

Pan-eukaryote ITS2 homologies revealed by RNA secondary structure

Annette W. Coleman*

Division of Biology and Medicine, Brown University, Providence, RI 02912, USA

Received January 29, 2007; Revised March 29, 2007; Accepted March 29, 2007

ABSTRACT

For evolutionary comparisons, phylogenetics and evaluation of potential interbreeding taxa of a species, various loci have served for animals and plants and protists. One [second internal transcribed spacer (ITS2) of the nuclear ribosomal DNA] is highly suitable for all. Its sequence is species specific. It has already been used extensively and very successfully for plants and some protists, and a few animals (where historically, the mitochondrial genes have dominated species studies). Despite initial impressions that ITS2 is too variable, it has proven to provide useful biological information at higher taxonomic levels, even across all eukaryotes, thanks to the conserved aspects of its transcript secondary structure. The review of all eukaryote groups reveals that ITS2 is expandable, but always retains in its RNA transcript a common core structure of two helices with hallmark characteristics important for ribosomal RNA processing. This aspect of its RNA transcript secondary structure can rescue difficult alignment problems, making the ITS2 a more powerful tool for phylogenetics. Equally important, the recognition of eukaryote-wide homology regions provides extensive and detailed information to test experimental studies of ribosomal rRNA processing.

INTRODUCTION

Comparative phylogenetics is a powerful tool, not just for establishing evolutionary relationships among taxa, but also for discerning biochemically significant aspects of a locus, versus those always present but variable. The 3' nuclear ribosomal transcribed spacer region [second internal transcribed spacer (ITS2)] sequence is much used for phylogenetic studies at the species and genus level. More recently, additional phylogenetically useful information from these sequences has emerged as a consequence of the solution of their putative transcript

secondary structure (1). Initial analysis of potential folding homologies was presented by Wolf *et al.* (2). From folding of transcripts of many additional phyla, it now appears that conservation of secondary structure of ITS2 itself among essentially all eukaryotes is far more remarkable than had ever been anticipated. Although the highly conserved helix regions revealed are presumably essential for rRNA processing, the details of this complex activity are not yet fully understood. Here we point out homologies as they apply across animals, plants and protistan phyla to call attention to the breadth of information already available to both phylogenetic studies and to detailed analyses of ITS processing.

Phylogenetic studies have relied on a variety of DNA loci. Protists present probably the most difficult choice of locus to sequence, for the broad and ancient variety of cell types encompassed is greater than in fungi, animals or plants. Among protists, plastid *rbcL* has been sequenced for many algae and mitochondrial *cox I* for many non-photosynthetic protists. However, some protists are photosynthetic and some not, even in the same class; and there are numerous examples of host cells with plastids transferred from a different eukaryote. As for mitochondria, some protists are even lacking standard mitochondrial genomes. These facts make organelle genes less appealing for broader comparisons.

A single common nuclear locus would seem most useful, as long as it is appropriate for the task.

What should be its attributes? Such a locus should be present in all of the chosen taxa, and there should be no known case of horizontal transfer; and it should identify the organism to a unique species. One candidate is the single most frequently sequenced DNA region, with variability suitable to the species level, the ITS2 in the nuclear ribosomal gene cistron (3). Here we review the ITS2 region across eukaryotes and find it is not just specific to species but also laden with surprisingly useful information concerning higher relationships, and clearly constrained in its evolution to maintain certain regions of transcript secondary structure universal among eukaryotes. Aspects of these conserved regions should be useful and very important to future studies of rRNA processing.

*To whom correspondence should be addressed. Tel: +1 401 863 3917; Fax: +1 401 963 2421; Email: Annette_Coleman@brown.edu

MATERIALS AND METHODS

For the various phyla of eukaryotes, sets of ITS2 sequences corresponding to clades of species, genera and families were sought. Sequences of ITS2 and adjacent nucleotides were then downloaded from GenBank and initially aligned using MacVector software (Kodak, International Biotechnologies Inc., New Haven, CT, USA). RNA transcript foldings were engendered by mfold Version 3.0 available at <http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi> (4), and multiple comparisons, searching for the common structure within the clade, were carried out. These were aided by the knowledge of conserved helix II structure, and by the existence of the highly conserved region of DNA on the 5' side of helix III. Each final alignment was adjusted to reflect pairings in common among the sequences, either identical or showing compensatory base changes that conserve the pairing, and the conservation of their secondary structures was compared.

RESULTS

Eukaryote-universal ITS2 secondary structure

From ITS sequencing and analysis of RNA transcript secondary structure, a set of near-universal eukaryote ITS2 structures has emerged, as highlighted in Figure 1 for a diatom, a green alga, a red alga and yeast. The features common to eukaryotes were first noted during analyses of the many genera and species of the green algal group Volvocaceae, and their comparison with terrestrial plants (5). ITS2 typically has four helices (Figure 1), and helices I and IV are the most variable (species and subspecies level specificity). At least within a genus, the basal pairings of helices I, II and IV are conserved while the most conserved portion of helix III is distal. The two helices that we term II and III, and their adjacent single-stranded regions, contain the most conserved regions of primary sequence (Figure 1, cartoon).

