Comparative analysis of the mitochondrial genomes in *Drosophila virilis* species group (Diptera: Drosophilidae)

Boris Andrianov,[1] Irina Goryacheva,[1] Nikolai Mugue,[2] Svetlana Sorokina,[2] Tatyana Gorelova,[1] Vladimir Mitrofanov[2]

[1]Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia; [2]Kol'tsov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia

Abstract

We present comparative analysis of mitogenomics data of the *Drosophila virilis* group based on newly obtained complete mt-genome sequence of *Drosophila littoralis*, previously published mt-genome sequence of *D. virilis* and fragments of mt-genomes of other *Drosophila* species belonging to the *virilis* group. *Drosophila* of the *virilis* group shared most recent common ancestry 40 MYA with *D. melanogaster*. Therefore, presented data help to overview the evolution of the genus. The mt-genome of *D. littoralis* is a circular molecule of 16,017 bp with a total A+T content of 76.2%. The gene order is consistent with other *Drosophila* genomes. All tRNAs can be folded in the form of a typical clover-leaf structure except for *tRNASer(AGN)*. In the *control region* of *D. littoralis* we found four conserved sequence elements: 275 bp highly conserved sequence element, two thymidylate stretches and a G-island. The most variable genes in *Drosophila* of the *virilis* group are *nad6, nad3* and *nad4L*. The most conservative is *cox1*. We revealed long intergenic sequences' $(TA)_n$ separating *atp6* and *cox3* genes in the mitochondrial genomes of *Drosophila* of the *virilis* group. In other insect species these genes have no or few separating nucleotides. We detected fragments of mitochondrial genes *atp6* and *cox3* in the nuclear genome of *D. virilis*. These mitochondrial pseudogenes are marked by site-specific insertions of *Tv1* retrotransposon in the $(TA)_n$ intergenic spacer sequences.

Introduction

Insect mitochondrial genomes are circular DNA molecules about 16 kb, encoding a canonical set of 37 genes (13 inner membrane proteins, 2 ribosomal RNAs and 22 transfer RNAs); characteristically these genes have no introns.[1] Mitochondrial genome organization of *Drosophila yakuba* is taken as a standard for insects.[2] Regulatory sequences involved in the initiation of replication and transcription have been identified in the control region.[3-5] Both strands of mtDNA are transcribed and precursor RNAs are processed to produce mature RNAs for individual genes.[1] Mitochondrial DNA sequences are useful molecular markers, often explored for population genetic, phylogenetic and ecological studies of different animal species.[6-8] Mitogenomics data have been used to investigate the evolutionary history of insects and molecular processes that drive the evolution of the mitochondrial genome.[9-12] In this study we describe the newly determined mitochondrial genome of the *Drosophila littoralis* (Meigen, 1830). This is the first complete mitochondrial genome of this species. *D. littoralis* is one of the twelve closely related *Drosophila* species forming the *virilis* group.[13] This group was the object of investigation in the field of microevolution and speciation.[14-16] *Virilis* group is monophyletic and belongs to the subgenus *Drosophila*.[17] Divergence time from *D. melanogaster*, based on the set of several nuclear genes, is estimated to be 62.9 MYA.[18] However, based on the results of 12 *Drosophila* species complete genome analysis, divergence time between subgenera *Sophophora* and *Drosophila* is now placed in late Eocene (40 MYA).[19] This estimation is supported by mtDNA analysis.[20] It is believed that the origin of the *virilis* group is located in South-East Asia, because the most archaic species of replete-virilis section are limited to the South-East Asia region, and these species are absent in North America.[13] Species differentiation within the group is dated back to 11 MYA[13,14,16] when two major clades within the group, *virilis* and *Montana*, were formed. On the other hand, speciation in this group is still ongoing, and the youngest species within the group, *D. novamexicana* and *D. americana*, have divergence time estimated as 0.38 MYA.[21]

The *virilis* group was also the object of research elucidating genetic control of thermotolerance and heat-shock response.[22-24] We chose *D. littoralis* to determine complete mitochondrial DNA sequence for several reasons. *D. littoralis* is a widespread species. Natural populations of *Drosophila littoralis* inhabit the temperate and subtropical zones of Eurasia from Iran to Finland. This species is the most common representative of the *virilis* group in Europe. *D. littoralis* populations were characterized by chromosome inversion polymorphism[25,26] and an allosyme polymorphism.[27] Molecular phylogeographic studies based on mitochondrial haplotypes polymorphism of the *D. littoralis* have contributed to the understanding of the population dynamics of the natural populations of *Drosophila*.[28]

Correspondence: Boris Andrianov, Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Gubkin St.3, Moscow, Russia. E-mail: andrianovb@mail.ru

Key words: complete mitochondrial genome, retrotransposones, NUMTs, *virilis* group, *Drosophila*.

Contributions: all authors contributed equally to this work, read and approved the final manuscript. BA is responsible for the integrity of the work as a whole.

Conflict of interest: the authors report no conflicts of interest.

Acknowledgments: we would like to thank Dr. M. Evgen'ev (Institute of Molecular Biology, Russian Academy of Sciences) for critical reading of the manuscript and fruitful discussion. This research was supported by grants from the Russian Foundation for Basic Researches, Projects 08-04-00958 to VGM and 09-04-01720-a to IIG 08-04-01351-a to NSM. Grant of President of RF Leading scientific schools 4442.2010.4, Subprogram Genofonds and genetic diversity of the Program of Presidium of the RAS Biodiversity and State Contract N. 10002-251/P-24/154-150/200404-111 to VGM.

Received for publication: 27 December 2009.
Revision received: 28 June 2010.
Accepted for publication: 5 July 2010.

Trends in Evolutionary Biology 2010; 2:e4
doi:10.4081/eb.2010.e4

Mitochondrial DNA sequences are frequently transferred to the nucleus giving rise to the so-called nuclear mitochondrial DNA (NUMTs).[29] NUMTs are not equally abundant in all species. Copy number ranges from few copies in *Anopheles, Caenorhabditis, Plasmodium*, *Drosophila*, and *Fugu* to more than 500 in humans.[30] Mechanisms controlling accumulation and loss of NUMTs are unknown but are thought to be species-specific.[30] We detected DNA fragments of mitochondrial genes *atp6* and *cox3* in the nuclear genome of *D. virilis*. Mitogenomic data for *D. littoralis* will facilitate investigations of the evolutionary history of fruit flies.

Materials and Methods

Fly strains and cell culture

All strains used in this work are from the collection of the Laboratory of Genetics, Kol'tsov Institute of Developmental Biology, Russian Academy of Sciences and National *Drosophila* Species Resource Center (Bowling

Green, United States): *D. americana americana* (Spencer) 405, *D. americana texana* (Stone, Griffen and Patterson) 423, *D. ezoana* (Takada and Okada) 572, *D. kanekoi* (Watabe and Higuchi) 1540, *D. lacicola* (Patterson) 0991.13, *D. littoralis* (Meigen) 06-17a, *D. montana* (Patterson, Stone, and Griffen) 1021.13, *D. novamexicana* (Patterson) 424, *D. virilis* (Sturtevant) B9. Each strain was founded by a single female fertilized in nature and maintained as a mass culture in vials. In the laboratory, the strains were maintained on a standard *drosophila* medium. We determined the mitochondrial genome of *Drosophila littoralis* isofemale line 06-17a, collected in 2006 on the bank of the Don River (Rostov oblast, Russian Federation). The line is available in the Laboratory of Genetics, Kol'tsov Institute of Developmental Biology, Russian Academy of Sciences. In addition to isofemale lines of different fly species, the transplantable embryonic cell culture *79f7Dv3g* (*D. virilis*) was used in experiments for mitochondrial pseudogenes characterization. This cell line has diploid male karyotype.[31]

Isolation of total cellular DNA

Etherized flies from the isofemale strains examined were ground manually in a homogenizer in a lysing solution (2% SDS; 400 mM Tris-HCl, pH 8.0; 200 mM EDTA). The lysate was deproteinized with phenol (pH 8.0) and then with a phenol-chloroform mixture and DNA was sedimented by ethanol precipitation. The isolated total DNA was used as a template for PCR.

