Forest Systems
27 (3), e016, 11 pages (2018)
eISSN: 2171-9845
https://doi.org/10.5424/fs/2018273-13688
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria O. A., M. P. (INIA)

RESEARCH ARTICLE

OPEN ACCESS

Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*

Sanna Olsson[1], Delphine Grivet[1,2] and Jerónimo Cid Vian[1,3]

[1]*Department of Forest Ecology & Genetics, Forest Research Centre, INIA-CIFOR, Carretera de la Coruña km 7.5, 28040 Madrid, Spain.* [2]*Sustainable Forest Management Research Institute, INIA, University of Valladolid, 34004 Palencia, Spain.* [3]*Technical University of Madrid, School of Forestry and Natural Resources*

Abstract

Aim of study: The identification of material of forest tree species using genetic markers was carried out. Two promising chloroplast barcode markers, *matK* and *ycf1*, were tested for species identification and reconstruction of phylogenetic relationships in pines.

Area of study: The present study included worldwide *Pinus* species, with a wide representation of European taxa.

Material and methods: All *matK* sequences longer than 1600 base pairs and *ycf1* sequences for the same species were downloaded from GenBank, aligned and subsequently analyzed to estimate alignment statistics, phylogenetic trees and substitution saturation signals.

Main results: We confirm the usefulness of the *ycf1* marker for barcoding purposes and phylogenetic studies in pines, especially in studies focusing at the within-genus level relationships, but caution in the use of the *matK* marker is recommended.

Research highlights: Incongruent phylogenetic signals between these two chloroplast markers are demonstrated in pines for the first time.

Additional keywords: barcoding, conifers, phylogeny.

Abbreviations used: posterior probabilities (PP), bootstrap (BS).

Authors´ contributions: SO and DG designed the study. JCV analysed the data with help from SO. SO wrote the manuscript together with DG and contributions from JCV. All authors approved the final version of the manuscript.

Citation: Olsson, S., Grivet, D., Cid-Vian, J. (2018). Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. Forest Systems, Volume 27, Issue 3, e016. https://doi.org/10.5424/fs/2018273-13688

Received: 11 Jul 2018. **Accepted:** 30 Oct 2018.

Funding: SO received funding from the Spanish Ministry of Economy and Competitiveness (MINECO) under PTA2015-10836-I contract.

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to sanna.olsson@inia.es

Introduction

In forest trees, diagnostic markers have diverse applications in biodiversity, conservation, restauration, trade control, or tree improvement. The identification of forest material is generally performed using molecular markers developed for different purposes, and therefore analysed at different hierarchical levels (species, provenances, families or clones). When the objective is the unambiguous identification of single species that are morphologically difficult to distinguish in their original state or because samples are transformed products (*e.g.* timber, furniture, barrel, processed food), barcoding technology, using short universal DNA sequences, can be applied (Lidder & Sonnino, 2011). At the species level, barcoding is central to a major field: the internationally traded timber and wood products. Forensic applications are directed towards identifying species that are illegally exported, high-value species that are falsely declared to be low value timbers and sold as such (Nielsen & Dahl, 2008), or protected species under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) regulations.

Species delineation is also of interest for establishing the relationships among species in phylogenetic studies. Apart from advancing our understanding in evolution and biodiversity, there are many practical applications of phylogenetics. For example, the knowledge of the species phylogenies may help understand the evolutionary trade-offs of life-history traits in pines (*e.g.* Grivet *et al.*, 2013) or assist strategies dealing with pine diseases and pests (*e.g.* Moreira *et al.*, 2016). In conservation biology, phylogenetic information can be used to select and prioritize populations (Volkmann *et al.*, 2014). Phylogenetic and phylogeographic methods

can be particularly useful to infer the origin of timber and wood products (Finkeldey *et al.*, 2010). Phylogenetic methods based on barcoding markers have successfully been applied to prevent illegal trade of protected species (Baker *et al.*, 2010; Ghorbani *et al.*, 2017). Furthermore, several applications are implemented at the intraspecific level for traceability of important tropical timber species (Tnah *et al.*, 2009, 2010; Degen *et al.*, 2010), following international agreements (*e.g.* FLEGT, the EU Forest Law Enforcement, Governance and Trade, regulation), or for trade control of forest reproductive material.

Chloroplast genomes, due to their characteristics, provide a good source of species-diagnostic markers. More specifically, they are present in multiple copies (facilitating PCR amplification), uniparentally inherited, and suitable for studies involving different taxonomic levels due to regions that evolve at different rates (Soltis & Soltis, 1998; Xu *et al.*, 2015). Species-diagnostic markers are deposited in public repositories of molecular sequence data that rassemble the information available for all species sequenced for a specific marker (*e.g.* Genbank). The use of novel diagnostic markers is therefore limited as it would require sequencing many species for that marker, and consequently the same established genetic markers are often used for both barcoding and phylogenetic purposes. Ideally these markers should be as generalizable across groups as possible without losing species resolution capacities (Kress *et al.*, 2009). The most suitable markers for barcoding in plants were selected among commonly used phylogenetic markers by the CBOL Plant Working Group (Hollingsworth *et al.*, 2009a).

In the present study our aim is to test diagnostic chloroplast markers in *Pinus*, a genus of huge ecological and economical importance (Price *et al.*, 1998). With over a hundred recognized species, *Pinus* is the largest genus of conifers and constitutes a major, often dominant component of multiple natural landscapes such as boreal, subalpine, temperate, tropical and arid woodlands (Richardson & Rundel, 1998). The economic importance of pines stems from their use as sources of wood, pulp, resins and charcoal. In addition, pines are currently the focus of biomass research as promising type of forest plantation for energy production (Álvarez-Álvarez *et al.*, 2018).