However, not all eukaryotes have the same number of helices, and only helices II and III are recognizable and common to essentially all. Hence, helix II (recognized by its hallmarks) may be the first helix in, for example, the typical ciliate ITS2 structure (6). In some groups, insects particularly, there is commonly a helix between the recognizable Helix II and Helix III, which we have termed IIA. Like all the other helices except II, it may be branched. For global comparisons, it is less confusing to ignore the total number of helices and concentrate rather on the two helices that have individually recognizable motifs.

The hallmark helices, helices II and III

Helix II is recognizable because it is short (rarely more than 12 pairings), it lacks any branching, and it possesses a pyrimidine–pyrimidine mismatch (arrows in Figure 1), which makes a bulge near the base of the helix. Figure 4 of (5) illustrates the many variant forms of helix II among angiosperms, all with the pyrimidine mismatch. The sole exception to the rule of shortness and no branching is

spectacular. In the genera of ticks studied by Hlinka *et al.* (7) there is an insertion of hundreds of nucleotides of long repeats at the tip of helix II, leading to complicated branching. These are the longest known ITS2 sequences.

Helix III, by contrast, is typically longer than helix II and is frequently branched. The region of greatest absolute sequence conservation in the whole ITS2 is on the 5' side of helix III near the tip. Even when helix III is branched, the highly conserved stretch of nucleotides is found on the 5' side of one of the branches. This sequence, marked in Figure 1, is typically conserved absolutely at the family or even higher level. Coelomates and some other taxa (e.g. 8) have a branched, sometimes much branched, helix III. Historically this characteristic has affected the research field, because mammalian and particularly human, helix III of ITS2 is long and multiply branched. Hence not only is the common secondary structure difficult to perceive, but the great number of relatively poorly conserved nucleotides in ITS2 initially gave the locus a reputation for high variability and perhaps uselessness for phylogenetic purposes.

Our secondary structure analyses and comparisons show that, no matter what the total length of ITS2, ~100–115 nt positions, consistent in their position in the secondary structure, are relatively conserved. These include the basal 10 pairings of helix II and, in helix III, those 18 pairings that include and surround the single most absolutely conserved sequence, that on the 5' side of helix III (see Figure 1 cartoon). The known or proposed cut sites for transcript processing in yeast and mammals (9–12) are all in the relatively highly conserved regions in the 5' half of ITS2 (approximately up to the tip of helix III). The significance of actual sequence nucleotides has as yet been explored experimentally only for the pyrimidine–pyrimidine bulge region of helix II in yeast (13).

The hallmarks of ITS2 secondary structure can now be recognized in the major groupings of algae and various other protists as well as animals, fungi and plants (see Table 1). We have derived the major transcript secondary characteristics of many clades, including in column 7 the most highly conserved sequence on the 5' side of helix III, using sequence comparisons. The sequence given in column 7 is common to all of the taxa cited in column 7. Significant omissions from Table 1, either for lack of sufficient sequences in GenBank or from lack of attention to transcript secondary structure, include various lesser known animal phyla, plus reptiles and birds. Among protists, the trypanosomes and euglenoids, some Heterokont groups (Dictyochophytes, Pinguiphytes) and various poorly represented (and some perhaps yet unknown) protistan groups of flagellates and amoeboid types (including the photosynthetic Chlorarachniophytes) are also missing. There is no reason to expect any exceptional structure of their ITS2; for trypanosomes, the locus has already proven useful in delineating species (14).

The utility of the table is severalfold. In total, it summarizes the remarkable uniformity among eukaryotes of the fundamental RNA transcript secondary structure features of ITS2. At a lower taxonomic level, it suggests an