D. littoralis mitogenome amplification in overlapping PCR fragments

Initial rounds of amplification for genome sequencing were performed using the set of heterologous primers that we have developed based on *Drosophila yakuba* mitogenome sequence. The sequences of *D. littoralis* PCR fragments obtained at this initial step were used to design specific primers for *D. littoralis* that allowed us to amplify the entire mitogenome in overlapping PCR fragments. Information about primers is shown in the Table 1 (Supplementary). PCR products were directly sequenced after purification, with the exception of fragment 46. This fragment has a length of 1.3 kb and contains the entire Control region. Fragment 46 was cloned into pGEM-T Easy Vector (Promega) and sequenced.

PCR amplification, cloning, and sequencing

Primers used to amplify spacer sequence at the *atp6/cox3* junction and the adjacent parts of the genes of the eight *drosophila* species of the *virilis* group: forward - 5'- AAGGAACCCCAGCAATTCTT - 3' (primer: *Dvir4.1F*) and reverse - 5' - TGCTGGGGATAAACTTCTGTG - 3' (primer: *Dvir3.2R*).

Primers used to amplify the mitochondrial pseudogenes associated with the insertion of *Tv1* retrotransposon (*atp6-Tv1* association): forward - *Dvir4.1F* and reverse - 5'-CTTTATTGCCCAAAGGGTCA-3' (primer: *Dvir4.2R*); (*cox3-Tv1* association): forward - 5'-GGAAGGTTCTTGTGCGGATA-3' (primer: *Dvir4*.7F) and reverse - *Dvir3.2R*.

PCR was carried out in a 25 μL volume containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1 mM $MgCl_2$, 0.5 mM dNTPs, 0.3 mM of forward and reverse primers, and 1.0 units of Taq polymerase (Eurogene) using the Gene Amp® PCR System 2700 (Applied Biosystem) and applying the following thermo profile: initial denaturation at 95°C for 180 sec, following 35 cycles with denaturation at 95°C for 15 sec, primer annealing at 60°C for 15 sec, and primer extension at 72°C for 90 sec. The final elongation step was continued for 4 min at 72°C. PCR amplifications were performed on a Gene Amp® PCR System 2700. PCR products were visualized by 1.0% agarose gel electrophoresis. PCR fragments were purified using the JETguick Gel Extraction Spin Kit (GENOMED).

The resulting PCR fragments were ligated into the pGEM-T Easy Vector using the DNA ligation Kit and transformed into *E. coli* cells JM109 (Promega) using standard protocols. Each clone was sequenced on both strands. DNA sequencing was performed using the BigDye Terminator Cycle Sequencing Kit and the ABI 3730XL Genetic Analyzer (PE Applied Biosystems), according to the recommendations of the manufacturer. All fragments were sequenced from both strands.

Sequence assembly, gene identification and genome analysis

Sequence chromatograms were proof-read using the program CHROMAS available at *http://www.technelysium.com.au.* Sequence alignment, genome assemblage, and nucleotide composition statistics were carried out with Mega 4.[32] The locations of protein-coding genes and rRNA genes were identified by determining sequence similarity with other *Drosophila*. Identification of tRNA genes and prediction of tRNAs secondary structure was made using the ARVEN server.[33] Individual gene sequences were compared with the homologous sequences of other *Drosophila* species available in GenBank and inspected for the presence of gene overlaps, non-canonical start codons and truncated termination codons. Basic sequence statistics, codon usage and genetic distances among *Drosophila* genes were calculated using MEGA4.[32] Calculation of the ratio of Kn/Ks in 13 protein coding genes between *D. littoralis* and *D. virilis* were made on the basis of the number of synonymous substitutions per synonymous site and non-synonymous substitutions per non-synonymous site. Kn/Ks value of 1.0 indicates that substitutions in this gene are selectively neutral. The bias of the base composition of an individual strand was described by skewness[34] which is calculated using the formulas: AT-skew = (A%-T%)/(A%+T%) and CG-skew = (C%-G%)/(C%+G%). The presence of repeated sequences was studied using the Repeat Masker Web Server (*http://www.repeatmasker.org/cgibin/WEBRepeatMasker*).

Phylogenetic analysis

Phylogenetic analyses were conducted in MEGA 4.[32] The evolutionary history of the *virilis group* was inferred from DNA sequences of mitochondrial genes using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the units of the number of base substitutions per site.

Results and Discussion

Gene content and genome organization

The mitochondrial genome of *D. littoralis* is a closed circular molecule of 16,017 bp in length. It contains the set of 37 genes usually found in animal mitochondrial genomes: 22 transfer RNA genes, 13 protein coding genes, and 2 ribosomal RNA genes. Besides, there is one major non-coding region, *control region*, or the *A+T - rich region*, localized between the genes for *srRNA* and *tRNA*^{Ile} (Figure 1).

The gene order is identical to *Drosophila*

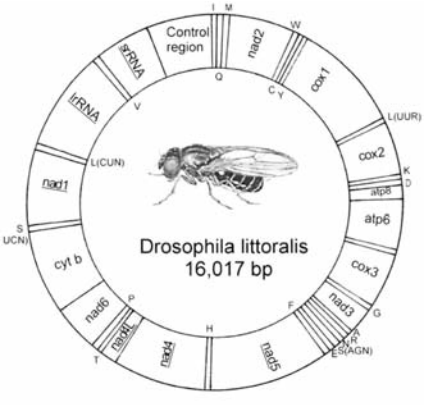

Figure 1. The mitochondrial genome organization of *D. littoralis* (GenBank ID: FJ447340). Genes for proteins and rRNAs are indicated with standard abbreviations, whereas those for tRNAs are designated by a single letter code for the corresponding amino acid. Genes oriented on the N-strand are underlined.

yakuba.[2] The majority of genes are located on the plus or J-strand, the remainder having opposite polarity and being oriented on the minus or N-strand (Figure 1 and Table 2). There are several short non-coding regions at the genes' junctions. The largest one 45 bp is located between *atp6* and *cox3*.

Protein coding genes

All protein-coding genes, except *cox1*, are found to have in-frame ATR methionine or ATY isoleucine codons as start signals. Seven start codons are coded by ATR: *cox2, atp6, cox3, nad4, nad4l, cytb*, and *nad1* and five by ATY: *nad2, atp8, nad3, nad5, nad6*. Recently, experimental data have cast doubt on previously accepted ideas about the starting codons for *Drosophila* genes *nad1* and *nad5*.[11] Mature mRNA of these genes turned out to be longer at the 5' ends than expected. *Nad1* mRNA includes three additional codons and so translation may be initiated from non-canonical codon UUG for leucine, *nad5* mRNA contains five additional codons. Therefore, translation may be initiated from non-canonical codon GUG for valine.[11] However, the results of mitochondrial mRNA sequencing do not exclude the possibility of translation initiation of these genes with the canonical start codons, so we annotate these genes in accordance with previously adopted concepts (Table 2).

