Diversity and evolution of telomere and subtelomere DNA sequences in insects

Vladimir A. Lukhtanov[1]*

[1]Department of Karyosystematics, Zoological Institute of the Russian Academy of Sciences, Universitetskaya nab. 1, St. Petersburg 199034, Russia

*Corresponding author. E-mail: lukhtanov@mail.ru

Abstract

In insects, two types of telomere length maintenance are known: telomerase-dependent, resulting in chromosome ends consisting of short nucleotide repeats (typically TTAGG), and telomerase-independent, resulting in chromosome ends consisting of long nucleotide repeats or transposon-like elements. However, only a few species have been previously studied with regard to their telomere DNA sequences. Here, based on analysis of chromosome-level genome assemblies, I present the data on telomere and subtelomere organization for 180 species from 148 genera, 53 families and 8 orders of insects.

Analysis of these taxa reveals that in fact chromosome ends of most insect species have an intermediate structure and consist of numerous arrays of short telomeric repeats interspersed with telomere-specific non-LTR retrotransposons. An unexpectedly high level of diversity of short telomeric motifs (22 variants ranging in length from 1 to 17 nucleotides) is documented. Telomeres consisting of long repeats (from 173 to 374 bp) are confirmed for flies (the order Diptera) and also found in the orders Odonata and Hymenoptera. The most unusual telomere structure is found in the bee *Lasioglossum lativentre*, in which the chromosomes possess the short telomeric repeat TTAGGTCTGGG at only one end, whereas opposing ends terminate with medium and long repeats.

I conclude that different types of telomere organization and numerous variants of long and short T-containing motifs, including the $(T)_n$ mononucleotide sequence, are compatible with the performance of telomere functions. I argue that both telomerase-dependent and telomerase-independent mechanisms for maintaining telomere length operate simultaneously in many insects. The balance between them and the exchange of sequences between telomeres and subtelomeres are most likely the key factors that determine the structure and evolution of telomeres.

Significance

Multilayer telomeres, resulted from numerous, site-specific insertions of retrotransposons into the region of short telomeric repeats, are not an aberrant type of organization, as previously thought. They are widely distributed among insects and can represent up to 30-40 % of eukaryotic species diversity. Accordingly, the telomere maintenance mechanism based on the joint work of telomerase-dependent and telomerase-independent mechanisms can also be extremely widespread in nature.

Key words: telomere, telomerase, telomeric motif, telomere-specific non-LTR retrotransposons, subtelomere, chromosome-level genome assembly

INTRODUCTION

Telomeres are protein-DNA complexes located at the ends of eukaryotic linear chromosomes (Blackburn, 1991; Osterhage & Friedman, 2009, Nandakumar & Cech, 2013; Shay & Wright, 2019). Telomeres protect the terminal regions of chromosomal DNA from progressive degradation and ensure the integrity of chromosomes (Zakian, 1995; Frydrychová et al., 2021). These essential functions are conserved in eukaryotic organisms and have not changed over hundreds of millions of years of their evolution (Saint-Leandre & Levine, 2020). The persistence of these functions correlates well with the stability in the primary structure of telomeric DNA. In most organisms, as diverse as protozoans, fungi, metazoans, and plants, telomeric DNA consists of the short TTAGGG motif repeated many times (Louis & Vershinin 2005; Červenák et al., 2021). This specific structure of telomere is maintained by telomerase, a telomere-specific reverse transcriptase that elongates the ends of DNA using a small region of its RNA subunit as a template for the telomeric motif (Greider & Blackburn, 1989).

Although telomerase RNA varies dramatically in size, primary sequence and secondary structure among different groups of eukaryotes (Logeswaran et al., 2021), its template part is conservative (Chen et al., 2000; Červenák et al., 2021). Changes in this template are rare in evolution, and as a rule, changed motifs mark higher taxa such as flowering plants (TTTAGGG), nematodes (TTAGGC) and arthropods (TTAGG) (Vítková et al., 2005; Kuznetsova et al., 2020).

However, in some groups of organisms such as alveolate protists (Blackburn & Gall, 1978, Klobutcher et al., 1981; Shirley, 1994; Sohanpal et al., 1995; Liu et al., 1998), ascomycetous yeasts (Červenák et al., 2021) and plant genera *Cestrum* and *Allium* (Peška et al., 2015; Fajkus et al., 2016; Peska & Garcia, 2020), the structure of telomeric motifs changes noticeably faster, and the motifs themselves are more diverse.

Insects are of particular interest for studying the organization and evolution of telomeres, as well as for studying the mechanisms for maintaining telomere length. The telomere motif $(TTAGG)_n$ is considered to be ancestral for the class Insecta and the entire phylum Arthropoda and is preserved in the majority of the studied insect orders and families (Frydrychová et al., 2004; Vítková et al., 2005; Lukhtanov & Kuznetsova, 2010; Kuznetsova et al., 2020). In some beetles (Coleoptera) it is known to be replaced by similar motifs TCAGG and TTAGGG (Prušáková et al., 2021). It can be assumed that the telomere length in these insects is maintained by the standard telomerase-dependent mechanism (Korandová et al., 2014) (see also Mason et al., 2011, 2016). However, for insects of the order Diptera, a completely different, telomerase-independent mechanism has been described, resulting in chromosome ends consisting of long repeats or transposon-like elements (Biessmann & Mason, 1997; Mason et al., 2008). Nematocera species, the so-called lower Diptera, have longer repeated sequences at their chromosome ends that may elongate telomeres by gene conversion (Nielsen & Edstrom, 1993; Rosén & Edström, 2002). *Drosophila* species (higher Diptera, suborder Brachycera) carry terminal retrotransposons that elongate telomeres by transposition specifically to chromosome termini (Biessmann et al., 2000; Casacuberta & Pardue, 2003).

Telomere-specific non-LTR retrotransposons are known to be involved in telomere organization and function in a few other, non-Diptera insects that display the standard TTAGG motif at the ends of chromosomes. In these insects, arrays of the main short motif are interspersed with (or contacted by) retrotransposons (Okazaki et al., 1995; Takahashi et al., 1997; Fujiwara et al., 2005; Garavís et al., 2013). However, the prevalence of this type of telomere organization and its possible role in maintaining telomere length are poorly understood (but see Garavís et al., 2013).

Another problem in the study of insect telomeres is the so-called "loss of the ancestral TTAGG motif", which has been repeatedly reported for members of several insect orders (Frydrychova & Marec, 2002; Mason et al., 2016; Kuznetsova et al., 2020; Prušáková et al., 2021), for example, for several families of Hemiptera (Frydrychová et al., 2004; Golub et al., 2015, 2017, 2018), Hymenoptera (Menezes et al., 2017; Gokhman & Kuznetsova, 2018) and Odonata (Kuznetsova et al., 2021). Reports of the loss of the TTAGG motif are based on numerous studies in which TTAGG (or similar motifs) were searched using fluorescent in situ hybridization (FISH) and/or southern blot hybridization techniques. Both techniques can detect a specific DNA sequence, but unfortunately are not available for detecting DNA fragment with unknown DNA sequence (Grozeva et al., 2011). Therefore, the true reasons for this "loss of the ancestral TTAGG motif" (transition to another short motif or a complete change in the structure of telomeres, for example, replacement by long repeats) has remained unknown.

Here, based on analysis of chromosome-level genome assemblies, I present the data on telomere and subtelomere organization for 180 species from 148 genera, 53 families and 8 orders of insects.

This approach, based on the analysis of nucleotide sequences of whole chromosomes, solves the problem of the "lost TTAGG motif", reveals the true diversity of telomeric and subtelomeric organization, and outlines approaches to solving the question of how insect telomeres evolved.

MATERIAL AND METHODS

Chromosome-level genome assemblies generated by the Darwin Tree of Life Project (https://www.darwintreeoflife.org/) (The Darwin Tree of Life Project Consortium, 2022) and freely available upon deposition in the European Nucleotide Archive (ENA) (https://www.darwintreeoflife.org/wp-content/uploads/2020/03/DToL-Open-Data-Release-Policy-1.pdf) were used for analysis of telomere and subtelomere sequences (Table S1). Published and freely available chromosome-level genome assemblies of *Formica selysi* (Brelsford et al., 2020), *Solenopsis invicta* (Yan et al., 2020), *Nasonia vitripennis* (Benetta et al., 2020) and *Apis mellifera* (Wallberg et al., 2019) were also used.

The chromosome-scale scaffolds confirmed by the Hi-C data were analyzed using the graphical sequence panel (format "Graphics") implemented in GenBank. Alternatively, in some cases, the chromosome-scale scaffolds were downloaded and analyzed in BioEdit (Hall, 1999). The terminal regions of chromosomes were first visually inspected. Then tandem repeats were searched using Tandem Repeats Finder (Benson, 1999). Finally, the position analysis and the comparison of putative telomeric and subtelomeric repeats were performed using the "Search" option in the BioEdit program and the "Search" and "BLAST" options implemented in GenBank.

In almost all species in which a distinct short telomere motif has been found in the majority of chromosomes, in some chromosomes the telomere motif was absent on one or even both chromosome ends. This situation has been previously described for the honey bee (Wallberg et al., 2019). As for the honey bee, I interpret these chromosome-scale scaffolds as truncated ones, but not as lacking arrays of telomere motifs. A telomere was considered present if its length exceeded 400 bp, the estimated minimal functional length required for t-loop formation (Watson et al., 2021).