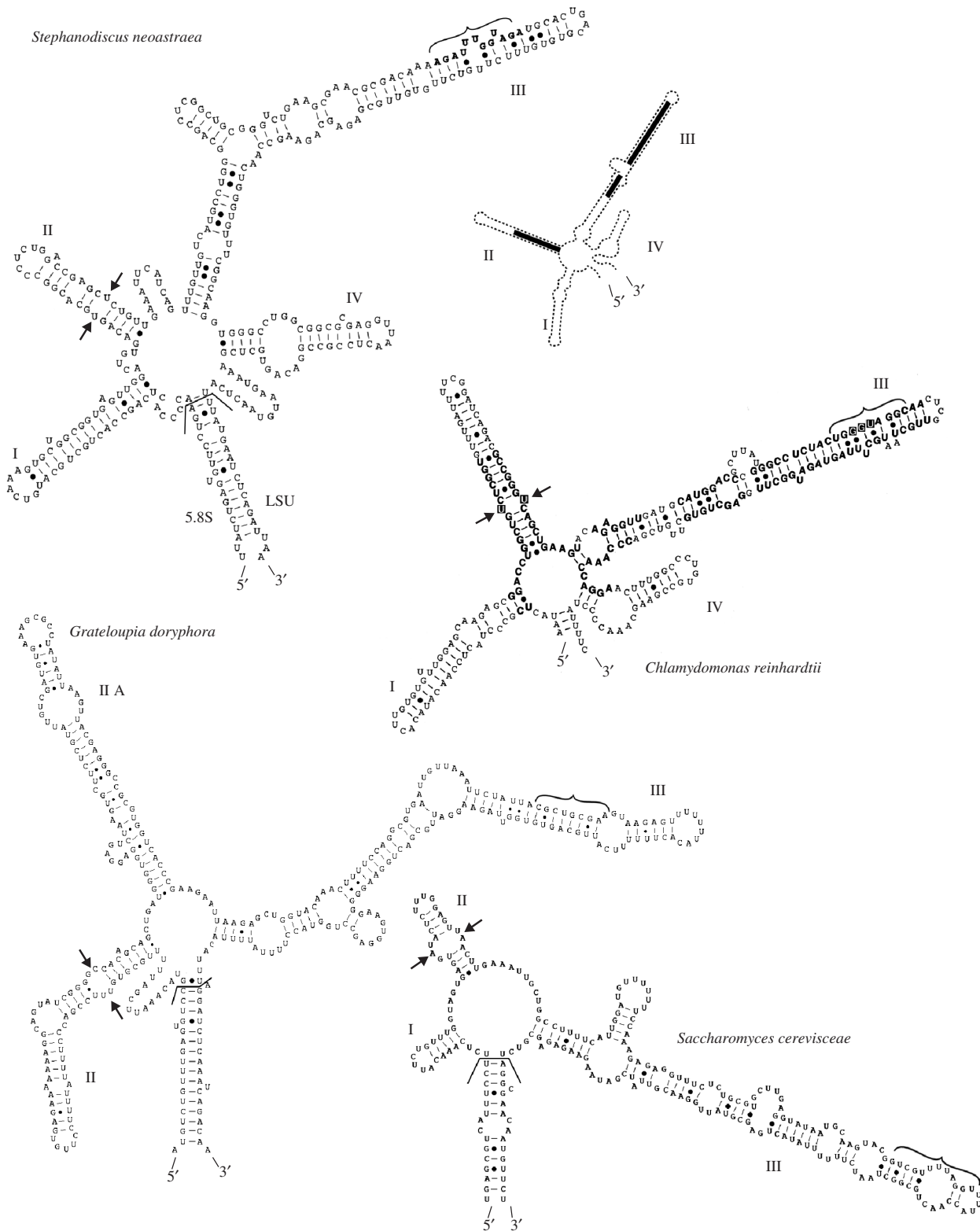


Figure 1. Exemplars of ITS2 RNA transcript secondary structures for a diatom (AF455267), a green alga (U66954), a red alga (AF412018) and yeast (AY130310), based on comparisons of compensatory base changes in pairings. All but *Chlamydomonas* have also the 5.8S-5'LSU association (9). Helix II, with its characteristic pyrimidine-pyrimidine bulge (arrows), is highly conserved in its basal region (see cartoon). Helix III has its most highly conserved sequence region on the 5' side, near the tip (bracket) and its most highly conserved pairings include this region (see cartoon). Cartoon shows with black fill the relatively conserved regions of sequence. It should be noted that the transcript folding pattern for yeast here is slightly different from either of the prior published examples, yet still satisfies the chemical and experimental molecular biology characteristics established for its secondary structure (12,13). The structure for *Stephanodiscus* is the same as in (42) except for helix IV.

