

Accésit de la SECF 2018 al mejor Proyecto o Trabajo Fin de Carrera o de Grado

Evaluación de las regiones del cloroplasto *matK* e *ycfl* como marcadores de diagnóstico para el género *Pinus*

Evaluation of the chloroplast regions *matK* and *ycfl* as diagnostic markers for the genus *Pinus*

Cid, J.^{1,2*}; Grivet, D.²; Olsson, S.²; Fernández, M.V.¹

¹*ETSI Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Ciudad Universitaria s/n 28040 Madrid, España.*

²*Departamento de Ecología y Genética Forestal, INIA. Centro de Investigación Forestal Carretera de La Coruña km 7.5. 28040 Madrid, España.*

Autor para correspondencia: jeronimocidvian@gmail.com

Resumen

A pesar de la importancia ecológica y económica mundial del género *Pinus*, aún no hay una filogenia de consenso del género que llegue hasta el nivel de especie. Asimismo tampoco existe un marcador “código de barras” de consenso para las plantas terrestres. Esto se debe principalmente a la reducida variabilidad genética de *Pinus* y/o a la inadecuada selección de marcadores. Para obtener mejores resultados, proponemos que el marcador *matK* se utilice en su longitud total en lugar de solo la región “código de barras” habitualmente utilizada. También evaluamos la variabilidad y susceptibilidad a la saturación de un marcador recientemente propuesto, *ycf1*.

Se descargaron todas las secuencias de *matK* presentes en GenBank del género *Pinus* de longitud igual o superior a 1600 nucleótidos, y todas las secuencias de *ycf1* para esas especies. Las estadísticas de los alineamientos de secuencias resultantes fueron analizadas y comparadas con otros estudios. A continuación se realizaron análisis filogenéticos mediante el método bayesiano y el de máxima verosimilitud. También se hizo un estudio de saturación del marcador *ycf1*.

El desempeño de *matK* mejoró significativamente al utilizar la longitud total, corroborando la hipótesis inicial. Otro hallazgo fue que la variabilidad de *matK* se reparte de manera relativamente uniforme por toda la longitud del marcador. Como se esperaba *ycf1* tuvo un nivel de variabilidad mucho mayor. No se constató saturación en *ycf1* pero se recomienda precaución en su uso por encima de la categoría taxonómica del género. También se comentan dos conflictos encontrados entre la señal filogenética de *matK* e *ycf1*.

Palabras clave: bioinformática, código de barras para la vida, ecología evolutiva, filogenia.

Abstract

The economic and ecological importance of genus *Pinus* has been universally recognised. However, there is still not a good consensus phylogeny that goes down to the species level. In addition there is not a consensus barcode marker for land plants. This is mainly due to low genetic variability in pines and/or inadequate selection of molecular markers. We propose the *matK* marker to be used in the whole length of the gene for better results instead of only the barcode region of *matK* used in many studies. We also evaluate the variability and substitution saturation of a new marker recently proposed, *ycf1*, and search for conflicting phylogenetic signals between the two markers.

We downloaded all the *matK* sequences for *Pinus* available on GenBank longer than 1600 base pairs and all *ycf1* sequences for those species. The statistics of the resulting alignments were analysed and compared to other studies. The alignments were then subjected to two types of phylogenetic analyses, maximum likelihood and bayesian inference.

The phylogenetic and barcode performance of the *matK* marker significantly increased when using its whole length. The phylogenetic performance of *ycf1* was not significantly affected by the removal of the supposedly saturated region but could have saturation issues if used above the genus level. Globally the *ycf1* marker was found to be more useful than *matK* for phylogenetic inference of pines as well as to serve as a barcode marker for the genus. Finally we provide a recommendation on how the two markers should be used for better resolved pine phylogenies.

Keywords: barcode for life, bioinformatics, evolutionary ecology, phylogeny.

1. Introducción

El género *Pinus* es uno de los más importantes del planeta ecológica y económicamente (Price *et al.*, 1998). Sus especies son componentes mayoritarios y muchas veces dominantes de muchos ecosistemas del hemisferio norte, donde cumplen importantes funciones ecológicas y hasta reguladoras del clima global. Especies de este género componen el 54% del área dedicada a plantaciones forestales en las regiones templadas y el 26% en las regiones tropicales (Brown, 2000), por lo que también se han plantado extensamente fuera de su rango ecológico. Los pinos son fuente de numerosos e importantes bienes, como madera, papel, fibras, resinas, carbón vegetal, alquitrán, taninos, frutos y usos ornamentales (Sealy-Fisher and Pizzi, 1992; Price *et al.*, 1998; Gernandt *et al.*, 2005; Earle, 2017). También nos proveen de numerosos servicios más difusos como captura de carbono, fijación del suelo, información como bioindicadores o su interés científico per se (Micieta and Murín, 1998; Grotkopp *et al.*, 2004; Morse, 2009; Wang and Ran, 2014). Sin embargo aún hay preguntas sin respuesta en la filogenia del género *Pinus*.

En las últimas décadas el genoma del género *Pinus* ha sido objeto de investigación con el fin de discernir su historia evolutiva. Los métodos filogenéticos basados en técnicas moleculares han resuelto muchos interrogantes en la circunscripción de especies a sus respectivos subgéneros, secciones y subsecciones. Sin embargo, y a pesar de los avances, no existe aún una filogenia consensuada que discierna las relaciones más recientes en la historia evolutiva de los pinos, es decir, que llegue hasta el nivel de especie (Gernandt *et al.*, 2009). Esto se debe principalmente a la reducida variabilidad genética de *Pinus* y/o a la inadecuada selección de marcadores moleculares o genéticos (zonas variables del genoma).

Otro ámbito en el que los marcadores moleculares y los pinos podrían ser de utilidad es en el proyecto *barcode* (código de barras). Esta iniciativa aspira a poder identificar cualquier muestra de ADN hasta el nivel de especie mediante el uso de marcadores moleculares. Algunos grupos como las aves o los peces ya tienen proyectos en marcha de búsqueda de marcadores de código de barras efectivos (Ratnasinghan, 2007), pero todavía no existe un marcador o conjunto de marcadores utilizable para las plantas terrestres (Hollingsworth *et al.*, 2009). Para ayudar a aclarar las dudas restantes en la filogenia de los pinos y avanzar en la identificación de pinos mediante marcadores código de barras será necesario establecer marcadores moleculares con un grado de variabilidad adecuado al nivel taxonómico para el que se desee averiguar información.

Previo a este estudio se han identificado dos marcadores moleculares del ADN del cloroplasto, denominados *matK* e *ycf1*. El primer marcador, *matK* ha sido largamente utilizado en *Pinus*. En este estudio se planteaba la posibilidad de que su rendimiento pudiera aumentar si se utilizara la totalidad del gen (unos 1600 nucleótidos) en lugar de la región habitualmente utilizada, la llamada región código de barras (unos 600 nucleótidos). Por su parte, dos estudios recientes (Hernández-León *et al.*, 2013; Dong *et al.*, 2015) han señalado al segundo marcador, *ycf1* (≈ 5400 nucleótidos) como notablemente variable a nivel de especie en todo el género *Pinus*; sin embargo podría causar problemas de saturación a los métodos estadísticos de

inferencia filogenética, consecuencia de una región de alta variabilidad (208 nucleótidos) encontrada durante los análisis preliminares.

2. Objetivos

Los objetivos generales de este trabajo son contribuir a la filogenia del género *Pinus* así como a la identificación de sus especies mediante un estudio bioinformático de dos marcadores genéticos conocidos. Los objetivos específicos son:

1. Evaluar la idoneidad del marcador del cloroplasto *yefl* para:
 - a) resolver la filogenia de los pinos.
 - b) servir como marcador de código de barras.
2. Comparar el desempeño del marcador *matK* en la resolución filogenética del género *Pinus* cuando se usa la longitud total del gen *vs.* cuando se utiliza la región código de barras.
3. Identificar señales filogenéticas conflictivas entre *matK* e *yefl*.
4. Proporcionar recomendaciones sobre qué marcadores utilizar para análisis filogenéticos en *Pinus*.

