

Evolution of the rDNA spacer, *ITS 2*, in the ticks *Ixodes scapularis* and *I. pacificus* (Acari: Ixodidae)

DENSON K. McLAIN†*, DAWN M. WESSON‡§, FRANK H. COLLINS‡ & JAMES H. OLIVER JR†

†Department of Biology & Institute of Arthropodology & Parasitology, Georgia Southern University, Statesboro, GA 30460-8042 and ‡Malaria Branch, DPD, CID, Centers for Disease Control, 1600 Clifton Road, C-22, Atlanta, GA 30333, U.S.A.

Evolution of the rDNA spacer, *ITS 2*, is examined by comparing 17 DNA sequences of the ticks, *Ixodes scapularis* and *I. pacificus*. The distribution of fixed interspecific differences and the relative frequency of base changes vs. insertions/deletions (indels) matches the distribution and relative frequency for intraspecifically variable sites. This suggests that most intraspecific variation is not effectively selected against. The base composition of the *ITS 2* transcript is G- and U-biased. But, 5-base regions enriched (> 80 per cent) for A or U occur more frequently than expected while G- and C-enriched regions occur less frequently than expected. Enriched sequences may be prone to replication slippage, accounting for the A/T bias in insertions. Slippage-mediated gains and losses of A/T-rich tandem repeats apparently account for most indels. Minimum-energy conformations of the two species' folded transcripts share major structural features. Structural inertia arises from intramolecular base pairing within stems that allows most mutations to be absorbed as new bulges off stems. Yet, there is evidence of selection to maintain the conformation. First, intraspecifically variable sites are concentrated at the ends of stems in loops and intersections, structures that do not contribute to intramolecular base pairing. Moreover, some indels that have become fixed in one species compensate for the presence of conformation-destabilizing indels. However, high rates of sequence evolution within stems and absence of compensatory base evolution contraindicates selective constraint. Degenerate dispersed and tandem copies of two subrepeats, each approximately 20 bases long, may account for much of the *ITS 2* sequence. These are approximate inverses of each other and are, consequently, capable of significant intramolecular hydrogen bonding to produce folded transcripts of low energy. Evolution of the *ITS 2* sequence may largely entail replication slippage-mediated gains and losses of these repeats or their composite subrepeats.

Keywords: compensatory base evolution, concerted evolution, *ITS 2*, replication slippage, subrepeats, tick.

Introduction

Arthropod rDNA consists of tandem or bipolar arrays (Robbins & Swanson, 1988) of a moderately repeating unit that contains genes for 18S, 5.8S, and 28S rRNA, two internal transcribed spacers (*ITS 1* and *ITS 2*) that flank the 5.8S gene (Beckingham, 1982), and a largely nontranscribed intergenic spacer composed of subrepeats (Lohe & Roberts, 1990). During transcription,

each repeating unit gives rise to a single transcript, the 45S pre-rRNA, that must be processed to liberate genic products from associated spacers (Beckingham, 1982).

Typically, arrays of rDNA repeating units occur at several sites within and among chromosomes, the nucleolar organizing regions (Beckingham, 1982). Yet, individual repeating units do not evolve independently, but, rather, in concert (Arnheim *et al.*, 1980; Dover *et al.*, 1982; Dallas *et al.*, 1988). Concerted evolution appears primarily to be driven by unequal exchange between arrays (Arnheim *et al.*, 1982; Gillings *et al.*, 1987; Dover, 1989; Lyckegaard & Clark, 1991) which

*Correspondence.

§Current address: School of Public Health & Tropical Medicine, 1501 Canal Street, Suite 505, Tulane University, New Orleans, LA 70122, U.S.A.

occurs at rates of 10^{-4} (Williams *et al.*, 1989) to 10^{-3} (Maddern, 1981) per gamete per generation. Secondly, gene conversion may promote concerted evolution (Ohta, 1983; Amstutz *et al.*, 1985; Curtis & Bender, 1991). Replication slippage, intrastrand exchange, and germ line magnification may also affect the copy number of particular repeat variants (Walsh, 1987; Lyckegaard & Clark, 1991).

Stochastic (e.g. unequal exchange) and biased (e.g. gene conversion) cellular events are responsible for molecular drive, the spreading of a variant repeat through a multigene family (homogenization) and a sexual population (Dover, 1982, 1986, 1989). Theoretical (Ohta & Dover, 1984) and empirical studies (Coen *et al.*, 1982; Lachance, 1990) indicate that the process of homogenization is rapid. Consequently, it is not necessary to sample many copies of a sequence to deduce relationships between populations and species.

rDNA contains very rapidly evolving sequences, spacers (Boncinelli *et al.*, 1983; Tautz *et al.*, 1987) and expansion segments within rRNA genes (Hancock & Dover, 1988; Hancock *et al.*, 1988), as well as some of the most evolutionarily conserved sequences known (Hassouna *et al.*, 1984; Gerbi, 1985; Larson & Wilson, 1989). Base usage bias and sequence simplicity suggest that sequence variation in expansion segments arises through replication slippage (Hancock & Dover, 1988). Nonetheless, hypervariable sequences may experience selection for compensatory base changes arising from the requirement that their transcripts form particular secondary structures for proper functioning and/or processing (Hancock *et al.*, 1988; Michot & Bachelierie, 1987).

In this study, we investigate sequence evolution within *ITS 2* of the ticks *Ixodes scapularis* and *I. pacificus*, members of the *I. ricinus* complex (Keirans *et al.*, 1992) by comparing sequences within and between species. Seventeen representatives of *ITS 2* were sequenced: four from *I. pacificus*, five from the southern part of the geographical range of *I. scapularis*, and eight from the northern part of the range of *I. scapularis*. *I. scapularis* from the northern part of the species range were previously described as a separate species, *I. dammini* (Spielman *et al.*, 1979), but this species has recently been synonymized with *I. scapularis* on the basis of a number of molecular, morphological, and reproductive criteria (Oliver *et al.*, 1993; Wesson *et al.*, 1993). Thus, sequence variation can be analysed with respect to the degree of relationship between populations of origin.

Comparisons of *ITS 2* within and between related genera of arthropods indicate that blocks of sequence homology are retained, preserving particular features of secondary structure, while other blocks, typically

GC-rich, appear to evolve at a high rate via replication slippage (Wesson *et al.*, 1992). Some constraint on the evolution of *ITS 2* sequences appears to derive from the role of the secondary structure of the *ITS 2* transcript in processing the 45S pre-rRNA transcript (Reddy *et al.*, 1983; Yeh & Lee, 1990; Thweatt & Lee, 1990). Here, we are also interested in determining the distribution of sequence variation with regard to predicted features of secondary structure based on minimal energy configurations for *I. scapularis* and *I. pacificus* (e.g. Wesson & Collins, 1992).

Materials and methods

Sources of ticks

Eggs of *I. scapularis* from Georgia and Massachusetts and of *I. pacificus* from California were obtained as the mixed clutches of several females from separate colonies maintained at the Institute of Arthropodology and Parasitology, Georgia Southern University. *I. scapularis* nymphs derived from eggs of single females from Maryland, Massachusetts, New Jersey, New York, and North Carolina were provided by J. Piesman (Division of Vector-Borne Infectious Diseases, Centers for Disease Control, Ft. Collins, Colorado). *I. pacificus* adults collected in Arizona were provided by J. M. C. Ribeiro (University of Arizona, Tucson, Arizona). Details on the date and site of collection and on the age of laboratory colonies are provided in McLain *et al.* (1995).

ITS 2 subclones

DNA was isolated from either 3 mg of eggs, of mixed parentage, or from single nymphs using the procedure of Livak (1984). Eggs in a late stage of development were frozen at -80°C prior to DNA isolation. Similarly, nymphs were frozen until time of DNA extraction. One-2500th of the sample was used to amplify the *ITS* using the polymerase chain reaction (Mullis *et al.*, 1986).

PCR amplification employed a Perkin Elmer DNA thermal cycler and Perkin Elmer GeneAmp reagents (Perkin Elmer Corporation, Norwalk, CT). DNA was amplified in a 100 μL volume for 25 cycles. Each cycle consisted of 1 min at 94°C , 2 min at 37°C , and 2 min at 72°C . Primers, flanking the *ITS*, were (1) 5'-GTAAGCTTCCTTTGTACACACCGCCCGT-3' where bases 3–8 are a *Hind*III site and bases 9–28 are a conserved sequence in the 18S rRNA gene (Olsen *et al.*, 1983) and (2) 5'-GTGAATTCTATGCTTAAATT-CAGGGGGT-3' where bases 3–8 are an *Eco*RI site and bases 9–28 are a conserved sequence in the 28S rRNA gene (Ware *et al.*, 1983).

