Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 44 (2007) 911–919
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Review

Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants

Gonzalo Nieto Feliner [a,*], Josep A. Rosselló [b]

[a] *Real Jardín Botánico, CSIC, Plaza de Murillo 2, E-28014 Madrid, Spain*
[b] *Jardín Botánico, Universidad de Valencia, c/Quart 80, E-46008 Valencia, Spain*

Received 15 September 2006; revised 3 December 2006; accepted 10 January 2007
Available online 15 February 2007

Abstract

The internal transcribed spacers (ITS) of the nuclear ribosomal 18S–5.8S–26S cistron continue to be the most popular non-plastid region for species-level phylogenetic studies of plant groups despite the early warnings about their potential flaws, which may ultimately result in incorrect assumptions of orthology. It has been gradually realized that the alternative target regions in the nuclear genome (low-copy nuclear genes, LCNG) are burdened with similar problems. The consequence is that, to date, developing useful LCNG for non-model organisms requires an investment in time and effort that hinders its use as a real practical alternative for many labs. It is here argued that ITS sequences, despite drawbacks, can still produce insightful results in species-level phylogenetic studies or when non-anonymous nuclear markers are required, provided that a thoughtful use of them is made. To facilitate this, two series of guidelines are proposed. One helps to circumvent problems of ITS amplification from the target organism, including spurious results from contaminants, paralogs and pseudogenes, as well as detection of sequencing artifacts. The other series helps to find out causes for unresolved clades in phylogenetic reconstruction, to integrate gene phylogenies, to distinguish horizontal transfer from lineage sorting, and to reveal if ITS phylogeny is not a good estimate of organism phylogeny.

Keywords: Low-copy nuclear genes; nrDNA ITS; Orthology; Plant phylogeny

1. Introduction

For over a decade, the internal transcribed spacers (ITS) of the nuclear ribosomal 18S–5.8S–26S cistron have been the most popular target region in the nuclear genome for evolutionary studies of diverse plant groups (Baldwin et al., 1995; Álvarez and Wendel, 2003; Hughes et al., 2006).

The are four main reasons for this widespread use. (i) The availability of several sets of universal (or near so) PCR primers working with a large diversity of taxonomic groups (White et al., 1990; Gardes and Bruns, 1993) is the first one. (ii) The multicopy structure facilitates PCR amplification even from herbarium specimens. (iii) The moderate size of the ITS (below 700 bp) usually allows amplification and sequencing without internal primers although they are remarkably long exceptions in several gymnosperm groups (Gernandt et al., 2001). (iv) Due to their levels of variation, ITS frequently provide enough molecular markers suitable for evolutionary studies at the species level. These include topics such as the origin of polyploid taxa, hybridization, introgression and, above all, phylogenetic inference (up to 40% sequence divergence in pair-wise comparisons between congeneric taxa reported in early studies, Baldwin et al., 1995).

Several other factors have apparently contributed to a spectacular rise in their use in plant phylogenetic studies. The lack of alternative variable regions within the three plant genomes that could provide useful markers for low taxonomic level studies was surely one of them. This situation contrasts with animal studies for which the mitochondrial DNA provides markers for a wide range of evolutionary

* Corresponding author. Fax: +34 914200157.
E-mail address: nieto@rjb.csic.es (G.N. Feliner).

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.01.013

questions, from intraspecific phylogeography (Avise et al., 1987) to deep origin of major vertebrate groups (Zardoya and Meyer, 1996). Other postulated potentially advantageous features of ITS sequences are their biparental inheritance compared to the uniparental inheritance of organellar DNAs, and the assumed intragenomic uniformity due to the active homogenization of repeat copies within and between loci that take place in these multicopy regions (but see below), known as concerted evolution (Zimmer et al., 1980; Arnheim, 1983). On the other hand, in hybrids and introgressants in which concerted evolution has not homogenized copies, ITS may help to identify progenitors or lineages involved (Sang et al., 1995).

Finally, the predominant background of ITS users (taxomically trained rather than molecularly trained scientists who welcome easy universal lab protocols) and a bandwagon effect, possible contributed to the use of ITS. This became so widely established in species-level phylogenetic investigations of plant groups that the profile of a study sampling ITS and a plastid non-coding region (Taberlet et al., 1991) became almost routine (Hughes et al., 2006).

The use of ITS regions has not been restricted to green plants. These markers have been also utilized across a wide scope of taxonomic diversity spanning virtually the tree of life, including fungi and lichenicolous fungi (Wu et al., 2000; Martín et al., 2003; Cubero et al., 2004) unicellular and pluricellular algae (Van Oppen et al., 2005; Leclerc et al., 1998), non-arthropod invertebrates (Dumont et al., 2005), arthropods (Harris and Crandall, 2000), and even vertebrates (Booton et al., 1999).

2. Drawbacks of the nuclear ITS region

Concomitant to the seminal paper of Baldwin et al. (1995), Dubcovsky and Dvorák (1995) warned against the use of ribosomal sequences in phylogenetic reconstruction. Their observations assessed the dynamic nature of NOR loci in Triticeae. This involved transposition of major loci within and among chromosomes possibly mediated by minor loci containing a few rDNA copies. These findings implied a violation of the assumed orthology of ITS sequences that could lead to wrong conclusions in analysis requiring a sound hypothesis of character homology. Surprisingly, this lucid caution was not echoed in the molecular systematist community until recently, when other voices have recommended routine utilization of low-copy nuclear genes given the potential pitfalls of the ITS region for inferring phylogenies (Álvarez and Wendel, 2003; Sang, 2002; Small et al., 2004). Most of the alleged flaws that dismiss the phylogenetic utility of ITS sequences stem from the molecular architecture of the ribosomal units, as pointed out by Álvarez and Wendel (2003).