The search of the telomere-specific non-LTR retrotransposons was conducted manually (Goubert et al., 2022) in 500000 bp terminal fragments of the 5′ ("left") and 3′ ("right") chromosome ends, by using the contact of the polyA (polyT) tract with short telomeric repeats (TTAGG or others) as a specific marker (Okazaki et al., 1995; Takahashi et al., 1997).

In accordance with common practice, hereinafter in the text (except Figure 2) I list telomeric and subtelomeric sequences as if they were found at the 3′ chromosome end ("right" telomere), even if in Genbank they are given for the 5′ chromosome end ("left" telomere) and have, respectively, a reverse complement orientation and spelling.

RESULTS AND DISCUSSION

Repeats forming the extreme ends of chromosomes (telomeric motifs)

In the studied species, 22 variants of the main short telomeric motifs were found at the very ends of chromosomes: 1bp repeat (T), 5 bp repeats (TTAGG, TCAGG, TTGGG), 6bp repeat (TTAGGG, TTCCTC, GGGTCT), 8 bp repeat (TTATTGGG), 9 bp repeat (TGGATGGGA), 10 bp repeats (TTAAAACGCC, TTGAAACGCC, TTAAAGCGCC, TTAAGGCGTT, TTAGGGTGGT, TTAGTTTGGG), 11 bp repeats (TTAGGTCTGGG, TTAGGTTGGGG, TTAGGTTCGGG, TTGGGTCTGGG, TTGCGTCTGGG, TTGCGTCAGGG) and 17 bp repeat AAAAAATTCTTTGATGC (Table S1). Of these motifs, three were previously known for insects (TTAGG, TCAGG, TTAGGG) (Frydrychová et al., 2004; Kuznetsova et al., 2020; Prušáková et al. 2021), and 19 are described here for the first time.

It is not surprising that the 5 bp motif TTAGG, "standard" for insects (Frydrychová et al., 2004; Kuznetsova et al., 2020) was found at the ends of chromosomes in most of the studied species (Table S1). Confirming already known data (Frydrychová et al., 2004; Mandrioli et al., 2012; Vershinina et al., 2015), this motif was found to be invariable within the orders Lepidoptera and Trichoptera (butterflies and moths, caddisflies). It was also found in the only studied species of the order Plecoptera, in most beetles, in one of the basal lineages of the order Hymenoptera (Tenthredinidae: *Tenthredo notha*) and in several representatives of the clade Aculeata (Hymenoptera) (Fig. 1). In beetles (the order Coleoptera), the modified 5 bp motif TCAGG was found in two species (Melyridae: *Malachius bipustulatus*; Pyrochroidae: *Pyrochroa serraticornis*), and the modified 5 bp motif TTGGG was found in the species *Apoderus coryli* (Attelabidae). The motif TTGGG was also found in the white-legged damselfly *Platycnemis pennipes* (Odonata, suborder Zygoptera: Platycnemididae).

Strongly modified 10 bp motifs were found in pentatomomorphic bugs *Aelia acuminata* (Hemiptera, Pentatomidae) (TTAGGGATGG) and *Acanthosoma haemorrhoidale* (Hemiptera,

Acanthosomatidae) (TTAGGGTGGT). (The loss of the TTAGG motif for the infraorder Pentatomomorpha was previously reported (Grozeva et al., 2011)).

The most intriguing results were obtained in the study of the order Hymenoptera (Fig. 1, Table S1). In the "primitive" family Tenthredinidae, which represents one of the first branches within the Hymenoptera, a common sawfly *Tenthredo notha* had the TTAGG motif, ancestral to all arthropods. This result is consistent with previously obtained, FISH-based data on two other species of this family (Gokhman & Kuznetsova, 2018).

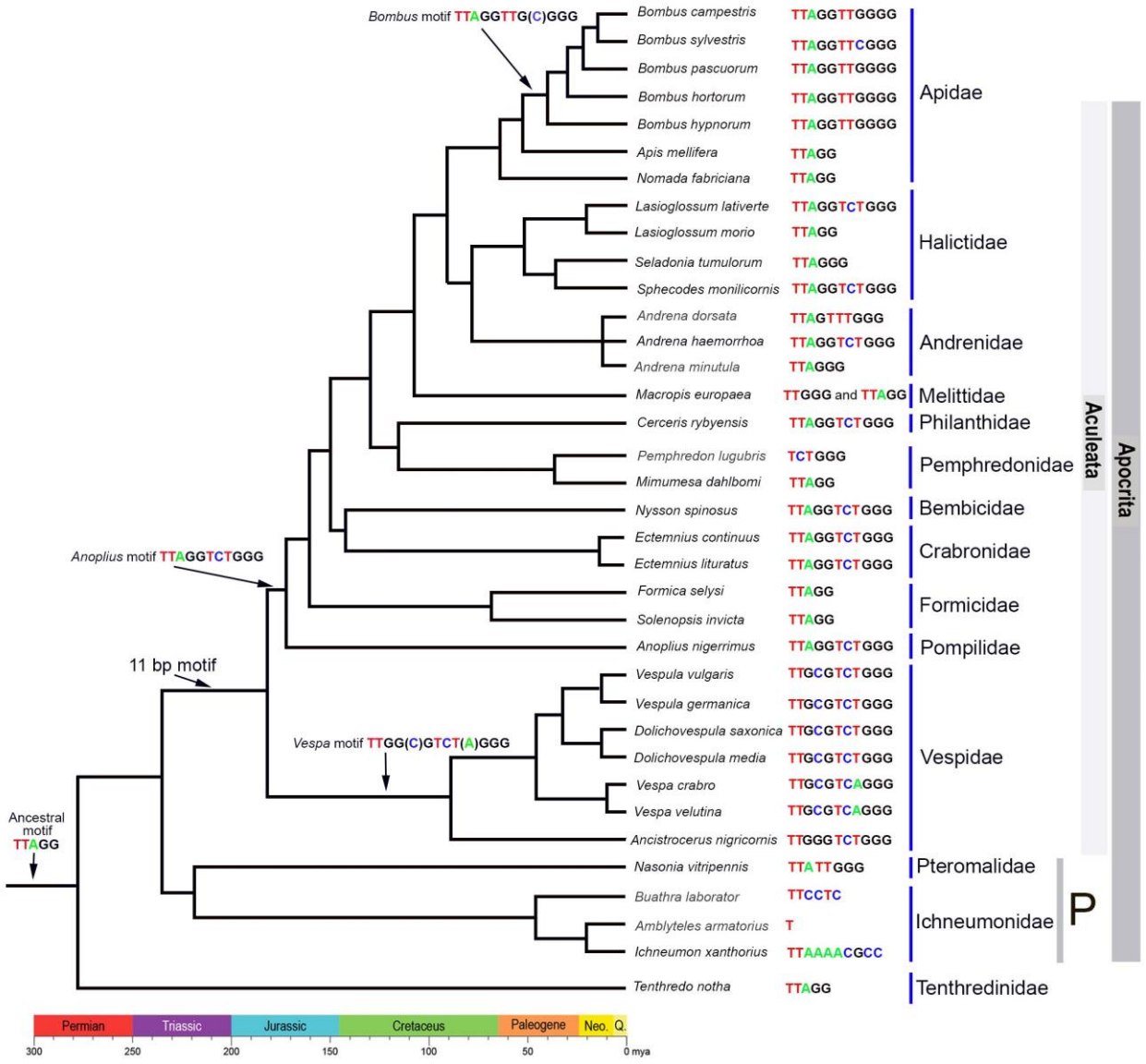

Figure 1. Diversity and evolution of short telomeric motifs in the order Hymenoptera. The phylogeny and dating are based on the published works (Cameron et al., 2007; Peters et al., 2017; Sann et al., 2018, 2021). P is Parasitica (Here: Pteromalidae + Ichneumonidae).

The clade Apocrita showed a clear trend to a high diversity of short telomeric motifs, with the prevalence of multi-nucleotide motifs (8-11 bp). In these multi-nucleotide motifs, the first five letters were often represented by the classic TTAGG (or slightly modified variants), while the G and T nucleotides predominated in the second part of the motifs.

In four studied species of parasitic wasps (Parasitica, a group of the infraorder Apocrita), the 5 bp TTAGG motif was not found (Table S1, Fig. 1), confirming the previous results of Gokhman et al. (2014). Instead, 6 bp motif TTCCTC was found in *Buathra laborator* (Ichneumonidae) and various motifs, with predominance of 10 bp motif TTAAAACGCC and the 8 bp motif TTATTGGG were found in *Ichneumon xanthorius* (Ichneumonidae) and *Nasonia vitripennis* (Pteromalidae) respectively.

The most unusual telomere structure was found in the parasitic wasp *Amblyteles armatorius* (Ichneumonidae). In this species a mononucleotide telomeric motif T was found. It consisted of the (T)n sequence at the 3′ chromosome end and (A)n sequence at the 5′ chromosome. The length of this mononucleotide telomere ranged from 306 to 629 bp, being comparable with the estimated minimal functional length required for t-loop formation (Watson et al., 2021).

The clade Aculeata, the group containing stinging wasps, ants and bees, was characterized by the presence of 11 nucleotide motifs. The TTAGGTCTGGG motif predominates in different families and is probably ancestral for the Aculeata.

