RESEARCH NOTES

Genomic nucleotide variation in the ITS1 rDNA spacer of land snails

Georg F. J. Armbruster[1] and Anja Korte[2]

[1]*University of Basel, Department of Integrative Biology, Conservation Biology Unit, St. Johanns-Vorstadt 10, CH - 4056 Basel, Switzerland;*
[2]*University of Leipzig, Institute of Zoology, Department of Molecular Evolution & Animal Systematics, Talstr. 33, 04103 Leipzig, Germany*

Ribosomal RNA is coded by the rDNA gene cluster. Hundreds of rDNA loci are consecutively arranged and form repeated motifs on chromosomes. The Internal Transcribed Spacer 1 (ITS1) is part of the eukaryotic rDNA locus. This spacer is a relatively fast evolving, 'non-coding' element. It is often used in phylogenetic reconstructions of closely related taxa. The ITS1 separates two functional regions: the 18S rDNA and the 5.8S rDNA (Fig. 1).

There are usually no or very few ITS1 nucleotide polymorphisms within an animal species (0–3% intraspecific variation[1–4]). The phenomenon of very low intraspecific rDNA variation is explained by the concept of concerted evolution in the gene pool of a species under molecular drive.[5–7] Yet, there are exceptions from the rule of very low intraspecific rDNA variation, particularly in crustaceans[8] and in some species-flocks. For example, in the *Cicindela dorsalis* complex (tiger beetle), in the *Simulium damnosum* group (West African black flies), in corals with reticulate evolution and in a few plant species the genomic variation in the ITS1 is extremely high, with some sequences showing as much as 20–35% intraspecific divergence.[9–12] In the following contribution, we analyse the extent of ITS1 variation of four taxonomically robust species of land snail of different family categories, i.e. within *Acanthinula aculeata* (Valloniidae *sensu lato*), *Arianta arbustorum* (Helicoidea), *Cochlodina fimbriata* (Clausiliidae), and *Vallonia costata* (Valloniidae *sensu stricto*). We show that one species (*A. aculeata*) exhibits a high level of ITS1 variation, whilst another species (*V. costata*) displays nearly no variation. The other two taxa display moderate variation. We also examine the ITS1 rDNA for highly conserved sequence motifs and investigate the extent of nucleotide variation within such motifs. Evolutionarily conserved regions in ribosomal spacers are suspected to be involved in the maturation of the precursor rRNA.[13–15] Finally, we discuss the possible occurrence of rDNA pseudogenes in PCR amplicons and genome databases.

DNA was extracted from 13 snail specimens. Polymerase chain reaction (PCR) of the ITS1 rDNA was performed with flanking 18S and 5.8S primers. PCR products were cloned in pCR blunt vectors (INVITROGEN), transformed in *E. coli*, and isolated and sequenced in both directions (Accession Nos AY267026-AY267030, AF124052-AF124053, AY546422-AY546468). Number of analysed sequences per species was as follows: *A. aculeata* = 23 cloned ITS1 molecules obtained from three individuals (collected in Germany and Switzerland), *A. arbustorum* = 17 from four individuals (two unknown Dutch sites, Switzerland and Estonia), *C. fimbriata* = 9 from one individual (Switzerland), and *V. costata* = 5 from five individuals (Germany, Great Britain and Austria). Details on sampling sites, DNA extraction and PCR are available in the Electronic Supplementary Data.

Correspondence: G. F. J. Armbruster; e-mail: g.armbruster@unibas.ch

The consensus alignment of the 54 sequences is 662 bp long, and includes ITS1 sequences of 522 bp to 599 bp (see Supplementary Data and EMBL/EBI Alignment Accession No. ALIGN_000906). Genetic distances (D) were calculated among all cloned sequences from the ITS1 regions using Kimura's 2-parameter method implemented in MEGA, version 2.1.[16] The D values give information about genomic variability within the samples. The highest genomic D was 0.061 (within *A. aculeata*), whilst *A. arbustorum*, *C. fimbriata* and *V. costata* showed lower D values (with maxima of 0.022, 0.018 and 0.002, respectively). A neighbour-joining tree was reconstructed using MEGA (Fig. 2). Each sequence was correctly assigned to its predicted species level, and the interspecific signal was not confounded by genomic D values (Fig. 2).

Approximately 190 nucleotide positions of the ITS1 are invariant in all cloned sequences (i.e. 32–36% of the alignment length). The invariant nucleotides are conserved among species from the Helicoidea, Clausiliidae and Valloniidae. The last common ancestor of these three families lived in the Upper-Mesozoic, >70 million years ago (for origin of extant land snail family lineages, see Ref. 17). It is possible that these invariant ITS1 nucleotide positions are evolving under functional constraints, i.e. they could be involved in the processing pathway of the precursor rRNA.

