Mitochondrial Gene Order is Not Conserved in Arthropods: Prostriate and Metastriate Tick Mitochondrial Genomes

William C. Black IV and Richard L. Roehrdanz†*

*Department of Microbiology, Colorado State University; and †United States Department of Agriculture, Agricultural Research Service, Biosciences Research Laboratory, Fargo, North Dakota

The entire mitochondrial genome was sequenced in a prostriate tick, *Ixodes hexagonus,* and a metastriate tick, *Rhipicephalus sanguineus.* Both genomes encode 22 tRNAs, 13 proteins, and two ribosomal RNAs. Prostriate ticks are basal members of Ixodidae and have the same gene order as *Limulus polyphemus*. In contrast, in *R. sanguineus,* a block of genes encoding NADH dehydrogenase subunit 1 (ND1), $tRNA^{Leu\ (UUR)}$, $tRNA^{Leu\ (CUN)}$, 16S rDNA, $tRNA^{Val}$, 12S rDNA, the control region, and the $tRNA^{Ile}$ and $tRNA^{Gln}$ have translocated to a position between the $tRNA^{Glu}$ and $tRNA^{Phe}$ genes. The $tRNA^{Cys}$ gene has translocated between the control region and the $tRNA^{Met}$ gene, and the $tRNA^{Leu\ (CUN)}$ gene has translocated between the $tRNA^{Ser\ (UCN)}$ gene and the control region. Furthermore, the control region is duplicated, and both copies undergo concerted evolution. Primers that flank these rearrangements confirm that this gene order is conserved in all metastriate ticks examined. Correspondence analysis of amino acid and codon use in the two ticks and in nine other arthropod mitochondrial genomes indicate a strong bias in *R. sanguineus* towards amino acids encoded by AT-rich codons.

Introduction

The metazoan mitochondrial genome (mtDNA) is a single-circular, double-stranded DNA molecule that ranges in size from 14–42 kb (Wolstenholme, 1992; Wolstenholme and Fauron 1995). Most or all of mtDNA has been sequenced in more than 50 metazoa. The phylum Arthropoda is the most speciated of all metazoa phyla and contains the large classes Insecta and Crustacea and the chelicerate class Arachnida. Mites (Acari) form a large subclass in Arachnida. The complete mtDNA sequence has been determined for six insects: *Drosophila yakuba* (Clary and Wolstenholme 1985), *D. melanogaster* (Lewis, Farr, and Kaguni 1995), *Anopheles gambiae* (Beard et al. 1993), *A. quadrimaculatus* (Mitchell, Cockburn, and Seawright 1993), *Apis mellifera* (Crozier and Crozier 1993), *Locusta migratoria* (Flook, Rowell, and Gellissen 1995), and one branchiopod crustacean, *Artemia franciscana* (Valverde et al. 1994). Complete sequence data has not been obtained from any chelicerate species; however, gene order has been identified in a number of taxa through partial sequences (Staton, Daehler, and Brown 1997; Boore, Lavrov, and Brown 1998).

Ticks are obligate hematophagous ectoparasites of a wide variety of terrestrial vertebrates, including amphibians, reptiles, birds, and mammals. Ticks form a suborder (Ixodida) in the order Parasitiformes of the subclass Acari. This suborder consists of two families, the Argasidae (soft ticks) and the Ixodidae (hard ticks). The Ixodidae is split into five subfamilies (Ixodinae, Amblyomminae, Haemaphysalinae, Hyalomminae and Rhipicephalinae). An important division of the Ixodidae consists of prostriate (Ixodinae) and metastriate ticks (the other four subfamilies). Analysis of the entire 18S rDNA gene provided strong support for monophyly of Argasidae and Ixodidae (Black, Klompen, and Keirans 1997). We further demonstrated monophyly of Ixodinae and Amblyomminae. However, few characters separated Haemaphysalinae or Rhipicephalinae, and, as in other studies (Black and Piesman 1994), Hyalomminae arose within Rhipicephalinae. We have sought to use other mtDNA sequences to obtain additional characters with which to resolve phylogenetic relationships among metastriate subfamilies (Black and Piesman 1994; Black, Klompen, and Keirans 1997). In the course of these studies, we have consistently had problems amplifying regions of the mtDNA using primers that are considered conserved within arthropods (Simon et al. 1994). We have sequenced the entire mitochondrial genome of a prostriate hard tick, *Ixodes hexagonus,* and a metastriate tick, *Rhipicephalus sanguineus,* to resolve this problem.

With few exceptions, the metazoan mtDNA contains 37 genes that encode 13 protein subunits, 22 tRNAs, and a small and a large ribosomal subunit RNA. Boore et al. (1995) and Boore, Lavrov, and Brown (1998) have shown that mtDNA gene order is generally highly conserved within arthropod groups and suggest that gene arrangements could be used to deduce deep-level phylogenetic relationships. We report here on extensive mtDNA gene rearrangements that occur in all metastriate ticks and that deviate dramatically from the ancestral arthropod gene order inferred by Boore et al. (1995). These rearrangements and a strong bias toward amino acids encoded by AT-rich codons may explain the failure of conserved arthropod primers to amplify many mitochondrial genes in metastriate ticks.

Materials and Methods

Total genomic DNA was isolated from a female specimen of *I. hexagonus* provided by L. Jones from a colony maintained at the Institute of Virology and Environmental Microbiology (IVEM; Oxford, Britain). DNA was isolated by a CTAB (hexadecyltrimethylammonium bromide) procedure (Black, Klompen, and Keirans 1997). Melanie Palmer (Department of Entomology,

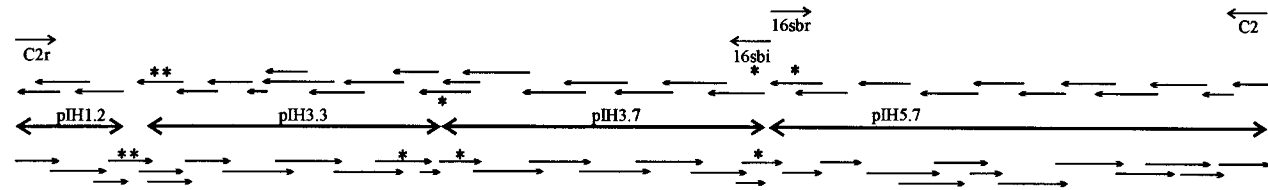
Key words: arachnid mitochondrial DNA, gene rearrangements, hard ticks, *Ixodes hexagonus, Rhipicephalus sanguineus,* amino acid and codon usage bias.

Address for correspondence and reprints: William C. Black IV, Department of Microbiology, Colorado State University, Fort Collins, Colorado 80523. E-mail: wcb4@lamar.colostate.edu.

Mol. Biol. Evol. 15(12):1772–1785. 1998
© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

Ixodes hexagonus



1,000 bp

Rhipicephalus sanguineus

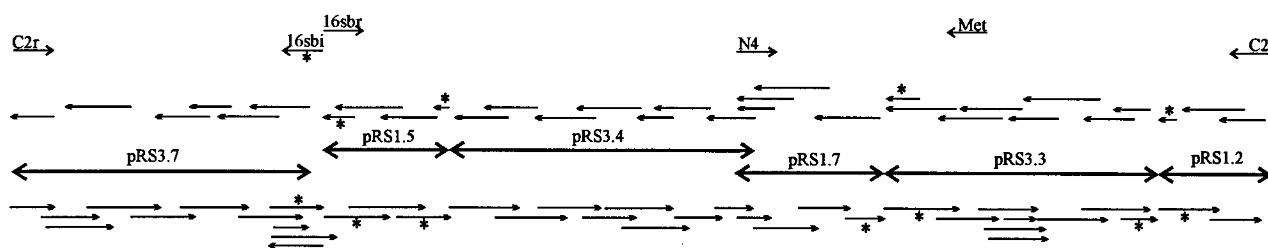


FIG. 1.—Locations of primers used for amplification of the *Ixodes hexagonus* and *Rhipicephalus sanguineus* mitochondrial genomes. The locations and sizes of forward and reverse sequences appear as arrows above and below the cloned fragments. Sequences obtained to confirm accurate joining of cloned fragments are indicated with an "*" or a "**." An '*' appears near sequences obtained by amplifying and sequencing the joined region directly from tick genomic DNA using primers located close to the ends of the cloned fragments. Sequence of the joined regions indicated with "**" was obtained using gel-purified long-PCR product as the template. Sequencing primers were located near the ends of the cloned fragments. The C2 and C2r primers are complementary and were found in the cloned sequences. However, a *Hind*III site was located ~200 bp 5′ to the 16Sbi in both species and was not, therefore, present in the cloned regions. pRS3.4 and pRS1.7 overlap, obviating the need to sequence the joined ends of these clones.

Oklahoma State University [OSU]) provided *R. sanguineus* DNA purified on a CsCl gradient.

The entire mitochondrial genome was amplified in two pieces in *I. hexagonus* and was amplified in three pieces in *R. sanguineus* (fig. 1) with the GeneAmp XL Kit (Perkin Elmer, Foster City, Calif.), following protocols and primers outlined in Roehrdanz (1995). In *I. hexagonus,* a 5.7-kb region was amplified with the 16Sbr primer (LR-J-12883; Xiong and Kocher 1991) from the 5′ end of the 16S rRNA gene and with the C2 primer (C2-N-3662; Simon et al. 1994) from the 3′ end of the cytochrome oxidase 2 (COII) gene. Both primers were modified to contain a CGC-clamp and a *Sal*I site in the 5′ end: 16Sbr (5′-CGC GTC GAC TCC GGT TTG AAC TCA GAT C-3′), C2(5′-CGC GTC GAC CAC AAA TTT CTG AAC ATT GAC C-3′). The remaining 8.8 kb of the mitochondrial genome was amplified with the 16Sbi primer (LR-N-12868; Simon et al. 1994) from the 5′ end of the 16S rRNA gene and with the C2r primer (C2-J-3684) partially complementary to the C2 primer. Both primers were modified to contain a CGC-clamp and a *Hind*III site in the 5′ end: 16Sbi (5′-CGC AAG CTT ACA TGA TCT GAG TTC AAA CC-3′), C2r (5′-CGC AAG CTT GGT CAA TGT TCA GAA ATT TGT GG-3′). The 16Sbr–C2–amplified fragment was purified using a trough elution procedure (Sambrook, Fritsch, and Maniatis 1989), mixed with 100 ng of pBluescript II KS (Stratagene), digested with *Sal*I, ligated, and transformed into competent DH5α *Escherichia coli* cells (Gibco BRL Inc.). The cloned fragment was labeled pIH5.7. The 16Sbi–C2r fragment contained four *Hind*III sites that produced five fragments when digested, of the following lengths: 1,237 bp; 281 bp; 3,346 bp; 3,741 bp; and 237 bp. The three largest fragments were cloned and were labeled pIH1.2, pIH3.3, and pIH3.7. The smaller regions were not cloned and were instead sequenced directly from trough-eluted polymerase chain reaction (PCR) product.