Groups of specific motifs, that are derived from TTAGGTCTGGG, were found in the family Vespidae (TTGGGTCTGGG, TTGCGTCAGGG, TTGCGTCTGGG) and in the bumble bees (genus *Bombus*) (TTAGGTTGGGG, TTAGGTTCGGG).

The 6 bp motifs were found in the bee *Seladonia tumulorum* (Halictidae) (TTAGGG) and the wasp *Pemphredon lugubris* (Pemphredonidae) (TCTGGG).

The 10 bp motif TTAGTTTGGG was found in the bee *Andrena dorsata* (Andrenidae).

In addition, six cases of possible reversion to the ancestral TTAGG motif were identified in Aculeata: (1) in the family Formicidae, (2) in *Mimumesa dahlbomi* (Pemphredonidae), (3) in *Macropis europea* (Melittidae), (4) in *Lasioglossum morio* (Halictidae), (5) in *Nomada fabriciana* and (6) *Apis mellifera* (Apidae). Alternatively, we can think that the TTAGG motif was ancestral for Aculeata; then we must assume the multiple cases of independent origin of the TTAGGTCTGGG motif in different branches of Aculeata.

Short telomeric repeats were not found in flies (Diptera), which is also in line with expectations (Rosén & Edström, 2002; Mason et al., 2008; Kuznetsova et al., 2020). Instead, in five species of the suborder Brachycera, long repeats were found at the ends of chromosomes. The length of these repeats varied from 173 to 374 bp. Alignment of these long telomeric repeats from different ends of the same chromosome, from different chromosomes of the same species, and from different species showed: (1) species specificity and small variations (95-98% similarity) in length and nucleotide composition within each species, and (2) sharp interspecific differences, while maintaining some similarity between the species, which most likely indicates the origin of these repeats from a common ancestor.

Short telomeric repeats were not found in the blue-tailed damselfly *Ischnura elegans* (Odonata, Zygoptera: Coenagrionidae). In this species, various sequences, including low-copy repeats of various length, were found at the very ends of autosomes, while 175 bp long repeats formed the very terminal regions of the X chromosome.

Intraspecific stability and variation of short telomere motifs

The length of the most terminal telomeric sequences in the studied species varied from 365 to 15724 bp and included from 33 to 1989 short repeats. Within this length, the structure of the main motifs was stable in most cases, sometimes with rare single altered variants, which can be considered as single substitutions/indels or sequencing errors.

However, in the order Hymenoptera, several species were found in which the main motifs were regularly interspersed with variant repeats.

In the bumble bee *Bombus campestris*, the main repeat 11 bp motif TTAGGTTGGGG was found to be interspersed with the sequences TTAGGTGGGGG, TTAGGTTGGTGGGG, TTAGGTTGGTTAGTGGGG, TTAGGTTGGGTAGGTTTGG, TTAGGTTGGGT and TTTGGGTTGGGG. In the bumble bee *B. sylvestris*, the main 11 bp repeat TTAGGTTCGGG was found to be regularly interspersed with 16 bp motif TTAGGTTAGGTTCGGG.

In the pteromalid wasp *Nasonia vitripennis*, the main motif TTATTGGG obviously prevailed, but the modified sequences TTATTGGGGG, TTTTTATTGGG, TTTTTTATTGGG, TTATTGG, TTATTGGAGGG, TTATTGGGG, TTATTATTGGG and TTTTGTTATTGGG were regularly found.

Only partially confirming previous analyses (Garavís et al., 2013; Wallberg et al., 2019), in the honey bee *Apis mellifera*, the TTAGG telomeres were found to be interspersed with variant repeats TTAGGGTT and TTATAGG.

In the bee *Macropis europaea*, the two main motives TTGGG and TTAGG were found.

In the beetle *Philonthus cognatus* the main motif TTAGG interspersed with TTTAGG.

In the parasitic wasp *Ichneumon xanthorius* the motif TTAAAACGCC was found to be fixed in chromosome 11 and to prevail in chromosomes 1 and 8 and in the left telomere of chromosome 2, in which it was interspersed with variant motives TTAAAACGCCT, TTAAACGCCT, TTACAACGCC, TTAGAACGCC, TTAAACCGCC, TTACAACGCC, TTACGACGCC and TTAAATCGCC. In chromosome 6, the motif TTAAAACGCC and the altered motif TTGAAACGCC occurred with approximately equal frequency. The motif TTAAAGCGCC was fixed in chromosome 9. In chromosome 7, the strongly altered motif AAAAAATTCTTTGATGC occurred. In chromosome 5, the strongly altered main motif TGGATGGGA was found to be interspersed with variant sequences CGGGA, TGGGA and TGGA. Finally, in the right telomere of chromosome 2, the main motif TTAAGGCGTT (which is actually a palindrome of the main motif in the left telomere!) was interspersed with variant repeats TTAAAGCCGTT, TTGTAAGGCG, TTCTAAAGGCG and TTAAAGGCGTT.

The most unusual telomere structure and intraspecific variations were found in chromosomes 1-13 of the bee *Lasioglossum lativentre*. In these chromosomes, one end is formed by arrays of a short motif TTAGGTCTGGG associated with subtelomeric sequence $(TTAGGTCTGCATCGCGG)_n$ or *SART* elements, and the other end is formed by medium and long repeats (similar to Diptera) (Table 1). The structure of the first ends (without the TTAGGTCTGGG motif) is difficult to explain by truncation, since they do not contain not only the telomeric TTAGGTCTGGG motifs, but also the specific subtelomeric sequences of the second ends (although independent confirmation of telomere structure in this unusual case would be desirable, for example, using the FISH method to reveal the structure of telomeres on microscopic preparations). Interestingly, a similar structure was recently discovered in the tapeworm

Hymenolepis microstoma (Platyhelminthes), in which the chromosomes possess telomeric repeats GGGATT at only one end, whereas opposing ends terminate with a novel repeat array with a median unit length of 179bp) (Olson et al., 2020).

Table 1. Telomere and subtelomere sequences in chromosomes of the bee *Lasioglossum lativentre*.

Chromosome #	First chromosome end: telomere	First end: subtelomere	Second chromosome end: telomere
1	TTAGGTCTGGG	*SART*-(TTAGGTCTGGG)$_n$	41 bp repeats
2	TTAGGTCTGGG	*SART*-(TTAGGTCTGGG)$_n$	304 bp repeats
3	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	150 bp repeats
4	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	118 bp repeats
5	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	304 bp repeats
6	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	Various short, medium and long repeats
7	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	Various short, medium and long repeats
8	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	Various short, medium and long repeats
9	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	40-41 bp repeats
10	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	41 bp repeats
11	TTAGGTCTGGG	*SART*-(TTAGGTCTGGG)$_n$	211 bp repeats
12	TTAGGTCTGGG	(TTAGGTCTGCATCGCGG)$_n$	Various short, medium and long repeats
13	TTAGGTCTGGG	*SART*-(TTAGGTCTGGG)$_n$	148 bp repeats
14	41 bp repeats		41 bp repeats

Sequences adjacent to the most terminal telomeric repeats (Telomere-Subtelomere structure)

Subtelomeric DNA consists of specific repetitive sequences that are in direct contact with telomeres and, accordingly, have a subterminal position in chromosomes (Louis & Becker, 2014). As a rule, subtelomeres are more diverse in structure of their DNA sequences, which is correlated with the diversity of subtelomeric proteins (Louis & Vershinin, 2005; Saint-Leandre & Levine, 2020).

In order to study the diversity of the subtelomeric region, in the species under study I inspected the sequences that were adjacent to the most terminal chromosomal array. This search revealed three types of the subterminal organization. The sequences of the first two types had

features typical of subtelomeres (Louis & Becker, 2014), while the third type is represented by multilayer telomeres that share some of the properties of subtelomeres (see below).

(Type 1) The subtelomeric region of the first type was found to include numerous tandem repeats of various lengths. This type was found in the caddisfly *Limnephilus lunatus* (Trichoptera), the bug *Aelia acuminata* (Hemiptera), in representatives of the orders Hymenoptera (*Anoplius nigerrimus*, *Bombus sylvestris*, *Ectemnius continuus*, *Nasonia vitripennis*, *Nomada fabriciana*, *Nysson spinosus*, *Seladonia tumulorum*, *Tenthredo notha*; part of chromosomes of *Andrena haemorrhoa*, *Lasioglossum morio* and *Mimumesa dahlbomi*) and Coleoptera (part of chromosomes of *Apoderus coryli*) (Table S1).

A variant of the first type were subtelomeres, which consisted only of long repeats. This variant was found in two species: in the right telomere of chromosome 6 in the beetle *Apoderus coryli* (100 repeats of the 171 bp motif) and in the bee *Macropis europaea*, in which the subtelomere consisted predominantly of 121 bp repeats (Table S1). This variant was previously reported for the aphid *Myzus persicae* (Hemiptera) (Spence et al., 1997). Such subtelomeres, consisting of long repeats, are structurally very similar to the telomeres found by me in five species of flies (Diptera: suborder Brachycera) (Table S1) and previously in *Chironomus* (Diptera: suborder Nematocera) (Rosén & Edström, 2002).