Amplified sequences and the plasmid *pBluescript II* (Stratagene Cloning Systems, La Jolla, CA) were digested with both *HindIII* and *EcoRI*, ethanol precipitated, resuspended in water, and ligated. *Escherichia coli*, strain XL1-Blue (Stratagene), was transformed with the ligation products using the CaCl_2 procedure (Sambrook *et al.*, 1989).

Seventeen *ITS 2* subclones were sequenced: four *I. pacificus* sequences (three from California and one from Arizona) and 13 *I. scapularis* sequences (four from Georgia, five from Maryland, and one each from Massachusetts, New Jersey, New York, and North Carolina).

DNA sequencing

Double-strand *ITS* inserts in *pBluescript II* were sequenced by the dideoxy chain-termination method (Sanger *et al.*, 1977) using Sequenase reagent kits (U.S. Biochemical Corporation, Cleveland, OH). Inserts were sequenced in both forward and reverse directions by employing a variety of primers complementary to internal sequences of *ITS 2*. When, on rare occasions, compressions were suspected, sequencing reactions were repeated utilizing dITP. Sequencing reaction products were separated in a 6 per cent acrylamide/urea gel for which the top and bottom buffers were $1 \times$ TBE (Sambrook *et al.*, 1989).

Sequence analysis

Sequence alignment, determination of base composition, and search for direct and inverted repeats were performed on a microcomputer using PC/GENE™ software (IntelliGenetics, Inc., 1991). For alignments using PC/GENE™'s NALIGN program, the open gap cost was 5 and the unit gap cost was 5. Transcript folding patterns that provide minimum energy structures for the entire *ITS 2* sequence were generated using the FOLD program (Genetics Computer Group, Version 7; Devereux *et al.*, 1984) which utilizes the method of Zuker and Steigler (1981) with energies as defined in Freier *et al.* (1986). Folding of subsequences was performed with the RNAFOLD program of PC/GENE™.

Statistics

Chi-square analyses are used to assess differences between sequences in the number of base changes and to assess differences in the distribution of variable sites within a sequence. Most analyses are based on the use of consensus sequences (e.g. for species comparisons) because as the number of sequences (n) increases the number of pairwise comparisons increases exponen-

tially ($=[n^2 - n]/2$), inflating the degrees of freedom and compromising the independence of data points.

When assessing the number of mutation events, it is assumed that contiguous insertions or deletions, relative to another sequence, represent a single event. However, contiguous base changes, transitions and/or transversions, are assumed to represent independent events.

Results

Secondary structure

The predicted secondary structure of the *ITS 2* of *I. scapularis* has extensive intramolecular hydrogen bonding (energy = -242.2 kcal) that produces five stem structures that terminate in small loops plus three stem-like connectors (Fig. 1). Note, however, that this is a minimum energy conformation while the actual conformation attained may be different if it is determined by interaction with proteins (e.g. Brimacombe *et al.*, 1983). One stem, designated no. 4 in Fig. 1 (also, see Table 1), encompasses over half of the approximately 700 bases of the *ITS 2* transcript. The secondary structure is folded with 41 bases from the 3' end of the 5.8S gene and with 22 bases from the 5' end of the 28S gene. These extragenic sequences base pair entirely between themselves, permitting the complete isolation of the *ITS 2* within the pre-rRNA as a series of stems exposed to processing enzymes.

The secondary structure of the *ITS 2* transcript of *I. pacificus* is very similar to that of *I. scapularis* (Fig. 1) although the transcript is 48 bases longer (727 vs. 679 bases). Still, 86.0 per cent of the *I. pacificus* sequence is in the same stem, loop, or intersection as the homologous *I. scapularis* sequence. When insertions in one species relative to the other are removed, the aligned sequences match at 603 of 658 bases (91.6 per cent; Table 1). The *I. pacificus* minimum energy secondary structure (-268.6 kcal) is obtained with some base pairing between the transcripts of the 28S and 5.8S genes, as is the case for *I. scapularis*.

The secondary structure of the *I. pacificus* transcript has an additional stem, relative to that of *I. scapularis*, designated no. 9 in Fig. 1. The size of this stem accounts for two-thirds of the difference in size of the *ITS 2* between the species. The origin of this stem is easily accounted for (1) by the presence of a large bulge from stem no. 8 in the *I. scapularis* transcript (positions 733–739 in Table 1) which becomes the loop of stem no. 9 in *I. pacificus*, and (2) by the presence of two insertions in the *I. pacificus* sequence, relative to that of *I. scapularis*, that are located on each side of the bulge/

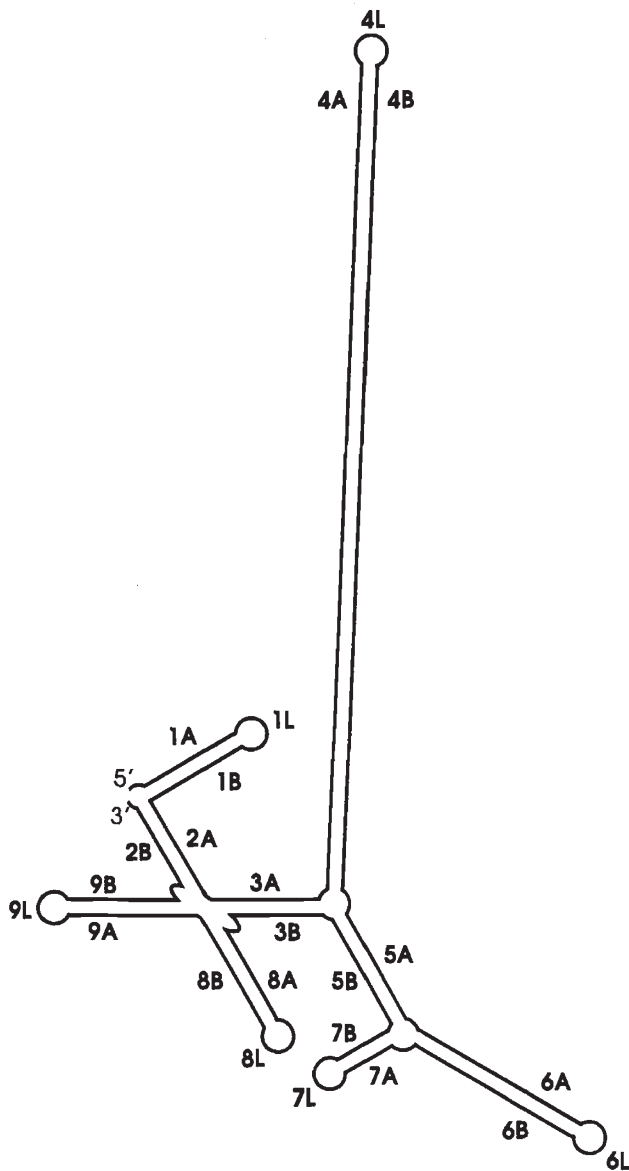


Fig. 1 Schematic diagram of the secondary structure of rRNA ITS 2 transcripts of *Ixodes scapularis* and *I. pacificus*. Stem 9 is unique to *I. pacificus*. The corresponding DNA sequence begins at the junction of 1A and 2B and proceeds in increasing numerical order as labelled. Stem and loop labels provide a reference to the base sequences in Table 1.

loop (positions 725–729 and 749–754 in Table 1) and that constitute much of stem no. 9.

Bulges from stems occur for bases that do not pair intramolecularly in the folded transcript. According to the minimum energy secondary structure, 27.8 per cent (= 189/679) of the *I. scapularis* and 24.5 per cent (= 178/727) of the *I. pacificus* sequences occur in bulges. Ninety-nine bases occur in bulges that are conserved between the species. Among these, the degree

of sequence conservation is high, 91.9 per cent, and similar to the degree for the remainder of the bases in stems, 90.1 per cent (excluding insertions in one species relative to the other).

