Mol. Cells *28*, 155-165, September 30, 2009
DOI/10.1007/s10059-009-0125-9

Molecules and Cells

Complete Mitochondrial Genome of Brown Marmorated Stink Bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and Phylogenetic Relationships of Hemipteran Suborders

Wonhoon Lee[1,5], Joongnam Kang[2,5], Chansik Jung[2], Kim Hoelmer[3], Si Hyeock Lee[4], and Seunghwan Lee[1,*]

The newly sequenced complete mitochondrial genome of the brown marmorated stink bug, *Halyomorpha halys* (Stål) (Hemiptera: Pentatomidae), is a circular molecule of 16,518 bp with a total A+T content of 76.4% and two extensive repeat regions in A+T rich region. Nucleotide composition and codon usage of *H. halys* are about average when compared with values observed in 19 other published hemipteran mitochondrial genomes. Phylogenetic analyses using these 20 hemipteran mitochondrial genomes support the currently accepted hypothesis that suborders Heteroptera and Auchenorrhyncha form a monophyletic group. The mitochondrial gene arrangements of the 20 genomes are also consistent with our results.

INTRODUCTION

Mitochondrial genomes of the superclass Hexapoda are generally circular, 14-16 kb in length and typically consist of 37 genes: 13 protein coding genes (PCGs), 22 transfer RNA genes (tRNAs), and 2 ribosomal RNA genes (rRNAs). A total of 132 completely sequenced mitochondrial genomes have been reported in Hexapoda (Lee et al., 2009). The mitochondrial genome has been considered as a useful marker for reconstructing the phylogeny and evolutionary relationships within Hexapoda (Kim et al., 2009). Generally, nucleotide sequences of single or multiple genes in mitochondrial genomes are used for phylogenetic analysis at low taxonomic levels, such as species, genus, or family, and gene rearrangements, e.g. gene order and content variability, are considered useful markers of phylogenetic signals at higher taxonomic levels, such as superfamily and suborder (Boore, 1999; Boore and Brown, 1998; Dowton, 1999; Rawlings et al., 2001)

The order Hemiptera is the largest group of the hemimetabolous insects (Schuh and Slater, 1995), including the three phylogenetically controversial suborders, Auchenorrhyncha, Sternorrhyncha, and Heteroptera (Carver et al., 1991). Before the term "Hemiptera" became common, most entomologists used the terms Heteroptera and Homoptera as two independent orders, and Sternorrhyncha and Auchenorrhyncha were considered suborders of Homoptera (Kristensen, 1991; von Dohlen and Moran, 1995). Early taxonomists considered Homoptera and Heteroptera to be either separate orders or equivalently ranked suborders of the order Hemiptera (Poisson and Pesson, 1951). However, many researchers have debated whether the suborder Sternorrhyncha and Auchenorrhyncha constitute a monophyletic group or not, and have suggested four putative theories about the phylogenetic relationship of three suborders in Hemiptera: 1) impossibility of determining the relationships of the three suborders (Hennig, 1981; Kristensen, 1973; 1981; Scudder, 1973; Szelegiewicz, 1971), 2) existence of sufficient synapomorphies to unite Heteroptera with Auchenorrhyncha (Carver et al., 1991; Cobben, 1978; Goodchild, 1966; Kukalová-Peck, 1991; Wootton and Betts, 1986), 3) constitution of paraphyly on Auchenorrhyncha (Goodchild, 1966; Hamilton, 1981; Ross, 1965), and 4) constitution of each monophyly on Sternorrhyncha and Auchenorrhyncha (Carver et al., 1991; Hennig, 1981; Kristensen, 1973; Szelegiewicz, 1971). Recent study based on 18S rDNA suggested that Sternorrhyncha is a sister group to a group comprising Auchenorrhyncha and the Heteroptera (von Dohlen and Moran, 1995).

In this study, we sequenced the mitochondrial genome of the brown marmorated stink bug, *Halyomorpha halys* (Stål) (Hemiptera: Pentatomidae), an important polyphagous horticultural pest in northeast Asia which recently became invasive in North America (Hoebeke and Carter, 2003). Our analyses focused on 1) the characterization of the evolution and structural organization of *H. halys* mitochondrial genome, 2) the evaluation of mitochondrial genome sequences for inferring hemip-

[1]Insect Biosystematics Laboratory, Research Institute for Agricultural and Life Sciences, Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea, [2]Department of Forest Insect Pests and Diseases, Korea Forest Research Institute, Korea, [3]USDA Agricultural Research Service, Beneficial Insect Introduction Research Laboratory, Newark, DE 19711, USA, [4]Insect Molecular Biology and Toxicology Laboratory, Research Institute for Agricultural and Life Sciences, Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea, [5]These authors contributed equally to this work.
*Correspondence: seung@snu.ac.kr

Received March 17, 2009; revised June 21, 2009; accepted August 4, 2009; published online September 4, 2009

Keywords: gene rearrangement, *Halyomorpha halys*, Hemiptera, mitochondrial genome, phylogeny

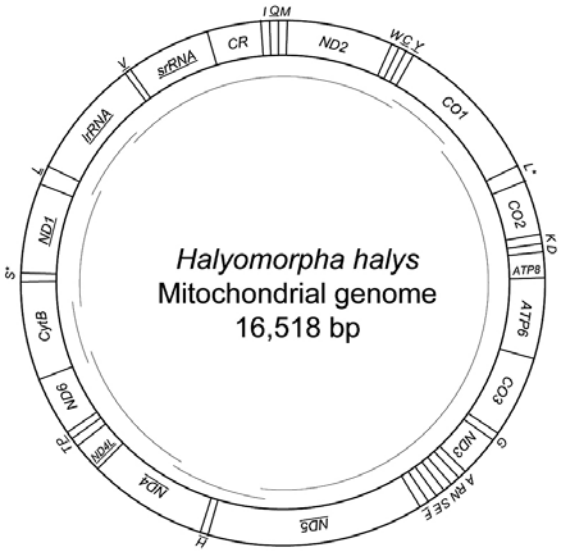

Fig. 1. Circular map of *Halyomorpha halys* mitochondrial genome. The abbreviations for the genes are as follows: *CO1-3*, cytochrome oxidase subunits I-III; *CytB*, cytochrome B; *ATP6* and *ATP8*, ATP synthase F0 subunits 6 and 8; *ND1-6*, NADH dehydrogenase subunits 1-6; *ND4L*, NADH dehydrogenase subunit 4L; *lrRNA*, large subunit ribosomal RNA; *srRNA*, small subunit ribosomal RNA; *CR*, A+T-rich region: *A*, $tRNA^{Ala}$; *C*, $tRNA^{Cys}$; D, $tRNA^{Asp}$; *E*, $tRNA^{Glu}$; *F*, $tRNA^{Phe}$; G, $tRNA^{Gly}$; H, $tRNA^{His}$; *I*, $tRNA^{Ile}$; K, $tRNA^{Lys}$; *L*, $tRNA^{Leu}$; *M*, $tRNA^{Met}$; *N*, $tRNA^{Asn}$; Q, $tRNA^{Gln}$; *R*, $tRNA^{Arg}$; *S*, $tRNA^{Ser}$; *T*, $tRNA^{Thr}$; *P*, $tRNA^{Pro}$; *V*, $tRNA^{Val}$; *W*, $tRNA^{Trp}$; *Y*, $tRNA^{Tyr}$. Gene names that are not underlined indicate a clockwise direction of transcription, while the underlined genes indicate counter clockwise transcription.

teran phylogeny, and 3) the comparison between the phylogenetic relationship and gene rearrangements to investigate the variability of mitochondrial genomes at the suborder level of Hemiptera.