The 16Sbi–C2r primers amplified a 3,681-bp region in *R. sanguineus* that contained a *Hind*III 169 bp from the 16Sbi primer (fig. 1). The clone containing this fragment is pRS3.7, and the 169-bp region was sequenced directly from trough-eluted PCR product. The remaining 11.2 kb of the *R. sanguineus* mtDNA was amplified weakly with 16Sbr–C2 primers. Instead, this region was amplified in two overlapping pieces with C2 and ND4 primer (N4-J-8944; from Simon et al. 1994) from the middle of the NADH dehydrogenase subunit 4 (ND4) gene. The remainder was amplified using Met (5′-TGG GGT ATG AAC CCA GTA GC-3′) (TM-N-193; Simon et al. 1994) and 16Sbr. The ND4 primer was modified to contain a CGC-clamp and a *Xho*I site in the 5′ end: N4 (5′-CGC CTC GAG GAG CTT CAA CAT GAG CTT T-3′). The amplified C2–N4 fragment was digested with *Xho*I and *Sal*I to generate three fragments of the following lengths: 1,249 bp; 3,353 bp; and 1,727 bp. Clones containing these fragments are, respectively, as follows: pRS1.2, pRS3.3 and pRS1.7. The Met–16Sbr fragment was digested with *Sal*I and *Hind*III to clone a 1,475-bp fragment containing portions of the 16S and 12S genes (pRS1.5). The Met–16Sbr fragment was also digested with *Eco*RI and *Hind*III to clone a 3,525-bp fragment (pRS3.5) containing the remainder of the 12S, all of the NADH dehydrogenase subunit 5 gene (ND5), and a portion of ND4. The ND4 region overlapped with pRS1.7.

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

Portions of 16S, 12S, cytochrome oxidase I (COI), and COII genes were amplified directly from tick DNA. The 16S was amplified using a forward primer, 5′-CCG GTC TGA ACT CAG ATC AAG T-3′ (LR-J-12887; as per Simon et al. 1994 but with an A at position 20), and a reverse primer, 16S (- 1) (5′-CTG CTC AAT GAT TTT TTA AAT TGC TGT GG-3′) (Black and Piesman 1994). The 12S was amplified using a forward primer, 5′-TAC TAT GTT ACG ACT TA-3′ (as per Kambhampati 1995 but without the 3′-T), and a reverse primer, 5′-AAA CTA GGA TTA GAT ACC C-3′ (SR-N-14582; Simon et al. 1994). The COI/COII genes were amplified from COI–RLR (5′-TTG ATT TTT TGG TCA YCC WGA AGT-3′) (C1-J-2195; Roehrdanz 1993) and the C2 primer. Amplified products were purified with the Wizard PCR kit (Promega) and were sequenced to confirm that tick mitochondrial DNA was amplified, cloned, and sequenced throughout the various procedures described above.

DNA Sequencing

Plasmid DNA was extracted and purified from each clone using the Wizard midiprep kit (Promega). Purified DNA was used as the template for *Taq* Fluorescent Sequencing (FS) Dye Deoxy Terminator cycle-sequencing reactions (Applied Biosystems Inc.), and sequencing was performed on an automated DNA sequencer Model 377 (Applied Biosystems Inc.) at the Macromolecular Resources Laboratory at Colorado State University. Sequences from each end of a cloned insert were obtained with the M13, T3, or T7 universal sequencing primers. The sequence was extended using primers that were complementary to the insert DNA and that were designed on OLIGO 5.0 (National Biosciences Inc.). Locations of internal primers and the lengths of the sequences obtained are indicated in figure 1.

Taxonomic Distribution of Gene Rearrangements

Once complete sequences were obtained for both tick species, primers were designed that flank the excision and insertion sites of the translocated region to define the taxonomic distribution of the rearrangement in Ixodidae. The break between the $tRNA^{Ser(UCN)}$ and ND1 genes was tested with a forward primer from $tRNA^{Ser(UCN)}$ (5′-ATT AAG TAT GTA TTT TGA A-3′) and an ND1 reverse primer (5′-WGC TCA AGT AAT TTC TTA TG-3′). Presence of the 5′ end of the insertion was tested for with a forward primer from the $tRNA^{Glu}$ gene (5′-CAC WTA ACW TTT TCR TTG TT-3′) and the ND1 reverse primer. The break on the opposite end of the translocated segment between $tRNA^{Ile}$ / $tRNA^{Gln}$ and $tRNA^{Met}$ was tested with a forward primer from the $tRNA^{Ile}$ gene (5′-GAR TAA AGG RTT ATC TTG AT-3′) and with a $tRNA^{Met}$ reverse primer (5′-GGT ATG AAC CCA MTA GCT T-3′). Presence of the 3′ end of the insertion was identified with a 12S forward primer (5′-AAA GWA ATG TAA TTC ACT TCA-3′) and with a $tRNA^{Phe}$ reverse primer (5′-AAA ATA YTA CAC TGA AAA TG-3′). The deletion was confirmed with a forward primer in the cytochrome *b* gene (Cytb) (5′-AAY AAA YTK GGG GGA GTA ATT GC-3′) and with the $tRNA^{Met}$ reverse primer.

Phylogenetic Analyses

Duplicate control regions discovered in metastriate ticks were subjected to a phylogenetic analysis to determine if the regions evolve through independent or concerted evolution. Control regions were aligned with CLUSTAL W (version 1.7) (Higgins and Sharp 1989). Pairwise alignments were performed with an open-gap penalty (OGP) of 15 and a gap extension penalty (GEP) of 6.66 and with the IUB DNA weight matrix. Multiple alignments used the same OGP and GEP values and a transition weight of 0.5. Phylogenetic relationships were assessed by maximum-likelihood (ML) analysis (Hasegawa, Kishino, and Yano 1985) in PAUP* version 4.0d61 (with permission). A check for the consistency with which ML branches were supported was conducted by bootstrap analysis with 100 replications.

The gene rearrangements discovered in *R. sanguineus* did not completely explain our inability to amplify certain genes using primers considered to be conserved within arthropods. As a next step, a phylogenetic analysis of nucleotide sequences was attempted for the two ticks, *Drosophila yakuba* (GenBank accession number X03240), *D. melanogaster* (U37541), *Anopheles gambiae* (L20934), *A. quadrimaculatus* (L04272), *Apis mellifera* (L06178), *Locusta migratoria* (X80245), *Artemia franciscana* (X69067), and *Limulus polyphemus* (AF002644). *Lumbricus terrestris* (U24570) was treated as an outgroup. Nucleotides were aligned with CLUSTAL W. This approach was abandoned when visual inspection of the multiply aligned sequences revealed very few conserved regions. Instead, amino acids and codons were inferred in each of the 11 taxa using the arthropod and platyhelminth mitochondrial code, and amino acid sequences were then aligned by CLUSTAL W. Pairwise alignments were performed with an OGP of 15 and a GEP of 6.66 and with the BLOSUM30 protein weight matrix. Multiple alignments used the same gap penalties. This process identified many conserved, aligned regions in each of the 13 genes.

The aligned amino acid sequences were subjected to maximum-parsimony analysis by PAUP* version 4.0d61 and PROTPARS in the PHYLIP version 3.5c package (Felsenstein 1993). Distance analyses were done in PHYLIP version 3.5c by PROTDIST with the Dayhoff PAM matrix (Dayhoff, Schwartz, and Orcutt 1978). Distance trees were constructed by neighbor-joining (Saitou and Nei 1987). The consistency with which the data set supported resolved branches was estimated by bootstrap analysis with 100 replications.

Codon and Amino Acid Usage

Codon and amino acid usage patterns were examined using correspondence analysis (Greenacre 1984) in the computer package CodonW (www.molbiol.ox.ac.uk/cu/culong). Correspondence analysis initially estimates principal components to identify patterns of similarity in amino acid or codon use among taxa and then attempts to identify the sources of these similarities by

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

Table 1
Codon usage table and inferred genetic code (aa = amino acid) for *Ixodes hexagonus (Ih)* and *Rhipicephalus sanguineus (Rs)*

Condon	aa	*Ih*	*Rs*	Condon	aa	*Ih*	*Rs*	Condon	aa	*Ih*	*Rs*	Condon	aa	*Ih*	*Rs*
AAA	K	98	126	ACA	T	50	67	AGA	S	64	72	ATA	M	232	293[b]
AAG	K	22	10[a]	ACG	T	7	1[a]	AGG	S	15	7	ATG	M	64	25
AAC	N[a]	47	29[a]	ACC	T	29	5[c]	AGC	S	2	2	ATC	I	125	42[c]
AAT	N[a]	115	179[c]	ACT	T	54	69	AGT	S	32	30	ATT	I	311	422[c]
CAA	Q	43	51	CCA	P	33	47	CGA	R	22	32	CTA	L	70	42[b]
CAG	Q	6	5	CCG	P	5	3	CGG	R	5	1	CTG	L	7	1
CAC	H	23	6[a]	CCC	P	36	17[b]	CGC	R	7	0[b]	CTC	L	37	6[c]
CAT	H	44	59	CCT	P	63	53	CGT	R	14	11	CTT	L	67	59
GAA	E	43	69[a]	GCA	A	41	41	GGA	G	87	79	GTA	V[a]	62	70
GAG	E	36	10[c]	GCG	A	3	2	GGG	G	54	21[c]	GTG	V[a]	36	5[c]
GAC	D	21	9[a]	GCC	A	18	14	GGC	G	6	13	GTC	V[a]	11	7
GAT	D	43	49	GCT	A	54	60	GGT	G	34	50	GTT	V[a]	74	57
TAA	<	8	10	TCA	S	99	129[a]	TGA	W	78	74	TTA	L	265	355[c]
TAG	<	0	0	TCG	S	6	1[c]	TGG	W	17	6[a]	TTG	L	52	24[b]
TAC	Y	24	19	TCC	S	58	25	TGC	C	6	8	TTC	F	100	51[c]
TAT	Y	98	111	TCT	S	110	89	TGT	C	27	28	TTT	F	287	344[a]

NOTE.—Total codons in *I. hexagonus* = 3599, unconventional stops = 5. Total codons in *R. sanguineus* = 3601, unconventional stops = 3.
[a] 2 × 2 contingency χ^2, $P \leq 0.05$
[b] 2 × 2 contingency χ^2, $P \leq 0.01$
[c] 2 × 2 contingency χ^2, $P \leq 0.001$

testing for a correlation (correspondence) between the taxon values in the first two principal components and for a number of indices of amino acid and codon use. Significant correlations indicate whether similar patterns of amino acid or codon use arise due to biased codon usage, nucleotide composition of codons, or properties of the encoded amino acids.

