

Evolution of the secondary structure of the rRNA internal transcribed spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda)

O Hlinka, A Murrell and SC Barker

Department of Microbiology and Parasitology, and Institute for Molecular Biosciences, The University of Queensland, Brisbane, 4072, Australia

ITS2 sequences are used extensively in molecular taxonomy and population genetics of arthropods and other animals yet little is known about the molecular evolution of ITS2. We studied the secondary structure of ITS2 in species from each of the six main lineages of hard ticks (family Ixodidae). The ITS2 of these ticks varied in length from 679 bp in *Ixodes scapularis* to 1547 bp in *Aponomma concolor*. Nucleotide content varied also: the ITS2 of ticks from the Prostriata lineage (*Ixodes* spp.) had 46–49% GC whereas ITS2 sequences of ticks from the Metastriata lineage (all other hard ticks) had 61–62% GC. Despite variation in nucleotide sequence, the secondary structure of the ITS2 of all of these ticks apparently has five domains. Stems 1, 3, 4 and 5 of this secondary structure were obvious in all of the species studied. However, stem 2 was not always obvious despite

the fact that it is flanked by highly conserved sequence motifs in the adjacent stems, stems 1 and 3. The ITS2 of hard ticks has apparently evolved mostly by increases and decreases in length of the nucleotide sequences, which caused increases, and decreases in the length of stems of the secondary structure. This is most obvious when stems of the secondary structures of the Prostriata (*Ixodes* spp.) are compared to those of the Metastriata (all other hard ticks). Increases in the size of the ITS2 may have been caused by replication slippage which generated large repeats, like those seen in *Haemaphysalis humerosa* and species from the Rhipicepalinae lineage, and the small repeats found in species from the other lineages of ticks. *Heredity* (2002) 88, 275–279. DOI: 10.1038/sj/hdy/6800040

Keywords: ITS2; secondary structure; ribosomal RNA; rRNA; Ixodida; ticks

Introduction

The genes that encode ribosomal RNA (rRNA) occur as tandem repeats on one or more chromosomes in the nuclear genome of all eukaryotes. Each repeat comprises three coding regions (small subunit, 18S; 5.8S; and large subunit, 28S) and four non-coding regions or spacers. Three of these spacers, internal transcribed spacer 1 (ITS1), internal transcribed spacer 2 (ITS2) and the external transcribed spacer (ETS), are transcribed into a 45S precursor molecule but later this precursor is cleaved at specific sites and the spacers are removed. The precise roles of the transcribed spacers are unknown. However, there is good evidence that transcribed spacers have important roles in the biogenesis of the large rRNA subunit and maturation of the small subunit (eg Good *et al*, 1997). When nucleotides were deleted from ITS2 in a species of yeast, *Saccharomyces cerevisiae*, production of the ribosomal subunits decreased or stopped, especially when the deletions altered the secondary structure of ITS2 (van Nues *et al*, 1995). It seems that the secondary structure of the ITS2 is needed to facilitate its cleavage

from the precursor to form the mature ribosomal subunit (eg van der Sande *et al*, 1992).

Despite the apparent functional importance of ITS2, its primary sequence varies greatly in length from 100 bp in corals (Odorico and Miller, 1997) to 1547 bp in hard ticks (this paper). Although many methods have been used to infer secondary structures of RNA such as electron microscopy (Gonzales *et al*, 1990), site-directed mutagenesis (van der Sande *et al*, 1992; van Nues *et al*, 1995) and chemical and structural probing (Yeh and Lee, 1990), the most common method is to infer the secondary structure with software that favours secondary structures with the smallest free energy values (Zucker and Steigler, 1981) and then to examine the secondary structures from different populations and closely related species for common features (eg McLain *et al*, 1995 for *Ixodes* spp.).

A secondary structure for ITS2 that has four domains has been proposed for green algae and flowering plants (Mai and Coleman, 1997), fruit flies (*Drosophila* spp., Schlotterer *et al*, 1994), and parasitic flatworms (Michot *et al*, 1993; Morgan and Blair, 1998). We studied species from each of the six main lineages of hard ticks (family Ixodidae) to see if their ITS2 secondary structure also had four domains. We found that the secondary structure of the ITS2 in hard ticks did not have four domains. Instead, despite extreme variation in the primary sequence, each of the sequences folded into a structure with five domains.

