Molecular Phylogenetics and Evolution 50 (2009) 197–203

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Review

Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide

Annette W. Coleman *

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

ARTICLE INFO

Article history:
Received 18 March 2008
Revised 6 October 2008
Accepted 8 October 2008
Available online 21 October 2008

Keywords:
Biological species
Interbreeding limits
ITS-2
rRNA secondary structure
Speciation
Taxonomy

ABSTRACT

DNA sequences, powerful for phylogeny, have not yet proven as rewarding for taxonomic categorization purposes. However, further analyses of one locus, the second Internal Transcribed Spacer of the nuclear ribosomal gene cistron, has suggested a high degree of predictability across eukaryotes. Comparison of the secondary structure of ITS2 transcripts reveals its most conserved region, on the 5′-side of helix III. Comparison of this 5′ 30 bp highly conserved region with the extent of sexual compatibility in a clade of organisms produces two useful predictions: identity of this region predicts meaningful intercrossing ability, and, difference of even one CBC pairing in this region predicts total failure of crossing. Previous to the appearance of the first CBC in the highly conserved portion, all gametic compatibility has been lost, thanks to the parallel evolutionary changes in genes controlling mating. These two landmark events help to delimit the level of interbreeding taxa.

1. Introduction

The very basis of our taxonomic system and its growth over the past 250 years is the presumption that individuals very similar in morphology are those most closely related, constituting a species. Yet the category of morphological "species" is in the eye of the beholder. Nature does not produce species, but rather individuals, populations and clades, of continuing evolutionary change. The second major approach to classification states that individuals capable of interbreeding are each others' nearest relatives, as so importantly emphasized by Dobzhansky (1940) and his coworkers more than 50 years ago in their recognition of the "biological" species. They emphasized the critical significance of the loss of interbreeding potential, the major step that dooms clades to remain genetically separate forever after. Hence, in some ideal taxonomic world, one morphology = one interbreeding group = one species; such an entity would be morphologically uniform, and sexually isolated from other clades.

Such is not our world. Many individual populations/clades are indeed, at this point in time, uniform in morphology and capable of interbreeding, and are also unable to cross with individuals of other clades. For many other clades, however, morphological uniformity does not completely parallel interbreeding capacity, or breeding status of a morphological group may remain uncertain/undetermined, challenging the taxonomist to seek other critera.

Given the known pitfalls of our primarily morphology-based taxonomic system, how great then must have been the hopes of taxonomists, with the advent of DNA sequencing, to find at last a molecular surrogate for some level of the speciation process. Alas, this hope has been drowned in the accumulation of sequencing data on various candidate loci. Nucleotide change does occur, genera differ more than do their component pairs of species, but the nucleotide change is continuous, with no gap, no point of reference correlatable with some facet of speciation.

One approach would be to concentrate on understanding the genes involved in sexual behavior, particularly those controlling gamete approach and fusion; this has the advantage of focussing on a process common to all eukaryotes. Activities such as the response to mutual attraction by modification of gamete surface components, specific agglutination, and gamete fusion, require the coordinated action of a considerable number of genes. Such genes might seem to be the ideal object of study for evaluating the evolution of the speciation process. However, our understanding of their comparative molecular aspects is still in its infancy (Swanson and Vacquier, 2002; Hamaji et al., 2008).

Recently, one such DNA region (the Internal Transcribed Spacer 2 region, ITS2, of the nuclear ribosomal cistrons) has been found to vary in sequence and secondary structure in a way that correlates highly with taxonomic classification. The ITS2 region of nuclear DNA provides a powerful tool because not only can sequence be compared but also aspects of the secondary structure formed by the initial RNA transcript. Müller et al. (2007) compared the entire ITS2 of 1373 sequences of species to see how homogeneous were the sequences within species compared to their nearest congeners.

* Fax: +1 401 863 2421.
E-mail address: Annette_Coleman@brown.edu

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.10.008

They found that for 93% of cases where two taxa differed somewhere in their ITS2 by a CBC (Compensatory Base Change, an altered pairing in a helix of the secondary structure of the ITS2 RNA transcript—see Fig. 1), they were classified as different species. Thus, in the vast majority of organisms they surveyed, a discrete property of the ITS2 (no CBC anywhere) characterized members of a species.

This result suggests an outer limit to the sequence variation accompanying current taxonomic assignment of species names, whatever their criteria. However, it says nothing about the level of biological species. Here we use the ITS2 region to explore further its information content, concentrating on characteristics that accompany separation of clades into two or more sexually isolated subclades. We evaluate its predictive powers for discriminating interbreeders, and derive a simplified sequenced-based technique for its application to evaluation of the biological species level among eukaryotes.

2. Properties of ITS2

ITS2 sequence comparisons are perhaps the most common source of phylogenetic reconstructions at the species, genus and family level among all eukaryotes (Alverez and Wendel, 2003; Bailey et al., 2003). The locus of interest is in the nuclear ribosomal cistrons, a region lying between the 5.8S RNA gene and the LSU RNA gene. The cistron is transcribed initially into a single long transcript. During "processing" in the nucleolus to release the ribosomal RNAs, this spacer region is degraded to nucleotides.