2.1. Multiple rDNA arrays

The ribosomal region is composed by the 18S, 5.8S, and 25S genes, two internal spacers (ITS-1 and ITS-2) and the intergenic spacer (IGS). The coding and ITS regions, and a part of the IGS (ETS) form the transcriptional unit that is further processed to produce the mature RNAs, which are part of the cytoplasmic ribosomes. Each locus is formed by hundreds to thousands of tandem copies of such transcriptional unit. Besides, several ribosomal loci, either transcriptionally active (NOR) or inactive, are usually present within plant genomes. The general assumption in phylogenetic reconstruction is that all ribosomal copies present within the genome have identical sequences due to functional constraints. The ribosomal multigene family usually evolves in concert (Arnheim, 1983) because all copies within and among ribosomal loci are expected to be homogenized through genomic mechanisms of turnover like gene conversion and unequal crossing over (Dover, 1994).

Thus, the non-cloned ITS sequence retrieved from a PCR product is the consensus of many targets sharing the same priming sites in one or several loci usually located in separate chromosomes. Therefore, the consensus ITS sequence that is used as raw data in phylogenetic reconstruction is a molecular phenotype from which the genotype of the organism cannot always be inferred. Specifically, neither the number of ribosomal loci nor the presence of allelic variants can be deduced, thus the homozygous or heterozygous condition for this marker cannot be deduced either. Cloning efforts, which unfortunately are infrequent in routine phylogenetic projects, may unveil divergent intragenomic copies thereby decomposing the alluded consensus sequence. However, these copies may have various possible origins that cannot be straightforwardly deduced. First, if the rate of mutation among copies is faster than the molecular forces driving the concerted evolution of the array, those sequences may correspond to divergent targets located within a single loci that has not fully homogenized all the arrays. Second, functional loci could be duplicated in the absence of polyploidy. In this case, new loci could be originated from amplification of pre-existent loci. Ectopic recombination between terminal chromosomal regions might be the mechanism responsible for this phenomenon (Pedrosa-Harand et al., 2006). Alternatively, repetitive sequences in the IGS or scattered along rDNA units, rather than chromosome rearrangements, could play an important role in the dispersion of NORs (Castro et al., 2001). Third, different ITS sequences may belong to homeologous (xenologous) loci incorporated into the nuclear genome through hybridization either or not involving polyploidy. Fourth, significantly different sequences may have originated in multiple loci, some of which have evolved without selective constraints, accumulating mutations and ultimately becoming non-functional (pseudogenes). Finally, divergent sequences may represent true allelic variants of a homologous locus.

The existence of orthologs and paralogs in ITS is thus real but a thorough analysis based on a representative sampling enables telling them apart provided that both orthologs and paralogs are detected (Mayol and Rosselló, 2001; Razafimandimbison et al., 2004). This usually requires

Table 1
Features of the nuclear ribosomal internal transcribed spacers (ITS) and low-copy nuclear genes (LCNG)

	ITS	LCNG
Multigene families	Always present	Usually not present
Duplicated loci	Usually present	Present in some genes (most readily detected for those genes coding for soluble enzymes)
Concerted evolution	Usually operating	Usually absent
Recombination	Present between alleles, arrays and loci	Present between alleles
Non-active loci and pseudogenes	Present	Present
Secondary structure	Present	Present in intron regions
Alignment problems	Usually between distant species and between genera	Usually between genera (at non-coding regions)
Universality of primers	Apply to most groups of organisms although specific primers work better for certain groups (e.g., fungi; Gardes and Bruns, 1993)	Cases of universal primers for LCNG are rare (e.g., Strand et al., 1997)

cloning and possibly also molecular cytogenetic techniques that can contribute to substantiate the occurrence of paralogs from different rDNA arrays across the genome.

2.2. Concerted evolution

Although mechanisms of concerted evolution tend to homogenize sequences in genomic nrDNA arrays in a process that operates within whole reproductive groups, this process is not always operating or found completed. When different ITS repeats are merged within a single genome via hybridization (including allopolyploidy) or introgression the speed and direction of homogenization cannot be predicted and is not consistent across different descendant lineages (Álvarez and Wendel, 2003). Genomic features tending to retard homogenization include the occurrence of several rDNA loci located in separate chromosomes (which is particularly frequent in allopolyploids). This is consistent with reports indicating that homogenization of the rDNA repeats through gene conversion and unequal crossing over occur more effectively within than between loci (Ohta and Dover, 1983; Schlotterer and Tautz, 1994). The chromosomal location of rDNA loci (either interstitial or terminal) may also play a role. In fact, it has been suggested that this factor has a more substantial impact than the number of loci on the tempo of concerted evolution (Zhang and Sang, 1999).

Agamic reproduction is also known to retard concerted evolution (Campbell et al., 1997) but, on the opposite direction, biased fast homogenization towards a parental ribosomal sequence is already detectable in artificial F_2 hybrids in *Armeria* (Fuertes Aguilar et al., 1999a). Three stages can be found depending on (i) whether the different ITS repeats are maintained without recombination or homogenization (e.g., Soltis and Soltis, 1991; Ritland et al., 1993), (ii) the different repeats undergo some degree of homogenization giving rise to chimeric sequences (e.g., Buckler et al., 1997; Nieto Feliner et al., 2004); or (iii) one repeat becomes dominant within the new genome (Wendel et al., 1995).