The non-canonical start of translation of *cox1*, found in *Drosophila* and some other, but not all, insect species has been discussed.[35,36] It has been suggested that canonical start codons are added to the open reading frame by splicing.[2] Recent experimental studies on *Drosophila* have not confirmed this assumption. Sequences of the 5' ends of the mature *cox1* mRNA of *D. virilis*, *D. melanogaster*, *D. simulans, D.erecta* and *D. mojavensis* lack a canonical start codon. In all cases, the open reading frame begins with the codon (T/C)CG which thus serves as the only possible initiating codon.[11] This is also the case for *D. littoralis* (Table 2). Canonical TAA and TAG termination codons are found in six genes: *nad2, cox1, atp8, atp6, cox3, nad3*. The remaining seven have incomplete termination codons (T or TA) and their functionality is probably recovered after a posttranscriptional polyadenylation.[37] Experimental analyses of cDNA pools have demonstrated that genes *atp8/atp6* and *nad4L/nad4* - are recovered as bicistronic units in *Drosophila*[38] and in the dipteran *Anopheles funestus*.[39] *Atp8* and *atp6* overlap by seven nucleotides in almost all animal mitogenomes[39] and are, therefore, in different frames, while gene pair *nad4L/nad4* may be composed of a single in-frame coding unit (the two genes are separated by 6 nucleotides)[40] or may be in different frames, as in the case of *D. littoralis*. Translation of *nad4L* from bicistronic RNA may solve the problem of incomplete termination codon. In this case, the gene should be longer by one nucleotide than as annotated in Table 2.

Table 2. General characteristics of the mitochondrial genome of *Drosophila littoralis*.

Gene	Strand	Position From	Position To	Sise (bp)	Intergenic nucleotides	Anticodon	Start codon	Stop codon
tRNA^{Ile}	J	1	65	65	30	GAU		
tRNA^{Gln}	N	96	164	69	0	UUG		
tRNA^{Met}	J	164	232	69	0	CAU		
nad2	J	233	1258	1026	4		ATT	TAG
tRNA^{Trp}	J	1263	1329	67	-8	UCA		
tRNA^{Cys}	N	1322	1384	63	24	GCA		
tRNA^{Tyr}	N	1409	1473	65	-2	GUA		
cox1	J	1472	3007	1536	2		TCG	TAA
tRNA$^{Leu(UUR)}$	J	3010	3076	67	5	UAA		
cox2	J	3082	3769	688	0		ATG	T
tRNA^{Lys}	J	3770	3840	71	17	CUU		
tRNA^{Asp}	J	3858	3927	70	0	GUC		
atp8	J	3928	4089	162	-7		ATC	TAA
atp6	J	4083	4757	675	45		ATG	TAA
cox3	J	4803	5591	789	12		ATG	TAA
tRNA^{Gly}	J	5604	668	65	0	UCC		
nad3	J	5669	6022	354	8		ATT	TAA
tRNA^{Ala}	N	6031	6095	65	39	UGC		
tRNA^{Arg}	J	6135	6197	63	0	UCG		
tRNA^{Asn}	J	6198	6263	66	0	GUU		
tRNA$^{Ser(AGN)}$	J	6264	6331	68	0	GCU		
tRNA^{Glu}	J	6332	6401	70	18	UUC		
tRNA^{Phe}	N	6420	6485	66	0	GAA		
nad5	N	6486	8205	1720	15		ATT	T
tRNA^{His}	N	8221	8285	65	0	GUG		
nad4	N	8286	9624	1339	0		ATG	T
nad4l	N	9625	9914	290	2		ATG	TA
tRNA^{Thr}	J	9917	9981	65	0	UGU		
tRNA^{Pro}	N	9982	10046	65	2	UGG		
nad6	J	10049	10572	524	0		ATT	TA
Cytb	J	10573	11707	1135	0		ATG	T
tRNA$^{Ser(UCN)}$	J	11708	11774	67	15	UGA		
nad1	N	11800	12736	937	10		ATA	T
tRNALeu$^{(CUN)}$	N	12747	12811	65	2	UAG		
lrRNA	N	12814	14137	1324	0			
tRNA^{Val}	N	14138	14209	72	0	UAC		
srRNA	N	14210	14994	785	0			
Control region		14995	16017	1023				

Transfer RNA genes

All the 22 tRNA genes typically found in metazoan mtDNAs were identified according to their secondary structure and primary sequence of the corresponding anticodon (Figure 2A and B). The anticodons of the *D. littoralis* tRNAs are identical to those in *Drosophila yakuba*.[2] All tRNAs have typical clover-leaf structure except *tRNA^{Ser}* (*AGN*). This tRNA has a simple DHU loop lacking an arm. Abnormal structure does not influence its function because codon (AGN) recognized by this tRNA is widely used in *D. littoralis* mitochondrial genes (Table 3). The anomalous structure of this tRNA is conserved in other insect mitochondrial genomes.[11] Several tRNA genes in *D. littoralis* genome have few mismatches: *tRNA^{Lys}, tRNA^{Arg}, tRNA^{Ala}, tRNA^{Tyr}*. Correct base pairing may be restored posttranscriptionally with an RNA-editing mechanism.[41] Overlapping was observed in two cases: between *tRNA^{Trp}* and *tRNA^{Cys}*, and between

tRNA^{Tyr} and *cox1* (Table 2). This is also the case in other *Drosophila*.[2]

Non-coding regions

Control region is the only major non-coding region in the mitochondrial genome of *Drosophila*. It contains replication origins and promoters for both strands of the mitochondrial genome.[42] Within the genus *Drosophila*, there are two groups of species, with short and with long control regions.[43-45] *D. littoralis* belongs to the group with a short *control region*. The *control region* of *D. littoralis* has a length of 1023 bp and an A+T content of 90.1%. Comparison of the *D. littoralis* and *D. virilis control regions* shows the presence of conservative domains and characteristic differences in the rates of evolution between different segments of the *control region* (Figure 3). Four DNA sequence elements are found to be highly conserved in *Drosophila control regions*. These include about 300-bp element in the central part of the region, two thymidylate stretches on opposite DNA strands and a G island. A key role in replication has been suggested for T stretches identified on opposite DNA strands,[42] one near the gene for *tRNA*^{Ile} ranging in length from 11 to 17 bp, the other on the opposite DNA strand with a mean length of 13 to 23 bp in the central part of the *control region*.[4] It is shown experimentally, that these poly T blocks are the origins of replication for both strands of the mitochondrial genome of four *Drosophila* species (*D. yakuba, D. obscura, D. albomicans,* and *D. virilis*).[4] In the central part of the control region, between the two poly T stretches, comparative analyses reveal a highly conserved sequence element (HCSE) of about 300 bp.[46,47] This region is easily identifiable in *D. littoralis* control region. The value of P distance calculating for the entire control regions of *D. littoralis* and *D. virilis* is 0.15±0.01. This is three times higher than that calculated only for the HCSE of these species (P=0.05±0.01). Nucleotide sequence variation between HCSE elements of *D. littoralis* and *D. virilis* is about the same as the variability of the protein coding genes of these species. The stretch of four Gs which is thought to be a part of putative replication termination signal for the N strand,[43,48] was found at the expected place, near the srRNA gene in the control region of *D. littoralis*. Animal mitochondrial genomes are very compact, with a high proportion of coding versus non-coding sequences. Intergenic spacers are usually limited in number and size, and their occurrence is believed to be the result of errors in the mtDNA replication system. Point mutations or duplications in mitochondrial genomes originate apparently due to slipped-strand mispairing.[49] Apart from the *control region*, non-coding sequences of *D. littoralis* mtDNA are 250 bp in sum and represent the most variable part of

