

# Genetic variation of the nuclear sequences of mitochondrial origin associated with retrotransposon *Tv1* insertions in *Drosophila* species of the *virilis* group

B.V. Andrianov , D.A. Romanov, T.V. Gorelova

Vavilov Institute of General Genetics, RAS, Moscow, Russia

Mitochondrial DNA sequences integrated into chromosomes are a promising object for designing genetic markers for studies of phylogenesis and genomic instability. Mitochondrial genomes of *D. virilis* and other *Drosophila* species of the *virilis* group contain (AT)<sub>n</sub> microsatellites in the spacer region between the *atp6* and *cox3* genes, and this microsatellite sequence is one of the hallmarks of the *virilis* group. The nuclear genome of *D. virilis* contains many extended fragments of mitochondrial DNA, which in total are several times longer than the mitochondrial genome. These nuclear sequences of mitochondrial origin contain all types of mitochondrial sequences, including mitochondrial genes and the aforementioned microsatellite sequence. The presence of the (AT)<sub>n</sub> microsatellite allows insertion of retrotransposon *Tv1*, which can transpose into the (AT)<sub>n</sub> microsatellite in a site-specific manner. The *Tv1* insertion into (AT)<sub>n</sub>, close to the *atp6* or *cox3* pseudogenes produces a unique sequence. This sequence is formed by retrotransposon *Tv1* and pseudogenes *atp6* or *cox3*. This unique sequence can be detected in the genome by a PCR-based method. We applied this method to the detection and analysis of the nucleotide variability of the pseudogenes *atp6* and *cox3* associated with *Tv1* insertions in a *D. virilis* cell culture and in the genomes of four *Drosophila* species of the *virilis* group: *D. virilis*, *D. montana*, *D. borealis*, and *D. laticola*. We discovered new events of mitochondrial sequence transfer to the nucleus in the transplanted cell culture of *D. virilis*, and new *Tv1* insertions, having emerged during the passage of this cell line were detected in the genome of the *D. virilis* transplanted cell culture. We found *atp6* and *cox3* pseudogenes associated with insertions of retrotransposon *Tv1* in the nuclear genomes of four *Drosophila* species from the *virilis* group. These chimeric sequences proved to be species-specific. The age of the *Tv1* insertion into the *atp6* and *cox3* pseudogenes is estimated at 1.50 Ma for *D. virilis*, 1.31 Ma for *D. laticola*, and 1.56 Ma for *D. borealis*. A specific situation was revealed for *D. montana*, in which *Tv1* insertions with nearly identical 5' and 3' long terminal repeats (LTRs) were present in accessions of flies from Europe and Asia. The age of this insertion was about 300 thousand years, and the insertion was absent from the *D. montana* fly line from North America.

Key words: mitochondrial genome; retrotransposons; numts; *Drosophila*; *virilis* group.

## Изучение изменчивости ядерных последовательностей митохондриального происхождения, ассоциированных с инсерциями ретротранспозона *Tv1*, у видов дрозофил из группы *virilis*

Б.В. Андрианов , Д.А. Романов, Т.В. Горелова

Институт общей генетики им. Н.И. Вавилова  
Российской академии наук, Москва, Россия

Митохондриальные последовательности, интегрированные в ДНК хромосом, – перспективный объект для разработки генетических маркеров филогенеза и геномной нестабильности. Митохондриальный геном *D. virilis* и других видов дрозофил из группы *virilis* содержит микросателлитные последовательности (AT)<sub>n</sub> в спейсерной области между генами *atp6* и *cox3*, что является отличительным признаком группы *virilis*. Ядерный геном *D. virilis* содержит большое количество протяженных фрагментов митохондриальной ДНК, которые в сумме в несколько раз длиннее митохондриального генома. Эти ядерные последовательности митохондриального происхождения содержат все типы митохондриальных последовательностей, в том числе митохондриальные гены и микросателлитные последовательности (AT)<sub>n</sub> в спейсерной области между генами *atp6* и *cox3*. Наличие микросателлита (AT)<sub>n</sub> обеспечивает возможность инсерции ретротранспозона *Tv1*, имеющего свойство встраиваться сайт-специфично в последовательности микросателлита (AT)<sub>n</sub>. В результате инсерции транспозона в микросателлит образуется уникальная последовательность, образованная ядерной копией гена *atp6* или *cox3* и ретротранспозоном *Tv1*, которая может быть выделена из генома методом ПЦР. Используя этот подход, мы выявили и проанализировали нуклеотидную изменчивость псевдогенов *atp6* и *cox3*, ассоциированных с инсерциями *Tv1*, в клеточной культуре *D. virilis* и у четырех видов дрозофил из группы *virilis*: *D. virilis*, *D. montana*, *D. borealis* и *D. laticola*. Выявлены новые события переноса митохондриальных последовательностей в ядро клетки в пересеваемой культуре *D. virilis* и новые события инсерций ретротранспозона *Tv1* в геноме клеток пересеваемой культуры, возникших в ходе пассирования данной клеточной линии. Показана видоспецифичность фрагментов митохондриальных псевдогенов *atp6* и *cox3*, ассоциированных с инсерциями ретротранспозона *Tv1*, в ядерном геноме видов дрозофил из группы *virilis*, позволяющая идентифицировать виды группы. Возраст инсерций *Tv1* в последовательности митохондриального происхождения у *D. virilis* равен 1.50 млн лет, *D. laticola* –

1.31 млн лет, *D. borealis* – 1.56 млн лет. Особая ситуация найдена для *D. montana*. У этого вида выявлена инсерция с практически идентичными 5' и 3' длинными концевыми повторами в линиях мух европейского и азиатского происхождения. Возраст инсерции около 300 тыс. лет, и она отсутствует у линии *D. montana* из Северной Америки.

Ключевые слова: митохондриальный геном; ретротранспозоны; numts; *Drosophila*; группа *virilis*.