Table 1. Comparisons of ITS2 characters among eukaryotes

Group of organisms	Number of helices	IIA?	STD II?	XS nt	2° structure source	Conserved sequence on 5' side of helix III
Green algae						
Prasinophytes	4–5	–	+	No		AGCGTGGTAG4 gen. (11)
Chlorophytes	4	–	+	No	1	GGTAGGY >50gen. (many)
Volvocales	4	–	+	No	5	YRGGTAGGC >25 gen. (many)
Charophytes	4	–	+	No		
Desmids	4	–	+	Yes	8,37	CCGGCGTGGACGA <i>Staurostrum</i> (21)
Siphonaceous greens	3?		+	+/-		
Terrestrial plants	4	–	+	No	1,5,39,40	NRTGGT Angiosperms >150 gen.
Stramenopiles (Heterokonts)						
Brown algae	4	–	+	+/-	41	GYKACGGM >50 gen. (~90)
Diatoms	4	+/-	+	+/-	42	AGRTTTGGTARA e.g. <i>Stephanodiscus</i> 4 gen. (32)
Chloromonads (Raphidiphytes)	4	–	+	+/-		GTGGTAGY 5 gen. (18)
Oomycetes	3–4	–	+	No		YGYGGTATG Peronosporales, Pythiales 4gen. (11)
Chrysophytes	3	–	+	+/-		*
Xanthophytes	4	–	+	+/-		*
Pelagophytes	5?	–	+	No		GAGGCGGGGT <i>Aureococcus</i> , <i>Pelagomonas</i> (4)*
Eustigmatophytes	3	–	+	No		*
Cryptophytes	4	–	+	No		TGTGCCAGCCT <i>Cryptomonas</i> , <i>Chilomonas</i> (16)
Haptophytes (Prymnesiophytes)	3	–	+	No		GTGCTAGY <i>Phaeocystis</i> , <i>Coccolithus</i> 3 gen. (14)
Alveolates						
Dinoflagellates	3–5	+	+	No	43	YGRYRRCAT Peridiniaceae 6 gen. (>15)
Ciliates	2–3	–	+	No	6	RGYRGTACAT Spirotrichea 19 gen. (32)
Red Algae	3–4	+	+	Yes		YGTGCGAA <i>Grateloupia</i> , <i>Gracilaria</i> 8 gen. (41)
Fungi	2–4	+/-	+	No	12,44	GTCGTTTTAGGT e.g. <i>Saccharomyces</i> , 2 gen. (13)
Animals						
Sponges	3–4	Rare	+	+/-	45	CAGCT(T)GGY <i>Leucosolenia</i> , <i>Crambe</i> 2 gen. (4)
Placozoa	4	–	+	No		GTGATTGGTATAGATCAGGC <i>Trichoplax</i> spp. (4)
Myxozoa	4	–	+	+/-		GTTGGTGA <i>Myxobolus</i> spp. (3)*
Comb jellies (Ctenophores)	4	–	+	No	25	CGGYGTGRTAG 10 gen. (18)
Corals (Cnidaria)	3–5	–	+	No	46	GCGRAGGC stony corals 19 gen. (28)
Trematodes	4	–	+	+/-	47,48	TCRTGGYTYART 9 gen. (42)
Nematodes	3–4	+	+	No	11,49	GATGTGRAC Molineoidea, Trichostrongyloidea (7)
Coelomates						
Molluscs	4	+/-	+	+/-	33,50	ARGCTGCGYGGA abalone (19)
Arthropods						
Ticks	5	–	+	Yes	7	GATGAATACTGG <i>Ixodes</i> (17)
Crustacea	5	+	+	Yes	12	GACCGGGYCGG crabs 6 gen. (8)
Insects						
Mosquitoes	3	–	+	No	51,52	GATAGTCAGRCG <i>Aedes</i> (5)
<i>Drosophila</i>	4+	+	+	+/-	34	GTCTAGCATA <i>Drosophila</i> , <i>Musca</i> 5 gen. (22)
Beetles	4–5	+	+	Yes	53	CGATCGTCGTG Chrysomelinae 5 gen. (49)
Echinoderms	4	–	+	Yes	12	CGCGCGGTGCAGG Echinacea 3 gen. (3)
Fish	4–5	+/-	+	Yes	12	YCGGTGGR Neopterygii 12 gen. (15)
Frog	4	–	+	Yes	12	GCGGCTGTCTGTGG <i>Xenopus</i> , <i>Rana</i> (3)
Mammal	4+	–	+	Yes	54	CGGCGCCGCCCCGCGG mice, rat 2 gen. (5)

*=too few ITS2 sequences to provide confirmation by sites of compensatory base changes. Column 1=designation of phylum and/or other higher taxonomic category, using the common English term, where possible, to be user friendly. Column 2=typical total number of helices in ITS2. Column 3=presence or absence of a helix between standard helix II and standard helix III. Column 4=presence of recognizable standard helix II. Column 5=typical ITS2 length, either less or more than ca. 325 nt. Column 6=references containing transcript secondary structure diagrams. Column 7=Where possible, an example, derived from analyses of a multiple alignment, of the very highly conserved nucleotide sequence on the 5' side of helix III, with indication of taxonomic subgroup and span, and number of sequences (parentheses) from which this is derived, if different from column 1. All except for Desmids were derived by the author.

even greater degree of uniformity within a phylum, and then within a class. Our choice of >325 nt as a criterion for excess ITS2 length is entirely arbitrary, providing the user some hint of expectation. There is no obvious correlation between ITS2 length and position in the eukaryote tree, and some phyla contain both long and short examples.

In all phyla, the relatively highly conserved hallmark sequences on the 5' side of helix III in Figure 1 display a certain sameness, a high purine content, particularly guanine, presumably of functional importance to

processing. In the majority of examples in Table 1, a YGGY can be found here. Comparison of transcript foldings at higher taxonomic levels fails to support any clade-constant positioning of pairing versus bulges in this most highly conserved sequence, implying some other aspect as the clue to processing. Finally, the Table serves not only as a guide to expectation when working on a group of taxa, but also as a source of citations concerning detailed RNA folding structure of the phylum, useful for resolving secondary structure in further taxa. These citations also serve as an entree into the phylogenetic

analysis literature. The full significance of the transcript secondary structure parallels will not become apparent until the biochemical details of processing are deciphered.

Exceptions

There are two general types of exception to the rule that ITS2 has a recognizable helix II and III. The first type of exception is that helix II is present, but a recognizable helix III seems not to be. Only three disparate groups exemplify this condition, the coral genus *Acropora* and its immediate relatives (as opposed to the remaining stony corals), at least some of the marine siphonaceous green algae, and four genera (*Tetramitas*, *Vahlkampfia*, *Naegleria* and *Willaertia*) of the Vahlkampfiidae of the Heterolobosea. Whether their total ITS2 region is long or quite short, nothing recognizably comparable to a standard helix III is found. These organisms all presumably undergo normal processing of their initial RNA transcript to produce SSU, 5.8S and LSU RNAs.