The *Pinus* genus is divided in subgenus *Strobus* and subgenus *Pinus*, the latter consisting of sections Pinus (subsections Pinus and Pinaster) and section Trifoliae (subsections Contortae, Ponderosae and Australes) (Gernandt *et al.*, 2005). Pine phylogenetic relationships are still partly unresolved, especially among terminal taxa in the subsections Strobus and Australes (Eckert & Hall 2006; Parks *et al.*, 2009; Gernarndt *et al.*, 2018). Furthermore, species complexes have been particularly debated groups and their exact composition and relationships have been questioned, as this is the case for instance for North-American *Pinus contorta-banksiana* (Yang *et al.*, 2007), Asian *Pinus kesiya* (Businský *et al.*, 2014), as well as European *Pinus mugo* (Christensen, 1987) and Mediterranean pines (Syring *et al.*, 2005; Grivet *et al.*, 2013). This species-delineation limitation poses problems when trying to identify forest materials at the species level based on solid timber products from species that are not well identified by wood traits, as is the case of the closely related *Pinus nigra*, *Pinus mugo* and *Pinus sylvestris* (Schoch *et al.*, 2004). Two promising species-diagnostic chloroplast markers in pines are *matK* and *ycf1*. The *matK* marker has been one of the most frequently used genes for inferring phylogeny in pines (Wang *et al.*, 1999, Geada López *et al.*, 2002; Gernandt *et al.*, 2003, 2005, 2008; Hernández-León *et al.*, 2013; Dong *et al.*, 2015). The more recently introduced *ycf1* was reported to be more variable than other chloroplastic markers commonly used in phylogenetic studies in pines (*rbcL, trnD-Y-E, trnH-psbA* and *matK*) as shown by Hernández-León *et al.*, (2013). Based on these premises, we tested the suitability of *matK* and *ycf1* for barcoding purposes and for resolving phylogenetic relationships in pines mostly from Europe.

Material and Methods

The approximately 1,550 base pairs (bp) long maturase K (*matK*) gene was shown to be one of the most promising barcode markers in all land plants (Hollingsworth *et al.*, 2009a). In pines, *matK* has been frequently used for inferring phylogeny (Wang *et al.*, 1999, Geada López *et al.*, 2002; Gernandt *et al.*, 2003, 2005, 2008; Hernández-León *et al.*, 2013; Dong *et al.*, 2015). These studies showed that *matK* is not variable enough in pines to fully resolve species level relationships. Efforts to develop more variable markers to clarify the remaining controversial relationships have been made. The marker *ycf1* was proposed as a promising marker for pines by Parks *et al.* (2009, 2011). Dong *et al.* (2015) confirmed *ycf1* to be the most variable plastid DNA barcode of land plants. However, the evolution of the gene was pointed as abnormal and probably under selection (Parks *et al.*, 2009). Furthermore, this uncommonly high variability could be an issue in higher taxonomic level in studies focusing on above-species level relationships. The few earlier studies comparing the use of *matK* and *ycf1* in resolving phylogenetic relationships in the genus *Pinus* (Hernández-León *et al.*, 2013; Dong *et al.*, 2015) did

not study the whole length of the *matK* marker but used only an approximately 800 bp long region.

In the present study, all the *matK* sequences longer than 1600 bp were downloaded from the GenBank, totalling 55 *Pinus* species (Table 1). The *ycf1* sequences for the same species were also downloaded. The GenBank Accession Number of each sequence is provided in Table 1. Only one sequence per species was used. The sequences were aligned using MAFFT (Katoh & Standley, 2013) to produce two alignments, one for *matK* and one for *ycf1*, and adjusted manually with PhyDE® v1.0 (Müller *et al.*, 2005). Statistics on the alignments were obtained with PhyDE plugin SeqState. Uncorrected pairwise distances were compared with maximum likelihood distances in PAUP v4.0b10 (Swofford, 2002) to detect any saturation signal in the markers, and checking for deviation from linearity of plots.

Two phylogenetic analyses were performed on the individual alignments and on a concatenated matrix. First, Bayesian analyses were performed with MrBayes v3.2.6 (Ronquist *et al.*, 2012) implemented at CIPRES Science Gateway (Miller *et al.*, 2010). Best-fit substitution models were inferred from jModeltest v.2.1.10 (Darriba *et al.*, 2012). Following the output from the jModeltest the GTR+Γ model was applied for both *matK* and *ycf1*. The *a priori* probabilities supplied were those specified in the default settings of the program. Four runs with four chains (1×10^6 iterations each) were run simultaneously. Chains were sampled every 1,000 iterations and the respective trees written to a tree file. Tracer v1.6 (Rambaut *et al.*, 2014) was used to analyze the output of the model parameters, more specifically to examine the sampling and convergence results. Calculations of the consensus tree and of the posterior probability of clades were performed

Table 1. *Pinus* sequences from 55 species downloaded from GenBank. The dataset corresponds to all *matK* sequences longer than 1600 base pairs and to all *ycf1* sequences for the same species. Asterisks (*) indicate those sequences where the *ycf1* region was extracted from the whole or partial chloroplast genome.

Species name	*matK*	*ycf1*	Species name	*matK*	*ycf1*
Pinus armandii	AB161002	KP089404	*Pinus maximinoi*	AB161010	KC157109
Pinus attenuata	AB080933	KC157134	*Pinus morrisonicola*	AF295031	JN854182*
Pinus banksiana	AB080922	KP089408	*Pinus mugo*	AB081087	JN854181*
Pinus brutia	AB161018	KP089932	*Pinus muricata*	AB080935	KC157153
Pinus canariensis	AB084494	KP089933	*Pinus nigra*	AB084498	KP089411
Pinus caribaea	AB080940	JN854222*	*Pinus oocarpa*	AB081084	KC157158
Pinus clausa	AB161003	KC157159	*Pinus palustris*	AB080937	KC157163
Pinus contorta	AB080921	FJ580260	*Pinus parviflora*	AB081086	KP089941
Pinus cooperi	AB161004	FJ580193	*Pinus patula*	AB080944	KP089936
Pinus coulteri	AB097785	JN854215*	*Pinus pinaster*	AB084493	FJ899583*
Pinus cubensis	AB080938	KC157114	*Pinus pinea*	AB084496	JN854173*
Pinus densata	AB097779	JN854209*	*Pinus pumila*	AB161013	JN854168*
Pinus densiflora	AB084497	KP089385	*Pinus pungens*	AB080932	JN854167*
Pinus douglasiana	AB080925	KJ152831	*Pinus radiata*	AB080934	KC157207
Pinus echinata	AB080936	KC157152	*Pinus resinosa*	AB080945	KC157078
Pinus elliottii	FM955321	JN854202*	*Pinus rigida*	AB080929	KC157183
Pinus engelmannii	AB080927	FJ580207	*Pinus roxburghii*	AB084495	JN854162*
Pinus fenzeliana	AB161005	KX255674*	*Pinus serotine*	AB080930	KC157199
Pinus halepensis	AB081089	JN854197*	*Pinus sibirica*	AB161014	FJ899558*
Pinus herrerae	AB080943	KC157155	*Pinus sylvestris*	AB097781	KP089937
Pinus hwangshanensis	AB161007	JN854194*	*Pinus tabuliformis*	AB161015	KP089380
Pinus jeffreyi	AB080926	KC157181	*Pinus taeda*	AB080928	KC157197
Pinus kesiya	AB161008	JN854191*	*Pinus taiwanensis*	AB161016	KP771703*
Pinus koraiensis	AB161009	AY228468*	*Pinus teocote*	AB097783	KC157202
Pinus lawsonii	AB097784	KC157176	*Pinus tropicalis*	AB080920	JN854156*
Pinus leiophylla	AB081085	KC157132	*Pinus virginiana*	AB080923	KC157196
Pinus massoniana	AB081088	KC427272*	*Pinus yunnanensis*	AB161017	JN854151*

