



Ribosomal ITS sequences and plant phylogenetic inference

I. Álvarez and J.F. Wendel*

Department of Botany, Iowa State University, Ames, IA 50011, USA

Received 21 November 2002; revised 4 April 2003

Abstract

One of the most popular sequences for phylogenetic inference at the generic and infrageneric levels in plants is the internal transcribed spacer (ITS) region of the 18S–5.8S–26S nuclear ribosomal cistron. The prominence of this source of nuclear DNA sequence data is underscored by a survey of phylogenetic publications involving comparisons at the genus level or below, which reveals that of 244 papers published over the last five years, 66% included ITS sequence data. Perhaps even more striking is the fact that 34% of all published phylogenetic hypothesis have been based *exclusively* on ITS sequences. Notwithstanding the many important contributions of ITS sequence data to phylogenetic understanding and knowledge of genome relationships, a number of molecular genetic processes impact ITS sequences in ways that may mislead phylogenetic inference. These molecular genetic processes are reviewed here, drawing attention to both underlying mechanism and phylogenetic implications. Among the most prevalent complications for phylogenetic inference is the existence in many plant genomes of extensive sequence variation, arising from ancient or recent array duplication events, genomic harboring of pseudogenes in various states of decay, and/or incomplete intra- or inter-array homogenization. These phenomena separately and collectively create a network of paralogous sequence relationships potentially confounding accurate phylogenetic reconstruction. Homoplasy is shown to be higher in ITS than in other DNA sequence data sets, most likely because of orthology/paralogy conflation, compensatory base changes, problems in alignment due to indel accumulation, sequencing errors, or some combination of these phenomena. Despite the near-universal usage of ITS sequence data in plant phylogenetic studies, its complex and unpredictable evolutionary behavior reduce its utility for phylogenetic analysis. It is suggested that more robust insights are likely to emerge from the use of single-copy or low-copy nuclear genes.

© 2003 Elsevier Inc. All rights reserved.

1. Introduction

As testified by the launching of the journal in which these words appear, molecular sequence data have revolutionized phylogenetic analysis. Since the late 1980s but at a seemingly ever-increasing pace over the last decade, molecular phylogenetic hypotheses are being forwarded for nearly all groups of organisms. In plants, the majority of sequenced-based molecular phylogenetic studies, particularly in the early years, were based exclusively on genes and spacers from the plastid genome (Catalán et al., 1997; Clegg, 1993; Olmstead and Palmer, 1994; Olmstead and Reeves, 1995; Soltis et al., 1998), most notably *rbcL* (Chase et al., 1993). With increasing recognition of the dangers inherent in relying exclusively on what typically are uniparentally inherited sequences

for phylogenetic inference (Rieseberg and Soltis, 1991; Rieseberg and Wendel, 1993), widespread enthusiasm developed in the plant systematics community for the inclusion of sequence data from nuclear markers. For reasons enumerated below but accelerated by sociological factors, a single kind of nuclear locus experienced a meteoric rise in popularity, becoming almost a *sine qua non* for phylogenetic inference at generic and infrageneric levels in plants. Accordingly, this tool, the internal transcribed spacer (ITS) region of the 18S–5.8S–26S nuclear ribosomal cistron, now is extensively employed around the globe, having first been utilized scarcely a decade ago (Baldwin, 1992, 1993).

To illustrate just how popular ITS sequence-based phylogenetic analyses have become since the early review by Baldwin et al. (1995), we surveyed plant phylogenetic publications during the last five years in several of the most prominent systematics and evolution journals. This tabulation revealed that of 244 papers, fully

* Corresponding author. Fax: 1-515-294-1337.

E-mail address: jfw@iastate.edu (J.F. Wendel).

two-thirds (66%) involving comparisons at the genus level or below included ITS sequence data. Perhaps even more striking is the fact that more than one third (34%) of all published phylogenetic hypothesis have been based *exclusively on ITS sequences*.

Why has ITS-based phylogenetic analysis come to dominate plant molecular phylogenetic methodology? Apparently for at least two sets of reasons, one based on its presumed advantageous properties for phylogenetic inference, but the other apparently from a rather powerful bandwagon effect, whereby ITS utilization was accelerated in the community by usage itself, without much explicit challenge of the appropriateness of the tool. We will not delve into this latter set of sociological factors further here, but instead reiterate the long-noted (Baldwin et al., 1995) properties of ITS loci that were claimed to be advantageous for purposes of phylogenetic reconstruction:

- *Biparental inheritance*. Since 18S–26S rDNA arrays reside in the nuclear genome, ITS sequences are biparentally inherited, and are thus distinguished from the cpDNA loci in widespread use. Some of the earlier studies demonstrated how valuable this property is for revealing past cases of reticulation, hybrid speciation, and parentage of polyploids (Baldwin, 1992; Baldwin et al., 1995; Kim and Jansen, 1994; Rieseberg et al., 1990; Rieseberg and Soltis, 1991; Rieseberg and Wendel, 1993; Wendel et al., 1995).

- *Universality*. White et al. (1990) described a set of primers that was useful for amplifying ITS sequences from most plant and fungal phyla. This obviated the need for primer design or prior sequence knowledge, meaning that ITS sequence data could be more readily obtained than perhaps any other nuclear marker.

- *Simplicity*. Nuclear ribosomal genes are constituents of individual 18S–5.8S–26S repeats, which typically are in the size range of about 10 kb. These repeats are tandemly reiterated at one or more chromosomal loci per haploid complement. Because there are hundreds to thousands of nuclear rDNA repeats in plant genomes, they are more easily isolated than most low-copy nuclear loci, requiring little experimental expertise to successfully amplify. In plants, ITS sequences vary in length from approximately 500–700 bp in angiosperms (Baldwin et al., 1995) to 1500–3700 bp in some gymnosperms (Bobola et al., 1992; Germano and Klein, 1999; Liston et al., 1996; Maggini et al., 2000; Marrocco et al., 1996). Excluding gymnosperms, both the high copy number and the small size of the target DNA fragment facilitate ITS amplification by PCR, even permitting the use of ancient material, herbarium specimens, and samples other than from living material (much as with cpDNA).

- *Intragenomic uniformity*. It has long been recognized that multigene families in general and ITS sequences in particular may be subject to a phenomenon termed concerted evolution (Ainouche and Bayer, 1997;

Brochmann et al., 1996; Elder and Turner, 1995; Francke and Mummenhoff, 1999; Fuertes Aguilar et al., 1999a; Hillis et al., 1991; Roelofs et al., 1997; Schlotterer and Tautz, 1994; Wendel et al., 1995; Zimmer et al., 1980, among others). Concerted evolution occurs when sequence differences among reiterated copies in the genome, which should be accumulating their own distinct mutations, become homogenized to the same sequence type by mechanisms such as high-frequency unequal crossing over or gene conversion. When carried to completion, this process eliminates both sequence variation within genomes and potentially confounding variation, leaving only species- and clade-specific character-state changes to inform phylogenetic reconstruction efforts.

- *Intergenomic variability*. An early observation was that ITS sequence variation levels are suitable for phylogenetic inference at the specific, generic or even family levels (Baldwin, 1992; Baldwin et al., 1995). Baldwin and others noted that the variation at hierarchical levels at which most phylogeneticists work (generic and sub-generic) is attributable mostly to nucleotide polymorphisms, but that insertion–deletion polymorphisms (indels) are also common. They further reported divergence values that ranged from 0 to 39% in pairwise comparisons between taxa, with 5–59% of these being potentially phylogenetically informative (Baldwin et al., 1995).

- *Low functional constraint*. It was thought that since the ITS sequences are removed via splicing during transcript processing, they would be subject to reasonably mild functional constraints, which in turn would offer a preponderance of nucleotide sites that would evolve essentially neutrally. The functionality of ITS is related to specific cleavage of the primary transcript within ITS-1 and ITS-2 during maturation of the small subunit (SSU), 5.8S, and the large subunit (LSU) ribosomal RNAs (Hadjiolova et al., 1984, 1994; Musters et al., 1990; Nashimoto et al., 1988; Veldman et al., 1981; van Nues et al., 1994). Although this maturation and splicing process depends on the secondary structure of ITS, implying some degree of conservation at the sequence or at the structure level (Mai and Coleman, 1997), the presumption of limited functional constraint was widely adopted and further justified by observations of extensive nucleotide and length variation.

The foregoing list of properties constitute an impressive set of advantages for experimental design, and so it is not surprising that ITS-based phylogenetics rapidly grew in popularity. Given the prevalence of ITS sequence data in plant phylogenetic analyses, it seems prudent to pause and reflect upon these and other molecular evolutionary properties that are relevant to its utilization. We were motivated by the realization that the several advantages noted above may be counterbalanced by phenomena that are expected to confound

phylogenetic analyses, and that a deeper understanding of process might lead to enhanced insight into evolutionary history. Some of the relevant phenomena have been revealed or informed by the actual process of phylogenetic inference, where unexpected results or incongruent topologies were recovered. We review here molecular genetic processes that impact ITS sequence variation, drawing attention to the implications for phylogenetic inference.

2. Phenomena that impact the phylogenetic utility of ITS sequences

2.1. Multiple rDNA arrays

It has long been known that 18S–26S rDNA arrays and their RNA products constitute an essential component of eukaryotic NORs (nucleolus organizing regions). The number and distribution of NOR loci in eukaryotic genomes are variable, as is their size (Brown et al., 1993; Panzera et al., 1996; Pedersen and Linde-Laursen, 1994; Tartof and Dawid, 1976; Vanzela et al., 1998; Worton et al., 1988). Moreover, the number and genomic location of NOR arrays is evolutionary labile within families and genera (Dubcovsky and Dvořák, 1995; Murray, 2002; Schubert and Wobus, 1985; Thomas et al., 2001). In the Triticeae, for example, Dubcovsky and Dvořák (1995) showed that there are both major and minor rDNA arrays whose chromosomal location varies among species. It appears that in some cases major arrays have arisen via magnification of minor arrays while in other cases major arrays have evolutionarily been reduced to minor arrays through the loss of repeats. Other authors (e.g., Pedersen and Linde-Laursen, 1994) have also suggested that minor sites and rDNA pseudogenes originated from major sites.

The generality of the phenomenon of evolutionarily lability of rDNA loci has become firmly established in recent years as a consequence of the widespread application of fluorescent in situ hybridization (FISH) and genomic in situ hybridization (GISH) techniques to plant genomes (Adams et al., 2000; Anamthawat-Jónsson, 2001; Anamthawat-Jónsson and Bödvarsdóttir, 2001; Dagne et al., 2000; Hizume, 1994; Ma et al., 1997; Mishima et al., 2002; Murray, 2002; Nakayama et al., 2001; Raina and Rani, 2001; Shishido et al., 2000; Thomas et al., 2001; Vanzela et al., 1998; Xu and Earle, 1996; Zhang et al., 1995; Zoller et al., 2001). Minor rDNA sites and variation in site location and number are common observations, and this variation is only partly correlated with ploidy level. In some cases, more loci than expected from ploidy relationships may be observed. For example, Mishima et al. (2002) determined by FISH the 5S and 18S–5.8S–26S rDNA site number in ten polyploid species of *Sanguisorba*. All $2n = 14$ taxa had two rDNA sites

except for *S. annua*, which exhibited four or six 18S–5.8S–26S rDNA sites, as did the $2n = 28$ species. The authors explain these results by suggesting that the rDNA sites were duplicated during ancient polyploidization events in *Sanguisorba* followed by loss of 5S rDNA sites but not 18S–5.8S–26S rDNA sites in *S. annua*.