MATERIALS AND METHODS

Mitochondrial DNA extraction, primer design, PCR, and sequencing

An adult individual of *H. halys* was collected on *Glycine max* Merrill (Leguminosae) from Seoul, Korea, in 2007, and preserved in 98% ethanol. Two hind legs of the specimen were used to extract total DNA, using the DNeasy Tissue Kit (QIAGEN), according to the manufacturer's protocols.

The results of amplification reactions of the mitochondrial genome of *H. halys* are presented as a genome map in Fig. 1. Firstly, the fragments of the four gene sets, cytochrome c oxidase subunits I (*CO1*), NADH dehydrogenase subunits 4 & 5 (*ND4* & *ND5*), cytochrome b (*CytB*), and the large subunit of ribosomal RNA (*lrRNA*), were amplified using previously published primers (Table 1). The amplification was performed on a PTC-0220 Peltier Thermocycler (MJ Research), with 1 ul of genomic DNA as a template, in a total reaction volume of 20 μl containing Ex *Taq*™ (Takara). The PCR was run with the following settings: 2 min initial denaturation at 94°C, followed by 30 cycles of 30 s denaturation at 94°C, 30 s annealing at 55°C (*CO1*, *CytB*, and *lrRNA*) or 45°C (*ND5*), 1-2 min extension at 72°C depending on the size of products and finalized by 7 min at 72°C. Secondly, on the basis of determined sequence information, new sets of primers were designed for the complete amplification of the mitochondrial genome using LA *Taq*™ (Takara) (Table 1). The amplification of long fragments was also performed on a PTC-0220 Peltier Thermocycler, with 2 ul of genomic DNA as a template, in a total reaction volume of 50 μl. The PCR was run with the following settings: 2 min initial denaturation at 94°C, followed by 30 cycles of 10 s at 98°C, 10 min at 60°C, and finalized by 7 min at 72°C.

All PCR fragments were sequenced directly after purification with QIAquick PCR purification Kit (QIAGEN), all of the long products were cloned into pCR4-TOPO sequencing vector (Invitrogen, USA), and the resulting plasmid DNA was isolated using the Wizard Plus SV Minipreps DNA Purification System (Promega). Internal primers were applied to all fragments to complete the sequences by primer walking. DNA sequencing was conducted using ABI 3730xl DNA analyzer and the ABI BigDye Terminators version 3.1 sequencing kit (Applied Biosystems). All fragments were sequenced from both strands.

Assembling annotation and sequence analysis

Sequences from overlapping fragments were assembled by aligning neighboring fragments using CLUSTAL X (Thompson et al., 1997). Individual *H. halys* mitochondrial genes and the A+T-rich region were determined by aligning homologous sequences of known full-length insect mitochondrial sequences using CLUSTAL X. The nucleotide sequences of PCGs were translated based on the invertebrate mtDNA genetic code. The secondary structures of tRNAs were predicted on the basis of the nucleotide sequences of the tRNAs of other insects, confirmed by tRNAScan-SE 1.1 program (Lowe and Eddy, 1997), and edited by eye. Codon usage analysis was carried out by IMGD (http://www.imgd.org; Lee et al., 2009). Major noncoding repeat regions between the small ribosomal subunit (*srRNA*) and the $tRNA^{Ile}$ gene were identified using Tandem Repeats Finder ver. 4.0 (Benson, 1999). The sequence data have been deposited into the GenBank database under the accession no. FJ685650.

Phylogenetic analysis

Nucleotide sequences of 13 PCGs in *H. halys* and 19 other known hemipteran mitochondrial genomes were analyzed in order to understand the phylogenetic relationships existing among three suborders of Hemiptera (Table 2). *Thrips imaginis* Bagnall (Thysanoptera: Thripidae) (Shao and Barker, 2003) was used as an outgroup. Alignments of nucleotide sequences of 13 individual PCGs were constructed using CLUSTAL X under default conditions. The well-aligned blocks from nucleotide sequences were then selected with GBlocks 0.91b (Castresana, 2000) under default conditions, resulting in a total length of 10,531 nucleotides, which is equivalent to 90% of the original length.

Phylogenetic reconstructions were done with maximum likelihood (ML) and Bayesian inference (BI) analyses. The ML trees were reconstructed using PAUP* 4.0b10 (Swofford, 2002) under the best-fit model (GTR + I + G), which was chosen as the most appropriate for the present case based on hierarchical likelihood tests by Modeltest 3.7 (Posada and Crandall, 1998). Robustness of each internal branch of the ML tree estimated was evaluated with 100 bootstrap replications (Felsenstein, 1985). The BI trees were obtained with MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001) using the Markov Chain Monte Carlo technique (MCMC). We chose the GTR + I + G model based on hierarchical likelihood-ratio test (hLRTs) by MrModeltest v2.2 (Nylander et al., 2004). Model parameter values were treated as unknown and were estimated for each analysis. Random starting trees were used, and analyses were run for 500,000 generations, sampling

Table 1. List of primers used for amplification in this study

Locus	Primer	Sequence (5′ → 3′)	References
CO1	LCO1490f	GGTCAACAAATCATAAAGATATTGG	(Folmer et al., 1994)
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	(Folmer et al., 1994)
ND5&*ND4*	N5-J-7502	CTAAAGTTGATAATGAACTAAAG	(Kambhampati and Smith, 1995)
	N4-N-8925	GCTCATGTTGAAGCTCC	(Jermiin and Crozier, 1994)
CytB	10933*	TAGGATATGTTTTACCTTGAGGAC	(Muraji et al., 2000)
	11388*	TCCTCCTAATTTATTAGGAATTG	(Muraji et al., 2000)
lrRNA	12866*	CCGGTTTGAACTCAGATCATGT	(Muraji and Tachikawa, 2000)
	13417*	CGCCTGTTTAACAAAAACAT	(Muraji and Tachikawa, 2000)
CO1-ND5	CN-F	GAGCACCTGATATAGCCTTCCCACGAT	
	CN-R	CTGCTATAGCTGCCCCAACTCCTGT	
ND5&*ND4-CytB*	NC-F	TAAATAATCTACGGCTTCTTCTACG	
	NC-R	ATAAAATTTTCCGGGTCTCCTA	
CytB-lrRNA	C16-F	CCATTCCACCCTTACTTTTCAACA	
	C16-R	AAGGCTGGAATGAACGGTCGGATGA	
lrRNA-CO1	16C-F	AATGGTCCTTTCGTACTAACTCACAAT	
	16C-R	TGCTAACACAGGTAAGGATAATAACAGAA	

*The names of the primers refer to the sequence location in *Drosophila yakuba*

every 100 generations. Bayesian posterior probabilities were then calculated from the sample points after the Markov Chain Monte Carlo (MCMC) algorithm began to converge. To ensure that our analyses were not trapped in local optima, four independent MCMC runs were performed. Topologies and posterior clade probabilities from different runs were compared for congruence.