based upon the trees sampled after chain convergence (< iteration 100,000). The second phylogentic method, a maximum likelihood (ML) analysis, was performed with RAxML (Stamatakis *et al.*, 2008) on the CIPRES Science Gateway using the GTR+CAT model with 1000 bootstrap replicates. Phylogenetic trees were displayed and edited using TreeGraph2 (Stöver & Müller, 2010).

Results

Alignment statistics

There were 1667 characters in the *matK* alignment, of which 586 belonged to the barcode region for *matK*. The *ycf1* alignment contained 2863 characters, including a visually observed hypervariable region of 208 bp. The regions are depicted in Figure 1. Details on the alignment are given in Table 2. Our alignment statistics for these two markers are consistent with earlier reported results (Hernández-León *et al.*, 2013; Dong *et al.*, 2015). No signal of saturation was observed, except for the *ycf1* marker including the hotspot region, for which very slight substitutional saturation was observed as illustrated with a slight desviation of the pairwise distance points from linearity (Figure 2).

The *ycf1* alignment was more variable than the *matK* alignments, with 17.5 % of parsimony informative sites (PIS) vs 7.5% and 5.8% for *matK*, depending whether the longer full *matK* region or only the barcode region was included, respectively. The hypervariable region observed by visual inspection of the *ycf1* marker had 32.2% of informative sites. Excluding this region lowered slightly the variability of the rest of the *ycf1* region (16.4 PIS %).

Phylogenetic trees

The majority rule consensus tree from the Bayesian inference had better resolution compared to the maximum likelihood tree (Figures 3-5). Therefore, the Bayesian trees are presented with confidence at the nodes indicated by posterior probabilities (PP) and complemented with bootstrap values (BS) of the maximum likelihood analysis when applicable. Following Alfaro *et al.* (2003) we consider PP > 0.95 or BS > 70 as statistically significant support for a clade.

The phylogenetic tree based on combined marker data is shown in Figure 3. The tree is fairly well resolved and supported. The relationships in subsection *Pinaster* are resolved and fully supported, but in subsection *Pinus* many of the placements do not receive statistically significant support. The topology of section *Trifoliae* is congruent with the phylogeny presented by Gernandt *et al.* (2018), with the formation of the same groups Contortae, Ponderosae, Attenuatae, Australes I and II. Australes II does not receive significant support (PP 0.87 / BS 62), though,

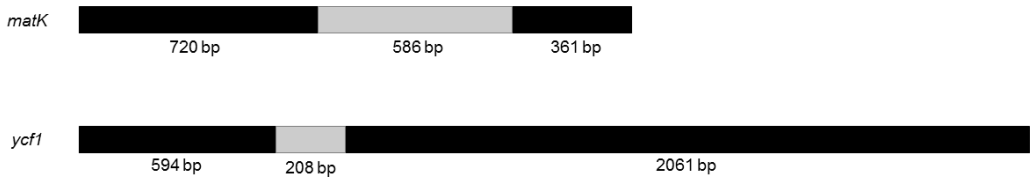

Figure 1. Depiction of the genetic regions *matK* and *ycf1* included in this study. The grey color in *matK* stands for a region used as barcoding marker and in *ycf1* for a hypervariable region. Regions are scaled by the length in base pairs (bp).

Table 2. Alignment statistics. Number of base pairs (bp), number of variable sites (VS), percentage of variable sites (VS %), number of parsimony informative sites (PIS) and percentage of parsimony informative sites (PIS %) are shown.

Region	bp	VS	VS %	PIS	PIS %
matK entire	1667	137	8.23	97	5.84
matK barcode region	586	55	9.39	44	7.51
ycf1	2863	629	21.97	502	17.53
ycf1 with no hotspot	2655	555	20.90	435	16.38
Hotspot region from *ycf1*	208	74	35.58	67	32.21
Combined *matK* + *ycf1*	4530	767	16.93	600	13.25