CodonW used the multivariate analysis source code of Thioulouse et al. (1995) and Thioulouse and Chevenet (1996) to perform a principal component analysis of all 13 protein coding genes in the 11 arthropod taxa. Relative synonymous codon usage (RSCU) (Sharp and Li 1987) and amino acid frequencies were analyzed. Taxon values in the first two principal components were plotted to identify patterns of similarity in overall amino acid or codon use among taxa. For each taxon, CodonW computed 14 codon and amino acid indices including the following: Codon Adaptation Index (CAI) (Sharp and Li 1987), Frequency of Optimal Codons (F_{op}) (Ikemura 1981), Codon Bias Index (CBI) (Bennetzen and Hall 1982), Effective Number of Codons (N_C) (Wright 1990), GC content of the gene, GC content of the third position of synonymous codons (GC_{3s}), base composition of silent sites (A_{3s}, C_{3s}, G_{3s}, and T_{3s}), frequency of synonymous codons (L_{sil}), number of translatable codons (L_{aa}), and, for the inferred amino acid sequence, a hydropathicity index (Gravy) (Kyte and Doolittle 1982) and an aromaticity score (Aromo) (Lobry and Gautier 1994). The first two principal components and all 14 codon use indices were combined into a single file, and SAS (SAS Institute, Inc. 1998) was used to calculate Pearson correlation coefficients (r) between the taxon values in the first two principal components and in each of the 14 codon usage indices. The two largest significant correlations and the smallest and largest indices were then listed in the plots of principal components. Heterogeneity χ^2 analyses comparing amino acid and codon frequencies among taxa were performed by SAS (SAS Institute, Inc. 1998).

For analysis of codon usage in a diversity of metastriate and prostriate species, a ~1,600-bp portion of the COI and COII genes was amplified in selected taxa with COI–RLR and C2. A 580-bp region of the COII was sequenced with a forward primer (5′-AAA YTC WCC AAT TAT AGA ACA-3′) and C2. CodonW was used to perform a correspondence analysis of amino acid and codon usage in the COII of 12 tick species and of the nine invertebrate taxa listed above.

Results

We initially amplified and directly sequenced the 16S, 12S, COI, and COII genes from the same specimen of *I. hexagonus* in which we had performed long PCR and in individual *R. sanguineus* to determine if tick mitochondrial DNA had been accurately amplified, cloned, and sequenced. Sequences of the cloned genes were identical to those obtained through direct sequencing of amplified products (data not shown). Annotated sequences of the entire mitochondrial genome are available from GenBank (*I. hexagonus*—AF081828; *R. sanguineus*—AF081829) or are available upon request from the author (W.C.B.). The *I. hexagonus* genome contains 14,539 bp, with a nucleotide composition of 37.5% adenine, 8.7% guanine, 18.7% cytosine, and 35.1% thymine. The *R. sanguineus* genome contains 14,710 bp, with a nucleotide composition of 37.6% adenine, 9.9% guanine, 12.1% cytosine, and 40.3% thymine.

Codon usage patterns for both species are indicated and frequencies of codons and amino acids are compared in Table 1. Figure 2 shows the structure of each of the 22 tRNA genes. Gene arrangements in the two mitochondrial genomes appear in figure 3. Gene order in *I. hexagonus* follows exactly that reported for *L. poly-*

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

Amino acid	Acceptor arm		D arm	Anticodon arm		UψC arm	Acceptor arm
Alanine	[AGAAAAG]	UA	(UUUAAAAAAAUAAA)	A(CCUAAUU*UGC*AAUUAGA)	AAUU	(GAAAUAUUC)	[CUUUUCUA]
	[AGAAAAG]	UA	(UUUAAAUUAUUAUAAA)	G(UCUAAUU*UGC*AUUUAGA)	AAUU	(GGUUAACC)	[CUUUUCUG]
Arginine	[AAAAUA]	AA	(GCGAUACUAUUGC)	A(AUCAGUU*UCG*ACCUGAU)	UUUA	(GACGAAGUC)	[UAUUUUU]
	[AAAAUA]	AA	(GCGAUAAUAAUUGC)	A(UUCAGUU*UCG*GCCUGAA)	UUUA	(GAAGUUUC)	[UAUUUUU]
Asparagine	[UUAAUAG]	AA	(GCCAAAACCCCCCGAGGC)	C(AUUCACU*GUU*AAUGAAU)	AAUU	(GAGUAUCUC)	[CUAUUAAA]
	[UUAAUAG]	AA	(GCCAAAAUAGAGGC)	G(UUUCACU*GUU*AAUGAA)	UUAAAU	(GAUUUUC)	[CUAUUAAA]
Aspartic acid	[AAAAAC]	UA	(GUUAAAUAUAAAUAAC)	A(AGAGAAU*GUC*AUUCUCU)	AAUU	(ACCAUCGGU)	[GUUUUUU]
	[AAAAAC]	UA	(GUUAAUCUAUAAC)	G(AAAGAAU*GUC*AUUCUUU)	UAUA	(ACCAAUAAAGGU)	[GUUUUUU]
Cystiene	[AAUCUUA]	UA	(UUUUAUAUAAAAU)	A(UUAAAUU*GCA*AAUUUAA)	AGAU	(GAUUUUUC)	[UAAGAUUU]
	[AAUCUUA]	UU	AAA	A(UUAAACU*GCA*AAUUUAA)	AAAUU	(GAUAAAUAUUUAUC)	[UAAGAUUU]
Glutamine	[AGUUUU]	UG	(GUGUAUAUGCAC)	A(AUAAAUU*UUG*AUUUUAU)	AAGA	(AAUAAUUUUGAAUUAUU)	[AAAACUA]
	[UUAACUUU]	GG	(GUGUCAUCAC)	A(AAAAAUU*UUG*AAUUUUU)	AGAA	(AAUAAUUAAUUUUAUU)	[AAAGUUAAU]
Glutamic acid	[AUUUUUA]	UA	(GUGUAAUUUAACAC)	A(UAACAUU*UUC*GUUGUUA)	AAAU	(GAUUUUUC)	[UAAAAAUA]
	[AUUUAAA]	UG	(GUGUAAUAAACAC)	U(UAACUUU*UUC*AUUGUUA)	AAUU	(CCUUUAGG)	[UUUAAAUU]
Glycine	[UUUAAU]	UA	(GUAUAUACUUAGUAC)	A(UUUAACU*UCC*AAUUAAA)	GAGA	(UUUCUCAAAAAA)	[AUUAAAC]
	[UUUAAU]	UA	(GUAUAAAAAGUAC)	A(UUUAAUU*UCC*AAUUAAA)	UGG	(UUUCUAAAGAA)	[AUUAAAU]
Histidine	[CUUAAU]	UA	(GUUUUAAUUUUUUAAAAU)	U(AUAAAUU*GUG*GAUUUAU)	AGAU	(GAUAUUUGUUC)	[AUUAAGA]
	[GUUUAAU]	UA	(GUUUAAAUUAAAAU)	U(AUAAAUU*GUG*GAUUUAU)	AGAU	(GAACGAUC)	[AUUAAACA]
Isoleucine	[AUAAAA]	UG	(CCUGAGUAAAGG)	A(UUAUCUU*GAU*AGGAUAA)	AUUAU	(GUAAGUAACCUUAC)	[UUUUAUU]
	[GUAAAA]	UG	(CCUGAAUAAAGG)	G(UUAUCUU*GAU*AGGAUAA)	AUAAU	(GUAAUUUAUAC)	[UUUUACU]
Leucine(CUN)	[ACUAAUU]	UG	(GCAGAAAAAUUGU)	G(UCAAAUU*UAG*AAUUUGA)	GUAU	(GGAUACCCC)	[AAUUAGUA]
	[UUCAGCU]	UG	(GCAGAUGAAUGC)	A(AUAAAUU*UAG*AAUUUAU)	UUAU	(GAGAUAAACUC)	[AGUUGAUA]
Leucine(UUR)	[UUAAAG]	UG	(GCAGAAUGGAGUAAUGC)	G(AGGAAUU*UAA*GCUUCCU)	UUAU	(GAAUAGUUUC)	[CUUUAAU]
	[UUAAAG]	UG	(GCAGAGAACUAAUGC)	A(AGGAAUU*UAA*GCUUCCU)	UUAC	(GAAAAUUUC)	[CUUUAAU]
Lysine	[CAUUGAA]	UG	(GCUGAAAUAUUUAAGC)	A(AUGGUCU*CUU*AAACCAC)	UUCAU	(AGUGAUUAUCACU)	[UUCAAUGG]
	[CAUUGAA]	UG	(GCUGAAAAAAGU)	G(AUGGUCU*CUU*AAACCAA)	UUUAU	(AGUUUAAUUUAACU)	[UUCAAUGA]
Methionine	[AUAAAG]	UC	(AGCUAAUCUUAAGCU)	A(GUGGGUU*CAU*ACCCCAA)	AAAU	(GAAAUAUAUUUUC)	[CUUUAUU]
	[GUAAAG]	UA	(AGCUAAAAUAAAGCU)	A(UUGGGUU*CAU*ACCCCAA)	AUAU	(GAUUAUAAAAUC)	[CUUUAUU]
Phenylalanine	[GUUCUUA]	UA	(GUUUAUAUAAAAAU)	A(UUACACU*GAA*AAUGUAA)	AGA	(AAUUAUU)	[UAAGAAUA]
	[GUUUUUA]	UA	(GUUUAAGUAAAAU)	A(CUACACU*GAA*AAUGUAG)	AGAU	(AUAUGU)	[UAAAAACA]
Proline	[CAGAAAA]	UA	(GUAUAUAGUAAGUAC)	A(UAAAUUU*UGG*AUAUUUA)	AAGA	(GAAUUUC)	[UUUUCUGA]
	[CAGAAAA]	UA	(AUAUAGAAAUUAGUAU)	A(UAAAUUU*UGG*GGAUUUA)	UUGG	(GAUAAUAUC)	[UUUUCUGA]
Serine(UCN)	[GACUUUU]	UA	(ACUAUGCAUUAAGU)	A(UGUAUUU*UGA*AAAUACA)	AAAA	(AGAACAUUUAUUCU)	[AAAAGUCU]
	[GACUUUU]	UA	(ACUAUAUAAUUAAGU)	A(UGUAUUU*UGA*AAAUAUA)	AAAA	(AGAAUUUUUCU)	[AAAAGUCU]
Serine(AGY)	[AAGAUUA]		ACCUU	.(AUAGACU*UCU*AAUCUAU)	UAUU	(AAUGAAAUUCAUU)	[UAAUCUUU]
	[AAGAUUA]		AUCAU	.(UUAGGCU*UCU*AACCUAA)	AAUU	(AAUGAUAUUUCAUU)	[UAAUCUUU]
Threonine	[GUUAUAA]	UA	(AUUUAAAACAAAAAAU)	A(UUGGUUU*UGU*AAACCAA)	ACUU	(GAACUAAUC)	[UUAUAACU]
	[GUUGAAA]	UA	(AUUUAAUAAAAAU)	G(UUGAUUU*UGU*AAAUCAA)	UUUU	(GACUAGGUC)	[UUUCAAUU]
Tryptophan	[AGAUUU]	UA	(AGUUAUAUAAAACU)	A(UAUGCCU*UCA*AAGUAUA)	AAAU	(AAUCAAAAUUGAUU)	[AAAUCUU]
	[AGAUUU]	UA	(AGUUAAUAAAAAACU)	A(UUUACCU*UCA*AAGUAAA)	AAUU	(AUUGUAAAU)	[AAAUCUU]
Tyrosine	[GUAAAA]	UG	(GCUGAAGUUUUAAGC)	U(AUAAAUU*GUA*AAUUUAU)	UAAU	(GAUGAUAUUC)	[UUUUACU]
	[GUAAAA]	UG	(GCUGAACUUUAGGC)	G(AUAAACU*GUA*AAUUUAU)	UUAU	(GAAUAAUUUC)	[UUUUACU]
Valine	[AAAAUG]	AA	(AUCAUUAGAU)	A(UUUCAUU*UAC*AAUGAAA)	GUUU	(GUUUGUAAGAAC)	[CAUUUUA]
	[AAAAUG]	AA	(AUCAUUAGAU)	G(UUUCACU*UAC*AAUGAAA)	AUUU	(GUAUAAUUUAC)	[CAUUUUU]