Correspondence: SC Barker, Department of Microbiology and Parasitology, University of Queensland, Brisbane, 4072, Australia.

E-mail: s.barker@imb.uq.edu.au

Received 27 February 2001; accepted 14 November 2001

Materials and methods

Taxa

We obtained ticks from both the Prostriata (*Ixodes* spp.) and the Metastriata (all other hard ticks) and chose to study species which represent the six main phylogenetic lineages of hard ticks (Figure 1): *Ixodes scapularis* from the main lineage of *Ixodes* spp.; *I. holocyclus* from the *I. tasmani* group, which lives exclusively in Australasia (see Dobson and Barker, 1999; Klompen *et al*, 2000); *Amblyomma vikirri* and *Aponomma fimbriatum* from the *Amblyomma* and typical *Aponomma* lineage (*Amblyomma* s.l.; Dobson and Barker, 1999; Klompen *et al*, 2000); *Ap. concolor* from the endemic Australian *Aponomma* (described as *Bothriocroton* by JSH Klompen *et al*, submitted); *Haemaphysalis humerosa* from the Haemaphysalinae lineage; and *Dermacentor variabilis* from the rhipicephaline lineage (Rhipicephalinae plus Hyalomminae) (see Murrell *et al*, 2001a and references therein). We generated the sequences of *Ap. concolor* (AF199116), *Ap. fimbriatum* (AF199113), *A. vikirri* (AF199112) and *H. humerosa* (AF199115) but not the sequences of *I. scapularis* (Genbank L22276), *I. holocyclus* (AF208344) and *D. variabilis* (S83088). We also examined, but in less detail, the ITS2 sequences of *I. uriae* (D88307), *I. pacificus* (L22280), *Boophilus microplus* (U97712), *Rhipicephalus zambeziensis* (U97709) and *R. appendiculatus* (U97704).

DNA amplification and sequencing

ITS2 was amplified by PCR with the forward primer 3SA (Barker, 1998) and the reverse primers ITS2R (Domanico *et al*, 1997) and JB9A (Barker, 1998). PCR fragments were purified with the Qiaquick PCR purification kit (Qiagen) or the Wizard PCR Preps purification system (Promega). Both strands were sequenced with the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Purified products were separated with Applied Biosystems 373A and 377 gene sequencers. The sequencing primers, some of which were designed with the aid of Oligo5, were: 3SA (Barker, 1998); ITS2FB, 5' CTCGTTTGGACCGYGTCCGC 3' (anneals to positions 50–69 in *Am. vikirri*); ITS2MF, 5' AGTCCGCCGTCGGTC CAAGTG 3' (anneals to positions 630–650 in *Ap. fimbriatum*); ITS2HAF, 5' GTGCAGTCGTCTCTGATG TTG 3' (anneals to positions 595–615 in *H. humerosa*); ITS-2APCF, 5' TGCGGTGCTGTGTGAGAAATA 3' (anneals to positions 452–472 in *Ap. concolor*); AMAPHYITS2, 5' GCTCTCCCTGCAGRTA 3' (anneals to positions 260–277 in *Am. vikirri*); APC500R, 5' CCGCTTCCCCGACCA CAC 3' (anneals to positions 417–434 in *Ap. concolor*); HA500R, 5' CCGTTCCCCACGGCACGG 3' (anneals to

positions 386–403 in *H. humerosa*); ITS2HAR, 5' TCCTGG GACGTTTATTCTGCG 3' (anneals to positions 741–761 in *H. humerosa*); ITS2R2, 5' ACTTGGACCGACGGCGG ACT 3' (anneals to positions 1074–1093 in *H. humerosa*); ITS2R (Domanico *et al*, 1997); JB9A (Barker, 1998).

Sequence analysis

The start of the ITS2 was identified by the motif ACTA-TAT whereas the end was identified as being 5 bp upstream of the motif ACCTCAGA (see Barker, 1998). Both of these motifs were found in all of the ticks we studied and were identified with respect to the sequences of the flanking rRNA genes, 5.8S and 28S. The programs Repeat (searches for repeats), Composition (calculates nucleotide composition, ie, mono, di and tri-nucleotide frequencies), Wordcount (searches for common 'words' in a sequence) and Tandem (searches for tandem repeats) from the GCG package on ANGIS (Australian National Genomic Information Service) were used to analyse the nucleotide sequences. Sequences were aligned with ClustalW1.74 (Thompson *et al*, 1994); default gap and extension penalties were used. Alignments were then adjusted by eye.