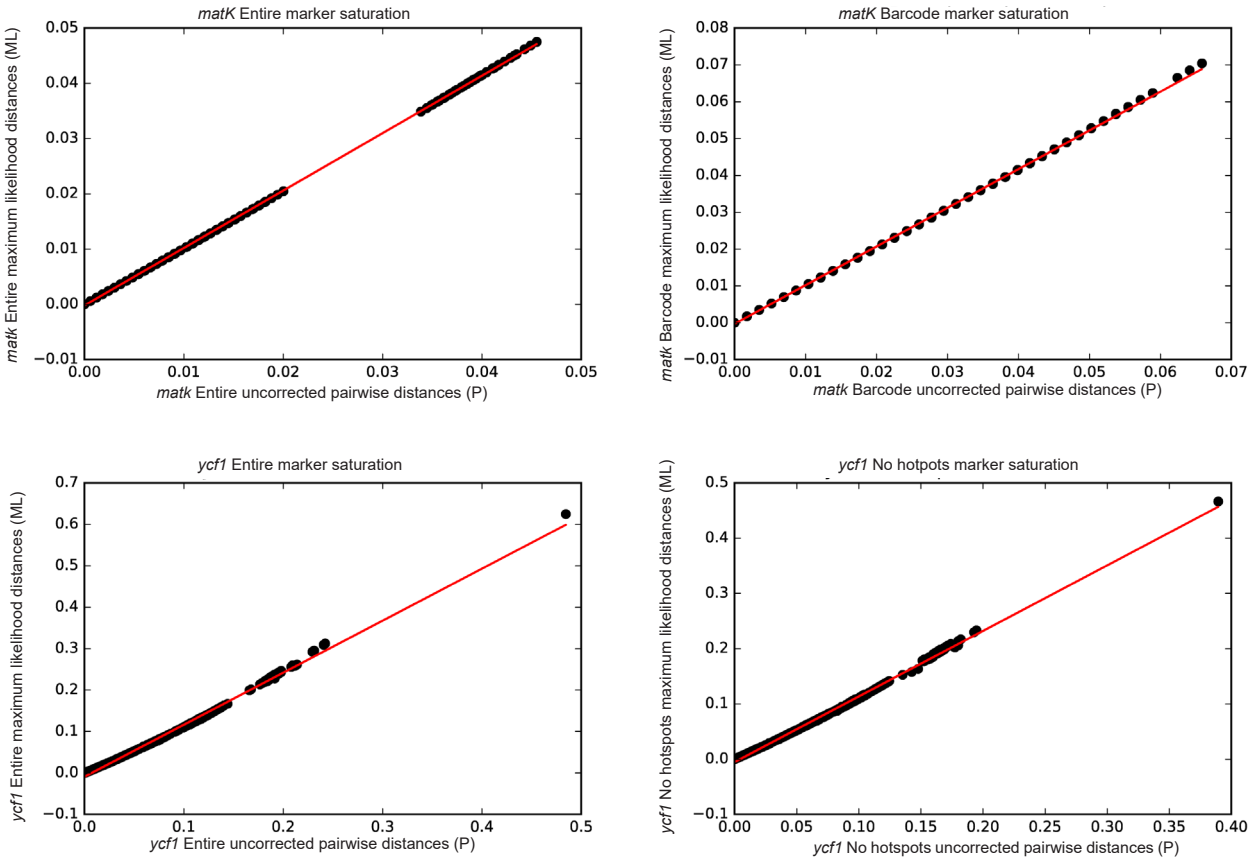

Figure 2. Plots of substitutional saturation in the markers. The uncorrected pairwise sequence distances ("P") were plotted against ML distances.

and Oocarpae is not resolved as a monophyletic group.

The relationships in the tree based on *matK* are poorly resolved from species level up to subsection level (Figure 4). The subsections *Pinaster* and *Pinus* are not resolved as individual clades, neither are the groups Attenuata, Oocarpa nor Australes.

The *ycf1* tree (Figure 5) is similar to that based on the combined marker data in both resolution and topology. The same subsections and groups are formed, and as in the combined tree, *Oocarpae* is not resolved as monophyletic clade. The support of Australes II clade is, however, significantly better supported than in the combined tree (PP 0.98 / BS 59). There were no significant differences between the phylogenetic trees based on *ycf1* with or without (data not shown) the hotspot region.

A few significant incongruences were detected when comparing the gene trees based on individual markers. The conflicting positions involve *P. attenuata, P. oocarpa, P. caribaea* and *P. tabuliformis. P. attenuata* is placed sister to *Pinus oocarpa* (PP 0.96 / BS 62) in the analysis based on *matK*, while *P. attenuata* more logically forms a clade together with *P. muricata* and *P. radiata* (Attenuatae or the California closed-cone pines) based on *ycf1* and the combined analysis. *P. caribaea* is placed in a clade with *P. leiophylla* and *P. patula* (PP 0.99 / BS 66) only in the analysis based on *matK*, while it is sister species to *P. elliottii* based by *ycf1* and the combined analysis. *P. tabuliformis* is sister species to *P. yunnanensis* (PP 0.96 / BS 65) based on *matK* but sister to *P. kesiya* (PP 0.98 / BS 63) based on *ycf1*. In the combined analysis *P. tabuliformis* is sister to *P. yunnanensis* with low support (PP 0.65 / BS 40).

Furthermore, the placement of some species present higher support values in one of the single marker trees. Most noteworthy, the relationships in the subsection *Pinus* are better resolved based on *ycf1* alone than on the combined data set. Based on *ycf1*, the positions of P. *resinosa, P. nigra, P. mugo, P. densiflora* and *P. sylvestris* are fully resolved with maximum support from the Bayesian analysis and mostly high bootstrap support from the maximum likelihood analysis. In the combined analysis, only the clade comprising *P. mugo*, *P. densiflora* and *P. sylvestris* receives statistically significant support values. This is because the main phylogenetic signal grouping those species comes from *ycf1*, while *matK* brings a conflicting signal.

