this approach, we provided a set of 331 unigenes that resolve Taxaceae and Cephalotaxaceae as distinctly separate clades, suggesting their distinctiveness at the family level. The phylogenetic methods used here gave the same segregation of the taxa, strongly supporting our inference. We believe that our analysis is robust and powerful due to the incorporation of a large number of loci, very strong statistical support, topological certainty for the data and convergence of the different phylogenetic methods to a similar result.

Acknowledgements

This study was financially supported by MoEF & CC (Ministry of Environment, Forests and Climate Change), India, under grant No. NMHS/SG-2016/011. AM acknowledges CSIR New Delhi for their financial assistance towards PhD study.

Conflict of interest

The authors declare no competing interests.

References

Borowiec, M. L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 4, e1660.

Capella-Gutiérrez, S., Silla-Martínez, J. M., Gabaldón, T., 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Chaw, S. M., Sung, H. M., Long, H., Zharkikh, A., Lie, W. H., 1995. The phylogenetic positions of the conifer genera *Amentotaxus*, *Phyllocladus*, and *Nageia* inferred from 18S rRNAsequences. J. Mol. Evol. 41, 224–230.

Chaw, S. M., Zharkikh, A., Sung, H. M., Lau, T. C., Li, W. H., 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNAsequences. Mol. Biol. Evol. 14, 56–68.

Chen, Z. K., Wang, F. H., 1984. On the systematic position of *Amentotaxus* from its embryological investigation. Acta Phytotax. Sin. 22, 269–276.

Cheng, Y., Nicolson, R. G., Tripp, K., Chaw, S. M., 2000. Phylogeny of Taxaceae and Cephalotaxaceae genera inferred from chloroplast matK gene and nuclear rDNA ITS region. Mol. Phylogenet. Evol. 14, 353–365.

Christenhusz, M., Reveal, J., Farjon, A., Gardner, M. F., Mill, R. R., Chase, M. W., 2011. A new classification and linear sequence of extant gymnosperms. Phytotaxa 19, 55–70.

Chuang, T. I., Hu, W. W., 1963. Study of *Amentotaxus argotaenia* (Hance) Pilger. Bot. Bull. Acad. Sinica 4, 10–14.