#### HOW TO CITE THIS ARTICLE:

Andrianov B.V., Romanov D.A., Gorelova T.V. Genetic variation of the nuclear sequences of mitochondrial origin associated with retrotransposon *Tv1* insertions in *Drosophila* species of the *virilis* group. Vavilovskii Zhurnal Genetiki i Selektii = Vavilov Journal of Genetics and Breeding. 2018;22(7):887-894. DOI 10.18699/VJ18.430

The transfer of mitochondrial DNA into the nucleus of the cell has been found in all eukaryotic species studied so far (Bensasson et al., 2001; Richly, Leister, 2004; Hazkani-Covo et al., 2010). Fragments of mitochondrial DNA in chromosomes were called numts (abbreviation of *nuclear sequences of mitochondrial origin*) (Lopez et al., 1994). In the literature, this term is variously capitalized and italicized. We interpret *numt* as a genetic term and write it in lowercase Roman letters. The numbers and lengths of numts vary among species, depending on the ratios of their acquisition and loss.

The main source of data on the origin and variability of numts is the comparative analysis of complete genomes. The numbers of numts in the genomes of different fruit fly species vary by an order of magnitude (Rogers, Griffiths-Jones, 2012). Such a significant variation is likely to be determined by different rates of numt acquisition and loss. In *Drosophila*, the frequency of fixation of a new numt type in the genome is 0.75 copies per million years (Rogers, Griffiths-Jones, 2012). It is natural to assume that the speed of occurrence of new copies of numts is highly variable in different species.

The number of numts detected in a complete genome depends on search parameters. Therefore, data reported by different authors are controversial. In the genome of *Drosophila melanogaster* Meigen (1830), the number of detectable numts varies from 3 to 6 (Bensasson et al., 2001; Richly, Leister, 2004; Pamilo et al., 2007; Rogers, Griffiths-Jones, 2012; Hazkani-Covo, Martin, 2017), and this is the least number of numts among the studied genomes of *Drosophila*. As reviewed in (Rogers, Griffiths-Jones, 2012), the numbers of numts found in other *Drosophila* species are 5 in *D. simulans* Sturtevant (1919), 9 in each of *D. yakuba* Burla (1954) and *D. grimshawi* Oldenberg (1914), 20 in *D. erecta* Tsacas, Lachaise (1974), 24 in *D. mojavensis* Patterson (1940), 25 in *D. sechellia* Tsacas, Bachli (1981), 26 in *D. ananassae* Doleschall (1858), 54 in *D. persimilis* Dobzhansky, Epling (1944), 59 in *D. virilis* Sturtevant (1916), and 67 in *D. willistoni* Sturtevant (1916). *D. virilis* belongs to the group of *Drosophila* species with the highest number of numts in the genome. In the nuclear genome of this species, all sequences of the mitochondrial chromosome are represented.

Since insertions of fragments of mitochondrial DNA into chromosomal DNA cause mutations, a significant increase in the number of numts in the genome can be a marker of genomic instability. The frequency of mitochondrial DNA fragment insertion into chromosomal DNA varies in the course of evolution (Hazkani-Covo, Martin, 2017). It has been suggested that in most cases the appearance of new numts in the genome is related to the speciation process (Gunbin et al., 2017). In this connection, it is of interest to study the as-

sociation of numts with retrotransposons, since the induction of transpositions of retrotransposons and numts can cause genomic instability. Numts are usually integrated into (AT)<sub>n</sub> microsatellite sites and are often flanked by retrotransposons, or *Alu* repeats in humans (Tsuji et al., 2012). In *Drosophila*, 45 % of numts are located close to retrotransposons of LINE type and LTR-containing type (Rogers, Griffiths-Jones, 2012).

The search for numts associated with retrotransposons in the complete genome of *D. virilis* revealed an insertion of retrotransposon *Tv1* into the spacer region between the *atp6* and *cox3* numts. Retrotransposon *Tv1* was found in the *D. virilis* genome and in the genomes of all species of *Drosophila* forming the *virilis* group (Andrianov et al., 1999). Retrotransposon *Tv1* is integrated in a site-specific manner into the microsatellite (AT)<sub>n</sub> sequence to form a direct duplication (AT)<sub>4</sub> at the site of insertion (Andrianov et al., 2010). The presence of the (AT)<sub>n</sub> microsatellite favors *Tv1* integration into this site. As a result of *Tv1* insertion into the numt, a unique sequence arises, which can be detected from the genome by a PCR-based method. Using this approach, we analyzed the variability of numts associated with *Tv1* insertions in *D. virilis* fly lines, in *D. virilis* cell culture, and in the genomes of four species of *Drosophila* of the *virilis* group: *D. virilis*, *D. montana* Stone, Griffen, Patterson (1941), *D. lacicola* Patterson (1944) and *D. borealis* Patterson (1952). All numts associated with the retrotransposon *Tv1* in fly lines are located on the Y chromosome. Our data reveal new events of mitochondrial DNA transfer into chromosomes and new events of *Tv1* retrotransposition in *D. virilis* cell culture. This finding brought us to the conclusion that the emergence of new numt–*Tv1* associations were specific markers characterizing genomic instability in *D. virilis* cell culture.

## Materials and methods

**Fly and cell lines.** *Drosophila* lines of species of the *virilis* group used in the work were obtained from the Genetic Collection of Insect Lines and Insect Cell Cultures of the Vavilov Institute of General Genetics of RAS (<http://vigg.ru/index.php?id=337>), the Collection of *Drosophila* Genetic Lines of the Koltzov Institute of Developmental Biology of the RAS (<http://idbras.comcor.ru/collection/Drosophila.pdf>), and the collection of the Tucson *Drosophila* Species Stock Center (<http://stockcenter.arl.arizona.edu>). All *Drosophila* lines were isofemale except for the old laboratory line L160. The characteristics of the fly lines studied are shown in Supplement 1<sup>1</sup>. The permanent *D. virilis* cell line 79f7Dv3g was obtained in 1979 (Braude-Zolotarjova et al., 1986). All *Drosophila* lines

<sup>1</sup> Supplementary Materials 1–4 are available in the online version of the paper: <http://www.bionet.nsc.ru/vogis/download/pict-2018-22/appx14.pdf>

