HOW UNIVERSAL ARE UNIVERSAL rDNA PRIMERS? A CAUTIONARY NOTE FOR PLANT SYSTEMATISTS AND PHYLOGENETICISTS

M. MÖLLER*

Phylogenetic relationships are frequently inferred from rDNA ITS sequences obtained employing universal primers. Whereas in most cases phylogenetic topologies inferred from ITS data make good biological sense, caution has to be used when a number of taxa vary greatly in PCR yield. The phylogenetic implications of such a scenario are discussed.

Keywords. Internal transcribed spacer, nucleolar organizer regions, PCR, ribosomal DNA, universal primers.

Since the invention of automated PCR and the innovation of direct cycle-sequencing, the use of molecular data in areas such as molecular systematics and phylogenetics has increased dramatically. Here, I would like to elaborate on an issue of a principally technical nature, but with potential serious implications for the correct interpretation of phylogenetic analyses.

White *et al.* (1990) published sequences of universal primers, designed for the amplification of fungal ribosomal DNA (rDNA) genes and spacers, that have been widely used, either unchanged (e.g. Baldwin, 1992; Wojciechowski *et al.*, 1993; Wendel *et al.*, 1995), or modified to fit plant sequences (e.g. Sang *et al.*, 1995; Möller & Cronk, 1997). The rDNA gene copies, including the two internal transcribed spacer (ITS) regions, are arranged in multicopy tandem repeat units in nucleolar organizer regions (NOR), and are thought to be homogenized by forces such as 'concerted evolution' (Hillis & Dixon, 1991), thought mainly to be a result of unequal crossing over (Smith, 1976) or gene conversion events (Arnheim, 1983). However, recent studies have shown that this 'molecular drive' appears to be less efficient than previously thought; the variation of individual copies in some cases is now undisputed, and can exceed interspecies variation (Karvonen *et al.*, 1994; Smith & Klein, 1994; Oxelman & Lidén, 1995). This may indicate past hybridization events (Karvonen *et al.*, 1994; Campbell *et al.*, 1997) or the accidental amplification of paralogues (Chaw *et al.*, 1995; Buckler *et al.*, 1997). The high sequence variation between ITS copies within amplifiable repeat units of *Zea* may give an indication of the possible variation of rDNA genes within a single NOR locus; out of more than 60 cloned copies none but two were identical (those belonging to different species!) (Buckler & Holtsford, 1996).

It is widely believed or assumed, that universal primer pairs amplify all, or at least the majority, of the existing ITS copies present. Failure of PCR-reactions to amplify

* Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK.

the desired product are usually attributed to deficiencies of DNA extractions (e.g. low DNA concentration, DNA degradation, high salt impurities, co-precipitation of inhibiting compounds) or unsuitable PCR conditions (e.g. annealing temperature or time, buffer composition) and can be overcome by elimination of those inadequacies. Buckler and Holtsford (1996) reported higher amplification of isolated ITS paralogues with low GC base contents (putative low stability pseudogenes) as compared with high stability functional copies with high GC contents (>70%). However, in further PCR experiments mixing high and low stability copies (which would reflect more realistically laboratory situations) the results were more equivocal (Buckler *et al*., 1997).

In certain cases it is not possible to obtain normal ITS amplicon yields under any PCR condition, although the amplification of other DNA fragments (e.g. chloroplast DNA; cpDNA) from the same DNA extraction appears to be unimpaired (Fig. 1). This could be attributable to pipetting tolerances in the set-up of individual PCR reactions, thus in differences in the number of starting templates, which will ultimately affect the PCR yield. However, an elegant method to exclude those deficiencies, and to introduce an internal calibration system, is to combine the amplification of two amplicons in one PCR reaction (the simultaneously amplification of amplicons of up to eight primer pairs have been demonstrated; Löffert *et al*., 1997). A double-check on non-complementarity across the individual pairs (to avoid primer dimer artefacts) and a unison annealing temperature are pre-requisites of such a system.