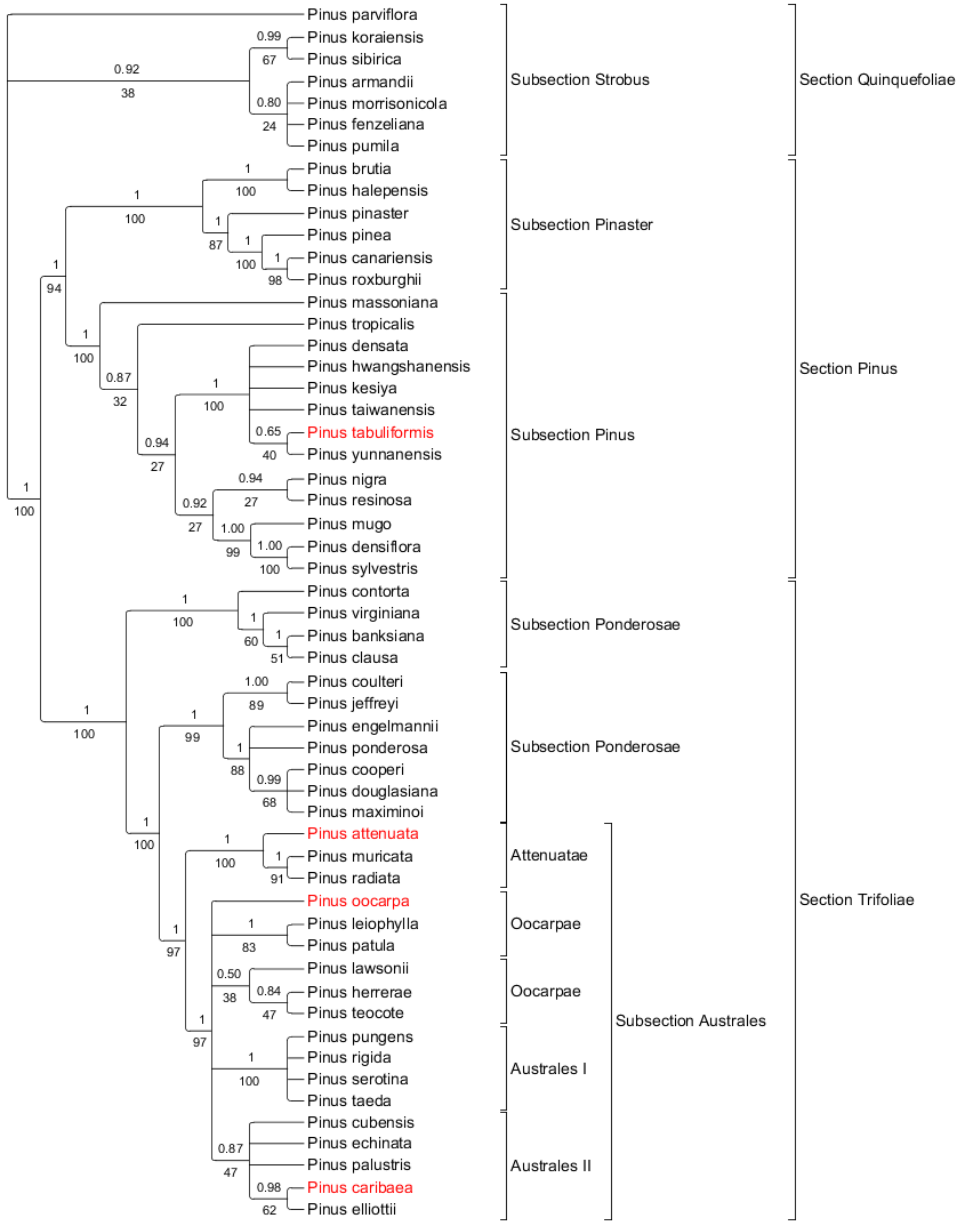

Figure 3. Phylogenetic tree based on combined data matrix of *matK* and *ycf1*. The tree represents the majority consensus of trees sampled after stationarity in the Bayesian analysis. Posterior probability values from the Bayesian inference are indicated above and the corresponding bootstrap values of the parsimony analysis are shown below when it was applicable. The labels indicating the taxonomic divisions following Gernandt *et al*. (2005) are shown. The taxa in red colour had incongruent positions between the individual analyses based solely on one marker.

Discussion

This study confirms the usefulness of the *ycf1* marker as diagnostic marker in pines. Although it has been suggested that *ycf1* does not correctly reflect phylogenetic relationships in plants (Parks *et al.*, 2009), its use for pine phylogenetic analyses resulted in expected taxonomic grouping in the present study. However, the hypervarible region of this marker could cause problems in homology assessment when it is used on a broader taxonomic scale. The marker *matK* should be used in pines with caution, because as shown in the present study, its phylogenetic signal does not reflect species relationships correctly in pines. In spite of this result, *matK* could be useful as a barcode marker with an intermediate level of variation in combination with other markers for species delineation (Bruni *et al.*, 2012; see also Celinsky *et al.*, 2017).

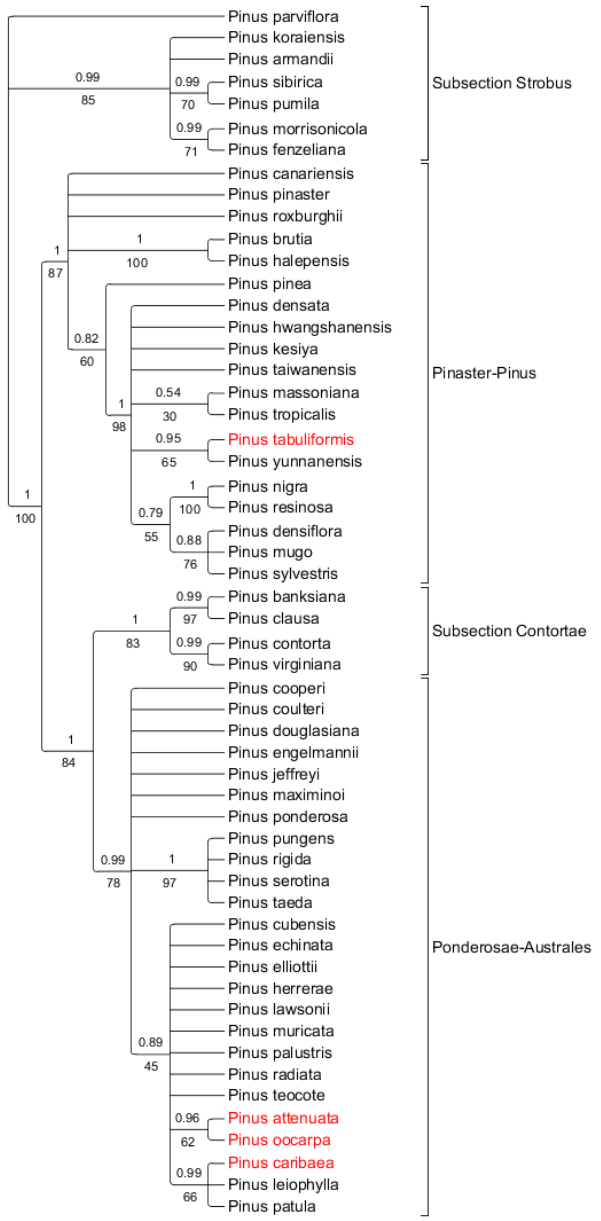

Figure 4. Phylogenetic tree based on the *matK* marker. The tree represents the majority consensus of trees sampled after stationarity in the Bayesian analysis. Posterior probability values from the Bayesian inference are indicated above and the corresponding bootstrap values of the parsimony analysis are shown below when it was applicable. The labels indicating taxonomic divisions into subsections following Gernandt *et al*. (2005) are shown. The taxa in red colour had different positions than in the analysis based on *ycf1*.