## PCR primers

Primer designation	Sequence 5'–3'	The genomic nucleotide sequence taken for primer design	Primer specificity	Species tested with the primer
Dvir0.1F	GTCAATGTTTCAGAAATCTGTGG	<i>D. littoralis</i> 06-17a FJ447340.1	<i>cox2</i>	<i>D. borealis</i> , <i>D. kanekoi</i> , <i>D. laticola</i> , <i>D. montana</i>
Dvir4.1F	AAGGAACCCCAGCAATTCTT	<i>D. virilis</i> 1051.87 BK006340.1	<i>atp6</i>	<i>D. virilis</i> , <i>D. virilis</i> cell line
Dvir5.1R	TCGTAAACCAATAGTTACAGGGTAAG	<i>D. virilis</i> 1051.87 BK006340.1	<i>cox3</i>	<i>D. borealis</i> , <i>D. laticola</i> , <i>D. lummei</i> , <i>D. montana</i>
Dvir5.2R	AGCAGGTGGTCATGATGCTC	<i>D. virilis</i> 1051.87 BK006340.1	<i>cox3</i>	<i>D. borealis</i> , <i>D. laticola</i> , <i>D. montana</i>
Dvir5.3R	TCAATATCATGCTGCTGCTTC	<i>D. virilis</i> 1051.87 BK006340.1	<i>cox3</i>	<i>D. borealis</i> , <i>D. kanekoi</i> , <i>D. laticola</i> , <i>D. lummei</i> , <i>D. montana</i>
Dvir5.4R	TTGTGAATGGTTTCCTTCCA	<i>D. virilis</i> 1051.87 BK006340.1	<i>cox3</i>	<i>D. virilis</i> cell line
Dvir6.1F	GTAATTCGACCAGGAACCTTAGC	<i>D. virilis</i> 1051.87 BK006340.1	<i>atp6</i>	<i>D. borealis</i> , <i>D. kanekoi</i> , <i>D. laticola</i> , <i>D. lummei</i> , <i>D. montana</i>
Dvir6.2F	CTCTTTTAGGACCCTCAGGTCA	<i>D. virilis</i> 1051.87 BK006340.1	<i>atp6</i>	<i>D. borealis</i> , <i>D. laticola</i> , <i>D. lummei</i> , <i>D. montana</i> , <i>D. virilis</i>
Dvir6.3F	TTCTGTATTTCGACCCATCAGC	<i>D. virilis</i> 1051.87 BK006340.1	<i>atp6</i>	<i>D. virilis</i>
Dvir6.4R	TTCCTTGAGGAACATAATGAGC	<i>D. borealis</i> 0961.00 KF669860.1	<i>atp6</i>	<i>D. borealis</i> , <i>D. kanekoi</i> , <i>D. laticola</i> , <i>D. montana</i>
Dvir7.1R	TTCCCGAATATGAACTGATTTATCT	<i>D. virilis Tv1</i> complete sequence AF056940.1	LTR <i>Tv1</i>	<i>D. virilis</i>
Dvir7.2R	CAATTCCTTTATTTTCACTAACGGCTAC	<i>D. virilis Tv1</i> complete sequence AF056940.1	LTR <i>Tv1</i>	<i>D. borealis</i> , <i>D. laticola</i> , <i>D. montana</i>
Dvir8.1F	GCACCCACCCTTTTCACAT	<i>D. virilis Tv1</i> complete sequence AF056940.1	LTR <i>Tv1</i>	<i>D. borealis</i> , <i>D. laticola</i> , <i>D. montana</i>
Dvir8.2F	CAAAGCATTTACGTAAGCAATGA	<i>D. virilis Tv1</i> complete sequence AF056940.1	UTR <i>Tv1</i>	<i>D. virilis</i> cell line

and the cell culture used in this study are available upon request to the corresponding collections or to the authors of the article.

**DNA isolation, PCR and sequencing of PCR fragments.** Genomic DNA was isolated from *Drosophila* imagoes and cultured cells by the conventional phenol-chloroform extraction method (Sambrook et al., 1989). PCR amplification was performed on a template of total DNA isolated from an individual *Drosophila* imago or from  $10^6$  cells of permanent cell culture.

The primers used to amplify the *atp6* and *cox3* numts associated with the insertion of retrotransposon *Tv1* and fragments of the mitochondrial genes *atp6* and *cox3* of *Drosophila* species of the *virilis* group are listed in the Table. PCR was performed on an Applied Biosystems (PCR System 2700) thermocycler with a universal Encyclo Plus PCR kit (Evrogen, Moscow) as recommended by the manufacturer, the reaction volume being 25  $\mu$ L.

PCR schedule: (1) predenaturation at 95 °C for 5 min; (2) 5 cycles: denaturation at 95 °C for 40 s, annealing at 50 °C for 40 s, and elongation at 72 °C for 60 s; (3) 30 cycles: denaturation at 95 °C for 30 s, annealing at 59 °C for 30 s, and elongation at 72 °C for 60 s; (4) postextension at 72 °C for 7 min.

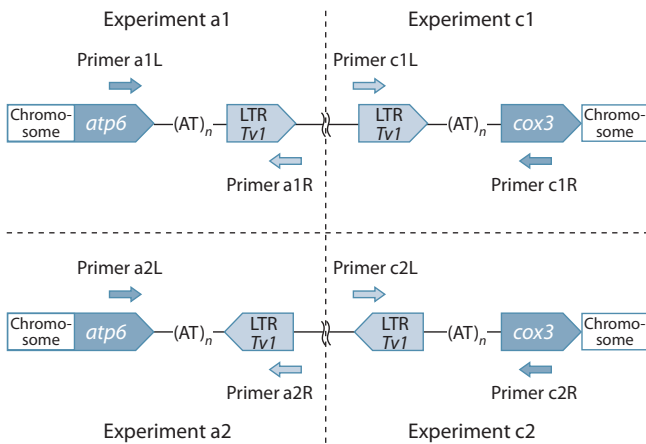
All conceivable structures of “chimeric” sequences resulting from the insertion of retrotransposon *Tv1* into the microsatellite (AT)<sub>n</sub> sequence between the *atp6* and *cox3* genes in the forward and reverse orientations relative to the orientation of the mitochondrial genes are presented in Figure 1. There may be four types of sequences, and they correspond to four types of experimental design to obtain a “chimeric” PCR fragments

**Experiment a1.** The following pairs of primers were used to amplify the *atp6* numts associated with the insertion of retrotransposon *Tv1* in the forward orientation: (1) Dvir6.1F and Dvir7.2R, (2) Dvir6.2F and Dvir7.2R, (3) Dvir6.3F and Dvir7.2R. In case of *D. virilis*, primers Dvir6.1F and Dvir7.2R were replaced by Dvir4.1F and Dvir7.1R.