Base composition

The 5' to 3' consensus sequences for the ITS 2 of *I. scapularis* and *I. pacificus* are presented in Table 1. The lengths of the consensus sequences for *I. scapularis* and *I. pacificus* are, respectively, 679 and 727 bases. For neither *I. scapularis* ($\chi^2_1 = 3.25$) nor *I. pacificus* ($\chi^2_1 = 2.31$) is the proportion of A/T bases in the sequence significantly greater than the proportion of G/C bases ($P > 0.05$). However, for both species, there is a bias toward G- and especially T-bearing nucleotides in the transcript (*I. scapularis*: T=0.296, G=0.269, A=0.238 and C=0.198; *I. pacificus*: T=0.310, G=0.281, A=0.218 and C=0.191). Base composition does not vary significantly between species ($\chi^2_3 = 1.09$, $P > 0.50$).

The various bases are not randomly distributed. The sequences of both species have more regions five bases in length that are enriched for A (80–100 per cent A) or enriched for T (80–100 per cent T) than is expected by chance, given the proportions of these bases in the overall sequence (Table 2). Because the number of regions enriched for C or G is slightly fewer than expected by chance, there is a significant difference in the frequency of regions enriched for A or T vs. enriched for G or C (*I. pacificus*: $\chi^2_1 = 27.69$, $P < 0.001$; *I. scapularis*: $\chi^2_1 = 30.64$, $P < 0.001$; four random sequences with *I. pacificus* ITS 2 base composition and length: $\chi^2_1 = 0.80$, $P > 0.10$; see Table 2 for the calculation of expected values). In the folded transcript, regions enriched for T occur in half the stem loops (loops 1, 4, and 8; not loops 6, 7, and 9) and in most large intersections (greater than five bases) between stems (between stems 3 and 4, 4 and 5, and 6 and 7 but not between stems 2 and 8; Table 1, Fig. 1).

Among 28 interspecifically conserved blocks of sequence, the G/C ratio (0.505, $n = 517$ bases) is significantly higher ($t = 3.30$, $P < 0.001$, t -test for proportions) than in more variable sequence blocks (0.384, $n = 365$ bases) as has been noted for Aedine mosquitoes (Wesson *et al.*, 1992). Here, a conserved sequence block is at least five bases long, has no consensus sequence differences between species, and has no base position that is variable in both species.

Interspecific sequence variation

Forty-four transitions/transversions occur in the *I. scapularis* consensus sequence relative to the aligned *I.*

Table 1 Continued

scap	-----6A-----									
pac	b b b b	b				b b b b b b b	b		b b b	
pac	b b b b	b				b b b b b b b	b		b b b	
pac	C T T C C G T C A G T C T A A G A C C T T C G C G G T C C C C G A T									
north
south
		550				560			570	
scap	-----6A] [-----6B-----									
pac	b b b b]l l l l l[b		b	b b	
pac	b b b b]l l l l l[b		b	b b	
pac	G A A T A C T G G A G C C A T C C A G T A G G G G T A T G C C G T T							A		
north
south	A	.	.
	580		590			600			610	
scap	-----6B] [---7A---] [---7B---][5B									
pac		b]libibibibibi[]l l l l l l l l[i]l[
pac		b]i[]l l l l l l l l[b b	i]l[
pac	G G A T T G C G C G C T T T C T T T T T G T C G A A T C G A G A G									
north	A . .
south	A . .
	620		630			640				
scap	-----5B] [---3B---] [---8A									
pac]i[]l		b		b b b b b		
pac]i[]l		b b b b]bibibibibibi[
pac	C G A T G C A T A G C G T G T G T A C A A C T C A T T T T T T T C									
north	.	*	.	A	.	.	.	A A C	.	G . . A G . . *
south	.	.	.	A	.	.	.	A A C	.	G . . A G . .
	650		660			670			680	
scap	-----8A] [-----8B									
pac	b		b]l l l l l l l l[b		
pac	b]l l l l l l l l l					l l l l[
pac	G C C A T C C T T T T T T T G T T - - - - - G G T A A G G A G									
north	.	.	*	.	.	*	.	- G . . . T T T T G A C A T		
south	.	.	*	- G . . . T T T T G A C A T		
		690				700			710	
scap	-----8B] [-----2B									
pac	b		b]bibibibibibibi[b		b b b b		
pac		i]l[---9A		b	9A---]l l l l l l l l[---9B			b		
pac	G C G A A A T T T T T A T T G T G T T G G A A A G C G C A A A A A A									
north	*	.	* C
south	*	.	* C
	720		730			740				
scap	-----2B]									
pac	b b b b b b	b]bibibi[
pac	-9B]i[---2B		--2B]bibi[
pac	A A A G A T T C G - A T T G									
north	- - - - -	.	*	A						
south	- - - - -	.	*	A						
	750		760							

Features of secondary structure are indicated for corresponding sequences as follows: (1)] and [delimit stems while numbers and associated upper case letters correspond to specific stems as labelled in Fig. 1; (2) i = intersection between stems; (3) l = loop between complementary sides of a stem; and (4) b = bulges from a stem for bases not complementarily paired. A dot indicates that a north or south sequence matches that of pac. An underlined sequence is intraspecifically variable as are scapularis sequence positions indicated by an asterisk. Numbers below each block of sequence permit the localization of sequence elements discussed in the text.

pacificus sequence. Among these base changes, 17 are apparently associated with intramolecular base pairing in the folded transcript of both species when non-canonical G/T pairs are permitted. However, there is not a single instance in which a base change at one point is associated with another change elsewhere to maintain/promote intramolecular hydrogen bonding in the folded transcript.

Most points of consensus base change between the species (61.4 per cent or 27/44) are associated with bulges from stems in one or both species. For the subset of bases that correspond to bulges conserved between the species, the rate of base change, 0.081 (= 8/99) is not significantly greater than the rate of change for the remainder of the sequence, 0.065 (= 36/553) ($t = 0.54$, $P > 0.50$, t -test for proportions).

Deletions and insertions

Based on the alignment of consensus sequences, a minimum of 44 deletion/insertion events (= indels) are represented by differences in the consensus sequences of *I. scapularis* and *I. pacificus*. The sizes of indels range from 1 to 22 bases, one being the mode (28/44 = 63.6 per cent). The base composition of insertions shows a higher A/T proportion than in the sequences as a whole ($\chi^2_1 = 5.79$, $P < 0.025$).

Twenty-four of the deletions/insertions are fixed (but may be of variable length), while 20 are variable in their occurrence within a species. Thirteen of 24 (= 54.2 per cent) fixed insertions (relative to the other species) give rise to bulges or expanded loops and intersections between stems, assuming the minimum energy conformations of the transcripts are correct.

Insertions (relative to the other species) of greater than one base in length were often associated with the presence of similar flanking sequences, suggesting that indels arise from replication slippage (Table 3). For instance, the 3-base deletion 'TCG' in *I. scapularis* (stem 4A; see Table 1) is preceded in *I. pacificus* by 5'-TGTGTCG-3'. Similarly, the deletion 5'-ACGCTTC-3' in *I. pacificus* is, in *I. scapularis*, one member of two tandem repeats. The first seven bases of the largest indel, 22 bases long, resemble a sequence motif that is repeated twice in tandem 5' to the indel while the last 15 bases resemble a motif repeated twice in tandem 3' to the indel (Table 3). Thus, the indel may be the product of two contiguous replication slippage events. These and other examples are presented in Table 3. The proportion of A and T bases among all the indels and their flanking tandem copies is significantly higher than for the overall *ITS 2* sequence ($t = 2.42$, $P < 0.025$).

The 22-base insertion may be capable of intramolecular pairing. Similarly, the sequences flanking the

Table 2 Observed (obs) and expected (exp) number of 5-base long regions with four or five nucleotides of the same base (i.e. enriched) in the *ITS 2* of *Ixodes scapularis* and *I. pacificus* and in the *28S* gene of *Drosophila melanogaster*

	<i>I. scapularis</i>		<i>I. pacificus</i>		Random	<i>D. melanogaster</i>	
	Obs	Exp	Obs	Exp	Obs	Obs	Exp
A	31	8.77	18	6.74	8.36	144	150.40
T	57	19.77	68	25.11	30.73	127	119.91
G	7	13.87	11	16.21	14.93	35	36.02
C	5	4.37	2	4.08	4.50	8	13.86

Random is the average number of enriched regions observed among four random sequences with the same base composition as the *ITS 2* of *I. pacificus*.

If p is the frequency of the base of interest, q is the frequency of all other bases, and s is the sequence length, then, assuming a random distribution of bases, the expected number of regions enriched with a given base is $[p^5 + 5p^4q][s - 4]$ where the first bracketed term is the probability that a random sequence will be enriched for the base while the second bracketed term is the number of different but overlapping 5-base regions within the sequence.

second largest insertion, 10 bases long, are capable of intramolecular hydrogen bonding (Fig. 2).

