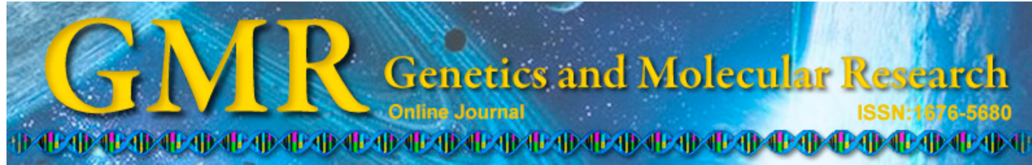

Consistency between molecular phylogeny and morphological classification of the *Salix matsudana* Koidz. complex (Salicaceae)

S.H. Du[1,2]*, Z.S. Wang[1,3]*, Y.X. Li[1], D.S. Wang[1] and J.G. Zhang[1,3]

[1]State Key Laboratory of Tree Genetics and Breeding,
Key Laboratory of Silviculture of the State Forestry Administration,
Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China
[2]College of Forestry, Shandong Agriculture University, Tai'an, Shandong, China
[3]Collaborative Innovation Center of Sustainable Forestry in Southern China,
Nanjing Forestry University, Nanjing, Jiangsu, China

*Theses authors contributed equally to this study.
Corresponding author: J.G. Zhang
E-mail: zhangjg@caf.ac.cn

Genet. Mol. Res. 14 (3): 8663-8671 (2015)
Received October 15, 2014
Accepted March 12, 2015
Published July 31, 2015
DOI http://dx.doi.org/10.4238/2015.July.31.15

ABSTRACT. The morphological species concept is based on morphological traits, which are often subject to subjectivity or artifact. Molecular evidence is needed to test the reliability of morphological classification of taxa that are controversial and to provide appropriate taxonomic delimitation. In this study, we used 15 single-copy nuclear loci and 2 chloroplast fragments to verify the morphological classification of the *Salix matsudana* Koidz. complex using phylogenetic approaches. Complete sequence alignment showed slight diversification in nuclear sequences and no variety in chloroplast DNA fragments. Phylogenetic trees revealed a monophyletic group consisting of all individuals of *S. matsudana* and 2 clades within this group, with a 100% bootstrap support value and 1.00 posterior probability. The topology of the phylogenetic trees was highly

 ©FUNPEC-RP www.funpecrp.com.br

consistent with the morphological classification of the *S. matsudana* complex. Verifying the genetic background of these classification units based on remarkable morphological differences will provide a foundation for future studies of *Salix* and the breeding of new horticultural varieties.

Key words: Morphology; Phylogeny; *Salix matsudana*; Species concept

INTRODUCTION

Species is the essential unit of concern in biodiversity, conservation, and other biological properties (Bradley and Baker, 2001). Various concepts of species have been developed (Fraser and Bernatchez, 2001). The morphological species concept is defined mainly based on the description of outlier taxonomic characters of different classification units. The phylogenetic species concept recognizes species as an irreducible cluster of individuals that is distinct from other clusters and within which there is an obvious pattern of ancestry and descent (Cracraft, 1989), or a species is a monophyletic group of common ancestry (de Queiroz and Gauthier, 1990). The development of molecular systematics over the past several decades has greatly advanced the phylogenetic species concept (Amato and Montresor, 2008). An inevitable aspect of morphological classification is the subjectivity or artifact, and thus molecular evidence is needed to test the reliability of morphological classification of taxa that are controversial (Su et al., 2012).