(Type 2) The second type, which is widespread in the order Hymenoptera (the families Apidae, Crabronidae, Formicidae, Halictidae, Philanthidae, Vespidae), was represented by subtelomeres consisting of modified telomeric sequences (Table S1). These modified sequences were found to include: (a) altered versions of the main motif; (b) elongated and varying in structure variants that retained similarities with the main motif, and (c) longer repeats (20-41 bp) that had a remote resemblance to the main motif. Interestingly, these longer repeats may have high specificity. For example, the 41 bp subtelomeric repeat AAAATACCTTATAATATAACGTAATATGCAGTGTATCATTG was detected in all 14 chromosomes of the bee *Lasioglossum lativentre*, but the BLAST function throughout the Genebank database did not reveal it in any other organisms. Кроме того, в субтеломерном региона могут встречаться отдельные копии неизмененного основного теломерного мотива. In addition, copies of the unaltered main telomeric motif may sporadically occur in the subtelomeric region, for example, in the wasp *Dolichovespula sylvestris.*

(Type 3) The most widespread among the studied species and the most intriguing was the third type of subterminal organization, which was found to be characterized by numerous arrays of short repeats (for example, TTAGG in Lepidoptera, TTAGGTTGGGG in bumble bees) interspersed with telomeric repeat-specific non-LTR retrotransposons of the *SART* and *TRAS* families.

TRAS is a family of retrotransposable, telomeric repeat-associated elements with a polyA (polyT) tail at the end (Okazaki et al., 1995). It is integrated in a highly specific manner into the short telomeric repeats at the same target sequence. In right telomere (3′ chromosome end) it can be easily recognized by its typical long (14-72 $(T)_n$ fragment, immediately starting after the GG dinucleotide of the short repeat (Fig. 2).

SART is a family of retrotransposable, telomeric repeat-associated elements with a polyA (polyT) tail at the end, in a reverse orientation from that of *TRAS* (Takahashi et al., 1997). It is inserted into the telomeric repeats in a highly specific manner. Therefore, in right telomere (3′ chromosome end) it can be easily recognized by its typical long (14-72 (A)n fragment contacting with the nucleotides GG(GGG) and then followed by standard short motifs (Figs 2 and 3).

For the silkworm *Bombyx mori*, the retrotransposons *TRAS* and *SART* are sometimes considered as subtelomeric elements because they do not occupy the most terminal position, being proximal to the most distal array of canonical short repeats (Kubo et al., 2001; Fujiwara, 2014; Louis & Becker, 2014; Mason et al., 2016). In fact, they are an integral part of the telomere, since they are inserted specifically and exclusively into telomeric repeats and cannot exist outside of short telomeric motifs (Osanai et al., 2006).

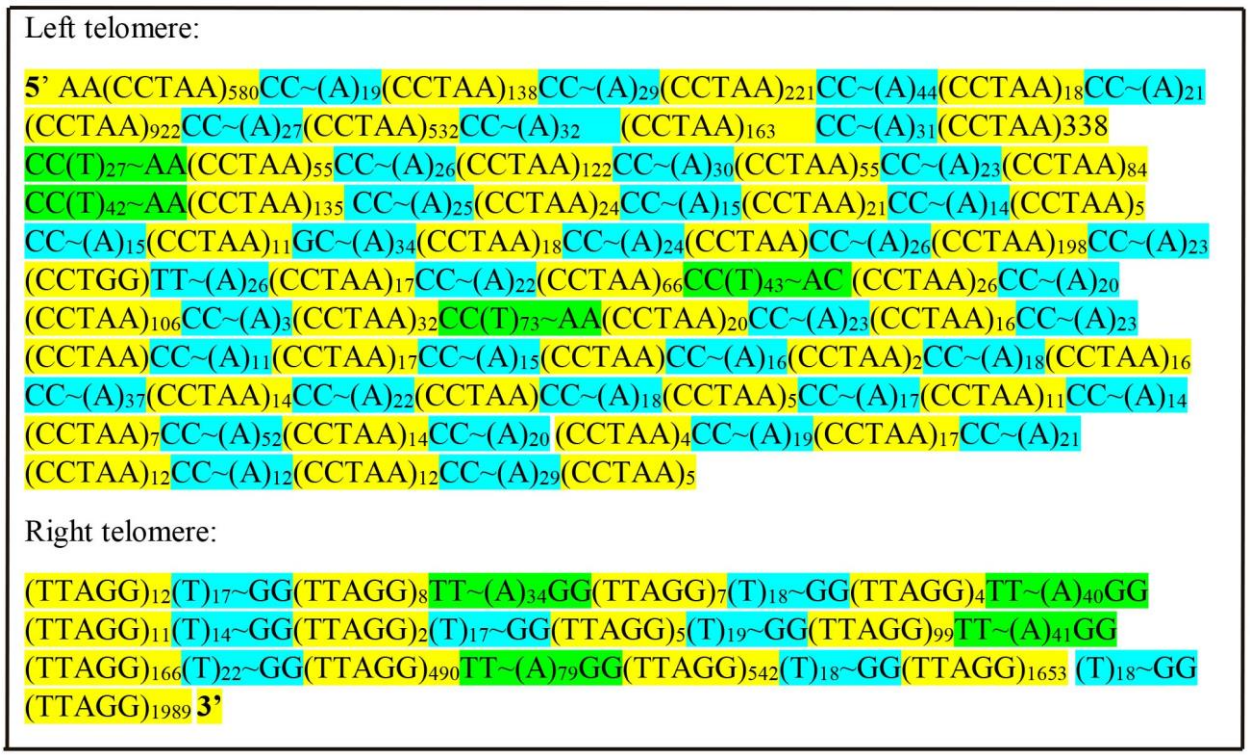

Figure 2. Telomere structure in chromosome 1 of *Orgya antiqua* (Lepidoptera).

Short telomeric repeats (CCTAA in left (5′) and TTAGG in right (3′) chromosome ends) (yellow) are interspersed with retrotransposable elements *TRAS* (blue) and *SART* (green). Left telomere (positions from 1 to 266133) includes 44 arrays of TTAGG interspersed with 39 *TRAS* and 4 *SART* retrotransposons. Right telomere (positions from 44592109 to 44675926, 83817 bp) includes 13 arrays of TTAGG interspersed with 8 *TRAS* and 4 *SART* retrotransposons. The tilde symbol (~) denotes the main body of *TRAS* and *SART* (without polyA/polyT tail); its length varies from 614 to 15369 bp, with a clear predominance of sequences from 6000 to 7000 bp.

To the best of my knowledge, this type of organization has previously been reported for only six species of insects: three species of moths (Kubo et al., 2001); flour beetle *Tribolium castaneum* (Osanai et al., 2006), pea aphid *Acyrthosiphon pisum* (Hemiptera) (International Aphid Genomics Consortium, 2010), human body louse *Pediculus humanus humanus* (Phthiraptera) (Kirkness et al., 2010). It was also known for the mite *Tetranychus urticae* (Grbić et al., 2011).

Here it was found in 113 species of Lepidoptera, 10 species of Coleoptera, 10 species of Hymenoptera, 3 species of Trichoptera and 1 species of Hemiptera (Table S1). The number of retrotransposons inserted in telomeric regions was found to vary from single ones found in some (not all) chromosomes (for example, in *Mimumesa dahlbomi* only in chromosome 9, in *Coccinella septempunctata* only in chromosome 8) to several tens at each chromosome end. In the latter case,

the massive integration of these elements into the proximal regions of the TTAGG, TTGGG, TCAGG, TTAGGTTGGGG, TTAGGTCTGGG, and TTGGGTCTGGG repeat arrays resulted in huge telomeric massive up to 260 kb in size (Fig. 2).

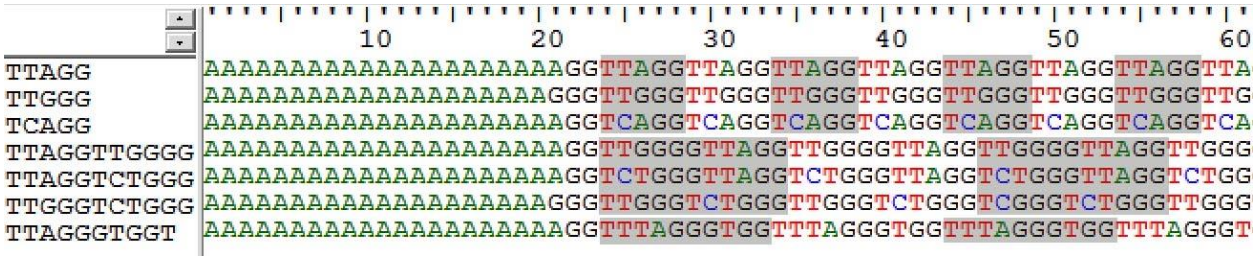

Figure 3. Contact area between the (A)*n*-tail of the *SART* retrotransposon and short telomeric repeats TTAGG, TTGGG, TCAGG, TTAGGTTGGGG, TTAGGTCTGGG, TTGGGTCTGGG, and TTAGGGTGGT.

The telomeric repeat-specific non-LTR retrotransposons seems to be highly species-specific. For example, in *Bombus hortorum*, one of the retrotransposons of the *SART* family has a length of 5951 bp. It was found in all 18 chromosomes of this species, and the similarity between chromosomes was 99-100%. The BLAST function revealed this (or similar) element in *Bombus vancouverensis*, *B. bifarius*, *B. vosnesenskii*, *B. sylvestris*, *B. campestris*, *B. terrestris*, *B. pascuorum*, as well as in all 12 chromosomes of *B. hypnorum*, but the similarity between different species was in the range of 69-80%.