The second type of exception is found among a few genera of parasites, the Diplomonadida *Giarda* and the genera of Microsporidia, *Nosema*, *Encephalitozoon* and *Vairiomorpha*. These organisms lack a free 5.8S RNA molecule, instead incorporating its homolog within the 5' end of the LSU gene, as in prokaryotes. The resulting transcript needs no processing of the type requiring guidance by ITS2 structure. In all of these exceptions, the region of the 5.8S and what lies between it and the standard LSU, though short, is still adequate for identifying to species (e.g. 15,16).

DISCUSSION

Internal transcribed spacers

The term ITS traditionally refers to the entire region between the nuclear genes for the ribosomal small subunit (SSU) and the ribosomal large subunit (LSU) RNAs; ITS1 is the first spacer, followed by the 5.8S RNA gene, and then the ITS2. Both the ITS1 and ITS2 regions of the long RNA transcript are removed, by 'processing' enzymes in the nucleolus, which produce the final SSU, 5.8S and LSU RNAs for ribosomes. The 'processing' is not yet fully understood in biochemical terms, but it is clear that the folding pattern, the secondary structure, of the initial RNA transcript plays a role in guiding processing (9–12). Hence, not only the sequence but also the transcript folding patterns of ITS1 and of ITS2 have been objects of study. The second spacer, ITS2, has received vastly more attention, for plants, animals and protists, than the ITS1. More than 100 000 ITS2 sequences can be found in GenBank, and numerous studies have used this region for phylogenetic analyses. When both ITS2 and another locus, plastid or mitochondrial, have been studied in parallel, the ITS2 has been found to contain at least as much information, and usually more (e.g. 17). Its sequence is unique to a species, and usually to the subspecies level.

Basically, the reason for this is that ITS2 combines both remarkably conserved with relatively labile stretches of nucleotides, as shown in the alignment of Hershkovitz and

Lewis (18). The explanation for this apparent alternation of conserved with variable regions was presented by Mai and Coleman (5). The regions of greatest sequence conservation contribute heavily to the pairings in the helices of the secondary structure that the initial RNA transcript takes on, as shown in the cartoon in Figure 1. This secondary structure, in turn, guides processing.

Contributing to the popularity of ITS2 for phylogenetic analyses is its ease of PCR and sequencing. Because its primer regions are in the 5.8S and the 5' region of the large subunit genes of the ribosomal DNA, both utterly essential genes, they are very highly conserved, yet offer the possibility of selecting phylum-specific primers. Also, ITS2 lengths are generally shorter than 350 nt, easily sequenced in one run.

One hesitation affecting phylogenetic ITS2 studies has been the question of alignment of differing sequences, a question that arises because for proteins, with their triplet amino acid signature in the DNA sequence, the alignment is immediately obvious. For DNA regions that do not code for protein products, alignment of subregions where sequence differs considerably presents difficult and often unjustifiable options, a facet that has sometimes led earlier workers to omit the more variable portions of ITS2 from an alignment. This problem has been largely overcome with the recognition of the role of transcript secondary structure, which dictates alignment of paired positions. The alignment can be done either manually or by using computer programs (see below).

Theoretical problems

An initial concern with ITS was the fact that there are typically several hundred copies of the ribosomal RNA locus in tandem in the nuclear genome (19)—hence there was the potential for intragenomic variation. It is now clear that, except as described below, these repeats are subject to 'concerted evolution' (3,20–22), the result of a poorly understood process of homogenization that renders the ITS repeats of an organism identical over very short evolutionary time. Intragenomic variation, if present at all, is typically only in very few extremely variable positions that are never paired in secondary structure. Thus, the ITS2 can be treated as a single gene (1). For an exhaustive analysis of ITS2 intragenomic variation, see Pröschold *et al.* (23).

The other potential problems, discussed at length in Alvarez and Wendel (3) and Bailey *et al.* (24), concern hybridization and polyploidy. Clearly if two organisms differing in their ITS2 sequences hybridize and crossover occurs, the outcome can confuse the phylogenetic analysis. There may ultimately be more than one locus of rDNA genes in the nucleus, there may be mixed ITS2 sequences within and between arrays, resulting from differing parents and from crossingover, and there may even be pseudogene sequences of ITS resulting from degeneration of one set. Non-functional pseudogenes, in fact, are readily recognizable by their imperfect 5.8S and for absence of some or all of the relatively conserved regions of ITS2, aspects obvious from transcript secondary structure knowledge. Hybridization, polyploidy and

their consequences, more common in plants than in animals, can engender confusion, but such situations are generally recognizable and already suspected in particular groups under study. The resulting ITS sequences, containing two or more repeat types, may then actually prove highly informative in resolving the phylogenetic problems.

Bonuses

The conservation of basic hallmarks of ITS2 structure across great taxonomic spans has inspired several novel approaches to handling major taxonomic categories. Recently, Landis and Gargas (submitted for publication) have proposed a method for identification of all fungi to species, using just PCR and a set of sequential 20-mer primers designed from the first half of the ITS2, reflecting the specificity and stability of helix I, II and the 5' portion of III.