**Experiment c1.** Primers for *cox3* numts associated with the insertion of *Tv1* in the forward orientation: (1) Dvir8.1F and Dvir5.1R, (2) Dvir8.1F and Dvir5.2R, (3) Dvir8.1F and Dvir5.3R.

**Experiment a2.** Primers for *atp6* numts associated with the insertion of *Tv1* in the reverse orientation: (1) Dvir6.1F and Dvir8.1F, (2) Dvir6.2F and Dvir8.1F, (3) Dvir6.3F and Dvir8.1F.

**Experiment c2.** Primers for *cox3* numts associated with the insertion of *Tv1* in the reverse orientation: (1) Dvir7.2R



**Fig. 1.** Four types of experiments (a1, a2, c1, and c2) for identification of *atp6* and *cox3* numts associated with the insertion of *Tv1*.

and Dvir5.1R, (2) Dvir7.2R and Dvir5.2R, (3) Dvir7.2R and Dvir5.3R.

**PCR of mitochondrial genes.** Primers for fragments of mitochondrial genes *atp6* and *cox3* of *Drosophila* of the *virilis* group (1) Dvir0.1F and Dvir6.4R for *D. borealis*, *D. kanekoi*, *D. lacicola*, and *D. montana*; (2) Dvir6.2F and Dvir5.1R or Dvir6.1F and Dvir5.3R for other *Drosophila* species of the *virilis* group.

The results of electrophoretic fractionation of PCR fragments formed by *atp6* and *cox3* numts associated with the insertion of retrotransposon *Tv1* are presented in Supplement 2. The sizes of the PCR fragments can be indicated only approximately, because *atp6* and *cox3* numts reveal indels in interspecific and, sometimes, intraspecific comparison, and the lengths of the spacer sequence between *atp6* and *cox3* differ among species.

**Cloning.** PCR products were run in agarose gel, eluted, and purified with an elution kit (Zymoclean™ Gel DNA Recovery Kit, Zymo Research, USA) according to manufacturer's recommendations. All PCR products were cloned prior to sequencing. PCR product cloning was performed using the pGEM®-T Easy Vector System according to standard protocols (Fermentas InsTAclone™ PCR Cloning Kit). The resulting clones were sequenced. At least three independent clones were sequenced for each PCR fragment. Sequencing of the amplification products was conducted with both primers on an ABI PRISM 3500 instrument using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, United States), according to manufacturer's recommendations.

**Phylogenetic analysis.** Alignment of the resulting sequences and phylogenetic analysis were carried out in the MEGA6.06 program (Tamura et al., 2013). For construction of dendrograms, we used the NJ method and the Kimura evolutionary model. The bootstrap support of 1000 replicas was used. Indels were removed from the compared sequences prior to the construction of dendrograms.

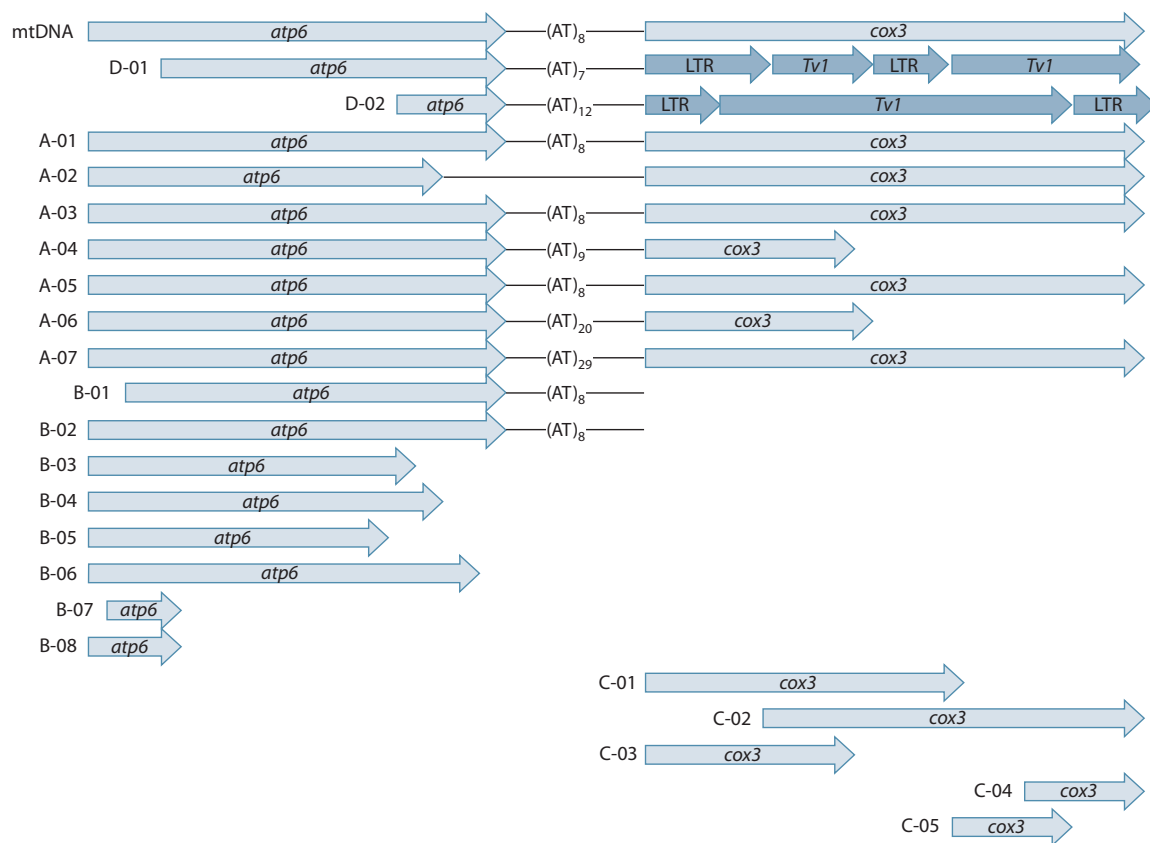
## Results and discussion

To provide a basis for further work, we characterized *atp6* and *cox3* numts and associated insertions of retrotransposon *Tv1* in the genome of *D. virilis* (Fig. 2).