Salix matsudana Koidz. has 1 *varietas* and 3 *forma*, including *S. matsudana* var. *pseudo-matsudana*, *S. matsudana* f. *pendula*, *S. matsudana* f. *tortuosa*, and *S. matsudana* f. *umbraculifera* (Wu, 1999). Compared with *S. matsudana* var. *matsudana*, the annual branches of *S. matsudana* var. *pseudo-matsudana*, *S. matsudana* f. *tortuosa*, and *S. matsudana* f. *pendula* are long and drooped. The pistillate flower of *S. matsudana* var. *pseudo-matsudana* contains 1 gland, while the other classification units contain 2. The branch of *S. matsudana* f. *tortuosa* is curly, which is distinctive in *Salix*. The tree form of *S. matsudana* f. *umbraculifera* is significantly different from other classification units, the crown of which is hemispherical and appears as a steamed bread. Because of the remarkably different morphological characters, it remains unclear whether these classification units should be treated as different species or as subspecies. Verifying the placement of these classification units and clarifying their evolutionary relationships will not only provide molecular evidence for the classification of controversial taxa but also set a foundation for future studies of functional genes related to horticultural traits. In this study, a phylogenetic approach based on the concordance of multiple single-copy orthologous nuclear loci and chloroplast sequences was utilized to testify the reliability of morphological classification and to evaluate the evolutionary relationships of different classification units in the *S. matsudana* complex.

MATERIAL AND METHODS

The biological material was sampled from the Beijing Botanical Garden and Summer Palace. Three to six individuals of each classification unit were sampled. Fresh leaves were

collected and dried immediately with silica gel. *Populus euphratica*, *Salix arbutifolia*, *Salix triandra*, and *Salix raddeana* were used as outgroups in phylogenetic analysis. Details of the materials sampled are listed in Table 1.

Table 1. Details of the materials sampled.

Samples	No. of individuals	Latitude	Longitude	Altitude (masl)
S. matsudana var. *matsudana*	5	39°59'N	116°12'E	73
S. matsudana var. *pseudo-matsudana*	4	45°48'N	126°31'E	52
S. matsudana f. *tortuosa*	5	39°59'N	116°16'E	67
S. matsudana f. *pendula*	5			
S. matsudana f. *umbraculifera*	5			
S. *arbutifolia*	1	40°43'N	124°47'E	433
S. triandra	1	47°22'N	87°47'E	20
S. raddeana	1	40°41'N	117°14'E	410
P. euphratica	1	47°41'N	86°49'E	670

masl, meters above sea level.

DNA extraction, polymerase chain reaction amplification, and sequencing

The primers used for amplifying and sequencing single-copy nuclear loci in this study were developed in our previous study (Du et al., 2014). Primers for chloroplast DNA (cpDNA) *rbcl* and *trnK* are listed in Table 2. All DNA sequences of this study have been submitted to GenBank (#KJ101449-KJ101504 and #KJ155027-KJ155446).

Table 2. Primers used for *rbcl* and *trnK* in this study (primers underlined are internal sequencing primers).

Locus	Primer	Sequence (5'-3')	Ta (°C)	References
trnK	trnKF	GGGTTGCCCGGGACTCGAAC	58	modified from Demesure et al. (1995)
	trnkR	ATTGGATTTGCTGTGATA		
	trnK-3F	ACTAATGGGATGTCCTACTG		Wang et al. (2014)
	trnK-3R	GATTTCTAGTCACCTATTAC		
rbcl	rbclF	ATGTCACCACAAACAGAAACT	56	Bobowski et al. (1999)
	rbclR	CTTCACAAGCAGCAGCTAGTTCAGGACTCC		

Total genomic DNA was extracted from 30 mg leaf tissue from each individual using a modified protocol that has been described previously (Doyle, 1987). Polymerase chain reaction was performed in a total volume of 30 μL containing 10-60 ng genomic DNA, 0.8 μM of each dNTP, 2.4 μM of each primer, 0.15 U exTaq DNA polymerase (TaKaRa, Shiga, Japan), and 2.0 mM $MgCl_2$. Amplification was carried out in a temperature gradient 96 U thermocycler (Applied Biosystems, Foster City, CA, USA) as follows: 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 60 s at 50-60°C (depending on the annealing temperature of specific primers and length of the amplified regions), 90 s at 72°C, and a final extension at 72°C for 10 min. Products were examined by agarose gel electrophoresis and purified using a DNA Purification kit (Amersham Pharmacia Biotech, Piscataway, NJ, USA). The purified DNA was directly sequenced using an ABI 3730XL DNA analyzer (Applied Biosystems). The same primers were utilized for amplification and sequencing.