FIG. 2.—Structure of the 22 transfer RNAs in *Ixodes hexagonus* (upper) and *Rhipicephalus sanguineus* (lower). Parentheses delimit a region involved in base pairing (known as ''stem'' formation) where one half of a stem is adjacent to the complementary half. Underlined nucleotides form base pairs with adjacent underlined nucleotides. Square brackets delimit the complementary halves of the acceptor arms, and underlined nucleotides form base pairs. Anticodon base pairs appear as italicized letters.

phemus (Staton, Daehler, and Brown 1997). However, in *R. sanguineus,* the block of genes encoding ND1, $tRNA^{Leu(UUR)}$, 16S rDNA, $tRNA^{Val}$, 12S rDNA, the control region, and the genes encoding $tRNA^{Ile}$ and $tRNA^{Gln}$ have translocated to a position between the $tRNA^{Glu}$ and $tRNA^{Phe}$ genes. However, possibly prior to this translocation, the gene ending $tRNA^{Leu(CUN)}$ translocated between the $tRNA^{Ser(UCN)}$ gene and the control region. In addition, the $tRNA^{Cys}$ gene inverted and translocated between the control region and the $tRNA^{Met}$ gene. Figure 3 also indicates that the control region (CR1) was retained during the translocation but that a second copy of the control region (CR2) was retained 5′ to the $tRNA^{Cys}$ and $tRNA^{Met}$ genes. CR1 and CR2 are nearly identical (figs. 4 and 5).

Rhipicephalus sanguineus is a member of the subfamily Rhipicephalinae, which, on the basis of relative substitution rates, Black, Klompen, and Keirans (1997) con-

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

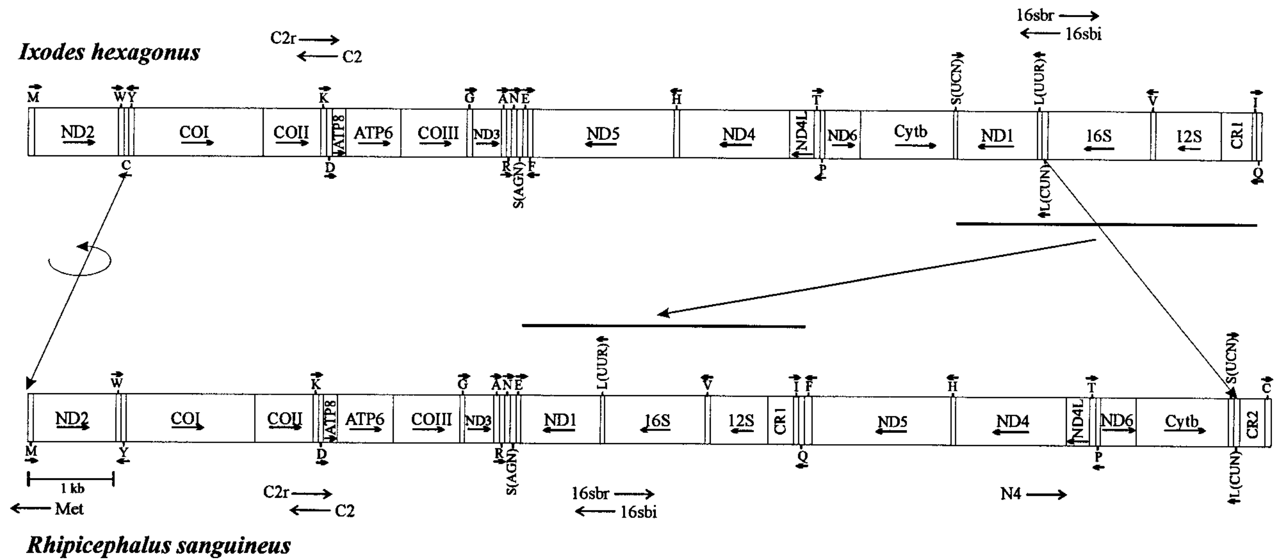


FIG. 3.—Gene arrangements in *Ixodes hexagonus* and *Rhipicephalus sanguineus.*. The insertion location of the translocated segment into the *R. sanguineus* mtDNA is shown, as is its original position in the *I. hexagonus* mtDNA. Origin and insertion of the transposed tRNA$^{Leu(CUN)}$ gene and of the inverted and transposed tRNA^{Cys} gene are also indicated. Abbreviations are as follows: ATPase8 = ATP8, ATPase6 = ATP6, Cytochrome oxidase I = COI, Cytochrome oxidase II = COII, Cytochrome oxidase III = COIII, Cytochrome *b* = Cytb, NADH dehydrogenase subunit 1 = ND1, NADH dehydrogenase subunit 2 = ND2, NADH dehydrogenase subunit 3 = ND3, NADH dehydrogenase subunit 4 = ND4, NADH dehydrogenase subunit 4L = ND4L, NADH dehydrogenase subunit 5 = ND5, NADH dehydrogenase subunit 6 = ND6, large ribosomal subunit = 16S, small ribosomal subunit = 12S. tRNAs are denoted by the one-letter amino acid abbreviation.

sidered to be the most recently derived metastriate subfamily. Primers were designed to determine the taxonomic distribution of the *R. sanguineus* gene rearrangements and of CR2 throughout the Ixodidae. The presence of the gene rearrangements was tested by standard PCR, with primers adjacent to break/insertion points. Prostriate ticks were selected from a diversity of *Ixodes* genera and included the following: *I.* (*Pholeoixodes*) *cookei* (Chesapeake Bay, Md., Rocky Mountain Lab [RML] number 120814), *I.* (*Ixodes*) *affinis* (Georgia Southern University lab colony), *I.* (*Sternalixodes*) *holocyclus* (RML 37729), *I.* (*Endopalpiger*) *tasmani* (RML 121322), and *I.* (*Ceratixodes*) *uriae* (RML 110572). *I.* (*Pholeoixodes*) *hexagonus* (IVEM lab colony) was used as a positive control. Representatives of the four metastriate subfamilies were also examined and included, among Amblyomminae: *Aponomma varanensis* (RML 121557), *A. glebopalma* (RML 121854), *A. latum* (RML 121580), *Amblyomma variegatum* (IVEM lab colony), *A. americanum* (OSU lab colony), *A. maculatum* (OSU lab colony), and *A. cajennense* (OSU lab colony). Included among Haemaphysalinae were *Haemaphysalis leporispalustris* (Yuma Co., Ariz.) and *H. inermis* (RML 122165); among Hyalomminae, *Hyalomma dromedarii* (IVEM lab colony) was included. And, among Rhipicephalinae, the following were included: *Dermacentor andersoni* (Larimer Co., Col.), *D. variabilis* (OSU lab colony), *Boophilus annulatus* (lab colony, USDA, Kerrville, Tex.), and *Rhipicephalus appendiculatus* (IVEM lab colony); *R. sanguineus* (lab colony, Hebrew University, Israel) served as a positive control. In all prostriate species examined, the tRNA$^{Ser(UCN)}$ forward primer and the ND1 reverse primer produced a ~580-bp fragment, and the tRNA^{Ile} forward and tRNA^{Met} reverse primer produced a ~160-bp fragment. Under the standard PCR conditions used, no products were amplified with these two primer pairs in any of the metastriate species. In all metastriate species examined, the forward tRNA^{Glu} primer and the ND1 reverse primer amplified a ~570-bp fragment, and the 12S forward primer and the tRNA^{Phe} reverse primer amplified a ~1,000-bp fragment. No products were amplified in any of the prostriate species. Presence of the deletion was tested for with the forward Cytb primer and with the tRNA^{Met} reverse primer. They amplified a ~780-bp fragment in all of the metastriate and none of the prostriate species. This analysis indicated that the gene rearrangements and, probably, the duplicated CR2 are distributed throughout and are exclusive to metastriate ticks.

To test for the presence of CR2 and to assess the similarity of the duplicate regions, products amplified from the 12S/tRNA^{Phe} primers and the Cytb/tRNA^{Met} primers were sequenced (fig. 4) from an individual *R. sanguineus,* from the congeneric species *R. appendiculatus,* and from *Aponomma varanensis,* a member of the basal metastriate subfamily Amblyomminae. CR2 was found in all amplified products and is very similar both within and among the three species. Furthermore, metastriate CR1 and CR2 aligned with the prostriate CR1 (fig. 5).

A phylogenetic analysis was performed on the aligned regions to determine if the duplications evolve independently or in a concerted fashion. If the genes evolve independently, then all CR1 sequences should fall within one phylogenetic group separate from the CR2 sequences. Alternatively, if the genes coevolve, CR1 and CR2 should cluster together within a species. Maximum-likelihood analysis (fig. 6) indicated that the