Experimental search for chimeric sequences of *atp6* numt – *Tv1* in fly lines of *D. virilis* of different geographical origins revealed the expected nucleotide sequence corresponding to the D-01 map (see Fig. 2) in all fly lines examined. We compared five *D. virilis* lines of different geographic origins (see Supplement 1). They all contain the same numt. The nucleotide divergence of this *atp6* numt from the *atp6* mitochondrial gene of *D. virilis* is 0.09. The amino acid sequence contained 21 substitutions. This numt has no internal termination codons. Insertions of *Tv1* are also the same in different fly lines, except for the insertion of *Tv1* in line L160. The *Tv1* associated with the *atp6* numt in this line has several point nucleotide substitutions and two small deficiencies in the LTR sequence. The lengths of numts in the genome of *D. virilis* flies and in the permanent cell line are constant, being the same as in the corresponding mitochondrial sequence. The genetic maps of all experimentally obtained numts of *D. virilis* associated with the insertion of *Tv1* are presented in Supplement 3. Differences in the length of PCR fragments, observed only on the DNA template of the cell culture, are determined by differences in the length of LTRs of retrotransposon *Tv1*. These differences in length are due to the presence or absence of 40-bp long duplications. Only one type of association of *atp6* numts with *Tv1* in direct orientation was found in fly genomes, and there were no associations of *cox3* numts with *Tv1*, whereas in the cell culture all the four possible types of associations were identified (see Supplements 2 and 3). Consequently, they emerged in the cell culture in the process of cultivation after the obtaining of this culture in 1979 (Braude-Zolotarjova et al., 1986). Unfortunately, the *D. virilis* cell line was established from a fly line that has been lost, and presently they cannot be compared. However, with regard to the fact that *D. virilis* is a nearly monomorphic species (Mirol et al., 2008), we can use extant fly lines for comparisons with the *D. virilis* cell line.

This raises the question of the origin of these numts and the origin of *Tv1* insertions. Theoretically, there are two possible sources of new numt insertions in the genome: mitochondrial DNA and previously arisen numts (Hazkani-Covo et al., 2003). To make it clear, we conducted a phylogenetic analysis of the variability of the obtained numts and the corresponding sequences of mitochondrial genes. The result of the comparison of mitochondrial genes with numts is shown in Figure 3.

We detected almost complete identity between the newly emerged numts in the cell culture and the sequence of the mitochondrial gene of *D. virilis*, and significant differences from the numts of *D. virilis* flies. Consequently, the numts that arise in the cell culture descend from mitochondrial DNA rather than from preexisting numts. The genome-derived numt associated with the insertion of *Tv1* belongs to the most divergent and probably the oldest insertions of numts in the *D. virilis* genome. Comparison of the variability of numts and associated copies of *Tv1* provides an answer to the question of the differences in age of these insertions. In doing it, we analyzed the *Tv1* LTR nucleotide variability from *Tv1* insertions in the numts. The result of the phylogenetic analysis of *Tv1* LTR variability is presented in Figure 4. *In silico* search revealed 12 types of *Tv1* LTRs in the *D. virilis* genome. We also found two currently active types of *Tv1* in the cell culture (Fig. 4). These types differ from the *Tv1* copy associated with the numt in the genome. Hence, both numts of *D. virilis* and



**Fig. 2.** Schematic representation of all *atp6* and *cox3* numts from the complete genome of *Drosophila virilis* line TSC # 15010-1051.87 (GenBank accession number: AANI00000000.1).

Each numt is assigned a unique name to the left of the map. Numts *atp6* D-01 and D-02 are associated with insertions of retrotransposon *Tv1* in the direct orientation.

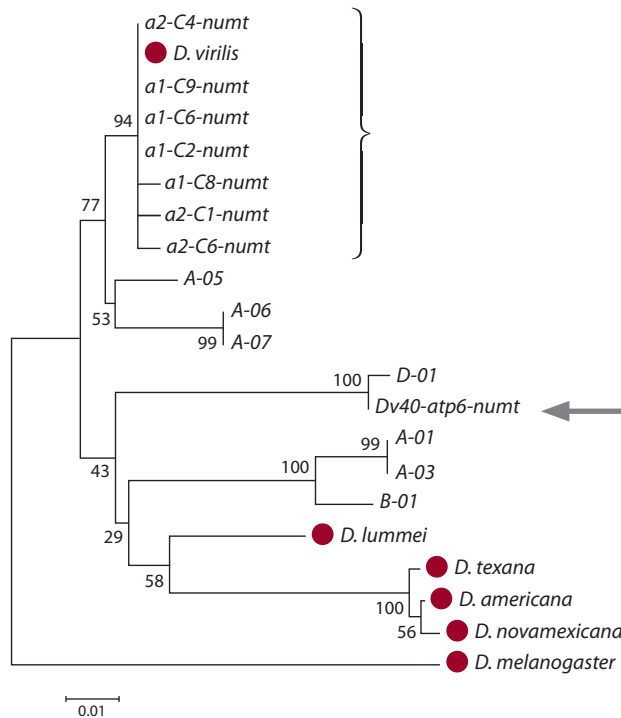
the associated *Tv1* LTRs in the cell culture are among the youngest sequences in their groups, and they arose during the cell culturing. The ages of the numt and *Tv1* insertions in the cell culture match.