The consequences of the foregoing for phylogenetic analysis are profound. A fundamental requirement for historical inference based on nucleic acid or protein sequences is that the genes compared are orthologous as opposed to paralogous. Two genes are said to be orthologous if their relationship originated from organismal cladogenesis (see Doyle, 1992; Wendel and Doyle, 1998 for discussion). Orthologous sequences are appropriate for phylogenetic analysis, in that their history may reveal divergence events among species. If, however, there has been a history of gene (or array) duplication, the duplicated sequences within and between lineages are paralogous. When paralogous genes are unknowingly included in phylogenetic analysis to the exclusion of appropriate orthologous comparisons, the resulting gene tree will confound organismal divergence events with a tracking of the history of duplication. Hence, erroneous assessments of orthology and paralogy will lead to phylogenetic incongruence, as will sampling an unintended mixture of orthologous and paralogous sequences in a study (reviewed in Doyle, 1992; Wendel and Doyle, 1998).

The widespread occurrence of “nomadic” rDNA loci demonstrates that it should not be assumed that all ribosomal sequences isolated in a study are truly orthologous. Instead, it may be that a suite of orthologous and paralogous loci exist (e.g., Buckler et al., 1997; Hartmann et al., 2001; Ko and Jung, 2002; Mayol and Rosselló, 2001; O'Donnell and Cigelnik, 1997; among others), which may or may not be sampled by a single PCR experiment and which often are not fully homogenized via concerted evolutionary processes (see below). Buckler et al. (1997), for example, demonstrated the occurrence of potentially confounding rDNA paralogous as well as recombinants in *Gossypium*, *Nicotiana*, *Tripsacum*, *Zea* and Winteraceae. Similarly, in *Quercus* (Mayol and Rosselló, 2001), the divergence of paralogs in *Q. acutissima*, *Q. rubra*, and *Q. suber* produces long branches in a neighbor-joining tree and contradictory relationships when compared with those inferred from orthologous counterparts. In addition, some ribosomal arrays (5S or 18S–26S) may be evolutionary lost (e.g., Danna et al., 1996; Dubcovsky and Dvořák, 1995; Leggett and Markand, 1995; Li and Zhang, 2002; Mishima et al., 2002; Snowdon et al., 1997; Vaughan et al., 1993), leading to the problematic consequence that they are not available for experimental sampling, and as described more fully below, repeats at some loci may undergo pseudogene formation, creating additional and

possibly erratically sampled sequence variation. Hence, it is often a non-trivial exercise to distinguish orthologs from paralogs among rDNA loci, and because of this there are inherent risks in relying exclusively on rDNA sequences for phylogenetic inference.

Although it may be impractical in most cases to conduct FISH or GISH studies to document locus number and location in any given taxon, in some cases this knowledge may inform phylogenetic results based on ITS sequences (Adams et al., 2000; Anamthawat-Jónsson and Bödvarsdóttir, 2001; Ma et al., 1997; Mishima et al., 2002; Murray, 2002; Thomas et al., 2001; Vanzela et al., 1998; Zhang et al., 1995). For example, an understanding of array number and location may facilitate evaluation of alternative explanations for “missing sequences,” such as locus loss vs. homogenization (Hodkinson et al., 2002; Li and Zhang, 2002; Wendel et al., 1995).

2.2. Concerted evolution

Ribosomal genes in plants exist in hundreds to thousands of copies that are reiterated tandemly at one or more chromosomal locations. One of the more remarkable properties of rDNA genes is that the individual copies may appear to evolve more or less in unison. That is, instead of each gene copy acquiring unique sequence variation due to the evolutionary accumulation of mutations, all repeat copies within an array (or genome) may jointly share the same set of mutations. This uniformity arises from one or more processes of inter-genic sequence homogenization that collectively are referred to as concerted evolution (Arnheim, 1983; Elder and Turner, 1995; Zimmer et al., 1980). Arising from mechanisms such as unequal crossing over and high-frequency gene conversion, the classical concept of concerted evolution is one whereby inter-repeat sequence variation within an organism is reduced to a negligible level or is completely reduced, so that the multigene family contains largely or only one unique kind of sequence.

This classical concept has long been thought to characterize ribosomal arrays, and indeed this attribute has often been touted in phylogenetic applications (Baldwin, 1992; Baldwin et al., 1995; Elder and Turner, 1995; Hillis and Dixon, 1991). In principle, concerted evolutionary mechanisms would act to eliminate paralogous sequences, thereby facilitating inference of true homology among taxa and accurate phylogenetic reconstruction. In the absence of complete homogenization due to concerted evolution, multiple divergent rDNA copies will constitute orthologs and paralogs whose very presence complicates efforts to reconstruct phylogenetic history. As noted above, multiple rDNA arrays are quite common, arising both from well-known organismal processes such as hybridization and poly-

ploidization and by genomic processes like gene and chromosome segment duplication and various forms of homologous and non-homologous recombination. As discussed below, sequence homogenization resulting from concerted evolutionary mechanisms may not keep pace with variation generating processes at the organismal and genomic levels, and hence, in any given phylogenetic application it cannot be assumed that only one kind of ITS sequence type exists.

Even in cases where concerted evolution is complete or nearly complete among repeats within and between multiple rDNA arrays, it cannot be assumed that strict orthology has been maintained among sequences amplified among a set of taxa. One reason for this is that following a reticulation, introgression, or polyploidization event giving rise to the evolutionarily ephemeral coexistence of divergent ITS repeat types, the subsequent direction of sequence homogenization may be different in various descendant lineages, such that only one repeat type is observed per genome (Hillis et al., 1991; Wendel et al., 1995). In these instances the component of evolutionary history associated with the eliminated sequence type is lost, while the sequence types that remain may represent a mixture of paralogs and orthologs that are difficult to diagnose as such in the absence of other data that generate potentially informative incongruence (Wendel and Doyle, 1998). Given the prevalence of reticulation (Comes and Abbott, 2001; Fuertes Aguilar et al., 1999b; Rieseberg and Wendel, 1993; Rieseberg et al., 2000; Sang et al., 1995) and polyploidization in plants (Grant, 1981; Masterson, 1994; Stebbins, 1950), this becomes an important consideration.

A realization that has emerged in the last decade is that divergent rDNA copies may experience a variety of fates following their merger in a single genome as a consequence of a reticulation event (Wendel, 2000). One possibility is that the divergent copies are maintained, evolving independently without recombination or inter-array “contact.” In this case, ITS sequence data may prove informative with respect to documenting the historical hybridization or polyploidization event, providing information on both the maternal and paternal progenitor lineages. In the allopolyploids *Tragopogon mirus* and *T. miscellus*, formed from the diploid progenitor pairs *T. dubius*/*T. porrifolius* and *T. dubius*/*T. pratensis*, respectively, Soltis and Soltis (1991) and Soltis et al. (1995) showed that both parental rDNA types are retained. Baumel et al. (2001) observed the same pattern using RFLP analysis in the *Spartina anglica*, the allopolyploid derivative of *S. maritima* and *S. alterniflora*. Allopolyploidy in both *Tragopogon* and *Spartina* is rather recent (around 100 years old) but it is clear that divergent repeat types may be maintained over tens of millions of years. In the Winteraceae, for example, apparently ancient paralogs have been maintained to the

present (Suh et al., 1993). Regardless of relative antiquity, in cases where two or more repeat types do not become homogenized so that different ribotypes may be sampled, the divergent ITS sequences may prove highly informative with respect to historical relationships (e.g., Ritland et al., 1993 in *Mimulus guttatus*; Sang et al., 1995 in *Paeonia*; Ainouche and Bayer, 1997 in *Bromus lanceolatus* and *B. secalinus*; Campbell et al., 1997 in *Amelanchier*; Vargas et al., 1999 in *Hedera*; Wissemann, 1999, 2002b in *Rosa*).

A second possible outcome of the reunion of divergent sequences following hybridization is that two or more repeat types are maintained but undergo various degrees of recombination. This leads to chimeric ITS sequences that in phylogenetic analysis will behave erratically and resolve in phylogenetic positions basal to either parental lineage (McDade, 1992, 1995). Genic recombination is thought to be common in hybrids (Barkman and Simpson, 2002; Buckler et al., 1997; Campbell et al., 1997), and has been offered as the explanation for the recovery of chimeric ITS sequences. For example, in the hybrid *Dendrochilum acufiferum*, a survey of 14 ITS cloned sequences showed within-individual variation (Barkman and Simpson, 2002). Two of these sequences had ITS1 identical to one putative progenitor, *D. stachyodes*, whereas their ITS2 sequences were almost identical to the other putative progenitor, *D. grandiflorum*, suggesting genic recombination between the parental ribotypes. Another example of sequence recombination is for the fungal genus *Trichaptum*, where comparison of several ITS sequences from *T. abietinum* revealed that while all repeats examined share the same ITS2 sequence, they exhibit three different types of ITS1 sequence (Ko and Jung, 2002). The authors suggest that these chimeric sequences came from different loci residing on different chromosomes. They further notice that one ITS1 sequence type occurs only among Korean samples, perhaps indicating the geographic place and temporal component of the origin of this particular paralog. As discussed more fully in Section 2.3, below, recombination may occur not only between divergent functional repeat types, but also between functional repeats and non-functional pseudogenes in various states of decay. It is not difficult to envision how this process may mask phylogenetic signal, particularly when one considers that pseudogenization, reticulation, and recombination may be repeated episodically on an evolutionary timescale.

A third evolutionary possibility following hybridization is that one and only one repeat type comes to dominate the rDNA repeat population within a genome as a consequence of concerted evolutionary mechanisms. This repeat type may represent either an uncontaminated descendant of one of the parental sequences or a chimeric type resulting from intergenomic recombination. An example of sequence elimination following

allopolyploid formation in the genus *Gossypium* was provided by Wendel et al. (1995), who demonstrated that rDNA arrays are homogeneous or nearly so in both the progenitor diploid and descendant allopolyploids and, since these arrays occur in several chromosomal loci, repeats must have become homogenized both within and between arrays dispersed in different genomic locations. Because models of both diploid progenitors are extant, and because multiple sources of evidence independently substantiate the genomic composition of the allopolyploid species (Wendel and Cronn, 2003), it was possible to demonstrate that approximately 3800 rDNA repeats, each circa 10 kb in length, were “overwritten” by intergenomic cross-talk following the allopolyploidization event 1–2 million years ago (Senchina et al., 2003; Wendel and Cronn, 2003). Importantly, this study also demonstrated that the interlocus concerted evolutionary process may be bi-directional; that is, one parental repeat type may be lost in some descendant lineages, while the alternative repeat type may be lost in other lineages. Because only one repeat type remains in each descendant allopolyploid, phylogenetic analysis of ITS sequences by themselves would not reveal the history of genomic merger, and the topology obtained for the allopolyploids would be misleading.