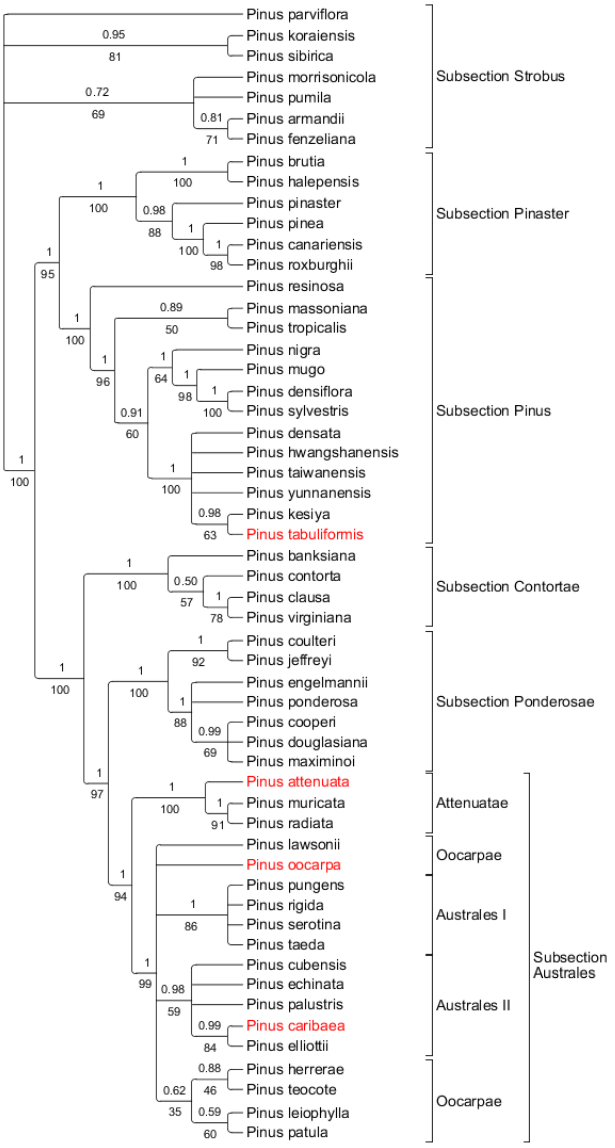

Figure 5. Phylogenetic tree based on the *ycf1* marker. The tree represents the majority consensus of trees sampled after stationarity in the Bayesian analysis. Posterior probability values from the Bayesian inference are indicated above and the corresponding bootstrap values of the parsimony analysis are shown below when it was applicable. The labels indicating taxonomic divisions following Gernandt *et al*. (2005) are shown. The taxa in red colour had different positions than in the analysis based on *matK*.

The present study is the first work which reports phylogenetic incongruences in pines between the chloroplast markers *matK* and *ycf1*. These incongruences were not detected in earlier studies because of the use of a shorter *matK* region resulting in a poorly resolved gene tree (*e.g.* Hernández-León *et al.*, 2013). Previous studies have shown that pine phylogenies based on chloroplast markers may be incongruent with phylogenies based on nuclear markers, as well as morphological and geographical classifications (*e.g.* Liston *et al.*, 2003; Syring *et al.*, 2005; Wilyard *et al.*, 2009; Gernarndt *et al.*, 2018).

One of the disadvantages of using chloroplast markers is chloroplast capture, defined as the movement of a chloroplast genome from one species to another through the process of introgression (Soltis & Soltis,

1998). This phenomenon has negative consequences on both phylogenetic inference and systematic efforts (Tsitrone, *et al.*, 2003), and it has been suggested to occur in pines (Gernarndt *et al.*, 2005; Liston *et al.*, 2007; Gernarndt *et al.*, 2018). Furthermore, different parts of the chloroplast have different phylogenetic topologies (Zeng *et al.*, 2014). To circumvent these limitations, few initiatives focused on developing new nuclear markers for pines (Syring *et al.*, 2005; Palme *et al.*, 2009; Grivet *et al.*, 2013; Gernarndt *et al.*, 2018), but their wide use is limited by the availability of multispecies sequence data from public databases.

Other reasons may impede pine phylogenies, such as reticulate evolution due to hybridization. Gernarndt *et al.* (2018) suggested that hybridization occurred in the Oocarpae ancestors, explaining the difficulties to place them taxonomically. The Oocarpae group appears polyphyletic in our analyses. Hybridization could also explain other aberrant phylogenetic grouping observed in this study in the analysis based on *matK*. While chloroplast markers may not succeed to discriminate species in a group of plants in which reticulate evolution is present, they might result useful to discern hybridization processes in interspecific hybrids by the presence or absence of selected chloroplast markers. The usefulness of the *matK* marker to identify hybrids remains to be investigated.

For all land plants, the establishment of a single DNA region as universal barcode is not a realistic goal, but accurate species delineation may be achieved by combining several *loci* used as barcode (Kress, 2017). However, the rate of successfully identified gymnosperm species using different combinations of the seven main candidate plastid regions for barcoding (*rpoC1, rpoB, rbcL, matK, trnH-psbA, atpF-atpH, psbK-psbI*) is low (Hollingsworth *et al.*, 2009b; Ran *et al.*, 2010). Species delineation with existing chloroplast markers in closely related conifer species is particularly problematic (Ortiz-Martínez & Gernandt, 2016; Celinski *et al.*, 2017). In spite of the challenges to barcode species in the genus *Pinus*, the present study shows that the marker *ycf1* is promising at the species level delineation. Consequently, this marker could be used to solve specific problems, such as the differentiation of the closely related *Pinus nigra*, *Pinus mugo* and *Pinus sylvestris*, which are difficult to identify based solely on wood traits (Schoch *et al.*, 2004).

Due to the importance of species-level identification in pines, it will be useful to further develop barcodes for specific sections and assess how to combine successfully species-level markers with population- or clonal-level markers. There is indeed a huge interest in forestry to identify forest material at the intra-specific level with genetic markers, more specifically to avoid fraud marketing of forest reproductive material (Nanson, 2001; Degen *et al.*, 2010). There already exist some examples of studies, in which material of specific origins at the infraspecific levels have been identified (Aragonés *et al.*, 1997; Ribeiro *et al.*, 2002; Deguilloux *et al.*, 2004; Tigabu *et al.*, 2005; Fidler *et al.*, 2006; Hernandez-Tecles *et al.*, 2017). Therefore, an awaiting challenge is to combine multilevel diagnostic markers that could respond to the many challenges facing forest product traceability.

Acknowledgements

This study formed part of the undergraduate thesis of Jerónimo Cid Vian for the Scool of Forestry and Natural Resources (E.T.S.I.), Madrid Polytechnic University. The authors would like to thank an anonymous reviewer and the associate editor for their constructive comments, which greatly improved the manuscript.