When analyzing species-level scenarios, incomplete concerted evolution is probably responsible for complex patterns following the merging of ITS repeats within a single genome via hybridization (including allopolyploidy) or introgression. Sampling all ITS copies, independently of their number is recommended (Rauscher et al., 2002). On the other hand, retention of ITS copies is a useful property for identifying reticulation. When examined together with the distribution of repeats across geographic areas and species, it can be effective to discard the possibility of lineage sorting (Fuertes Aguilar et al., 1999b; Nieto Feliner et al., 2004).

2.3. Pseudogenes

When concerted evolution is not fully operating, duplicate ribosomal loci do not necessarily remain functional and some arrays may degenerate into pseudogenes. If neglected, this can lead to wrong inferences of the phylogenetic relationships at the organism level, as shown by Mayol and Rosselló (2001).

However, ITS pseudogenes do not seem to pose serious problems to phylogenetic inference because they can be detected when a careful scrutiny is followed (Buckler et al., 1997; Mayol and Rosselló, 2001; Hughes et al., 2002). In fact, several recent studies have shown that putative pseudogenes can be useful for phylogenetic analyses of related species where functional ITS copies are less likely to provide variation (Razafimandimbison et al., 2004; Besnard et al., 2007).

2.4. Secondary structure

The ribosomal ITS are subjected to evolutionary constraints related to the maintenance of specific secondary structures necessary for the correct processing of mature RNAs (Mai and Coleman, 1997). This could imply the occurrence of compensatory base mutations in positions located on stem structures that violate the assumptions of neutrality and independence of characters (Liu and Schardl, 1994; Hillis and Dixon, 1991). Like pseudogenization, evolutionary constraints related to the maintenance of specific secondary structures can be detected by inspecting low-energy secondary structure models (Wolf et al., 2005).

2.5. Alignment and homoplasy

It has been argued that not being protein-coding genes and frequently containing indels, ITS sequences may be

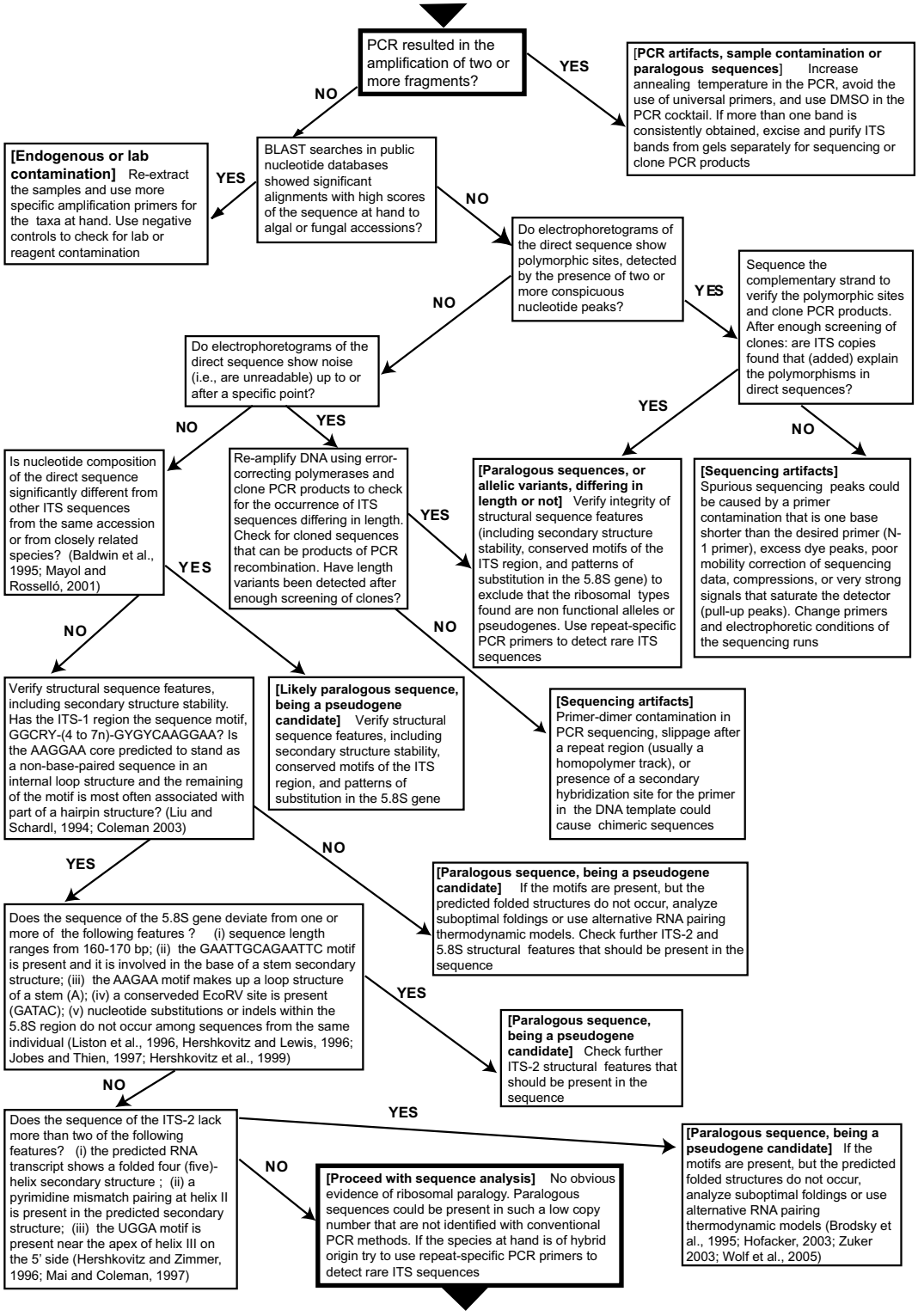

Fig. 1. Chart 1: Guidelines for obtaining a reliable ITS sequence in plants. (See above-mentioned references for further information.)

difficult to align across unrelated species and lead to incorrect homology assessments. According to Álvarez and Wendel (2003), this may be a reason why ITS sequence data are reported to contain more homoplasy than other markers, thereby hindering phylogenetic signal and ultimately resolution of the phylogenies. However, indels seldom pose serious problems in practical phylogenetic reconstruction at the species-level where close relationships facilitate alignment and where the need for useful molecular markers is more imperative.