Gene order comparison of Hemipteran mitochondrial genomes

Gene order and content were analyzed on 13 PCGs, 22 tRNAs, and 2 rRNAs of the 20 hemipteran mitochondrial genomes comparing that of *Drosophila yakuba* known as an ancestral form in Hexapoda (Boore, 1999; Boore et al., 1995; Rokas and Holland, 2000). Gene order and content were compared with the BI tree analyzed in this study.

RESULTS AND DISCUSSION

Features of *Halyomorpha halys* mitochondrial genome

The mitochondrial genome size of *H. halys* is 16,518 bp, well within the range of 19 other published hemipteran genomes, which extend from 14,496 bp (*N. andropogonis*) to 18,414 bp (*T. vaporariorum*) and have an average value of 15,998 bp (Table 2). This genome resembles that of the known ancestral species *D. yakuba* in terms of its structural organization and composition: 13 protein coding genes (PCGs), 22 transfer RNAs (tRNAs), and 2 ribosomal RNAs (rRNAs) (Fig. 1). Fifteen of the 19 genomes (78.95%) contained 37 genes like that of *H. halys*, whereas the rest of the genomes have different numbers only of tRNAs, probably due to gene deletion events (Fig. 4; Table 5). For example, *N. andropogonis* (Sternorrhyncha) includes 18 tRNAs, but *A. aceris* (Sternorrhyncha) contains 21 tRNAs.

The mitochondrial genes of *H. halys* overlap in a total of 62 bp at 11 locations, with the longest three measuring 8 bp located in $tRNA^{Trp}$-$tRNA^{Cys}$, $tRNA^{Gly}$-*ND3*, and $tRNA^{Leu}$-*lrRNA*, and a total of 119 bp intergenic spacer sequences occur and are spread over 18 regions, ranging in size between 1 and 34 bp. The longest spacer sequence (34 bp) is located between $tRNA^{Ser}$-*ND1* (Table 3). Similarly-sized intergenic spacer sequences were detected in $tRNA^{Asp}$-*ATP8* (36 bp, *A. aceris*), $tRNA^{Pro}$-*ND6* (34 bp, *A. aceris*), $tRNA^{Pro}$-*ND6* (34 bp, *B. tabaci*), and *ATP6-CO3* (36 bp, *T. vaporariorum*), *ND4L*-$tRNA^{Pro}$ (33 bp, *D. cingulatus*) (Hua et al., 2008; Thao et al., 2004). The control region of *H. halys* was observed to be 1,815 bp in length and includes two repeat regions. The first repeat region is situated from 15,178 bp to 15,469 bp, including four repeat units, and the second region is situated from 15,638 bp to 16,513 bp, including twelve repeat units (Table 3).

In *H. halys*, conventional Met start codons could be assigned to nine of the 13 PCGs, whereas three genes, *ND2*, *ND4L*, and *ND5*, used Ile start codon and *CO1* use Leu start codon (Table 3). The Met and Ile start codons were reported in the other 19 hemipteran genomes; however, Leu start codon was reported in only 11 of these genomes: *C. bifaria*, *Y. parallelus*, *M. inconspicuous*, *H. longirostris, R. pedestris*, *D. cingulatus*, *P. gutta*, and *N parus* (*CO1*); *N. viridula* (*CO1* and *ATP8*); *P. spumarius* (*CO2*, *ND2* and *ND3*); the *T. acacia* (*CO2*). Conventional stop codons could be assigned to most of the putative protein sequences, but only the *CO2* gene terminated into a single T residue abutting directly to adjacent tRNAs. Identical situations have been described for all 19 hemipteran genomes (Table 3) (Hua et al., 2008; Thao et al., 2004).

All tRNA sequences could be folded into typical cloverleaf secondary structures exposing appropriate anticodon triplets (Fig. 2). The standard 22 tRNA genes were identified in the same relative genomic positions as observed for the utative ancestral insect, *D. yakuba* (Boore, 1999; Boore et al., 1995; Rokas and Holland, 2000). Because the tRNAScane-SE was unable to detect $tRNA^{Arg}$, $tRNA^{Ser}$(AGN), and $tRNA^{His}$, these tRNAs were identified through alignment of other insect tRNA genes to unassigned nucleotide gaps corresponding to the