Interestingly, the three types of subterminal organization of chromosomes found are not necessarily fixed within a species or even individual chromosomes (Table S1); for example, the beetle *Apoderus coryli* has telomeres of all three of these types. In general, subtelomeric sequences are much more variable than telomeric sequences.

CONCLUSION

This study solves the problem of the lost of the TTAGG telomeric motif in insects (Frydrychova & Marec, 2002; Mason et al., 2016; Kuznetsova et al., 2020, 2021; Prušáková et al., 2021) and shows that the absence of this motif in representatives of the orders Hemiptera and Hymenoptera is explained by evolutionary transitions to other, previously unknown short telomeric repeats. In the order Odonata (suborder Zygoptera), the absence of the TTAGG motif (Kuznetsova et al., 2021) can be explained by both nucleotide substitutions (as it is observed in *Platycnemis pennipes*, which have TTGGG instead of the ancestral TTGGG motif) and by a complete change in the telomere structure due to loss of short telomeric repeats (as seen in *Ischnura elegans*).

The study revealed an extremely high diversity in the structure of telomeric nucleotide sequences in insects. This diversity is manifested in:

(1) numerous (22) variants of short telomeric motifs found in different orders, especially in Hymenoptera (Table S1),

(2) in presence of telomeres consisting of long repeats, especially in Diptera, but also in Odonata and Hymenoptera (Table S1), and

(3) different types of intra-telomeric organization (Fig. 4).

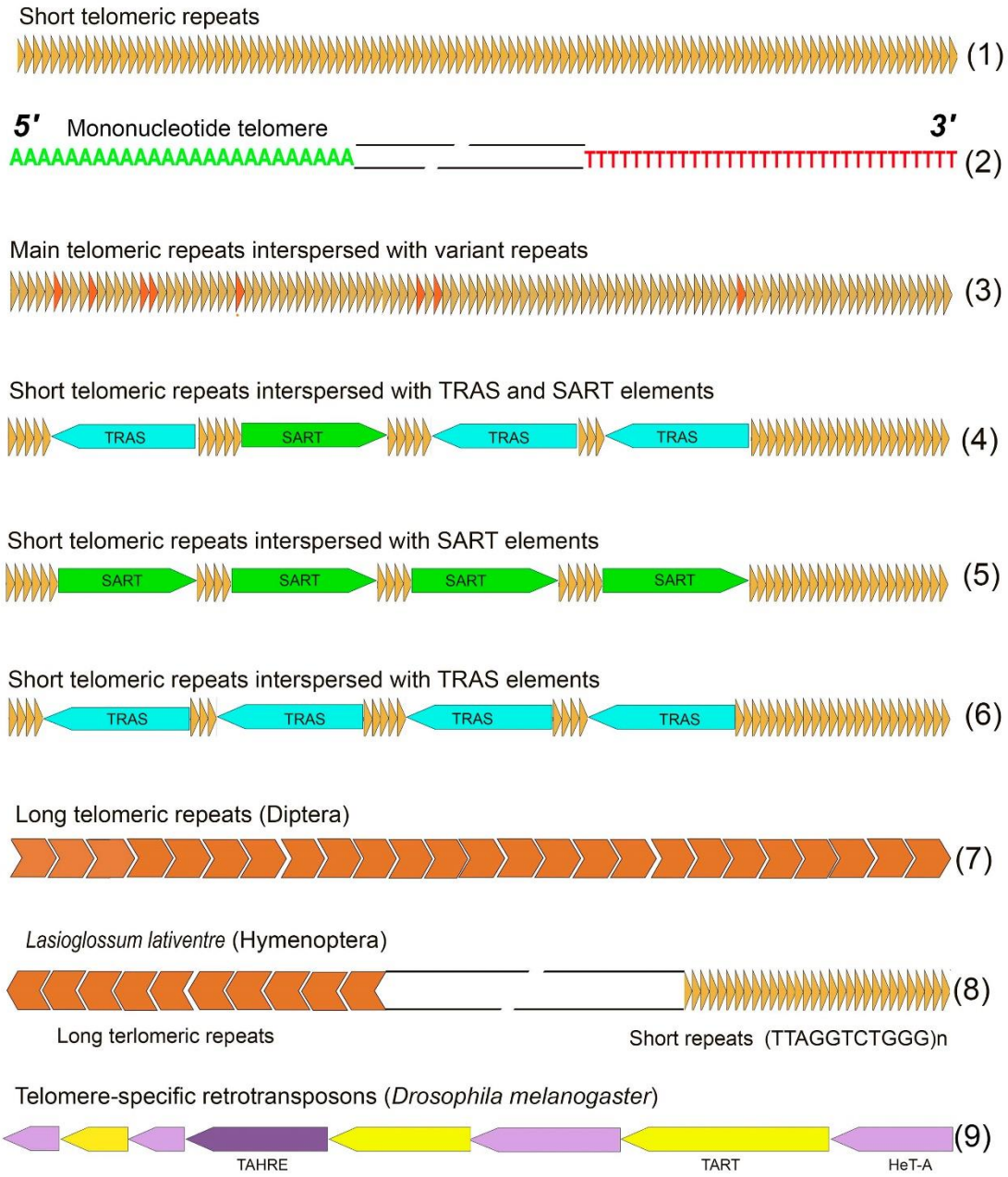

Figure 4. Types of telomere structures and telomere sequences in insects.

(1) short (5-17 bp) repeats (canonical telomere); (2) mononucleotide telomere, consisting of long (306-629 bp) $(A)_n$ tail at the 5′ chromosome end and $(T)_n$ tail at the 3′ chromosome end (*Amblyteles armatorius*, Ichneumonidae); (3) main short repeats interspersed with variant short repeats; (4) main short repeats interspersed with telomere-specific non-LTR retrotransposons of the *TRAS* and *SART* families; (5) main short repeats interspersed with telomere-specific non-LTR retrotransposons of the *SART* family; (6) main short repeats interspersed with telomere-specific non-LTR retrotransposons of the *TRAS* family; (7) long (173-374 bp) repeats, (8) one telomere consists of long repeats (similar to Diptera) and the other telomere if formed by arrays of short repeats TTAGGTCTGGG; (9) telomere-specific non-LTR retrotransposons of the *HeT-A*, *TAHRE* and *TART* families.

The study confirms that the pentanucleotide TTAGG motif is widespread among insects. However, telomeres consisting solely of the TTAGG repeats (type 1 in Fig. 4) turned out to be not as common (7 out of 180 species studied) as it is previously believed (Frydrychová et al., 2004; Vítková et al., 2005; Lukhtanov & Kuznetsova, 2010; Kuznetsova et al., 2020). Other types dominate or are found in six insect orders (Fig. 4): type 6 in Trichoptera, type 4 in Lepidoptera, type 5 in Coleoptera and Hemiptera, types 3 and 5 in Hymenoptera, and types 7 and 9 in Diptera. Interestingly, the last five orders are the richest in species number. Together they comprise over 800,000 described species, i.e. about 80% of the known insect species richness (Stork, 2018). The "non-canonical" telomeric organizations actually characterize a significant part (~40%) of all *ca*1900,000 described eukaryotic species (Chapman, 2009).

The most unexpected telomere sequence, consisting exclusively of the nucleotide T at the 3′ chromosome end, was found in the parasitic wasp *Amblyteles armatorius* (type 2 in Fig. 4). A common feature of all identified telomeric motifs (Table S1, Fig. 1) is the predominance of the nucleotide T. Telomeric motifs are generally thought to be saturated with G nucleotide (Osterhage & Friedman, 2009; Garavís et al., 2013); however, telomeric motifs without guanine (TTCCTC and T) were found in two parasitic wasp species (Fig. 1). Thus, different types of telomere organization and different variants of T-containing short motifs, including mononucleotide sequences, are compatible with the performance of telomere functions.

Therefore, the length and nucleotide composition of short telomeric motifs in themselves do not appear to be significant constrains in the evolution of telomeres. In this situation, when there is no mandatory selection for only one type of motif, and if telomerase is lost, as suggested by some theories (Mason et al., 2016), one would expect rapid evolution of telomere nucleotide sequences as a result of genetic drift/selection. However, this expectation is generally not met. In fact, the ancestral motif TTAGG is preserved in most arthropods over hundreds of millions of years of their evolution (Vítková et al, 2005; Kuznetsova et al., 2020).

Altered variants of telomeric motifs are also phylogenetically stable, although to a lesser degree. For example, in the order Hymenoptera, the modified 11 bp motifs characterize high-level clades, families, genera, and groups of species (Fig. 1). Such evolutionary stability of the motifs is easily explained by their maintenance through the telomerase-dependent mechanism: the telomeric template region of the telomerase RNA subunit is conservative (Chen et al., 2000; Červenák et al., 2021); therefore, its changes are extremely rare, and if they occur, they mark higher taxa.

On the other hand, in some Hymenoptera species, the detected high intra-telomeric and, even more so, inter-telomeric and inter-chromosomal variability, is hardly compatible with the telomerase-dependent mechanism of telomere length maintenance. It is hard to imagine that within the same organism there are many differentiated RNA templates that support the polymorphism of telomeric motifs. Such intraspecific variations seem to be more compatible with traditional Darwinian evolution, in which new motifs arise from nucleotide substitutions and then propagate and become fixed in individual chromosomes and genomes through genetic drift and/or selection.