Intraspecific variation

With the exception of one transition and one transversion (in *I. pacificus*), all intrapopulation deviations from the appropriate consensus were present in only one sequence; that is, virtually all intraspecific sequence variation is unique to single sequences.

The consensus sequence folded transcript of *I. pacificus* has 59 bases in secondary structures for which there is no hydrogen bonding, loops at the ends of stems and intersections between stems, and 668 bases in stems, 73.4 per cent (= 490/668) of which are intramolecularly paired and 26.6 per cent (= 178/668) of which are present as bulges from stems in the folded transcript (Table 4). However, intraspecifically variable sites are disproportionately represented among the loops and intersections ($\chi^2_1 = 17.98$, $P < 0.001$). In *I. scapularis*, the number of intraspecifically variable sites is too small for a similar analysis although the number of transitions and transversions within loops and intersections is twice the number expected, based on the proportion of total bases within these structures (observed = 5, expected = 2.45; Table 4).

However, intraspecific variation is not disproportionately represented in the sequence regions that form interspecifically conserved stem bulges in the folded

Table 3 Insertions/deletions and their 5' and/or 3' flanking sequences

Stem/loop		5' flanking	Deletion/insert	3' flanking
4A	P	TG TG <u>TCG</u>	<u>TCG</u>	
	S	AAAC	CAAAAAAAAAATC	TTTC
	P	CTTTCTTGCGAT		
		CGTTT-C-AT	-GTTCTT-CG--	TAGCAAATGCACG-AGT
		GAGCAAGT-ACGGACGT	GTGCGATTGCACG-TGT	
4B	P	AA	AA	A
	P	TTTTCTT	TG	GGCGTGG
	P	<u>G</u> A G	A <u>GA</u>	
	P	<u>GAT</u> G	AT	
	P	T	TTT	
	P	AC	AC	AC
	P		CTT CTT GCT	
	S	<u>ACGCTTC</u>	<u>ACGCTTC</u>	
	P	GGAGGAAAGT	GAC TT	TTTTT
	8L	S	TTTTTTTGTT	TTTT GACATG
8B/9A	P	TTTT	TATTGTG	TTGG
2B/9B	P	AAAAAA	AAA GAT	TGATT

Stem and loop numbers refer to designations in Fig. 1 (also, Table 1).

P, for *Ixodes pacificus*, and S, for *I. scapularis*, identify the species in which the 'insertion' occurs.

Spaces and underlining emphasize repeats while dashes indicate missing bases in a possible repeat sequence.

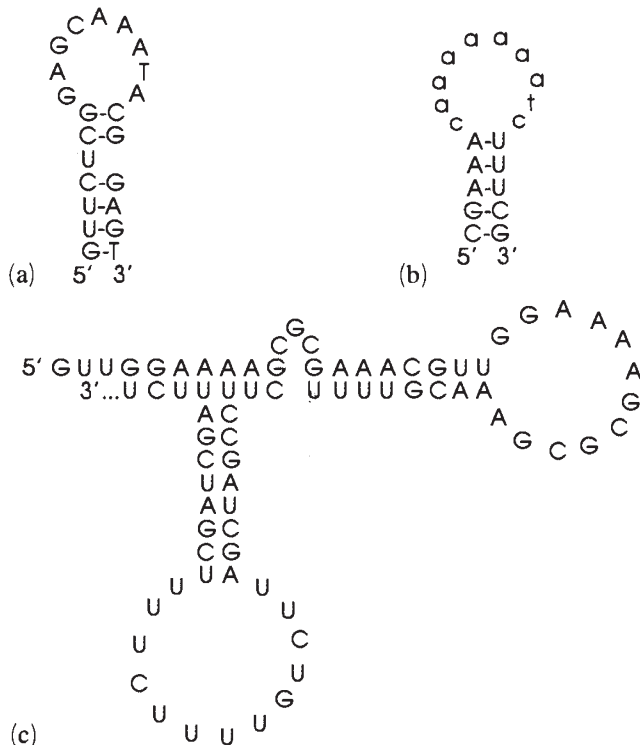


Fig. 2 Potential intramolecular pairing associated with indels and putative subrepeats. (a) Possible stem structure associated with 22-base interspecific indel. (b) Stem resulting from folding of sequences flanking 10-base interspecific indel (lower-case). (c) Folded transcript composed of two tandem copies of the putative 18- and 21-base subrepeat (see Table 7).

Table 4 Number of intraspecific transitions and transversions as a function of secondary structure

Species	Stems	Intersections	Loops
<i>Ixodes pacificus</i>			
No. bases	667	22	40
P(bases)	91.5%	3.0%	5.5%
Differences	72	10	8
P(differ.)	80.0%	11.1%	8.9%
Ratio	0.87	3.70	1.62
<i>I. scapularis</i>			
No. bases	629	31	27
P(bases)	91.5%	4.5%	4.0%
Differences	24	5	0
P(differ.)	82.8%	17.2%	0%
Ratio	0.90	3.82	0.00

P(bases) is the percentage of all bases in a structural feature, P(differ.) is the percentage of total sequence differences (from the appropriate consensus), and ratio = P(differ.) / P(bases).

minimum energy transcript (*I. pacificus*: $\chi^2_1 = 3.59$, $P > 0.05$; *I. scapularis*: $\chi^2_1 = 0.02$, $P > 0.50$). Yet these regions are expected to experience no sequence specific selective constraint as they do not participate in intramolecular base pairing.

The kind of sequence variation between sequences within a single locality is similar to that observed

Table 5 Type of sequence variation as a function of the level of comparison

Species	Source of variation	N	Transitions	Transversions	Indels (bases)		
					1	2	> 2
<i>I. pacificus</i>	Within CA population	3	18	15	15	7	5
	Between populations	2	10	7	7	3	2
<i>I. scapularis</i>	Within MA population	3	8	2	0	1	0
	Within GA population	4	8	2	2	0	0
<i>I. pacificus</i> / <i>I. scapularis</i>	Between species	2	21	23	18	6	10

N is the number of sequences compared.

Consensus sequences are compared between populations and species.

between locality consensus sequences and between species consensus sequences (Table 5). The level at which comparisons are made (within a population, between populations, or between species) is not associated with differences (1) in the frequency of transitions/transversions vs. indels ($\chi^2_2 = 1.09$, $P > 0.50$), (2) in the frequency of transitions vs. transversions ($\chi^2_2 = 2.68$, $P > 0.10$) and (3) in the length of indels, one, two, or more than two bases ($\chi^2_4 = 2.98$, $P > 0.50$). Note, however, that the number and size of indels can vary with the specifics of the alignment program, especially if sequence variation is sufficiently concentrated to render the alignment suspect.

Intraspecifically variable sites are significantly more likely than expected to lie near (i.e. within two bases) consensus interspecifically variable sites than farther away (i.e. more than two bases away). For instance, the consensus sequence for *I. scapularis* has 201 and 410 bases, respectively, within and greater than two bases from a site of interspecific variation. Yet, 44 of 56 intraspecific deviations from consensus (among 13 sequences) are within two bases of an interspecific sequence difference although the expected number is only 18.42 ($\chi^2_1 = 52.93$, $P < 0.001$). Similarly, in *I. pacificus* 66 of 119 intraspecifically variable sites (among four sequences) are near sites of consensus interspecific variation while the expected number is 39.15 ($\chi^2_1 = 27.44$, $P < 0.001$). The coincidence of intra- and interspecific variation has also been observed in mosquitoes (Wesson *et al.*, 1992).

Subrepeats within ITS 2

Small tandem repeats are rare within the ITS 2 of either *I. scapularis* or *I. pacificus*. In the consensus sequence of *I. scapularis*, the small subrepeat TCC is observed four times in tandem beginning at base position 170; two of these are found in a bulge from a stem. In *I. pacificus*, the same set has apparently mutated to TCC

TCT TCC TAC, all of which can intramolecularly base pair in the corresponding transcript. No other subsequence appears more than three times in tandem, unless degeneracies are assumed (e.g. *I. pacificus*: at base 13, GA GA G- GA GA; at base 61, GCA GCA -C- GCA; at base 407, GTTG GTTG Gcgc GTTG; e.g. *I. scapularis*: at base 13, GA GA G- GA; at base 61, GCA GCA -C- GCA; at base 394, GTgGGa cgga GTTGGC GTTGGC gc GTTGaa; at base 635, TCGAA TCGAAa gCGAt TgcAA; at base 656, Gc GT GT GT AC A- AC A- AC).