Similar results have since been obtained in other systems (e.g., Brochmann et al., 1996; Ferguson et al., 1999; Franzke and Mummenhoff, 1999; Fuertes Aguilar et al., 1999a,b; Roelofs et al., 1997). Brochmann et al. (1996), for example, demonstrate the hybrid origin of the allopolyploid *Saxifraga osloensis*, which shows an additive pattern with respect to its two progenitors (*S. tridactylites* and *S. adscendens*) using RAPDs. When they analyzed ITS sequences of these three species, no additive pattern was observed in *S. osloensis*. They noted that the *S. osloensis* and *S. adscendens* ITS sequences are virtually identical but are divergent from those of *S. tridactylites*, and therefore conclude that the absence of both repeat types in the allopolyploid *S. osloensis* is due to biased concerted evolution towards the maternal progenitor (inferred from cpDNA evidence to be *S. adscendens*). A particularly interesting example of biased concerted evolution concerns the relatively recent and putatively hybrid origin of *Armeria villosa* subsp. *longiaristata* (Fuertes Aguilar et al., 1999a). Using extensive sampling of taxa and sequences, a geographical rather than purely taxonomic pattern was observed in the ITS tree, which was attributed to biased gene conversion following introgression events.

The three possible evolutionary fates discussed above, i.e., maintenance of both repeat types, generation of new repeat types, and loss of repeat types via concerted evolution, are not mutually exclusive. In *Bromus*, for example, Ainouche and Bayer (1997) demonstrated that the allopolyploid *B. hordeaceus* exhibits exclusively a single ITS sequence type, presumably due to interlocus

concerted evolution, whereas other allopolyploids (*B. lanceolatus* and *B. secalinus*) contain more than one repeat type reflecting the parental ITS sequences. In a detailed analysis of ITS polymorphisms in *Paeonia*, Sang et al. (1995) observed different degrees of additivity. In five hybrid species the ITS sequences show a clear polymorphism pattern indicative of the maintenance of both parental repeat types. In 10 other hybrids, however, ITS sequences combine nucleotides to different degrees in ITS1 from two putative groups of progenitors, while the ITS2 sequences show no variation and are identical to one of the two potential parental groups. The authors explain that in these 10 hybrids some of the variable sites, including ITS2, have been homogenized towards one progenitor (i.e.; loss of one repeat type via concerted evolution), while most ITS1 polymorphisms are maintained. This uneven pattern of homogenization across the ITS region implies the existence either of intergenic recombination between different repeat types (as in Ko and Jung, 2002) or a greater underlying complexity of rDNA organization than is realized. Another example of complex patterns of nrDNA evolution and hybridization is in *Rosa* (Wissemann, 1999, 2002b), where the allopolyploid *R. jundzillii* shows two types of ITS1 sequences from *R. gallica* plus one ITS1 type from *R. canina*, clarifying its hybridogenic origin.

A special situation arises when ITS sequences contain sub-repeats. In some gymnosperms there are tandem sub-repeats in ITS1 that vary in size (68–72 bp in *Larix* and *Pseudotsuga* to 215–237 bp in *Pinus pinea*) and number, leading to a threefold total length variation (Gernandt and Liston, 1999; Marrocco et al., 1996; Vining and Campbell, 1997; Vining et al., 1998). Gernandt et al. (2001) demonstrate within-individual heterogeneity of the ITS1 subrepeats, suggesting slow rates of concerted evolution. In the present context, the tandem repeats may promote recombination between paralogs and also between non-homologous subrepeats from orthologous copies, in the process producing chimeric ITS sequence types that could confound phylogenetic analysis.

Given the observation of great diversity among plants in the number of rDNA arrays and their sequence diversity, it is of interest to consider the factors that might influence the degree to which concerted evolution homogenizes sequences in any given case. At present we have little understanding of the life-history and genomic features that may quantitatively and qualitatively affect unequal crossing over and gene conversion, but presumably the genomic location and number of rDNA arrays should play an important role. It may be, for example, that when multiple arrays are located near the telomeres, interlocus recombination events are either promoted or are tolerated more than they would be if arrays were interstitial, in which case crossing over might lead to deleterious chromosomal recombinations

(Arnheim, 1983; Fulneček et al., 2002; Wendel et al., 1995). Generation time may also be an important factor, as with molecular evolutionary rates in general (Gaut, 1998), in that longer generation times may be associated with retarded rates of homogenization. In this respect we note that in some woody groups, there may be partial to little homogenization (Sang et al., 1995, in *Paeonia*; Campbell et al., 1997, in *Amelanchier* agamic complex; Wissemann, 1999, 2002b in *Rosa*) whereas in other woody groups interlocus homogenization may be complete or near complete (Wendel et al., 1995, in *Gossypium*). The time since reticulation may also be relevant, but even this might not be a particularly useful predictor of the extent of homogenization. As commented above, relatively young allopolyploids such as those in *Tragopogon* (Soltis and Soltis, 1991; Soltis et al., 1995) and *Spartina* (Baumel et al., 2001) maintain both parental repeat types. In artificial *Armeria* hybrids, however, the expected additive pattern was observed in the F1 generation, but surprisingly, in the F2 generation homogenization towards one parental type was detected (Fuentes Aguilar et al., 1999a). Thus, concerted evolution may act even in the very early stages of the hybrid formation (see also Franzke and Mummenhoff, 1999, in the genus *Cardamine*).

The many examples cited above demonstrate that concerted evolution cannot be assumed to operate instantaneously and completely to homogenize all rDNA repeats within a genome, and that the phenomenon may operate incompletely, at uneven evolutionary rates, and at the intralocus and interlocus levels. The corollary is that genomes may harbor a diversity of sequence types, reflecting not only the history of organismal recombination events but the vagaries of genomic interactions following the merger of divergent ITS sequence types. ITS sequences may be lost, recombined, or maintained. Clearly, insights into organismal reticulation and intergenic recombination can emerge only when ITS sequences are generated not from single PCR amplification reactions, which will mask any residual sequence diversity, but from exploration of sequence diversity within the amplicon pool following cloning and sequencing of multiple clones. Even when sampling is exceptionally thorough, however, one or more ITS repeat types and/or loci may become eliminated in one or more descendant taxa, thereby being lost as a possible source of historical evidence. This possibility, discussed further below, is tantamount to the loss of an ortholog, and hence phylogenetic comparisons for the affected taxa will by necessity include paralogous comparisons. Bailey et al. (2002; pers. comm., unpubl.) discuss in detail issues of assessing orthology and paralogy in nrDNA, based on gene-tree relationships, exhaustive sampling, and patterns of nucleotide diversification. In some cases this type of analysis may facilitate recovery of phylogenetically useful information, depending on

the taxonomic level and timing of duplication events. In any event, the paralogy problem increases with ploidy level, and may be exacerbated by both episodic intergenic recombination events and the phenomenon of pseudogene formation, discussed in the following section.

2.3. Pseudogenes

Because 18S–5.8S–26S repeats exist in hundreds to thousands of copies at one or more than one chromosomal location, and because the genomic location of rDNA arrays is evolutionarily labile (as discussed in Section 2.1, above), one might imagine that not all repeats remain functional on an evolutionary timescale, but that instead some copies degenerate into pseudogenes. Unless these sequences immediately become deleted from the genome or become so badly decayed that they no longer amplify, or unless the pseudogenes are “rescued” by interlocus concerted evolution, the consequence is that genomes become graveyards for dead and perhaps dying repeats of various ages. These pseudogenes may be harbored within the genome and perhaps evolve independently and at a different rate than their functional counterparts once they are released from functional constraints. Their existence, however, may pose serious challenges for phylogenetic analysis, in that the assumption of complete intergenic homogenization via concerted evolutionary processes may no longer be assumed, and instead, taxa included in phylogenetic studies may possess a multiplicity of sequence types.

Abundant evidence now substantiates the widespread existence of rDNA (including ITS) pseudogenes in plant genomes. Following the detailed and insightful discoveries in *Zea* by Buckler and Holtsford (1996a,b), pseudogenes have been detected from throughout the angiosperms (Buckler et al., 1997; Hartmann et al., 2001; Kita and Ito, 2000; Mayol and Rosselló, 2001; Muir et al., 2001; Yang et al., 1999). Many of these studies show and discuss the phylogenetic consequences of the existence of pseudogenes, and in addition, discuss some of the properties that can lead to their identification. The latter include the obvious indications arising from large indels, but also changes in predicted secondary structure, GC content, sequence divergence, and methylation patterns (Buckler and Holtsford, 1996a,b; Buckler et al., 1997). Using examples from several plant groups these latter authors illustrate that compared to functional genes, ITS pseudogenes have lowered secondary structure stability, an increase in AT content via deaminations, and a higher relative substitution rate in conserved regions. Bailey et al. (2002, unpubl.) discuss these methods and propose in addition a new tree-based approach using patterns of nucleotide diversification among putatively conserved and less constrained parts of the ITS region.

Although it may be possible to identify pseudogenes, particularly older ones, in the present context the most important issue is the effect of their existence on phylogenetic inference. Clearly, if multiple divergent sequences are sampled from within a single genome, the host taxon will appear at multiple places in a resulting phylogenetic tree (e.g., Buckler and Holtsford, 1996b; Buckler et al., 1997; Hartmann et al., 2001; Kita and Ito, 2000; Muir et al., 2001). In addition, the accelerated rates of substitution, indel accumulation, and antiquity may all contribute to the recovery of isolated or difficult to place long branches (e.g., Kita and Ito, 2000 for the genus *Aconitum*).

One of the most striking examples of the effects of ITS pseudogenes on phylogenetic analysis is that of Mayol and Rosselló (2001). They contrast the very different and well-supported results obtained by two different teams of researchers, namely Samuel et al. (1998) and Manos et al. (1999), who each studied the genus *Quercus*. When Mayol and Rosselló (2001) compared the sequences reported in the studies, the data set from the Samuel team exhibited high levels of variation in GC content, a great deal of length variation due to indels, higher rates of substitutions (even in conserved motifs), and lowered secondary structure stability. In addition, intraspecific divergence levels were very high (up to 42% in ITS2 for *Q. suber*, and up to 27% in ITS1 for *Q. rubra*), clearly demonstrating the absence of the sequence homogenization that is often assumed for rDNA repeats. The analyses by Samuel et al. (1998) and Manos et al. (1999) generated different phylogenies, apparently due to the existence of non-functional paralogs (pseudogenes) in the Samuel et al. data set. Mayol and Rosselló (2001) argue that the data could reflect divergent functional paralogs, but this would not account for the observed differences in GC content, the loss of secondary structure stability, and the higher rates of substitution in conserved motifs. Accordingly, the most convincing explanation for the phylogenetically disparate results is the inclusion of pseudogenes in one of the two analyses.