Site-specific insertions of the *SART* and *TRAS*-retrotransposons into the telomeric region is found in most of the studied species, leading to the formation of multilayer telomeres (types 4-6 in Fig. 4) in five of eight studied orders. This variant of the "non-canonical" telomeric organization is very common in the order Hymenoptera, prevails in the order Coleoptera and is the only type in the order Lepidoptera. The prevalence of this "non-canonical" telomeric organization

in other insect orders remains to be studied, but it can be assumed to be common, given that it is already found in the orders (Hemiptera) (International Aphid Genomics Consortium, 2010) and (Phthiraptera) (Kirkness et al., 2010) and also outside the class Insecta in the mite *Tetranychus urticae* (Grbić et al., 2011). Interestingly, the multilayer telomeric organization correlates with low telomerase activity and loss of introns and GQ domains from TERT genes (Fujiwara, 2014). Most likely, the telomere length in these insects is partially maintained through transpositions and terminal recombination (Fujiwara et al., 2005; Osanai et al., 2006).

However, a pure *SART*/*TRAS*-based telomere maintenance mechanism seems impossible without contribution of the telomerase-based processes. The reason is that retrotransposons can integrate into telomeric regions in a highly specific manner. Therefore, if telomere motifs disappear (which is inevitable if telomerase activity is absent, due to shortening of telomeres in a series of cell divisions), this will also lead to the disappearance of the telomere-specific retrotransposons *SART* and *TRAS* (Fujiwara et al., 2005).

Single copies of unaltered or modified telomeric motifs are often found in subtelomeres. I hypothesize that the telomeric repeats can enter the subtelomeric region due to recombination and be modified there due to nucleotide substitutions. Through recombination, these modified repeats can re-enter telomeric regions, resulting in the structure, in which main short repeats interspersed with variant short repeats (type 3 in Fig. 4), as seen in some Hymenoptera.

With a complete loss of telomerase activity, there should be a loss of short telomeric motifs, and then, inevitably, a loss of the *SART* and *TRAS* elements. The intermediate stage of this process is observed in the bee *Lasioglossum lativentre* (loss of part of the telomeres and their replacement by subtelomeres, type 8 in Fig. 4), and the final stage is observed in Diptera (type 7 in Fig. 4).

Thus, the balance between telomerase-dependent and telomerase independent mechanisms and the exchange of sequences between telomeres and subtelomeres are most likely the key factors that determine the structure and evolution of telomeres in insects.

ACKNOWLEDGEMENTS

I thank Valentina Kuznetsova (Zoological Institute of the Russian Academy of Sciences) for numerous discussions on the organization of telomeres in insects over the past 10 years. I am grateful to Antonina Smorkacheva (St. Petersburg State University) for help in making drawings. This study was supported by Ministry of Science and Higher Education of the Russian Federation (grant no. 075-15-2021-1069) (analysis of telomere and subtelomere structure). The study of retrotransposons was supported by the Russian Science Foundation, grant no. 19-14-00202 to the Zoological Institute of the Russian Academy of Sciences.

DATA AVAILABITY

All the analyzed chromosome-level assemblies are available via the GenBank links provided (Table S1).

ORCID Vladimir A. Lukhtanov https://orcid.org/0000-0003-2856-2075

REFERENCES

Benetta E.D., Antoshechkin I., Yang T., Nguyen H.Q.M., Ferree P.M., Akbari O.S. 2020. Genome elimination mediated by gene expression from a selfish chromosome. Science Advanced 6(14). doi: 10.1126/sciadv.aaz9808

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27(2): 573–580. doi: 10.1093/nar/27.2.573.

Biessmann H., Mason J.M. 1997. Telomere maintenance without telomerase. Chromosoma 106: 63–69. doi: 10.1007/s0041 20050225

Biessmann H., Zurovcova M., Yao J. G., Lozovskaya E., Walter M. F. 2000. A telomeric satellite in *Drosophila virilis* and its sibling species. Chromosoma 109: 372–380. doi: 10.1007/s004120000094

Blackburn E.H., Gall J.G. 1978. A tandemly repeated sequence at the termini of extrachromosomal ribosomal RNA genes in *Tetrahymena*. Journal of Molecular Biology 120(1): 33–35. doi: 10.1016/0022-2836(78)90294-2.

Blackburn E.H. 1991. Structure and function of telomeres. Nature 350: 569–573. doi: 10.1038/350569a0

Brelsford A., Purcell J., Avril A., Van P.T., Zhang J., Brütsch T., Sundström L., Helanterä H., Chapuisat M. 2020. An ancient and eroded social supergene is widespread across *Formica* ants. Current Biology 30(2): 304–311. https://doi.org/10.1016/j.cub.2019.11.032

Cameron S.A., Hines H.M., Williams P.H. 2007. A comprehensive phylogeny of the bumble bees (*Bombus*). Biological Journal of the Linnean Society 91(1): 161–188. https://doi.org/10.1111/j.1095-8312.2007.00784.x

Casacuberta E., Pardue M. L. 2003. Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the *virilis* group. Proceedings of the National Academy of Sciences of the United States of America 100: 3363–3368. https ://doi.org/10.1073/pnas.02303 53100

Červenák F., Sepšiová R., Nosek J., Tomáška L. 2021. Step-by-step evolution of telomeres: lessons from yeasts. Genome Biology and Evolution 13(2): evaa268. https://doi.org/10.1093/gbe/evaa268

Chapman A. D. 2009. Numbers of living species in Australia and the World (2nd ed.). Canberra: Australian Biological Resources Study. pp. 1–80. ISBN 978-0-642-56861-8.

Chen JL, BlascoMA, Greider CW. 2000. Secondary structure of vertebrate telomerase RNA. Cell 100(5): 503–514.

Fajkus P., Peška V., Sitová Z., Fulnečková J., Dvořáčková M., Gogela R., Sýkorová E., Hapala J., Fajkus J. 2016. *Allium* telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. The Plant Journal 85(3): 337–347. doi: 10.1111/tpj.13115

Frydrychová R.Č., Mason J.M., Vratislav Peska V. 2021. Editorial: telomere flexibility and versatility: a role of telomeres in adaptive potential. Frontiers in Genetics. https://doi.org/10.3389/fgene.2021.771938

Frydrychova R., Marec F. 2002. Repeated losses of TTAGG telomere repeats in evolution of beetles (Coleoptera). Genetica 115: 179–187. https://doi.org/10.1023/A:1020175912128

Frydrychová R., Grossmann P., Trubač P., Vítková M., Marec F.2004. Phylogenetic distribution of TTAGG telomeric repeats in insects. Genome 47: 163–178. https://doi.org/10.1139/g03-100

Fujiwara H., Osanai M., Matsumoto T., Kojima K.K. 2005. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. Chromosome Research 13(5): 455–467. doi: 10.1007/s10577-005-0990-9

Fujiwara H. 2014. Accumulation of telomeric-repeat-specific retrotransposons in subtelomeres of *Bombyx mori* and *Tribolium castaneum*. In: Louis E., Becker M. (eds) Subtelomeres. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41566-1_13

Garavís M., González C., Villasante A. 2013. On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution. Genome Biology and Evolution 5(6): 1142–1150. https://doi.org/10.1093/gbe/evt079

Gokhman V.E., Anokhin B.A., Kuznetsova V.G. 2014. Distribution of 18S rDNA sites and absence of the canonical TTAGG insect telomeric repeat in parasitoid Hymenoptera. Genetica 142: 317–322. https://doi.org/10.1007/s10709-014-9776-3

Gokhman V.E., Kuznetsova V.G. 2018. Presence of the canonical TTAGG insect telomeric repeat in the Tenthredinidae (Symphyta) suggests its ancestral nature in the order Hymenoptera. Genetica 146: 341–344. https://doi.org/10.1007/s10709-018-0019-x

Golub N.V., Golub V.B., Kuznetsova V.G. 2015. Variability of 18rDNA loci in four lace bug species (Hemiptera, Tingidae) with the same chromosome number. Comparative Cytogenetics 9(4): 513–522. https://doi.org/10.3897/CompCytogen.v9i4.5376

Golub N.V., Golub V.B., Kuznetsova V.G. 2017. Distribution of the major rDNA loci among four hemipteran species of the family Tingidae (Heteroptera, Cimicomorpha). Folia Biologica (Kraków) 65: 155–158. https://doi.org/10.3409/fb65_3.155

Golub N.V., Golub V.B., Kuznetsova V.G. 2018. New data on karyotypes of lace bugs (Tingidae, Cimicomorpha, Hemiptera) with analysis of the 18S rDNA clusters distribution. Comparative Cytogenetics 12(4): 515–528. https://doi.org/10.3897/CompCytogen.v12i4.30431

Goubert C., Craig R.J., Bilat A.F., Peona V., Vogan A.A., Protasio A.V. 2022. A beginner′s guide to manual curation of transposable elements. Mobile DNA 13, 7. https://doi.org/10.1186/s13100-021-00259-7

Grbić M., Van Leeuwen T., Clark R. et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. Nature 479: 487–492. https://doi.org/10.1038/nature10640.

