ONLINE APPENDIX S2. Detailed methods.

Methods

Treating 2ISPs as informative

Intra-individual site polymorphisms (2ISPs) may occur whenever there is more than one copy of a gene region within the genome, which is the case for the majority of nuclear DNA regions in diploid organisms. In the presence of multiple copies, substitution events can occur in different ways. In the simplest case, where nuclear genes are present at two homologous loci (alleles), the shift from a C to a T, for example, can occur in two ways: (1) under asexual reproduction, two mutation events are needed: C-C to C-T to T-T, or (2) the fixation of the T variant and elimination of the C variant in a population via sexual reproduction, genetic drift and/or natural selection. However, in high copy nuclear regions, such as ITS, the matter is complicated by the additional requirement that the mutation must spread onto enough copies to be detected by cloning or direct sequencing. Thus, a substitution of a C by a T in a direct-PCR sequence requires a mutation on an ITS copy, then a subsequent spread of the T accompanied by the loss of the C across copies within the individual (intra- and inter-array homogenisation), followed or accompanied by a spread of the fixed T variant through the population (allelic homogenisation). An optimal transition probability model would need to be highly complex if it should accommodate both mutation (substitution) and homogenisation (including concerted evolution) probabilities, both likely to be variable from lineage to lineage and through time. Creating a general model for this case may be an impossible task. Nevertheless, the example from standard phylogenetic analysis of sequence data demonstrates that despite the regular violation of many model assumptions (see Chapters 9, 13, and 16 in Felsenstein, 2004) molecular trees still produce reasonable results in the majority of cases. Even inferences that assume entirely different transition probabilities can yield similar topologies.

Therefore, we start with the following simplified working hypothesis. Shifts from one monomorphic state to another via a polymorphic state are treated as ordered substitution events (i.e. C↔Y↔T). Thus, a step matrix representing the substitutions required to shift from one state to another (Fig. 1 in main text) can allow this process

to easily be incorporated into distance- and parsimony - based phylogenetic approaches, as well as into distance-based splits networks. For distance algorithms, the step matrix is used to calculate a modified uncorrected *p*-distance incorporating polymorphisms (hereafter termed polymorphism *p*-distance). Uncorrected *p*-distances may be considered problematic as genetic distances increase between samples; however, Göker & Grimm (2008) observed no significant differences (M. Göker, pers. com.) between uncorrected-*p* or model-based distances across a range of datasets, suggesting that this is not a generality at the taxonomic level in which we are interested here. Also, a clear advantage of uncorrected *p*-distances is that they are straightforward and directly comparable to the maximum parsimony results. Furthermore, we use uncorrected *p*-distances for our analyses as it would be difficult to separate the biases and branching artefacts due to the data structure versus those due to assumptions of the underlying model when using model-based distances.

However, the proposed step matrix (Fig. 1 in main text) may be an over-simplification, since the probabilities of fixation or loss of a 2ISP are likely to differ, therefore warranting a maximum likelihood (ML) framework. Thus, of necessity, we utilise an *ad hoc* 2ISP-informative implementation in the ML approach by treating each IUPAC code (i.e. monomorphic and polymorphic bases) as a unique character (*sensu* multi-state analysis of categorical data). This estimates the rate of transitions between characters (i.e. estimating the rate of transitions between states instead of the single steps shown in Fig. 1). We use this approach as it is currently impossible to independently estimate a substitution model that accommodates multicopy processes (e.g. concerted evolution). Furthermore, the strength of ML in comparison to maximum parsimony (MP) and distance-based (i.e. neighbour joining: NJ) tree-inference methods lies in the flexibility of the substitution model, and it would be counter-productive to constrain the model in the same fashion as done for parsimony and distance inferences. For example, an ML model that *a priori* assigns high probability to change of C/T↔Y and low to change of C↔T would simply mirror the MP step-matrix. The multi-state analysis retains the model flexibility while still producing reliable results.