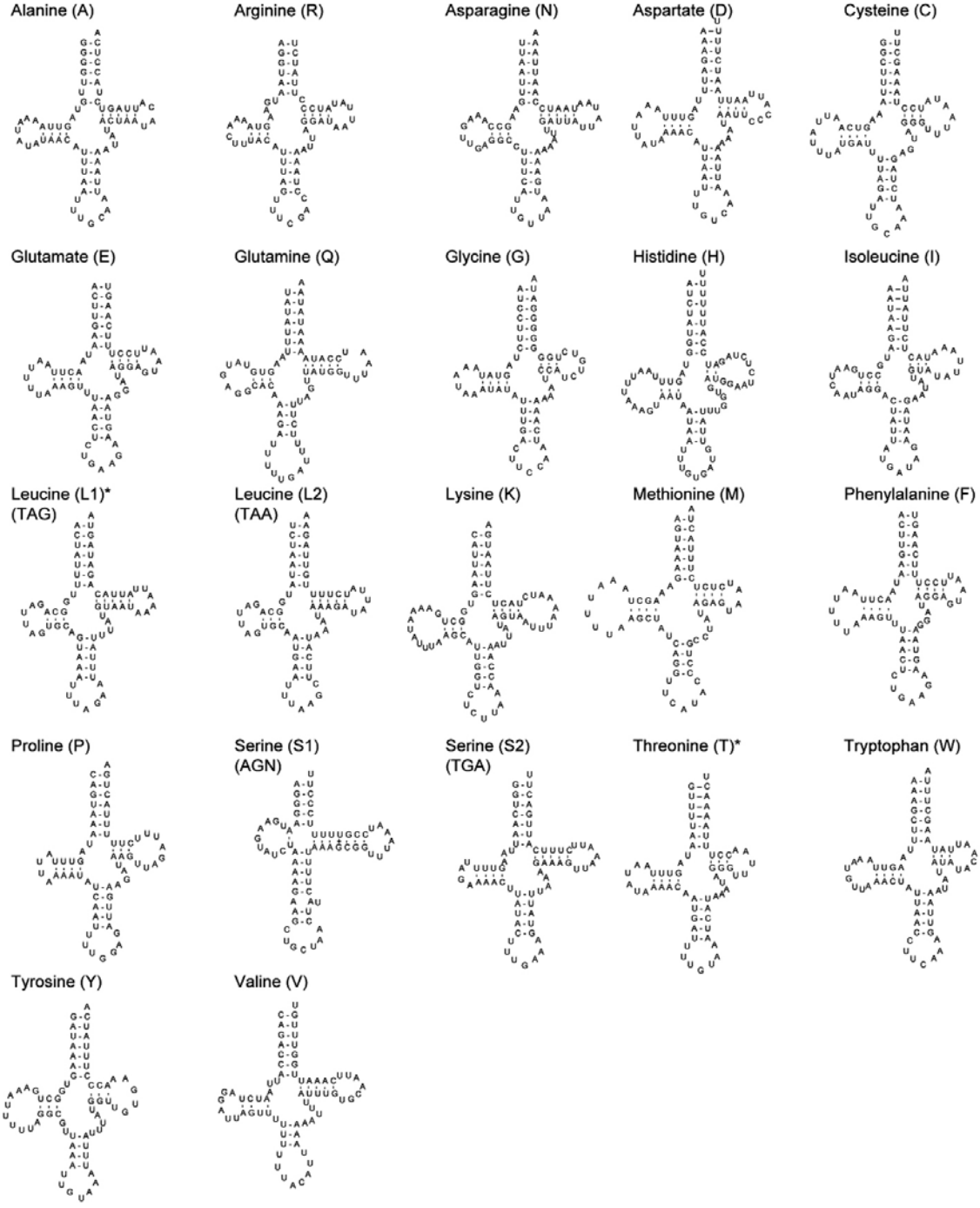

Fig. 2. Putative secondary structure folds for the 22 tRNA genes of *Halyomorpha halys*.

conserved relative position of these missing tRNA genes.

Base composition, codon usage, and amino acid composition

The mitochondrial genome of *H. halys* is heavily biased towards a high A + T content. The overall A + T content was 76.2% for *H. halys*, and this value corresponds well to the AT bias generally observed in the 19 hemipteran mitochondrial genomes, which ranges from 68.9% (*N. parus*) to 86.4% (*A. dugesii*) with an average value of 76.2% (Table 2).

The codon usage of *H. halys* reveals the genome-wide base compositional bias for AT (Table 4). All codons with A or T in the third position are generally overrepresented compared to G or C in third position. The ratio of GC-rich codons (Arg, Pro, Gly, and Ala) versus AT-rich codons (Lys, Asn, Tyr, Phe, Ile, and Met) of amino acids was 0.32 in *H. halys*, which is within the range from 0.16 (*A. dugesii*) to 0.43 (*N. parus* and *T. dimidiata*) with average value of 0.31 in the 19 genomes (Table 4; Sup-

Table 2. Characteristics of the 20 hemipteran mitochondrial genomes

Suborder	Superfamily	Species	Genome size (bp)	PCGs	tRNAs	rRNAs	A(%)	T(%)	G(%)	C (%)	AT-skew	Accession	References
Heteroptera	Pentatomoidea	*Halyomorpha halys*	16,518	13	22	2	43.1	33.3	10.1	13.6	0.128	FJ685650	This study
	Pentatomoidea	*Coptosoma bifaria*	16,179	13	22	2	40.9	30.4	12.5	16.2	0.147	EU427334	(Hua et al., 2008)
	Pentatomoidea	*Nezara viridula*	16,889	13	22	2	43.2	33.6	9.8	13.3	0.125	NC_011755	(Hua et al., 2008)
	Aradoidea	*Neuroctenus parus*	15,354	13	22	2	41.3	27.6	12.1	19.0	0.199	EU427340	(Hua et al., 2008)
	Coreoidea	*Hydaropsis longirostris*	16,521	13	22	2	41.6	33.9	9.2	15.4	0.102	EU427337	(Hua et al., 2008)
	Coreoidea	*Riptortus pedestris*	17,191	13	22	2	41.9	34.7	9.4	14.0	0.093	EU427344	(Hua et al., 2008)
	Lygaeoidea	*Malcus inconspicuus*	15,575	13	22	2	44.3	33.5	8.9	13.3	0.138	EU427339	(Hua et al., 2008)
	Lygaeoidea	*Yemmalysus parallelus*	15,747	13	22	2	42.1	35.1	10.3	12.5	0.091	EU427346	(Hua et al., 2008)
	Pyrrhocoroidea	*Dysdercus cingulatus*	16,249	13	22	2	44.1	33.6	8.7	13.6	0.135	EU427335	(Hua et al., 2008)
	Pyrrhocoroidea	*Physopelta gutta*	14,935	13	22	2	44.9	29.6	10.1	15.4	0.206	EU427343	(Hua et al., 2008)
	Reduvioidea	*Triatoma dimidiata*	17,019	13	22	2	40.6	28.9	11.2	19.3	0.168	NC 002609	(Dotson and Beard, 2001)
	Leptopodoidea	*Saldula arsenjevi*	15,324	13	22	2	43.2	31.4	10.9	14.5	0.158	EU427345	(Hua et al., 2008)
Auchenorrhynca	Cercopoidea	*Philaenus spumarius*	16,324	13	22	2	40.9	36.1	10.7	12.4	0.063	NC 005944	(Stewart and Beckenbach, 2005)
Sternorrhyncha	Aphidoidea	*Schizaphis graminum*	15,721	13	22	2	44.8	39.2	6.0	10.1	0.067	NC 006158	(Thao et al., 2004)
	Aleyrodoidea	*Aleurochiton aceris*	15,388	13	21	2	33.7	44.2	12.5	9.6	-0.135	NC 006160	(Thao et al., 2004)
	Aleyrodoidea	*Aleurodicus dugesii*	15,723	13	20	2	39.7	46.7	8.1	5.5	-0.081	NC 005939	(Thao et al., 2004)
	Aleyrodoidea	*Bemisia tabaci*	15,322	13	22	2	33.1	42.5	13.2	11.1	−0.124	NC 006279	(Thao et al., 2004)
	Aleyrodoidea	*Neomaskellia andropogonis*	14,496	13	18	2	35.4	45.9	11.3	7.4	0.016	NC 006159	(Thao et al., 2004)
	Aleyrodoidea	*Tetraleurodes acaciae*	15,080	13	19	2	30.2	41.8	15.6	12.5	−0.161	NC 006292	(Thao et al., 2004)
	Aleyrodoidea	*Trialeurodes vaporariorum*	18,414	13	22	2	29.6	42.8	16.6	11.1	−0.182	NC 006280	(Thao et al., 2004)

Table 3. Annotation and gene organization of *Halyomorpha halys* mitochondrial genome