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

A)

```
Rs(C) ATTGTTTTTATAGTGGTATACAAATTGAATTTACAAATTTAAGTAAGATT-AAGTGTTTTATCCATTAAAGAACAAATTCCTCTGAAAAGCTTA
Rs(i) ..................................................T.G...................G.....................
Ra    ........A..............G....T..G..................T.G...................G.....................
Av    ...A....A...................TA.A..T.....T......G..T.G...................G.............T...T...
      12S[8300]→

AAATACCGCCATAATTTTTTGCTTTCGTAATTTTTATTTACTAACAATATTTACCTCTTAAATAATAGGGTATCTAATCCTAGTTTAATTTTAGAATTTC
.................................C..............G...................................................
................................A....A....T..........G.................................T.A.A......AT
..G..................T......T...-..................A.G....AT.............................AA.GA.T...T

AATTATATTTTTATTATCTTAAAAAAAAAATTTTACCTTGATTTTTTAATAAGTTT--TTATAATTATTCGTTAAATTTCTTAAAGAAAGAATTTTTCTA
................CT...............................................AG..T...............A..............
....T......A....AT..T.....T......C.....A......-..C..AA.......CC.CAT..TT.A..T.C.A...G.AT.......G.....
.T.ATA...A..T.CTAT...TTT.........C.....TT..AAA..TA..C...AT...A..AAT..AAA...T...A..TT.A...T..AAAAAT.G

TTTGTTTAACCGCTGCTGCTGGCACAAATTTAGCTAGATTATTTTTCTAATTCTTAATAATTTTTTATTATTACATAAAATAAATTTTCTACTTTAATTT
.........................................................................A..........................
.........................................................................A..................AAA.....
.....C......................C..T....-....CC..A-...A.T.C......A...........AT....T.T........GTC..TTC..

-TTTTTTAAAAAATAGTATAAAAAAATTAGAAGGATTTT-CTTAAAAACCAAACCAAATTCCAGAA-ATGATTTAATTAAAAATTTAATTGCGGTTATTT
A......T...........................C.................T.......T..............................A.......
.---..AT......T...........C..AG.A......T.............T......AA.TTTT..T.AAC.TA.T....A........C.CA.AA.
.---...T.....A.AC........T...TTT-.T....A..ATTT.......AT.....AGTCGG.--TT...TC......T.......TTAAA.....

TAAA-------TTTTGTTAAAACTCGTGTCTATCGGTTATCTCGATATATAAAAGGAAACGTATGCCAGACTTTACTGGGCAAATCTCCCTTATT-TGAA
A..............T.........A..........................................................................
...T...........T.C.......A........................................T........A..............A....C....
....AAACTTT....T...C....T..............A..G..C.G.......-...T..........AC...-.AC.TC...T...TCAC.AA.TT.
           [----------------------------------------------------------------------------------------
           CR1→

AATATGTCCTATAATTTGAGCAAAATGTAACCATCTCCCCTTTTT-CTCCGAATTTATTAATTAGTAGATGTGTACATGACTTAGTATGAACCTTTGTTA
.............................................T..............G.............................C.........
....G.A................G....G.A.....TTTT....................G.............................C.........
...C...TT..--..AA....T.....CCTT.....TTTT....C.-.............T.A.T.........GTCA............C.........
----------------------------------------------------------------------------------------------------

CTCTCCTATGGGTCAATAGATGAGACAGCCGGTCGTCGCCCCTTATTTACAATAGATGAGATAATATTCCGGCATAGTAAAATGCCTGAATA---AAGGG
....................................................................................................
..................A...................A.........CTA.......TA..T...........C.........................
.ATCTA.GCA.TC..G..A...C..T..A...AG...............A........TA...C.....A.....................A.TGA...A
---------------------------------------------------------------------------][tRNA(Ile)→-----------

TTATCTTGATAGGATAAATAATGTAA---TTTATACTTTTACTA-----------TATTAACTTTAATAAAATTAA-TTATTTTCTAAAAATTCAAAATT
....................................................................................................
.........................-......C...........GGAGCTCTAGAA............................................
..................A.......ATT.A.T...........CATT......AA..............T.....A......CTA.....A........
-------***--------------------------------]            [-------------------------------------***----

TTTTGTG-----ATGACACCCAAAGTTAATTGCAT
............T......................
..............A.............T......
.......CTTTA..T..G..A.....A.T.AT.T.
-----------------←tRNA(Gln)]     [9,190]
```

FIG. 4.—Alignment of the control regions downstream from the (*A*)12S rRNA (CR1) and (*B*) Cytochrome *b* genes (CR2). Sequences are from the *Rhipicephalus sanguineus*–cloned DNA (Rs(C)) and are amplified from an individual tick (Rs(i)). Ra denotes *Rhipicephalus appendiculatus,* and Av denotes *Aponomma varanensis.* A dot indicates an identical nucleotide to the one that appears on the top line. Gaps inserted to improve the alignment are indicated with a dash (-). The boundaries of the 12S, CR1, $tRNA^{Ile}$, and $tRNA^{Gln}$ genes are indicated in (*A*), and the boundaries of the Cytb, $tRNA^{Ser(UCN)}$, $tRNA^{Leu(CUN)}$, CR2, $tRNA^{Cys}$, and $tRNA^{Met}$ genes appear in (*B*).

regions coevolve within species. However, note that the CR2 sequences cluster together in *R. sanguineus* and share specific substitutions in common (fig. 5). This suggests that CR1 and CR2 may evolve independently to a limited extent within species but that some mechanism maintains sequence similarity between the two copies over the long-term course of evolution.

Inferred amino acid sequences were multiply aligned among all 10 arthropod taxa and *Lumbricus terrestris* for each of the 13 genes. Alignments were then assembled into a single file for phylogenetic analysis. There were 3,829 characters in the aligned dataset, and 2,027 of these were parsimony informative. Maximum-parsimony analysis was performed with PAUP* using the

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

B)

```
Rs(C) TCAATTATTATTATTTTAAGTTTTTGCTTTACAATAAATAATAAAATATCATCTTTTTATTTTAATATTTCATTCAAATTAATATTTTGAATTT
Rs(i) ..............................................................................................
Ra    ......GG.TA.CA...........CT.......................C..A................T..T..A.T...............
Av    ......T.A.........TT.A.CCCA...TTTC..GC......T....A.GC.........G..TTAAA.....T.AT......T...
      S  I  I  I  I  L  S  F  C  F  T  M  N  N  K  M  S  S  F  Y  F  N  I  S  F  K  L  M  F  W  I  L
            G  Y  H     F  I  S     I  S  K  H     I        C           S  L  K     I  N        F
            L                 P                                            F        F
      Cytochrome b[14,000]→

TAGTAAATTGTTTTTTTATACTCACATATCTTGGAGCAATGCCTATTGAATATCCTTTTGATTTAATAAGTAAAATTACTACAATTATTTATTTTTTAAT
.......................T.................A..C.......C.........C......................................
..A.C........A.C...T.A..C.......G..T..A...............A...AACT....A......GT....G..G....C...A....
....T...AT....A.A..TT.A..T...T.A..T..T..A.........T.....A....A.TT....A...T.A.TA..TTC.GC............
 V  N  C  F  F  M  L  T  Y  L  G  A  M  P  I  E  Y  P  F  D  L  M  S  K  I  T  T  I  I  Y  F  L  M
 I     I     I  I                                F     Y  E  S  L        L  V     V  V        M
             M                                               T              M     S  A
                                                             I

ATTTTTTATAATTCCTTTATGTTAAGACTTTTTAACTATATAATTAAGTATGTATTTTGAAAATATAAAAAAGA--ATTTTTCTAAAAGTCTATTATCAA
....................................................A............................................C..G
....A..T.......A...AT..................-G.........A.......................A...............A.T...G
.......C..T.AT.....CA...-..T..........----.......A.....................GA....C......A....A.T....
 F  F  M  I  P  L  C  <  [tRNA(Ser(UCN))==>--------------***------------------------------]   [-----
    I  L  L  S     I
       L           H

CTGAGTTT----ATCTCATAAATAAATTCTAAATTTATTGCATTCATCTGCCAAGCTGAATTTAAAT-TTTTTATTTAATAAAGGGGTTTTTAAAATTT
..........................................T................................G.....................
....AG.-......T......G...............C......T.................AA.GC.A......T......TGCC.CAAAA....--...
T...A...TTTT.GT.......................CA....TT....T...AT....AAA...C....----------------------------..
----------------------------***-----------<==tRNA(Leu(CUN))]Control region→                     [-
                                                                                                 CR2→
TTGTTAAAACTCATGTCTATCGGTTATCTCGATATATAAAAGGAAACGTATGCCAGACTTTACTGGGCAAATCTCCCTTATT-TGAAAATATGTCCTATA
..-.................................................................................................
..T.C.....................................T........A..............A....C........G.A.....
..T...C....TG.............A..G..C.G.......-...T..........AC...TAC.T...--T.T...C.C.A.TT....C...TT...-
----------------------------------------------------------------------------------------------------

ATTTGAGCAAAATGTAACCATCTCCCCTTTTT-CTCCGAATTTATTAATTAGTAGATGTGTACATGACTTAGTATGACCCTTTGTTACTCTCCTATGGGT
................................T...............G...................................................
..........G....G.A.....TTTT....................G....................................................
-..A....T.....CCTT.....TTTT....................T.A.T.........GTCA.......................ATCTA.GCA.TC
----------------------------------------------------------------------------------------------------

CAATAGATGAGACAGCCGGTCGTCGCCCCTTATTTACAATAGATGAGATAATACATTGCATGCATA------AATCTTATTA---AAATTAAACTGCAAA
....................................................................................................
.....A...................A........CTA........TA..T...TTCC.....A.......................-.............
..G..A...C..T..A...AG...............A........TA...C..TTAC.GCGTA...ATTCCA.......CA.CAA...............
-------------------------------------------------------------------]        [tRNA(Cys)→-----------***--

TTTAAAAAT--TGATAAATATTTATCTAAGATTTTAAGATTTT-AAGTAAAGTAAGCTAAA
.............................................................
.......T.GA.A..T..A........................G..A..............
...........GA..T...T.C.---..................TT.A......C...C...
-----------------------------------------]    [tRNA(Met)[1]→
```

FIG. 4 *(Continued)*

heuristic search option with Tree-Bisection-Reconnection branch swapping (fig. 7). A single tree of 9,843 steps was found, with a consistency index of 0.843 and a retention index of 0.611. Identical phylogenies were derived, whether using maximum-parsimony or distance analysis, and most branches had ≥99% bootstrap support. Temporarily ignoring the placement of *Apis mellifera* (honeybee), there was strong support for monophyly of the Insecta. Crustacea and Insecta formed a monophyletic group, as did the three chelicerate taxa. But both clades had weak bootstrap support. Figure 7 illustrates an interesting example of convergence in amino acid sequences. *Apis mellifera* belongs to the suborder Apocrita of the order Hymenoptera, considered one of the most recently evolved and derived groups of Insecta and yet, based upon mitochondrial amino acid sequences, falls within a well-supported clade with hard ticks.