Subsequently, the ITS1 was analysed on variable nucleotide positions: (1) in phylogenetically very 'old' and conserved regions; and (2) in phylogenetically unconserved segments. (1) Six highly conserved regions of the stylommatophoran-basommatophoran split 150 million years ago[17] have been found previously.[18] These six regions were concatenated and are referred to as 'StyloBaso' block (black boxes in Fig. 1). These six regions can be aligned across stylommatophoran land snails and basommatophoran freshwater snails,[18] and are relatively constant in length. The concatenated StyloBaso box ranged in length from 110 bp (in *A. aculeata*) to 115 bp (in *C. fimbriata*). In *A. aculeata*, 18 variable positions (nucleotide substitutions and indels) occurred in the StyloBaso region (Table 1). Thus, even in sequence motifs of 150 million years, intraspecific ITS1 variation was found. (2) In the remaining part of the ITS1 of *A. aculeata* (length between 450 and 480 bp), 84 variable sites were detected. This means that a total of 102 polymorphic nucleotide positions occurred in the ITS1 of this species, i.e. 17–19% of the nucleotide positions were variable (see Table 1). The other land snail species showed lower percentage values of variable sites, with 0.2% (*V. costata*), 3% (*C. fimbriata*), and 7% (*A. arbustorum*). These percentage values are only estimates because the probability of detection of variable nucleotide positions depends on the number of cloned sequences, and on sequence length. We therefore performed a partial rarefaction analysis (with the MEGA sequence editor). As target groups we focused on *A. aculeata* (23 sequences; ITS1 alignment length = 603 bp) and

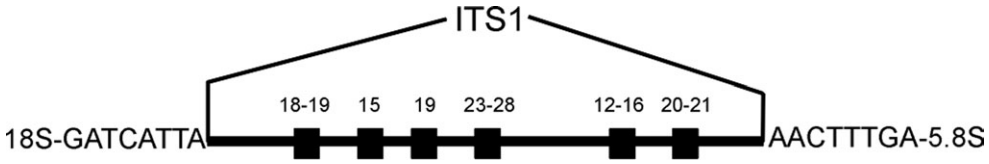

Figure 1. The ITS1 region with flanking 18S and 5.8S rDNA. The 3′-end of the 18S rDNA exhibits the animal-specific motif GATCATTA. The 5′-terminus of the 5.8S rDNA starts in land snails with the signal AACTTTGA. Relative position and length of the six 150-million-year-old StyloBaso blocks are shown with black boxes (length of each region in bp indicated above each box; see text and Table 1). The StyloBaso consensus regions are inherited from the last common ancestor of terrestrial pulmonate gastropods (Stylommatophora) and pulmonate freshwater snails (Basommatophora).[18]

A. arbustorum (17 sequences; ITS1 alignment length = 589 bp). In each species, we analysed two randomly chosen sequences and divided the number of observed polymorphic nucleotides by alignment length. This was repeated with twenty replicates for each species. Finally, we calculated mean and standard deviation for both species. The mean value of the twenty replicates in *A. aculeata* was 0.029 (SD = 0.012), whilst *A. arbustorum* had a mean value of 0.016 (SD = 0.007). The rarefaction analysis indicated that there were significant differences in nucleotide variation between *A. aculeata* and *A. arbustorum* (*t*-test with unequal variance, $P < 0.001$).

Focusing on the concept of concerted evolution, the question arises of how effective is genomic homogenization of the ITS1 rDNA? In experimental studies, a relaxation of concerted evolution was found in arbuscular mycorrhizal fungi.[19] Also in the land snail *A. aculeata*, concerted evolution seems both relaxed and incomplete (with high nucleotide variation in PCR amplicons of three individuals from Switzerland and Germany). *Arianta arbustorum* and *C. fimbriata* also showed a tendency towards incomplete homogenization. In *V. costata*, however, concerted evolution and genomic homogenization is nearly complete, indicated by just a single polymorphic nucleotide in five