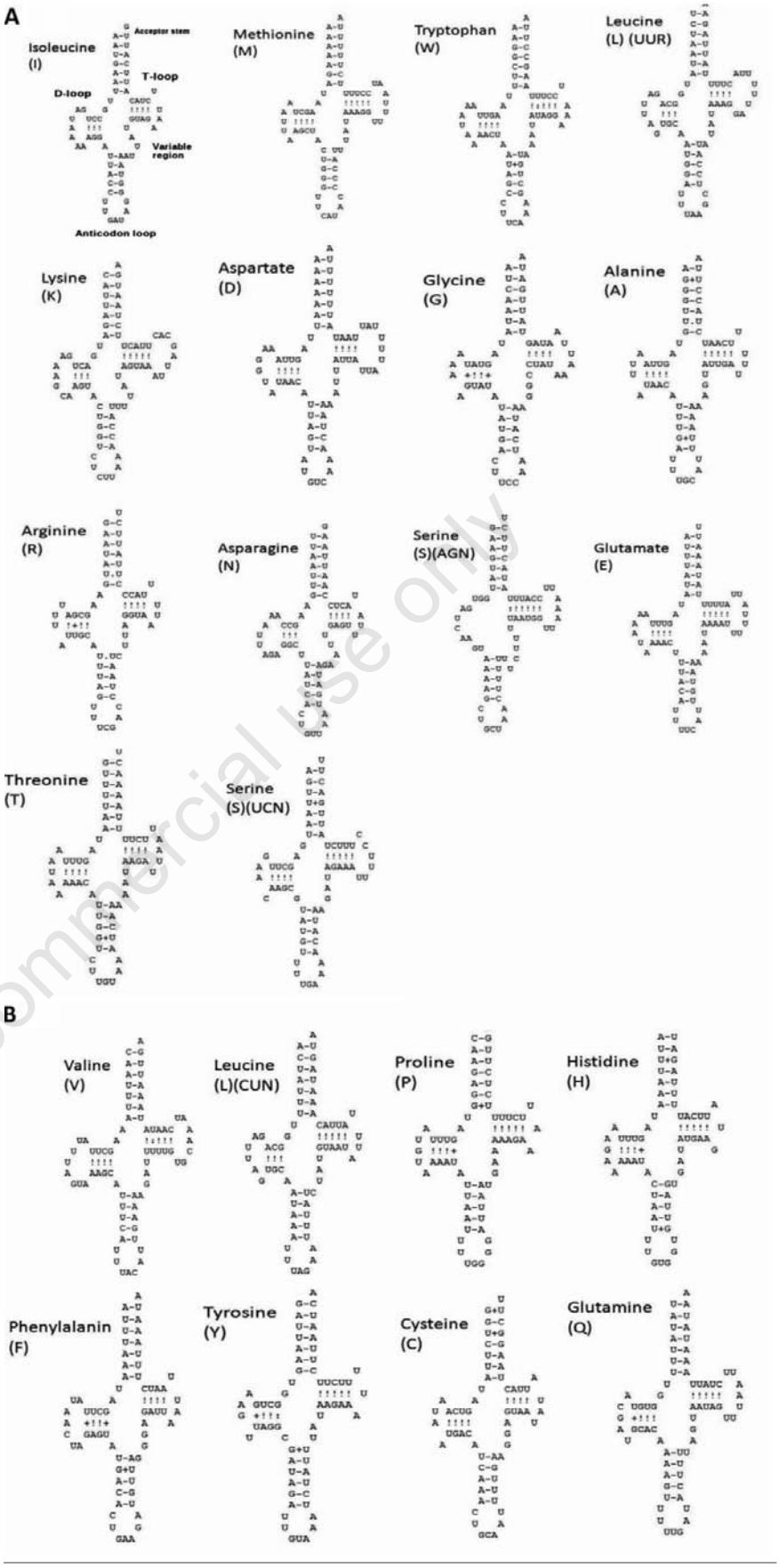

Figure 2. Putative secondary structure folds for the tRNAs of *D. littoralis* mt-genome. Watson-Crick base pairs designated by "–" or "!" and G–T base pairs by "+". (A) Majority coding strand tRNAs. (B) Minority coding strand tRNAs.

the genome. Within non-coding regions, three $(TA)_n$ microsatellites can be observed, at the genes' junctions: *tRNA^{lys}/tRNA^{Asp}*, *tRNA^{Ala}/tRNA^{Arg}*, *atp6/cox3*. All these regions are highly variable in *Drosophila* mitogenomes.

Base composition and codon usage

A remarkable molecular feature of mtDNAs is the asymmetry in the composition of the nucleotide content between the two strands.[34] Usually, in insect mitochondrial genomes A% and C% are higher than T% and G% on the J-strand.[40,50] Asymmetry in nucleotide composition among strands may be due to the mitochondrial DNA asynchronous replication {Bogenhagen, 2003 #12; Reyes, 1998 #40. The bias of the base composition of an individual strand can be described by skewness.[34] We observed strong CG-skew (0.181) in the J strand of *D. littoralis* (Table 3). In the J-strand, cytosine always occurs more frequently than the guanine. This is true for protein and RNA coding genes and non-coding regions, but the value of CG-skew varies greatly. In the control region, HCSE has very low CG-skew, while hyper variable fragment of the control region has extremely high CG-skew (Table 3). All protein coding genes have distinctly negative AT-skew, while the value of AT-skew for RNA coding genes is near zero. It is clear, that the codon usage preference of *D. littoralis* drives AT-skew to negative values. Codons for leucine (UUA), isoleucine (AUU) and phenylalanine (UUU) are the most frequently used in *D. littoralis* mitogenome, accounting for 14.2%, 9.6%, 8.9%, respectively, of the total number of codons (Table 4). Sequences of these codons have negative AT-skew. Codon usage, in turn, may be influenced by selection for efficiency and accuracy of translation.[51]

Nucleotide variability of *D. littoralis*

Comparison of nucleotide variability of individual genes from closely related species reveals conservative and polymorphic genes. These data provide the basis for conclusions of types of natural selection in the resent evolution of the species. The nucleotide variability of each mitochondrial gene has been estimated by calculating the ratio of Kn/Ks in all 13 protein-coding genes between *D. littoralis* and *D. virilis* (GenBank ID: BK006340). (Ks) - is the number of synonymous substitutions per synonymous site, (Kn) - is the number of non-synonymous substitutions per non-synonymous site. Kn/Ks value of 1.0 indicates that substitutions in this gene are selectively neutral. Comparisons with other *Drosophila* is less informative due to multiple reverse mutations at variable sites accumulated during long evolutionary periods. The ratio of Kn/Ks in all 13 protein-coding genes between *D. littoralis* and *D. virilis* is much less than one, which indicates a strong stabilizing

Table 3. Nucleotide composition and skews of *Drosophila littoralis* mitochondrial genes.