I have used this method in my research on the *Gesneriaceae*. The ITS copies of a group of Madagascan *Streptocarpus* species that had previously repeatedly failed to amplify properly (despite changes in various PCR parameters) whereas a mainland African counterpart was unproblematic in this respect. In both cases chloroplast

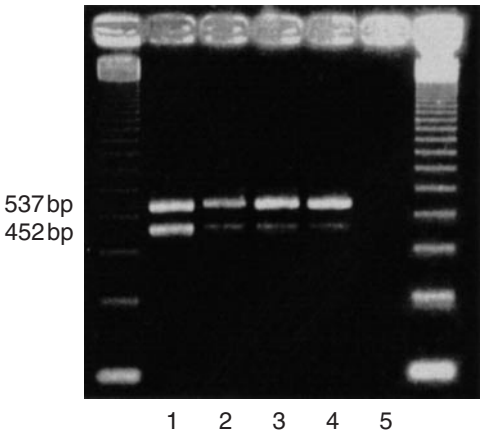

FIG. 1. Simultaneous PCR amplification of cpDNA (*trn*L intron, 537bp) and nuclear DNA (ITS1, 452bp) for the African *Streptocarpus rexii* (lane 1) and the Madagascan *S. muscosus* (lane 2), *S. levis* (lane 3) and *S. tanala* (lane 4). Negative control (lane 5). Left and right lanes are 123bp size marker.

DNA amplified equally well in the same PCR reaction (Fig. 1). The situation in *Streptocarpus* is not an isolated phenomenon. It also occurs in other, not necessarily closely related, plant families, such as *Zingiberaceae* (A. Rangsiruji, personal communication) and thus may present a more common but, probably unconsciously, underestimated or neglected aspect of direct rDNA sequencing. Direct sequencing approaches were used in the majority of molecular phylogenetic studies using multicopy rDNA (e.g. Baldwin, 1992; Wojciechowski *et al*., 1993; Sang *et al*., 1995; Soltis & Kuzoff, 1995; Wendel *et al*., 1995). The results are in general majority rule sequences, were individual base determination is governed by the base prevailing amongst all templates amplified. This 'hides' any lower level base variation between individual copies within an individual (see above).

What could be the explanation for the differential amplification of rDNA genes? There are several hypothetical scenarios: 1, a bias in the DNA extraction favouring cpDNA in some species, resulting in a higher number of starting templates of this molecule. However, the CTAB method usually employed generally extracts total DNA (Doyle & Doyle, 1987). Further, there is no compelling reason why the DNA extraction in one group of species should be preferentially affected, while closely related species are not. 2, a reflection of differences in the total number of ITS copies assembled in the NORs. This is theoretically possible, but in the present case practically unlikely. If the number of copies amplified correlates with the number of starting copies, comparison by fluorometric quantification with well amplified species suggest that the Madagascan species would have only c.20–30% of the copies of the mainland Africa species (Fig. 1). This low copy number seems unlikely to be sufficient to support any organism, giving the role of ribosomes in cell function. 3, a more worrying but likely explanation could be a partial amplification of the total number of ITS copies present. There can be various reasons for partial amplification, such as a mutation in a PCR primer site in combination with incomplete homogenization by concerted evolutionary forces. Apparently conserved regions flanking both ITS1 and 2 can have considerable variation, enough to allow possible mutation hits in primer sites; e.g. the 5.8S rDNA gene, the location of internal primer sites, has up to 3.7% divergence between *Brassica* species (Suh *et al*., 1992). This has apparently happened in the genus *Alpinia* where a specific internal sequencing primer had to be designed (A. Rangsiruji, personal communication). The third scenario would effectively result in a selective sampling of the total rDNA copies, introducing a 'bias'.