 ©FUNPEC-RP www.funpecrp.com.br

Data analysis

The assembled contigs of each individual were aligned using CLUSTAL X (Thompson et al., 1997) and refined manually in BioEdit (Hall, 1999). Homogeneity across nuclear DNA was tested using the incongruence length difference test (Farris et al., 1994) as implemented in PAUP* 4.0b10* (Swofford, 2003). A heuristic search was performed with 10 random addition replicates and tree-bisection-reconnection branch swapping. Phylogenetic trees were reconstructed using maximum parsimony (MP) and Bayesian methods. MP analysis was conducted in PAUP* 4.0b10* (Swofford, 2003), with all characters equally weighted and treated as unordered and gaps treated as missing data. A heuristic search was performed with tree-bisection-reconnection branch swapping, MULPARS option, and RANDOM stepwise addition with 1000 replicates. Topological robustness was assessed by bootstrap analysis with 1000 replicates using simple taxon addition (Felsenstein, 1985). An appropriate nucleotide substitution model for each locus was determined using Modeltest 3.7 (Posada, 2003). Models were chosen mainly based on the Akaike information criterion and used for subsequent Bayesian analysis. Bayesian analysis was conducted using Mrbayes 3.2.1 (Ronquist and Huelsenbeck, 2003). Two independent runs of Metropolis-coupled Markov chain Monte Carlo were conducted simultaneously, with each run including 1 cold chain and 3 incrementally heated chains, starting randomly in the parameter space. A total of 2,000,000 generations were run, and trees were sampled once every 100 generations. The Tracer v1.5 program (Drummond and Rambaut, 2007) was utilized to evaluate stationary characteristics. The first 25% of sampled trees were discarded as burn-ins, and the posterior probabilities were calculated from the remaining trees. The phylogenetic trees were viewed in FigTree v 1.3.1 (Rambaut, 2008).

RESULTS

Sequence characteristics

We successfully obtained all 15 nuclear DNA sequences and 2 cpDNA fragments for all individuals. The characteristics of the 17 loci and their most appropriate nucleotide substitution models are summarized in Table 3. After removing regions showing ambiguous alignment, the aligned length of the nuclear loci ranged from 329 to 1054 bp with a total length of 8599 bp. The number of variable and informative sites ranged from 1-34 and 1-32, with total numbers of 218 and 203, respectively. The sequence length of *rbcl* ranged from 978 to 1329 bp, with an aligned length of 924 bp. The aligned length of *trnK* was 2440 bp, and the sequencing length varied from 2445 to 2562 bp, in which there was an indel of 21 bp. Unexpectedly, no variable sites were detected in the *rbcl* and *trnK* sequences of all 24 individuals of *S. matsudana*. To confirm this, we repeated the amplification and sequencing for all individuals and observed the same result.

Phylogenetic analysis of the nuclear loci

Phylogenetic trees generated based on individual nuclear DNA sequences were

Table 3. Characteristics of nuclear and cpDNA loci and the combined data (excluding outgroups).

Locus (without outgroups)	Aligned length	No. of variable sites (%)	No. of informative sites (%)	Nucleotide substitution model	Gamma shape parameter
Nuclear DNA					
DSL 1	426	25 (5.87)	24 (5.63)	HKY+I+G	0.2473
DSL 2	329	3 (0.91)	2 (0.61)	F81	equal
DSL 3	540	24 (4.44)	24 (4.44)	JC	equal
DSL 4	427	4 (0.94)	4 (0.94)	JC	equal
DSL 5	403	6 (1.49)	6 (1.49)	F81	equal
DSL 6	473	12 (2.54)	12 (2.54)	F81	equal
DSL 7	390	10 (2.56)	9 (2.31)	HKY	equal
DSL 8	476	1 (0.21)	1 (0.21)	F81	equal
DSL 24	1054	25 (2.37)	22 (2.09)	F81+I+G	0.6230
DSL 25	635	10 (1.57)	9 (1.42)	F81	equal
DSL 26	504	2 (0.40)	1 (0.20)	JC	equal
DSL 28	825	28 (3.39)	27 (3.27)	F81+I+G	1.1356
DSL 29	834	34 (4.07)	32 (3.84)	TrN+I+G	0.8756
DSL 33	531	23 (4.33)	21 (3.95)	F81+I+G	0.5468
DSL 35	752	11 (1.46)	9 (1.20)	F81	equal
Concatenated data	8599	218 (2.54)	203 (2.36)	HKY+I+G	0.6165
cpDNA					
rbcl	924	0	0		
trnK	2440	0	0		