Further, homoplasy in ITS-based phylogenies depends on several causes, e.g., on the degree of homogenization of

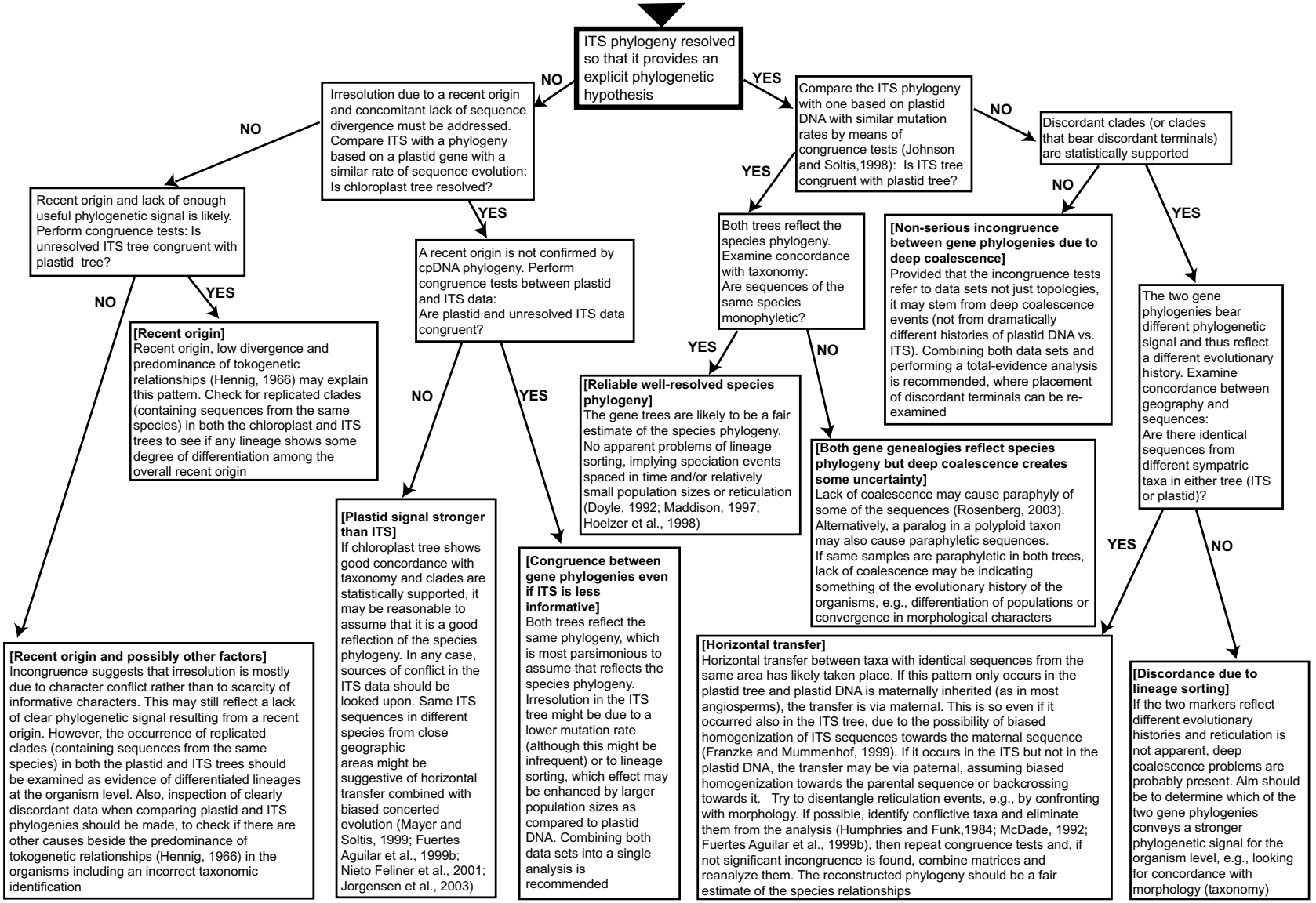

Fig. 2. Chart 2: Guidelines for analyzing ITS sequences as estimators of species-level organism phylogeny. (See above-mentioned references for further information.)

copies. In fact, it tends to be low in advanced stages of concerted evolution (Nieto Feliner et al., 2001) as predicted in simulations (Sanderson and Doyle, 1992). Besides, homoplasy ultimately results from mixing different historical signals (Doyle, 1996) and when a significant amount is detected, it may be revealing an undetected process.

Finally, contamination caused by the universality of the primers has also been pointed as a potential drawback for ITS sequences (Zhang et al., 1997). But this is usually an anecdote rather than a concern.