For the four-helix model of ITS2, using the hallmarks of the helix II pyrimidine mismatch and the longer helix III with TGGT on the 5' side, Schultz *et al.* (25) have automated GenBank searching for ITS2 sequences and their folding. Subsequently, Wolf *et al.* (2) and Schultz *et al.* (26) have extended automation and set up a website (<http://its2.bioapps.biozentrum.uni-wuerzburg.de>) with eukaryote-wide representation of exemplar ITS2 folds. In addition, there is a growing literature on programs to handle simultaneously sequence alignments and their secondary structure characteristics, as in Siebert and Backofen (27), Seibel *et al.* (28) and Wolf *et al.* (29). Biffen *et al.* (30) provide a recent example of the application of such programs, plus an interesting analysis of the types and rates of compensatory base change in SSU versus ITS. These methods appear to work best for relatively short ITS2 sequences, averaging ~200 nt, and thus have proven particularly applicable to plants and green algae, fungi, dinoflagellates and some metazoan groups.

A unique advantage of the ITS2 as a choice for sequencing is that the resulting alignment contains information related to the level of the biological species (31). This is an empirical observation that has borne up for all eukaryote groups investigated so far. The fundamental correlation arises from the highly conserved regions of sequence. In taxonomic groups with fairly short ITS2 sequences, where these are all identical, the organisms are observed to be able to intercross experimentally, and if no compensatory base change is present, at least to some degree. For groups with longer sequences, the additional regions appear to show lesser evolutionary constraint, so that one must limit the comparison to only the most conserved paired positions (the 10 basal pairings in II and the 18 pairings including and immediately surrounding the highly conserved 5' sequence of helix III); these should be identical, or lack any compensatory base change, for any interbreeding to be possible. Such an analysis requires a plethora of data, not only ITS2 sequences but experimental interbreeding data; yet examples have been found among protists (6,8,32), plants (5) and animals (33,34) to test the hypothesis. An additional

data set arises from the clades of angiosperms endemic to either the Hawaiian Islands or to Macaronesia, where only a limited evolutionary time has been available to produce the endemic genera and species groups now present. Breeding studies have suggested that none of these groups has managed to evolve to the point of sexual isolation (35). Comparisons of the ITS2 sequences, in the nine genus and species swarms where ITS2 is available, now agree that all have the expected ITS2 nucleotide identity (Coleman, in preparation).

In sum, ITS2 is present in essentially all eukaryotes, and even when truncated, the region is still sufficient to allow identification to species and lower. There is no evidence of any horizontal gene transfer (36). ITS2 PCR and sequencing are straight forward, and there is rarely any excessive length. A recognizable short pyrimidine bulge-containing helix ('helix II') and downstream, a longer helix with highly conserved nucleotide motif on the 5' side ('helix III') are essentially universally present. For the full biochemical understanding of how ribosomal RNA processing occurs, the many ITS alignments available should prove invaluable to test models, and those few eukaryotes groups (e.g. *Acropora* corals) lacking helix III and hence its detailed guidance role in processing pose a further challenge since they obviously produce ribosomes. Furthermore, thanks to its conserved secondary structure aspects, one has a guide, from sequence and structure alone, to the group of taxa probably capable of interbreeding. This correlation allows interesting comparisons of breeding potential with the idiosyncracies of taxonomic practice at the species level across eukaryote groups. The ITS2, once considered a highly variable and largely uninteresting locus, has proven in fact to be one containing eukaryote-wide homology, undoubtedly a consequence of its guidance role in what must be a eukaryote-wide biochemistry of ribosome formation.

ACKNOWLEDGEMENTS

This work would not have been possible without the aid of the mfold website, <http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi> and its patient master, Dr M. Zucker of Rensselaer Polytechnic Institute. Funding to pay the Open Access publication charges for this article was provided by A.W.Coleman.

Conflict of interest statement. None declared.

REFERENCES

1. Coleman, A.W. (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.*, **19**, 370–375.
2. Wolf, M., Achtziger, M., Schultz, J., Dandekar, T. and Müller, T. (2005) Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA*, **11**, 1616–1623.
3. Alvarez, I. and Wendel, J.R. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.*, **29**, 417–434.
4. Zuker, M., Mathews, D.H. and Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In Barciszewski, J. and Clark, B.F.C. (eds),