References

Alfaro ME, Zoller S, Lutzoni F, 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol Biol Evol 20: 255-266. https://doi.org/10.1093/molbev/msg028

Álvarez-Álvarez P, Pizarro C, Barrio-Anta M, Cámara-Obregón A, Bueno JLM, Álvarez A, Gutiérrez I, Burslem DFRP, 2018. Evaluation of tree species for biomass energy production in Northwest Spain. Forests 9(4): 160. https://doi.org/10.3390/f9040160

Aragonés A, Barrena I, Espinel S, Herrán A, Ritter E, 1997. Origin of Basque populations of radiata pine inferred from RAPD data. Ann Sci For 54(8): 697-703. https://doi.org/10.1051/forest:19970801

Baker SC, Steel D, Choi Y, Lee H, Kim KS, Choi SK, Ma Y-U, Hambleton C, Psihoyos L, Brownell RL, Funahashi N, 2010. Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. Biol Lett 6(5): 647-650. https://doi.org/10.1098/rsbl.2010.0239

Bruni I, De Mattia F, Martellos S, Galimberti A, Savadori P, Casiraghi M, Nimis PL, Labra M, 2012. DNA Barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy). PLoS ONE 7(9): e43256. https://doi.org/10.1371/journal.pone.0043256

Businský R, Frantik T, Vit P, 2014. Morphological evaluation of the Pinus kesiya complex (Pinaceae). Plant Syst Evol. 300: 273-285. https://doi.org/10.1007/s00606-013-0880-0

Celinski K, Kijak H, Wojnicka-Półtorak A, Buczkowska-Chmielewska K, Sokołowska J, Chudzińska E, 2017.

Effectiveness of the DNA barcoding approach for closely related conifers discrimination: A case study of the Pinus mugo complex. C R Biol 340(6-7): 339-348. https://doi.org/10.1016/j.crvi.2017.06.002

Christensen KI, 1987. Taxonomic revision of the Pinus mugo complex and P. rhaetica (P. mugo sylvestris) (Pinaceae). Nord J Bot 7 383-408. https://doi.org/10.1111/j.1756-1051.1987.tb00958.x

Darriba D, Taboada GL, Doallo R, Posada, 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9(8): 772-772.

Degen B, Höltken A, Rogge M, 2010. Use of DNA-fingerprints to control the origin of forest reproductive material. Silvae Genet 59(6): 268-273. https://doi.org/10.1515/sg-2010-0038

Deguilloux M-F, Pemonge M-H, Petit RJ, 2004. DNA-based control of oak wood geographic origin in the context of the cooperage industry. Ann For Sci 61(1): 97-104. https://doi.org/10.1051/forest:2003089

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S, 2015. ycf1, the most promising plastid DNA barcode of land plants. Sci Rep 5(1): 8348. https://doi.org/10.1038/srep08348

Eckert A, Hall B, 2006. Phylogeny, historical biogeography, and patterns of diversification for Pinus (Pinaceae): Phylogenetic tests of fossil-based hypotheses. Mol Phylogenet Evol 40(1): 166-182. https://doi.org/10.1016/j.ympev.2006.03.009

Fidler F, Burgman MA, Cumming G, Buttrose R, Thomason N, 2006. Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. Conserv Biol 20(5): 1539-1544. https://doi.org/10.1111/j.1523-1739.2006.00525.x

Finkeldey R, Leinemann L, Gailing O, 2010. Molecular genetic tools to infer the origin of forest plants and wood. Appl Microbiol Biotechnol 85: 1251-1258. https://doi.org/10.1007/s00253-009-2328-6

Geada López G, Kamiya K, Harada K, 2002. Phylogenetic relationships of Diploxylon pines (Subgenus Pinus) based on plastid sequence data. Int J Plant Sci 163(5): 737-747. https://doi.org/10.1086/342213

Gernandt D, Liston A, Piñero D, 2003. Phylogenetics of Pinus Subsections Cembroides and Nelsoniae Inferred from cpDNA Sequences. Syst Bot 28(4): 657-673.

Gernandt DS, López GG, García SO, Liston A, 2005. Phylogeny and classification of Pinus. Taxon 54: 29-42. https://doi.org/10.2307/25065300

Gernandt D, Magallón S, Geada López G, Zerón Flores O, Willyard A, Liston A, 2008. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae Phylogeny. Int J Plant Sci: 169(8): 1086-1099. https://doi.org/10.1086/590472

Gernandt DS, Aguirre Dugua X, Vázquez-Lobo A, Willyard A, Moreno Letelier A, Pérez de la Rosa JA, Piñero D, Liston A, 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in Pinus subsection Australes. Am J Bot 105: 1-15. https://doi.org/10.1002/ajb2.1052

Ghorbani A, Gravendeel B, Selliah S, Zarré S, de Boer H, 2017. DNA barcoding of tuberous Orchidoideae: a resource for identification of orchids used in Salep. Mol Ecol Res 17(2): 342-352. https://doi.org/10.1111/1755-0998.12615

Grivet D, Climent J, Zabal-Aguirre M, Neale D, 2013. Adaptive evolution of Mediterranean pines. Mol Phylogenet Evol 68(3): 555-566. https://doi.org/10.1016/j.ympev.2013.03.032

Hernández-León S, Gernandt D, Pérez de la Rosa J, Jardón-Barbolla L, 2013. Phylogenetic relationships and species delineation in Pinus Section Trifoliae inferred from plastid DNA. PLoS ONE 8(7): e70501. https://doi.org/10.1371/journal.pone.0070501

Hernández-Tecles, de las Heras J, Lorenzo Z, Navascués M, Alia R, 2017. Identification of gene pools used in restoration and conservation by chloroplast microsatellite markers in Iberian pine species. For Syst 26(2): e058.

Hollingsworth PM, Forrest L, Spouge J, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, *et al.*, 2009a. A DNA barcode for land plants. Proc Natl Acad Sci, 106(31): 12794-12797. https://doi.org/10.1073/pnas.0905845106

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM, 2009b. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9: 439-457. https://doi.org/10.1111/j.1755-0998.2008.02439.x

Katoh K, Standley D, 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. Mol Biol Evol 30(4): 772-780. https://doi.org/10.1093/molbev/mst010

Kress W, 2017. Plant DNA barcodes: Applications today and in the future. JSE IBC Special Issue on Frontiers in Plant Systematics and Evolution. J Syst Evol 55(4): 291-307. https://doi.org/10.1111/jse.12254

Kress W, Erickson D, Jones F, Swenson N, Perez Sanjur O, Bermingham E, 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci 106(44): 18621-18626. https://doi.org/10.1073/pnas.0909820106

Lidder P, Sonnino A, 2011. Background study paper no. 52. Biotechnologies for the management of genetic resources for food and agricuture. Commission on Genetic Resources for Food and Agriculture.