3. Low-copy nuclear genes as alternative

Concern about the above potential pitfalls of ITS regions has prompted recommendations to survey alternative regions containing useful phylogenetic signal at the species level (Álvarez and Wendel, 2003; Sang, 2002; Small et al., 2004). Theoretically, low-copy nuclear genes (LCNG) offer advantageous properties for species-level phylogenies. Overall, rates of mutation are potentially higher than in organellar genes and thus LCNG are expected to contain more variable sites for reconstructing evolution of genes and organisms (Gaut, 1998).

For this purpose, species-level phylogenetic inference has focused on non-coding regions such as introns although recent studies are finding similar or even faster rates in 3rd codon positions (Small and Wendel, 2002). Further, wide variation in mutation rates in LCNG (Senchina et al., 2003) could provide enough phylogenetic signal for addressing different evolutionary questions. The existence of tens of thousands of LCNG across the genome represents a theoretical wealth of independent loci, and thus an immense source of markers for phylogenetic inference. Biparental inheritance is another desirable property, as compared to plastid and mitochondrial sequences, since it provides information on the maternal and paternal parentage, which is crucial when the evolutionary history of a group involves reticulation events.

To date, the real utility of LCNG for phylogenetic reconstruction is however less encouraging than their potential advantages would predict. This is related to the fact that their use in this field is in its "relative infancy" (Small et al., 2004). It is not surprising perhaps that the main difficulty for the use of LCNG has to do with the identification of orthologous genes, just as reported for nuclear ribosomal DNA (Álvarez and Wendel, 2003).

Two genes are orthologs if their relationship originated from organismal cladogenesis (Wendel and Doyle, 1998). The LCNG tend to exist in gene families, i.e., in multiple copies related by events of genomic duplication (i.e., paralogs) or capture through endosymbiontic events associated to the acquisition of plastids and mitochondria from prokaryote ancestors (Allen, 2003). Thus, the percent of available single-copy nuclear genes, not just low-copy, which would be the ideal targets for phylogenetic analysis is probably low. Given the need to compare orthologs to infer phylogenies at the organismal level, it is recommended to characterize gene family composition prior to phylogenetic analysis. But this may represent considerable time and effort particularly when allopolyploidy (a key mechanism in plant evolution) is involved. Even though several tests have been suggested (Small et al., 2004), the concern for identifying true orthologs remains until the phylogenetic analysis of the sequences is completed (Doyle and Doyle, 1999).

Intraspecific variation is inherent to LCNG as has been routinely detected by surveys of electrophoretic variation of genes coding for soluble enzymes (Hamrick and Godt, 1989; Loveless and Hamrick, 1984). Allelic variation within individuals, populations and species may be large in LCNG due to large effective population sizes as compared to organellar genes. This allelic variation may not hinder phylogenetic inference in the relatively infrequent instance where alleles found in related species are reciprocally monophyletic. However, this is not often the case possibly due to deep coalescence favored by large effective population sizes, as compared to much smaller organellar genomes (Small et al., 2004). Additionally, natural selection favors trans-species polymorphism in certain genes, e.g., the S genes involved in self-incompatibility (Charlesworth and Awadalla, 1998; Klein et al., 1998). When the recommendable goal of sampling multiple independent nuclear loci is achieved, a potential concern arises caused by the different mutation rates across loci. Specifically, a highly variable locus may override the signal contained in less-variable sampled loci and possibly distort the analysis (Doyle et al., 2003; Hughes et al., 2006).

The use of LCNG is thus not devoid of problems in phylogenetic inference, to the point that Hughes et al. (2006) consider that the development of sequence loci "has been somewhat of a lottery" requiring considerable investment and not always generating enough data. If the main alternative to ITS as a source of nuclear data for reconstructing phylogenies (LCNG) is not yet methodologically available for a high percent of labs that have previously used ITS data, should ITS "no longer be routinely utilized for phylogenetic analysis" as suggested by Álvarez and Wendel (2003)?

4. A careful use of ITS

We think that the systematic and phylogenetic plant community is not yet ripe for abandoning ITS as an important and easy-to-work, albeit not *the only*, nuclear marker. A survey of 244 plant phylogenetic papers at the genus level or below between 1998 and 2002 found that 66% used ITS, and 34% used ITS as the only marker (Álvarez and Wendel, 2003). We have surveyed the same four publications (*American Journal of Botany, Molecular Phylogenetics and Evolution, Systematic Botany, Taxon*) for papers appeared during 2005 containing phylogenetic analysis of plants and fungi below the familial level. Of the 114 papers meeting that requirement, 75 used ITS sequence data, which is exactly the same percentage (66.4%) as in the period 1998–2002. But in 2005 only 18 of those papers (15.7%) used ITS

as the only marker. In our view, this shows that the current trend is the most reasonable, i.e., to continue to use ITS but in combination with plastid and mitocondrial sequences, and certainly, where feasible, with LCNG.

Public nucleotide databases contain many thousands of plant ITS sequences (e.g., more than 4500 entries of 444 different genera in Asteraceae, Gemeinholzer et al., 2006), that constitute an invaluable raw data for taxonomic identification methods using DNA sequences. A recent evaluation of the potential of ITS1 sequences for species, genus, and family identification in Asteraceae (tribes Lactuceae and Anthemideae) has given very promising results (Gemeinholzer et al., 2006). In fact, these authors echoed the proposal of the Plant Working Group to the Consortium of the Barcode of Life to evaluate the nuclear ITS region and the plastid *trn*H-*psb*A spacer as the target markers to use in plant identification.

Just as primary disadvantages of LCNG come from their "complex genetic architecture and evolutionary dynamics" (Small et al., 2004), problems with ITS stem ultimately from complexity derived from the evolution of these markers at several levels, from the gene to the population. In fact, most of the problems with the ITS region could potentially be also present when using LCNG (Table 1).