To find out whether *Tv1* insertions always occur in newly formed numts, we searched some *Drosophila* species of the *virilis* group for associations of numts with *Tv1* according to the experimental design shown in Figure 1. Numts marked with *Tv1* insertions were identified in *D. virilis*, *D. borealis*, *D. lacicola*, and *D. montana*. Genetic maps of the PCR fragments are presented in Supplement 4. Insertions of *Tv1* in the direct orientation with respect to *atp6* and *cox3* were found in *D. lacicola*. The nucleotide divergence of the *cox3* numt from the sequence of the corresponding mitochondrial gene of the same species is 0.05, and for the two detected *atp6* numts, the nucleotide divergences are 0.06 and 0.07. The numbers of amino acid substitutions for these numts are 6, 5, and 9, respectively. None of the nucleotide substitutions generated termination codons. These two *D. lacicola atp6* numts are associated with different copies of *Tv1* (GenBank accession numbers: KX399470, KX399471). As inferred from the intercomparison of these numts, they diverged after the transfer to the nucleus. A total of 14 nucleotide substitutions were found; of them, 5 in the first position, 6 in the second, and only 3 in the third position of codons, respectively. Note that in the mitochondrial genome most nucleotide substitu-

tions fall in the third position. It is reasonable to suggest that the observed differences accumulated after the transfer of the sequences to the nuclear genome.

Insertions of *Tv1* in the reverse orientation relative to *atp6* and *cox3* were found in the genome of *D. montana*. The nucleotide divergence of the detected numts from the corresponding mitochondrial genes of this species is 0.09. This level of nucleotide divergence suggests that the divergence of the numts is ancient. We found 12 amino acid substitutions and 1 termination codon. The *cox3* numt of the same species has 15 amino acid substitutions and 3 termination codons. In total, four lines of *D. montana* were analyzed. In the 1021.13, 20 OL8 and KR 13-09 fly lines, *atp6* and *cox3* numts are associated with retrotransposon *Tv1* in the reverse orientation. The 5' and 3' LTRs of this *Tv1* are nearly identical, which suggests a recent insertion of retrotransposon *Tv1* into the ancient numt. In *D. montana* line 1021.19 from North America, these numts were not found. In this line, we found another *cox3* numt, associated with the retrotransposon *Tv1* in the reverse orientation. The nucleotide divergence of this numt is 0.05; it has seven amino acid substitutions and one termination codon.

An insertion of *Tv1* in the reverse orientation with respect to *cox3* was found in *D. borealis*. The nucleotide divergence of the detected numt from the corresponding mitochondrial DNA is also large, 0.12. There are 14 amino acid substitu-

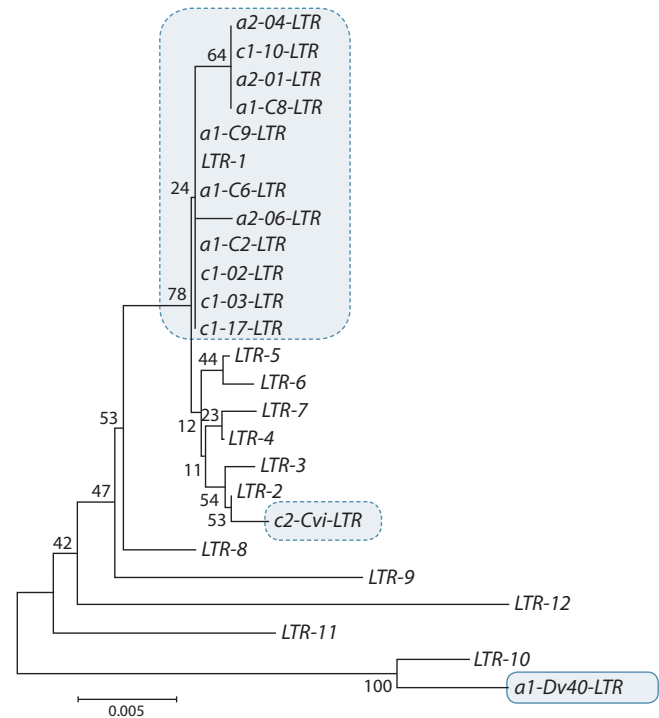


**Fig. 3.** Reconstruction of the phylogenetic tree of *atp6* numts from the genome of the permanent cell culture of *D. virilis* and *D. virilis* fly lines, in comparison with the mitochondrial *atp6* genes of various *Drosophila* species of the *virilis* group.

As an external group, we use *atp6* of *D. melanogaster*. Each of the *atp6* numts of *D. virilis* found during *in silico* analysis of the complete genome is indicated with a capital letter and two digits. The genetic maps of these numts are shown in Figure 2. The *atp6* numt from the genome of *D. virilis* flies of the Dv40 line is marked with an arrow. The *atp6* numts from the genome of the cell culture of *D. virilis* isolated in our experiments are indicated with a curly brace. The first two characters in the names of the experimentally obtained numts indicate the type of experiment in which the nucleotide sequences were obtained. Nucleotide sequences were submitted to GenBank under accession numbers JX560766–JX560769 and KF669862–KF669864. Nucleotide sequences of mitochondrial genes of the corresponding species of *Drosophila* of the *virilis* group were submitted to GenBank with accession numbers KX399463, FJ536196, FJ536199, FJ536203, and FJ536204. Mitochondrial sequences are marked with black circles.