	Percentage of nucleotides				%(A+T)	AT skew	CG skew
	A	C	G	T			
nad2	0.345	0.123	0.090	0.441	0.786	-0.122	0.155
cox1	0.288	0.167	0.162	0.383	0.670	- 0.140	0.015
cox2	0.330	0.148	0.134	0.387	0.717	- 0.079	0.050
atp8	0.370	0.136	0.056	0.437	0.807	- 0.083	0.417
atp6	0.314	0.179	0.108	0.398	0.712	- 0.118	0.243
cox3	0.300	0.170	0.146	0.383	0.683	- 0.122	0.076
nad3	0.314	0.127	0.088	0.471	0.785	- 0.200	0.181
nad5	0.308	0.089	0.160	0.443	0.751	- 0.180	- 0.285
nad4	0.308	0.084	0.147	0.460	0.768	- 0.198	- 0.273
nad4L	0.306	0.069	0.124	0.500	0.806	- 0.241	- 0.285
nad6	0.366	0.126	0.052	0.456	0.822	- 0.109	0.416
cob	0.310	0.169	0.130	0.390	0.700	- 0.114	0.130
nad1	0.274	0.089	0.153	0.483	0.757	- 0.276	- 0.264
lrRNA	0.403	0.063	0.116	0.417	0.820	- 0.017	- 0.296
srRNA	0.399	0.080	0.134	0.387	0.786	0.015	- 0.252
J strand tRNA genes	0.368	0.113	0.127	0.392	0.760	- 0.032	- 0.057
N strand tRNA genes	0.374	0.081	0.155	0.390	0.764	- 0.021	- 0.314
HCSE of the Control Region	0.440	0.058	0.051	0.451	0.891	-0.012	0.064
Hyper variable fragment of the	0.503	0.072	0.016	0.408	0.911	0.104	0.636
Control Region Intergenic non coding regions	0.448	0.084	0.048	0.420	0.868	0.032	0.273
Total J strand	0.386	0.140	0.097	0.376	0.762	0.013	0.181

Figure 3. Alignment of the nucleotide sequences of the CR of *D. littoralis* and *D. virilis* (GenBank ID: X05914.1). A dot indicates a nucleotide that is the same as that in *D. littoralis*. A dash indicates a nucleotide that is absent. A letter indicates a substitution. Highly conserved sequence element is underlined. Conservative domains are in upper case letters.

selection (Figure 4). It is clear that nad6, nad3 and nad4L are the most variable genes in this *Drosophila* species. The most conservative *D. littoralis* gene is *cox1*.

Nucleotide variability of mitochondrial protein coding genes in the *D. melanogaster* species group revealed similar results.[52] *Nad3* and *nad4L* are more conservative in the *D. melanogaster* species group than in the *D. virilis* species group, while nad6 is the most variable gene in both *Drosophila* groups.

Sequence variation in *atp6/cox3* intergenic spacer in *Drosophila* of the *virilis* group

Intergenic spacer sequences are the most variable part of mitogenome. A spacer of 45 nucleotides separating *atp6* and *cox3* is the longest one in *D. littoralis*. To characterize species-specific variability in the *Drosophila virilis* group, we determined nucleotide sequences of this region for nine *Drosophila* of this group together with the adjacent parts of *atp6* and *cox3* (Table 5). In all analyzed cases intergenic sequence can be folded in the form of a hairpin (Figure 5). We observed the minimal length of this hairpin in *D. kanekoi*. This allows identification of a minimal or basic element of the hairpin. Other *Drosophila* have longer stems of the hairpin due to the addition of several (TA) dinucleotides, sometimes with a few mismatches. This hairpin, of variable length, is a specific feature of the *virilis* group. In other insects and *Drosophila*, *atp6* and *cox3* have no few separating nucleotides or abut directly. This observation drew our attention to the possibility of exploiting $(TA)_n$ microsatellites in the mitogenome of *Drosophila virilis* group to find and characterize mitochondrial pseudogenes.

Mitochondrial DNA in the nucleus

Gene transfer from mitochondria to nuclear genomes is detected in many species.[30] These mitochondrial pseudogene (NUMTs) sequences may accumulate in genome regions with low recombination.[53] Although the molecular mechanism of NUMTs integration has not been revealed, NUMTs are often associated with transposones.[54] Previously we described and characterized transpositionally active retrotransposon, *Tv1* in the *Drosophila virilis* group.[55] According to the recent classification, *Tv1* is a member of *errantiviruses*.[56] *Errantiviruses* are usually site specific and duplicate four nucleotides at the site of insertion. In the case of retrotransposon *gypsy*, there is experimental evidence that Integrase encoded by this retrotransposon is capable not only to insert, but also precisely excise *gypsy*, with the original nucleotide sequence of the target site being completely restored.[57] *Tv1* chromosome copies are always flanked by the octanu-

Table 4. Codon usage in 13 protein genes of *Drosophila littoralis* mitochondrial DNA (total 3,724 codons).

Amino acid corresponding to one of the mitochondrial tRNAs	%	Codon	N	Amino acid corresponding to one of the mitochondrial tRNAs	%	Codon	N
Ala	4.6	GCA	47	Lys	2.3	AAA	74
	GCC	22			AAG	10	
	GCG	12	Met	5.9	AUA	193	
	GCU	92			AUG	27	
Arg	1.6	CGA	35	Phe	8.9	UUC	27
	CGC	1			UUU	303	
	CGG	4	Pro	3.5	CCA	33	
	CGU	19			CCC	17	
Asn	5.5	AAC	34			CCG	9
	AAU	170			CCU	73	
Asp	1.8	GAC	15	Ser (AGN)	2.7	AGA	68
	GAU	52			AGC	9	
Cys	1.2	UGC	2			AGG	1
	UGU	41			AGU	22	
Gln	2.0	CAA	72	Ser (UCN)	6.2	UCA	102
	CAG	3			UCC	13	
Glu	2.1	GAA	70			UCG	10
	GAG	7			UCU	105	
Gly	6.0	GGA	105	Thr	5.0	ACA	87
	GGC	5			ACC	21	
	GGG	52			ACG	3	
	GGU	63			ACU	76	
His	2.1	CAC	25	Trp	2.7	UGA	89
	CAU	52			UGG	12	
Ile	9.6	AUC	41	Tyr	4.5	UAC	43
	AUU	316			UAU	124	
Leu (CUN)	2.1	CUA	33	Val	5.5	GUA	79
	CUC	3			GUC	10	
	CUG	3			GUG	11	
	CUU	41			GUU	102	
Leu (UUR)	14.2	UUA	476				
	UUG	50					

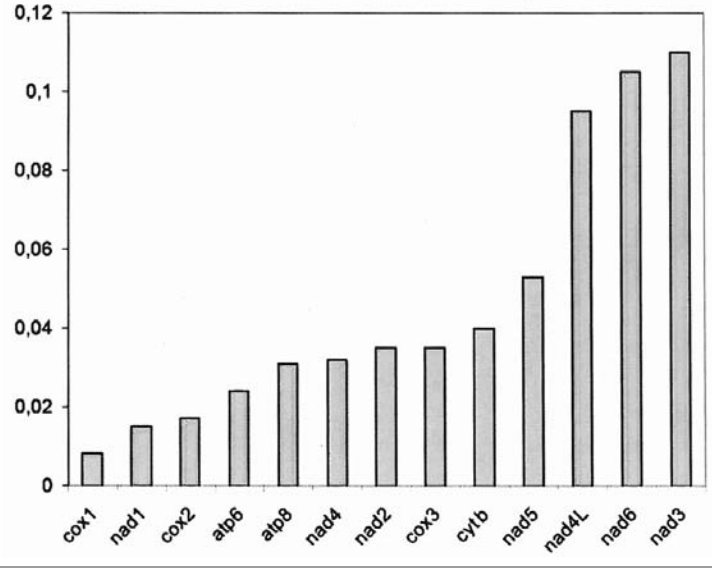

Figure 4. The ratio of non-synonymous to synonymous substitutions (Kn/Ks) of all 13 mitochondrial protein coding genes of *D. littoralis* and *D.virilis*.