We are now far from the early molecular phylogenetic times in which analysis of DNA sequences were expected to yield the true phylogenies of the organisms. It has been realized in a perceptive paper that phylogenetic incongruence can offer a window into previously undetected evolutionary processes (Wendel and Doyle, 1998). Phylogenies based on ITS data may cause incongruence with those based on other markers due to the various mechanisms that influence ITS. But, to the extent that such incongruence can be further looked upon, it has a true potential to render insightful results, as was the case in most of the studies on which the warnings about ITS were based (references in Álvarez and Wendel, 2003).

In conclusion, we think that the proposal to abandon ITS is not realistic at present. LCNG should be the future main source of nuclear data in species-level phylogenetic studies but while more empirical work is done on a wider sample of loci across natural groups, it seems premature to give up using a marker like ITS. We recommend that ITS sequences continue to be used for species-level phylogenies, albeit no routinely nor in exclusivity, keeping an awareness of the potential problems summarized in Álvarez and Wendel (2003) and with three general recommendations: representative samplings following prospective pilot studies, careful lab protocols, mindful analysis.

5. Guidelines

To help handling carefully ITS sequences, we here provide two series of guidelines in the form of flow-charts, which aim at improving the quality of ribosomal data and their interpretation in the context of organism phylogenies of plant groups. They are not intended to be followed strictly as a dichotomous key but to be used more freely, so as to offer ideas that may assist when using these markers. Chart 1 aims at helping to solve problems of amplification, detection of pseudogenes and paralogs, contamination and sequence artifacts (Fig. 1). Chart 2 aims at helping to find out causes for unresolved clades, to integrate gene phylogenies, to detect horizontal transfer and lineage sorting, and to reveal if ITS phylogeny is not a good estimate of organism phylogeny (Fig. 2). The term *resolved* in the entrance to Fig. 2 is intended here in relative terms. Irresolution does not refer to sequences from the same taxon that are monophyletic but gathered in a polytomy, obviously one of the best possible situations one can find in terms of congruence between taxonomy and gene phylogeny. Also, a large tree with a few small terminal unresolved clades containing more than one species is not considered in this context, a phylogeny unresolved either.

Acknowledgments

We thank Inés Álvarez, Pilar Catalán, Javier Fuertes, Mari Paz Martín, Pablo Vargas, Jonathan F. Wendel and two anonymous reviewers for their constructive comments to an earlier draft of the manuscript. This work has been supported by grants BOS2001-1839 (to GNF) and CGL2004-00223/BOS (to JAR) from the Spanish Ministry of Science and Education.

References

Allen, J.F., 2003. Why chloroplasts and mitochondria contain genomes. Comp. Funct. Genom. 4, 31–36.

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434.

Arnheim, N., 1983. Concerted evolution of multigene families. In: Nei, M., Koehn, R. (Eds.), Evolution of Genes and Proteins. Sinauer, Sunderland, MA, pp. 38–61.

Avise, J.C., Arnold, J., Ball Jr., R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reed, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18, 489–522.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. 82, 247–277.

Besnard, G., Rubio de Casas, R., Vargas, P., 2007. Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). J. Biogeogr. doi:10.1111/j.1365-2699.2006.01653.x.

Booton, G.C., Kaufman, L., Chandler, M., Oguto-Ohwayo, R., Duan, W., Fuerst, P., 1999. Evolution of the ribosomal RNA internal transcribed spacer one (ITS-1) in cichlid fishes of the Lake Victoria region. Mol. Phylogenet. Evol. 11, 273–282.

Brodsky, L.I., Ivanov, V.V., Kalai, Y.L., Leontovich, A.M., Nikolaev, V.K., Feranchuk, S.I., Drachev, V.A., 1995. GeneBee-NET: Internet-based server for analyzing biopolymers structure. Biochemistry 60, 923–928.

Buckler, E.S.I., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: Divergent paralogous and phylogenetic implications. Genetics 145, 821–832.

Campbell, C.S., Wojciechowski, M.F., Baldwin, B.G., Alice, L.A., Donoghue, M.J., 1997. Persistent nuclear ribosomal DNA sequence polymor-

phism in the *Amelanchier* agamic complex (Rosaceae). Mol. Biol. Evol. 14, 81–90.

Castro, J., Rodríguez, S., Pardo, B.G., Sánchez, L., Martínez, P., 2001. Population analysis of an unusual NOR-site polymorphism in brown trout (*Salmo trutta* L.). Heredity 86, 291–302.

Charlesworth, D., Awadalla, P., 1998. Flowering plant self-incompatibility: the molecular population genetics of Brassica S-loci. Heredity 81, 1–9.

Coleman, A., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 19, 370–375.

Cubero, O.F., Crespo, A., Esslinger, T.L., Lumbsch, H.T., 2004. Molecular phylogeny of the genus *Physconia* (Ascomycota, Lecanorales) inferred from a Bayesian analysis of nuclear ITS rDNA. Mycol. Research 108, 498–505.

Dover, G., 1994. Concerted evolution, molecular drive and natural selection. Curr. Biol. 4, 1165–1166.

Doyle, J.J., 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17, 144–163.

Doyle, J.J., 1996. Homoplasy connections and disconnections: genes and species, molecules and morphology. In: Sanderson, M.J., Hufford, L. (Eds.), Homoplasy. In: The Recurrence of Similarity in Evolution Academic Press, pp. 67–89.

Doyle, J.J., Doyle, J.L., 1999. Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: Some examples from Leguminosae. In: Hollingsworth, P., Bateman, R., Gornall, R. (Eds.), Molecular Systematics and Plant Evolution. Taylor and Francis, London.