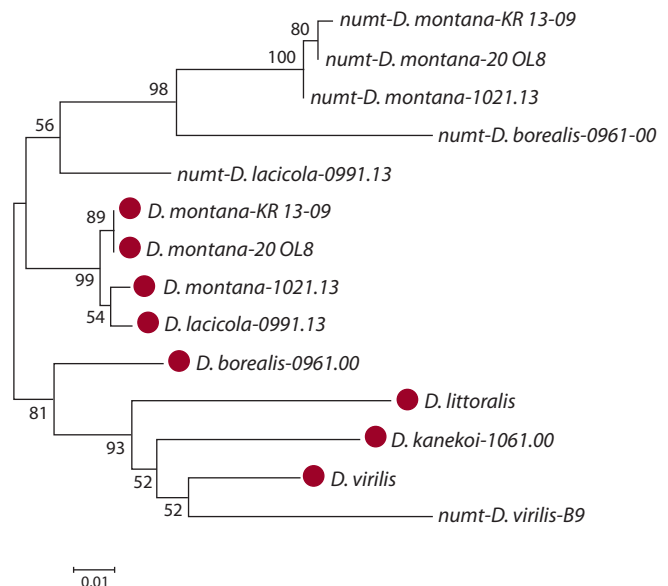
tions and 3 termination codons. In addition, *D. borealis* has an unusual insertion of retrotransposon *Tv1*, located at the beginning of the *atp6* gene rather than between the *atp6* and *cox3* genes. The detected numt includes *atp6* and *cox3* sequences associated with *Tv1* in the direct orientation. The nucleotide divergence of this *atp6* numt from the *D. borealis* mitochondrial *atp6* gene is 0.14; it has 31 amino acid substitutions and 1 termination codon. The nucleotide divergence of this *cox3* numt from the *D. borealis* mitochondrial *cox3* gene is 0.13; it has 11 amino acid substitutions and 2 termination codons. All the detected *atp6* and *cox3* numts are associated with retrotransposon *Tv1* only in males (see Supplement 2), which points to their location on the Y chromosome. It can be assumed that the Y chromosome is the preferred place of the preservation of numts, whereas in other parts of the genome these sequences are rapidly lost.

Phylogenetic analysis reveals characteristic differences in the time of the emergence of numts and insertions of *Tv1* (Figures 5 and 6). Figure 5 shows the phylogenetic reconstruction



**Fig. 4.** Reconstruction of the phylogenetic tree of *Tv1* LTR sequences associated with *atp6* numts.

LTMs from insertions into the numts of the cell culture are enclosed with a dashed line, and the LTR associated with the numt from the fly genome is boxed. The nucleotide sequences of LTR *Tv1* were obtained by *in silico* analysis of the complete genome of *D. virilis* line TSC # 15010-1051.87 (GenBank accession number: AANI00000000.1). The schematic representation and GenBank accession numbers of the nucleotide sequences of PCR fragments containing *Tv1* LTRs are presented in Supplement 3.



**Fig. 5.** Reconstruction of the phylogenetic tree of *atp6* numts associated with insertions of *Tv1* in the genomes of flies of the *virilis* group.

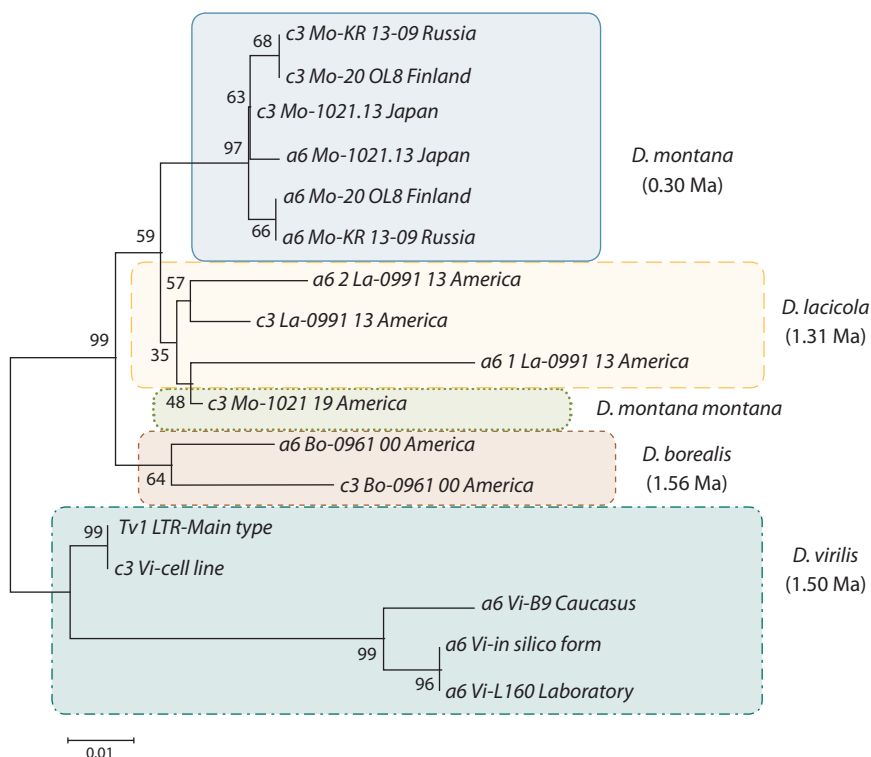
The GenBank accession numbers of the numt sequences are given in Supplement 3. Nucleotide sequences of mitochondrial genes of the corresponding *Drosophila* lines were submitted to GenBank under accession numbers KX399453–KX399459. The sequences of the *atp6* genes of *D. virilis* and *D. littoralis* are taken from the complete mitochondrial genomes of these species, GenBank accession numbers BK006340 and FJ447340. Mitochondrial genes are marked with black circles.

tion of the divergence of *atp6* numts in *Drosophila* of the *virilis* group. In all examined cases, the time of divergence between numts and the modern forms of the mitochondrial genes is about several million years. The ages of insertions of retrotransposon *Tv1* can be estimated by comparing different copies of LTRs associated either with the *atp6* or *cox3* numts in a particular line of flies. The analysis is illustrated in Figure 6. If we assume the rate of fixation of nucleotide substitutions to be 0.016 substitutions per site for one million years, the typical value for noncoding *Drosophila* sequences (Bowen, McDonald, 2001), the age of *Tv1* insertions into the numt of *D. virilis* is estimated to be 1.5 million years. A similar situation is observed in *D. borealis* and *D. laticola*. The opposite is found in *D. montana*. Insertions with nearly identical 5' and 3' LTRs are found in the 1021.13, 20 OL8 and KR 13-09 fly lines. The age of this *Tv1* insertion is about 300000 years, and this insertion is absent from line 1021.19, which has another insertion of *Tv1*.

It is important to note that there is no coincidence between the ages of numts and of the associated insertion of *Tv1* in *D. montana*. Young copies of *Tv1* are associated with the ancient numt. Therefore, the transfer of these elements occurred independently and at different times. This fact points to probable differences in the molecular mechanisms of the appearance of numt-*Tv1* associations in somatic cells of the cell culture and germline cells.