Liston A, Gernandt DS, Vining TF, Campbell CS, Piñero D, 2003. Molecular phylogeny of Pinaceae and Pinus. Acta Hort 615: 107-114. https://doi.org/10.17660/ActaHortic.2003.615.7

Liston A, Parker-Defeniks M, Syring JV, Willyard A, Cronn R, 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: a case study in Pinus lambertiana. Mol Ecol 16: 3926-3937. https://doi.org/10.1111/j.1365-294X.2007.03461.x

Miller MA, Pfeiffer W, Schwartz T, 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14th November, 2010, New Orleans, LA, 1-8. https://doi.org/10.1109/GCE.2010.5676129

Moreira X, Sampedro L, Zas R, Pearse I, 2016. Defensive traits in young pine trees cluster into two divergent syndromes related to early growth rate. PLOS ONE: 11(3): e0152537. https://doi.org/10.1371/journal.pone.0152537

Müller KF, Quandt D, Müller J, Neinhuis C, 2005. PhyDE ® 0.995: Phylogenetic Data Editor. http://www.phyde.de.

Nanson A, 2001. The new OECD scheme for the certification of forest reproductive materials. Silvae Genet 50(5-6): 181-187.

Nielsen LR, Dahl KE, 2008. Tracing Timber from Forest to Consumer with DNA Markers. Copenhagen: Danish Ministry of the Environment, Forest and Nature Agency. Available from http://www.skovognatur.dk/udgivelser

Ortiz-Martínez, Gernandt DS, 2016. Species diversity and plastid DNA haplotype distributions of Pinus subsection Australes (Pinaceae) in Guerrero and Oaxaca. TIP Rev Esp Cienc Quím Biol 19(2): 92-101.

Palmé AE, Pyhäjärvi T, Wachowiak W, Savolainen O, 2009. Selection on nuclear genes in a Pinus phylogeny. Mol Biol Evol 26(4): 893-905. https://doi.org/10.1093/molbev/msp010

Parks M, Cronn R, Liston A, 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol 7: 84. https://doi.org/10.1186/1741-7007-7-84

Parks M, Liston A, Cronn R, 2011. Newly developed primers for complete ycf1 amplification in Pinus (Pinaceae) chloroplasts with possible family-wide utility. Am J Bot 98(7): e185-188. https://doi.org/10.3732/ajb.1100088

Price RA, Liston A, Strauss SH, 1998. Phylogeny and systematics of Pinus. In: Ecology and Biogeography of Pinus; Richardson DM (ed.). pp. 49-68. Cambridge University Press, NY, USA.

Rambaut A, Suchard MA, Xie D, Drummond AJ, 2014. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer

Ran JH, Wang PP, Zhao HJ, Wang XQ, 2010. A test of seven candidate barcode regions from the plastome in Picea (Pinaceae). J Integr Plant Biol 52: 1109-1126. https://doi.org/10.1111/j.1744-7909.2010.00995.x

Ribeiro MM, Le-Provost G, Gerber S, Vendramin GG, Anzidei M, Decroocq S, Marpeau A, Mariette S, Plomion C, 2002. Origin identification of maritime pine stands in France using chloroplast simple-sequence repeats. Ann Forest Sci. 59(1): 53-62. https://doi.org/10.1051/forest:2002100

Richardson DM, Rundel, 1998. Pine ecology and biogeography – An introduction. In: Ecology and Biogeography of Pinus; Richardson DM (ed.). pp. 49-68. Cambridge University Press, NY, USA. pp 3-46. Cambridge University Press, NY, USA.

Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP, 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539-542. https://doi.org/10.1093/sysbio/sys029

Schoch W, Heller I, Schweingruber FH, Kienast F, 2004. Wood anatomy of central European Species. Online version http:www.woodanatomy.ch

Soltis D, Soltis P, 1998. Molecular systematics of plants II. DNA sequencing. Kluwer Academic Publishers, NY, USA. 42 pp. https://doi.org/10.1007/978-1-4615-5419-6

Stamatakis A, Hoover P and Rougemont J, 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 57(5): 758-771. https://doi.org/10.1080/10635150802429642

Stöver BC, Müller KF, 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11: 7. https://doi.org/10.1186/1471-2105-11-7

Swofford DL, 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0. Sinauer Assoc, Sunderland, MA, USA.

Syring J, Willyard A, Cronn R, Liston A, 2005. Evolutionary relationships among Pinus (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot 92(12): 2086-2100. https://doi.org/10.3732/ajb.92.12.2086

Tigabu M, Oden PC, Lindgren D, 2005. Identification of seed sources and parents of Pinus sylvestris L. using visible-near infrared reflectance spectra and multivariate analysis. Trees-Struct Funct 19: 468-476. https://doi.org/10.1007/s00468-005-0408-5

Tnah LH 2009, Lee SL, Ng KKS, Tani N, Bhassu S, Othman RY, 2009. Geographical traceability of an important tropical timber (Neobalanocarpus heimii) inferred from chloroplast DNA. Forest Ecol Manag 258(9): 1918-1923. https://doi.org/10.1016/j.foreco.2009.07.029

Tnah LH, Lee Soon LL, Ng KKS, Zaman FQ, Faridah-Hanum I, 2010. Forensic DNA profiling of tropical timber species in Peninsular Malaysia. Forest Ecol Manag 259(8): 1436-1446. https://doi.org/10.1016/j.foreco.2010.01.017

Tsitrone A, Kirkpatrick M, Levin D, 2003. A model for chloroplast capture. Evolution 57(8): 1776. https://doi.org/10.1111/j.0014-3820.2003.tb00585.x

Volkmann L, Martyn I, Moulton V, Spillner A, Mooers AO, 2014. Prioritizing populations for conservation using

phylogenetic networks. PLoS ONE 9(2): e88945. https://doi.org/10.1371/journal.pone.0088945

Wang XR, Tsumura Y, Yoshimaru H, Nagasaka K, Szmidt AE, 1999. Phylogenetic relationships of Eurasian pines (Pinus, Pinaceae) based on chloroplast rbcL, MATK, RPL20-RPS18 spacer, and TRNV intron sequences. Am J Bot 86(12): 1742-53. https://doi.org/10.2307/2656672

Willyard A, Cronn R, Liston A, 2009. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. Mol Phylogenet Evol 52: 498-511. https://doi.org/10.1016/j.ympev.2009.02.011

Xu J-H, Liu Q, Hu W, Wang T, Xue Q, Messing J, 2015. Dynamics of chloroplast genomes in green plants. Genomics 106(4): 221-231. https://doi.org/10.1016/j.ygeno.2015.07.004

Yang RC, Yeh FC, Ye TZ, 2007. Multilocus structure in the Pinus contorta-Pinus banksiana complex. Can J Bot 85: 774-784. https://doi.org/10.1139/B07-054

Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H, 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nat Commun 5:4956. https://doi.org/10.1038/ncomms5956