In some cases insights into the complexity of the orthology/paralogy relationships that are generated by an evolutionary history of duplication, divergence, and pseudogenization may emerge from thorough genomic sampling. A number of studies have explored the diversity of ITS sequences recovered from more exhaustive sampling of single genomes than typically is employed in phylogenetic studies. In a study of *Zea* (Buckler and Holtsford, 1996a,b), 78 ITS cloned sequences were analyzed, including thirteen clones amplified without DMSO, which led to the recovery of four sequences that were identified as pseudogenes because of their lower than expected GC content (62–65% vs. 70–73%). In contrast, all the clones amplified with DMSO had GC content within the expected range and no

pseudogenes were recovered. In *Quercus*, Muir et al. (2001) sequenced 70 clones for diploid *Quercus petraea* and *Q. robur*. This sampling led to the recovery of three classes, or families, of sequences in both species. Inspection of the sequences indicated that only one of the three sequence classes is likely to be functional, with the others corresponding to pseudogenes. In *Leucaena*, Hughes et al. (2002) analyzed a total of 87 sequences from 65 accessions, identifying 26 of them as pseudogenes. These examples from *Zea*, *Quercus*, and *Leucaena* demonstrate that individual plant genomes may harbor abundant sequence variation for ITS sequences, reflecting a long history of duplication, incomplete homogenization, and pseudogenization, thereby generating complex patterns of paralogy. Given this observation, it would seem to be incumbent upon investigators to explore this possibility prior to using ITS sequences for phylogenetic inference. To do otherwise clearly can lead to erroneous phylogenies, from orthology–paralogy conflation, unexpected isolation of pseudogenes, or both.

An additional complication that initially was revealed by in-depth sampling of sequences is that pseudogenes may recombine with functional repeats, leading to the presence in genomes of chimeric sequences. One may envision that such sequences could be functional or remain as pseudogenes, depending on the degree of decay in the pseudogene and the portion of the sequence that experiences recombination or gene conversion with a functional repeat. In addition, one might envision cases where some of the rescued products carry embedded vestiges of pseudogenized paralogs. A possible example of this phenomenon is discussed in Buckler et al. (1997), who argued that incongruence between ITS1 and ITS2 from the same sequence was due to genic recombination. One additional possibility is that of PCR-mediated, as opposed to in vivo recombination, whereby paralogs, perhaps involving but not necessarily involving pseudogenes, recombine due to template switching or incomplete template extension during PCR. This phenomenon has recently been demonstrated for a variety of genes by Cronn et al. (2002a). Regardless of the mode of origin, recombinants clearly are problematic for phylogenetic analysis.

2.4. Secondary structure and compensatory base changes

An assumption of phylogenetic analysis of nucleotide sequences is that each position is independent of other positions. While this assumption probably is never fully met in a strict sense, in many instances it may not be too seriously violated. For some types of sequences, however, structural considerations predict that the assumption of independence is problematic. Perhaps best known in this respect are genes that encode ribosomal RNAs, which are single-stranded but have a secondary

structure that includes stemmed regions containing base pairs. Because rRNA is divided into domains where bases are either paired or unpaired, it is likely that different evolutionary constraints operate in each case. As noted in the introduction, ITS sequences are subject to evolutionary constraints related to maintenance of specific secondary structures that provide functionality (Liu and Schardl, 1994; Mai and Coleman, 1997; Schlötterer et al., 1994; Torres et al., 1990). High levels of GC content provide stability of the DNA and RNA secondary structures and are associated with the formation of stem–loop and more complex secondary structures. The high GC content of ITS sequences, particularly in conserved regions, and the stem–loop structural features suggest that compensatory base changes are frequent among different nucleotide sites (Mai and Coleman, 1997). It is also clear that there are differences in the shape and number of stem–loop structures among different organisms (Venkateswarlu and Nazar, 1991; Baldwin, 1992), complicating any effort to identify a priori where compensatory changes might be expected to occur.

This feature of the ITS sequences has important consequences. Stemmed bases are subject to selection for compensatory mutations so that base-pairing is maintained, although phylogenetic evidence indicates that an unpaired base may persist for some time following a mutation, even if it does ultimately become compensated (Gatesy et al., 1994). Because mutations typically become compensated, however, compensatory base changes may lead to homoplasy, thereby obscuring phylogenetic signal. Some thought has been given to differential weighting of paired and unpaired nucleotide positions for purposes of phylogenetic analysis (e.g., Hillis and Dixon, 1991; Soltis et al., 1998; Wheeler and Honeycutt, 1988). Dixon and Hillis (1993) studied 28S RNA genes from selected vertebrates, and recommend reducing the weight accorded stem characters by 20% relative to loop characters, whereas Springer et al. (1995), in a study of 12S rRNA gene sequences from mammals, suggest a more extreme weighting of stemmed positions (circa 40% down-weighting). Though these weighting recommendations differ quantitatively, they underscore the possibility of non-independence of nucleotide positions in sequence data. In principle, one might attempt to infer the secondary structures of the ITS sequences included in a study, and from these down-weight during phylogenetic analysis the sites that are deduced to be subject to compensatory changes. While this approach may seem reasonable, many alternative secondary ITS structures may have similar free energy estimates, and some compensatory changes cannot be detected because they depend on nucleotides other than from directly base-paired positions (Baldwin et al., 1995). Dixon and Hillis (1993) further show how alternative weighting can modify the phylogenetic

topology. At any rate, the structural component inherent in the ITS molecule raise the possibility of non-independence, increased homoplasy, and possible discordance with phylogenetic estimates based on other data.

2.5. Alignment, accuracy, and rooting

One methodological advantage of using protein-encoding genes for phylogenetic analysis is that the researcher is provided an “alignment and error check” every three nucleotides, due to the partitioning of protein-coding sequences into codons (at least for exons). This advantage facilitates interpretation of trace files and autoradiograms, particularly regarding decisions on sequence accuracy and possible single nucleotide indels in regions of compression; given triplets and considerations of frame, these decisions become straightforward. Because ITS sequences are not protein-encoding, these natural alignment guides are unavailable, and hence the process of scoring sequence data becomes similar to that of cpDNA spacer and intron regions. With ITS, however, the sequence scoring problem is magnified by its propensity to accumulate indels, as well as its generally high GC content, the latter causing more compression than with the typically AT-rich nuclear introns. Indels are everpresent in ITS data sets, many occurring at problematic single nucleotide repeats and possibly arising from DNA replication slippage (Hancock and Vogler, 2000; Levinson and Gutman, 1987), but longer indels also are often observed.

These interrelated issues of sequencing accuracy, high GC content, and indel accumulation individually and collectively impact the process of ITS sequence alignment. Thus, assembly of a correct and aligned ITS data set may be more challenging than with other sequences such as nuclear protein-encoding genes. Many sequencing errors will be relatively immaterial, in that they will appear as autapomorphies, but others may lead to spuriously scored synapomorphies as well as erroneously inferred indels. All of these fates result in increased measures of homoplasy. In reading published papers that use ITS it is common to find authors discussing alignment ambiguities and the various treatments of resulting gaps during phylogenetic analysis, and it is not at all uncommon for entire stretches of nucleotides to be excluded from the data sets due to alignment issues. Because sequence accuracy, alignment and gap treatments all impact phylogenetic results (Baum et al., 1994, 1998; Davis et al., 1998; Kelchner and Clark, 1997; Kim and Jansen, 1994; Raymunde et al., 2002), this issue becomes an important consideration. Some useful approaches have been proposed for gap treatment when alignments are unambiguous (Simmons and Ochoterena, 2000; Simmons et al., 2001) that might lead to decreased homoplasy and better phylogenetic resolution.

ITS sequences evolve relatively quickly (Baldwin et al., 1995; Cronn et al., 2002b; Small et al., 1998), both with respect to indel accumulation and nucleotide substitutions. As a consequence, it is often the case that divergence is too great to permit confident alignment between ingroup and outgroup sequences (Alvarez Fernández et al., 2001; Ashworth, 2000; Kim and Jansen, 1996; Möller and Cronk, 1997; Ray, 1995; Suh et al., 1993). This issue also arises with cpDNA spacers and introns (e.g., van Ham et al., 1994, in *Crassulaceae* using the *trnF*(GAA) intergenic spacer; Kelchner and Clark, 1997, in *Bambusoideae* using the *rlp16* intron), leading to situations where phylogenetic analysis needs to be performed without outgroup rooting, or perhaps more problematically, where alignments become forced simply to accommodate one or more outgroups. Thus, the opportunity to generate added confidence in a particular topology through observations like ingroup stability under various outgroup choices is unavailable with ITS sequence data. As an alternative methodological choice, protein-encoded genes that have introns offer a nice compromise, because exons provide convenient alignment guides as well as triplets, whereas introns offer a higher level of sequence variation (e.g., Bailey and Doyle, 1999; Bortiri et al., 2002; Cronn et al., 2002a,b). The limitation of phylogenetic breadth of ITS relative to other genes has been shown in a number of publications (e.g., Bailey and Doyle, 1999; Bortiri et al., 2002; Cronn et al., 2002b; Rossetto et al., 2002; Seelanan et al., 1997).

2.6. Universality and contamination

Universality is a convenient feature of ribosomal sequences, and this represents one of the primary reasons why systematists, especially without training in molecular biology, choose ITS for molecular phylogenetic analyses. We note with some irony, therefore, that the same universality that accords ITS sequences their broad utility also has generated some wildly unexpected phylogenetic results. Specifically, because of the near-universality of the amplification primers commonly used in PCR for ITS sequences, any contamination of the reaction cocktail with tissue from a different plant (or even a fungus) could lead to preferential amplification of the contaminant. Many laboratories include in their local lore stories of “PCR nightmares” involving unexpected amplification products, and some of these entail erroneous amplification of ITS sequences. Much of this contamination can readily be avoided by appropriate precautions to ensure that the only template DNA in the reaction tube is from the desired organism. In addition, it is good scientific practice to replicate any specific phylogenetic result that is particularly surprising (although in many groups without a priori phylogenetic expectations, such results may go unnoticed). Remarkably, however, even the most stringent laboratory

hygiene standards and replication practices are not sufficient to guard against unintended amplification of fungal epiphytes or endophytes (Camacho et al., 1997; Liston and Alvarez-Buylla, 1995; Zhang et al., 1997). For example, Camacho et al. (1997) found that several *Picea* ITS1 sequences used for phylogenetic purposes by Smith and Klein (1994) might represent endophytic fungi. In fact, the supposed *Picea engelmannii* sequence exhibited 97.6% identity to the endophytic fungus *Hormonema dematioides*, in which case contamination clearly is implicated. In an attempt to amplify ITS sequences from several genera of woody bamboos, Zhang et al. (1997) obtained repeatedly and under a variety of PCR conditions sequences that were more similar to unidentified fungi, most likely basidiomycetes. Additionally, angiosperm ITS sequences were obtained when fresh leaves were sterilized prior to DNA extraction, suggesting that the fungal contaminants were epiphyllous, as also shown by scanning electron microscopy. At present, fungal contamination is more of a nuisance than a phylogenetic misdirection, as it would be easily detected in database searches using the “odd” sequences as the queries. In these cases, however, it may be necessary to redesign amplification primers or modify amplification conditions so that the angiosperm sequences of choice are preferentially amplified.