Folding of sequences into putative secondary structures

Sequences were folded with mfold-3 at M Zuker's web page (www.ibc.wustl.edu/zuker). Default values (37°C with 5% suboptimal folding) were used to fold the ITS2 rDNA. When available, about 20 bases of flanking sequence (5.8S and 28S rRNA) were included because these have been shown to be important in the folding of ITS2 in other ITS2 sequences (Veldman *et al*, 1981; Morgan and Blair, 1998). Structures inferred by mfold-3 were examined for common stems, loops and bulges. Sequences which formed markedly atypical stems and loops were resubmitted to mfold-3 with constraints that forced the extension of some stems or forced canonical base pairs that were present in most other species of ticks. In this way, stem 1 was extended in most species so that TCAAG (positions 9–13 in *I. holocyclus*) bound with TTTGA (positions 44–48 in *I. holocyclus*) because this increased the number of canonical base pairings in stem 1. For species from the subfamily Rhipicephalinae for which the 20 bp of flanking sequence (5.8S and 28S) were not available, ACA (positions 4–6 in *B. microplus*) was forced to bind with TGT (positions 1138–1140 in *B. microplus*). Putative stems 1, 3, 4, and 5 were also folded independently of the entire ITS2 molecule in an attempt to optimise the secondary structures. Local alignments of conserved sequences from stems from the different species were examined for base substitutions that were compensatory, ie, base substitutions that maintained the secondary structure.

Results

In the hard ticks we studied, the ITS2 varied in length from 679 bp in *I. scapularis* to 1547 bp in *Ap. concolor* (Figure 1). Species from the Prostriata lineage (*Ixodes* spp.) were 46–49% GC rich whereas the rest of the ticks, the Metastriata, were 61–62% GC rich (Figure 1). Apart from *Ap. fimbriatum* and *Am. vikirri* (65% homology), *I. holocyclus* and *I. uriae* (78% homology) and the four rhipicephalinae ticks (*D. variabilis*, *B. microplus*, *R. zambeziensis* and *R. appendiculatus*: 64–97% homology), the sequences

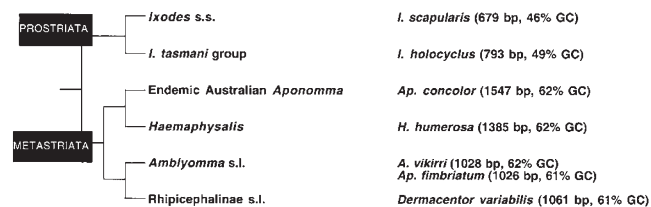


Figure 1 Phylogenetic relationships of the six major lineages of hard ticks (according to Klompen *et al*, 2000). The species studied by ourselves are on the right. Size and GC content of ITS2 sequences are shown in parentheses. s.s., *sensu stricto*; s.l., *sensu lato*.

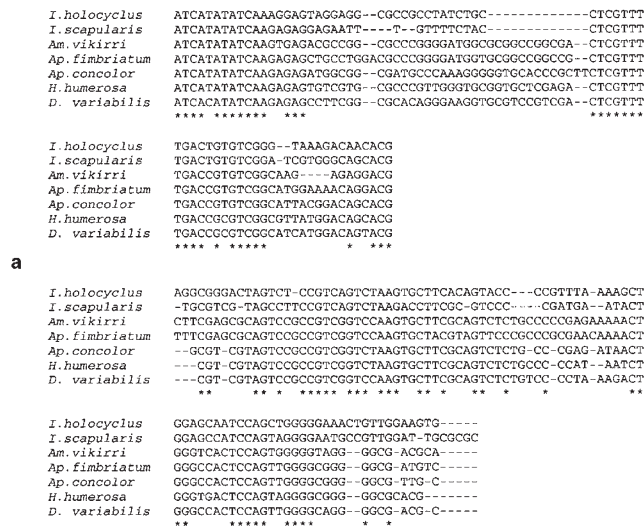


Figure 2 (a) 88 bp alignment of conserved sequences at the 5' end of ITS2 of ticks (5'–3'). (b) 99 bp alignment of conserved sequences in the last quarter of ITS2 (5'–3').