However, the number of nontandem simple repeats within the ITS 2 is higher than expected by chance. The numbers of 5-base subrepeats (with $n > 4$) in the consensus ITS 2 of *I. scapularis* and *I. pacificus* are, respectively, six and eight, with three common to both species (Table 6). The observed number of subrepeats with $n > 4$ is far outside the 95 per cent confidence interval, 0.323–1.28, derived from 10 random 700-base sequences with the same base composition as the ITS 2 in *Ixodes* spp.

Subrepeat sequences have a significantly higher proportion of A/T bases than does the remainder of the ITS 2. For *I. scapularis*, the proportion of A/T among 166 bases in subrepeats is 0.657 while the proportion for the remainder of the ITS 2 is 0.495 ($\chi^2_1 = 13.33$, $P < 0.001$). For *I. pacificus*, the proportion of A/T among 178 bases in subrepeats is also 0.657 while the proportion for the remainder of the ITS 2 is 0.486 ($\chi^2_1 = 15.79$, $P < 0.001$).

The dispersed repeats listed in Table 6 are not necessarily distributed independently of each other; the 5' end of one sequence may share a site with the 3' end of another sequence. This suggests that some repeats are parts of larger repeats that have degenerated. Two degenerated sequences appear to occur dispersed and in tandem within the ITS 2 of both *I. scapularis* and *I. pacificus* (Table 7). The consensus sequences are: (1) 5'-GTTTTCTTTTCGATCGATTCT-3' (= 21 bases),

Table 6 Dispersed repeats with a copy number of at least five

Species	Sequence	Location in <i>ITS 2</i>
<i>Ixodes scapularis</i>	GTTTT	26 38 71 94 188 349 522 695 720
	GCGTT	87 92 302 411 419 456
	TTTCT	28 81 96 330 351 624
	GGAAA	212 243 385 518 734
	TTTGT	23 238 631 689 723
	AAACG	215 311 324 425 494
<i>I. pacificus</i>	GTTTT	23 38 71 94 236 349
	TTTCT	28 81 330 351 502 624
	TTTGT	23 73 189 238 631 692
	TTTTG	39 72 237 355 463 531 630 697
	CGATG	88 144 225 575 647
	CGTTG	275 303 420 457 609
	GTTGG	304 407 411 610 695 731
	TTTTT	528 628 672 688 721

Location of repeats refers to the position in Table 1.

which has internal tandem repetition, NTTTT NTTTT and CGAT CGAT; and (2) 5'-GTTGGAAAAGCGC-GAAAC-3' (= 18 bases), which may have degenerate internal tandem repetition, GGAAA^{AGC} G^CGA-AAC.

The GTTTTCTTTTCGATCGATTCT repeat could largely be composed of the following 5-base repeats (of $n > 4$; Table 6): GTTTT, TTTCT and CGATG. The GTTGGAAAAGCGCGAAAC repeat could be composed, in part, of other dispersed repeats, CGTTG and GGAAA (all of $n > 4$; Table 6).

The 5' first 15 bases of the putative, dispersed 21-base repeat (see above) approximate the inverse of the putative, dispersed 18-base repeat; compare GTTT---CTTTTCGATC (21-base repeat) to GTTTCGCGCTTTTCCAAC (inverse of 18-base repeat; where underlining indicates sequence match). Consequently, tandem copies of the repeats are capable of intramolecular base pairing with one another (Fig. 2). In fact, a sequence composed of two tandem copies of each repeat pairs internally to a significantly greater extent than random sequences of the same length and base composition (extent of intramolecular pairing of tandem repeats = 64.1 per cent; 95 per cent CI for 10 random sequences = 35.2–43.5 per cent).

The alignment of either the *I. pacificus* or *I. scapularis* *ITS 2* with a hypothetical *ITS 2* composed only of 21- and 18-base subrepeats matches at 60 per cent of the bases. By comparison, the identity of *I. pacificus* with *I. scapularis* is 80.6 per cent. The hypothetical *ITS 2* has 21- and 18-base subrepeats in the following

order: 18, 21, 18, 21, 21, 21, 21, 21, 18, 21, 18, 21, 18, 18, 18, 21, 18, 18, 18, 21, 18, 18, 21, 21, 18, 18, 18, 21, 18, 18, 21, 18, 18. Except for (1) the first 120 bases and (2) the gain/loss of an interior subrepeat, the hypothetical *ITS 2* is composed of three tandem copies of this set of 21- and 18-base subrepeats: 21, 21, 18, 18, 18, 21, 18, 18, 18.

When the hypothetical *ITS 2* is folded, the conformation is very similar to that of *I. scapularis* (Fig. 3). The structure has an energy of -184 kcal/mol and has intramolecular pairing among 440 of 693 bases. On average, the *I. scapularis* and *I. pacificus* transcripts pair at 467 of 703 bases.

Base pairing and energy

Ten random sequences of 380 bases, the length of stem no. 4 (see Fig. 1) in *I. scapularis*, and of the same base composition as the *I. scapularis* *ITS 2* sequence were folded into their minimum energy conformations. The 95 per cent confidence interval (CI) for number of pairing bases in the random sequences was 231.4 to 265.8 and included the number observed for stem no. 4 of *I. scapularis*, 248. However, the 95 per cent CI for the minimum energy (kcal/mol), -88.33 to -76.25 , did not include that observed for stem no. 4 of *I. scapularis*, -115.4 . There was no relationship between the number of paired bases and energy among the random sequences ($R^2 = 0.01$, $F_{1,8} = 0.08$, $P = 0.78$).

Discussion

Transcript conformation and natural selection

The *ITS 2* transcripts of *I. scapularis* and *I. pacificus* average 703 bases in length with 2/3 of these capable of intramolecular pairing if the transcript assumes its minimum energy conformation. When the transcript sequences are aligned, they match at 600 positions. Consequently, the folded secondary structures are similar. Yet little in the data argues for selective constraint on either the base sequence or the particular transcript conformation.

For instance, there is no evidence of compensatory base evolution to maintain a high degree of intramolecular pairing between bases within stems of the folded transcript. Nor is there evidence that the rate of base substitution is less within stem regions that otherwise engage in intramolecular pairing vs. stem bulges that do not. The only evidence of natural selection on the base sequence is the significant over-representation of intraspecifically variable sites in stem loops and intersections between stems. But even this evidence is

Table 7 *Ixodes pacificus* sequence, 5' to 3', arranged to reveal subsequences representing degenerate tandem and dispersed copies of putative 21-base and 18-base subrepeats (in bold type)

Subrepeat spacers	21-base subrepeat	18-base subrepeat
	GTTTTCCTTTTCGATCGATTCT	GTTGGAAAAGCGCGAAAC
ATc AT AT ATc A- AG AG AG -G AG AT tT CAG CAcG CAG	GTTTTCTac ^c TCGt ^t TtGAcTgT CTTTTgTT ^{Bc} TT ^c CttgCGATgC- GTTTTC-----AT-G-TTCT	GTcGGAt---CGtGgg--
TCGGAG caag TaCGGAG		GTT ^B GG ^c AAAtGCaCGA--- GT-GGt---GCCAt ^{gc} AC G-TGt---GCCG ¹ GA ^{gcc} AA-
cAAAAAAAA	GTccTCTT-cCtAc ^c CG---- GTTT----gTCG-TCGA ^{ac} ATCT -----TCGATgGA ^{gaccg} TTtT ---TCTTT-CG ^c A-CG-TT-- --TtGTTgaCG-TCGgTgC-	c ^a TaGGAAAA-GCGAAA- GT-GGAAAAtCcCGAAAC ---GAAcgGCGCGc--- GTTGGAAAAaCG-GAgA- -TtGAAA-CG-----
GGAc GGA	GTTT-CTTTTCGATCGATTCT GTTTTCTTTT ^{BBS} CG-TgGAT ^{BTC} - cTTTgggga-CG ^g AT-GAT---	GTTcGAAA-G ^a C----- GT ^T TGcAAt-----
g CA -A CA CA CTT CTT	GaaaTTCTgT ^{Bg} CG-A-GATTtT	GcTGG--gGaGaGAA-C GTT ^{ct} GGgA---G-GAAA-
gtga	cTTTTtTTg-CG-TCG ¹ A ^c TTC- -----CG-TC-AgTCT	
aa gg AGCC AtCC a	Gac ^c TTC ^{Bcgg} TcccCGAT-GAaT ^B CT	GTaGGggtA ¹ Gcc----- GTTGGAtt-GCGCG----
GA GA Gc GA tgcatagc GT GT GT AC A- AC tca	-cTTTCTTTTtTgTCGAaTC-	
TTTTTTT TTTTTaTTgT	-TTTT-TTTTCG--CcATcCT	GTTGGAtAAGaGCGAAA- GTTGGAAA-GCGCaAAA-

Note that subrepeats are sometimes separated by simpler tandem repeats.