very similar to each other (results not shown). The incongruence length difference test showed significant incongruence among the nuclear DNA sequences (P = 0.03). However, previous studies have examined the limitations of this statistical test (Hipp et al., 2004; Quicke et al., 2007). Therefore, we combined the 15 individual nuclear DNA loci into a single data set to sufficiently reconstruct the phylogeny of classification units in *S. matsudana*.

The combined data set generated the 121 most parsimonious trees of 837 steps with a consistency index of 0.81 and a retention index of 0.71 using the MP method, which recovered a fully resolved monophyletic group consisting of all 24 individuals with a 100% bootstrap support value (Figure 1). This group was also yielded through Bayesian analysis with a 1.00 posterior probability (Figure 2). Within the monophyletic group, all individuals were subdivided into 2 clades (I and II) with 100% bootstrap support and 1.00 posterior probability. Individuals of *S. matsudana* var. *pseudo-matsudana* clustered in clade I, while the sister clade, clade II, comprised the remaining individuals. Within this clade, individuals of different classification units were scattered and most braches were polytomic. In the Bayesian phylogenetic tree, several branches with relatively high posterior probability were recovered. However, not all individuals of the same classification unit were clustered together in 1 branch or individuals from different classification units linked to each other in 1 clade. For example, 3 individuals of *S. matsudana* f. *pendula* grouped into 1 clade, but the remaining samples clustered with *S. matsudana* f. *umbraculifera* and *S. matsudana* var. *matsudana* in another clade. Furthermore, these branches were not supported or showed relatively low bootstrap support values in the MP tree.

DISCUSSION

With a large data set of combined 15 single-copy nuclear loci and 2 cpDNA fragments, the phylogeny of the *S. matsudana* complex was reconstructed (Figures 1 and 2). All