The power of PCR to (theoretically) exponentially amplify DNA fragments may disguise the extent of this bias as long as a 'sufficiently' high yield is obtained, and its degree only becomes apparent when using an internal calibration system, that allows an estimation and comparison of the effective starting templates. Thus, knowledge of rDNA gene organization, the exact number of NOR loci as well as an accurate estimation of rDNA repeat numbers is necessary to approximate the extent of the excluded rDNA copies. This is comparable with the 'dark matter' of the universe; we infer it is there, but it cannot be directly detected. Approaches to clarify the organization of rDNA within a genome may involve the physical mapping of

NOR loci by fluorescent in-situ hybridization (FISH) (Maluszynska & Heslop-Harrison, 1991), and quantification of total nuclear rDNA tandem repeats by restriction mapping and southern blotting (Copenhaver & Pikaard, 1996; Campbell *et al*., 1997). In any case, where large discrepancies in PCR yield between samples are encountered, caution should be executed, and internal checks should be employed to estimate the proportion of templates amplified.

Can this situation be rectified? When only a small subset of rDNA copies is sampled the sequences obtained may not be representative for the taxon analysed. Possible strategies to include a larger sample of rDNA repeats in PCR reactions would involve lowering the stringency by decreasing the annealing temperature, or the use of alternative conserved primer sites. The first may result in a higher amplicon concentration, but can result in the co-amplification of non-specific products that will cause problems during direct sequencing. The latter would create the possible problem of strict homology between sequences if a matrix is assembled from sequences obtained with different primer sets; different primer pairs may draw differential sets of copies. The addition of denaturants, such as dimethylsulphoxide (DMSO) has been suggested to improve the individual amplification of high stability ITS copies as opposed to low stability products (presumed non-functional paralogues); however, this approach was found to give inconsistent results in competition amplifications (Buckler *et al*., 1997). (Non-functional paralogues can be more effectively eliminated by sequence determination from functional rRNA genes indirectly by the extraction of total RNA, from which cDNA is synthesized; Chaw *et al*., 1995.)

What are the consequences for phylogenetic studies? Where multigene families are completely homogenized (and assuming there is no between individual variation) selective sampling would not represent a problem, as any single copy could represent a taxon (Sanderson & Doyle, 1992; Doyle & Davis, 1998). Consequently, in cases where primer site mutations reach fixation in an individual by concerted evolution no PCR amplification will occur. However, in cases of incomplete homogenization involving mutations at primer sites a smaller number of copies will be amplified. This could be compared with lineage sorting or 'gene extinction'. Where only very few copies are sampled (i.e. indicated as low PCR yield), the subsequent inferred taxic phylogenetic relationships can be compromised (Doyle & Davis, 1998). The solution of employing several alternative primer sets in the assemblage of one matrix will raise the problem of strict homology, as the sequences may stem from differential sampling of ITS copies from the whole gene pool and may thus compare non-homologous copies. The use of an alternative primer set within a defined range of taxa, however, may be acceptable.

This paper is primarily concerned with intra-individual polymorphisms and conservation of primer sites. Of course where hybridization occurs and/or intraspecific variability exists phylogenetic estimates can be confounded. These topics are beyond the scope of this paper but are also other important potential error sources.

In conclusion, direct PCR-based sequencing is a convenient and rapid method to obtain numerous characters. Whereas in most cases phylogenetic topologies inferred

from ITS data make good biological sense and are corroborated by other lines of evidence, caution has to be used when a number of taxa vary greatly in PCR yield. If PCR parameters can be excluded as potential source of the problem, the choice of an alternative gene sequence for phylogenetic analyses may be necessary.

MATERIALS, METHODS AND RESULTS

Plant material was from the living collection held at the Royal Botanic Garden Edinburgh (RBGE), and vouchered as described previously, as were DNA extraction, PCR and electrophoresis (Möller & Cronk, 1997). Addition of two primer pairs to one PCR reaction, amplifying the ITS1 (primers ITS5P+ITS2G), and the intron of the chloroplast gene *trn*L (primers c+d; Taberlet *et al*., 1991) resulted in two strong bands in the African species, but one strong band (cpDNA) and one weak band (ITS1) for the Madagascan species (Fig. 1).

ACKNOWLEDGEMENTS

I would like to thank the Royal Botanic Garden staff, in particular, R. M. Bateman, Q. C. B. Cronk, P. M. Hollingsworth and R. T. Pennington for helpful discussions on the topic. I further extend my thanks to the Leverhulme Trust, and the FWF for financial contributions.