Correspondence analysis was used to determine if this convergence was associated with similar amino acid or codon usage. A plot of principal components indicates that *A. mellifera* is more similar in amino acid usage to *R. sanguineus, I. hexagonus,* and *D. melanogaster* than to any of the other arthropods examined (figure 8*A*). Correspondence analysis of the first principal component indicates that this similarity was associated with a

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

```
Rs(C)-CR1   TTTTGTTAAAACTCGTGTCTATCGGTTATCTCGATATATAAAAGGAAACGTATGCCAGA---CTTTACTGGGCAAATCTCCCTTATT-
Rs(C)-CR2   ..............A........................................................................
Rs(i)-CR1   ....T.........A........................................................................
Rs(i)-CR2   ....-.........A........................................................................
Ra-CR1      ....T.C.......A.......................................T...........A..............A....C
Ra-CR2      ....T.C.......A.......................................T...........A..............A....C
Av-CR1      ....T...C....T..............A..G..C.G......-...T..............AC..TAC.T...T.TC.T.ACTAA.
Av-CR2      ....T...C....T..............A..G..C.G......-...T..............AC..TAC.T...T.TC.T.ACTA-.
Ih          CAAAA..GT..G..A.TC....GC..CTA..TCG....GT..TTG..TA..ATTA.T.TAGA...C.CT.AT..GGGAA.AA.GC-.

TGAAAATATGTCCTATAATTTGAGCAAAATGTAACCATCTCCCCTTTTT-CTCCGAATTTATTAATTAGTAGATGTGTACATGACTTAGTATGAACCTTT
..............................................................................................C.....
.................................................T..............G.............................C.....
.................................................T..............G.............................C.....
........G.A................G....G.A.....TTTT....................G.............................C.....
........G.A................G....G.A.....TTTT....................G.............................C.....
.TT....C...TT...--..AA...T.....CCTT.....TTTT....-...............T.A.T.........GTCA............C.....
.TT....C...TT...--..A....T.....CCTT.....TTTT....................T.A.T.........GTCA............C.....
.CT.G..T.TAT...A.-...ATTTG..GCT..CTTG..GT.T..A.C.....T.GGAA...ATTCA.A..T...AAAC....T....A-..T.TAA...

GTTACTCTCCTATGGGTCAATAGATGAGACAGCCGGTCGTCGCCCCTTAT-TTACAATAGATGAGATAATATTCCGGCATA--
.......................................................................CATT.-...GCA
...................................................................................
.......................................................................CATT.-...GCA
......................A...................A.........CTA........TA..T...........C...
......................A...................A.........CTA........TA..T........CATG.A.
.....ATCTA.GCA.TC..G..A...C..T..A...AG................A........TA...C.....A........
.....ATCTA.GCA.TC..G..A...C..T..A...AG................A........TA...C....A....G..A.
C.C.G...-A.T.T.ACA....A...TA.TT.GA...G.AT..T.....CA..T.T...A...TT.C.T....TTTTTT.TTT
```

FIG. 5.—Alignment of control regions from the *Rhipicephalus sanguineus*–cloned DNA (Rs(C)) and amplified from an individual tick (Rs(i)). Ra denotes *Rhipicephalus appendiculatus,* Av denotes *Aponomma varanensis,* and Ih denotes *Ixodes hexagonus.* A dot indicates an identical nucleotide to the one that appears on the top line. Gaps inserted to improve the alignment are indicated with a dash (-).

negative CBI, chiefly in the excess use of Asn (AAY), Ile (ATY), and Met (ATR), and with low GC content of the codons in these four taxa. The first principal component was also significantly ($|r| > 0.7$, $P = 0.05$) correlated with A_{3s}, C_{3s}, G_{3s}, T_3s, F_{op}, N_C, GC_{3s}, Gravy, and Aromo. The second principal component was only significantly correlated with Aromo, indicating that it separates taxa based upon aromaticity (frequency of Phe, Tyr, and Trp) of the encoded amino acids. Heterogeneity χ^2 analyses showed that *A. mellifera* differed from *R. sanguineus* in having a lower frequency of Ala (GCN) and a higher frequency of Leu (CUN, UUR) and Tyr (TAY). Frequencies of the other 17 amino acids did not differ significantly between the two species. *Apis mel-*

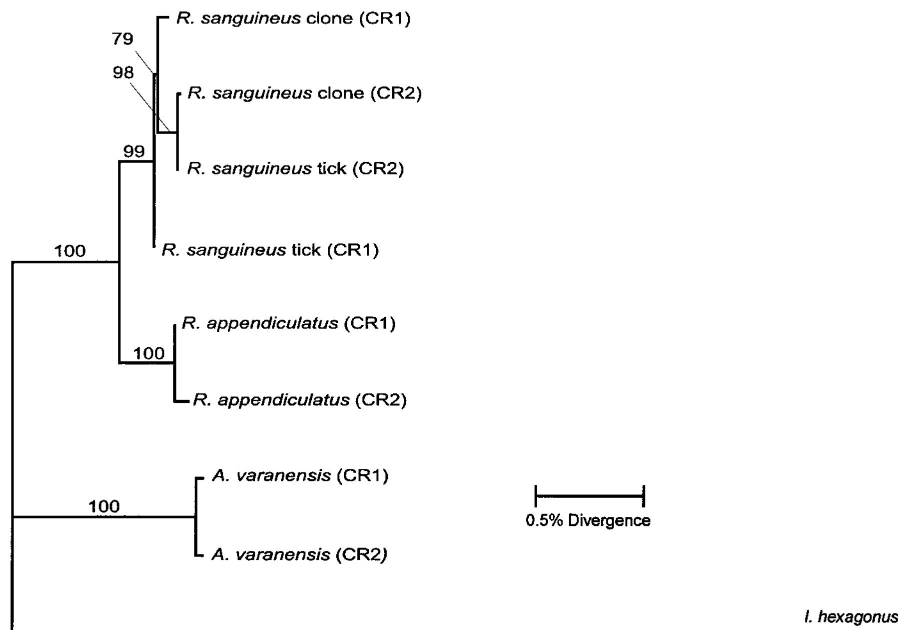


FIG. 6.—Phylogenetic relationships among aligned control regions (fig. 5) determined by maximum-likelihood analysis. The percentages of 100 bootstrap replications that support each branch are indicated.

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

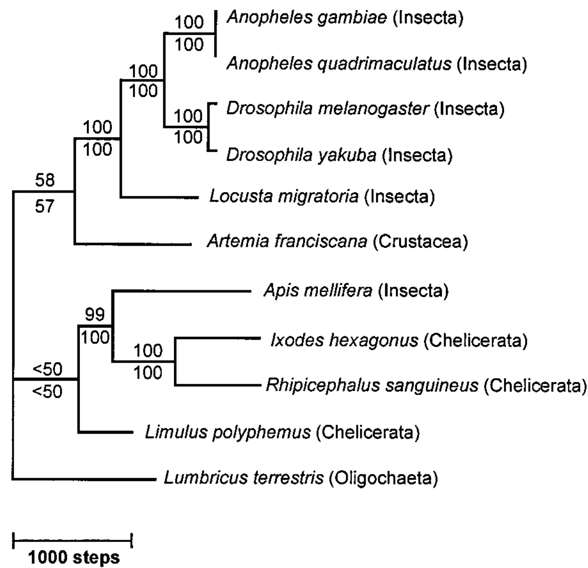
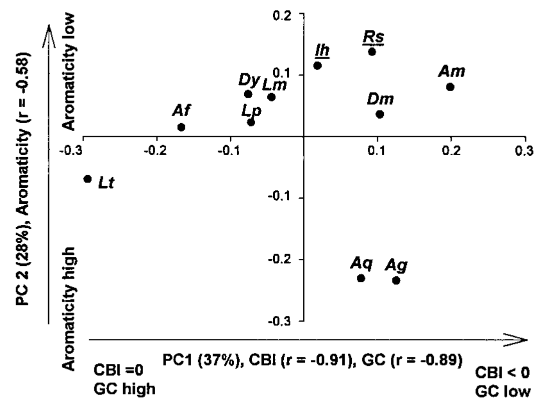


FIG. 7.—Phylogenetic relationships among 11 taxa for which all or most of the mtDNA has been sequenced. The number above each branch is the percent of 100 bootstrap replications that supported that branch when the data set was analyzed by PROTDIST in PHYLIP version 3.5c (Felsenstein 1993). The number below each branch is the bootstrap support for the branch when the data set was analyzed by PROTPARS in PHYLIP version 3.5c.



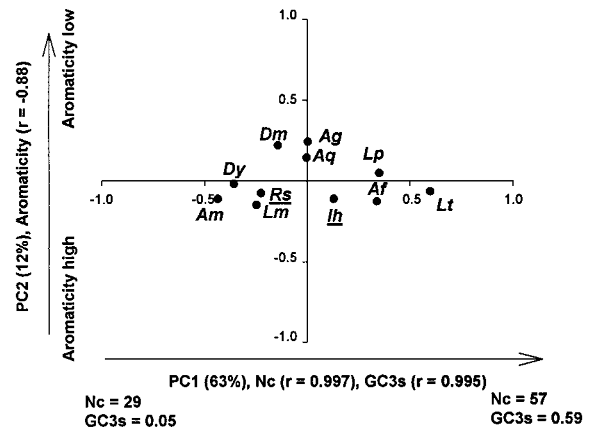


FIG. 8.—Correspondence analysis among amino acids and codons in all 13 mtDNA protein-coding genes in 11 taxa in figure 7. (*A*)Plot of first and second principal components (PC) derived from analysis of amino acid frequencies in the following: Af = *Artemia franciscana,* Ag = *Anopheles gambiae,* Aq = *A. quadrimaculatus,* Am = *Apis mellifera,* Dy = *Drosophila yakuba,* Dm = *D. melanogaster,* Ih = *Ixodes hexagonus,* Lm = *Locusta migratoria,* Lp = *Limulus polyphemus,* Lt = *Lumbricus terrestris* and Rs = *Rhipicephalus sanguineus.* (*B*) Plot of first and second PC derived from analysis of RSCU indices. The percentages of the total variance in amino acid frequencies or in RSCU accounted for by the first and second PC are indicated. The two largest significant correlations and the smallest and largest indices are listed alongside each PC.

lifera differed from *I. hexagonus* in the frequency of eight amino acids and from *D. melanogaster* in the frequency of ten amino acids. In all cases, *A. mellifera* had a significantly higher frequency of amino acids encoded by AT-rich codons and a lower frequency of amino acids encoded by GC-rich codons.

Correspondence analysis of RSCU (fig. 8*B*) indicates that codon use is similar among *A. mellifera, D. yakuba, L. migratoria,* and *R. sanguineus.* The first principal component accounted for the majority (63%) of the variance in RSCU among all 11 taxa and was most strongly associated with a low N_C and with a low GC_{3s}. However, all indices except Aromo, $L_{sil,}$ and L_{aa} were strongly correlated ($|r| > 0.9$, $P \leq 0.0001$) with the first principal component. As with the amino acid analysis, the second principal component was only significantly correlated with Aromo. In general, the *A. mellifera, D. yakuba, L. migratoria,* and *R. sanguineus* mtDNAs use a lower diversity of codons and exhibit biases towards codons with a higher AT content. Heterogeneity χ^2 analyses showed that *A. mellifera* differed from *R. sanguineus* in the frequency of 18 codons and from *I. hexagonus* in the frequency of 41 codons. *Ixodes hexagonus* differed from *R. sanguineus* in the frequency of 27 codons (table 1), and, in each case, *R. sanguineus* used a higher frequency of AT-rich codons and a lower frequency of GC-rich codons than did *I. hexagonus.* Crozier and Crozier (1993) reported that *A. mellifera* has an unusually high A-T content and a very high frequency of TTT (Phe), TTA (Leu), ATT(Ile), ATA(Met), AAA(Lys), AAT(Asn), GAT(Asp), and GAA(Glu) codons.