Greider C.W., Blackburn E.H. 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. Nature 337(6205): 331–337. https://doi.org/10.1038/337331a0

Grozeva S., Kuznetsova V., Anokhin B. 2011. Karyotypes, male meiosis and comparative FISH mapping of 18S ribosomal DNA and telomeric (TTAGG)n repeat in eight species of true bug (Hemiptera, Heteroptera). Comparative Cytogenetics 5(4): 355–374. https://doi.org/10.3897/CompCytogen.v5i4.2307

Hall T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

Kirkness E.F., Haas B.J., Sun W. et al. Pittendrigh B.R. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proceedings of the National Academy of Sciences of the United States of America 107(27): 12168–12173. https://doi.org/10.1073/pnas.1003379107

Klobutcher L.A., Swanton M.T., Donini P., Prescott D.M. 1981. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3′ teminus. Proceedings of the National Academy of Sciences of the United States of America 78(5): 3015–3019. Doi: 10.1073/pnas.78.5.3015

Korandová M., Krůček T., Vrbová K., Frydrychová R.Č. 2014. Distribution of TTAGG-specific telomerase activity in insects. Chromosome Research 22: 495–503. doi: 10.1007/s10577-014-9436-6

Kubo Y., Okazaki S., Anzai T., Fujiwara H. 2001. Structural and phylogenetic analysis of *TRAS*, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. Molecular Biology and Evolution 18(5): 848–857. doi: 10.1093/oxfordjournals.molbev.a003866

Kuznetsova V., Grozeva S., Gokhman V. 2020. Telomere structure in insects: A review. Journal of Zoological Systematics and Evolutionary Research 58: 127–158. doi: 10.1111/jzs.12332

Kuznetsova V., Maryańska-Nadachowska A., Anokhin B., Shapoval N., Shapoval A. 2021. Chromosomal analysis of eight species of dragonflies (Anisoptera) and damselflies (Zygoptera) using conventional cytogenetics and fluorescence in situ hybridization: Insights into the karyotype evolution of the ancient insect order Odonata. Journal of Zoological Systematics and Evolutionary Research 59: 387–399. doi: 10.1111/jzs.12429

Liu C., Schroeder A.A., Kapur V., Abrahamsen M.S. 1998. Telomeric sequences of *Cryptosporidium parvum*. Molecular and biochemical parasitology 94(2): 291–296. doi: 10.1016/s0166-6851(98)00072-3

Logeswaran D., Li Y., Podlevsky J.D., Chen J. 2021. Monophyletic origin and divergent evolution of animal telomerase RNA. Molecular Biology and Evolution 38(1): 215–228. https://doi.org/10.1093/molbev/msaa203

Louis E.J., Vershinin A.V. 2005. Chromosome ends: different sequences may provide conserved functions. BioEssays 27(7): 685–697. doi: 10.1002/bies.20259.

Louis E.J., Becker M.M. (eds.). 2014. Subtelomeres. Springer Verlag: Berlin – Heidelberg. doi: 10.1007/978-3-642-41566-1_1

Lukhtanov V.A., Kuznetsova V.G. 2010. What genes and chromosomes say about the origin and evolution of insects and other arthropods? Russian Journal of Genetics 46: 1115–1121. https://doi.org/10.1134/S1022 79541 0090279

Mandrioli M., Monti V., Manicardi G.C. 2012. Starting at the end: Telomeres and telomerase in arthropods. BioMolecular Concepts 3: 465–470. https://doi.org/10.1515/bmc-2012-0008

Mason J.M., Reddy H.M., Frydrychova R.C. 2011. Telomere maintenance in organisms without telomerase. In book: DNA Replication-Current Advances. Herve Seligmann (Editor) Publisher: Intechopen p. 323–346. doi: 10.13140/2.1.2417.1525

Mason J.M., Frydrychova R.C., Biessmann H. 2008. *Drosophila* telomeres: an exception providing new insights. BioEssays 30(1): 25–37. doi:10.1002/bies.20688

Mason J.M., Randall T.A., Frydrychova R.C. 2016. Telomerase lost? Chromosoma 125: 65–73. https://doi.org/10.1007/s00412-015-0528-7

Menezes R.S.T., Bardella V.B., Cabral-de-Mello D.C., Lucena D.A.A., Almeida E.A.B. 2017. Are the TTAGG and TTAGGG telomeric repeats phylogenetically conserved in aculeate Hymenoptera? The Science of Nature 104, 85. https://doi.org/10.1007/s00114-017-1507-z

Nandakumar J., Cech T.R. 2013. Finding the end: recruitment of telomerase to the telomere. Nature Reviews Molecular Cell Biology 14(2): 69–82. doi:10.1038/nrm3505.

Nielsen L., Edström J.E. 1993. Complex telomere-associated repeat units in members of the genus *Chironomus* evolve from sequences similar to simple telomeric repeats. Molecular and Cell Biology 13: 1583–1589. https://doi.org/10.1128/MCB.13.3.1583

Okazaki S., Ishikawa H., Fujiwara H. 1995. Structural analysis of *TRAS1*, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. Molecular and Cellular Biology 15(8): 4545–4552. doi: 10.1128/MCB.15.8.4545

Olson P.D., Tracey A., Baillie A., James K., Doyle S.R., Buddenborg S.K., Rodgers F.H., Holroyd N., Berriman M. 2020. Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. BMC Biology 18:165. https://doi.org/10.1186/s12915-020-00899-w

Osterhage J.L., Friedman K.L. 2009. Chromosome end maintenance by telomerase. The Journal of Biological Chemistry 284(24): 16061–16065. doi: 10.1074/jbc.R900011200

Osanai M., Kojima K.K., Futahashi R., Yaguchi S., Fujiwara H. 2006. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). Gene 376(2): 281–289. https://doi.org/10.1016/j.gene.2006.04.022

Peška V., Fajkus P., Fojtová M., Dvořáčková M., Hapala J., Dvořáček V., Polanská P., Leitch A.R., Sýkorová E., Fajkus J. 2015. Characterisation of an unusual telomere motif (TTTTTTAGGG)n in the plant *Cestrum elegans* (Solanaceae), a species with large genome. The Plant Journal 82(4):644–654. doi: 10.1111/tpj.12839

Peška V., Garcia S. 2020. Origin, diversity, and evolution of telomere sequences in plants. Frontiers in Plant Science 11: 117. doi: 10.3389/fpls.2020.00117

Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., Kozlov A., Podsiadlowski L., Petersen M., Lanfear R., Diez P.A., Heraty J., Kjer K.M., Klopfstein S., Meier R., Polidori C., Schmitt T., Liu S., Zhou X., Wappler T., Rust J., Misof B., Niehuis O. 2017. Evolutionary history of the Hymenoptera. Current Biology 27(7): 1013–1018. doi: 10.1016/j.cub.2017.01.027.

Prušáková D., Peska V., Pekár S., Bubeník M., Čížek L., Bezděk A., Frydrychová R.Č. 2021. Telomeric DNA sequences in beetle taxa vary with species richness. Scientific Reports 11(1), 13319 https://doi.org/10.1038/s41598-021-92705-y

Rosén M., Edström J.E. 2002. Chromosome ends in *Chironomus tentans* do not have long single-stranded overhangs characterizing canonical telomeres. Chromosome Research 10(1): 21–31. https://doi.org/10.1023/A:1014257808705

Saint-Leandre B., Levine M.T. 2020. The telomere paradox: stable genome preservation with rapidly evolving proteins. Trends in Genetics 36(4): 232–242. https://doi.org/10.1016/j.tig.2020.01.007

Sann M., Meusemann K., Niehuis O., Mokrousov M., Ohl M., Pauli T., Schmid-Egger C. 2021. Reanalysis of the apoid wasp phylogeny with additional taxa and sequence data confirms the placement of Ammoplanidae as sister to bees. Systematic Entomology 46: 558–569. doi: 10.1111/syen.12475

Sann M., Niehuis O., Peters R.S., Mayer C., Kozlov A., Podsiadlowski L., Bank S., Meusemann K., Misof B., Bleidorn C., Ohl M. 2018. Phylogenomic analysis of Apoidea sheds new light on the sister group of bees. BMC Evolutionary Biology 18(1): 71. doi: 10.1186/s12862-018-1155-8.

Shay J.W., Wright W.E. 2019. Telomeres and telomerase: three decades of progress. Nature Review Genetics 20: 299–309. doi: 10.1038/s41576-019-0099-1

Shirley M.W. 1994. The genome of *Eimeria tenella*: further studies on its molecular organisation. Parasitology Research 80(5): 366–373. doi: 10.1007/BF00932373.

Sohanpal B.K., Morzaria S.P., Gobright E.I., Bishop R.P. 1995. Characterisation of the telomeres at opposite ends of a 3 Mb *Theileria parva* chromosome. Nucleic Acids Reseach 23(11): 1942–1947.

Spence J.M., Blackman R.L., Testa J.M., Ready P.D. 1998. A 169 bp tandem repeat DNA marker for subtelomeric heterochromatin and chromosomal rearrangements in aphids of the *Myzus persicae* group. Chromosome Research 6(3): 167–175. doi: 10.1023/a:1009251415941.

Stork N. 2018. How many species of insects and other terrestrial arthropods are there on Earth? Annual Review of Entomology 63: 31–45. doi:10.1146/annurev-ento-020117-043348

Takahashi H., Okazaki S., Fujiwara H. 1997. A new family of sitespecific retrotransposons, *SART1*, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. Nucleic Acids Research 25(8): 1578–1584. doi: 10.1093/nar/25.8.1578.