Upper case letters designate bases that match when subsequences are aligned with a repeat consensus; lower case letters designate mismatches.

Superscripted letters designate bases that are inserted relative to a consensus sequence.

equivocal. These regions are expected to be free of selective constraint since they do not base pair. Yet, it is possible that any given stem loop or pair of inter-sections flanking a stem could be smaller and the stem correspondingly longer if bases in these non-stem structures could pair among themselves. Thus, any base changes at the ends of stems could reduce intramolecular pairing, leading to larger loops and inter-sections that are then, necessarily, enriched for base changes.

Of course, the foregoing analyses are valid only if the folded transcript assumes a conformation identical or very similar to the minimum energy conformation. This is not the case for some rDNA sequences (Brimacombe *et al.*, 1983). However, among *Aedine* (Wesson *et al.*, 1992) and *Anopheline* mosquitoes (Paskewitz *et al.*, 1993), the minimum energy structure for the *ITS 2* shares core structural features with other constrained, low energy structures. The tick minimum energy structure resembles the mosquito *ITS 2* conformation (Fig.

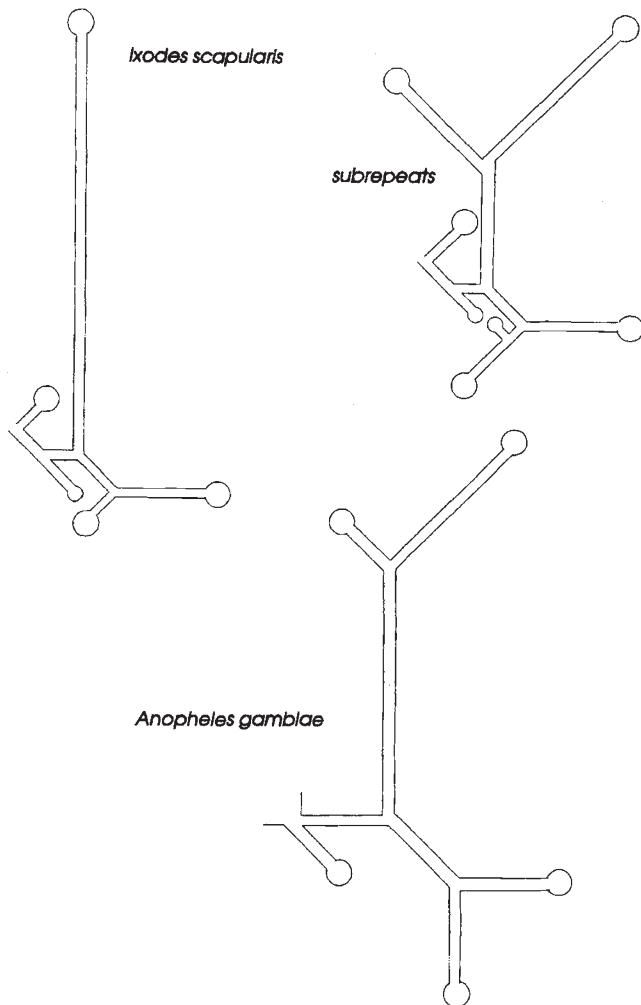


Fig. 3 Schematic representations of folded transcripts of (1) the *ITS 2* of *Ixodes scapularis*, (2) an artificial *ITS 2* composed of putative subrepeats (see text and Table 7), and (3) the *ITS 2* of *Anopheles gambiae* (after Paskewitz *et al.*, 1993).

3). Moreover, the tick structure contains the central stem (no. 4) that is observed in most animals (Yeh & Lee, 1990). Thus, the minimum energy foldings of the *Ixodes* spp. transcripts probably closely approximate the actual conformation attained.

In any case, the *ITS 2* experiences relaxed selection for sequence conservation. For instance, the aligned consensus sequences of *I. scapularis* and *I. pacificus* differ by 14.7 per cent (1-[600/703]) compared to only a 0.65 per cent difference for the adjacent 5.8S gene (one difference among 154 bases; Wesson *et al.*, 1993).

The data do not address whether or not the secondary conformation is conserved independently of the base sequence because the similarity in *ITS 2* transcript conformation between *I. scapularis* and *I. pacificus*

must, in part, simply reflect their sequence similarity. Thus, the presence of many shared stem bulges with a high degree of sequence conservation suggests the recent divergence of *I. scapularis* and *I. pacificus* from their common ancestor, not the selective conservation of these features. The great number of stem bulges that are unshared between the species and the presence of bulges of variable occurrence within a species argue against any functional significance for bulges.

Base changes and indels, whether between or within species, are usually absorbed as bulges from stems and do not cause the appearance or disappearance of particular stem structures. But this is not surprising given the high degree of intramolecular pairing that is already present and that constrains structural rearrangements. Some indels, however, are associated with new intramolecular base pairings. The presence of an insertion that hydrogen bonds with other bases is often associated with flanking bulges that represent bases that have been displaced from their pairing partners in the folded transcript (e.g. see insertions at positions 101, 475, 506, and 525 in Table 1). These occurrences and the large proportion of bases in stem bulges indicate that the basic conformation is resistant to change (Gonzalez *et al.*, 1985; Paskewitz *et al.*, 1993). Therefore, selection on secondary conformation should not be inferred from similarity in conformation.

The high degree of intramolecular pairing that occurs in the folded transcript does not constitute evidence of selection on the transcript conformation because random sequences intramolecularly pair to the same extent as observed in the *ITS 2* of the tick species. However, the observation that the energy of the tick folded transcript is much less than that for random sequences may be indicative of adaptation for stability of the secondary conformation.

However, low energy structures may evolve as a nonselective consequence of replication slippage. If replication slippage were facilitated by the ability of the 3' end of newly synthesized DNA to form low-energy fold-back structures, re-exposing the template strand to replication machinery, the copy number would increase for low energy subsequences. A number of such sequences might contribute to conformations of relatively low energy. In fact, the two largest duplications within the *ITS 2* involve sequences capable of intramolecular bonding.

Most suggestive of selection on the secondary conformation of the folded transcript is the fixation in one species relative to the other of compensating indels. But even here their number is few. For instance, in stem no. 4 (see Fig. 1; Table 1), a 22-base indel at positions 102–123 is paired in the folded transcript with a 10-base indel at positions 475–484. In *I. pacificus*, six

bases in the longer insertion occur as bulges and the insertion is flanked by a 4-base bulge. Thus, the remainder of the stem no. 4 can intramolecularly pair as it did before the indels occurred and without the introduction of new stem structures or especially large bulges.

In addition, a pair of indels is associated with the origin of stem no. 9 (see Fig. 1) in *I. pacificus* or its loss in *I. scapularis* (the loop is retained as a bulge). A single insertion/deletion event may have selected for a compensating indel to preserve the stems no. 8 and no. 2. Both stems are near points where *ITS 2* is excised from the 45S transcript (Reddy *et al.*, 1983) and may be necessary for efficient transcript processing as certain conformational features facilitate cleavage (Savino & Gerbi, 1990).

Natural selection on the conformation of the folded *ITS 2* transcript is suggested by the occurrence of compensating indels that preserve structural features and, perhaps, by the higher rate of change among bases within loops and intersections, where intramolecular base pairs do not form, vs. within stems where pairing is essential to maintain the structures. Natural selection is contraindicated by the high proportion of stem bases found in bulges that distort stems by introducing bends and twists, by the variable presence of stem no. 9, by the absence of compensatory base evolution, by the similar rate of change among bases in stem trunks vs. stem bulges, and by the overall high rate of base change in the *ITS 2* vs. in the adjacent 5.8S gene. Thus, natural selection on the sequence and on the conformation of its folded transcript appears, at most, very weak.