It was early reconized that ITS2 shares common features of sequence and secondary structure, conserved even above the family level (Hershkovitz and Zimmer, 1996). Analysis of the secondary structure formed by the RNA transcript as it folds back on itself at transcription, has generally been less commonly done, but has proven extremely useful in aiding proper sequence alignment (Mai and Coleman, 1997). As shown in Fig. 1, the typical ITS2 of eukaryotes forms four helices (Coleman, 2003). This secondary structure can be derived by applying the RNA folding program Mulfold (http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi) to all the sequences of a group of closely related organisms, comparing the multiple folds produced by the program to find the single type common to all the taxa, and then checking positions of nucleotide variation. Essentially all variation in positions paired in helices (Fig. 1) should show only those substitutions that preserve the pairing, both one-sided (hemi-CBC) and two sided (CBC). These characteristics are considered "proof" of the proposed RNA secondary structure, as first applied to studies of the ribosomal RNAs (Gutell et al., 1994). The realization that the same four helix structure characterizes the ITS2 transcript in essentially all eukaryotes has only been recognized recently (Coleman, 2003, 2007; Müller et al., 2007). This universality has led to computer automation of

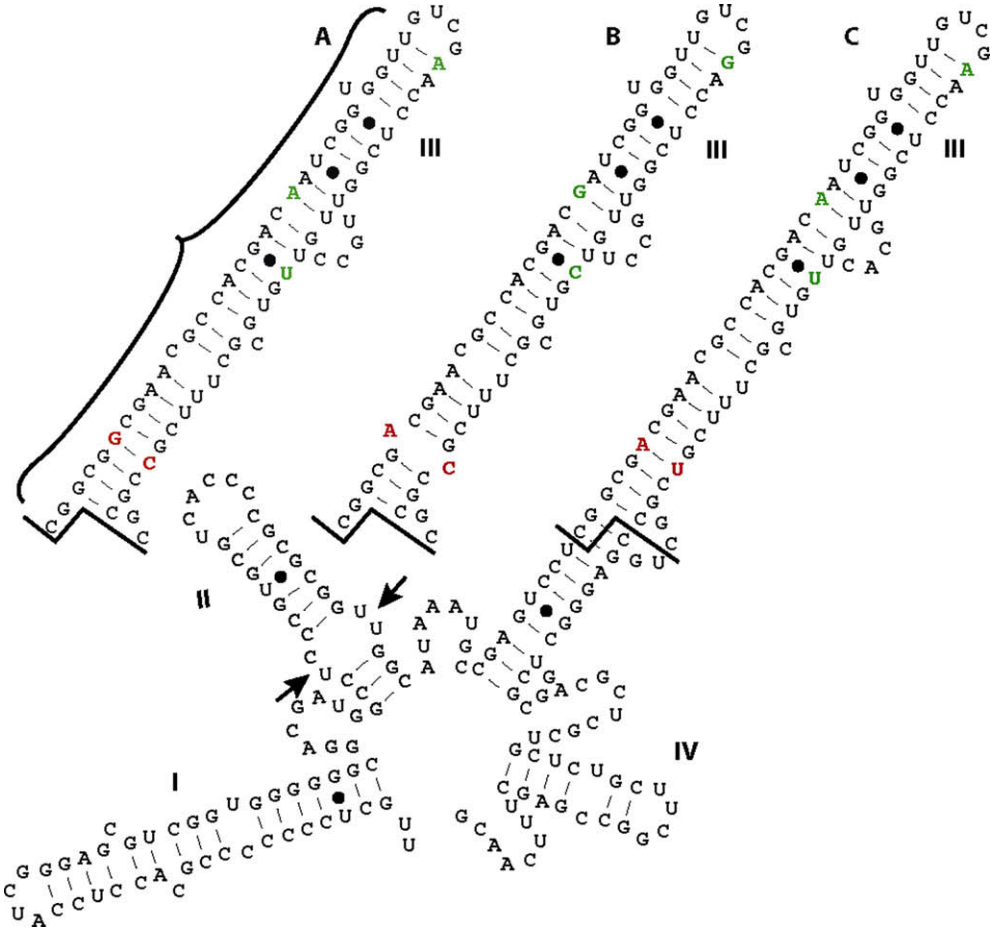

Fig. 1. Secondary structure of ITS2, and comparisons of helix III among examples from the apple subfamily Maloideae, (A) *Amelanchier*, (B) *Mespilus* and (C) *Photinia*. The entire ITS2 structure of *Photinia* is shown to illustrate the typical ITS2 secondary structure of eukaryotes; hallmark characteristics are marked in helix II by arrowheads (pyrimidine–pyrimidine bulge) and in helix III by a bracket (5′ 30 most conserved nucleotide positions). Helix III pairing variants among species of the family include one CBC (red) and three hemi-CBCs (green). *Amelanchier* fails to cross with any of the several genera that differ by the CBC (Robertson et al., 1991). (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)