3.1. Materiales: Secuencias y programas utilizados

Para probar estas hipótesis se descargaron todas las secuencias de *matK* presentes en el servidor global GenBank (Benson *et al.*, 2014) del género *Pinus* de longitud igual o superior a 1600 nucleótidos, y todas las secuencias del marcador *yefl* para esas especies. Esta búsqueda arrojó 55 secuencias de especies del género *Pinus* para las que se satisfacían los criterios arriba descritos (*Tab. 1*).

A continuación se sometieron las secuencias a un estudio bioinformático en el cual se emplearon diversos softwares gráficos, de cálculo y de visualización. Los programas utilizados en este estudio fueron: PhyDE (Müller *et al.*, 2010), MAFFT (Kato and Standley, 2013), JModeltest (Darriba *et al.*, 2012), MrBayes (Huelsenbeck and Ronquist, 2001), Tracer (Rambaut *et al.*, 2014), RAxML (Stamatakis, 2014), TreeGraph (Stöver, and Müller, 2010), PAUP* (Swofford, 2000) y el lenguaje de programación Python. Tres de los programas (JModeltest, MrBayes, RAxML) están disponibles en el portal científico CIPRES (Miller *et al.*, 2010).

3.2. Métodos: Fundamentos y Procedimiento

Las secuencias descargadas se combinaron de manera metódica para dar lugar a seis alineamientos diferentes de 55 especies que reflejaran las distintas condiciones de estudio que se querían poner a prueba (*Tab. 2*).

Tabla 1. Lista de especies y números de acceso de GenBank de todas las secuencias utilizadas. Los datos se obtuvieron de la base de datos de secuencias (National Center for Biotechnology Information – US National Library of Medicine – National Institute of Health, 2017). Los asteriscos indican las secuencias donde la región *ycf1* se extrajo del genoma del cloroplasto global o parcial.

ID	Nombre de la especie	Número de Acceso GenBank		ID	Nombre de la especie	Número de Acceso GenBank	
		<i>matK</i>	<i>ycf1</i>			<i>matK</i>	<i>ycf1</i>
1	<i>Pinus parviflora</i>	AB081086	KP089941	29	<i>Pinus lawsonii</i>	AB097784	KC157176
2	<i>Pinus koraiensis</i>	AB161009	AY228468*	30	<i>Pinus leiophylla</i>	AB081085	KC157132
3	<i>Pinus armandii</i>	AB161002	KP089404	31	<i>Pinus massoniana</i>	AB081088	KC427272*
4	<i>Pinus sibirica</i>	AB161014	FJ899558*	32	<i>Pinus maximinoi</i>	AB161010	KC157109
5	<i>Pinus morrisonicola</i>	AF295031	JN854182*	33	<i>Pinus mugo</i>	AB081087	JN854181*
6	<i>Pinus fenzeliana</i>	AB161005	KX255674*	34	<i>Pinus muricata</i>	AB080935	KC157153
7	<i>Pinus pumila</i>	AB161013	JN854168*	35	<i>Pinus nigra</i>	AB084498	KP089411
8	<i>Pinus attenuata</i>	AB080933	KC157134	36	<i>Pinus oocarpa</i>	AB081084	KC157158
9	<i>Pinus banksiana</i>	AB080922	KP089408	37	<i>Pinus palustris</i>	AB080937	KC157163
10	<i>Pinus brutia</i>	AB161018	KP089932	38	<i>Pinus patula</i>	AB080944	KP089936
11	<i>Pinus canariensis</i>	AB084494	KP089933	39	<i>Pinus pinaster</i>	AB084493	FJ899583*
12	<i>Pinus caribaea</i>	AB080940	JN854222*	40	<i>Pinus pinea</i>	AB084496	JN854173*
13	<i>Pinus clausa</i>	AB161003	KC157159	41	<i>Pinus ponderosa</i>	AB080924	KC157177
14	<i>Pinus contorta</i>	AB080921	FJ580260	42	<i>Pinus pungens</i>	AB080932	JN854167*
15	<i>Pinus cooperi</i>	AB161004	FJ580193	43	<i>Pinus radiata</i>	AB080934	KC157207
16	<i>Pinus coulteri</i>	AB097785	JN854215*	44	<i>Pinus resinosa</i>	AB080945	KC157078
17	<i>Pinus cubensis</i>	AB080938	KC157114	45	<i>Pinus rigida</i>	AB080929	KC157183
18	<i>Pinus densata</i>	AB097779	JN854209*	46	<i>Pinus roxburghii</i>	AB084495	JN854162*
19	<i>Pinus densiflora</i>	AB084497	KP089385	47	<i>Pinus serotina</i>	AB080930	KC157199
20	<i>Pinus douglasiana</i>	AB080925	KJ152831	48	<i>Pinus sylvestris</i>	AB097781	KP089937
21	<i>Pinus echinata</i>	AB080936	KC157152	49	<i>Pinus tabuliformis</i>	AB161015	KP089380
22	<i>Pinus elliottii</i>	FM955321	JN854202*	50	<i>Pinus taeda</i>	AB080928	KC157197
23	<i>Pinus engelmannii</i>	AB080927	FJ580207	51	<i>Pinus taiwanensis</i>	AB161016	KP771703*
24	<i>Pinus halepensis</i>	AB081089	JN854197*	52	<i>Pinus teocote</i>	AB097783	KC157202
25	<i>Pinus herrerae</i>	AB080943	KC157155	53	<i>Pinus tropicalis</i>	AB080920	JN854156*
26	<i>Pinus hwangshanensis</i>	AB161007	JN854194*	54	<i>Pinus virginiana</i>	AB080923	KC157196
27	<i>Pinus jeffreyi</i>	AB080926	KC157181	55	<i>Pinus yunnanensis</i>	AB161017	JN854151*
28	<i>Pinus kesiya</i>	AB161008	JN854191*				

Tabla publicada en:

Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

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

Tabla 2. Representación gráfica de los seis alineamientos confeccionados a partir de los dos marcadores. Leyenda: El marcador *matK* se representa en **rojo** excepto la región *barcode*, que aparece en **verde**. El marcador *ycf1* se representa en **azul** excepto su región de alta variabilidad, para la que se emplea **amarillo**.

Nr. de alineamiento	Qué incluye	Longitud total (nucleótidos)	Representación
1	El marcador <i>matK</i> entero	1668	
2	El marcador <i>ycf1</i> entero	2863	
3	La región "barcode" del marcador <i>matK</i>	583	
4	El marcador <i>ycf1</i> sin la región de alta variabilidad	2655	
5	Combinación de los alineamientos nr. 1 y nr. 2	4530	
6	Combinación de los alineamientos nr. 1 y nr. 4	4322	

Tabla publicada en:

Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

Para responder a las preguntas iniciales de la forma más completa posible se evaluó a los marcadores según tres criterios diferentes. Por un lado, se evaluaron los árboles filogenéticos que produjeron los alineamientos representando un criterio holístico, pues el principal uso de estos marcadores en este estudio es precisamente producir árboles filogenéticos de calidad. Por otro lado, se hizo un análisis comparativo de las estadísticas de los alineamientos, representando un criterio más reduccionista. Por último, se hizo un análisis de saturación como criterio específico para detectar un problema frecuente en los marcadores filogenéticos.

Los seis alineamientos se sometieron a los mismos análisis mediante dos métodos filogenéticos: el bayesiano y el de máxima verosimilitud. Esto es habitual en

filogenética y se hace con el fin de que los árboles combinados carezcan de las debilidades particulares de cada método. Según este criterio, la utilidad de los marcadores se evaluó principalmente en función de la calidad de los árboles filogenéticos que produjeron, en concreto el nivel de apoyo estadístico de todos los clados y la resolución de los clados terminales.

Las estadísticas de los diferentes alineamientos se analizaron con vistas a responder a las preguntas de los objetivos, comparándolas sistemáticamente con las estadísticas de otros cuatro estudios que utilizaron esos marcadores en *Pinus* (Wang *et al.*, 1999; Gernandt *et al.*, 2005; Hernández-León *et al.*, 2013; Dong *et al.*, 2015).

También se hizo un estudio de saturación para determinar si el marcador *ycf1* tenía problemas de saturación, y con el fin de delimitar el ámbito de aplicación taxonómico de cada marcador.