Doyle, J.J., Doyle, J.L., Harbison, C., 2003. Chloroplast expressed glutamine synthetase in *Glycine* and related Leguminosae: phylogeny, gene duplication and ancient polyploidy. Syst. Bot. 28, 567–577.

Dubcovsky, J., Dvorák, J., 1995. Ribosomal RNA multigene loci: nomads of the Triticeae genomes. Genetics 140, 1367–1377.

Dumont, H.J., Vanfleteren, J.R., De Jonckheere, J.F., Weekers, P.H.H., 2005. Phylogenetic relationships, divergence time estimation, and global biogeographic patterns of calopterygoid damselflies (odonata, zygoptera) inferred from ribosomal DNA sequences. Syst. Biol. 54, 347–362.

Franzke, A., Mummenhoff, K., 1999. Recent hybrid speciation in *Cardamine* (Brassicaceae)-conversion of nuclear ribosomal ITS sequences in statu nascendi. Theor. Appl. Genet. 98, 831–834.

Fuertes Aguilar, J., Rosselló, J.A., Nieto Feliner, G., 1999a. Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). Mol. Ecol. 8, 1341–1346.

Fuertes Aguilar, J., Rosselló, J.A., Nieto Feliner, G., 1999b. Molecular evidence for the compilospecies model of reticulate evolution in *Armeria* (Plumbaginaceae). Syst. Biol. 44, 735–754.

Gardes, M., Bruns, T.D., 1993. ITS primers with enhanced specificity for basidiomycetes: application to the identification of mycorrhizae and rusts. Mol. Ecol. 2, 113–118.

Gaut, B.S., 1998. Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht, M.K. (Ed.), Evolutionary Biology. Plenum Press, New York, pp. 93–120.

Gemeinholzer, B., Oberprieler, C., Bachmann, K., 2006. Using GenBank data for plant identification: possibilities and limitations using the ITS1 of Asteraceae species belonging to the tribes Lactuceae and Anthemideae. Taxon 55, 173–187.

Gernandt, D.S., Liston, A., Piñero, D., 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. Mol. Phylogenet. Evol. 21, 449–467.

Hamrick, J.L., Godt, M.J.W., 1989. Allozyme diversity in plant species. In: Brown, H.D., Clegg, M.T., Kahler, A.L., Weir, B.S. (Eds.), Plant Population Genetics, Breeding and Genetic Resources. Sinauer Associates, Sunderland, MA, pp. 43–64.

Harris, D.J., Crandall, K.A., 2000. Intragenomic variation within ITS1 and ITS2 of crayfish (Decapoda, Cambaridae): implications for phylogenetic and microsatellite studies. Mol. Biol. Evol. 17, 284–291.

Hennig, W., 1966. Phylogenetic Systematics. University of Illinois Press, Urbana, IL.

Hershkovitz, M.A., Lewis, L.A., 1996. Deep-level diagnostic value of the rDNA-ITS region. Mol. Biol. Evol. 13, 1276–1295.

Hershkovitz, M.A., Zimmer, E.A., 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. Nucleic Acids Res. 24, 2857–2876.

Hershkovitz, M.A., Zimmer, E.A., Hahn, W.J., 1999. Ribosomal DNA sequences and angiosperm systematics. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), Molecular Systematics and plant evolution. Taylor & Francis, London, pp. 268–326.

Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Q. Rev. Biol. 66, 411–453.

Hoelzer, G.A., Wallman, J., Melnick, D.J., 1998. The effects of social structure, geographical structure and population size on the evolution of mitochondrial DNA. II. Molecular clocks and the lineage sorting period. J. Mol. Evol. 47, 21–31.

Hofacker, I.L., 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429–3431.

Hughes, C.E., Bailey, C.D., Harris, S.A., 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. Am. J. Bot. 89, 1057–1073.

Hughes, C.E., Eastwood, R.J., Bailey, C.D., 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. Phil. Trans. R. Soc. B 361, 211–225.

Humphries, C.J., Funk, V.A., 1984. Cladistics methodology. In: Heywood, V.H., Moore, D.M. (Eds.), Current Concepts in Plant Taxonomy. Academic Press, London, pp. 323–362.

Jobes, D.V., Thien, L.B., 1997. A conserved motif in the 5.8S Ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) Sequences. Plant Mol. Biol. Rep. 15, 326–334.

Johnson, L.A., Soltis, D.E., 1998. Assessing congruence: empirical examples from molecular data. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II. DNA Sequencing. Kluwer Academic Publishers, Boston, MA, pp. 297–348.

Jorgensen, J.L., Stehlik, I., Brochmann, C., Conti, E., 2003. Implications of ITS sequences and RAPD markers for the taxonomy and biogeography of the *Oxytropis campestris* and *O. arctica* (Fabaceae) complexes in Alaska. Am. J. Bot. 90, 1470–1480.

Klein, J., Sato, A., Nagl, S., O'huigin, C., 1998. Molecular trans-species polymorphism. Ann. Rev. Ecol. Syst. 29, 1–21.

Leclerc, M.C., Barriel, V., Lecointre, G., de Reviers, B., 1998. Low divergence in rDNA ITS sequences among five species of *Fucus* (Phaeophyceae) suggests a very recent radiation. J. Mol. Evol. 46, 115–120.

Liston, A., Robinson, W.A., Oliphant, J.M., Alvarez Buylla, E.R., 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering plants. Syst. Bot. 21, 109–120.

Liu, J.-S., Schardl, C.L., 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol. Biol. 26, 775–778.