REFERENCES

ARNHEIM, N. (1983). Concerted evolution of multigene families. In: NEI, M. & KOEHN, R. K. (eds) *Evolution of Genes and Proteins*, pp. 38–61. Sunderland, MA: Sinauer.

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.

BUCKLER, E. S. & HOLTSFORD, T. P. (1996). *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 paralogues and phylogenetic implications. *Genetics* 145: 821–832.

CAMPBELL, C. S., WOJCIECHOWSKI, M. F., BALDWIN, B. G., ALICE, L. A. & DONOGHUE, M. J. (1997). Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol. Biol. Evol.* 14: 81–90.

CHAW, S.-M., SUNG, H.-M., LONG, H., ZHARKIKH, A. & LI, W.-H.(1995). The phylogenetic positions of the conifer genera *Amentotaxus*, *Phyllocladus*, and *Nageia* inferred from 18S rDNA sequences. *J. Mol. Evol*. 41: 224–230.

COPENHAVER, G. P. & PIKAARD, C. S. (1996). RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant Journal* 9: 259–272.

DOYLE, J. J. & DAVIS, J. I. (1998). Homology in molecular phylogenetics: a parsimony perspective. In: SOLTIS, D. E, SOLTIS, P. S. & DOYLE, J. J. (eds)

Molecular Systematics of Plants II: DNA Sequencing, pp. 101–131. London: Kluwer Academic Press.
DOYLE, J. J. & DOYLE, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11–15.
HILLIS, D. M. & DIXON, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66: 411–453.
KARVONEN, P., SZMIDT, A. E. & SAVOLAINEN, O. (1994). Length variation in the internal transcribed spacers of ribosomal DNA in *Picea abies* and related species. *Theor. Appl. Genet.* 89: 969–974.
LÖFFERT, D., SCHAFFRATH, N., BERKENKOPF, M., KARGER, S. & KANG, J. (1997). Muliplex PCR with Qiagen. *Qiagen News* 4: 1+20.
MALUSZYNSKA, J. & HESLOP-HARRISON, J. S. (1991). Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *Plant Journal* 1: 159–166.
MÖLLER, M. & CRONK, Q. C. B. (1997). Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *Amer. J. Bot*. 84: 956–965.
OXELMAN, B. & LIDÉN, M. (1995). Generic boundaries in the tribe Sileneae (Caryophyllaceae) as inferred from nuclear rDNA sequences. *Taxon* 44: 525–542.
SANDERSON, M. J. & DOYLE, J. J. (1992). Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Syst. Biol.* 41: 4–17.
SANG, T, CRAWFORD, D. J. & STUESSY, T. F. (1995). Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. USA* 92: 6813–6817.
SMITH, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528–535.
SMITH, D. E. & KLEIN, A. S. (1994). Phylogenetic inference 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.
SOLTIS, D. E. & KUZOFF, R. K. (1995). Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49: 727–742.
SUH, Y., THIEN, L. B. & ZIMMER, E. A. (1992). Nucleotide sequences of the internal transcribed spacers and 5.8S rRNA gene in *Canella winterana* (Magnoliales; Canellaceae). *Nucleic Acids Res.* 20: 6101–6102.
TABERLET, P., GIELLY, L., PAUTOU, G. & BOUVET, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17: 1103–1109.
WENDEL, J. F., SCHNABEL, A. & SEELANAN, T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* 92: 280–284.
WHITE, T. J., BRUNS, T., LEE, S. & TAYLOR, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: INNIS, M. A., GELFAND, D. H., SNINSKY, J. J. & WHITE, T. J. (eds) *PCR Protocols*, pp. 315–322, London: Academic Press.
WOJCIECHOWSKI, M. F., SANDERSON, M. J., BALDWIN, B. G. & DONOGHUE, M. J. (1993). Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *Amer. J. Bot.* 80: 711–722.

Received 25 October 1999; accepted with revision 22 December 1999