En este trabajo también se hace una introducción extensa para aclarar toda la terminología necesaria que permite entender las razones científicas y estadísticas existentes detrás de todo el proceso.

4. Resultados y discusión

A continuación se exponen brevemente los resultados principales.

4.1. Estadísticas de los alineamientos

Las variables medidas se explican en la *Tabla 3*, y las estadísticas se muestran en la *Tabla 4*.

Tabla 3. Resumen de las variables utilizadas en el análisis de las estadísticas y en la comparación con otros estudios.

Variable	Abreviatura	Definición
Número de nucleótidos (pares de bases)	bp	El número de columnas del alineamiento, correspondiente cada una a un nucleótido o un par de bases de ADN.
Número de sitios variables	VS	El número de sitios para los que al menos un taxón tiene un nucleótido diferente a los demás.
Porcentaje de sitios variables	%VS	El porcentaje de sitios variables respecto al total de sitios (columnas) del alineamiento
Número de sitios informativos	PIS	El número de sitios para los que al menos dos taxones tienen un nucleótido diferente a los demás.
Porcentaje de sitios informativos	%PIS	El porcentaje de sitios informativos respecto al total de sitios (columnas) del alineamiento

Tabla publicada en:

Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

Tabla 4. Estadísticas de los alineamientos y de los cuatro autores con los que se compararon.

Referencia	Marcador	Alineamiento	Nr. de taxones	Longitud (bp)	Sitios Variables		Sitios informativos		
					Nr.	%	Nr.	%	
Este estudio	<i>matK</i>	1. <i>matK</i> _Entero	55	1668	137	8.3	97	5.8	
	<i>matK</i>	<i>matK</i> _Entero Subgen. <i>Pinus</i>	48	1668	88	5.3	48	2.9	
	<i>matK</i>	<i>matK</i> _Entero Subgen. <i>Strobus</i>	7	1668	7	0.4	2	0.1	
	<i>matK</i>	<i>matK</i> _Entero Subsect. <i>Australes</i>	9	1668	10	0.6	5	0.3	
	<i>matK</i>	<i>matK</i> _Entero Subsect. <i>Contortae</i>	4	1668	4	0.2	3	0.2	
	<i>matK</i>	<i>matK</i> _Entero Subsect. <i>Ponderosae</i>	7	1668	1	0.1	0	0.0	
	<i>matK</i>	<i>matK</i> _Entero Sect. <i>Trifoliae</i>	29	1668	28	1.7	15	0.9	
	<i>ycf1</i>	2. <i>ycf1</i> _Entero	55	2863	629	22.0	502	17.5	
	<i>ycf1</i>	<i>ycf1</i> _Entero Subsect. <i>Australes</i>	9	2863	19	0.7	6	0.2	
	<i>ycf1</i>	<i>ycf1</i> _Entero Subsect. <i>Contortae</i>	4	2863	26	0.9	2	0.1	
	<i>ycf1</i>	<i>ycf1</i> _Entero Subsect. <i>Ponderosae</i>	7	2863	16	0.6	10	0.4	
	<i>ycf1</i>	<i>ycf1</i> _Entero Sect. <i>Trifoliae</i>	29	2863	151	5.3	110	3.8	
	<i>matK</i>	3. <i>matK</i> _región barcode	55	583	55	9.4	44	7.5	
	<i>ycf1</i>	4. <i>ycf1</i> _sin región de alta variabilidad	55	2655	555	20.9	435	16.4	
	<i>matK</i> + <i>ycf1</i>	5. Combinación_Entero	55	4530	767	16.9	600	13.2	
	<i>matK</i> + <i>ycf1</i>	6. Combinación_sin región de alta variabilidad	55	4322	693	16.0	533	12.3	
	<i>ycf1</i>	Región de alta variabilidad del alineamiento 2	55	208	74	35.6	67	32.2	
	Wang et al., 1999	<i>matK</i>	Subgen <i>Pinus</i>	17	1052	58	5.5	19	1.8
		<i>matK</i>	Subgen <i>Strobus</i>	15	1046	23	2.2	14	1.3
		<i>matK</i>	Género <i>Pinus</i>	32	1052	91	8.7	51	4.9
Gernandt et al., 2005	<i>matK</i>	<i>matK</i>	101	1555	182	11.7	157	10.1	
	<i>matK</i>	Subsect. <i>Australes</i>	93	812	3	0.4	3	0.4	
	<i>matK</i>	Subsect. <i>Contortae</i>	10	812	0	0.0	0	0.0	
	<i>matK</i>	Subsect. <i>Ponderosae</i>	85	812	6	0.7	3	0.4	
	<i>matK</i>	Sect. <i>Trifoliae</i>	188	812	10	1.2	7	0.9	
	<i>ycf1</i>	Subsect. <i>Australes</i>	95	2484	91	3.7	81	3.3	
Hernández-León et al., 2013	<i>ycf1</i>	Subsect. <i>Contortae</i>	10	2430	45	1.9	35	1.4	
	<i>ycf1</i>	Subsect. <i>Ponderosae</i>	86	2451	37	1.5	28	1.1	
	<i>ycf1</i>	Sect. <i>Trifoliae</i>	191	2541	200	7.9	184	7.2	
	<i>matK</i>	<i>matK</i>	131	795	79	9.9	61	7.7	
Dong et al., 2015	<i>ycf1</i>	<i>ycf1</i> b	138	1654	498	30.1	435	26.3	

Tabla publicada en:

Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

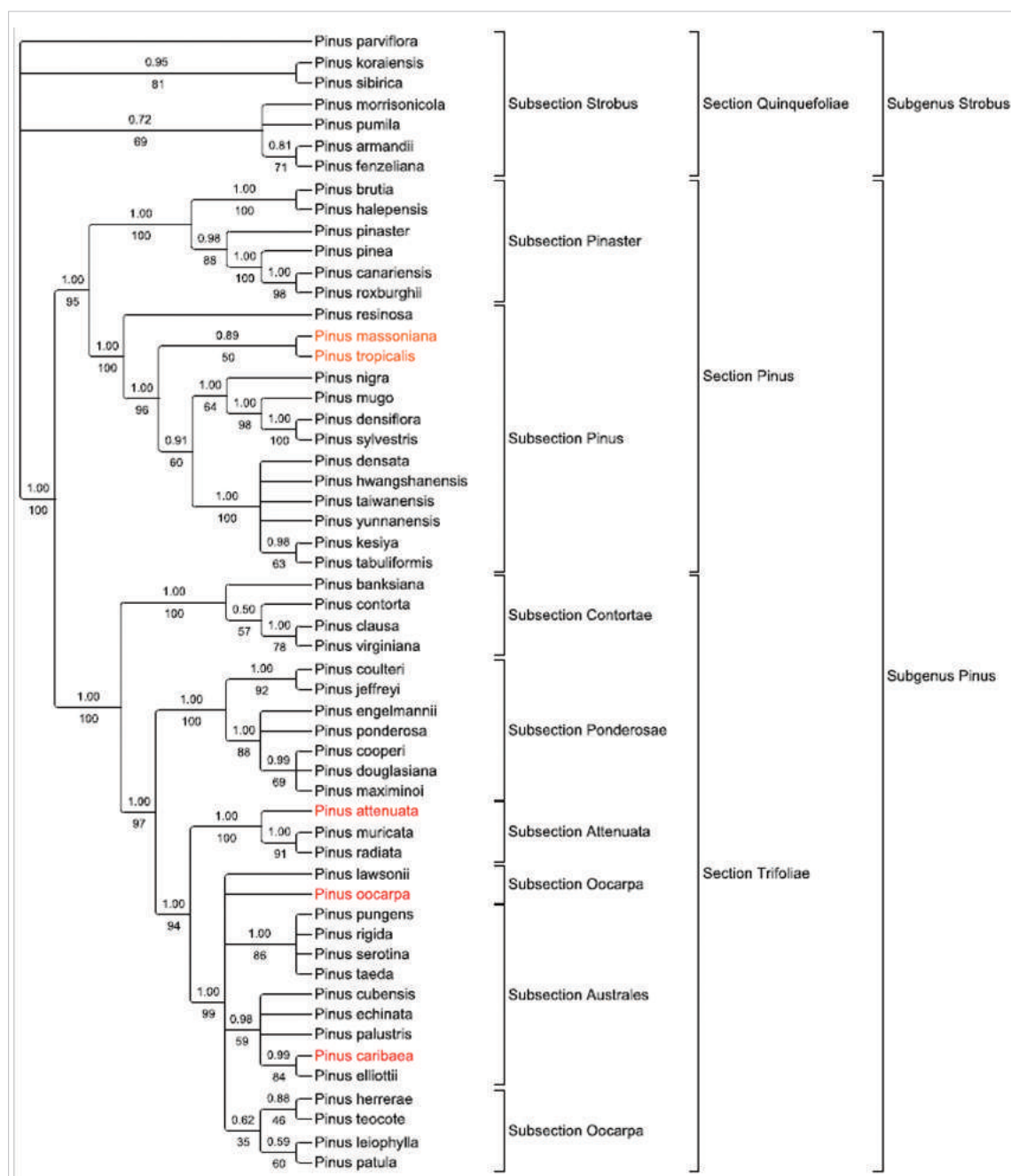


Figura 1. Árbol combinado entero proveniente del análisis del alineamiento 5. Los valores de apoyo del método bayesiano se muestran encima de la rama como fracción de 1, mientras que los correspondientes al método de máxima verosimilitud se muestran bajo la rama como un porcentaje. Los subgéneros, secciones y subsecciones se han marcado según Gernandt *et al.* (2005). Las especies para las que hay un conflicto se han marcado en rojo, mientras que las diferencias intra-subsección sin apoyo se han marcado en naranja.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Con algunas modificaciones gráficas en Olsson, S.; Gri-vet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

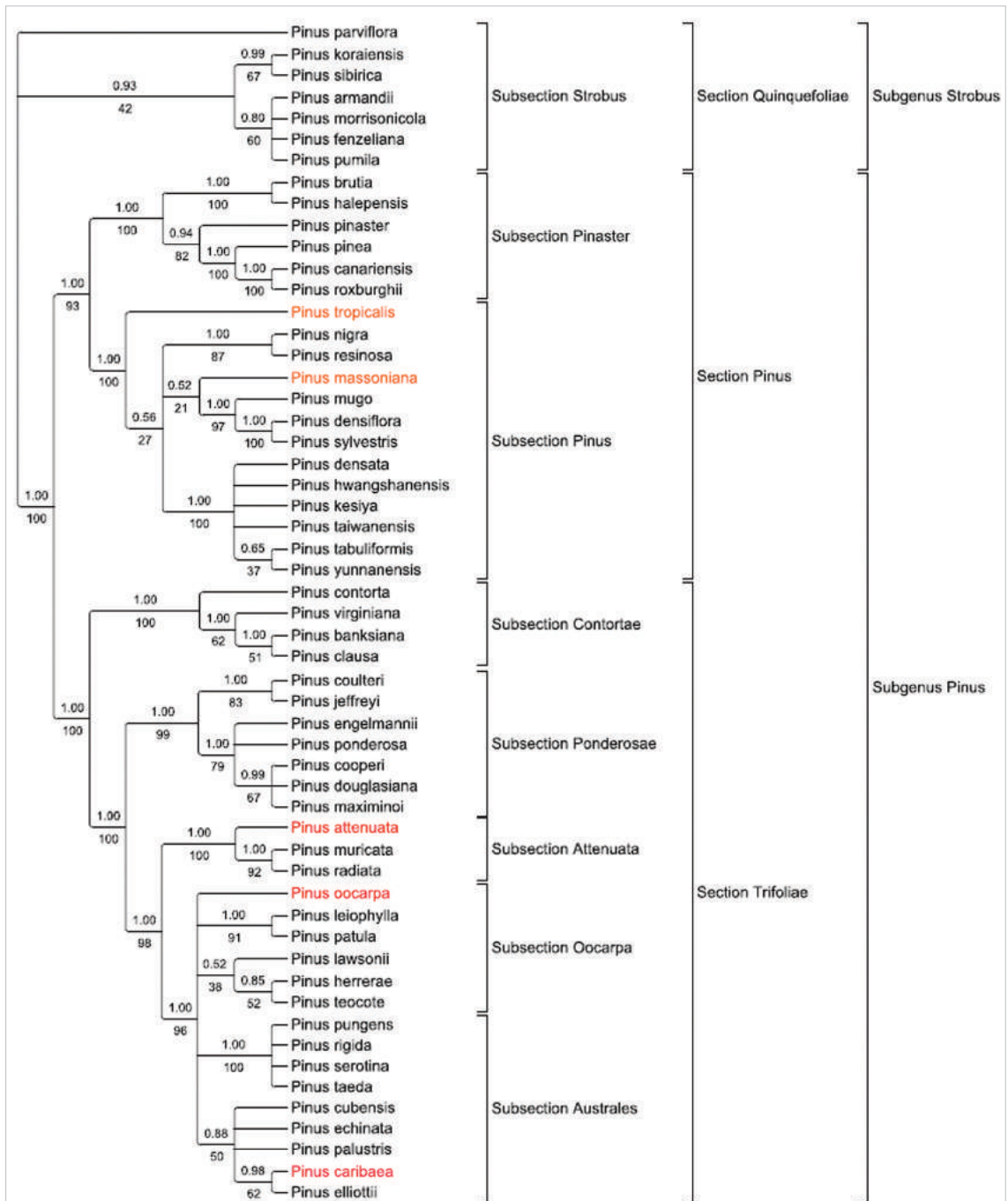


Figura 2. Árbol combinado de *matK* entero e *ycf1* sin la región de alta variabilidad proveniente del análisis del alineamiento 6. Los valores de apoyo del método bayesiano se muestran encima de la rama como fracción de 1, mientras que los correspondientes al método de máxima verosimilitud se muestran bajo la rama como un porcentaje. Los subgéneros, secciones y subsecciones se han marcado según Gernandt *et al.* (2005). Las especies para las que hay un conflicto se han marcado en rojo, mientras que las diferencias intra-subsección sin apoyo se han marcado en naranja.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

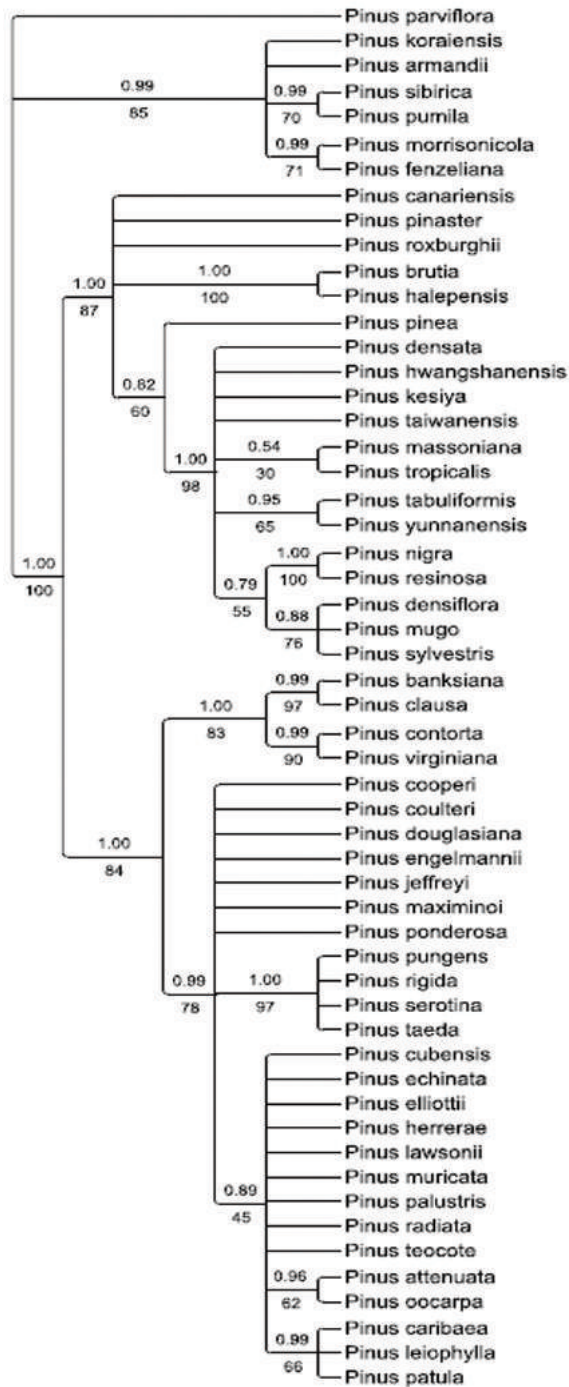


Figura 3. Árbol de matK entero proveniente del análisis del alineamiento 1. Los valores de apoyo del método bayesiano se muestran encima de la rama como fracción de 1, mientras que los correspondientes al método de máxima verosimilitud se muestran bajo la rama como un porcentaje.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Con algunas modificaciones gráficas en Olsson, S.; Gri-vet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