### Conclusions

We investigated the variability of *atp6* and *cox3* numts associated with site-specific insertions of retrotransposon *Tv1* in *Drosophila* of the *virilis* group and *D. virilis* permanent cell culture. The method of numt detection was based on the ability of retrotransposon *Tv1* to transpose into the microsatellite (AT)<sub>n</sub> sequence and on the presence of this microsatellite in the spacer region between the mitochondrial genes *atp6* and *cox3* in the mitochondrial genomes of *Drosophila* of the *virilis* group. In the *D. virilis* cell line, we found new events of mitochondrial DNA transfer to the nucleus and new *Tv1* insertions. Most of the new insertions of retrotransposon *Tv1* in the cell culture occur in the newly emerged numts. As a result, the ages of



**Fig. 6.** Reconstruction of the phylogenetic tree of LTRs of retrotransposons *Tv1* associated with *atp6* and *cox3* numts in *Drosophila* of the *virilis* group.

The first two characters in the name of a sequence indicate the type of experiment in which the LTR sequence was obtained. The next two letters indicate the species of *Drosophila*: Mo for *D. montana*, Bo for *D. borealis*, La for *D. laticola*, and Vi for *D. virilis*. The next letters indicate the *Drosophila* line name. The bootstrap support values are listed at the nodes of the phylogram. Trees are drawn in scale. The lengths of branches correspond to the frequencies of nucleotide substitutions per site. Estimation of the age of *Tv1* insertions is provided by the formula  $T = d/2k$ , where  $T$  is age, Ma;  $d$ , nucleotide divergence;  $k = 0.016$  (Bowen, McDonald, 2001). The LTR sequences of *Tv1* and the associated sequences of numts were submitted to GenBank under accession numbers KX399470–KX399482.

retrotransposon insertions and numt insertions in the cell line are the same. The opposite situation was found in *Drosophila* species of the *virilis* group. Insertions of *Tv1* occur in ancient numts in the genomes of flies. As a result, the ages of retrotransposon insertions and numts are different. The *atp6* and *cox3* numts, which are associated with site-specific insertions of retrotransposon *Tv1*, are species-specific in *Drosophila* of the *virilis* group.

### Acknowledgements

The work was supported by project AAAA-A16-116111610180-3 “Study of variability of autonomous genetic elements of insects and development of markers of genome instability”, contract 0112-2016-0001.

### Conflict of interest

The authors declare no conflict of interest.

### References

- Andrianov B., Goryacheva I., Mugue N., Sorokina S., Gorelova T., Mitrofanov V. Comparative analysis of the mitochondrial genomes in *Drosophila virilis* species group (Diptera: Drosophilidae). Trends Evol. Biol. 2010;2:e4. DOI 10.4081/eb.2010.e4.
- Andrianov B.V., Zakharyev V.M., Reznik N.L., Gorelova T.V., Evgen'ev M.B. Gypsy group retrotransposon *Tv1* from *Drosophila virilis*. Gene. 1999;239:193-199.
- Bensasson D., Zhang D.-X., Hartl D.L., Hewitt G.M. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 2001;16(6):314-321.
- Bowen N.J., McDonald J.F. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res. 2001;11(9):1527-1540. DOI 10.1101/gr.164201.

- Braude-Zolotarjova T.Ya., Kakpakov V.T., Schuppe N.G. Male diploid embryotic cell line of *Drosophila virilis*. In vitro. 1986;22:481-484.
- Gunbin K., Peshkin L., Popadin K., Annis S., Ackermann R.R., Khrapko K. Integration of mtDNA pseudogenes into the nuclear genome coincides with speciation of the human genus. A hypothesis. Mitochondrion. 2017;34:20-23. DOI 10.1016/j.mito.2016.12.001.
- Hazkani-Covo E., Martin W.F. Quantifying the number of independent organelle DNA insertions in genome evolution and human health. Genome Biol. Evol. 2017;9(5):1190-1203. DOI 10.1093/gbe/evx078.
- Hazkani-Covo E., Sorek R., Graur D. Evolutionary dynamics of large numts in the human genome: rarity of independent insertions and abundance of post-insertion duplications. J. Mol. Evol. 2003;56:169-174. DOI 10.1007/s-00239-002-2390-5.
- Hazkani-Covo E., Zeller R.M., Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 2010;6(2):e1000834. DOI 10.1371/journal.pgen.1000834.
- Lopez J.V., Yuhki N., Masuda R., Modi W., O'Brien S.J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. 1994;39:174-190.
- Mirol P.M., Routtu J., Hoikkala A., Butlin R.K. Signals of demographic expansion in *Drosophila virilis*. BMC Evol. Biol. 2008;8:59. DOI 10.1186/1471-2148-8-59.
- Pamilo P., Viljakainen L., Vihavainen A. Exceptionally high density of NUMTs in the honeybee genome. Mol. Biol. Evol. 2007;24(6):1340-1346. DOI 10.1093/molbev/msm055.
- Richly E., Leister D. NUMTs in sequenced eukaryotic genomes. Mol. Biol. Evol. 2004;21(6):1081-1084. DOI 10.1093/molbev/msh110.
- Rogers H.H., Griffiths-Jones S. Mitochondrial pseudogenes in the nuclear genomes of *Drosophila*. PLoS One. 2012;7(3):e32593. DOI 10.1371/journal.pone.0032593.
- Sambrook J., Fritsch E.F., Maniatis T. Molecular cloning: a laboratory manual. N. Y., 1989.
- Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 2013;30(12):2725-2729. DOI 10.1093/molbev/mst197.
- Tsuji J., Frith M.C., Tomii K., Horton P. Mammalian NUMT insertion is non-random. Nucleic Acids Res. 2012;40:9073-9088. DOI 10.1093/nar/gks424.

---

#### ORCID ID

B.V. Andrianov [orcid.org/0000-0002-0064-4696](https://orcid.org/0000-0002-0064-4696)

D.A. Romanov [orcid.org/0000-0003-3340-9278](https://orcid.org/0000-0003-3340-9278)

T.V. Gorelova [orcid.org/0000-0003-3777-0626](https://orcid.org/0000-0003-3777-0626)