were unalignable except for 88 bp at the 5' end of ITS2 (stem 1) and 99 bp in the last quarter of ITS2, which form part of stem 3 (Figure 2). Numerous short imperfect tandem repeats of up to 20 bp were found in all species studied (data not shown). Near perfect repeats were found in *D. variabilis*, 104 bp separated from its 84 bp incomplete copy by 39 bp, and in *H. humerosa*, 144 bp separated from its 159 bp copy by 480 bp (Figure 3). Repeats that may be homologous to the repeats found in *D. variabilis* are also in the ITS2 of *B. microplus*, *B. decoloratus*, *R. zambezensis*, *R. appendiculatus* and *R. evertsi* (Barker, 1998). The molecular evolution of these repeats in the rhipicephaline ticks and their relatives was studied by Murrell *et al* (2001b).

The putative secondary structures of the seven species from the six main lineages of hard ticks studied have five domains (see Figure 4 for *I. holocyclus*). The 20 bp of 5.8S and 28S rDNA that were immediately adjacent to the 5' and 3' ends of the ITS2 apparently form canonical bonds with each other and about the first 10 bp of the ITS2 to form stem 0 (see Figure 4 for *I. holocyclus*). Stems 1, 3, 4

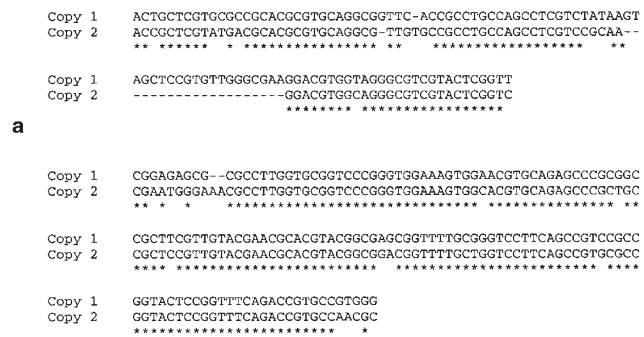


Figure 3 (a) The two copies of the long repeat in *Dermacentor variabilis* at base positions 803–906 and 946–1029 respectively. (b) The two copies of the long repeat in *Haemaphysalis humerosa* at base positions 251–397 and 878–1026 respectively.



Figure 4 The proposed secondary structure of *Ixodes holocyclus* which shows the five putative domains proposed for the ticks studied here; minimum free energy = –342.1 (–351.7 initially). Stem 0 consists of 20 bp of sequence that flanks (5'–3') the ITS2. Note that this structure is very similar to that proposed for *I. scapularis* by McLain *et al* (1995). The proposed secondary structures of the other species of hard ticks studied by ourselves are not shown.

and 5 were obvious in all species studied. However, stem 2 was not always obvious despite the fact that it was flanked by highly conserved sequence motifs in the adjacent stems, stems 1 and 3. In fact, stem 2 varied greatly among the seven species studied. In *Am. vikirri*, *Ap. fimbriatum* and *D. variabilis*, stem 2 had two additional terminal stems (stems 2a and 2b, not shown) whereas in *Ap. concolor* and *H. humerosa*, the initial part of stem 2 was short but then it branched into three to four other shorter stems (not shown). All four *Ixodes* spp. (*I. scapularis*, *I. pacificus*, *I. uriae* and *I. holocyclus*) had a long stem 2 that had no side or terminal branches (Figure 4). The ITS2 of the four rhipicephaline sequences (*D. variabilis*, *B. microplus*, *R. zambezensis* and *R. appendiculatus*) also folded into five domains; stems 2a and 2b were also present in these species. However, some of the sub-optimal foldings for *D. variabilis* had six domains; this was due to the long repeat in this species. Of the five stems, only stems 1 and 3 had sequence homology greater than 70% among ticks from the six lineages. In all cases all stems were shorter in species from the Prostriata lineage (*Ixodes* spp.) compared to the species from the other main lineage of hard ticks, the Metastrata (all hard ticks except the *Ixodes* spp.).