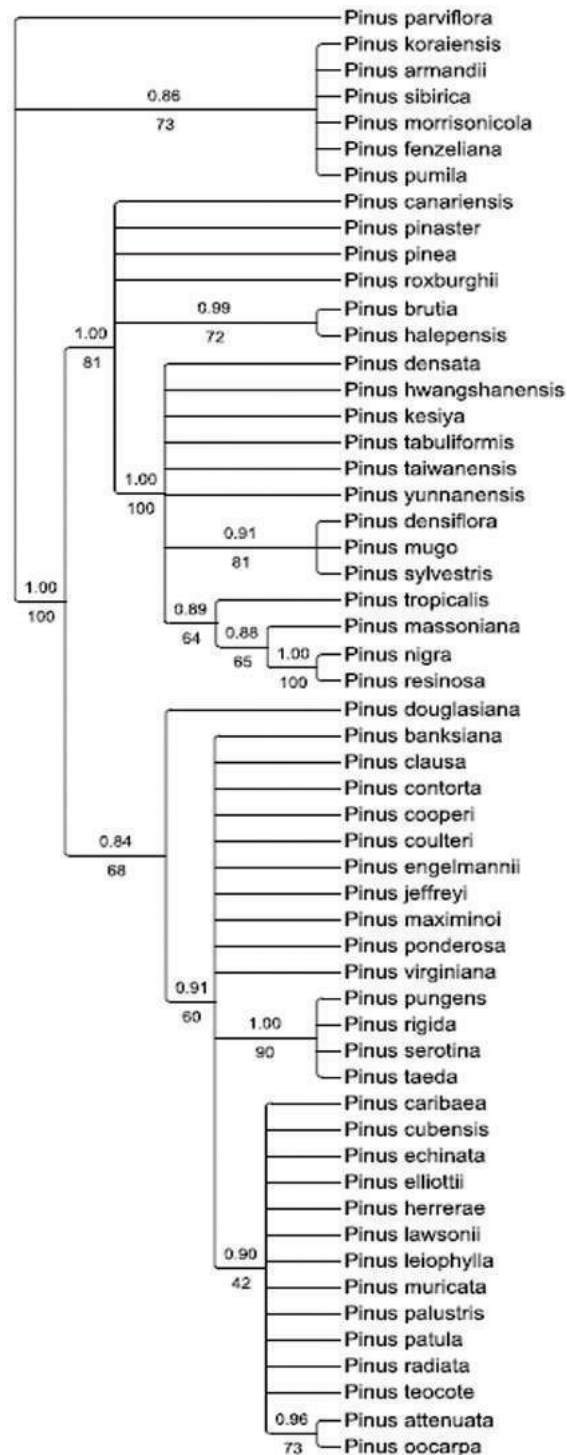


Figura 4. Árbol de la región *barcode* de *matK* proveniente del análisis del alineamiento 3. Los valores de apoyo del método bayesiano se muestran encima de la rama como fracción de 1, mientras que los correspondientes al método de máxima verosimilitud se muestran bajo la rama como un porcentaje.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

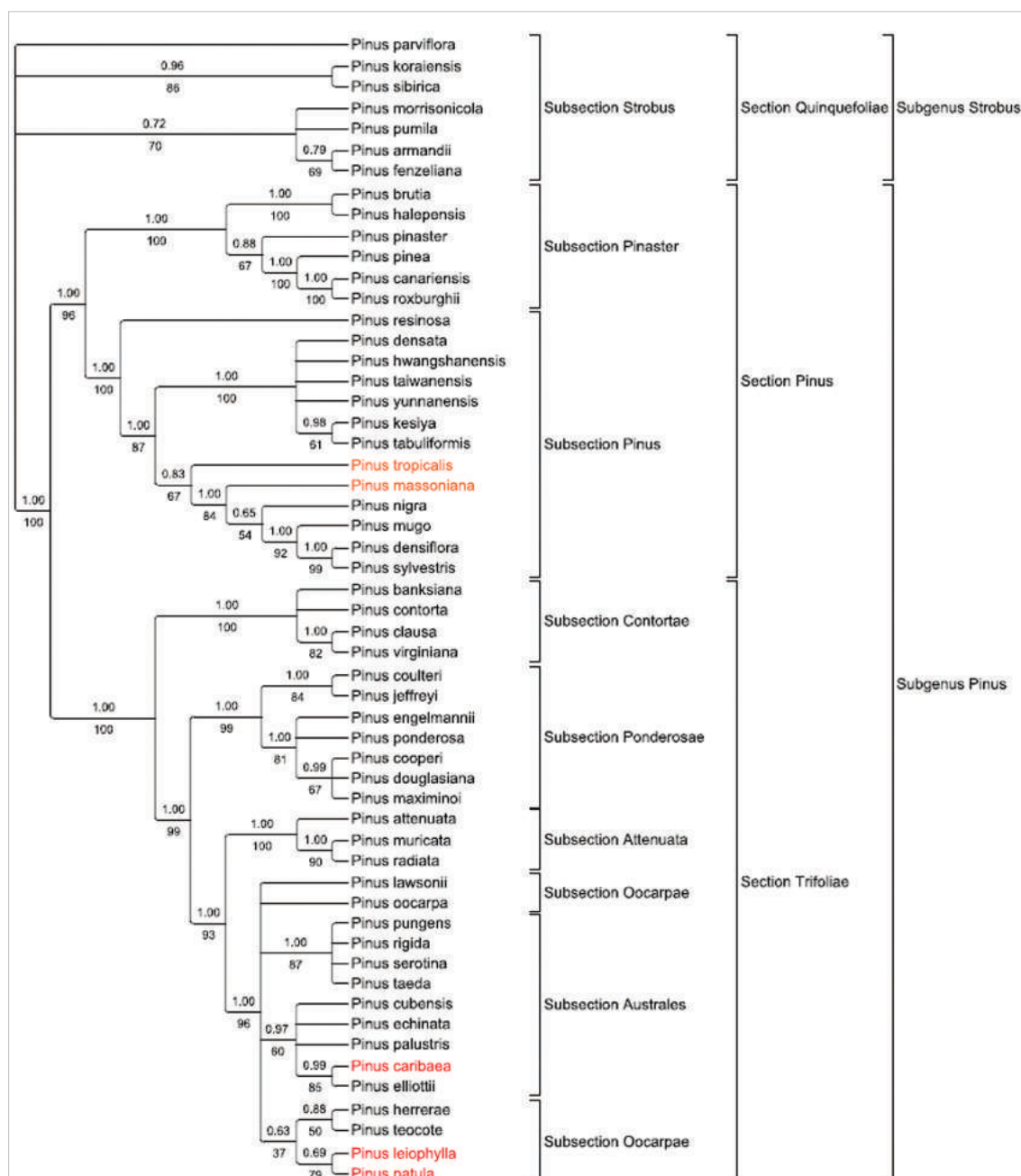


Figura 5. Árbol de *ycf1* entero proveniente del análisis del alineamiento 2. Los valores de apoyo del método bayesiano se muestran encima de la rama como fracción de 1, mientras que los correspondientes al método de máxima verosimilitud se muestran bajo la rama como un porcentaje. Los subgéneros, secciones y subsecciones se han marcado según Gernandt *et al.* (2005). Las especies para las que hay un conflicto se han marcado en rojo, mientras que las diferencias intra-subsección sin apoyo se han marcado en naranja.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Con algunas modificaciones gráficas en Olsson, S.; Gri-vet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