- RNA Biochemistry and Biotechnology*, NATO ASE Series. Kluwer Academic Publishers, Hingham, MA, pp. 11–43.
5. Mai, J.C. and 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.
 6. Coleman, A.W. (2005) *Paramecium aurelia* revisited. *J. Eukaryotic Microbiol.*, **52**, 68–77.
 7. Hlinka, O., Murrell, A. and Barker, S.C. (2002) Evolution of the secondary structure of the rRNA internal transcribed spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda). *Heredity*, **88**, 275–279.
 8. Denboh, T., Ichimura, T., Hendrayanti, D. and Coleman, A.W. (2003) *Closterium moniliferum-ehrenbergii* (Charophyceae, Chlorophyta) species complex viewed from the 1506 Group I intron and ITS2 of nuclear rDNA. *J. Phycol.*, **39**, 960–977.
 9. Veldman, G.M., Klootwijk, J., van Heerikhuizen, H. and 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.
 10. Van Nues, R.W., Rientjes, J.M.J., Morre, S.A., Mollee, E., Planta, R.J., Venema, J. and Raue, H.A. (1995) Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. *J. Mol. Biol.*, **250**, 24–36.
 11. Hung, G.-C., Chilton, N.B., Beveridge, I. and Gasser, R.B. (1999) Secondary structure model for the ITS-2 precursor rRNA of strongloid nematodes of equids: implications for phylogenetic inference. *Int. J. Parasitol.*, **29**, 1949–1964.
 12. Joseph, N., Krauskopf, E., Vera, M.I. and Michot, B. (1999) Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res.*, **27**, 4533–4540.
 13. Cote, C.A., Greer, C.L. and Beculis, B.A. (2002) Dynamic conformational model for the role of ITS2 in pre-RNA processing in yeast. *RNA*, **8**, 786–797.
 14. Beltrame-Botelho, I.T., Gaspar-Silva, D., Steindel, M., Davila, A.M.R. and Grisard, E.C. (2005) Internal transcribed spacers (ITS) of *Trypanosoma rangeli* ribosomal DNA (rDNA): a useful marker for inter-specific differentiation. *Infect. Genet. Evol.*, **5**, 17–28.
 15. Peyretailade, E., Biderre, C., Peyret, P., Duffieux, F., Metenier, G., Gouy, M., Michot, B. and Vivares, C.P. (1998) Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. *Nucleic Acids Res.*, **26**, 3513–3520.
 16. Gatehouse, H.S. and Malone, L.A. (1998) The ribosomal RNA gene region of *Nosema apis* (Microspora): DNA sequence for small and large subunit rRNA genes and evidence of a large tandem repeat unit size. *J. Invert. Pathol.*, **761**, 97–105.
 17. Duran, S., Giribet, G. and Turon, X. (2004) Phylogeographical history of the sponge *Crambe crambe* (Porifera, Poecilosclerida): range expansion and recent invasion of the Macaronesian islands from the Mediterranean Sea. *Mol. Ecol.*, **13**, 109–122.
 18. Hershkovitz, M.A. and Lewis, L.A. (1996) Deep-level diagnostic value of the rDNA-ITS region. *Mol. Biol. Evol.*, **13**, 1276–1295.
 19. Prokopowich, C.E., Gregory, T.R. and Crease, T.J. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, **46**, 48–50.
 20. Arnheim, N. (1983) Concerted evolution in multigene families. In Nei, M. and Koehn, R. (eds), *Evolution of Genes and Proteins*, Sinauer, Sunderland, MA, pp. 38–61.
 21. Elder, J.F. and Turner, B.J. (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.*, **70**, 297–320.
 22. Ganley, A.R.D. and Kobayashi, T. (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.*, (online).
 23. Pröschold, T., Harris, E.H. and Coleman, A.W. (2005) Portrait of a species: *Chlamydomonas reinhardtii*. *Genetics*, **170**, 1601–1610.
 24. Bailey, C.D., Carr, T.G., Harris, S.A. and Hughes, C.E. (2003) Characterization of angiosperm rDNA polymorphism, paralogy, and pseudogenes. *Mol. Phylogenet. Evol.*, **29**, 435–455.
 25. Schultz, J., Maisel, S., Gerlach, D., Müller, T. and Wolf, M. (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA*, **11**, 361–364.
 26. Schultz, J., Müller, T., Actziger, M., Seibel, P.N., Dandekar, T. and Wolf, M. (2006) The internal transcribed spacer 2 Database - a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res.*, **34**, 704–707.
 27. Siebert, S. and Backofen, R. (2005) MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics*, **205**, 3352–3359.
 28. Siebel, P.N., Müller, T., Dandekar, T., Schultz, J. and Wolf, M. (2006) 4SALE: a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC. Bioinformatics*, **7**, 498.
 29. Wolf, M., Friedrich, T., Dandekar, T. and Müller, T. (2005) CBCAnalyser: inferring phylogenies based on compensatory base changes in RNA secondary structures. *In Silico Biol.*, **5**, 0027.
 30. Biffin, E., Harrington, M.G., Crisp, M.D., Craven, L.A. and Gadek, P.A. (2006) Structural partitioning, paired-sites models and evolution of the ITS transcript in *Syzygium* and Myrtaceae. *Mol. Phylogenet. Evol.*, **22**, (Aug).
 31. Coleman, A.W. (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist*, **151**, 1–9.
 32. Fabry, S., Kohler, A. and Coleman, A.W. (1999) Intraspecific analysis: comparison of ITS sequence data and gene intron sequence data with breeding data for a world-wide collection of *Gonium pectorale*. *J. Mol. Evol.*, **48**, 94–101.
 33. Coleman, A.W. and Vacquier, V.D. (2002) Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (*Haliotis*). *J. Mol. Evol.*, **54**, 246–257.
 34. Young, I. and Coleman, A.W. (2004) The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Mol. Phylogenet. Evol.*, **30**, 236–242.
 35. Stuessy, T.F. and Ono, M. (1998) *Evolution and Speciation of Island Plants* Cambridge University Press, NY, USA.
 36. Rothschild, L.J., Ragan, M.A., Coleman, A.W., Heywood, P. and Gerbi, S.A. (1986) Are rRNA sequence comparisons the rosetta stone of phylogenetics? *Cell*, **47**, 640.
 37. Van Hannen, E.J., Fink, P. and Lurling, M. (2002) Revised ITS-2 secondary structure model for *Scenedesmus* and *Desmodesmus*. *Eur. J. Phycol.*, **37**, 203–208.
 38. Gontcharov, A.A. and Melkonian, M. (2005) Molecular phylogeny of *Staurostrum* Meyen ex Ralfs and related genera (Zygnematophyceae, Streptophyta) based on coding and noncoding rDNA sequence comparisons. *J. Phycol.*, **41**, 887–899.
 39. Goertzen, L.R., Cannone, J.J., Gutell, R.R. and Jansen, R.K. (2003) ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Mol. Phylogenet. Evol.*, **29**, 216–234.
 40. Denduangboripant, J. and Cronk, Q.C.B. (2001) Evolution and alignment of the hypervariable arm 1 of *Aeschynanthus* (Gesneriaceae) ITS2 nuclear ribosomal DNA. *Mol. Phylogenet. Evol.*, **20**, 163–172.
 41. Yotsukura, N., Denboh, T., Motomura, T., Horiguchi, T., Coleman, A.W. and Ichimura, T. (1999) Little divergence in ribosomal DNA internal transcribed spacer -1 and -2 sequences among non-digitate species of *Laminaria* (Phaeophyceae) from Hokkaido, Japan. *Phycol. Res.*, **47**, 71–80.
 42. Wolf, M. (2004) The secondary structure of the ITS2 transcript in *Cyclotella* and *Stephanodiscus* (Thalassiosiraceae, Bacillariophyta). *Diatom Res.*, **19**, 135–142.
 43. Gottschling, M. and Plotner, J. (2004) Secondary structure models of the nuclear internal transcribed spacer regions and 5.8S rRNA in Calciodinelloideae (Peridiniaceae) and other dinoflagellates. *Nucleic Acids Res.*, **32**, 307–315.
 44. Hausner, G. and Wang, X. (2005) Unusual compact rDNA gene arrangements within some members of the Ascomycota: evidence for molecular co-evolution between ITS1 and ITS2. *Genome*, **48**, 648–660.
 45. Schmitt, S., Hentschel, U., Zea, S., Dandekar, T. and Wolf, M. (2005) ITS-2 and 18S rRNA gene phylogeny of Aplysiniidae (Verongida, Demospongiae). *J. Mol. Evol.*, **60**, 327–336.