2.7. Homoplasmy

One of the consequences of alignment and sequencing problems, as well as the possibilities of compensatory base changes, paralogy, pseudogenes and lack of complete concerted evolution, is that homoplasmy is increased in phylogenetic data sets. Because the aforementioned problems are more acute with ITS sequences than for most other genetic loci, the ironic possibility is raised where the single most popular source of plant molecular phylogenetic data is also the most homoplasious. To evaluate and quantify this possibility, we tabulated the two most widely used measures of homoplasmy, i.e., the consistency index (CI) (Kluge and Farris, 1969) and the retention index (RI) (Farris, 1989), for 22 recent studies wherein both ITS and other loci were sequenced for the same set of samples (so that homoplasmy estimates could be fairly compared). While this list is not intended to be exhaustive, it is unbiased in the sense that the only criterion used is selecting papers to include was that they employed multiple sources of molecular sequence data, one of which was ITS. To provide a little more context for the meaning of the numbers we also tabulated the number of parsimony-informative sites. Clearly, the amount of homoplasmy observed in any given study will be highly correlated with the number of taxa included as well as on their degree of genetic divergence. Although we made no effort to statistically stratify our sampling in this respect or to use regression methods, Table 1 shows

that ITS exhibits a higher percentages of parsimony-informative sites than other markers at all levels of divergence.

As shown by the data in Table 1, ITS data yield lower CI and RI values than other loci in the vast majority of studies, indicating a higher level of homoplasmy than obtained with other markers. Homoplasmy does not necessarily affect the accuracy of a phylogeny if the homoplasious characters are not informative. But, when homoplasmy is distributed such that it resolves as synapomorphy, misleading phylogenetic results may arise. Although we did not study the distribution of homoplasmy in the studies in Table 1, in principle one might evaluate this possibility by examining incongruence between tree topologies obtained with ITS and other markers, although clearly there are many other factors that cause incongruence (Wendel and Doyle, 1998). In Table 1, the vast majority of the markers compared to ITS are cpDNA sequences, and thus any observed incongruence might reflect the uniparental origin of cpDNA vs. the biparental origin of ITS, differential rates of evolution, ITS homoplasmy, or other factors. In all 10 examples from Table 1 (Cronn et al., 2002a,b; Miller et al., 1999) where ITS is compared to other nuclear genes (excluding ribosomal DNA), levels of homoplasmy (CI) are higher in the ITS datasets. Hence, at least part of the higher percentage of variable and parsimony-informative sites observed for ITS must arise from homoplasmy rather than useful variation. Only in one of 34 comparisons (3%) was the CI for ITS judged to be higher than the other molecular marker used (i.e., between ITS and the *trnL-trnF* spacer). In one other case, the amount of homoplasmy is equivalent for ITS and ETS (rDNA external transcribed spacer), which is not unexpected given that these two spacers are part of the same molecule.

In most of the studies included in Table 1, only one or two additional markers are used. Cronn et al. (2002b), however, compared ITS against 12 different genes, both nuclear and chloroplast, yielding the same result, namely, that ITS exhibited more homoplasmy than most other gene sequences. It is noteworthy that in Cronn et al. (2002b) the exception to the generalization is for the cpDNA spacer *trnT-trnL* that share the properties of absence of codons, unbalanced nucleotide compositions, and a high frequency of indel accumulation. This suggests properties that in general should be avoided in selecting genes for phylogenetic analysis.

In an effort to quantify the degree to which homoplasmy is higher for ITS than other markers in plant phylogenetic studies, we used the data in Table 1 to calculate the mean percentage decrease (rarely, increase) in CI and RI for all pairwise comparisons between genes in each study and averaged across these values across studies. The estimates obtained (14.6 and 12.6% for CI and RI, respectively) illustrate unambiguously the

Table 1
Relative utility of internal transcribed spacer sequences in phylogenetic studies

Study	Sequence	Total number of characters	Parsimony informative characters (%)	Consistency index	Retention index	
Baum et al. (1998) (<i>Adansonia</i>)	ITS	787	10.8	0.91	0.86	
	<i>rpl16</i>	1347	0.8	0.97	0.97	
Bailey and Doyle (1999) (<i>Sphaerocardamum</i>)	ITS	639	14.2	0.76	0.79	
	<i>trnL</i> intron	556	2.3	0.81	0.87	
Miller et al. (1999) (<i>Ipomoea</i>)	ITS	573	28.5	0.46	0.65	
	<i>waxy</i>	651	13.2	0.60	0.77	
Molvray et al. (1999) (<i>Korthalsella</i>)	ITS	717	21.8	0.67	0.87	
	<i>trnL-F</i>	978	13.6	0.64	0.83	
Seelanan et al. (1999) (<i>Gossypium</i>)	ITS	688	5.7	0.82	–	
	<i>rpl16</i>	1138	0.8	0.92	–	
Clevinger and Panero (2000) (<i>Silphium</i>)	ITS	496	39.9	0.60	0.71	
	ETS	472	30.9	0.59	0.82	
Erdogan and Mehlenbacher (2000) (<i>Corylus</i>)	ITS	666	3.3	0.92	0.94	
	<i>matK</i>	1231	0.8	1	1	
Whitten et al. (2000) (Maxillarieae)	ITS	910	49	0.41	0.61	
	<i>matK</i>	1379	19	0.56	0.69	
	<i>trnL-F</i>	1282	22	0.58	0.64	
Álvarez Fernández et al. (2001) (<i>Doronicum</i>)	ITS	496	11.1	0.61	0.72	
	<i>trnL-trnF</i>	397	0.8	0.90	0.94	
Bortiri et al. (2001) (<i>Prunus</i>)	ITS	759	22.9	0.57	0.71	
	<i>trnL-trnF</i>	563	11.7	0.73	0.83	
Gravendeel et al. (2001) (<i>Coelogyne</i>)	ITS	720	16.3	0.58	0.53	
	<i>matK</i>	1921	13.6	0.74	0.78	
Roalson et al. (2001) (Cariceae)	ITS	691	30.5	0.34	0.63	
	<i>trnT-L-F</i>	2050	14.4	0.66	0.77	
Zomlefer et al. (2001) (Melanthieae)	ITS	860	35.9	0.78	0.89	
	<i>trnL-F</i>	1052	11.5	0.85	0.91	
Barber et al. (2002) (<i>Sideritis</i>)	ITS	645	19.8	0.79	0.87	
	<i>trnT-L</i>	1279	5.9	0.81	0.93	
Beardsley and Olmstead (2002) (<i>Mimulus</i>)	ITS + ETS	1187	30.8	0.54	0.66	
	<i>trnL-F</i>	1024	17.7	0.79	0.82	
Compton and Culham (2002) (Ranunculaceae)	ITS	715	15.5	0.78	0.59	
	<i>trnL-F</i>	1094	6.5	0.85	0.70	
Cronn et al. (2002b) (<i>Gossypium</i>)	ITS	695	4.1	0.76	0.76	
	A1341	681	0.6	0.95	0.95	
	A1713	595	0.5	0.95	0.95	
	A1751	858	0.3	1	1	
	AdhA	953	0.4	0.80	0.79	
	CesA1	1086	0.4	0.93	0.92	
	CesA1b	1177	0.5	0.96	0.96	
	G1121	748	0.6	1	1	
	G1134	547	1	0.95	0.95	
	G1262	934	0.2	0.95	0.95	
	<i>matK</i>	2553	0.8	0.92	0.92	
	<i>ndhF</i>	2064	0.5	0.93	0.93	
	<i>trnT-trnL</i>	1596	1	0.78	0.78	
	Kress et al. (2002) (Zingiberaceae)	ITS	814	39.7	0.37	0.72
		<i>matK</i>	3249	14.8	0.63	0.89
Mort et al. (2002) (Crassulaceae)	ITS	684	26	0.69	0.89	
	<i>matK</i> + <i>psbA-trnH</i> + <i>trnL-trnF</i>	2467	5.3	0.71	0.82	
Smedmark and Eriksson (2002) (<i>Geum</i>)	ITS	678	28.6	0.54	0.67	
	<i>trnL-trnF</i>	1252	12.5	0.76	0.90	
Wagstaff and Wege (2002) (Stylidiaceae)	ITS	782	30.6	0.68	0.81	
	<i>rbcL</i>	1402	7.1	0.76	0.87	
Zimmer et al. (2002) (Gesnerioideae)	ITS	731	44.1	0.44	0.64	
	<i>trnL-F</i> + <i>trnT-L</i>	1928	10.9	0.78	0.82	

higher levels of homoplasy in ITS datasets vs. other markers. Although homoplasy itself does not always negatively impact phylogenetic analyses, high levels of homoplasy increase the risk of an incorrect inference. Thus, despite the near universal usage of ITS and its potential to provide informative phylogenetic signal, the utility of ITS sequences, as well as other ribosomal DNA sequences, introns, and spacers, is limited relative to other options available (e.g., Cronn et al., 2002b).

2.8. Technical considerations

Given the foregoing account of the myriad forces and phenomena that shape the suite of ITS sequences that may exist in a given plant genome, it may be useful to highlight some technical considerations that may enhance phylogenetic insight. Some of these have been alluded to elsewhere, such as the desirability of having access to FISH or GISH data to reveal array number and chromosomal distribution. While this type of analysis would almost always be informative, in most applications it is impractical. However, a great deal can be learned from routine application of methods designed to explore the possibility of multiple divergent ITS sequences in the taxa under study. As discussed above, these multiple sequence types may arise from ancient or recent array duplication events, genomic harboring of pseudogenes in various states of decay, or incomplete intra- or inter-array homogenization. The net effect of this multiplicity of forms, however, is the same regardless of the mode of genesis, in that its existence complicates phylogenetic inference due to orthology/paralogy conflation.

Fortunately, progress in understanding possible sequence complexity may be achieved irrespective of the underlying cause. In general, this simply entails cognizance of the possibility of sequence variation and some simple experimental modifications. Given what we know about rDNA arrays, it is clear that direct sequencing of single PCR may yield misleading results because of numerical inequality among the spectrum of repeat types that exist within the genome and the preferential amplification of a single repeat type (or quantitative masking of minority repeats on autoradiograms or electropherograms). The phenomenon is a general one, and is referred to as PCR bias or PCR selection (Wagner et al., 1994). Biased PCR occurs when certain members of a gene family become favored for amplification by the reaction conditions. Thus, the same products from the selected repeat will be obtained in separate reactions under the same conditions. Also, amplification may be directed by PCR drift (Wagner et al., 1994), i.e., random factors affecting the selection of one type of repeat during the first amplification cycles, in which case different products may be obtained in separate reactions.

It is known that PCR specificity can be improved through the addition of additives to the amplification cocktail or through other modifications of PCR methodology (Bechmann et al., 1990; Innis, 1990; Varadaraj and Skinner, 1994; Winship, 1989; Zhang et al., 1994). Analogs to dGTP such as dITP or 7-deaza-dGTP may reveal cases of bias due to GC compression, and dimethyl sulfoxide (DMSO) may relax secondary structure during amplification, which can lead to an increase in the sequence diversity revealed. Different sequence types may be recovered and sequencing may be improved with such modifications (Buckler and Holtsford, 1996b; Buckler et al., 1997; Kim and Jansen, 1994). One remarkable example is from work in *Zea* (Buckler and Holtsford, 1996b; Buckler et al., 1997), where the addition of DMSO or 7-deaza-dGTP yielded different results; no pseudogenes were recovered using DMSO, whereas both functional and non-functional repeats were obtained with 7-deaza-dGTP. Finally, we draw attention to the results obtained in two different studies of *Quercus* (Samuel et al., 1998 vs. Manos et al., 1999), as reviewed by Mayol and Rosselló (2001). One of the methodological differences between the two teams was the use of universal primers vs. specific angiosperm primers, respectively, leading to the amplification of some pseudogenes by Samuel's team, while functional repeats were obtained by Manos' team.