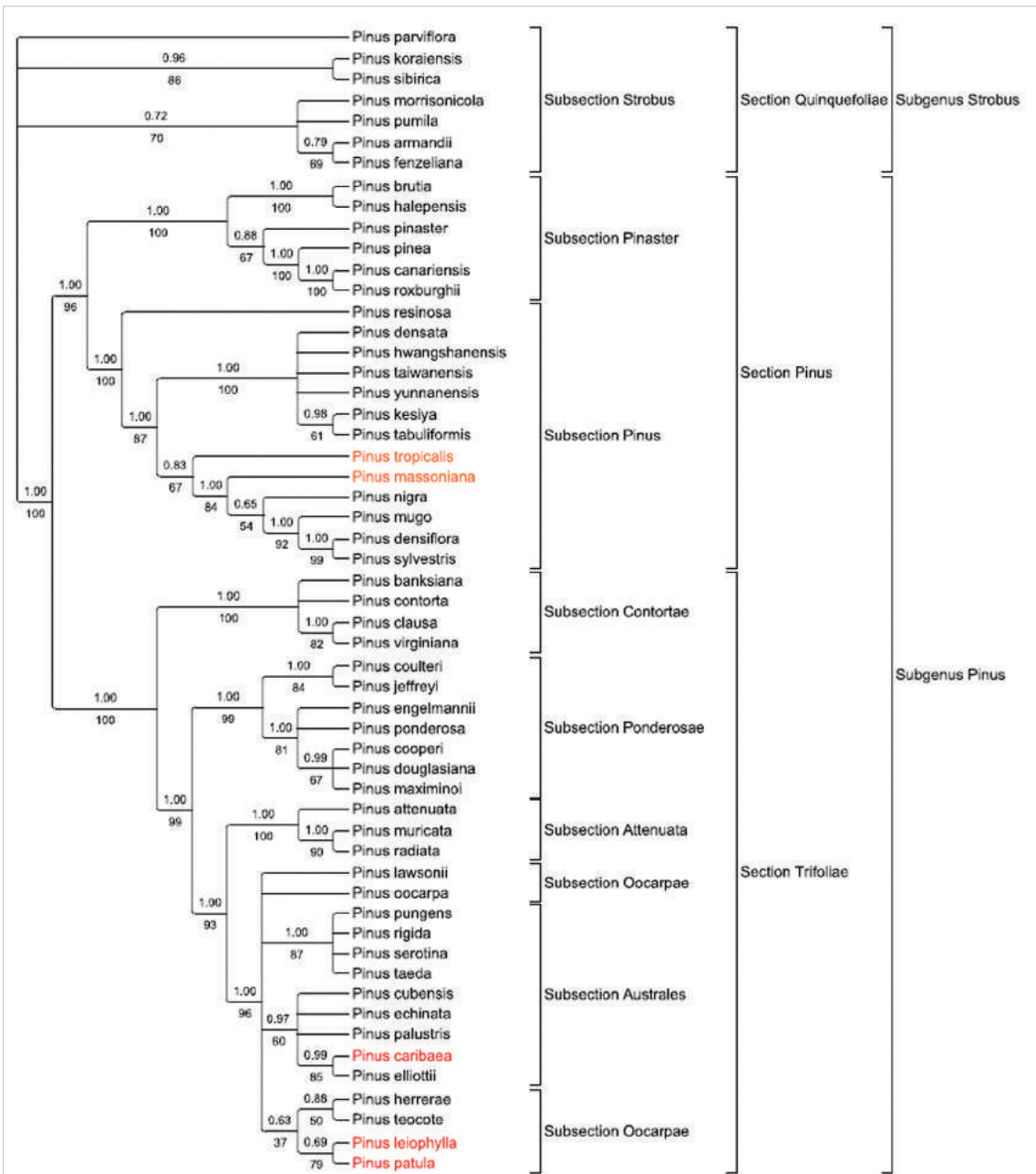


Figura 6. Árbol de *ycfI* sin la región de alta variabilidad proveniente del análisis del alineamiento 4. Los valores de apoyo del método bayesiano se muestran encima de la rama como fracción de 1, mientras que los correspondientes al método de máxima verosimilitud se muestran bajo la rama como un porcentaje. Los subgéneros, secciones y subsecciones se han marcado según Gernandt et al. (2005). Las especies para las que hay un conflicto se han marcado en rojo, mientras que las diferencias intra-subsección sin apoyo se han marcado en naranja.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycfI* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.

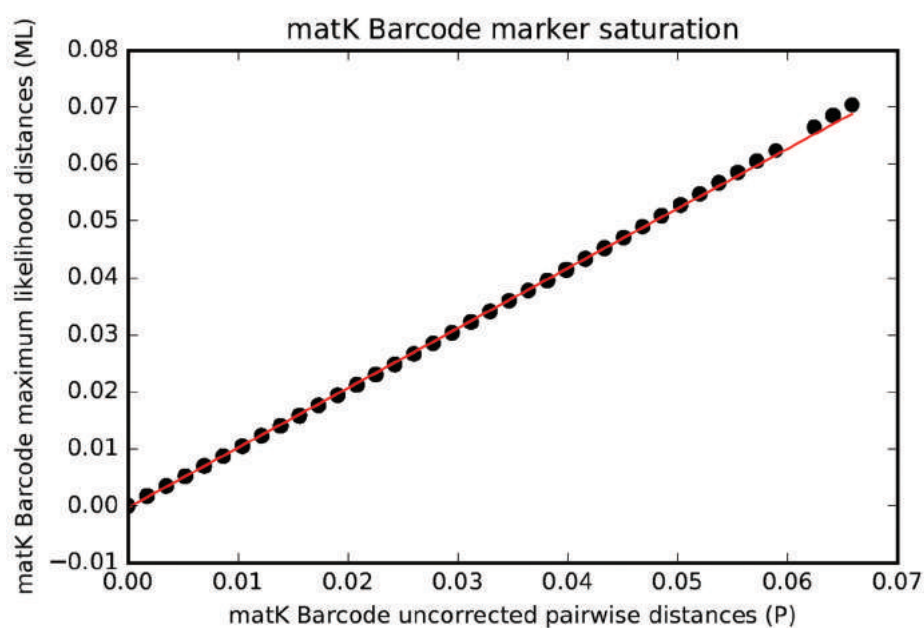


Figura 7. Resultado del análisis de saturación de la región barcode de *matK*.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Olsson, S.; Grivet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

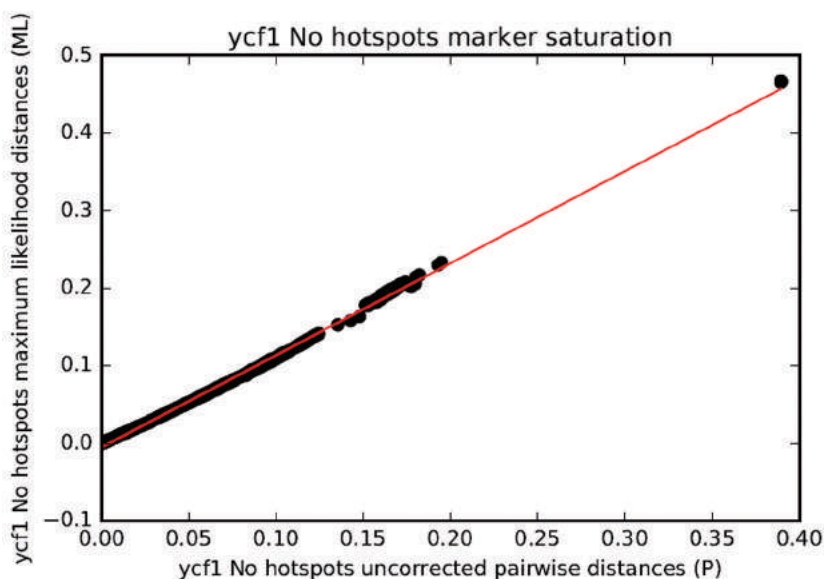


Figura 8. Resultado del análisis de saturación de *matK* entero.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Olsson, S.; Grivet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