46. Odorico, D.M. and Miller, D.J. (1997) Variation in the ribosomal internal transcribed spacers and 5.8S RNA among five species of *Acropora* (Cnidaria; Scleractinia): patterns of variation consistent with reticulate evolution. *Mol. Biol. Evol.*, **14**, 465–473.
47. Michot, B.L., Despres, L., Bonhomme, F. and Bachellerie, J.-P. (1993) Conserved secondary structures in the ITS2 of trematode pre-rRNA. *FEBS Lett.*, **316**, 247–252.
48. Morgan, J.A.T. and Blair, D. (1998) Trematode and monogenean rRNA ITS2 secondary structures support a four-domain model. *J. Mol. Evol.*, **47**, 406–419.
49. Chilton, N.B. and Gasser, R.B. (1999) Sequence differences in the internal transcribed spacers of DNA among four species of hookworm (Ancylostomatoidea: *Ancylostoma*). *Int. J. Parasitol.*, **29**, 1971–1977.
50. Oliviera, M., Cervelli, M. and Mariottini, P. (2002) ITS2 rRNA evolution and its congruence with the phylogeny of muricid neogastropods (Caenogastropoda, Muricoidea). *Mol. Phylogenet. Evol.*, **25**, 63–69.
51. Wesson, D.M., Porter, C.H. and Collins, F.H. (1992) Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). *Mol. Phylogenet. Evol.*, **1**, 253–269.
52. Severini, C., Silvestrini, F., Mancini, P., LaRosa, G. and Marinucci, M. (1996) Sequence and secondary structure of the rDNA second internal transcribed spacer in the sibling species *Culex pipiens* L. and *Cx. quinquefasciatus* Say (Diptera: Culicidae). *Insect Mol. Biol.*, **5**, 181–186.
53. Gomez-Zurita, J., Juan, C. and Petitpierre, E. (2000) 2000 Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the *Timarcha* leaf beetles (Coleoptera: Chrysomelidae). *Insect Mol. Biol.*, **9**, 591–604.
54. Michot, B., Joseph, N., Mazan, S. and Bachellerie, J.-P. (1999) Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. *Nucleic Acids Res.*, **27**, 2271–2282.