Putative compensatory changes were found in all

stems. However, only those compensatory changes from stems 1 (approximately 120 bp at the 5' end of ITS2) and 3 (approximately 60 bp at positions 605–665 in *I. holocyclus*) are likely to be true compensatory changes due to the substantial evolution of the primary sequence in all other regions; indeed, the sequences of these other regions were unalignable. The apparent compensatory changes in the other regions are probably artefacts due to chance associations of nucleotides.

Discussion

The ITS2 of hard ticks ranges in size from 679 bp (*I. scapularis*) to 1547 bp (*Ap. concolor*); the ITS2 of the latter is the largest ITS2 known. Given the range in size of ITS2 in ticks, it is not surprising that the ITS2 sequences we studied were almost entirely unalignable (10% homology overall) with the exception of *Ap. fimbriatum* and *Am. vikirri* (65% homology), *I. holocyclus* and *I. uriae* (78% homology), and the four rhipicephaline ticks (*D. variabilis*, *B. microplus*, *R. zambeziensis* and *R. appendiculatus*: 64–97% homology). Note that the sequences of *Ap. fimbriatum* and *Am. vikirri* could be aligned whereas the sequence of *Ap. concolor* could not easily be aligned with *Ap. fimbriatum* nor with *Am. vikirri*, and is over 500 bp longer than those sequences. This is consistent with the phylogenetic hypothesis of Dobson and Barker (1999) and JSH Klompen *et al* (submitted) that *Ap. concolor* and other Australian *Aponomma* spp. from reptiles and the echidna (a monotreme) are from a lineage that is phylogenetically well removed from the lineage that has the rest of *Aponomma* spp., including *Ap. fimbriatum*. The sequence of *I. uriae* could be aligned with that of *I. holocyclus* (78% homology) but not to the sequences of *I. pacificus* and *I. scapularis*. In phylogenies of *Ixodes* spp. inferred using ITS2 sequences (Fukunaga *et al*, 2000) *I. uriae* was more closely related to *I. holocyclus* than to *I. scapularis* and *I. pacificus*. However, morphology-based phylogenies indicate that *I. uriae* is more closely related to *I. scapularis* and *I. pacificus* than it is to *I. holocyclus* (Klompen *et al*, 1996). We found two highly conserved regions in all species studied. One of these, 99 bp at the 5' end of stem 3 (Figure 2b), may be a ribosomal processing site since this sequence has substantial homology (75%) to the ribosomal processing site of *S. cerevisiae* (see also Reddy *et al*, 1983; Morgan and Blair, 1998).

Despite much divergence in nucleotide sequences and large differences in length, the ITS2 of all hard ticks studied by us could be folded into five domains. This indicates that the evolution of the secondary structure of ITS2 in hard ticks has been constrained; the exception is stem 2, which is highly variable in length and form. The secondary structure of ITS2 of a range of animals and plants is thought to have four domains (algae and flowering plants, Mai and Coleman, 1997; *Drosophila* spp., Schlottterer *et al*, 1994; parasitic flatworms, Michot *et al*, 1993, and Morgan and Blair, 1998). However, the ITS2 of ticks apparently has five domains. The ITS2 of the ticks we studied had a long central stem. Thus, this stem may be functionally important. Despite the ITS2 sequences of animals and plants varying greatly in length they all apparently have a four, or in the case of ticks and *Gyrodactylus salaris*, a flatworm (Morgan and Blair, 1998), five domains and a long central stem.

The ITS2 of hard ticks has apparently evolved mostly

by increases and decreases in the length of the nucleotide sequence, which cause increases and decreases, in the length of stems of the secondary structure. This is most obvious when the secondary structures of the Prostriata (*Ixodes* spp.) are compared with those of the Metastriata (all other hard ticks). The Prostriata have much shorter stems than the Metastriata. In particular, stems 1 and 5 in the Prostriata are only half as long as those of the Metastriata. The increases and/or decreases in the size of the ITS2 may have been caused by replication slippage events which generated large repeats like those seen in *H. humerosa* and *D. variabilis* (Figure 3) and the small repeats found in the other species. Apparently, a high rate of point mutations has made most of these repeats undetectable.

Acknowledgements

We thank Mathew Shaw for access to his ITS2 sequence of *Ixodes holocyclus* before it was submitted to Genbank; James Lever, manager of VISAC (Visualisation and Advanced Computation Laboratory), for use of SGI computers; Francis Clark and Larry Croft for help with programming; and Sue Dobson and Nick Campbell for reading the manuscript and providing helpful comments.