Gene	Direction	Nucleotide number	Overlap	Interspace	Size	Anticodon		Start codon	Stop codon
$tRNA^{Ile}$	F	1-67			67	32-34	GAT		
$tRNA^{Gln}$	R	65-133	3		69	101-103	TTG		
$tRNA^{Met}$	F	140-206		6	67	173-175	CAT		
ND2	F	207-1178			972	·	·	ATT (Ile)	TAA
$tRNA^{Trp}$	F	1192-1258		13	67	1225-1227	TCA		
$tRNA^{Cys}$	R	1251-1318	8		68	1283-1285	GCA		
$tRNA^{Tyr}$	R	1322-1388		3	67	1353-1355	GTA		
CO1	F	1397-2938		8	1542	·	·	TTG (Leu)	TAA
$tRNA^{Leu}$	F	2934-2999	5		66	2963-2965	TAA		
CO2	F	3000-3678			679	·	·	ATA (Met)	T(AA)
$tRNA^{Lys}$	F	3679-3753			75	3701-3703	CTT		
$tRNA^{Asp}$	F	3761-3829		7	67	3794-3796	GTC		
ATP8	F	3830-3988			159	·	·	ATA (Met)	TAA
ATP6	F	3982-4656	7		675	·	·	ATG (Met)	TAA
CO3	F	4661-5449		4	789	·	·	ATG (Met)	TAA
$tRNA^{Gly}$	F	5455-5524		5	70	5486-5488	TCC		
ND3	F	5517-5873	8		357	·	·	ATA (Met)	TAA
$tRNA^{Ala}$	F	5879-5946		5	68	5911-5913	TGC		
$tRNA^{Arg}$	F	5953-6016		6	64	5981-5983	TCG		
$tRNA^{Asn}$	F	6021-6090		4	70	6052-6054	GTT		
$tRNA^{Ser}$	F	6091-6157			67	6117-6119	GCT		
$tRNA^{Glu}$	F	6158-6222			65	6189-6191	TTC		
$tRNA^{Phe}$	R	6222-6288	1		67	6253-6255	GAA		
ND5	R	6296-8008		7	1713		·	ATT (Ile)	TAA
$tRNA^{His}$	R	8003-8074	6		72	8038-8040	CAC		
ND4	R	8077-9405		2	1329	·	·	ATG (Met)	TAA
ND4L	R	9399-9686	7		288	·	·	ATT (Ile)	TAA
$tRNA^{Thr}$	F	9689-9753		2	65	9720-9722	TGT		
$tRNA^{Pro}$	R	9754-9818			65	9787-9789	TGG		
ND6	F	9822-10301		3	480	·	·	ATG (Met)	TAA
CytB	F	10306-11445		4	1140	·	·	ATG (Met)	TAA
$tRNA^{Ser}$	F	11451-11519		5	69	11481-11483	TGA		
ND1	R	11554-12480		34	927	·	·	ATA (Met)	TAA
$tRNA^{Leu}$	R	12475-12541	6		67	12510-12512	TAG		
lrRNA	R	12534-13824	8		1291	·	·		
$tRNA^{Val}$	R	13822-13889	3		68	13858-13860	GTA		
srRNA	R	13891-14703		1	813	·	·		
A+T-rich region		14704-16518			1815	·	·		

Table 4. Codon usage in the *Halyomorpha halys* mitochondrial genome

Amino acid	Codon	*n*	RSCU	Amino acid	Codon	*n*	RSCU
Leu	TTA	357	0.10	Ser	TCT	100	0.03
	TTG	41	0.01		TCA	102	0.03
	CTG	4	0.00		TCC	17	0.00
	CTC	2	0.00		TCG	4	0.00
	CTT	45	0.01		AGA	98	0.03
	CTA	47	0.01		AGT	36	0.01
Gln	CAG	8	0.00		AGG	2	0.00
	CAA	44	0.01		AGC	6	0.00
Trp	TGA	85	0.02	Cys	TGC	3	0.00
	TGG	13	0.00		TGT	42	0.01
Val	GTA	80	0.02	Tyr	TAC	28	0.01
	GTC	8	0.00		TAT	155	0.04
	GTG	7	0.00	Phe	TTC	68	0.02
	GTT	73	0.02		TTT	267	0.07
Asp	GAC	11	0.00	Gly	GGC	15	0.00
	GAT	55	0.01		GGA	99	0.03
Arg	CGC	3	0.00		GGT	64	0.02
	CGT	11	0.00		GGG	21	0.01
	CGA	36	0.01	Ala	GCA	54	0.01
	CGG	5	0.00		GCC	15	0.00
His	CAT	52	0.01		GCT	44	0.01
	CAC	16	0.00		GCG	2	0.00
Pro	CCG	5	0.00	Ile	ATC	47	0.01
	CCA	52	0.01		ATT	370	0.10
	CCT	57	0.02	Thr	ACT	69	0.02
	CCC	15	0.00		ACC	15	0.00
Glu	GAG	6	0.00		ACG	2	0.00
	GAA	80	0.02		ACA	86	0.02
Lys	AAA	102	0.03	Met	ATG	30	0.01
	AAG	17	0.00		ATA	291	0.08
Asn	AAT	166	0.04	*	TAG	0	0.00
	AAC	31	0.01		TAA	12	0.00

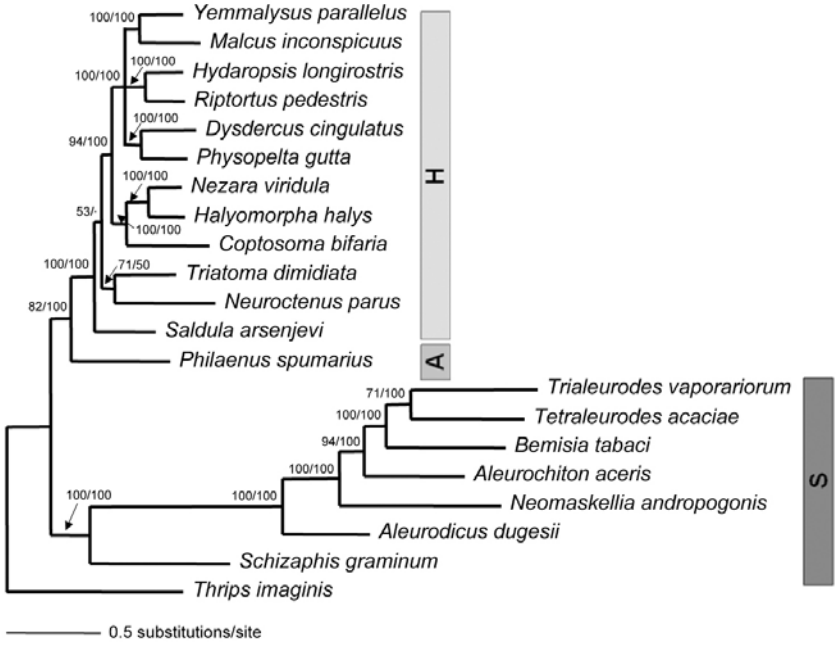

Fig. 3. Phylogenetic relationships of Hemiptera. Maximum likelihood (ML) tree for 20 hemipteran species with one outgroup, *Thrips imaginis* (Thysanoptera: Thripidae), based on 13 PCGs. Numbers on branches refer to ML bootstraps followed by BI posterior probabilities. Species membership in the three suborders of Hemiptera are indicated by the bars adjacent to the trees as follows: H, Heteroptera; A, Auchenorrhyncha; S, Sternorrhyncha.