Simulations: inherited variation

We used simulations to explore the effects of treating 2ISPs as ambiguous versus informative (additional) characters on phylogenetic inference when the source of this variation was due to inheritance. To simulate polymorphism induced by two sets of independently evolving sequences (e.g. ITS homoeologues in a broad sense, Cronn et al., 2002), a random tree was generated, two DNA datasets (A_1 and A_2) were simulated onto this tree (across a range of substitution rates), and the tip sequences from these two datasets were merged to form a polymorphic, combined variant, dataset ($A_{1\&2}$). This process is elaborated below. The A_1 and A_2 DNA datasets can be considered independent histories of two co-dominant variants present in the gene pool of a lineage. In the case of ITS (or ETS or 5S-IGS data) they could either represent intra-array (paralogy in a broad sense), allelic (between homologous nucleolus organiser regions [NORs]), homoeologous *s.str.* (between orthologous NORs/5S loci) or paralogous *s.str*. variation (between NORs/5S loci originated by duplication and translocation). The $A_{1\&2}$ datasets represent the substitution and inheritance of polymorphisms along branches and branching events. This is a highly simplified model of multicopy inheritance as it does not include the effects of concerted or reticulate evolution.

Random trees (*sensu* Paradis, 2012, pg. 313), with 20 tips each, were generated using the APE library version 2.7.1 (Paradis et al., 2004) in R version 2.13.0 (R Development Core Team, 2011). The A_1 and A_2 datasets were simulated onto each tree with only monomorphic bases (i.e. A, G, C and T) using the JC69 model of DNA evolution with a rate of 0.010, 0.025, 0.050, 0.100 or 0.200 substitutions per site; this represents a range of signal-poor to signal-rich datasets. The DNA dataset were simulated using the PHANGORN library version 1.3.1 (Schliep, 2011) in R. All datasets were largely free of any homoplasies. The starting sequence used for each simulation was based on the ITS-1 and ITS-2 regions (490 bases in total) of a sequence from *Nymania capensis* (Meliaceae) extracted from GenBank (DQ861633). For each substitution rate, 200 datasets were simulated (two independent datasets per tree; A_1 and A_2); each paired dataset was combined to make 100 additional polymorphic datasets ($A_{1\&2}$).

Phylogenetic trees were inferred using NJ, MP and ML. Bayesian inference was not included as existing software explicitly treats 2ISPs as ambiguities or missing data (e.g. MRBAYES and BEAST); a 2ISP-informative implementation is not currently possible in this software. NJ was implemented in PAUP version 4b10 and the APE library in R for the 2ISP-ambiguous (NJ-A) and 2ISP-informative (NJ-I) treatments, respectively. Genetic distances for NJ-A were calculated as uncorrected *p*-distances in PAUP* with the default option MissDist = Infer (missing data and ambiguity sites are inferred by distributing them proportionally to unambiguous bases), whereas the distances for NJ-I were calculated with the PHANGORN library in R using polymorphism *p*-distances. Bipartition support was evaluated using 1000 bootstrap replicates with random sequence addition.

MP was implemented in PAUP*. Intra-individual site polymorphisms were treated as either ambiguities (MP-A), the default treatment, or as informative (MP-I) using the cost matrix as in Figure 1. Bipartition support was evaluated with 1000 bootstrap replicates, each using random sequence addition and heuristic searching for most-parsimonious trees with the default branch swapping and character optimisation options (TBR and ACCTRAN, respectively). Note that the use of ACCTRAN or DELTRAN does not make any tangible differences to the results (data not shown). For computational efficiency, no more than one of the equally parsimonious trees found per bootstrap replicate was stored (MULTREES=NO); this prohibited time-consuming searches of equally parsimonious trees – a problem predominantly with the MP-A treatment when the dataset contained 2ISPs – and gave each bootstrap replicate equal weighting in the bipartition analyses.