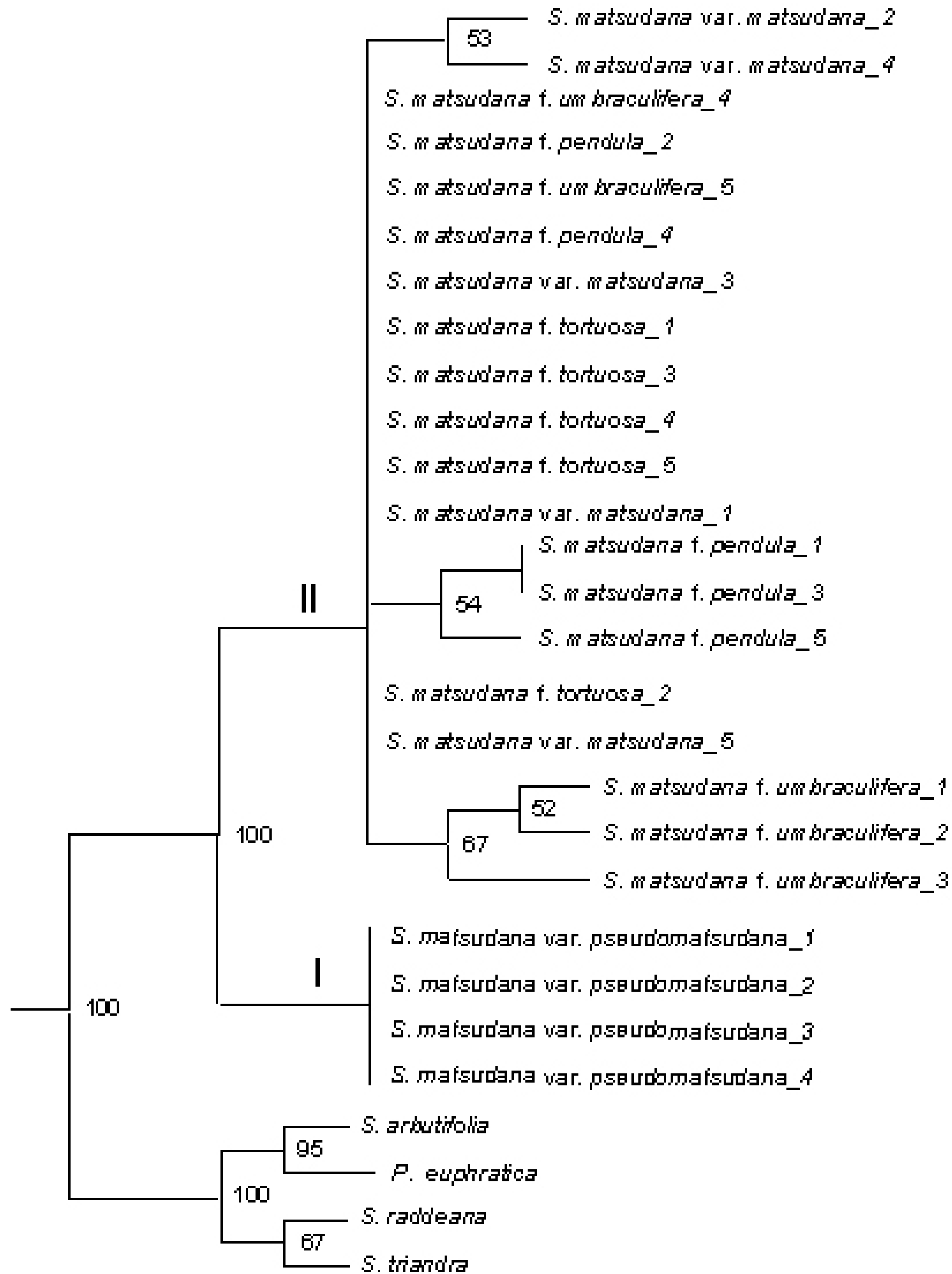

Figure 1. Phylogenetic tree reconstructed based on the MP method. Numbers after scientific name indicate the individual order. Bootstrap support values are listed next to nodes.