Cope, E. A., 1998. Taxaceae: the genera and cultivated species. Bot. Rev. 64, 291–322.
Deng, H., Zhang, G. Q., Lin, M., Wang, Y., Liu, Z. J., 2015. Mining from transcriptomes: 315 single-copy orthologous genes concatenated for the phylogenetic analyses of Orchidaceae. Ecol. Evol. 5, 3800–3807.
Doyle, J., Brennan, M., 1971. Cleavage polyembryony in conifers and taxads–a survey. 1. Podocarps, taxads and taxodioids. Roy. Dublin Soc. Sci. Proc. Ser. 4(A), 57–88.
Eichler, A. W., 1889. Coniferae. In: Engler, A., Prantl, K. (Eds.). Die Naturlichen Pflanzenfamilien, Engelmann, Leipzig, Vol. 1, pp. 28–116.
Elpe, C., Knopf, P., Stützel, T., Schulz, C., 2018. Diversity and evolution of leaf anatomical characters in Taxaceae sl—fluorescence microscopy reveals new delimitating characters. J. Plant. Res. 131 (1), 125–141.
Florin, R., 1948. On the morphology and relationships on the Taxaceae. Bot. Gaz. 110 (1), 31–39.
Florin, R., 1954. The female reproductive organs of conifers and taxads. Biol. Rev. 29, 367–389.
Ghimire, B., Heo, K., 2014. Cladistic analysis of Taxaceae sl. Plant Syst. Evol. 300, 217–223.
Ghimire, B., Lee, C., Heo, K., 2014. Leaf anatomy and its implications for phylogenetic relationships in Taxaceae s. l. J. Plant. Res. 127, 373–388.
Grabherr, MG, Haas, BJ, Yassour, M, Levin, JZ, Thompson, DA, Amit, I, Adiconis, X, Fan, L, Raychowdhury, R, Zeng, Q, Chen, Z, Mauceli, E, Hacohen, N, Gnirke, A, Rhind, N, di Palma, F, Birren, BW, Nusbaum, C, Lindblad-Toh, K, Friedman, N, Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Hart, J. A., 1987. A cladistic analysis of conifers: preliminary results. J. Arnold. Arbor. 26, 9–307.
Hittinger, C. T., Johnston, M., Tossberg, J. T., Rokas, A., 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. Proc. Natl Acad. Sci. USA 107, 1476–1481.
Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., Le, S. V., 2017. UFBoot2: improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol., 35, 518–522.
Hu, Y. S., Wang, F. H., 1989. Anatomy and affinities of *Cephalotaxus* (Cephalotaxaceae). Cathaya 1, 37–48.
Janchen, E., 1949. Das System der Konifern. Akad. Wien. Sitz-ber. 158, 155–262.
Johnson, K. P., Walden, K. K., Robertson, H. M., 2013. Next-generation phylogenomics using a target restricted assembly method. Mol. Phylogenet. Evol. 66, 417–422.
Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., Jermiin, L. S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods, 14, 587.
Katoh, K., Standley, D. M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Keng, H., 1969. Aspects of morphology of *Amentotaxus formosana* with a note on the taxonomic position of the genus. J. Arnold. Arbor. 50, 432–448.
Koidzumi, G., 1932. Notes on Amentotaxaceae. Acta Phytotax. Geobot. 1, 185.
Kudo, Y., Yamamoto, Y., 1931. Amentotaxaceae. Mater. Fl. Formosa IV Soc. Trop. Agric. 3, 110–111.
Kumar, S., Nei, M., Dudley, J., Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9, 299–306.
Lang, X. D., Su, J. R., Lu, S. G., Zhang, Z. J., 2013. A taxonomic revision of the genus *Cephalotaxus* (Taxaceae). Phytotaxa 84, 1–24.
Langmead, B., Salzberg, S. L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357.
Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., Prohaska, S. J., 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC Bioinformatics 12, 124.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.
Lu, Y., Ran, J. H., Guo, D. M., Yang, Z. Y., Wang, X. Q., 2014. Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. PLoS ONE 9, e107679.
Lyu, M. J. A., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., Hibberd, JM, Stata, M, Sage, RF, Lu, H, Wei, X, Wong, GK, Zhu, XG, 2015. RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications. BMC Evol. Biol. 15, 116.
Minh, B. Q., Nguyen, M. A. T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195.
Neger, F.W., 1907 Cephalotaxaceae. In: Göschen, G.J. (Ed.) Die Nadelhölzer (Koniferen) und übrigen. Gymnospermen. Leipzig, Vol. 00, pp. 23.
Nguyen, L. T., Schmidt, H. A., vonHaeseler, A., Minh, B. Q., 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Page, C. N., 1990. Coniferophytina. In: Kramer, K. U., Green, P. S. (Eds.). The Families and Genera of Vascular Plants. Pteridophytes and Gymnosperms, Springer-Verlag, Berlin.
Pilger, R. 1903. Taxaceae. In: Engler, A. (Ed.) Das Pflanzenreich 5 (Heft 18), 4th edn. Wilhelm Engelmann, Leipzig, pp. 99–105.
Pilger, R., 1926. Coniferae. In: Engler, A., Prantl, K. (Eds.). Die Natürlichen Pflanzenfamilien. Engelmann, Leipzig, vol. 13, pp. 121–407.
Price, R. A., 1990. The genera of Taxaceae in the southeastern United States. J. Arnold. Arbor. 71, 69–91.
Puigbò, P., Garcia-Vallvé, S., McInerney, J. O., 2007. TOPD/FMTS: a new software to compare phylogenetic trees. Bioinformatics, 23, 1556–1558.
Quinn, C. J., Price, R. A., Gadek, P. A., 2002. Familial concepts and relationships in the conifer based on rbcL and matK sequence comparisons. Kew. Bull., 57, 513–531.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol., 61, 539–542.
Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.
Singh, H., 1961. The life history and systematic position of *Cephalotaxus drupaceae* Sieb. et Zucc. Phytomorphol. 11, 153–197.
Singh, V. P., 2006. Gymnosperm (naked seeds plant): structure and development. Sarup & Sons, New Delhi.
Stefanoviac, S., Jager, M., Deutsch, J., Broutin, J., Masselot, M., 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. Am. J. Bot. 85, 688–688.
Takhtajan, A. L., 1953. Phylogenetic principles of the system of higher plants. Bot. Rev. 19, 1–45.
Tang, Z. X., Chen, Z. K., Wang, F. H., 1986. Investigation on sexual reproductive cycle in *Torreya grandis*. Acta Phytotax. Sin. 24, 447–453.
Tomlinson, P. B., Zacharias, E. H., 2001. Phyllotaxis, phenology and architecture in *Cephalotaxus*, *Torreya* and *Amentotaxus* (Coniferales*)*. Bot. J. Linn. Soc. 135, 215–228.
Van Tieghem, M. P., 1891. Structure et affinities des *Cephalotaxus*. Bull. Soc. Bot. Fr. 38, 184–190.
Wang, X., Shu, Y., 2000. Chloroplast matK gene phylogeny of Taxaceae and Cephalotaxaceae, with additional reference to the systematic position of *Nageia*. Acta Phytotax. Sin. 38, 201–210.
Wilde, M. H., 1976. A new interpretation of microsporangiate cones in Cephalotaxaceae and Taxaceae. Phytomorphology. 25, 434–450.
Xi, Y. Z., 1986. Studies on pollen morphology of Taxaceae of China. Acta Phytotax. Sinica 24, 247–252.
Xiao, P. G., Huang, B., Ge, G. B., Yang, L., 2008. Interspecific relationships and origins of Taxaceae and Cephalotaxaceae

revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences. Plant Syst. Evol. 276, 89–104.
Zou, M., Guo, B., Tao, W., Arratia, G., He, S., 2012. Integrating multi-origin expression data improves the resolution of deep phylogeny of ray-finned fish (Actinopterygii). Sci. Rep. 2, 665.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig S1. Maximum-parsimony phylogenetic tree constructed by MEGA. The tree was obtained using the SPR search method with search level 1 for 500 replicates.

Fig S2. Maximum-likelihood tree constructed by IQ-TREE; the order of values is UltraFast Bootstrap/SH-aLRT. The analysis involved 10000 replicates.

Fig S3. Bayesian inference tree constructed by Mr. Bayes showing posterior probability values. The analysis consisted of two runs for 10 00 000 generations.