We sequenced a 580-bp region of the COII gene from seven prostriate and five metastriate ticks to determine the taxonomic range of the similarity in codon usage with honeybee and *R. sanguineus.* Phylogenetic analysis of the amino acid sequences (not shown) indicated strong support for the monophyly of the metastriate and prostriate taxa. Principal component analysis of amino acid use indicates that *A. mellifera* is more similar to ticks than to any of the other arthropods examined (fig. 9*A*). The first principal component accounted for 51% of the variance in amino acid use among all 20 taxa and was only significantly correlated with a high aromaticity of the encoded amino acids. This was chiefly associated with a higher frequency of Phe (UUY) in the COII gene of ticks and of *A. mellifera.* The second principal component was significantly ($P \leq 0.05$) correlated with $N_{C,}$ $G_{3s,}$ and Gravy and was highly correlated ($P \leq 0.0001$) with $L_{aa,}$ but this is entirely due to partial sequence availability for *L. polyphemus.*

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

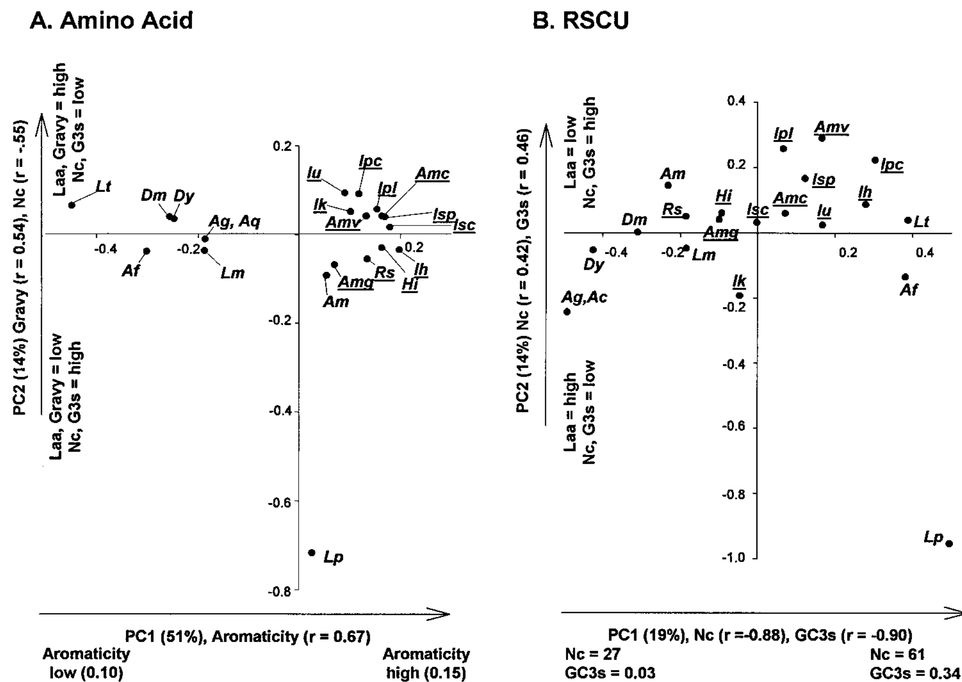


FIG. 9.—Correspondence analysis among amino acids and codons in 580 bp of the COII gene in the 11 taxa in figure 7, in addition to seven prostriate and five metastriate tick species. The origins of Iu = *Ixodes uriae,* Amv = *Amblyomma variegatum,* Amc = *A. cajennense,* and Hi = *Haemaphysalis inermis* are listed in the text. Other species analyzed are as follows: Amg = *A. glauerti* (RML 121854,) Ipl = *I. pilosus* (lab colony, Onderstepoort Veterinary Clinic, South Africa), Ipc = *I. pacificus* (CDC lab colony, Yuma, Ariz.), Isc = *I. scapularis* (RML 121468), Isp = *I. spinipalpis* (RML 122441), and Ik = *I. kopsteini* (RML 65388). (*A*) Plot of first and second PC derived from analysis of amino acid frequencies. (*B*) Plot of first and second PC derived from analysis of RSCU indices. The percentages of total variance in amino acid frequencies or RSCU accounted for by the first and second PC are indicated. The two largest significant correlations and the smallest and largest indices are listed alongside each PC.

The first and second principal components account for a small proportion of the total variance in RSCU in COII among all 21 taxa (fig. 9*B*), and no specific clusters are evident. *Rhipicephalus sanguineus* is most similar to *A. mellifera, H. inermis, A. glauerti,* and *D. melanogaster.* As in RSCU analysis of all genes (fig. 8*B*), the first principal component was strongly associated with a low N_C and a low GC in codons and was strongly correlated ($P \leq 0.0001$) with all indices except Gravy and Aromo. Note, however, that this trend does not extend to all metastriates: *A. maculatum* and *A. tuberculatum* were more similar to prostriate ticks, *L. terrestris,* and *Artemia* in having a higher N_C and a lower frequency of AT-rich codons. The second principal component was only significantly correlated with L_{aa}, an artifact of the partial *L. polyphemus* sequence.

Discussion

Prostriate ticks are basal members of Ixodidae and possess the basic chelicerate gene order first reported for *L. polyphemus* (Staton, Daehler, and Brown 1997). In contrast, a major gene rearrangement and a duplication of the control region occurred in the mitochondrial genome of the ancestor of metastriate ticks. Evidence suggesting major mtDNA gene rearrangements has been detected in a centipede, a sowbug, while another arachnid (harvestman, Opiliones spp.) appears to be more like prostriate ticks and *L. polyphemus* (unpublished data). These observations argue for careful and thorough taxon sampling when using mtDNA gene arrangements to deduce deep-level phylogenetic relationships among taxa.

In general, few insertions are found between genes in animal mitochondrial genomes, and the gene content of metazoan mtDNAs is remarkably constant. There are some examples of duplicated sequences in animal mtDNA. These involve partially duplicated rRNA, tRNA, or protein genes (Moritz and Brown 1987; Moritz 1991; Azevedo and Hyman 1993; Stanton et al. 1994). *Mytilus* mtDNAs contain an extra $tRNA^{Met(AUA)}$ and a $tRNA^{Ser(UCN)}$ pseudogene (Hoffmann, Boore, and Brown 1992). Octocoral mitochondrial genomes contain a gene that is very similar to a bacterial *mutS* gene (Pont-Kingdon et al. 1995), and sea anemone COI and ND5 genes contain group I introns and only two tRNA genes (Beagley, Okada, and Wolstenholme 1996; Beagley, Okimoto, and Wolstenholme 1998). Noncoding repeated sequences in the control region account for the larger size (up to 42 kb) of some metazoan mtDNA (Okimoto et al. 1991; Wolstenholme 1992; Casane et al. 1994; Zhang and Hewitt 1997). Tandem duplications of the control region have been identified in gekkos (Moritz and Brown 1987; Zevering et al. 1991) and in *Cnemidophorus* lizards (Moritz and Brown 1986). Kumazawa et al. (1996) reported a separate and duplicated control region present in a diversity of snake species and, as in the present study, found evidence for concerted evolution between the copies.

CR1 and CR2 appear to be undergoing concerted evolution in the metastriate mtDNA. Replication slippage

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

has been suggested as a mechanism to explain copy number variation and concerted evolution in adjacent repeats in the control regions of the mtDNA of some animals (Zhang and Hewitt 1997). However, other mechanisms must be invoked to explain concerted evolution in nonadjacent duplications. Unequal crossing-over is unlikely because there is no evidence of recombination between animal mitochondrial genomes in which heteroplasmy has been found (Solignac et al. 1984; Densmore, Wright, and Brown 1985; Harrison, Rand, and Wheeler 1985; Hale and Singh 1986). There is no evidence for recombination in most animal mtDNA (Brown 1983). Instead, gene conversion both within and among individual circular molecules during mitochondrial replication seems to be a more likely mechanism. Kumazawa et al. (1996) also argue for gene conversion as a mechanism to explain the concerted evolution exhibited between duplicate control regions in snakes. However, concerted evolution requires a nucleotide mismatch error correction mechanism that, to date, has not been found in a metazoan mitochondria.

The inferred structures of the 22 tRNAs in figure 2 agree with those previously inferred in other arthropods. The DHU arm of all metazoan $tRNA^{Ser(AGY)}$ is replaced with a loop of several nucleotides (Wolstenholme et al. 1987; Okimoto and Wolstenholme 1990; Boore and Brown 1995). The DHU arm of the $tRNA^{Cys}$ was missing in all metastriate species examined, but this has also been previously reported in a reptile mtDNA (*Sphenodon punctatus*; Seutin et al. 1994).

As is the case for other animal mtDNAs, it is possible to identify extensive secondary structure in tick sequences. Potential stem-and-loop structures can be formed at the junctions of protein-coding regions where the start codon for the second gene is located in the 3′ portion of the stem or a few bases 3′ of the stem. For example, at the *I. hexagonus* ATPase6–COIII junction, a 10-bp stem with a 12-base loop can be formed with the 3′ end of the stem four bases upstream of the ATA start codon for cytochrome oxidase 3 (COIII). In *R. sanguineus,* the most energetically favorable structures contain the ATGATA start codons in the stem. Similar structures can be formed at the junctions of COI–COII, ATPase 8–ATPase 6, ND4L–ND4, and NADH dehydrogenase 6–Cytb. Coding-region junctions created by gene rearrangements have been reported in other invertebrates (Boore and Brown 1994; Valverde et al. 1994; Hatzoglou, Rodakism, and Lacanidou 1995). The control regions also have the potential for extensive secondary structure, including RNA-like folding of the whole control sequence, tRNA-like cloverleaf configurations, and simple stem-and-loop figures. Zhang and Hewitt (1997) describe an array of structural elements (sequence and secondary structure) that seem to maintain a conserved order in two groups of insects. For the metastriate ticks, stem and loop structures can be identified which have GAT or GA-rich sequences within 10 bases of the 3′ end of the stem. Some of the stems also have flanking 5′ TATA or poly(T). None of the metastriate control regions completely match Zhang and Hewitt's ''conserved structure.'' The prostriate control region also can form stems and loops but does not appear to exhibit any of Zhang and Hewitt's flanking elements. A stem-and-loop structure located in the control region has been implicated as an origin of replication, but the requirement for and function of flanking sequences remains to be discovered.

Our phylogenetic analysis of mtDNA amino acid sequences in arthropods revealed a remarkable example of convergent molecular evolution. Eight *A. mellifera* amino acid sequences (ATPase6, COI, COII, COIII, Cytb, ND2, ND4L, and ND5) were more similar to tick sequences than they were to other insect sequences, and *A. mellifera* ATPase8, ND4, and ND1 were more similar to chelicerate sequences than they were to other insects. Only *A. mellifera* ND6 and ND3 sequences clustered with insects. Correspondence analysis provides strong evidence that this convergence is caused by a restricted and biased amino acid use in both *R. sanguineus* and *A. mellifera* (figs. 8*A* and 9*A*). However, this amino acid bias appears to be driven in turn by biased use of fewer codons with a high AT content (figs. 8*B* and 9*B*). The exact reasons for this are unclear because it is believed that the 22 tRNAs encoded in metazoan mtDNAs are able to translate transcripts of mitochondrial protein genes through unusual wobble effects that permit the anticodon of one tRNA to recognize all codons in a family (Barrell, Bankier, and Drouin 1979; Barrell et al. 1980). If true, this would reduce selection on codon usage bias, permitting a wide variety of mutations in third positions.