Table 5. Check list of gene inversions, deletions, and translocations on 20 hemipteran mitochondrial genomes

Suborder	Supterfamily	Species	tRNAs			PCGs		rRNAs	Total
			Inversion	Translocation	Deletion	Inversion	Translocation	Inversion	
Heteroptera	Pentatomoidea	*Halyomorpha halys*	·	·	·	·	·	·	0
	Pentatomoidea	*Coptosoma bifaria*	·	·	·	·	·	·	0
	Pentatomoidea	*Nezara viridula*	·	·	·	·	·	·	0
	Aradoidea	*Neuroctenus parus*	·	Q	·	·	·	·	1
	Coreoidea	*Hydaropsis longirostris*	·	·	·	·	·	·	0
	Coreoidea	*Riptortus pedestris*	·	·	·	·	·	·	0
	Lygaeoidea	*Malcus inconspicuus*	·	·	·	·	·	·	0
	Lygaeoidea	*Yemmalysus parallelus*	·	·	·	·	·	·	0
	Pyrrhocoroidea	*Dysdercus cingulatus*	·	T (or P)	·	·	·	·	1
	Pyrrhocoroidea	*Physopelta gutta*	·	T (or P)	·	·	·	·	1
	Reduvioidea	*Triatoma dimidiata*	·	·	·	·	·	·	0
	Leptopodoidea	*Saldula arsenjevi*	·	·	·	·	·	·	0
Auchenorrhynca	Cercopoidea	*Philaenus spumarius*	·	·	·	·	·	·	0
Sternorrhyncha	Aphidoidea	*Schizaphis graminum*	·	·	S_1	·	·	·	1
	Aleyrodoidea	*Aleurochiton aceris*	R, A, G, N, S_1	G, A, R, N, Q	I	ND3, CO3	CO3, ND3	·	15
	Aleyrodoidea	*Aleurodicus dugesii*	·	C (or Y)	S_1, Q	·	·	·	3
	Aleyrodoidea	*Bemisia tabaci*	N, R, A, G, S_1, E, D	G, Q, D, N, R, A	·	CO3, ND3	CO3, ND3	·	17
	Aleyrodoidea	*Neomaskellia andropogonis*	G, S_1, S_2	K, G, P, C, Y, V	A, R, N, I	CO3, ND3	CO3, ND3	SrRNA	18
	Aleyrodoidea	*Trialeurodes vaporariorum*	·	G, Q (or I), C (or Y)	·	·	·	·	3
	Aleyrodoidea	*Tetraleurodes acaciae*	E, R, D, G	D, G, A, R, V, Q	N, S_1, I	CO3, ND3	CO3, ND3	·	17

To present the position of gene rearrangements, the abbreviations described below were used: A, $tRNA^{Ala}$; C, $tRNA^{Cys}$; D, $tRNA^{Asp}$; E, $tRNA^{Glu}$; G, $tRNA^{Gly}$; I, $tRNA^{Ile}$; K, $tRNA^{Lys}$; N, $tRNA^{Asn}$; P, $tRNA^{Pro}$; Q, $tRNA^{Gln}$; R, $tRNA^{Arg}$; S, $tRNA^{Ser}$; T, $tRNA^{Thr}$; V, $tRNA^{Val}$; Y, $tRNA^{Tyr}$; CO3, cytochrome c oxidase subunit III; ND3, NADH dehydrogenase subunit 3; srRNA, small subunit of ribosomal RNA gene; PCG, protein coding gene; rRNA, ribosomal RNA gene; tRNA, transfer RNA gene.

plementary Table 1). Interestingly, the ratio of GC/AT rich codons shows an association with suborders. For example, the ratio of GC/AT codons of Sternorrhyncha range from 0.16 to 0.33, whereas that of Auchenorrhyncha is 0.32 and those of Heteroptera are higher than 0.30 (Supplementary Table 1).

The AT-skew of *H. halys* was 0.128, within the range of -0.182 to 0.206 and near the average value of 0.054 seen for the 19 hemipteran mitochondrial genomes (Table 2). The AT-skew also shows an association with suborders, unlike the AT bias: the AT-skews of Sternorrhyncha are lower than 0.067 and that of Auchenorrhyncha is 0.063, whereas those of Heteroptera range from 0.091 to 0.206. The G content in *H. halys* is mostly underrepresented compared to C content, which is in line with the general trend in the mitochondrial genome towards lower G content (Lessinger et al., 2000). Comparing average G and C content in the 19 hemipteran genomes reveals interesting results: all six species of Aleyrodoidea in Sternorrhyncha, *A. aceris*, *A. dugesii*, *B. tabaci*, *T. acaciae*, *T. vaporariorum*, and *N. andropogonis*, show higher G content than C content, but the other six species possess higher C content than G content (Table 2).

Phylogenetic relationships of three suborders in Hemiptera

The phylogenetic structure of Hemiptera has been controversial at the suborder level. In our analysis of the nucleotide sequences from 13 PCGs, we confirmed the currently accepted theory that Sternorrhyncha and Auchenorrhyncha each constitute monophyletic groups (Carver et al., 1991; Hennig, 1981; Kristensen, 1973; Szelegiewicz, 1971), with Sternorrhyncha as a sister group to a group comprising Auchenorrhyncha and the Heteroptera (von Dohlen and Moran, 1995).

The use of PCG sequences of the mitochondrial genomes as a reliable molecular marker has become an informative strategy for inferring phylogenetic relationships (Boore et al., 2005). The mitochondrial genome sequence of *H. halys* provided us with additional information to address whether the combined 13 PCG sequences were suitable for resolving ambiguous suborder level phylogeny in Hemiptera.

Figure 3 shows the phylogenetic relationships among the 13 PCG sequences of 20 hemipteran and 1 thysanopteran species. The sequences used in this analysis represent three suborders of Hemiptera (Table 2). In the ML tree, the deepest splits supported a monophyletic Auchenorrhyncha, Heteroptera, and Sternorrhyncha, respectively. The phylogenetic analyses reveal possible alternative origins of Heteroptera with respect to Auchenorrhyncha with high confidence bootstrap value in ML tree (Fig. 3): Heteroptera and Auchenorrhyncha were not only sister taxa but also a monophyletic group. In this study, phy-