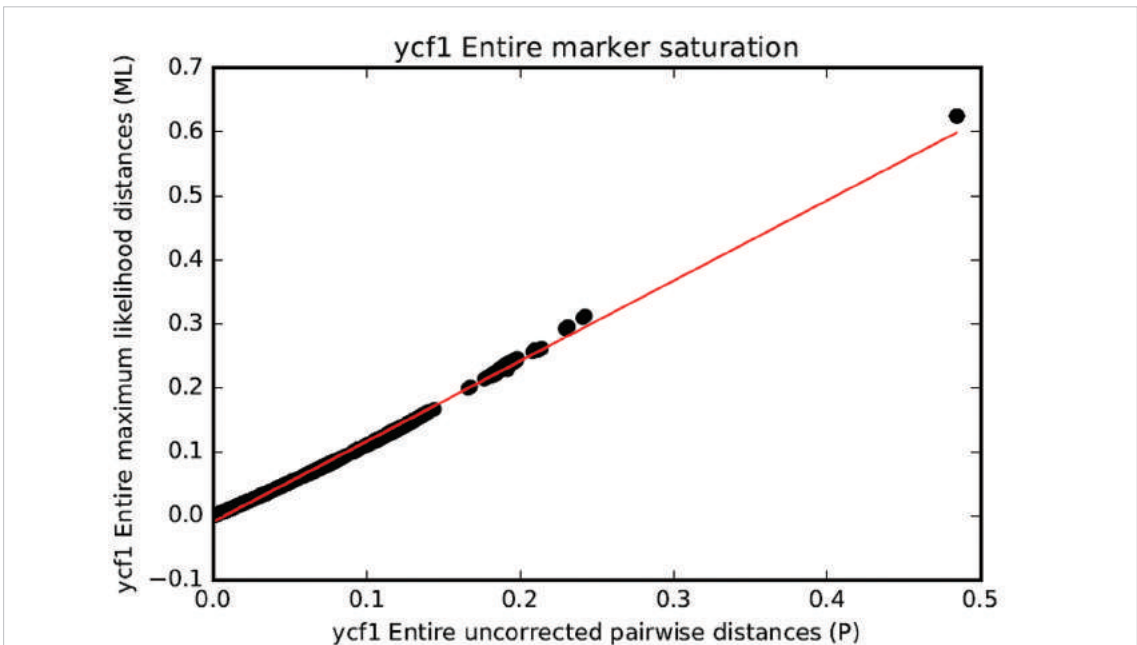


Figura 9. Resultado del análisis de saturación de *ycf1* entero.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Olsson, S.; Grivet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

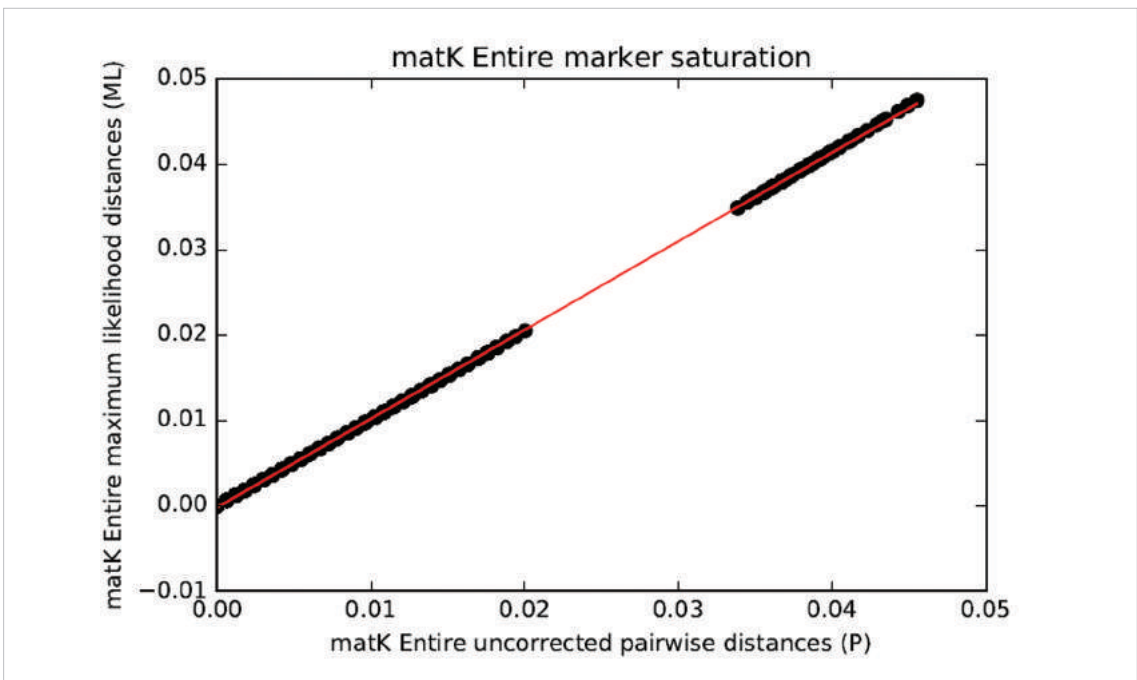


Figura 10. Resultado del análisis de saturación de *ycf1* sin la región de alta variabilidad.

Figura publicada en: Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid. Olsson, S.; Grivet, D.; and Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. doi: 10.5424/fs/2018273-13688. © INIA.

El nivel de significación estadística se estableció en un 5%. Hubo dos hallazgos principales:

1. El marcador *ycf1* es significativamente más variable que el marcador *matK*, con un 21.97% de sitios variables respecto a un 8.23%. En contraste la región variable de *ycf1* tuvo un 35.58% de sitios variables respecto al 21.97% del marcador *ycf1* entero. Esto confirma los resultados obtenidos por los dos estudios que lo utilizaron, Hernández-León *et al.* (2013) y Dong *et al.* (2015).
2. La región código de barras de *matK* no es significativamente más variable que la totalidad del gen (9.39% respecto a 8.23%).

4.2. Árboles filogenéticos

Los árboles filogenéticos producidos por cada alineamiento se muestran en las Figuras 1 a 6. Proporcionan información cuando los clados comparados tienen:

- a) Apoyo estadístico suficiente (>0.95 en el valor superior en cada nodo, correspondiente al método bayesiano y >85 en el valor situado debajo del nodo, correspondiente al método de máxima verosimilitud).
- b) Suficiente resolución, entendida como nodos bifurcantes, en lugar de trifurcantes o polifurcantes.

Hubo cuatro hallazgos principales y dos conflictos encontrados entre ambos marcadores.

1. No hubo diferencias significativas entre los árboles confeccionados a partir de la combinación de los alineamientos de ambos marcadores; uno contenía los dos marcadores completos y al otro se le había extraído la región más variable del marcador *ycf1*.
2. Los dos árboles generados con los alineamientos del marcador *matK* tenían mala resolución en ambos casos, pero significativamente peor en el que se basaba solo en la región código de barras.
3. Los dos árboles confeccionados con los alineamientos del marcador *ycf1* (el que se utilizó completo y al que se le retiró la región variable) no mostraron diferencias significativas.
4. Sin embargo ambos alineamientos del marcador *ycf1* generaron árboles de calidad significativamente mayor que los dos árboles de *matK*.

Solo se consideraron conflictos entre subsecciones, no en la topología interna de las subsecciones. Estos conflictos implican que los marcadores sugieren historias evolutivas diferentes.

1. Conflicto *Pinus attenuata* + *Pinus oocarpa*: *P. attenuata* aparece en la subsección *Attenuatae* y *P. oocarpa* aparece en la subsección *Oocarpae* en los dos árboles combinados, en ambos casos con alto apoyo estadístico. Sin embargo, en los dos árboles de *matK* estas dos especies forman un clado con apoyo de alto a moderado.
2. Conflicto *Pinus caribaea*: De menor entidad. *P. caribaea* aparece emparejado con *P. leiophylla* y *P. patula* en el árbol de *matK* entero con apoyo de alto a moderado. Sin embargo, en los dos árboles combinados estas dos especies forman un clado con apoyo de alto a moderado mientras que *P. caribaea* está en la subsección *Australes* con apoyo de alto a moderado.