Instead, in honeybee and *R. sanguineus* mtDNA, codon usage bias appears to be driven by unidentified constraints imposed by high AT content. Codons in *A. mellifera* contained 83% AT, and those in *R. sanguineus* contained 78% AT. These two genomes are more similar in AT content than were any of the other nine species examined in figure 7. Dowton and Austin (1995) examined the COI gene in a diversity of Hymenoptera and demonstrated that the AT-rich genome found in *A. mellifera* occurs throughout the entire suborder Apocrita. They reported extreme codon usage bias in the AT-rich codons TTT (Phe), TTA (Leu), ATT (Ile), ATA (Met), AAA (Lys), AAT (Asn), GAT (Gln), and GAA (Glu). The same first six codons are also the most abundant in the *R. sanguineus* genome (table 1: TTT [9.5%], TTA [9.8%], ATT [11.7%], ATA [8.1%], AAA [3.5%], and AAT [5.0%]) and collectively account for 47.6% of all codons. Codon usage bias could cause an amino acid use bias and ultimately lead to the observed convergence in amino acid sequences between *R. sanguineus* and apocritan Hymenoptera.

Acknowledgments

We thank Amy Paddock (Macromolecular Resources) for her excellent technical expertise in sequencing. Thanks to Dave Swofford for permission to use PAUP*4.0 version d61 in the phylogenetic analyses. This research was supported by grant DEB-9420658 from the National Science Foundation to W.C.B., J.S.H. Klompen, and J. E. Keirans.

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

LITERATURE CITED

AZEVEDO, J. L. B., and B. C. HYMAN. 1993. Molecular characterization of lengthy mitochondrial DNA duplications from the parasitic nematode *Romanomermis culicivorax.* Genetics **133**:933–942.

BARRELL, B. G., S. ANDERSON, A. T. BANKIER et al. (15 coauthors). 1980. Different pattern of codon recognition by mammalian mitochondrial tRNAs. Proc. Natl. Acad. Sci. USA **77**:3164–3166.

BARRELL, B. G., A. T. BANKIER, and J. DROUIN. 1979. A different genetic code in human mitochondria. Nature **282**: 189–194.

BEAGLEY, C. T., N. A. OKADA, and D. R. WOLSTENHOLME. 1996. Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile:* one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc. Natl. Acad. Sci. USA **93**:5619–5623.

BEAGLEY, C. T., R. OKIMOTO, and D. R. WOLSTENHOLME. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes and a near-standard genetic code. Genetics **148**:1091–1108.

BEARD, C. B., D. M. HAMM, and F. H. COLLINS. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae,* DNA sequence, genome organization and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. **2**:103–104

BENNETZEN, J. L., and B. D. HALL. 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.

BLACK, W. C. IV, J. S. H. KLOMPEN, and J. E. KEIRANS. 1997. Phylogenetic relationships among tick subfamilies (Ixodida: Ixodidae: Argasidae) based on the 18S nuclear rDNA gene. Mol. Phylo. Evol. **7**:129–144.

BLACK, W. C. IV, and J. PIESMAN. 1994. A phylogeny of hard and soft tick taxa based on mitochondrial 16S ribosomal DNA sequences. Proc. Natl. Acad. Sci. USA **91**:10034–10038.

BOORE, J. L., and W. M. BROWN. 1994. Complete DNA sequence of the mitochondrial genome of the black chiton *Katharina tunicata.* Genetics **138**:423–443.

———. 1995. Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris.* Genetics **141**: 305–319.

BOORE, J. L., T. M. COLLINS, D. STANTON, L. L. DAEHLER, and W. M. BROWN. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature **376**:163–165.

BOORE, J. L., D. V. LAVROV, and W. M. BROWN. 1998. Gene translocation links insects and crustaceans. Nature **392**: 667–668.

BROWN, W. M. 1983. Evolution of animal mitochondrial DNA. Pp. 62–88 *in* M. NEI and R. K. KOEHN, eds. Evolution of genes and proteins. Sinauer, Sunderland, Mass.

CASANE, D., N. DENNEBOUY, H. DEROCHANBEAU, J. C. MOUNOLOU, and M. MONNEROT. 1994. Genetic analysis of systematic mitochondrial heteroplasmy in rabbits. Genetics **138**:471–480.

CLARY, D. O., and D. R. WOLSTENHOLME. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. **22**:252–271.

CROZIER, R. H., and Y. C. CROZIER. 1993. The mitochondrial genome of the honey bee *Apis mellifera:* Complete sequence and genome organization. Genetics **133**:97–117.

DAYHOFF, M. O., R. M. SCHWARTZ, and B. C. ORCUTT. 1978. A model of evolutionary change in proteins. Pp. 345–352 *in* M. O. DAYHOFF, ed. Atlas of protein sequence and structure. Vol. 5, suppl. 3. National Biomedical Research Foundation, Silver Spring, Md.

DENSMORE, L. D., J. W. WRIGHT, and W. M. BROWN. 1985. Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics **110**:689–707

DOWTON, M., and A. D. AUSTIN. 1995. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. J. Mol. Evol. **41**: 958–965.

FELSENSTEIN, J. 1993. PHYLIP (phylogenetic inference package). Version 3.5c Department of Genetics, University of Washington, Seattle.

FLOOK, P. K., C. H. F. ROWELL, and G. GELLISSEN. 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. **41**:928–941.

GREENACRE, M. J. 1984. Theory and applications of correspondence analysis. Academic Press, London.

HALE, L. R., and R. S. SINGH. 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **83**:8813–8817.

HARRISON, R. G., D. M. RAND, and W. C. WHEELER. 1985. Mitochondrial DNA size variation within individual crickets. Science **228**:1446–1448.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **21**:160–174.

HATZOGLOU, E., G. C. RODAKIS, and R. LECANIDOU. 1995. Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea.* Genetics **140**:1353–1366.

HIGGINS, D. G., and P. M. SHARP. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5**:151–153

HOFFMANN, R. J., J. L. BOORE, and W. M. BROWN. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis.* Genetics **131**:397–412.

IKEMURA, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* system. J. Mol. Biol. **151**:389–409.

KAMBHAMPATI, S. 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. Proc. Natl. Acad. Sci. USA **92**:2017–2020.

KUMAZAWA, Y., H. OTA, M. NISHIDA, and T. OZAWA. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Mol. Biol. Evol. **13**:1242–1254.

KYTE, J., and R. DOOLITTLE. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.

LEWIS, D. L., C. L. FARR, and L. S. KAGUNI. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol. Biol. **4**:263–278

LOBRY, J. R., and C. GAUTIER. 1994. Hydrophobicity, expressivity and aromaticity are the major trends of amino acid usage in 999 *Escherichia coli* chromosome encoded genes. Nucl. Acid Res. **22**:3174–3180.

MITCHELL, S. E., A. F. COCKBURN, and J. A. SEAWRIGHT. 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome **36**:1058–1073.

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022

MORITZ, C. 1991. Evolutionary dynamics of mitochondrial DNA duplication in parthenogenetic geckos, *Heteronotia binoei.* Genetics **129**:221–230.

MORITZ, C., and W. M. BROWN. 1986. Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233**:1425–1427.

———. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84**:7183–7187.

OKIMOTO, R., H. M. CHAMBERLIN, J. L. MACFARLANE, and D. R. WOLSTENHOLME. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes *(Meloidogyne):* nucleotide sequences, genome location and potential for host-race identification. Nucl. Acids Res. **19**: 1619–1626.

OKIMOTO, R., and D. R. WOLSTENHOLME. 1990. A set of tRNAs that lack either the TyC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. EMBO J. **9**:3405–3411.

PONT-KINGDON, G. A., N. A. OKADA, J. L. MACFARLANE, C. T. BEAGLEY, T. CAVALIER-SMITH, D. R. WOLSTENHOLME, and G. D. CLARK-WALKER. 1995. A coral mitochondrial *mutS* gene. Nature **375**:109–111.

ROEHRDANZ, R. L. 1993. An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. Insect Mol. Biol. **2**:89–91.

———. 1995. Amplification of complete insect mitochondrial genome in two easy pieces. Insect Mol. Biol. **4**:169–172.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SAS INSTITUTE, INC. 1998. SAS. Version 6.12, Cary, N.C.

SEUTIN, G., B. F. LANG, D. P. MINDELL, and R. MORAIS. 1994. Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11**:329–340.

SHARP, P. M., and W. H. LI. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucl. Acids Res. **15**: 1281–1295.

SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, and P. FLOOK. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. **87**:651–701

SOLIGNAC, M., J. GENERMONT, M. MONNEROT, and J. C. MOUNOLOU. 1984. Mitochondrial genetics of *Drosophila:* mtDNA segregation in heteroplasmic strains of *D. mauritiana.* Mol. Gen. Genet. **197**:183–188.

STANTON, D. J., L. L. DAEHLER, C. C. MORITZ, and W. M. BROWN. 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics **137**:233–241.

STATON, J. L., L. L. DAEHLER, and W. M. BROWN. 1997. Mitochondrial gene arrangement of the horseshoe crab *Limulus polyphemus* L.: Conservation of major features among arthropod classes. Mol. Biol. Evol. **14**:867–874.

THIOULOUSE, J., and F. CHEVENET. 1996. Netmul, a world-wide-web user interface for multivariate analysis software. Comp. Stat. Data Anal. **21**:369–372.

THIOULOUSE, J., S. DOLEDEC, D. CHESSEL, and J. M. OLIVER. 1995. ADE software: multivariate analysis and graphical display of environmental data. Pp. 57–61 *in* G. GUARISO and A. RIZZOLI, eds. Sofware per l'Ambiente, Patron editor, Bolonia.

VALVERDE, J. R., B. BATUECAS, C. MORATILLA, R. MARCO, and R. GARESSE. 1994. The complete mitochondrial DNA sequence of the crustacean *Artemia franciscana.* J. Mol. Evol. **39**:400–408.

WOLSTENHOLME, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Pp. 173–216 *in* D.R. WOLSTENHOLME and K.W. JEON, eds. Mitochondrial genomes. International review of cytology, Vol. 141. Academic Press, New York.

WOLSTENHOLME, D. R., and C. FAURON. 1995. Mitochondrial genome organization. Pp. 1–59 *in* C. S. LEVINGS III and I. K. VASIL, eds. Advances in cellular and molecular biology of plants, Vol. 3: molecular biology of the mitochondria. Academic Publishers, Dordrecht, The Netherlands.

WOLSTENHOLME, D. R., J. L. MACFARLANE, R. OKIMOTO, D. O. CLARY, and J. A. WAHLEIGHNER. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Sci. USA **84**:1324–1328.

WRIGHT, F. 1990. The effective number of codons used in a gene. Gene **87**:23–29.

XIONG, B., and T. D. KOCHER. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies Diptera Simulidae. Genome **34**: 306–311.

ZEVERING C. E., C. MORITZ, A. HEIDEMAN, and R. A. STURM. 1991. Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*; Gekkonidae). J. Mol. Evol. **33**: 431–441

ZHANG, D. X., and G. M. HEWITT. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Evol. **2**:99–120.

RICHARD H. THOMAS, reviewing editor

Accepted August 24, 1998

Downloaded from https://academic.oup.com/mbe/article/15/12/1772/963112 by guest on 16 August 2022