Loveless, M.D., Hamrick, J.L., 1984. Ecological determinants of genetic structure in plant populations. Ann. Rev. Ecol. Syst. 15, 65–95.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Mai, J.V., Coleman, A., 1997. The Internal Transcribed Spacer 2 exhibits a common secondary structure in green algae and flowering plants. J. Mol. Evol. 44, 258–271.

McDade, L.A., 1992. Hybrids and phylogenetic systematics. II. The impact of hybrids on cladistic analysis. Evolution 46, 1329–1346.

Martín, M.P., LaGreca, S., Schmitt, I., Lumbsch, H.T., 2003. Molecular phylogeny of *Diploschistes* inferred from ITS sequence data. Lichenologist 35, 27–32.

Mayer, M.S., Soltis, P.S., 1999. Intraspecific phylogeny analysis using ITS sequences: insights from studies of the *Streptanthus glandulosus* complex (Cruciferae). Syst. Bot. 24, 47–61.

Mayol, M., Rosselló, J.A., 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. Mol. Phylogenet. Evol. 19, 167–176.

Nieto Feliner, G., Fuertes Aguilar, J., Rosselló, J.A., 2001. Can extensive reticulation and concerted evolution result in a cladistically structured molecular data set? Cladistics 17, 301–312.

Nieto Feliner, G., Gutiérrez Larena, B., Fuertes Aguilar, J., 2004. Fine scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). Ann. Bot. 93, 189–200.

Ohta, T., Dover, G.A., 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. Proc. Natl. Acad. Sci. USA 80, 4079–4083.

Pedrosa-Harand, A., Souza deAlmeida, C.C., Mosiolek, M.M., Blair, M.W., Schweizer, D., Guerra, M., 2006. Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. Theoret. Appl. Genet. 112, 924–933.

Rauscher, J.T., Doyle, J.J., Brown, H.D., 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine* tomentella (Leguminosae) polyploid complex. Mol. Ecol. 11, 2691–2702.

Razafimandimbison, S.G., Kellogg, E.A., Bremer, B., 2004. Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleeae (Rubiaceae). Syst. Biol. 53, 177–192.

Ritland, C.E., Ritland, R.K., Straus, N.A., 1993. Variation in the ribosomal internal transcribed spacers (ITS1 and ITS2) among eight taxa of the Mimulus guttatus species complex. Mol. Biol. Evol. 10, 1273–1288.

Rosenberg, N.A., 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution 57, 1465–1477.

Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst. Biol. 41, 4–17.

Sang, T., Crawford, D.J., Stuessy, T.F., 1995. Documentation of reticulate evolution in peonies (Paeonia) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. Proc. Natl. Acad. Sci. USA 92, 6813–6817.

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37, 121–147.

Schlotterer, C., Tautz, D., 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr. Biol. 4, 777–783.

Senchina, D.S., Álvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A., Wendel, J.F., 2003. Rate variation among nuclear genes and the age of polyploidy in Gossypium. Mol. Biol. Evol. 20, 633–643.

Small, R.F., Wendel, J.F., 2002. Differential Evolutionary Dynamics of Duplicated Paralogous *Adh* Loci in Allotetraploid Cotton (*Gossypium*). Mol. Biol. Evol. 19, 597–607.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. Aust. Syst. Bot. 17, 145–170.

Soltis, P.S., Soltis, D.E., 1991. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. Syst. Bot. 16, 407–413.

Strand, A.E., Leebens-Mack, J., Milligan, B.G., 1997. Nuclear DNA-based markers for plant evolutionary biology. Mol. Ecol. 6, 113–118.

Taberlet, P., Gielly, L., Patou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Pl. Mol. Biol. 17, 1105–1109.

Van Oppen, M.J., Mieog, J.C., Sanchez, C.A., Fabricius, K.E., 2005. Diversity of algal endosymbionts (zooxanthellae) in octocorals: the roles of geography and host relationships. Mol. Ecol. 14, 2403–2417.

Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II. DNA Sequencing. Kluwer Academic Publishers, Boston, MA, pp. 265–296.

Wendel, J.F., Schnabel, A., Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. USA 92, 280–284.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, pp. 315–322.

Wolf, M., Achtziger, M., Schultz, J., Dandekar, T., Müller, T., 2005. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. RNA 11, 1616–1623.

Wu, Q.-X., Mueller, G.M., Lutzoni, F.M., Huang, Y.-Q., Guo, S.-Y., 2000. Phylogenetic and biogeographic relationships of eastern Asian and eastern North American disjunct *Suillus* species (Fungi) as inferred from ITS sequences of nuclear ribosomal RNA ITS sequences. Mol. Phylogenet. Evol. 17, 37–47.

Zardoya, R., Meyer, A., 1996. Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. Proc. Natl. Acad. Sci. USA 93, 5449–5454.

Zhang, D., Sang, T., 1999. Physical mapping of ribosomal RNA genes in paeonies (Paeonia, Paeoniaceae) by fluorescent in situ hybridization: implications for phylogeny and concerted evolution. Am. J. Bot. 86, 735–740.

Zhang, W., Wendel, J.F., Clark, L.G., 1997. Bamboozled again: Inadvertent isolation of fungal rDNA sequences from Bamboo (Poaceae: Bambusoideae). Mol. Phylogenet. Evol. 8, 205–217.

Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W., Wilson, A.C., 1980. Rapid duplication and loss of genes coding for the alphachains of hemoglobin. Proc. Natl. Acad. Sci. USA 77, 2158–2162.

Zuker, M., 2003. MFOLD web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.