4.3. Análisis de saturación

A la vista de los análisis de saturación no se puede afirmar que exista saturación en ninguno de los alineamientos (*Figuras 7 a 10*). No obstante, el marcador *ycfI* sin retirar la región más variable (*Fig. 9*), presenta una pequeña desviación de la recta de pendiente 1:1, lo que pudiera ser un indicio de problemas de saturación por encima del nivel de género.

5. Conclusiones

A la vista de los resultados se pueden extraer cuatro conclusiones, elaboradas y ampliadas en Cid, 2018 y en Olsson *et al.*, 2018:

1. Se confirma la hipótesis de que el desempeño de *matK* mejora significativamente al aumentar su longitud utilizada.
2. El marcador *ycfI* puede utilizarse en filogenética y estudios de código de barras sin eliminar la zona de alta diversidad. Sin embargo se recomienda cautela al emplearlo en niveles taxonómicos superiores al género pues podría tener problemas de saturación.
3. El desempeño de ambos marcadores mejora cuando se emplean juntos. La topología de los árboles combinados es principalmente la del marcador *ycfI*, con mayor resolución y apoyo debido a la adición de *matK*. No obstante, hay señales filogenéticas conflictivas que conviene conocer para evitar artefactos (falsas relaciones).

4. Los marcadores más prometedores en *Pinus* son insuficientes por sí solos para resolver la filogenia de las especies de aparición más tardía. Tendrán que buscarse nuevos marcadores, especialmente nucleares, como ya ha empezado a hacerse recientemente (Gernandt *et al.*, 2018).

Agradecimientos

Quiero agradecer a mi codirectora, Sanna Olsson todo el tiempo y esfuerzo empleado en formarme para llevar a cabo este proyecto, así como a mi directora, Delphine Grivet, por su dirección del estudio, orientación y correcciones. También agradezco a mi tutora sus valiosos consejos de toda índole y revisiones.

También me gustaría dar las gracias a la Escuela de Ingeniería de Montes, Forestal y del Medio Natural por su formación, apoyo y reconocimiento. Por último, agradecer a la Sociedad Española de Ciencias Forestales el reconocimiento con este accésit al premio al mejor trabajo de fin de grado de temática forestal de España.

5. Bibliografía

- Benson, D.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.; Ostell, J.; Sayers, E.; 2014. Gen Bank. *Nucleic Acids Research*, 43(D1), D30-D35. <https://doi.org/10.1093/nar/gku1216>
- Brown, C.; 2000. Global Forest Products Outlook Study. The global outlook for future wood supply from forest plantations. FAO: Working Paper No: GFPOS/WP/03. Chapter 2 (Forest Plantation Trends and Current Status). <http://www.fao.org/docrep/003/X8423E/X8423E00.HTM>- Accessed. 9 July 2019
- CBOL plant Working Group1, Hollingsworth, P.M.; Forrest, L.; Spouge, J.; Hajibabaei, M.; *et al.*, 2009. A DNA barcode for land plants. *Proc Natl Acad Sci*, 106(31), 12794-12797. <https://doi.org/10.1073/pnas.0905845106>
- Cid, J.; 2018. Evaluation of the chloroplast regions *matK* and *ycf1* as diagnostic markers for the genus *Pinus*. Trabajo de Fin de Grado. Universidad Politécnica de Madrid.
- Darriba, D.; Taboada, G.L., Doallo, R. and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*, 9(8), 772-772. <https://doi.org/10.1038/nmeth.2109>
- Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; *et al.*, 2015. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep.*, 5(1). Article Number: 8348. <https://doi.org/10.1038/srep08348>
- Earle, C.; 2017. Gymnosperms - The Gymnosperm Database. Conifers.org. <http://www.conifers.org/zz/gymnosperms.php>. Accessed 9 July 2019.
- Gernandt, D.; Aguirre Dugua, X.; Vázquez-Lobo, A.; Willyard, A.; *et al.*, 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *Am J Bot*, 105(4), 711-725. <https://doi.org/10.1002/ajb2.1052>
- Gernandt, D.; Hernández-León, S.; Salgado-Hernández, E.; Pérez de La Rosa, J.; 2009.

- Phylogenetic Relationships of *Pinus* Subsection *Ponderosae* Inferred from Rapidly Evolving cpDNA Regions. *Syst Bot*, 34(3), 481-491. <https://doi.org/10.1600/036364409789271290>
- Gernandt, D.; López, G.; García, S.; Liston, A.; 2005. Phylogeny and classification of *Pinus*. *TAXON*, 54(1), 29-42. <https://doi.org/10.2307/25065300>
- Grotkopp, E.; Rejmánek, M.; Sanderson, M.; Rost, T.; 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution*, 58(8), 1705-1729. <https://doi.org/10.1111/j.0014-3820.2004.tb00456.x>
- Hernández-León, S.; Gernandt, D.; Pérez de la Rosa, J.; Jardón-Barbolla, L.; 2013. Phylogenetic Relationships and Species Delimitation in *Pinus* Section *Trifoliae* Inferred from Plastid DNA. *PLOS ONE*, 8(7), Article Number: e70501. <https://doi.org/10.1371/journal.pone.0070501>
- Huelsenbeck, J.; Ronquist, F.; 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Katoh, K.; Standley, D.; 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>
- Miller, M.A.; 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>
- Micieta, K.; Murín, G.; 1998. Three Species of Genus *Pinus* Suitable as Bioindicators of Polluted Environment. *Water Air Soil Pollut*, 104(3/4), 413-422. <https://doi.org/10.1023/A:1004984121831>
- Morse, A.; Peterson, D.; Islam-Faridi, M.; Smith, K.; *et al.*, 2009. Evolution of Genome Size and Complexity in *Pinus*. *PLOS ONE*, 4(2), e4332. <https://doi.org/10.1371/journal.pone.0004332>
- Müller, J.; Müller, K.; Neinhuis, C.; Quandt, D.; 2010. PhyDE -Phylogenetic Data Editor (Version 0.9971). PhyDE® <http://www.phyde.de/index.html>. Accessed 9 July, 2019.
- Olsson, S.; Grivet, D.; Cid-Vian, J.; 2018. Species-diagnostic markers in the genus *Pinus*: evaluation of the chloroplast regions *matK* and *ycf1*. *Forest Syst.*, 27(3), p.e016. <https://doi.org/10.5424/fs/2018273-13688>
- Price, R.A.; Liston, A.; Strauss, S.H.; 1998. Phylogeny and systematics of *Pinus*. pp. 49-68 In: Richardson, D.M. (ed.), *Ecology and Biogeography of Pinus*. Cambridge University Press. ISBN-10: 0-521-55176-5.
- Python Software Foundation. Python Language Reference, version 3.5. <http://www.python.org/>. Accessed 9 July, 2019.
- Rambaut, A.; Suchard, M.A.; Xie, D.; Drummond, A.J.; 2014. Tracer v1.6. <http://beast.bio.ac.uk/Tracer>. Accessed 9 July, 2019.
- Ratnasinghan, S.; Herbert, P.N.; 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes*, 7, 355-364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Sealy-Fisher, V.; Pizzi, A.; 1992. Increased pine tannins extraction and wood adhesives development by phlobaphenes minimization. *Holz als Roh- und Werkstoff*, 50(5), 212-220. <https://doi.org/10.1007/BF02663290>
- Stamatakis, A.; 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis

- of large phylogenies. *Bioinformatics*, 30(9), 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stöver, B.; Müller, K.; 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, 11(1), 7. <https://doi.org/10.1186/1471-2105-11-7>
- Swofford, D.; 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0. Sinauer Associates, Sunderland, Massachusetts. <http://www.agromontpellier.fr/sppe/Recherche/JFM/PaupUp/main.htm>. Accessed 9 July, 2019.
- Wang, X.; Ran, J.; 2014. Evolution and biogeography of gymnosperms. *Mol Phylogenet Evol*, 75, 24-40. <https://doi.org/10.1016/j.ympev.2014.02.005>
- Wang, X.; Tsumura, Y.; Yoshimaru, H.; Nagasaka, K.; Szmidt, A.; 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. <https://doi.org/10.2307/2656672>