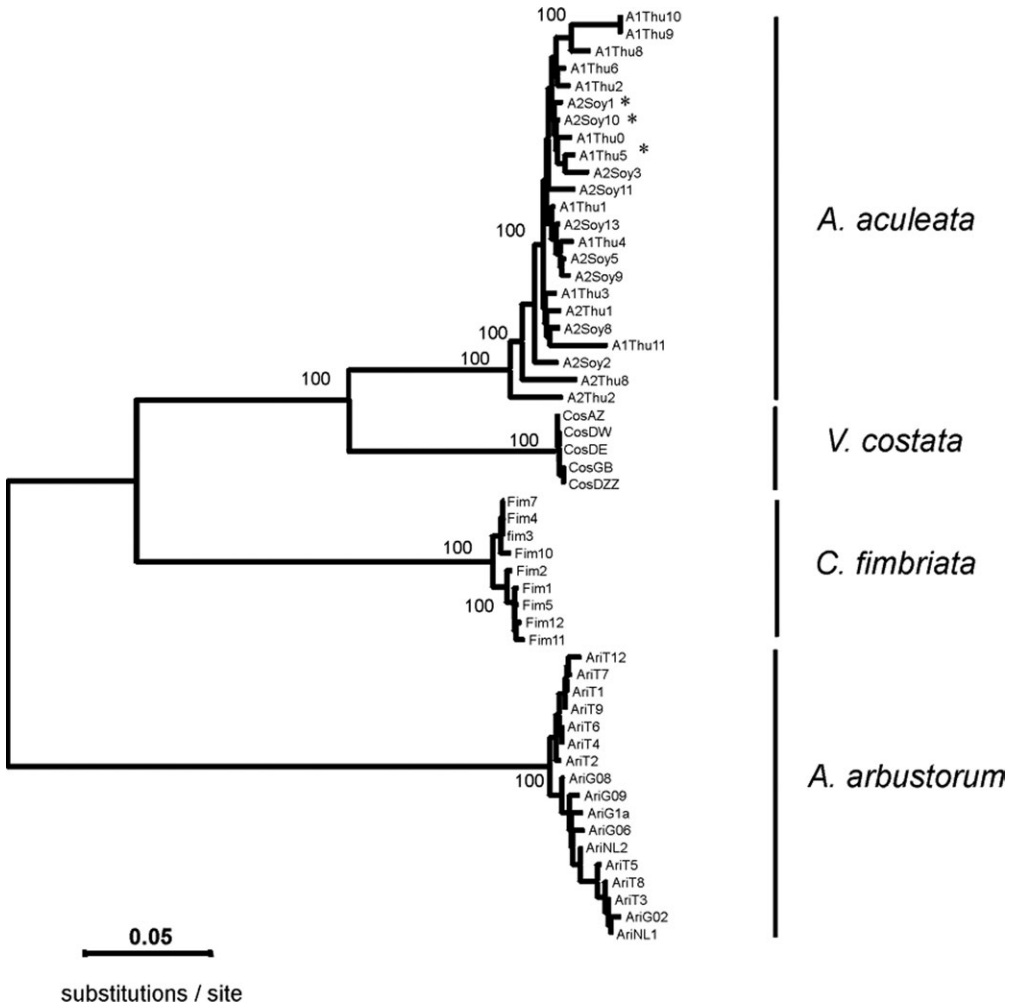

Figure 2. Neighbour-joining tree of 54 cloned ITS1 sequences of four land snail species. Bootstrap values higher than 95% are indicated (based on 1000 replicates). Asterisks display clones with a complete lack of the flanking 18S termination signal GATCATTA, i.e. putative genomic pseudogenes. Names refer to cloned sequences; see Supplementary Data for detailed information.

Table 1. Summary of variable nucleotide positions in the ITS1 rDNA of the four species (see text).

Species (no. of individuals analysed/no. of cloned sequences)	Consensus regions of land snails and basommatophoran freshwater snails (six black boxes)		Remaining part of the ITS1	
	Concatenated length (bp)	Variable positions	Concatenated length (bp)	Variable positions
A. aculeata (3/23)	110–113	18	450–480	84
A. arbustorum (4/17)	111–112	11	443–462	32
C. fimbriata (1/9)	115	2	450–452	15
V. costata (5/5)	114	1	407	0

sequences from individuals from Great Britain, Austria and Germany (see Supplementary Data).

Beside relaxed concerted evolution, an additional explanation for the observed intragenomic ITS1 variation may lie elsewhere. Polymorphisms within genomes may be attributed to solitary, incognito, pseudogene variants that 'escape' the homogenization process of the rRNA gene cluster. Several studies focus on this topic, for example in plants[12,20] and humans.[21–22] Three sequences in our data set could have been incognito variants since they show a lack of the animal-specific termination motif GATCATTA of the 18S rDNA (see Fig. 1; Accession Nos AY546427, AY546443, AY546444). These three sequences are marked with an asterisk in Figure 2. A lack of this termination motif provides evidence that rDNA pseudogene sequences could have been cloned from the PCR amplicons. Yet, all ITS1 sequences do group conspecifically (Fig. 2), pointing to little background noise in phylogenetic reconstructions. We therefore suggest that signals of rDNA pseudogenes should be further studied.