RAxML 7.2.6 was used to compute trees and perform bootstrap analyses under ML (Stamatakis, 2006). Whilst a standard RAxML analysis includes polymorphic bases in analysis, this can still lead to flattening of the likelihood surface, making it more difficult to determine the best-known tree and reducing support values. Thus, the standard analysis was treated as the analogue to the 2ISP-ambiguous treatment under NJ and MP (ML-A). RAxML includes a multi-state analysis for any kind of categorical data (-m MULTIGAMMA -k GTR); this would treat each IUPAC code as a unique character, thereby estimating the rate of transitions between characters (i.e. estimating the rate of transitions between states instead of the single steps shown in

Fig. 1). This was considered the 2ISP-informative treatment of intra-individual site polymorphisms (ML-I). The GTR-Γ model (Rodriguez et al., 1990) was used for all datasets under ML (ML-A and ML-I); this is the least-constrained model available for standard DNA characters and the only substitution model implemented in RAxML. The logic behind this is that model selection procedures (AIC/BIC-based: Posada and Crandall, 1998; Minin et al., 2003; Nylander et al., 2004) identify the GTR model, or the slightly more constrained HKY model, as optimal for the majority of data sets (see original literature for the analysed data - Table 2 in main text). In addition, RAxML optimises the tree topology and model parameters during the run. Thus, if the data complies with an HKY model, then RAxML will optimise the GTR model parameters so that they approach – in the best case – the constraints of a HKY model (Stamatakis, pers. comm., 2006–2012). The performance of RAxML in our simulations demonstrates that the optimisation also works for extremely simplified models like the JC69 model used here. Branch support was assessed using 1000 rapid bootstraps (Stamatakis et al., 2008) .

Bootstrap bipartition support was used to evaluate the phylogenetic outcomes between the ambiguous and informative treatments. Bipartitions were paired between the 2ISP-ambiguous and 2ISP-informative treatments for each method per dataset. In order to avoid the majority of bipartition pairs with only low support values, a threshold of 70% was used to eliminate any bipartition pairs where support values were both lower than the threshold. The number of bipartitions supported above the threshold for each simulated dataset (A_1, A_2 and $A_{1\&2}$) was calculated and converted to a percentage of the supportable bipartitions present in the true tree (number of tips minus two in all cases, therefore 18). The nonparametric Mann-Whitney signed rank test (Hollander and Wolfe, 1973, Pg. 68–75) was used (as the distributions of the bipartition support results were often severely skewed) to determine whether there were overall significant differences in the percentage of supported bipartitions between datasets (A_1, A_2, and $A_{1\&2}$) within each algorithm and treatment combination (e.g. NJ-A). The test was two-sided and significance determined as $\alpha < 0.05$. To investigate the occurrence of false positives (i.e. supported bipartitions that are not found on the 'true' tree), the bipartitions supported above the 70% threshold were also compared to the tree used to simulate the data. The false positive rate was calculated as the proportion of incorrectly supported bipartitions to the sum of all supported bipartitions.

Published datasets

Performances of the 2ISP-informative and 2ISP-ambiguous treatments were compared using 21 previously published DNA alignments and a novel dataset generated for this study (see Case Studies below). Included are datasets with 2ISPs found in direct-sequenced PCR products as well as strict individual-consensus sequences based on a collection of clones. Strict individual-consensus sequences are constructed by coding any sites found to be polymorphic across the population of clones sampled from an individual with IUPAC ambiguity codes. Other consensus approaches may use frequency thresholds to determine whether the consensus sequence are coded with the dominant base or IUPAC ambiguity codes (see Göker and Grimm, 2008, for further consensus construction details). As 2ISPs occur on many other nuclear gene regions, datasets were not limited to ITS. Aligned sequences were either obtained directly from the authors or downloaded from TreeBase (www.treebase.org). If indel coding was included in the dataset, then these were kept for subsequent analyses. Three datasets were also treated as case studies to explore the advantages and limitations of the 2ISP-informative approach, and are discussed in the next section.

All datasets were analysed using NJ, MP and ML using the same software employed for the simulation studies; NJ-I analyses included pairwise deletion of missing characters (such characters were inferred for NJ-A). NJ-A and NJ-I analyses used 10,000 bootstrap replicates to assess branch support. Assessment of MP-A and MP-I bootstrap support followed the suggestions of Müller (2005), with 10,000 bootstrap replicates composed of a single random sequence replicate and TBR branch swapping. ML-A and ML-I analyses used 10,000 rapid bootstraps to assess branch support. Summary information for each dataset was calculated using the APE AND PHANGORN libraries in R.