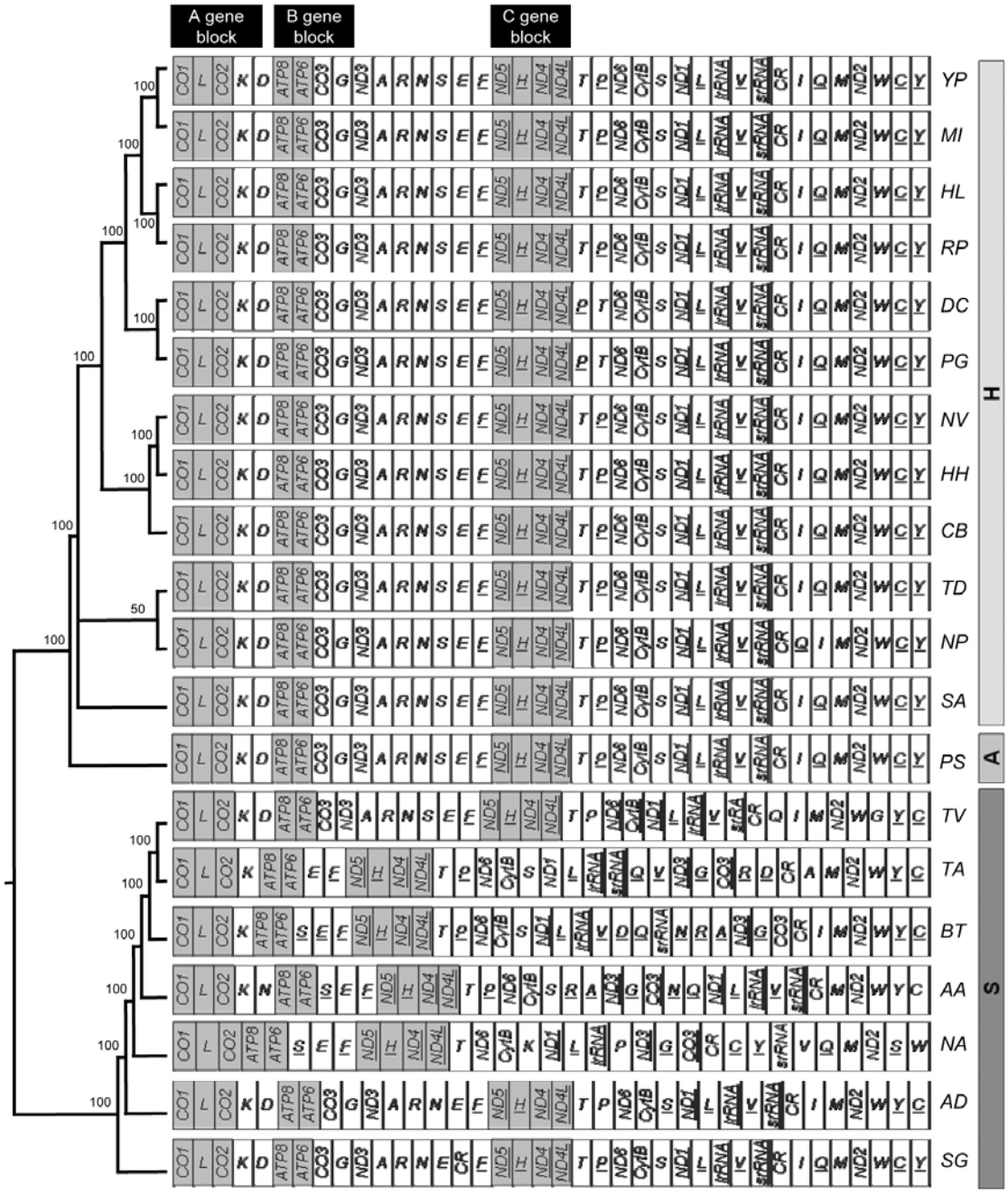

Fig. 4. Comparison between gene rearrangements of the 20 hemipteran mitochondrial genomes and the BI tree analyzed in this study. On the left side, the BI tree based on the 13 PCGs is shown. A, B, and C gene blocks show conserved gene arrangements within the 20 genomes (these genes are highlighted in graytone). Each gene of mitochondrial genome is shown following the abbreviations in Fig. 1. Along the right side, three suborders of Hemiptera and 20 species are indicated: A, Auchenorrhyncha; H, Heteroptera; S, Sternorrhyncha; AA, *Aleurochiton aceris*; AD, *Aleurodicus dugesii*; BT, *Bemisia tabaci*; CB, *Coptosoma bifaria*; DC, *Dysdercus cingulatus*; HH, *Halyomorpha halys*; HL, *Hydaropsis longirostris*; MI, *Malcus inconspicuous*; NA, *Neomaskellia andropogonis*; NP, *Neuroctenus parus*; NV, *Nezara viridula*; PG, *Physopelta gutta*; PS, *Philaenus spumarius*; RP, *Riptortus pedestris*; SA, *Saldula arsenjevi*; SG, *Schizaphis graminum*; TA, *Tetraleurodes acacia*; TD, *Triatoma dimidiate*; TV, *Trialeurodes vaporariorum*; YP, *Yemmalysus parallelus*.

logenetic relationships of the three suborders were broadly congruent, showing all same topologies about the three suborders. The relationships revealed for Hemiptera agree with previous molecular (von Dohlen and Moran, 1995) and morphological studies (Carver et al., 1991; Kukalová-Peck, 1991).

Gene rearrangement of mitochondrial genomes in Hemiptera

The order of genes in the *H. halys* mitochondrial genome is the same as in those of *C. bifaria*, *H. longirostris*, *M. inconspicuous*, *N. viridula*, *R. pedestris*, *P. venusta*, *S. arsenjevi T. dimidiata*,

and *Y. parallelus*, all belonging to the suborder Heteroptera. A comparison of gene orders and contents reveals that Hemiptera has three conserved gene blocks shared by all 20 species: A gene block, *CO1-tRNA^{Leu}-CO2*; B gene block, *ATP8-ATP6*; C gene block, *ND5-tRNA^{His}-ND4-ND4L* (Fig. 4), whereas the remaining mitochondrial genes show different gene rearrangements according to their suborders. Gene orders and contents of both Heteroptera and Auchenorrhyncha are mostly conserved, whereas those of Sternorrhyncha show extreme rearrangement (Table 5): 1) no gene rearrangement in Auchenorrhyncha, 2) only translocations of tRNAs in Heteroptera, 3) translocations and inversions of PCGs and/or tRNAs in Sternorrhyncha, and 4) the number of gene rearrangements in the three suborders increasing from Auchenorrhyncha and Heteroptera to Sternorrhyncha, indicating the gene orders and contents of Auchenorrhyncha are more similar to those of Heteroptera than Sternorrhyncha. This analysis confirmed that the evolutionary trend in Sternorrhyncha toward gene rearrangement is different from both Auchenorrhyncha and Heteroptera, and based on the extreme gene rearrangements, we were also able to ascertain the historical sequences of mitochondrial genes of Sternorrhyncha through the phylogenetic analyses.

CONCLUSION

We determined the mitochondrial genome of *H. halys* and compared it with 19 other published hemipteran mitochondrial genomes. Sequence analyses showed that the ranges of the distinctive AT-skew and the ratio of GC/AT-rich codons in Auchenorrhyncha and Heteroptera were different than in Sternorrhyncha. In phylogenetic analyses, both ML and BI trees supported the idea that both Auchenorrhyncha and Heteroptera constitute a monophyletic group. Finally, mitochondrial gene rearrangements indicate that Sternorrhyncha has different molecular signatures from Heteroptera and Auchenorrhyncha. Although the number of species with completely sequenced mitochondrial genomes is still inadequate to completely define the phylogenetic relationships within Hemiptera, the evolutionary trend within Sternorrhyncha was shown to be completely different from the other two suborders.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

We thank Dr. Iksoo Kim, Chonnam National University, for kindly advising many valuable comments on molecular analysis. This research was supported by the grants from Biogreen21 Project (2007030134032) and Agricultural R&D Promotion Center, and Specific Cooperative Agreement #58-1926-7-154F between the U.S. Department of Agriculture - Agricultural Research Service and Seoul National University.