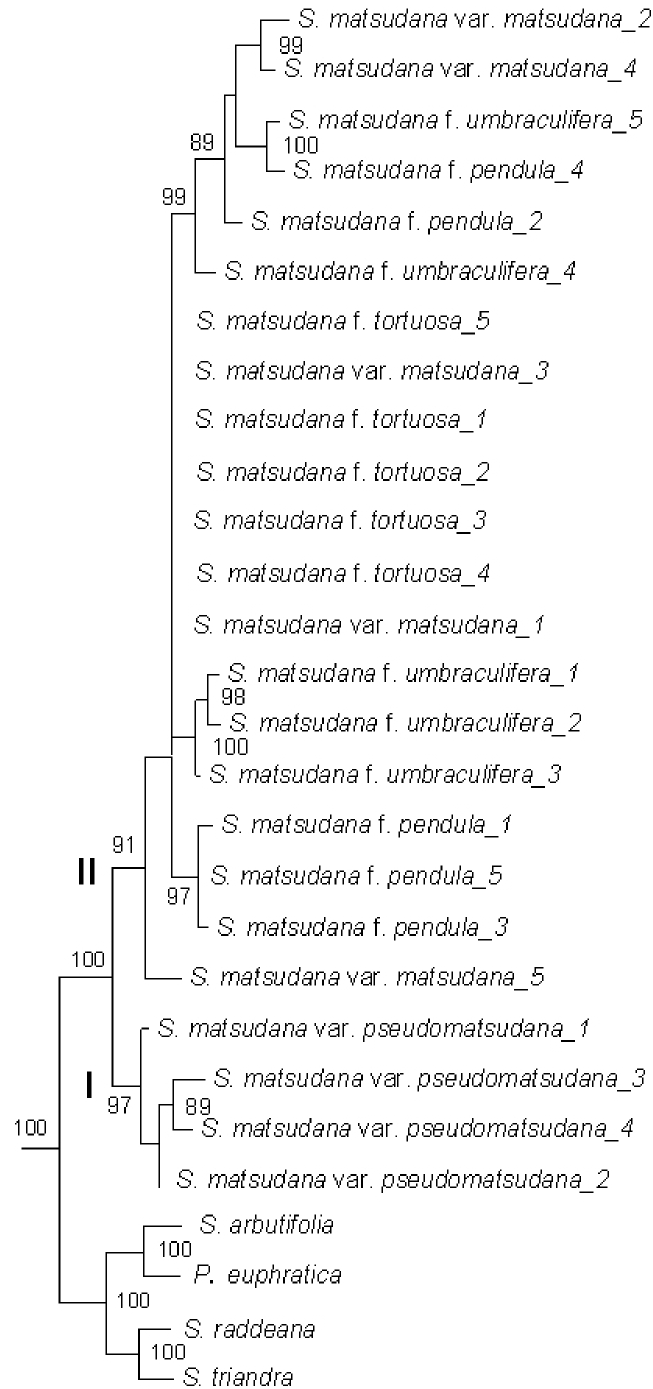

Figure 2. Phylogenetic tree reconstructed based on the Bayesian method. Numbers after scientific names indicate the individual order. Posterior probabilities are listed next to the nodes.

individuals of *S. matsudana* grouped together as a monophyletic group separate from the out-group species with 100% bootstrap support value or 1.00 posterior probability, revealing the close relationships between classification units.

In the 2 phylogenetic trees, individuals of *S. matsudana* var. *pseudo-matsudana* formed a single clade sister to other remaining individuals with 100% bootstrap and 1.00 posterior probability (Figures 1 and 2). Furthermore, *S. matsudana* var. *pseudo-matsudana* showed significant morphological differences from other classification units as the only *varietas* in the *S. matsudana* complex. The female flower has only 1 gland, while the others have 2 (Wu, 1999). Based on the nuclear phylogenetic trees and no variable sites in cpDNA fragments, we support the traditional classification of *S. matsudana* var. *pseudo-matsudana* as a *varietas* in *S. matsudana*.

The remaining classification units showed significant morphological differences; however, slight differences were detected at the molecular level (clade II in Figures 1 and 2). Furthermore, no variable sites existed in cpDNA fragments. Although some nodes were supported by high posterior probability (>95%) in the Bayesian phylogenetic tree, it is well-known that posterior probability calculated from actual data for some clades often appears to be too high (Yang and Rannala, 2012). Furthermore, internal nodes within clade II were supported by a low bootstrap support value (~50%) in the MP tree. Overall, the classification units were not sufficiently diversified at the molecular level and thus were classified as different *forma* with close evolutionary relationships.

The number of variable sites in the 15 single-copy nuclear loci of the 24 individuals was relatively high; however, extremely weak resolution in the phylogenetic trees showed that most of the variation was commonly shared by these classification units. Furthermore, there were no variable sites in the cpDNA *rbcl* and *trnK* sequences, which have been widely utilized in phylogenetic studies at the species or genus level (Kusumi et al., 2000; Soltis et al., 2001; Fan and Xiang, 2003; Guo and Ge, 2005; Tang et al., 2010). Shared variation in nuclear loci and no variation in cpDNA sequences of the classification units indicated that these classification units were not sufficiently diverse to be treated as different taxonomic species and that they belonged to 1 species, *S. matsudana*, both according to the morphological and phylogenetic species concept. The conspicuous morphological differences among the *varietas* and *forma* of *S. matsudana* may result from the differential expression of a small number of genes or a unique gene. Determining the genetic background of the morphological differences will set the stage for a broad range of future studies in *S. matsudana* and the breeding of new horticultural varieties.

ACKNOWLEDGMENTS

We thank Dr. Zou Xinhui and Dr. Tang Liang for phylogenetic analysis. We also thank Dr. Zeng Yanfei for data analysis and suggestions on a preliminary version of this paper. Research supported by the Specialized Research Fund for Young Scholars of the Research Institute of Forestry, Chinese Academy of Forestry (#RIF 2012-06).

REFERENCES

Amato A and Montresor M (2008). Morphology, phylogeny, and sexual cycle of *Pseudo-nitzschia mannii* sp. nov. (Bacillariophyceae): a pseudo-cryptic species within the *P. pseudodelicatissima* complex. *Phycologia* 47: 487-497.