We compared the difference of support values between the ambiguous versus informative treatment (Δ support) for each algorithm against information content within standard DNA and 2ISPs across the datasets. The Δ support values were calculated as the difference in paired (equivalent) bipartition support between the 2ISP-informative and 2ISP-ambiguous treatments where at least one of the pairs

received support above a low, moderate or high bootstrap support threshold (>50%, >70% or >90%, respectively). The thresholds were used to avoid calculating Δ support values for bipartition pairs that only received weak support (<50%); such pairs often dominated the pairwise bipartition comparisons (low support may be due to other causes: chance effects, lack of signal strength or competing signals, [e.g. inflicted by hybrids]). The Δ support values were calculated per dataset, and the non-parametric Wilcoxon signed rank test was used (as the bipartition results often did not follow a normal distribution and/or sample sizes were low) to determine if the distribution of Δ support values were significantly different from zero within each dataset (two-sided test; $\alpha<0.05$). Significance was not determined if there were less than five bipartitions greater than the threshold between the two treatments for a given dataset. The value and significance of Δ support were compared to the information content of the underlying dataset.

The information content within each dataset was characterised using a parsimony-informative sites index (P-index), which was calculated as follows:

$$P = \frac{PI_{2ISP} - PI_{std}}{PI_{2ISP} + PI_{std}}$$

where PI_{2ISP} and PI_{std} are the number of parsimony-informative sites for 2ISP and standard DNA characters, respectively. The P-index ranges from -1 to 1, where all parsimony informative sites are either exclusively standard or 2ISP DNA characters, respectively. All statistical analyses were performed in R.

To detect potentially incompatible or ambiguous signals in the datasets, such as those caused by hybridisation or allopolyploidisation, we inferred neighbour-net splits graphs (Bryant and Moulton, 2004) and consensus networks (Holland and Moulton, 2003) based on the bootstrap samples ("bootstrap networks", e.g. Grimm et al., 2006) using SPLITSTREE version 4.8 (Huson and Bryant, 2006) for a subset of published studies and the *Nymania* dataset. As SplitsTree does not treat polymorphisms as informative, a distance matrix based on the polymorphism *p*-distance was used to produce the neighbour-net splits graphs. Alternative support for competing phylogenetic splits was investigated using the bootstrap tree replicate samples obtained under NJ, MP and ML with the different 2ISP-treatments. The NJ, MP and

ML bootstrap analyses were conducted using the same settings used for the published dataset analyses. The tree consensus network module implemented in SPLITSTREE was then used to calculate bootstrap networks, i.e. multidimensional graphs in which the edge lengths are proportional to the frequency of the corresponding phylogenetic splits in the bootstrap sample ("Edge Weight" option set to "COUNT") using a threshold of 0.2. The threshold value ensures that only splits that were found in at least 20% of the bootstrap trees were represented in the bipartition networks. The support of alternative phylogenetic splits in the bootstrap networks for each combination of algorithm and treatment was then manually plotted onto the neighbour-net splits graphs based on polymorphism *p*-distances.

Case studies: Acer, Hieracium *and* Nymania

We used data from three angiosperm groups, *Acer* sect. *Acer* (Sapindaceae; Grimm et al., 2007; Göker and Grimm, 2008), *Hieracium* L. s.l. (Asteraceae; Fehrer et al., 2009) and *Nymania capensis* (Meliaceae; this study), that exhibited intra- and inter-individual 2ISP variability. The *Hieracium* and *Nymania* datasets were generated from direct-PCR sequencing and polymorphic sites were coded if they occurred in both reading directions. In contrast, the *Acer* dataset comprised strict individual-consensus sequences of ITS clones (see Göker and Grimm, 2008, for consensus construction details). The *Hieracium* and *Acer* datasets contain numerous putative hybrids (identified by 2ISPs and clone sequences, respectively) and were used to explore the effects of hybrids on phylogenetic support. The *Hieracium* dataset contained 60 sequences from the 5' external transcribed spacer (ETS) of the 35S rDNA. The putative hybrids in the *Hieracium* dataset were between two genetically and geographically divergent clades (Fehrer et al., 2009). The *Acer* dataset contained ITS consensus sequences from 27 individuals, including five that showed evidence of reticulation/lineage crossing with signals that were analogous to F1-hybrids (Grimm et al., 2007).