Bias in base composition

The *ITS 2* sequence is composed almost equally of A/T vs. G/C bases. Yet, they are distributed differently as there are significantly more regions of five bases enriched for A or T than expected and less that are enriched for G or C than expected. A- or T-enriched regions may arise from replication slippage events. The weaker hydrogen bonding between A and T vs. G and C may render slippage more likely for A/T rich sequences. In fact, indels have a significantly higher A/T composition than the *ITS 2* sequence as a whole suggesting that replication slippage may be more likely to duplicate or delete A- or T-rich subsequences. Also, where an insertion is associated with a tandem copy (see Table 3), suggesting its origin via slippage, the proportion of A/T is significantly higher than for the sequence as a whole. Moreover, dispersed subrepeats, that themselves may be parts of larger, degenerate tandem repeats, are significantly biased in their base composition towards A and T. Higher prevalence of

replication slippage in A/T rich regions is consistent with the significantly higher A/T ratio within variable vs. conserved sequence blocks.

Generation of A- or T-enriched regions via replication slippage does not address the under-representation of G- and C-enriched regions. Perhaps, natural selection for a flexible conformation acts against enrichment for G and C because of the stronger hydrogen bonding between these bases vs. between A and T. No under-representation of G- or C-enriched regions relative to A- or T-enriched regions is present in the greater than 4000 bases of the 28S genes of *Drosophila melanogaster* (using the sequence in Tautz *et al.* [1988]) ($\chi^2_1 = 0.85$, $P > 0.10$; Table 2). Consequently, the pattern observed for the tick *ITS 2* appears to be an adaptation for this sequence and not a general characteristic of rDNA.

Stem loops and intersections are often enriched for T. Thus, they may arise via replication slippage along initially small poly T tracts. Similar structures, stem bulges, appear sometimes to arise from slippage events as evidenced by the presence of copies of tandem repeats in bulges. Alternatively or in addition, there may be a mutational bias towards T in sequence stretches corresponding to secondary structural features under the least selective constraint, as is observed in *Drosophila melanogaster* (Hancock *et al.*, 1988). Mutational bias could explain the relatively high frequency of U throughout the *ITS 2* transcript.

Replication slippage

Relative to the consensus *ITS 2* sequences of each species, indels of greater than one base pair are associated with tandem repetition, albeit often degenerate. This suggests replication slippage as the source of mutations that increase or decrease the copy number of tandem repeats (Levinson & Gutman, 1987; e.g. Wesson *et al.*, 1992; Cummings *et al.*, 1994). The sequences of both *I. scapularis* and *I. pacificus* appear to contain tandem repeats of 2–21 bases in length. However, the copy number of any specific tandem repeat usually varies between but not within species, suggesting a high rate of fixation of slippage-induced gains and losses.

The *ITS 2* of *Ixodes* spp. is enriched with dispersed direct repeats relative to random sequences of the same base composition. The mosquito *ITS 2* also contains simple direct repeats (Wesson *et al.*, 1992). In ticks, some subsequences, putative 18- and 21-bp repeats, are composed of various smaller repeats that are common within the *ITS 2*. Thus, the ancestral *ITS 2* may have been composed of 18- and 21-bp repeats that have since mutationally deteriorated. Because the

5' first 15 bases of the dispersed 21-base repeat are approximately the inverse of the dispersed 18-base repeat, a high degree of intramolecular base pairing would have been possible in ancestral transcripts composed entirely of these subsequences.

Concerted evolution

The degree of intraspecific variation in the *ITS 2* sequences of both *I. scapularis* and *I. pacificus* is very small, three to four unique mutations per sequence, relative to the number of fixed differences between the two species, namely the 42 transitions, transversions, and indels. Thus, the many rDNA repeating units within a genome are evolving in concert (e.g. Stewart & Baker, 1994). Molecular drive is the proposed mechanism of concerted evolution (Dover, 1982), entailing the coincident (1) homogenization of multi-gene families via non-Mendelian turnover processes and (2) spread of repeat variants through populations via chromosome shuffling associated with sexual reproduction (Dover, 1986). Homogenization of large rDNA units is presumably driven by unequal crossing-over (e.g. Strachan *et al.*, 1985; reviewed in Dover, 1986, 1989).

Progressive homogenization of the rDNA multigene family for mutationally deteriorated sequences could give rise to concomitantly stronger natural selection against further homogenization if specific patterns of folding of the *ITS 2* transcript are required for effective processing of the 45S pre-rRNA (see Reddy *et al.*, 1983; Ware *et al.*, 1983; Gerbi, 1985; Fujiwara & Ishikawa, 1986). Thus, drive-intensified selection could impede the fixation of processing-ineffective mutants at some level of homogenization but would not act to impede the random fixation of one among a subset of processing-effective rDNA units that possess a long-term, ultimate advantage. The under-representation of C- and G-enriched regions and the presence of compensating indels that preserve major features of transcript conformation are consistent with some, probably weak, natural selection on the *Ixodes* spp. *ITS 2* sequences.

Conclusions

Our data suggest that the *ITS 2* is a mutationally degenerated array of tandem and dispersed subrepeats of approximately 20 bases in length that are themselves composed of various smaller repeats. In addition, the longer subrepeats may have been organized into only a few larger tandem repeats that together composed the great majority of the ancestral *ITS 2* (e.g. Paskewitz *et al.*, 1993). A consequence of the highly repetitious

nature of the *ITS 2*, coupled with the absence of sequence-specific selective constraint, is a high rate of sequence evolution, attributable in large part to replication slippage. Replication slippage is observed to occur within short, A/T-rich tandem subrepeats and is inferred to have contributed to the degeneracy of larger, composite subrepeats. Thus, while some slippage-induced indels may be selected to compensate for the effects of other indels on conformational energy, the sequence is largely free to evolve because intramolecular pairing permits many mutations to be absorbed, with respect to transcript conformation, as bulges from stems.

References

- AMSTUTZ, H., MUNZ, P., HEYER, W.-D., LEUPOLD, U. AND KOHLI, J. 1985. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell*, **40**, 879–886.
- ARNHEIM, N., KRYSZAL, M., SCHMICKEL, R., WILSON, G., RYDER, O. AND ZIMMER, E. 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 7323–7327.
- ARNHEIM, N., TRECO, D., TAYLOR, B. AND EICKER, E. M. 1982. Distribution of ribosomal gene length variants among mouse chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 4677–4680.
- BECKINGHAM, K. 1982. Insect rDNA. *Cell Nucl.*, **10**, 205–263.
- BONCINELLI, E., BORHESE, A., GRAZIANI, F., LA MANTIA, G., MANZI, A., MIRIANI, C. AND SIMEONE, A. 1983. Inheritance of the rDNA spacer in *D. melanogaster*. *Mol. Gen. Genet.*, **189**, 370–374.
- BRIMACOMBE, R., MALY, P. AND ZWIEB, C. 1983. The structure of ribosomal RNA and its organization relative to ribosomal protein. *Nucl. Acids Res. Mol. Biol.*, **28**, 1–48.
- COEN, E., STRACHAN, T. AND DOVER, G. A. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Biol.*, **158**, 17–35.
- CUMMINGS, M. P., KING, L. M. AND KELLOGG, E. A. 1994. Slipped-strand mispairing in a plastid gene: *rpo C2* in grasses (Poaceae). *Mol. Biol. Evol.*, **11**, 1–8.
- CURTIS, D. AND BENDER, W. 1991. Gene conversion in *Drosophila* and the effects of the meiotic mutants *mei-9* and *mei-218*. *Genetics*, **127**, 739–74.
- DALLAS, J. F., BARTON, N. H. AND DOVER, G. A. 1988. Interracial rDNA variation in the grasshopper, *Podisma pedestris*. *Mol. Biol. Evol.*, **5**, 660–674.
- DEVEREUX, J., HAEBERLI, P. AND SMITHIES, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.*, **12**, 387–395.
- DOVER, G. A. 1982. Molecular drive: a cohesive mode of species evolution. *Nature*, **299**, 111–117.
- DOVER, G. A. 1986. Molecular drive in multigene families: how biological novelties arise, spread, and are assimilated. *Trends Genet.*, **2**, 159–165.