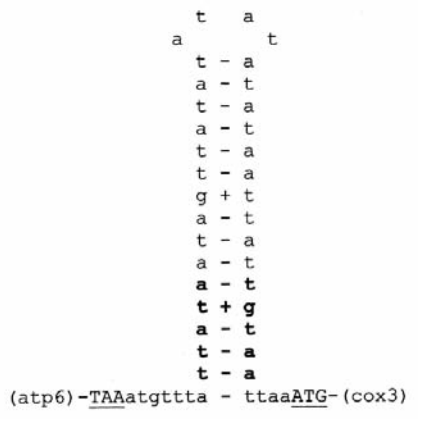

Figure 5. *D. littoralis* hairpin structure in the spacer sequence at the atp6/cox3 junction. Basic element of the hairpin is in bold. Stop and start codons of adjacent genes are underlined. Watson-Crick base pairs designated by "–" and G–T base pairs by "+".

cleotide - 5' ATATATAT 3', thus $(AT)_n$ microsatellite is a putative *Tv1* integration site. We proposed that *Tv1* insertions in $(TA)_n$ sites of NUMTs will tag them and allow them to be identified via a simple PCR technique. To check this hypothesis, we developed two pairs of primers to the expected sequence of a NUMT (*atp6/cox3*) with inserted *Tv1* in direct and reverse orientation (see Materials and Methods). In silico PCR with these primers and *D. virilis* genome (*UCSC web server: http://genome.ucsc.edu/index.html*) match no results. The results of an experimental check (Figure 6) allow detection of positive signals only in the case of males but not in females of *D. virilis*. Cloning positive bands into pGEM-T Easy Vector (Promega) and sequencing individual clones confirmed their "chimerical" *Tv1*-mitchondrial nature. These results allow mapping of identified *atp6/cox3* NUMT to the Y chromosome of *D.virilis*.

Positive PCR results for both primer pairs were detected also with the *D. virilis* cell culture 79f7Dv3g (*data not shown*). This cell line has male karyotype.[31] Negative results of these PCR test for females of *D. virilis* do not exclude the possibility that *atp6/cox3* NUMTs are present in their genome but they are not marked by *Tv1* insertions. Negative results for *D. littoralis* may be due to non-specificity of the used primers. To map the exact sites of *Tv1* insertions in the NUMT sequence, we exclude *Tv1* parts from the sequences of "chimerical" PCR fragments and align the resulted sequences with the mitochondrial sequence of *atp6/cox3* junction determined for the same *D. virilis* flies (line B9). In all these different NUMTs sequences, *Tv1* insert at the *atp6/cox3* gene junction and in the microsatellite region, exactly after the sequence 5' ATATATAT 3'

Table 5. Length variation of $(TA)_n$ microsatellite in the spacer sequence at the *atp6/cox3* junction.

Drosophila species	Basic element of the hairpin	Species specific length of the hairpin (N. of TA pairs of the hairpin)	GenBank ID: of the sequence of the corresponding PCR fragment
Drosophila kanekoi	5'-ATATT -3'	2	FJ536197
Drosophila a. americana	5'-TTATT -3'	7	FJ536199
Drosophila novamexicana	5'- TTATT -3'	9	FJ536203
Drosophila littoralis	5'- TTATA -3'	10	FJ536201
Drosophila virilis	5'- TTATA -3'	11	FJ536196
Drosophila a. texana	5'- TTATT -3'	11	FJ536204
Drosophila ezoana	5'- TTATA -3'	12	FJ536198
Drosophila montana	5'- TTATT -3'	17	FJ536202
Drosophila lacicola	5'- TTATT -3'	25	FJ536200

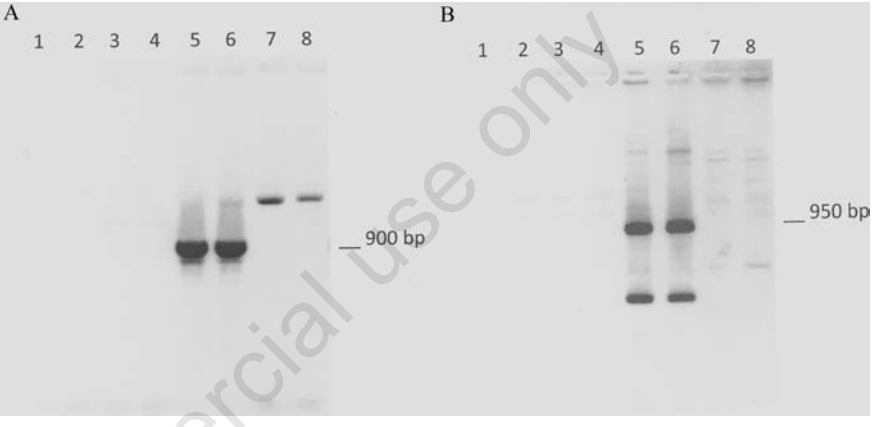

Figure 6. PCR identification of NUMT sequences. (A) PCR identification of NUMT (*atp6*) of *D. virilis* associated with *Tv1* retrotransposon. (B) PCR identification of NUMT (*cox3*) of *D. virilis* associated with *Tv1* retrotransposon. Each line shows the analysis of an individual fly. Bar indicates mobility of PCR fragment of expected size. Lines 1, 2 – males *D. littoralis*; 3, 4 – females *D. littoralis*; 5, 6 - males *D. virilis*; 7, 8 - females *D. virilis*.

Figure 7. Alignment of *D. virilis* NUMTs at the junction of *atp6/cox3* genes. *D. virilis* – mitochondrial sequence of fly line B9 (GenBank ID: FJ536196), *D. virilis* NUMT atp6 – mitochondrial pseudogene of fly line B9 (GenBank ID: FJ536205), *Cvi* NUMT atp6 - mitochondrial pseudogene of cell line *79f7Dv3g* (GenBank ID: FJ536206), *Cvi* NUMT *cox3* - mitochondrial pseudogene of cell line *79f7Dv3g* (GenBank ID: FJ539165), Translation initiation codons for *cox3* gene and termination codons for *atp6* gene are underlined, '-' represents inferred gaps.

(Figure 7). Detection of *atp6/cox3* NUMTs in the *D. virilis* genome raises the question of the time of their arrival and the frequency of this process. To answer these questions, we performed phylogenetic analysis of *atp6* and *cox3* in *Drosophila* of the *virilis* group, including sequences of mitochondrial pseudogenes (Figure 8).