The *Nymania capensis* dataset comprised 30 individuals sampled across three primary drainage basins in the Albany Subtropical Thicket biome which spans the Western and Eastern Cape Provinces of South Africa (collection details are given in Table

S2.1). Ten individuals were sampled per drainage basin. Two additional individuals from the disjunct northern distribution of the species were used as outgroups (BOL48535 and BOL60966). Genomic DNA was extracted from silica-dried leaf material using a modified version of the method specified by (Gawel and Jarret, 1991). Polyvinylpyrrolidone-40 (PVP) was added when grinding the leaf material in liquid nitrogen using a mortar and pestle. Nuclear variation was sampled for the ITS-1, 5.8S and ITS-2 region using the primers ITS5m (Sang et al., 1995) and ITS4 (White et al., 1990). PCR reactions were performed in 25 µl, with 5 µl 1 × KAPA HiFi Buffer, 0.75 mM dNTPs, 0.75 mM forward primer, 0.75 mM reverse primer, 0.4 µl of the proofreading KAPA HiFi DNA polymerase (2 Units) and 1.2 µl template DNA (~1–5 ng). PCR was conducted using a GeneAmp 2700 PCR System thermocycler (Applied Biosystems, USA) under the following conditions: initial denaturation and polymerase activation at 98°C for 20 seconds (s) followed by 30 cycles of 94°C for 45 s, 58°C for 30 s, 72°C for 30 s; and a final extension at 72°C for 1 minute. All sequences were aligned using CODON CODE ALIGNER version 3.5.7 (Codon Code Corp, http://www.codoncode.com). The following steps were followed in order to identify polymorphic sites across and within sequences: (1) each base-call was assigned a quality score using the automated base-calling program PHRED (Ewing et al., 1998), (2) sites containing secondary peaks greater than 30% of the primary peaks were scored as polymorphic using the `Call second peaks' option in CODON CODE ALIGNER, and (3) all polymorphic sites were verified by eye. Following Fehrer et al. (2009), overlapping and non-overlapping peaks were coded in capitals letters and small letters, respectively. In order to determine if pseudogenes were present in ITS we checked for four conserved angiosperm motifs, one in ITS-1 (Liu and Schardl, 1994) and three within 5.8S (Harpke and Peterson, 2008) for mutations.

TABLE S2.1. Collection details of *Nymania capensis* samples. Voucher samples are stored at the Bolus Herbarium.