A final strategy for discovering infra-genomic ITS diversity is exemplified by a recent study in the *Glycine tomentella* allopolyploid complex (Rauscher et al., 2002). In this study repeat-specific PCR primers were used in mismatch amplifications to successfully recover low-copy ITS repeats that otherwise could not be obtained by direct sequencing or cloning. This strategy may prove effective in obtaining rare repeat types when a previous hypotheses of hybridization exists and when sequences from the putative parents are available.

At the minimum, the realization of the possibility of ITS sequence complexity calls for cloning of amplification products obtained under a variety of PCR conditions, followed by sampling of clones to assess sequence diversity. Although this exploration is more labor-intensive, costly, and time-consuming than the traditional approach of sequencing a single amplification pool, the latter would appear to be a risky phylogenetic protocol. As an alternative, a strategy of sequence variation exploration and detection may lead not only to increased phylogenetic accuracy but an enhanced understanding of genome history in the group under study.

3. Conclusions

Our purpose in writing this review was to illuminate and bring attention to the many molecular evolutionary

and organism-level processes that may impact sequence variation for ITS repeats in plants, and thereby hopefully contribute to a more informed utilization in phylogenetic analyses. Although we have discussed a rather lengthy list of phenomena that may generate intragenomic sequence variation, to a certain extent the problem this creates for phylogenetic analysis is the same regardless of the underlying cause: incorrect or at least unsubstantiated assumptions of orthology. To the extent that this is construed as a negative message, it certainly is an important one that has not sufficiently permeated the systematics community, as testified by the dominance of ITS sequence data as a nuclear marker in present phylogenetic applications and the fact that most routine applications fail to adequately explore the possibility of multiple, divergent repeat types. ITS sequence data have and may continue to provide insights into phylogenetic history, polyploid ancestry, genome relationships, historical introgression, and other evolutionary questions (Bailey et al., 2002; Wissemann, 2002a), but the data generated will have the most value and the most lasting value only if issues such as those raised in this review are experimentally addressed.

As an alternative to the routine use of ITS sequences as the molecule of choice for phylogenetic analysis, we urge routine utilization of single-copy nuclear genes. Single-copy nuclear genes are becoming increasingly used for phylogenetic analysis, are biparentally inherited, with apparently rare exception are not subject to concerted evolution (Cronn et al., 1999; Senchina et al., 2003), and contain codons to limit alignment ambiguity and facilitate homologous comparisons (Bailey and Doyle, 1999; Bortiri et al., 2002; Doyle and Doyle, 1999; Sang, 2002). In addition, single-copy nuclear genes have lower homoplasy relative to ITS data (Table 1), and provide, for all intents and purposes, a nearly limitless supply of characters. For these reasons, we recommend that ITS no longer be routinely utilized for phylogenetic analysis, opting instead for using several or more different single-copy nuclear loci (e.g., Cronn et al., 2002b).

Inferring the history of life is a noble but challenging enterprise fraught with methodological and analytical unknowns and difficulties. In this era, however, it seems that one should attempt to select phylogenetic tools with properties that are more likely to provide historically meaningful signal than ITS. Publicly available databases are bursting at the seams with gene sequences from plants, which are readily available for purposes of experimental design. Because the genomes of *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000) and *Oryza sativa* (Goff et al., 2002; Yu et al., 2002) have been sequenced, and because hundreds of thousands of ESTs (Expressed Sequence Tags) have been deposited in GenBank representing scores of other plant species, it is now a relatively straightforward process to download homologous sequence for alignment, primer design, and

use in phylogenetic analysis (Small et al., 2003). The use of multiple nuclear genes is nearly certain to vastly improve phylogenetic understanding in years to come as these many diverse sources of evidence become increasingly employed.

Acknowledgments

Financial support was provided by the National Science Foundation and the Spanish Ministry of Education, Culture and Sports.

References

- Adams, S.P., Leitch, I.J., Bennet, M.D., Chase, M.W., Leitch, A.R., 2000. Ribosomal DNA evolution and phylogeny in *Aloe* (Asphodelaceae). *Am. J. Bot.* 87, 1578–1583.
- Ainouche, M.L., Bayer, R.J., 1997. On the origins of the tetraploid *Bromus* species (section *Bromus*, Poaceae): Insights from the internal transcribed spacer sequences of nuclear ribosomal DNA. *Genome* 40, 730–743.
- Álvarez Fernández, I., Fuertes Aguilar, J., Panero, J.L., Nieto Feliner, G., 2001. A phylogenetic analysis of *Doronicum* (Asteraceae, Senecioneae) based on morphological, nuclear ribosomal (ITS), and chloroplast (*trnL-F*) evidence. *Mol. Phylogenet. Evol.* 20, 41–64.
- Anamthawat-Jónsson, K., 2001. Molecular cytogenetics of introgressive hybridization in plants. *Methods Cell Sci.* 23, 139–148.
- Anamthawat-Jónsson, K., Bödvarsdóttir, S.K., 2001. Genomic and genetic relationships among species of *Leymus* (Poaceae: Triticeae) inferred from 18S–26S ribosomal genes. *Am. J. Bot.* 88, 553–559.
- 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.
- Ashworth, V.E.T.M., 2000. Phylogenetic relationships in Phoradendreae (Viscaceae) inferred from three regions of the nuclear ribosomal cistron. I. Major lineages and paraphyly of *Phoradendron*. *Syst. Bot.* 25, 349–370.
- Bailey, C.D., Doyle, J.J., 1999. Potential phylogenetic utility of the low-copy nuclear gene *pistillata* in dicotyledonous plants: Comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. *Mol. Phylogenet. Evol.* 13, 20–30.
- Bailey, C.D., Hughes, C.E., Harris, S.A., Carr, T.G., 2002. Characterization of angiosperm nrDNA polymorphism: Paralogy and pseudogenes. Unpublished manuscript.
- Baldwin, B.G., 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. *Mol. Phylogenet. Evol.* 1, 3–16.
- Baldwin, B.G., 1993. Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA: Chromosomal and morphological evolution reexamined. *Am. J. Bot.* 80, 222–238.
- 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. Missouri Bot. Gard.* 82, 247–277.
- Barber, J.C., Francisco-Ortega, J., Santos-Guerra, A., Turner, K.G., Jansen, R.K., 2002. Origin of Macaronesian *Sideritis* L. (Lamiaceae: Lamiaceae) inferred from nuclear and chloroplast sequence datasets. *Mol. Phylogenet. Evol.* 23, 293–306.

- Barkman, T.J., Simpson, B.B., 2002. Hybrid origin and parentage of *Dendrochilum acufiferum* (Orchidaceae) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. *Syst. Bot.* 27, 209–220.
- Baum, D.A., Sytsma, K.J., Hoch, P.C., 1994. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. *Syst. Bot.* 19, 363–388.
- Baum, D.A., Small, R.L., Wendel, J.F., 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Syst. Biol.* 47, 181–207.
- Baumel, A., Ainouche, M.L., Lévassieur, J.E., 2001. Molecular investigations in populations of *Spartina anglica* C.E. Hubbard (Poaceae) invading coastal Brittany (France). *Mol. Ecol.* 10, 1689–1701.
- Beardsley, P.M., Olmstead, R.G., 2002. Redefining Phrymaceae: The placement of *Mimulus*, tribe Mimuleae, and *Phryma*. *Am. J. Bot.* 89, 1093–1102.
- Beckmann, B., Luke, W., Hunsmann, G., 1990. Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res.* 18, 1309.
- Bobola, M.S., Smith, D.E., Klein, A.S., 1992. Five major nuclear ribosomal repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. mariana*. *Mol. Biol. Evol.* 9, 125–137.
- Bortiri, E., Oh, S.-H., Jiang, J., Baggett, S., Granger, A., Weeks, C., Buckingham, M., Potter, D., Parfitt, D.E., 2001. Phylogeny and Systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Syst. Bot.* 26, 797–807.
- Bortiri, E., Oh, S.-H., Gao, F.-Y., Potter, D., 2002. The phylogenetic utility of nucleotide sequences of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). *Am. J. Bot.* 89, 1697–1708.
- Brochmann, C., Nilsson, T., Gabrielsen, T.M., 1996. A classic example of postglacial allopolyploid speciation re-examined using RAPD markers and nucleotide sequences: *Saxifraga osloensis* (Saxifragaceae). *Symb. Bot. Ups.* 31, 75–89.
- Brown, G.R., Amarasinghe, V., Kiss, G., Carlson, J.E., 1993. Preliminary karyotype and chromosomal localization of ribosomal DNA sites in white spruce using fluorescence in situ hybridization. *Genome* 36, 310–316.
- Buckler, E.S., Holtsford, T.P., 1996a. *Zea* ribosomal repeat evolution and substitution patterns. *Mol. Biol. Evol.* 13, 623–632.
- Buckler, E.S., Holtsford, T.P., 1996b. *Zea* systematics: Ribosomal ITS evidence. *Mol. Biol. Evol.* 13, 612–622.
- Buckler, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: Divergent paralogous and phylogenetic implications. *Genetics* 145, 821–832.
- Camacho, F.J., Gernandt, D.S., Liston, A., Stone, J.K., Klein, A.S., 1997. Endophytic fungal DNA, the source of contamination in spruce needle DNA. *Mol. Ecol.* 6, 983–987.
- Campbell, C.S., Wojciechowski, M.F., Baldwin, B.G., Alice, L.A., Donoghue, M.J., 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol. Biol. Evol.* 14, 81–90.
- Catalán, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndhF* gene sequences. *Mol. Phylogenet. Evol.* 8, 150–166.
- Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.-L., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Sytsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedren, M., Gaut, B.S., Jansen, R.K., Kim, K.-J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.-Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Williams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn, G.H., Graham Jr., S.W., Barrett, S.C.H., Dayanandan, S., Albert, V.A., 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcl*. *Ann. Missouri Bot. Gard.* 80, 528–580.
- Clegg, M.T., 1993. Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA* 90, 363–367.
- Clevinger, J.A., Panero, J.L., 2000. Phylogenetic analysis of *Silphium* and subtribe Engelmanniinae (Asteraceae: Heliantheae) based on ITS and ETS sequence data. *Am. J. Bot.* 87, 565–572.
- Comes, H.P., Abbott, R.J., 2001. Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution Int. J. Org. Evolution* 55, 1943–1962.
- Compton, J.A., Culham, A., 2002. Phylogeny and circumscription of tribe Actaeae (Ranunculaceae). *Syst. Bot.* 27, 502–511.
- Cronn, R.C., Small, R.L., Wendel, J.F., 1999. Duplicated genes evolve independently after polyploid formation in cotton. *Proc. Natl. Acad. Sci. USA* 96, 14406–14411.
- Cronn, R.C., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002a. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor. Appl. Genet.* 104, 482–489.
- Cronn, R.C., Small, R.L., Haselkorn, T., Wendel, J.F., 2002b. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* 89, 707–725.
- Dagne, K., Cheng, B., Heneen, W.K., 2000. Number and sites of rDNA loci of *Guizotia abyssinica* (L. f.) Cass. as determined by fluorescence in situ hybridization. *Hereditas* 132, 63–65.
- Danna, K.J., Workman, R., Coryell, V., Keim, P., 1996. 5S rRNA genes in tribe Phaseoleae: Array size, number, and dynamics. *Genome* 39, 445–455.
- Davis, J.I., Simmons, M.P., Stevenson, D.W., Wendel, J.F., 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: An example from the monocotyledons using mitochondrial *atp A* sequences. *Syst. Biol.* 47, 282–310.
- Dixon, M.T., Hillis, D.M., 1993. Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* 1, 256–267.
- Doyle, J.J., 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.* 17, 144–163.
- 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.
- Dubcovsky, J., Dvořák, J., 1995. Ribosomal RNA multigene loci: Nomads of the Triticeae genomes. *Genetics* 140, 1367–1377.
- Elder, J.F., Turner, B.J., 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Quart. Rev. Biol.* 70, 297–320.
- Erdogan, V., Mehlenbacher, S.A., 2000. Phylogenetic relationships of *Corylus* species (Betulaceae) based on nuclear ribosomal DNA ITS region and chloroplast *matK* gene sequences. *Syst. Bot.* 25, 727–737.
- Farris, J., 1989. The retention index and the rescaled consistency index. *Cladistics* 5, 417–419.
- Ferguson, C.J., Krämer, F., Jansen, R.K., 1999. Relationships of Eastern North American *Phlox* (Polemoniaceae) based on ITS sequence data. *Syst. Bot.* 24, 616–631.
- 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.
- Fulneček, J., Lim, K.Y., Leitch, A.R., Kovarik, A., Matyasek, R., 2002. Evolution and structure of 5S rDNA loci in allotetraploid