Although phylograms were constructed using only short mitochondrial fragments, they are in good agreement with known phylogenetic relations of *Drosophila* for this group.[13] All identified NUMT sequences clustered with *D. virilis* original mitochondrial sequence indicating their recent origin. The case with *atp6* is especially informative. NUMT from culture cells are closer to B9 flies mitochondrial *atp6* than NUMT from the genome of the B9 flies. This demonstrates that NUMT generation is a frequent ongoing process at least in the genome of *D. virilis* culture cells. This may correlate with the elevated activity of retrotrans-

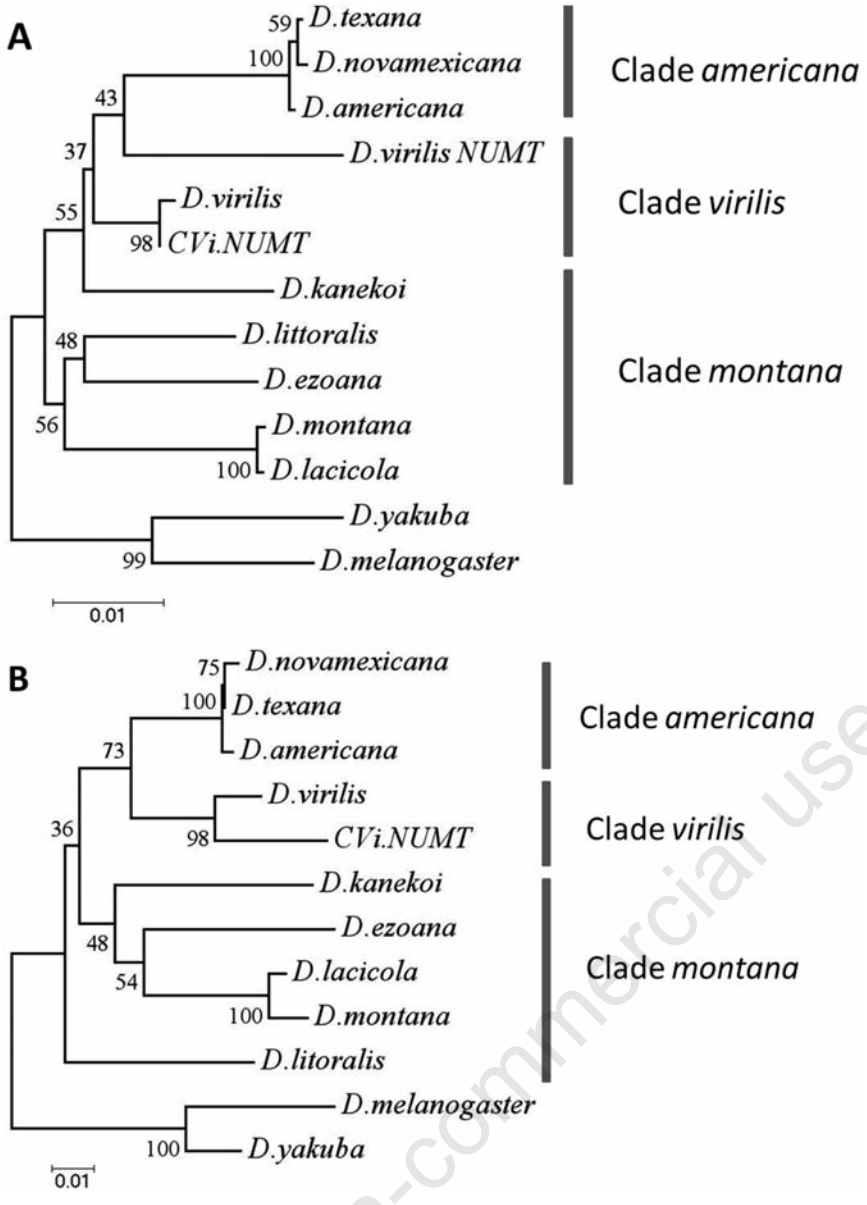

Figure 8. The evolutionary history of *Drosophila* of the *virilis* group inferred from DNA sequences of mitochondrial genes using the Neighbor-Joining method. The phylogenetic tree was constructed in MEGA 4. *D. melanogaster* and *D. yakuba* were chosen as an outlier group. (A) Phylogenetic analysis of partial sequences of atp6 and NUMTs (*atp6*). Fragment length of 261 nucleotides in total (nucleotide numbers: 4478-4738) of the complete *D. virilis* genome sequence - (GenBank ID: BK006340). (B) Phylogenetic analysis of partial sequences of *cox3* and NUMT (*cox3*). Fragment length of 301 nucleotides in total (nucleotide numbers: 4791-5091) of the complete *D. virilis* genome sequence. Sequence data comes from PCR fragments of nine *Drosophila* species of the virilis group and *D. virilis* cell culture. For GenBank ID see Table 5 and Figure 7 legend.

posons in culture cells. We had previously found a 10-fold amplification of *Tv1* copy numbers in this cell line.[55]

Conclusions

The description and analysis of the complete mtDNA genome sequence of *D. littoralis* has provided new insights into the mitogenomic evolution of *Drosophila*. Unusual for insects, a non-coding region of variable length was observed at the site of the *atp6/cox3* junction in *Drosophila* of the *virilis* group. Such intergenic spacers may contain regulatory signals involved in the transcription and processing of the mitochondrial transcripts, although additional data will be needed to clarify their function. Segments of these intergenic regions can be folded in typical stem-loop structures due to (TA) dinucleotides expansion. This suggests that the gene junction *atp6/cox3* may represent "hot spot" for mutations in the mitogenome of *Drosophila* of the *virilis* group. Another interesting feature, observed in the case of *Drosophila virilis*, is an ongoing process of gene transfer from mitochondria to the nuclear genome. It is not clear whether or not these two phenomena are somehow linked by a common molecular mechanism based on transpositions of *Tv1* retrotransposon. Alternatively, newly occurring NUMTs are simply effective targets of retrotransposon integration.

References

1. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res 1999;27:1767-80.
2. Clary DO, Wolstenholme DR. The mitochondrial DNA molecular of Drosophila yakuba: nucleotide sequence, gene organization, and genetic code. J Mol Evol 1985; 22:252-71.
3. Arunkumar KP, Nagaraju J. Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. PLoS One 2006;1:e110.[PubMed] 4. Saito S, Tamura K, Aotsuka T. Replication origin of mitochondrial DNA in insects. Genetics 2005;171:1695-705.
5. Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta 1999; 1410:103-23.
6. Boore JL. Complete mitochondrial genome sequence of Urechis caupo, a representative of the phylum Echiura. BMC Genomics 2004;5:67.
7. Nardi F, Spinsanti G, Boore JL, et al. Hexapod origins: monophyletic or paraphyletic? Science 2003;299:1887-9.
8. Salvato P, Simonato M, Battisti A, Negrisolo E. The complete mitochondrial genome of the bag-shelter moth Ochrogaster lunifer (Lepidoptera, Notodontidae). BMC Genomics 2008;9: 331.
9. Carapelli A, Lio P, Nardi F, et al. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol Biol 2007;7 Suppl 2:S8.
10. Cook CE, Yue Q, Akam M. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proc Biol Sci 2005;272:1295-304.
11. Montooth KL, Abt DN, Hofmann JW, Rand DM. Comparative genomics of Drosophila mtDNA: Novel features of conservation