Drainage basin	Province	District	Collection Number	Lattitude	Longitude	Voucher Sample
Gouritz	Western Cape	Ladismith	AJP0050	-33.593233	21.201433	Y
		Oudtshoorn	AJP0071	-33.548433	22.463717	
		Oudtshoorn	AJP0074	-33.487867	22.561767	
		Ladismith	AJP0083	-33.515213	21.137982	
		Ladismith	AJP0121	-33.536824	20.748489	
		Oudtshoorn	AJP0224	-33.610121	22.405158	
		Oudtshoorn	AJP0229	-33.635754	22.403977	
		Uniondale	AJP0238	-33.549492	22.802571	
		Uniondale	AJP0243	-33.491221	23.282956	
	Eastern Cape	Willowmore	AJP0774	-33.275740	23.288660	Y
Gamtoos	Eastern Cape	Willowmore	AJP0270	-33.513244	23.780121	
		Hankey	AJP0517	-33.806600	24.728710	Y
		Hankey	AJP0518	-33.820320	24.729750	Y
		Steytlerville	AJP0540	-33.309960	24.358470	Y
		Steytlerville	AJP0545	-33.226750	24.198500	
		Steytlerville	AJP0551	-33.276000	24.134070	Y
		Steytlerville	AJP0553	-33.228690	24.088070	Y
		Willowmore	AJP0555	-33.143820	23.841080	Y
		Willowmore	AJP0780	-33.399970	23.675520	Y
Sundays	Eastern Cape	Somerset East	AJP0465	-33.251306	25.442722	
		Jansenville	AJP0466	-33.210417	24.839944	
		Jansenville	AJP0492	-32.839710	24.713340	Y
		Jansenville	AJP0495	-33.000480	24.745880	
		Uitenhage	AJP0500	-33.341830	24.909610	Y
		Uitenhage	AJP0532	-33.342660	24.873420	Y
		Jansenville	AJP0628	-33.092010	24.886390	Y
		Jansenville	AJP0632	-33.077230	25.035360	Y
		Somerset East	AJP0633	-33.070510	25.176080	
		Somerset East	AJP0637	-33.039110	25.282500	Y
		Uitenhage	AJP0822	-33.542490	25.119780	Y
Outgroup	Northern Cape	Namakwaland	BOL48535	-28.256006	17.241669	Y
	Northern Cape	Namakwaland	BOL60966	-28.316668	17.249999	Y

References

Bryant, D., and V. Moulton. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution 21:255-265.

Cronn, R., M. Cedroni, T. Haselkorn, C. Grover, and J. F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theoretical and Applied Genetics 104:482-489.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Research 8:175-185.

Fehrer, J., K. Krak, and J. Chrtek. 2009. Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise. BMC Evolutionary Biology 9:239.

Gawel, N. J., and R. L. Jarret. 1991. A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Molecular Biology Reporter 9:262-266.

Göker, M., and G. Grimm. 2008. General functions to transform associate data to host data, and their use in phylogenetic inference from sequences with intra-individual variability. BMC Evolutionary Biology 8:86.

Grimm, G. W., T. Denk, and V. Hemleben. 2007. Evolutionary history and systematics of *Acer* section *Acer* - a case study of low-level phylogenetics. Plant Systematics and Evolution 267:215-253.

Grimm, G. W., S. S. Renner, A. Stamatakis, and V. Hemleben. 2006. A nuclear ribosomal DNA phylogeny of *Acer* inferred with maximum likelihood, splits graphs, and motif analysis of 606 sequences. Evolutionary Bioinformatics 2:7-22.

Harpke, D., and A. Peterson. 2008. 5.8S motifs for the identification of pseudogenic ITS regions. Botany 86:300-305.

Holland, B., and V. Moulton. 2003. Consensus networks: a method for visualising incompatibilities in collections of trees. Pages 165-176 *in* Algorithms in Bioinformatics, WABI 2003 (G. Benson, and R. D. M. Page, eds.). Springer Verlag, Berlin, Germany.

Hollander, M., and D. A. Wolfe. 1973. Nonparametric Statistical Methods. John Wiley & Sons, New York.

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23:254-267.

Liu, J., and C. Schardl. 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Molecular Biology 26:775-778.

Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. Systematic Biology 52:674-683.

Müller, K. 2005. The efficiency of different search strategies in estimating parsimony jackknife, bootstrap, and Bremer support. BMC Evolutionary Biology 5:58.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. N. Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47-67.

Paradis, E. 2012. Analysis of Phylogenetics and Evolution with R. Springer, New York.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289-290.

Posada, D., and K. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria.

Rodriguez, F., J. L. Oliver, A. Marin, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142:458-501.

Sang, T., D. J. Crawford, and T. F. Stuessy. 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. Proceedings of the National Academy of Sciences, U.S.A. 92:6813-6817.

Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592-593.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. Systematic Biology 57:758-771.

White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 *in* PCR protocols: A guide to methods and applications (M. Innis, D. Gelfand, J. Sninsky, and T. White, eds.). Academic Press, San Diego.