- Nicotiana tabacum* and its putative parental species. *Heredity* 88, 19–25.
- Gatesy, J., Hayashi, C., Desalle, R., Vrba, E., 1994. Rate limits for mispairing and compensatory change: The mitochondrial ribosomal DNA of antelopes. *Evolution* 4, 188–196.
- 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.
- Germano, J., Klein, A.S., 1999. Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana*, and *P. rubens*. *Theor. Appl. Genet.* 99, 37–49.
- Gernandt, D.S., Liston, A., 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). *Am. J. Bot.* 86, 711–723.
- 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.
- Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalima, T., Oliphant, A., Briggs, S., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100.
- Grant, V., 1981. *Plant Speciation*, second ed. Columbia University Press, New York.
- Gravendeel, B., Chase, M.W., de Vogel, E.F., Roos, M., Mes, T.H.M., Bachmann, K., 2001. Molecular phylogeny of *Coleogyne* (Epidendroideae; Orchidaceae) based on plastid RFLPs, *matK*, and nuclear ribosomal ITS sequences: Evidence for polyphyly. *Am. J. Bot.* 88, 1915–1927.
- Hadjiolova, K.V., Georgiev, O.I., Nosikov, V.V., Hadjiolov, A.A., 1984. Localization and structure of endonuclease cleavage sites involved in the processing of the rat 32S precursor to ribosomal RNA. *Biochem. J.* 220, 105–116.
- Hadjiolova, K.V., Normann, A., Cavaillé, J., Soupène, E., Mazan, S., Hadjiolov, A.A., Bachelier, J.-P., 1994. Processing of truncated mouse or human rRNA transcribed from ribosomal minigenes transfected into mouse cells. *Mol. Cell Biol.* 14, 4044–4056.
- Hancock, J.M., Vogler, A.P., 2000. How slippage-derived sequences are incorporated into rRNA variable-region secondary structure: Implications for phylogeny reconstruction. *Mol. Phylogenet. Evol.* 14, 366–374.
- Hartmann, S., Nason, J.D., Bhattacharya, D., 2001. Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*. *J. Mol. Evol.* 53, 124–134.
- Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quart. Rev. Biol.* 66, 411–453.
- Hillis, D.M., Moritz, C., Porter, C.A., Baker, R.J., 1991. Evidence of biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251, 308–310.
- Hizume, M., 1994. Allopolyploid nature of *Allium wakegi* Araki revealed by genomic in situ hybridization and localization of 5S and 18S rDNAs. *Jpn. J. Genet.* 69, 407–415.
- Hodkinson, T.R., Chase, M.W., Takahashi, Ch., Leitch, I.J., Bennett, M.D., Renevoize, S.A., 2002. The use of DNA sequencing (ITS and *trnL-F*), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). *Am. J. Bot.* 89, 279–286.
- 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.
- Innis, M.A., 1990. PCR with 7-deaza-2'-deoxyguanosine triphosphate. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), *PCR protocols: A guide to methods and applications*. Academic Press, San Diego, CA, pp. 54–59.
- Kelchner, S.A., Clark, L.G., 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Mol. Phylogenet. Evol.* 8, 385–397.
- Kim, K.-J., Jansen, R.K., 1994. Comparisons of phylogenetic hypothesis among different data sets in dwarf dandelions (*Krigia*): Additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. *Plant Syst. Evol.* 190, 157–185.
- Kim, K.-J., Jansen, R.K., 1996. Phylogenetic implications of *rbcL* and ITS sequence variation in the Berberidaceae. *Syst. Bot.* 21, 381–396.
- Kita, Y., Ito, M., 2000. Nuclear ribosomal ITS sequences and phylogeny in East Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae), with special reference to extensive polymorphism in individual plants. *Plant Syst. Evol.* 225, 1–13.
- Kluge, A., Farris, J., 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18, 1–32.
- Ko, K.S., Jung, H.S., 2002. Three nonorthologous ITS1 types are present in a polypore fungus *Trichaptum abietinum*. *Mol. Phylogenet. Evol.* 23, 112–122.
- Kress, W.J., Prince, L.M., Williams, K.J., 2002. The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Am. J. Bot.* 89, 1682–1696.
- Leggett, J.M., Markand, G.S., 1995. The genomic identification of some monosomics of *Avena sativa* L. cv. Sun II using genomic in situ hybridization. *Genome* 38, 747–751.
- Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.
- Li, D., Zhang, X., 2002. Physical localization of the 18S–5.8S–26S rDNA and sequence analysis of ITS regions in *Thinopyrum ponticum* (Poaceae: Triticeae): Implications for concerted evolution. *Ann. Bot.* 90, 445–452.
- Liston, A., Álvarez-Buylla, E.R., 1995. Internal transcribed spacer sequences of conifers: There is a fungus among us. *Inoculum* 46, 26.
- Liston, A., Robinson, W.A., Oliphant, J.M., Álvarez-Buylla, E.R., 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed 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.
- Ma, Y., Islam-Faridi, M.N., Crane, C.F., Ji, Y., Stelly, D.M., Price, H.J., Byrne, D.H., 1997. In situ hybridization of ribosomal DNA to rose chromosomes. *J. Hered.* 88, 158–161.
- Maggini, F., Frediani, M., Gelati, M.T., 2000. Nucleotide sequence of the internal transcribed spacers of ribosomal DNA in *Picea abies* Karst. *DNA Seq.* 11, 87–89.
- Mai, J.C., Coleman, A.W., 1997. The Internal Transcribed Spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J. Mol. Evol.* 44, 258–271.
- Manos, P., Doyle, J.J., Nixon, K.C., 1999. Phylogeny, biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Mol. Phylogenet. Evol.* 12, 333–349.
- Marrocco, R., Gelati, M.T., Maggini, F., 1996. Nucleotide sequence of the internal transcribed spacers and 5.8S region of ribosomal DNA in *Pinus pinea* L. *DNA Seq.* 6, 175–177.
- Masterson, J., 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* 264, 421–423.

- Mayol, M., Rosselló, J.A., 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Mol. Phylogenet. Evol.* 19, 167–176.
- McDade, L.A., 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. *Evolution* 46, 1329–1346.
- McDade, L.A., 1995. Hybridization and phylogenetics. In: Hoch, P.C., Stephenson, A.G. (Eds.), *Experimental and Molecular Approaches to Plant Biosystematics*. Missouri Botanical Gardens, St. Louis, Missouri, pp. 305–331.
- Miller, R.E., Rausher, M.D., Manos, P.S., 1999. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and *waxy* sequences. *Syst. Bot.* 24, 209–227.
- Mishima, M., Ohmido, N., Fukui, K., Yahara, T., 2002. Trends in site-number change of rDNA loci during polyploid evolution in *Sanguisorba* (Rosaceae). *Chromosoma* 110, 550–558.
- Möller, M., Cronk, Q.C.B., 1997. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *Am. J. Bot.* 84, 956–965.
- Molvray, M., Kores, P.J., Chase, M.W., 1999. Phylogenetic relationships within *Korthalsella* (Viscaceae) based on nuclear ITS and plastid *trnL-F* sequence data. *Am. J. Bot.* 86, 249–260.
- Mort, M.E., Soltis, D.E., Soltis, P.S., Francisco-Ortega, J., Santos-Guerra, A., 2002. Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Syst. Bot.* 27, 271–288.
- Muir, G., Fleming, C.C., Schlötterer, C., 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol. Biol. Evol.* 18, 112–119.
- Murray, B.G., 2002. Trees, maps and FISH: The application of genome based technologies to the analysis of chromosome evolution. *Curr. Genom.* 3, 539–550.
- Musters, W., Boon, K., van der Sande, C.A.F.M., van Heerikhuizen, H., Planta, R.J., 1990. Functional analysis of transcribed spacers of yeast ribosomal DNA. *EMBO J.* 9, 3989–3996.
- Nakayama, S., Fujishita, M., Sone, T., Ohyamak, T., 2001. Additional locus of rDNA sequence specific to the X chromosome of the liverwort, *Marchantia polymorpha*. *Chromosome Res.* 9, 469–473.
- Nashimoto, M., Ogata, K., Mishima, Y., 1988. In vitro sequence-specific cleavage in transcribed spacer of mouse precursor ribosomal RNA. *J. Biochem.* 103, 992–997.
- O'Donnell, K., Cigelnik, E., 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* 7, 103–116.
- Olmstead, R.G., Palmer, J.D., 1994. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* 81, 1205–1224.
- Olmstead, R.G., Reeves, P.A., 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. *Ann. Missouri Bot. Garden* 82, 176–193.
- Panzer, F., Giménez-Abián, M.I., López-Sáez, J.F., Giménez-Martín, G., Cuadrado, A., Shaw, P.J., Beven, A.F., Cánovas, L., de la Torre, C., 1996. Nucleolar organizer expression in *Allium cepa* L. chromosomes. *Chromosoma* 105, 12–19.
- Pedersen, C., Linde-Laursen, I., 1994. Chromosomal locations of four minor rDNA loci and a marker microsatellite sequence in barley. *Chrom. Res.* 2, 65–71.
- Raina, S.N., Rani, V., 2001. GISH technology in plant genome research. *Methods Cell Sci.* 23, 83–104.
- 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.
- Ray, M.F., 1995. Systematics of *Lavatera* and *Malva* (Malvaceae, Malveae) – a new perspective. *Plant Syst. Evol.* 198, 29–53.
- Raymundez, M.B., Mathez, J., Xena de Enrech, N., Dubuisson, J.Y., 2002. Coding of insertion-deletion events of the chloroplastic intergene *atp beta-rbcL* for the phylogeny of the *Valerianeae* tribe (Valerianaceae). *C. R. Biol.* 325, 131–139.
- Rieseberg, L.H., Beckstrom-Sternberg, S., Doan, K., 1990. *Helianthus annuus* ssp. *texasus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proc. Natl. Acad. Sci. USA* 87, 593–597.
- Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plant* 5, 65–84.
- Rieseberg, L.H., Wendel, J.F., 1993. Introgression and its consequences in plants. In: Harrison, R. (Ed.), *Hybrid Zones and the Evolutionary Process*. Oxford University Press, Oxford, pp. 70–109.
- Rieseberg, L.H., Baird, S.J., Gardner, K.A., 2000. Hybridization, introgression, and linkage evolution. *Plant Mol. Biol.* 42, 205–224.
- Ritland, C.E., Ritland, 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.
- Roalson, E.H., Columbus, J.T., Friar, E.A., 2001. Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and *trnT-L-F* (cpDNA) region sequences: Assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis*. *Syst. Bot.* 26, 318–341.
- Roelofs, D., van Velzen, J., Kuperus, P., Bachmann, K., 1997. Molecular evidence for an extinct parent of the tetraploid species *Microseris acuminata* and *M. campestris* (Asteraceae, Lactuceae). *Mol. Ecol.* 6, 641–649.
- Rossetto, M., Jackes, B.R., Scott, K.D., Henry, R.J., 2002. Is the genus *Cissus* (Vitaceae) monophyletic? Evidence from plastid and nuclear ribosomal DNA. *Syst. Bot.* 27, 522–533.
- Samuel, R., Bachmair, A., Jobst, J., Ehrendorfer, F., 1998. ITS sequences from nuclear rDNA suggest phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (Fagaceae). *Plant Syst. Evol.* 211, 129–139.
- 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.
- Schlötterer, C., Hauser, M.-T., von Haesler, A., Tautz, D., 1994. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Mol. Biol. Evol.* 11, 513–522.
- Schlötterer, C., Tautz, D., 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Current Biol.* 4, 777–783.
- Schubert, I., Wobus, U., 1985. In situ hybridization confirms jumping nucleolus organizing regions in *Allium*. *Chromosoma* 92, 143–148.
- Seelanan, T., Schnabel, A., Wendel, J.F., 1997. Congruence and consensus in the cotton tribe (Malvaceae). *Syst. Bot.* 22, 259–290.
- Seelanan, T., Brubaker, C.T., Stewart, J. McD., Craven, L.A., Wendel, J.F., 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Syst. Bot.* 24, 183–208.
- 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.
- Shishido, R., Sano, Y., Fukui, K., 2000. Ribosomal DNAs: An exception to the conservation of gene order in rice genomes. *Mol. Gen. Genet.* 263, 586–591.
- Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381.
- Simmons, M.P., Ochoterena, H., Carr, T.G., 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst. Biol.* 50, 454–462.
- Small, R.L., Cronn, R.C., Wendel, J.F., 2003. Nuclear genes and plant phylogeny reconstruction. *Australian Syst. Bot.* (in review).

- Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *Amer. J. Bot.* 85, 1301–1315.
- Smedmark, J.E.E., Eriksson, T., 2002. Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and *trnL-trnF* regions. *Syst. Bot.* 27, 303–317.
- Smith, D.E., Klein, A.S., 1994. Phylogenetic inferences on the relationship of North American and European *Picea* species based on nuclear ribosomal 18S sequences and the internal transcribed spacer 1 region. *Mol. Phylogenet. Evol.* 3, 17–26.
- Snowdon, R.J., Köhler, W., Köhler, A., 1997. Chromosomal localization and characterization of rDNA loci in the *Brassica* A and C genomes. *Genome* 40, 582–587.
- Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), 1998. *Molecular systematics of plants II: DNA sequencing*. Kluwer Academic, Boston.
- Soltis, P.S., Soltis, D.E., 1991. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. *Syst. Bot.* 16, 407–413.
- Soltis, P.S., Plunkett, G.M., Novak, S.J., Soltis, D.E., 1995. Genetic variation in *Tragopogon* species: Additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). *Am. J. Bot.* 82, 1329–1341.
- Springer, M.S., Hollar, L.J., Burk, A., 1995. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.* 12, 1138–1150.
- Stebbins, G.L., 1950. *Variation and evolution in plants*. Columbia University Press, New York.
- Suh, Y., Thien, L.B., Reeve, H.E., Zimmer, E.A., 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *Am. J. Bot.* 80, 1042–1055.
- Tartof, K.D., Dawid, I.B., 1976. Similarities and differences in the structure of X and Y chromosome rRNA genes of *Drosophila*. *Nature* 263, 27–30.
- The *Arabidopsis* Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Thomas, H.M., Harper, J.A., Morgan, W.G., 2001. Gross chromosome rearrangements are occurring in an accession of the grass *Lolium rigidum*. *Chrom. Res.* 9, 585–590.
- Torres, R.A., Ganal, M., Hemleben, V., 1990. GC balance in the internal transcribed spacers ITS 1 and ITS 2 of nuclear ribosomal RNA genes. *J. Mol. Evol.* 30, 170–181.
- van Ham, R.C., Hart, H., Mes, T.H., Sandbrink, J.M., 1994. Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. *Curr. Genet.* 25, 558–566.
- van Nues, R.W., Rientjes, J.M.J., van der Sande, C.A.F.M., Zerp, S.F., Sluiter, C., Venema, J., Planta, R.J., Raué, H.A., 1994. Separate structural elements within Internal Transcribed Spacer 1 of *Saccharomyces cerevisiae* ribosomal DNA. *J. Mol. Biol.* 223, 899–910.
- Vanzela, A.L.L., Cuadrado, A., Jouve, N., Luceño, M., Guerra, M., 1998. Multiple locations of the rDNA sites in holocentric chromosomes of *Rhynchospora* (Cyperaceae). *Chrom. Res.* 6, 345–349.
- Varadaraj, K., Skinner, D.M., 1994. Denaturants or cosolvents improve the specificity of PCR amplification of a G+C-rich DNA using genetically-engineered DNA-polymerases. *Gene* 140, 1–5.
- Vargas, P., McAllister, H.A., Morton, C., Jury, S.L., Wilkinson, M.J., 1999. Polyploid speciation in *Hedera* (Araliaceae): Phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. *Plant Syst. Evol.* 219, 165–179.
- Vaughan, H.E., Jamilena, M., Ruiz Rejón, C., Parker, J.S., Garrido-Ramos, M.A., 1993. Loss of nucleolus-organizer regions during polyploid evolution in *Scilla autumnalis*. *Heredity* 71, 574–580.
- Veldman, G.M., Klootwijk, J., van Heerikhuizen, H., Planta, R.J., 1981. The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. *Nucleic Acids Res.* 9, 4847–4862.
- Venkateswarlu, K., Nazar, R., 1991. A conserved core structure in the 18–25S rRNA intergenic region from tobacco, *Nicotiana rustica*. *Plant Mol. Biol.* 17, 189–194.
- Vining, T.F., Campbell, C.S., 1997. Phylogenetic signal in sequence repeats within nuclear ribosomal DNA internal transcribed spacer 1 in *Tsuga*. *Am. J. Bot.* 84, 702.
- Vining, T.F., Gernandt, D.S., Liston, A., Campbell, C.S., 1998. Nuclear ribosomal DNA (nrDNA) internal transcribed spacer 1 (ITS1) subrepeats and phylogeny of Pinaceae. *Am. J. Bot.* 85, 186.
- Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M., Pendleton, J., 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Syst. Biol.* 43, 250–261.
- Wagstaff, S.J., Wege, J., 2002. Patterns of diversification in New Zealand Styliidiaceae. *Am. J. Bot.* 89, 865–874.
- Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. In: Soltis, P., Soltis, D., Doyle, J. (Eds.), *Molecular Systematics of Plants II*. Kluwer Academic, Dordrecht, 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.
- Wendel, J.F., 2000. Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249.
- Wendel, J.F., Cronn, R.C., 2003. Polyploidy and the evolutionary history of cotton. *Adv. Agronomy* 78, 139–186.
- Wheeler, W.C., Honeycutt, R.L., 1988. Paired sequence difference in ribosomal RNAs: Evolution and phylogenetic implications. *Mol. Biol. Evol.* 5, 90–96.
- 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.
- Whitten, W.M., Williams, N.H., Chase, M.W., 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: Combined molecular evidence. *Am. J. Bot.* 87, 1842–1856.
- Winship, P.R., 1989. An improved method for direct sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* 17, 1266.
- Wissemann, V., 1999. Genetic constitution of *Rosa* Sect. *Caninae* (*R. canina*, *R. jundzillii*) and Sect. *Gallicanae* (*R. gallica*). *J. Appl. Botany* 73, 191–196.
- Wissemann, V., 2003. Hybridization and the evolution of the nrITS spacer region. In: Sharma, A.K., et al. (Eds.), *Plant genome, biodiversity and evolution I: Phanerogams*. Science Publishers, Inc. Enfield, New Hampshire 57–71 (in press).
- Wissemann, V., 2002. Molecular evidence for allopolyploid origin of the *Rosa canina* complex (Rosaceae, Rosoideae). *J. Appl. Bot.* 76, 176–178.
- Worton, R.G., Sutherland, J., Sylvester, J.E., Willard, H.F., Bodrug, S., Dube, I., Duff, C., Kean, V., Ray, P.N., Schmickel, R.D., 1988. Human ribosomal RNA genes: Orientation of the tandem array and conservation of the 5' end. *Science* 239, 64–68.
- Xu, J., Earle, E.D., 1996. High resolution physical mapping of 45S (5.8S, 18S, and 25S) rDNA gene loci in the tomato genome using a

- combination of karyotyping and FISH of pachytene chromosomes. *Chromosoma* 104, 545–550.
- Yang, Y.W., Lai, K.N., Tai, P.Y., Ma, D.P., Li, W.H., 1999. Molecular phylogenetic studies of *Brassica*, *Rorippa*, *Arabidopsis*, and allied genera based on the internal transcribed spacer region of 18S–25S rDNA. *Mol. Phylogenet. Evol.* 13, 455–462.
- Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., Yang, H., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92.
- Zhang, J., Friebe, B., Gill, B.S., 1995. Detection of maize DNA sequences amplified in wheat. *Genome* 38, 946–950.
- Zhang, W., Reading, C., Deisseroth, A.B., 1994. Improved PCR sequencing with formamide. *Trends Genet.* 8, 332.
- 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 alpha-chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* 77, 2158–2162.
- Zimmer, E.A., Roalson, E.H., Skog, L.E., Boggan, J.K., Idnurm, A., 2002. Phylogenetic relationships in the Gesnerioideae (Gesneriaceae) based on nrDNA ITS and cpDNA *trnL-F* and *trnE-T* spacer region sequences. *Am. J. Bot.* 89, 296–311.
- Zoller, J.F., Yang, Y., Herrmann, R.G., Hohmann, U., 2001. Comparative genomic in situ hybridization (cGISH) analysis on plant chromosomes revealed by labelled *Arabidopsis* DNA. *Chromosome Res.* 9, 357–375.
- Zomlefer, W.B., Williams, N.H., Whitten, W.M., Judd, W.S., 2001. Generic circumscription and relationships in the tribe Melanthieae (Liliales, Melanthiaceae), with emphasis on *Zigadenus*: Evidence from ITS and trnL-F sequence data. *Am. J. Bot.* 88, 1657–1669.