In summary, the high number of invariant ITS1 nucleotide positions could make land snails potential model organisms for further investigations of ITS1 rRNA processing. Nevertheless, efficiency of concerted evolution of ITS1 rDNA seems to vary between land snail species. Extent of ITS1 variation should always be elucidated, particularly if direct sequencing is used and 'unreadable' sequencing profiles appear from PCR templates. The occurrence of rDNA pseudogenes requires further study. For example, in the human genome database we found a full-length 18S copy on the Y-chromosome, which is an escaped variant of the rRNA gene cluster of chromosome number 13, 14, 15, 21 or 22 (data are available on request). The Y-chromosomal sequence shows 89% nucleotide identity to the original autosomal 18S rDNA (see Ref. 22: pp. 894-895). Hence, escaped gene copies can show a relatively high similarity to their orthologous counterparts. Such copies might also appear in PCR amplicons.

ACKNOWLEDGEMENT

We thank B. Gautschi and B. Koller (Ecogenics GmbH, Schlieren, Switzerland) for excellent cooperation. Three reviewers gave valuable help to improve the manuscript. The study was funded by the NLU budget of the Conservation Biology Unit of the University of Basel.

REFERENCES

1. SCHLÖTTERER, C., HAUSER, M.T., VON HAESELER, A. & TAUTZ, D. 1994. *Mol. Biol. Evol.*, **11**: 513–522.
2. ZHUO, L., SAJDAK, S.L. & PHILLIPS, R.B. 1994. *Genome*, **37**: 664–671.
3. HOLDER, K., MONTGOMERIE, R. & FRIESEN, V.L. 1999. *Evolution*, **53**: 1936–1950.
4. GANDOLFI, A., BONILAURI, P., ROSSI, V. & MENOZZI, P. 2001. *Heredity*, **87**: 449–455.
5. DOVER, G. 1982. *Nature*, **299**: 111–117.
6. ELDER, J.F. Jr, & TURNER, B.J. 1995. *Quart. Rev. Biol.*, **70**: 297–320.
7. DOVER, G. 2002. *Trends Gen.*, **18**: 587–589.
8. HARRIS, D.J. & CRANDALL, K.A. 2000. *Mol. Biol. Evol.*, **17**: 284–291.
9. VOGLER, A.P. & DESALLE, R. 1994. *Mol. Biol. Evol.*, **11**: 393–405.
10. TANG, J., TOE, L., BACK, C. & UNNASCH, T.R. 1996. *Mol. Biol. Evol.*, **13**: 244–252.
11. ODORICO, D.M. & MILLER, D.J. 1997. *Mol. Biol. Evol.*, **14**: 465–473.
12. BUCKLER IV, E.S., IPPOLITO, A. & HOLTSFORD, T.P. 1997. *Genetics*, **145**: 821–832.
13. MAI, J.C. & COLEMAN, A.W. 1997. *J. Mol. Evol.*, **44**: 258–271.
14. JOSEPH, N., KRAUSKOPF, E., VERA, M.I. & MICHOT, B. 1999. *Nucl. Acids Res.*, **27**: 4533–4540.
15. VENEMA, J. & TOLLERVEY, D. 1999. *Ann. Rev. Genet.*, **33**: 261–311.
16. KUMAR, S., TAMURA, K., JAKOBSEN, I.B. & NEI, M. 2001. *Bioinformatics*, **17**: 1244–1245.
17. TILLIER, S., MASSELOT, M. & TILLIER, A. 1996. In: *Origin and evolutionary radiation of the Mollusca*. (J.D. Taylor ed.), 267–284. Oxford University Press, Oxford.
18. ARMBRUSTER, G.F.J., VAN MOORSEL, C.H.M. & GITTENBERGER, E. 2000. *J. Moll. Stud.*, **66**: 570–573.
19. PAWLOWSKA, T.E. & TAYLOR, J.W. 2004. *Nature*, **427**: 733–737.
20. RUGGIERO, M.V. & PROCACCINI, G. 2004. *J. Mol. Evol.*, **58**: 115–121.
21. GONZALEZ, I.L. & SYLVESTER, J.E. 1997. *Genome Res.*, **7**: 65–70.
22. INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2001. *Nature*, **409**: 860–921.

doi:10.1093/mollus/eyi056
Advance Access Publication: 23 September 2005