- DOVER, G. A. 1989. Linkage disequilibrium and molecular drive in the rDNA gene family. *Genetics*, **122**, 249–252.
- DOVER, G. A., BROWN, S., COEN, E. S., DALLAS, J., STRACHAN, T. AND TRICK, M. 1982. The dynamics of genome evolution and species differentiation. In: Dover, G. A. and Flavell, R. B. (eds) *Genome Evolution*, pp. 343–372. Academic Press, New York.
- FREIER, S. M., KIERZEK, R., JAEGER, J. A., SUBIMOTO, N., CARUTHERS, M. H., NEILSON, T. AND TURNER, D. H. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9373–9977.
- FUJIWARA, H. AND ISHIKAWA, H. 1986. Molecular mechanism of introduction of the hidden break into the 28S rRNA of insects: implication based on structural studies. *Nucl. Acids Res.*, **14**, 6393–6401.
- GERBI, S. A. 1985. Evolution of ribosomal DNA. In: MacIntyre, R. J. (ed.) *Molecular Evolutionary Genetics*, pp. 419–517. Plenum, New York.
- GILLINGS, M. R., FRANKHAM, R., SPEIRS, J. AND WHALLEY, M. 1987. X–Y exchange and the coevolution of the X and Y rDNA arrays in *Drosophila melanogaster*. *Genetics*, **116**, 241–251.
- GONZALEZ, I. L., GORSKI, J. L., CAMPEN, T. J., DORNEY, D. J., ERICKSON, J. M., SYLVESTER, J. E. AND SCHMICKEL, R. D. 1985. Variation among human 28S ribosomal RNA genes. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 7666–7670.
- HANCOCK, J. M. AND DOVER, G. A. 1988. Molecular coevolution among cryptically simple expansion segments in eukaryotic 26S/28S rRNA. *Mol. Biol. Evol.*, **5**, 377–392.
- HANCOCK, J. M., TAUTZ, D. AND DOVER, G. A. 1988. Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol. Biol. Evol.*, **5**, 393–414.
- HASSOUNA, N., MICHOT, B. AND BACHELLERIE, J.-P. 1984. The complete nucleotide sequence of mouse 28S rRNA: implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucl. Acids Res.*, **12**, 3563–3583.
- KEIRANS, J. E., OLIVER, J. H. AND NEEDHAM, G. R. 1992. The *Ixodes ricinus/persulcatus* complex defined. In: Mundetloh, U. G. and Kurtti, T. J. (eds) *First International Conference on Tick-Borne Pathogens at the Host-Vector Interface*, p. 302. University of Minnesota, St. Paul.
- LACHANCE, M.-A. 1990. Ribosomal DNA spacer variation in the cactophilic yeast *Clavispora opuntiae*. *Mol. Biol. Evol.*, **7**, 178–193.
- LARSON, A. AND WILSON, A. C. 1989. Patterns of ribosomal RNA evolution in salamanders. *Mol. Biol. Evol.*, **6**, 131–154.
- LEVINSON, G. AND GUTMAN, G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.*, **4**, 203–221.
- LIVAK, K. 1984. Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics*, **107**, 611–634.
- LOHE, A. R. AND ROBERTS, P. A. 1990. An unusual Y chromosome of *Drosophila simulans* carrying amplified rDNA spacer without rDNA genes. *Genetics*, **125**, 399–406.
- LYCKEGAARD, E. M. S. AND CLARK, A. G. 1991. Evolution of ribosomal RNA gene copy number on the sex chromosomes of *Drosophila melanogaster*. *Mol. Biol. Evol.*, **8**, 458–474.
- MADDERN, R. H. 1981. Exchange between ribosomal RNA genes of X and Y chromosomes in *Drosophila melanogaster* males. *Genet. Res.*, **38**, 1–7.
- McLAIN, D. K., WESSON, D. M., OLIVER, J. H., JR. AND COLLINS, F. H. 1995. Variation in rDNA ITS 1 among eastern populations of *Ixodes scapularis* (Acari: Ixodidae). *J. Med. Entomol.*, **32**, 353–360.
- MICHOT, B. AND BACHELLERIE, J.-P. 1987. Comparisons of large subunit rRNAs reveal eukaryote-specific elements of secondary structure. *Biochimie*, **69**, 11–23.
- MULLIS, K. B., FALOONA, F. A., SCHARF, S. J., SAIKI, R. K., HORN, G. T. AND ERLICH, H. A. 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 263–273.
- OHTA, T. 1983. On the evolution of multigene families. *Theor. Pop. Biol.*, **23**, 216–240.
- OHTA, T. AND DOVER, G. A. 1984. The cohesive population genetics of molecular drive. *Genetics*, **108**, 501–521.
- OLIVER, J. H., JR., OWSLEY, M. R., HUTCHESON, H. J., JAMES, A. M., CHEN, C., IRBY, W. S., DOTSON, E. M. AND McLAIN, D. K. 1993. Conspecificity of the ticks *Ixodes scapularis* and *I. dammini* (Acari: Ixodidae). *J. Med. Entomol.*, **30**, 54–63.
- OLSEN, G. J., McCARROL, R. AND SOGIN, M. L. 1983. Secondary structure of *Dictyostelium discoideum* small subunit ribosomal RNA. *Nucl. Acids Res.*, **11**, 8037–8049.
- PASKEWITZ, S. M., WESSON, D. M. AND COLLINS, F. H. 1993. The internal transcribed spacers of ribosomal DNA in five members of the *Anopheles gambiae* species complex. *Insect Mol. Biol.*, **2**, 247–257.
- REDDY, R., ROTHBLUM, L. I., SUBRAHMANYAM, C. S., LIU, M.-H., HENNING, D., CASSIDY, B. AND BUSCH, H. 1983. The nucleotide sequence of 8S RNA bound to preribosomal RNA of Novikoff hepatoma. *J. Biol. Chem.*, **258**, 584–589.
- ROBBINS, L. G. AND SWANSON, E. E. 1988. Rex-induced recombination implies bipolar organization of the ribosomal RNA genes of *Drosophila melanogaster*. *Genetics*, **120**, 1053–1059.
- SAMBROOK, J., FRITSCH, E. F. AND MANIATIS, T. 1989. *Molecular Cloning*. Cold Spring Harbor Press, Plainview, NY.
- SANGER, F., NICKLEN, S. AND COULSEN, A. R. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5464–5468.
- SAVINO, R. AND GERBI, S. A. 1990. *In vivo* disruption of *Xenopus* U3 snRNA affects ribosomal RNA processing. *EMBO J.*, **9**, 2299–2308.
- SPIELMAN, A., CLIFFORD, C. M., PIESMAN, J. AND CORWIN, M. D. 1979. Human babesiosis on Nantucket Island, USA: description of the vector, *Ixodes (Ixodes) dammini*, n. sp. (Acarina: Ixodidae). *J. Med. Entomol.*, **15**, 218–234.
- STEWART, D. T. AND BAKER, A. J. 1994. Patterns of sequence variation in the mitochondrial D-loop region of shrews. *Mol. Biol. Evol.*, **11**, 9–21.
- STRACHAN, T., WEBB, D. AND DOVER, G. A. 1985. Transition stages of molecular drive in multi-copy DNA families in *Drosophila*. *EMBO J.*, **4**, 1701–1708.

- TAUTZ, D., HANCOCK, J. M., WEBB, D. A., TAUTZ, C. AND DOVER, G. A. 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol. Biol. Evol.*, **5**, 366-376.
- TAUTZ, D., TAUTZ, C., WEBB, D. AND DOVER, G. A. 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. *J. Mol. Biol.*, **195**, 525-542.
- THWEATT, R. AND LEE, J. C. 1990. Yeast precursor ribosomal RNA: molecular cloning and probing the higher-order structure of the internal transcribed spacer 1 by kethoxal and dimethylsulfate modification. *J. Mol. Biol.*, **211**, 305-320.
- WALSH, J. B. 1987. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. *Genetics*, **115**, 553-567.
- WARE, V. C., TAGUE, B. W., CLARK, C. C., GOURSE, R. L., BRAND, R. C. AND GERBI, S. A. 1983. Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*. *Nucl. Acids Res.*, **11**, 7795-7817.
- WESSON, D. M. AND COLLINS, F. H. 1992. Sequence and secondary structure of 5.8S rRNA in the tick, *Ixodes scapularis*. *Nucl. Acids Res.*, **20**, 1139.
- WESSON, D. M., McLAIN, D. K., OLIVER, J. H., PIESMAN, J. AND COLLINS, F. H. 1993. Investigation of the validity of species status of *Ixodes dammini* (Acari: Ixodidae) using rDNA. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 10 221-10 225.
- WESSON, D. M., PORTER, C. H. AND COLLINS, F. H. 1992. Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). *Mol. Phylogen. Evol.*, **1**, 253-269.
- WILLIAMS, S. M., KENNISON, J. A., ROBBINS, L. G. AND STROBECK, C. 1989. Reciprocal recombination and the evolution of the ribosomal gene family of *Drosophila melanogaster*. *Genetics*, **122**, 617-624.
- YEH, L.-C. C. AND LEE, J. C. 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. AND STEIGLER, P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.*, **9**, 133-148.