REFERENCES

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573-580.

Boore, J.L. (1999) Animal mitochondrial genomes. Nucleic Acids Res. *27*, 1767-1780.

Boore, J.L., and Brown, W.M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. *8*, 668-674.

Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., and Brown, W.M. (1995). Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature *376*, 163-165.

Boore, J.L., Macey, J.R., and Medina, M. (2005). Sequencing and comparing whole mitochondrial genomes of animals. Meth. Enzymol. *395*, 311-348.

Carver, M., Gross, G., and Woodward, T. (1991). Hemiptera (bugs, leafhoppers, cicadas, aphids, scale insects etc.). In: CSIRO (ed) The insects of Australia. A textbook for students and research workers, Vol. I, 2nd ed. (Carlton: Melbourne University Press), p. 429.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540-552.

Cobben, R.H. (1978). Evolutionary trends in Heteroptera. Part II. Mouthpart-structures and feeding strategies (The Netherlands, Wageningen: Mededelingen Landbouwhogeschool).

Dotson, E.M., and Beard, C.B. (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. Insect Mol. Biol. *10*, 205-215.

Dowton, M. (1999). Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. Mol. Phylogenet. Evol. *11*, 283-287.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution *39*, 783-791

Folmer, O., Black, M., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxydase subunit I from diverse metazoan invertebrates. Mol. Marine Biol. Biotechnol. *3*, 294-299.

Goodchild, A. (1966). Evolution of the alimentary canal in the Hemiptera. Biol. Rev. *41*, 97-140.

Hamilton, K. (1981). Morphology and evolution of the rhynchotan head (Insecta: Hemiptera, Homoptera). Can. Entomol. *113*, 953-974.

Hennig, W. (1981). Insect phylogeny (New York: J Wiley and Sons)

Hoebeke, E.R., and Carter, M.E. (2003). *Halyomorpha halys* (Heteroptera: Pentatomidae): a polyphagous plant pest from Asia newly detected in North America. Proc. Entomol. Soc. Wash. *105*, 225-237.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., and Bu, W. (2008). Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics *9*, 610.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754-755.

Jermiin, L.S., and Crozier, R.H. (1994). The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. J. Mol. Evol. *38*, 282-294.

Kambhampati, S., and Smith, P.T. (1995). PCR primers for the amplification of four insect mitochondrial gene fragments. Insect Mol. Biol. *4*, 233-236.

Kim, K.-G., Hong, M.Y., Kim, M.J., Im, H.H., Kim, M.I., Bae, C.H., Seo, S.J., Lee, S.H., and Kim, I. (2009). Complete mitochondrial genome sequence of the yellow-spotted long-horned beetle *Psacothea hilaris* (Coleoptera: Cerambycidae) and phylogenetic analysis among coleopteran insects. Mol. Cells *27*, 429-441.

Kristensen, N. (1973). The phylogeny of hexapod "orders." A critical review of recent accounts. J. Zool. Syst. Evol. Res. *13*, 1-44.

Kristensen, N. (1981). Phylogeny of insect orders. Annu. Rev. Entomol. *26*, 135-157.

Kristensen, N.P. (1991). Phylogeny of extant hexapods, In: CSIRO (ed) The insects of Australia. A textbook for students and research workers, Vol. I, 2nd ed. (Carlton: Melbourne University Press), p. 125.

Kukalová-Peck, J. (1991). Fossil history and the evolution of hexapod structures. In: CSIRO (ed) The insects of Australia. A textbook for students and research workers, Vol. I, 2nd ed. (Carlton: Melbourne University Press), p. 141.

Lee, W., Park, J., Choi, J., Jung, K., Park, B., Kim, D., Lee, J., Ahn, K., Song, W., Kang, S., et al. (2009). IMGD: an integrated platform supporting comparative genomics and phylogenetics of insect mitochondrial genomes. BMC Genomics *10*, 148.

Lessinger, A.C., Martins Junqueira, A.C., Lemos, T.A., Kemper, E.L., da Silva, F.R., Vettore, A.L., Arruda, P., and Azeredo-Espin, A.M. (2000). The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). Insect Mol. Biol. *9*, 521-529.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955-964.

Muraji, M., and Tachikawa, S. (2000). Phylogenetic analysis of water striders (Hemiptera:Gerroidea) based on partial sequences of mitochondrial and nuclear ribosomal RNA genes. Entomol. Sci. *3*, 615-626.

Muraji, M., Kawasaki, K., and Shimizu, T. (2000). Phylogenetic utility of nucleotide sequences of mitochondrial 16S ribosomal RNA and cytochrome b genes in anthocorid bugs (Heteroptera: Anthocoridae). Appl. Entomol. Zool. (Jpn.) *35*, 293-300.

Nylander, J.A., Ronquist, F., Huelsenbeck, J.P., and Nieves-Aldrey, J.L. (2004). Bayesian phylogenetic analysis of combined data. Syst. Biol. *53*, 47-67.

Poisson, R., and Pesson, P. (1951). Super-ordre des Hémiptéroïdes (Hemiptera Linné, 1758, Rhynchota Burrueister, 1835), In: Grassé P-P (ed) Traité de Zoologic: Anatomic, Systématique, Biologic. Tome X. Insectes superieurs, et Hémiptéroïdes.2 vols (Masson, Paris), pp. 1-975, 976-1948

Posada, D., and Crandall, K.A. (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics *14*, 817-818.

Rawlings, T.A., Collins, T.M., and Bieler, R. (2001). A major mitochondrial gene rearrangement among closely related species. Mol. Biol. Evol. *18*, 1604-1609.

Rokas, A., and Holland, P.W. (2000). Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. *15*, 454-459.

Ross, H. (1965). A textbook of entomology, 3rd ed. (New York: Wiley).

Schuh, R.T., and Slater, J.A. (1995). True bugs of the world (Hemiptera, Heteroptera): Classification and Natural History. (New York: Cornell University Press).

Scudder, G. (1973). Recent advances in the higher systematics and phylogenetic concepts in entomology. Can. Entomol. *105*, 1251-1263.

Shao, R., and Barker, S.C. (2003). The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol. Biol. Evol. *20*, 362-370.

Stewart, J.B., and Beckenbach, A.T. (2005). Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome *48*, 46-54.

Swofford, D.L. (2002). PAUP*: Phylogenetic analysis using parsimony (*and other methods) Sunderland: Sinauer Associates.

Szelegiewicz, H. (1971). Autapomorphous wing characters in the recent subgroups of Sternorrhyncha (Hemiptera) and their significance in the interpretation of the Paleozoic members of the group. Ann. Zool. Fenn. *29*, 1-67.

Thao, M.L., Baumann, L., and Baumann, P. (2004). Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol. Biol. *4*, 25.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876-4882.

von Dohlen, C.D., and Moran, N.A. (1995). Molecular phylogeny of the Homoptera: a paraphyletic taxon. J. Mol. Evol. *41*, 211-223.

Wootton, R.J., and Betts, C.R. (1986). Homology and function in the wings of Heteroptera. Syst. Entomol. *11*, 389-400.