References

- Barker SC (1998). Distinguishing species and populations of rhipicephalinae ticks with ITS 2 ribosomal RNA. *J Parasitol* **84**: 887–892.
- Dobson SJ, Barker SC (1999). Phylogeny of the hard ticks (Ixodidae) inferred from 18S rRNA indicates that the genus *Aponomma* is paraphyletic. *Mol Phylogenet Evol* **11**: 288–295.
- Domanico MJ, Phillips RB, Oakley TH (1997). Phylogenetic analysis of Pacific salmon (genus *Oncorhynchus*) using nuclear and mitochondrial DNA sequences. *Can J Fish Aquat Sci* **54**: 1865–1872.
- Fukunaga M, Yabuki M, Hamase A, Oliver Jr JH, Nakao M (2000). Molecular phylogenetic analysis of ixodid ticks based on the ribosomal DNA spacer, internal transcribed spacer 2, sequences. *J Parasitol* **86**: 38–43.
- Gonzales IL, Sylvester JE, Smith TF, Stambolian D, Schmickel RD (1990). Ribosomal RNA gene sequences and hominid phylogeny. *Mol Biol Evol* **7**: 203–219.
- Good L, Intine RVA, Nazar RN (1997). Interdependence in the processing of ribosomal RNAs in *Schizosaccharomyces pombe*. *J Mol Biol* **273**: 782–788.
- Klompen JSH, Black IV WC, Kierans JE, Oliver Jr JH (1996). Evolution of Ticks. *Ann Rev Ent* **41**: 141–146.
- Klompen JSH, Black WC, Keirans JE, Norris DE (2000). Systematics and biogeography of hard ticks, a total evidence approach. *Cladistics* **16**: 79–102.
- Mai JC, Coleman AW (1997). The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J Mol Evol* **44**: 258–274.
- McLain DK, Wesson DM, Oliver Jr JH, Collin FH (1995). Evolution of the rDNA spacer, ITS2, in the ticks *Ixodes scapularis* and *I. pacificus* (Acari: Ixodidae). *Heredity* **75**: 303–319.
- Michot B, Despres L, Bonhomme F, Bachelier JP (1993). Conserved secondary structures in the ITS2 of trematode pre-rRNA. *FEBS Lett* **316**: 247–252.
- Morgan JA, Blair D (1998). Trematode and monogenean rRNA ITS2 secondary structures support a four-domain model. *J Mol Evol* **47**: 406–419.
- Murrell A, Campbell NJH, Barker SC (2001a). A total-evidence phylogeny of ticks provides insights into the evolution of life cycles and biogeography. *Mol Phylogenet Evol* **21**: 244–258.
- Murrell A, Campbell NJH, Barker SC (2001b). Recurrent gains

- and losses of large (84–109 bp) repeats in the rDNA internal transcribed spacer 2 (ITS2) of rhipicephaline ticks. *Insect Mol Biol* **10**: 587–596.
- Odorico DM, Miller DJ (1997). Variation in the ribosomal internal transcribed spacers and 5.8S rDNA among five species of *Acropora* (Cnidaria; Scleractinia): patterns of variation consistent with reticulate evolution. *Mol Biol Evol* **14**: 465–473.
- Reddy R, Rothblum LI, Subrahmanyam CS, Liu MH, Henning D, Cassidy B, Busch H (1983). The nucleotide sequence of 8 S RNA bound to preribosomal RNA of *Novikoff hepatoma*. The 5'-end of 8 S RNA is 5.8 S RNA. *J Biol Chem* **258**: 584–589.
- Schlotterer C, Hauser MT, Von Haeseler A, Tautz D (1994). Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Mol Biol Evol* **11**: 513–522.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Van Der Sande CA, Kwa M, Van Nues RW, Van Heerikhuizen H, Raue HA, Planta RJ (1992). Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. *J Mol Biol* **223**: 899–910.
- Van Nues RW, Rientjes JM, Morre SA, Mollee E, Planta RJ, Venema J *et al* (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.
- Veldman GM, Klootwijk J, Van Heerikhuizen H, Planta RJ (1981). The nucleotide sequence of the intergenic region between 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.
- Yeh LC, Lee JC (1990). Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. *J Mol Biol* **211**: 699–712.
- Zuker M, Stiegler P (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**: 133–148.