and change across functional domains and lineages. J Mol Evol 2009;69:94-114.
12. Song N, Liang A. The complete mitochondrial genome sequence of Geisha distinctissima (Hemiptera: Flatidae) and comparison with other hemipteran insects. Acta Biochim Biophys Sin (Shanghai) 2009;41:206-16.
13. Throckmorton LH. The virilis species group. In: Ashburner M, Carson HL and Thompson JN (eds). The genetics and biology of Drosophila Academic Press, London 1982;3b:227-96.
14. Nurminsky DI, Moriyama EN, Lozovskaya ER, Hartl DL. Molecular phylogeny and genome evolution in the Drosophila virilis species group: duplications of the alcohol dehydrogenase gene. Mol Biol Evol 1996;13:132-49.
15. Orsini L, Huttunen S, Schlötterer C. A multilocus microsatellite phylogeny of the Drosophila virilis group. Heredity 2004;93:161-5.
16. Spicer GS, Bell CD. Molecular Phylogeny of the Drosophila virilis Species Group (Diptera: Drosophilidae) Inferred from Mitochondrial 12S and 16S Ribosomal RNA Genes. Annals of the Entomological Society of America 2002;95:156-61.
17. van der Linde K, Houle D. A supertree analysis and literature review of the genus Drosophila and closely related genera (Diptera, Drosophilidae). Insect Systematics & Evolution 2008;39:241-67.
18. Tamura K, Subramanian S, Kumar S. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol 2004;21:36-44.
19. Bhutkar A, Russo SM, Smith TF, Gelbart WM. Genome-scale analysis of positionally relocated genes. Genome Res 2007;17:1880-7.
20. Wang BC, Park J, Watabe HA, et al. Molecular phylogeny of the Drosophila virilis section (Diptera: Drosophilidae) based on mitochondrial and nuclear sequences. Mol Phylogenet Evol 2006;40:484-500.
21. Spicer GS. Molecular evolution and phylogeny of the Drosophila virilis species group as inferred by two-dimensional electrophoresis. J Mol Evol 1991;33:379-94.
22. Evgen'ev MB, Zatsepina OG, Garbuz D, et al. Evolution and arrangement of the hsp70 gene cluster in two closely related species of the virilis group of Drosophila. Chromosoma 2004;113:223-32.
23. Garbuz D, Evgenev MB, Feder ME, Zatsepina OG. Evolution of thermotolerance and the heat-shock response: evidence from inter/intraspecific comparison and interspecific hybridization in the virilis species group of Drosophila. I. Thermal phenotype. J Exp Biol 2003;206:2399-408.
24. Garbuz DG, Iushenova IA, Evgen'ev MB, Zatsepina OG. [Comparative analysis of hsp70 gene cluster in Drosophila virilis species group]. Mol Biol (Mosk) 2009;43:44-52.
25. Mitrofanov VG, Poluectova EV. Inversion Polymorphism in the natural populations of Drosophila littoralis Sokolov (Drosophila littoralis Meig). Genetika (in Russian) 1982;8:1849-55.
26. Poluectova EV, Mitrofanov VG, Korochkin LI, et al. Influence of karyotype polymorphism on viability in Drosophila imeretensis Sokolov (Drosophila littoralis Meig). Genetika (in Russian) 1984;20:564-9.
27. Goncharenko GG, Mitrofanov VG, Korochkin LI. Localisation of four genes encoding of the biochemical traits on second chromosomes of Drosophila imeretensis (in Russian). Doklady Academii Nauk of the USSR 1984;276:732-6.
28. Adrianov BV, Sorokina SY, Miuge NS, et al. Dynamics of Mitochondrial Polymorphism in a Natural Population of Drosophila littoralis. Genetika 2008;44:195-201.
29. Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 2001;16:314-21.
30. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 2004;21:1081-4.
31. Braude-Zolotarjova TY, Kakpakov VT, Schuppe NG. Male diploid embryonic cell line of Drosophila virilis. In Vitro 1986; 22:481-4.
32. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596-9.
33. Laslett D, Canbäck B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 2008;24:172-5.
34. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 1995;41:353-8.
35. Cha SY, Yoon HJ, Lee EM, et al. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, Bombus ignitus (Hymenoptera: Apidae). Gene 2007;392: 206-20.
36. Wilson K, Cahill V, Ballment E, Benzie J. The complete sequence of the mitochondrial genome of the crustacean Penaeus monodon: are malacostracan crustaceans more closely related to insects than to branchiopods? Mol Biol Evol 2000;17:863-74.
37. Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. Nature 1981;290470-4.
38. Berthier F, Renaud M, Alziari S, Durand R. RNA mapping on Drosophila mitochondrial DNA: precursors and template strands. Nucleic Acids Res 1986;14:4519-33.
39. Krzywinski J, Grushko OG, Besansky NJ. Analysis of the complete mitochondrial DNA from Anopheles funestus: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol 2006;39:417-23.
40. Carapelli A, Comandi S, Convey P, et al. The complete mitochondrial genome of the Antarctic springtail Cryptopygus antarcticus (Hexapoda: Collembola). BMC Genomics 2008;9:315.
41. Lavrov DV, Brown WM, Boore JL. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede Lithobius forficatus. Proc Natl Acad Sci USA 2000;97:13738-42.
42. Lewis DL, Farr CL, Farquhar AL, Kaguni LS. Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA. Mol Biol Evol 1994;11:523-38.
43. Brehm A, Harris DJ, Hernández M, et al. Structure and evolution of the mitochondrial DNA complete control region in the Drosophila subobscura subgroup. Insect Mol Biol 2001;10:573-8.
44. Clary DO, Wolstenholme DR. Drosophila mitochondrial DNA: conserved sequences in the A + T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J Mol Evol 1987;25:116-25.
45. Inohira K, Hara T, Matsuura ET. Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in Drosophila simulans and Drosophila mauritiana. Mol Biol Evol 1997;14:814-22.
46. Monforte A, Barrio E, Latorre A. Characterization of the length polymorphism in the A + T-rich region of the Drosophila obscura group species. J Mol Evol 1993;36:214-23.
47. Sugihara K, Yui R, Ibaragi Y, Matsuura ET. Complete nucleotide sequence of the A+T-rich region of Drosophila mauritiana mitochondrial DNA. Genes Genet Syst 2006; 81:21-8.
48. Oliveira MT, Azeredo-Espin AM, Lessinger AC. The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. J Mol Evol 2007;64:519-27.
49. Gray MW, Burger G, Lang BF. Mitochondrial evolution. Science 1999; 283:1476-81.
50. Stewart JB, Beckenbach AT. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the mead-

ow spittlebug Philaenus spumarius (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome 2005;48:46-54.
51. Stoletzki N, Eyre-Walker A. Synonymous codon usage in Escherichia coli: selection for translational accuracy. Mol Biol Evol 2007;24:374-81.
52. Ballard JW. Comparative genomics of mitochondrial DNA in members of the Drosophila melanogaster subgroup. J Mol Evol 2000;51:48-63.
53. Sawamura K, Koganebuchi K, Sato H, et al. Potential gene flow in natural populations of the Drosophila ananassae species cluster inferred from a nuclear mitochondrial pseudogene. Mol Phylogenet Evol 2008;48:1087-93.
54. Mishmar D, Ruiz-Pesini E, Brandon M, Wallace DC. Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. Hum Mutat 2004;23:125-33.
55. Andrianov BV, Zakharyev VM, Reznik NL, et al. Gypsy group retrotransposon Tv1 from Drosophila virilis. Gene 1999;239:193-9.
56. Terzian C, Pelisson A, Bucheton A. Evolution and phylogeny of insect endogenous retroviruses. BMC Evol Biol 2001;1:3.
57. Nefedova LN, Liubomirskaia NV, Il'in Iu V, et al. [Precise excision of long terminal repeats of the gypsy (mdg4) retrotransposon of Drosophila melanogaster detected in Escherichia coli cells is explained by its integrase function]. Genetika 2006;42:1656-63.