Bobowski B, Hole D, Wolf P, Bryant L, et al. (1999). Identification of roots of woody species using polymerase chain

reaction (PCR) and restriction fragment length polymorphism (RFLP) analysis. *Mol. Ecol*. 8: 485-491.

Bradley RD and Baker RJ (2001). A test of the genetic species concept: cytochrome-b sequences and mammals. *J. Mammal*. 82: 960-973.

Cracraft J (1989). Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Speciation and its Consequences (Otte D and Endler JA, eds.). Sinauer, Sunderland, 28-59.

de Queiroz K and Gauthier J (1990). Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. *Syst. Biol*. 39: 307-322.

Demesure B, Sodzi N and Petit R (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol*. 4: 129-134.

Doyle JJ (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*. 19: 11-15.

Drummond AJ and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol*. 7: 214.

Du SH, Wang ZS and Zhang JG (2014). A novel set of 15 single-copy nuclear DNA markers for genetic studies of Salicaceae. *Genet. Mol. Res*. 13: 4911-4917.

Fan C and Xiang QYJ (2003). Phylogenetic analyses of Cornales based on 26S rRNA and combined 26S rDNA-matK-rbcL sequence data. *Am. J. Bot*. 90: 1357-1372.

Farris JS, Källersjö M, Kluge AG and Bult C (1994). Testing significance of incongruence. *Cladistics* 10: 315-319.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

Fraser DJ and Bernatchez L (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol. Ecol*. 10: 2741-2752.

Guo YL and Ge S (2005). Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *Am. J. Bot*. 92: 1548-1558.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser*. 41: 95-98.

Hipp AL, Hall JC and Sytsma KJ (2004). Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. *Syst. Biol*. 53: 81-89.

Kusumi J, Tsumura Y, Yoshimaru H and Tachida H (2000). Phylogenetic relationships in Taxodiaceae and Cupressaceae *sensu stricto* based on matK gene, chlL gene, trnL-trnF IGS region, and trnL intron sequences. *Am. J. Bot*. 87: 1480-1488.

Posada D (2003). Using Modeltest and PAUP* to select a model of nucleotide substitution. *Curr. Protoc. Bioinformatics* 6.5. 1-6.5. 14.

Quicke DL, Jones OR and Epstein DR (2007). Correcting the problem of false incongruence due to noise imbalance in the incongruence length difference (ILD) test. *Syst. Biol*. 56: 496-503.

Rambaut A (2008). FigTree v1. 1.1: Tree figure drawing tool. Available at [http://tree.bio.ed.ac.uk/software/figtree/]. Accessed October 30, 2006.

Ronquist F and Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

Soltis DE, Tago-Nakazawa M, Xiang QY, Kawano S, et al. (2001). Phylogenetic relationships and evolution in *Chrysosplenium* (Saxifragaceae) based on matK sequence data. *Am. J. Bot*. 88: 883-893.

Su JX, Wang W, Zhang LB and Chen ZD (2012). Phylogenetic placement of two enigmatic genera, *Borthwickia* and *Stixis*, based on molecular and pollen data, and the description of a new family of Brassicales, Borthwickiaceae. *Taxon* 61: 601-611.

Swofford DL (2003). PAUP*: phylogenetic analysis using parsimony (*and other methods), Sinauer Associates, Sunderland.

Tang L, Zou X, Achoundong G, Potgieter C, et al. (2010). Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. *Mol. Phylogenet. Evol*. 54: 266-277.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25: 4876-4882.

Wang Z, Du S, Selvadurai D, Wang D, et al. (2014). Phylogeny reconstruction and hybrid analysis of *Populus* (Salicaceae) based on nucleotide sequences of multiple single-copy nuclear genes and plastid fragments. *PLoS One* 9: e103645. doi: 10.1371/journal.pone.0103645.

Wu Z (1999). Flora of China. Science Press, Beijing.

Yang Z and Rannala B (2012). Molecular phylogenetics: principles and practice. *Nat. Rev. Genet*. 13: 303-314.

 ©FUNPEC-RP www.funpecrp.com.br