The Darwin Tree of Life Project Consortium, Barnes, I., Berriman, M., Broad, G., Durbin, R., Flicek, P., Gaya, E., Hall, N., Hart, M., Holland, P., Hollingsworth, P., Howe, K., Kersey, P., Lawniczak, M. K. N., Lewis, O., Mieszkowska, N., Richards, T., Twyford, A. D., & Wilson, W. (2022). Sequence locally, think globally: The Darwin tree of life project. *Proceedings of the National Academy of Sciences* of the United States of America, 119(4), [e2115642118]. https://doi.org/10.1073/pnas.2115642118

Vershinina A.O., Anokhin B.A., Lukhtanov V.A. 2015. Ribosomal DNA clusters and telomeric (TTAGG)n repeats in blue butterflies (Lepidoptera, Lycaenidae) with low and high chromosome numbers. Comparative Cytogenetics 9(2): 161–171. https ://doi.org/10.3897/CompCytogen.v9i2.4715

Vítková M., Král J., Traut W., Zrzavý J., Marec F. 2005. The evolutionary origin of insect telomeric repeats, (TTAGG)n. Chromosome Research 13(2): 145–156. https ://doi.org/10.1007/s10577-005-7721-0

Wallberg A., Bunikis I., Pettersson O.V., Mosbech M.B., Childers A.K., Evans J.D., Mikheyev A.S., Robertson H.M., Robinson G.E., Webster M.T.2019. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. BMC Genomics 20(1): 275. doi: 10.1186/s12864-019-5642-0.

Watson J.M., Trieb J., Troestl M., Renfrew K., Mandakova T., Fulnecek J., Shippen D.E., Riha K. 2021. A hypomorphic allele of telomerase uncovers the minimal functional length of telomeres in *Arabidopsis*. Genetics 219(2), iyab126. doi: 10.1093/genetics/iyab126

Yan Z., Martin S.H., Gotzek D., Arsenault S.V., Duchen P., Helleu Q., Riba-Grognuz O., Hunt B.G., Salamin N., Shoemaker D., Ross K.G., Keller L. 2020. Evolution of a supergene that regulates a trans-species social polymorphism. Nature Ecology and Evolution 4(2): 240–249. doi: 10.1038/s41559-019-1081-1.

Zakian V.A. 1995. Telomeres: Beginning to understand the end. Science 270: 1601–1607. https://doi.org/10.1126/scien ce.270.5242.1601

Appendix

Table S1. Telomeric motifs and other sequences found in telomeric and subtelomeric regions in chromosomes of the studied insect species.

Order, Family	Species	Telomere motif (repeats found at the most terminal positions of chromosome ends)	Sequences adjacent to the most terminal chromosomal array	GenBank Accession #
ODONATA				
Zygoptera: Platycnemididae	*Platycnemis pennipes*	**TTGGG**	Combination of various repeats with prevalence of 129, 286-289 and 434 bp long repeats	OW121847.1-OW121846.1
Zygoptera: Coenagrionidae	*Ischnura elegans*	There is no short telomeric motif. Various sequences, including low-copy repeats of various length in autosomes. 175 bp repeats in X chromosome.s		OV121100.1-OV121106.1
PLECOPTERA				
Nemouridae	*Nemoura dubitans*	**TTAGG**	Various sequences including short repeat AACGG	OV121074.1- OV121073.1
HEMIPTERA				
Pentatomidae	*Aelia acuminata*	**TTAGGGATGG**	Combination of various short, medium and long repeats	OU426978.1-OU426985.1
Acanthosomatidae	*Acanthosoma haemorrhoidale*	**TTAGGGTGGT**	Combination of various short, medium and long repeats or SART - (TTAGGGTGGT)n	OV884011.1- OV884010.1
LEPIDOPTERA				
Blastobasidae	*Blastobasis adustella*	**TTAGG**	SART - (TTAGG)n - TRAS	OU026114.1-OU026144.1
Blastobasidae	*Blastobasis lacticolella*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990039.1-LR990069.1
Cossidae	*Zeuzera pyrina*	**TTAGG**	SART - (TTAGG)n - TRAS	OU015617.1-OU015648.1
Crambidae	*Chrysoteuchia culmella*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342641.1-OU342672.1
Crambidae	*Parapoynx stratiotata*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342457.1-OU342487.1
Drepanidae	*Habrosyne pyritoides*	**TTAGG**	SART - (TTAGG)n - TRAS	OU015585.1-OU015616.1
Drepanidae	*Thyatira batis*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990485.1-LR990516.1
Erebidae	*Eilema depressum*	**TTAGG**	SART - (TTAGG)n - TRAS	OU612012.1-OU612042.1

Erebidae	*Eilema sororculum*	**TTAGG**	SART - (TTAGG)n - TRAS	OU618532.1-OU618562.1
Erebidae	*Euproctis similis*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990103.1-LR990125.1
Erebidae	*Hypena proboscidalis*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990126.1-LR990157.1
Erebidae	*Laspeyria flexula*	**TTAGG**	SART - (TTAGG)n - TRAS	LR989949.1-LR989980.1
Erebidae	*Lymantria monacha*	**TTAGG**	SART - (TTAGG)n - TRAS	LR991081.1-LR991109.1
Erebidae	*Orgyia antiqua*	**TTAGG**	SART - (TTAGG)n - TRAS	OU779861.1-OU779875.1
Erebidae	*Schrankia costaestrigalis*	**TTAGG**	SART - (TTAGG)n - TRAS	FR997825.1-FR997856.1
Erebidae	*Spilarctia lutea*	**TTAGG**	SART - (TTAGG)n - TRAS	OU696470.1-OU696502.1
Erebidae	*Spilosoma lubricipeda*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992274.1-HG992304.1
Geomatridae	*Agriopis aurantiaria*	**TTAGG**	SART - (TTAGG)n - TRAS	OU611981.1-OU612011.1
Geometridae	*Biston betularia*	**TTAGG**	SART - (TTAGG)n - TRAS	FR989862.1-FR989893.1
Geometridae	*Campaea margaritaria*	**TTAGG**	SART - (TTAGG)n - TRAS	OU538788.1-OU538819.1
Geometridae	*Crocallis elinguaria*	**TTAGG**	SART - (TTAGG)n - TRAS	OU026065.1-OU026082.1
Geometridae	*Ennomos fuscantarius*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992023.1-HG992054.1
Geometridae	*Ennomos quercinarius*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342488.1-OU342519.1
Geometridae	*Erannis defoliaria*	**TTAGG**	SART - (TTAGG)n - TRAS	FR990066.1-FR990095.1
Geometridae	*Hydriomena furcata*	**TTAGG**	SART - (TTAGG)n - TRAS	OU538820.1-OU538848.1
Geometridae	*Hylaea fasciaria*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990192.1-LR990223.1
Geometridae	*Idaea aversata*	**TTAGG**	SART - (TTAGG)n - TRAS	OU026083.1-OU026113.1
Geometridae	*Peribatodes rhomboidaria*	**TTAGG**	SART - (TTAGG)n - TRAS	OU452166.1-OU452197.1
Hesperiidae	*Erynnis tages*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990071.1-LR990102.1
Hesperiidae	*Hesperia comma*	**TTAGG**	(TTAGG)n - TRAS	FR990012.1-FR990041.1
Hesperiidae	*Ochlodes sylvanus*	**TTAGG**	SART - (TTAGG)n - TRAS	FR990122.1-FR990152.1
Hesperiidae	*Pyrgus malvae*	**TTAGG**	SART - (TTAGG)n - TRAS	OU426946.1-OU426977.1
Hesperiidae	*Thymelicus sylvestris*	**TTAGG**	SART - (TTAGG)n - TRAS	OU426885.1-OU426913.1
Lycaenidae	*Aricia agestis*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990256.1-LR990279.1
Lycaenidae	*Celastrina argiolus*	**TTAGG**	SART - (TTAGG)n - TRAS	LR994577.1-LR994603.1
Lycaenidae	*Cyaniris semiargus*	**TTAGG**	SART - (TTAGG)n - TRAS	LR994546.1-LR994570.1
Lycaenidae	*Glaucopsyche alexis*	**TTAGG**	SART - (TTAGG)n - TRAS	FR990042.1-FR990065.1
Lycaenidae	*Lycaena phlaeas*	**TTAGG**	SART - (TTAGG)n - TRAS	HG995163.1-HG995187.1
Lycaenidae	*Lysandra bellargus*	**TTAGG**	SART - (TTAGG)n - TRAS	HG995319.1-HG995365.1
Lycaenidae	*Lysandra coridon*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992055.1-HG992145.1
Lycaenidae	*Plebejus argus*	**TTAGG**	SART - (TTAGG)n - TRAS	FR989926.1-FR989949.1
Noctuidae	*Abrostola tripartita*	**TTAGG**	SART - (TTAGG)n - TRAS	HG996486.1-HG996517.1
Noctuidae	*Acronicta aceris*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342756.1-OU342788.1
Noctuidae	*Agrochola circellaris*	**TTAGG**	SART - (TTAGG)n - TRAS	OU611839.1-OU611869.1
Noctuidae	*Amphipyra berbera*	**TTAGG**	SART - (TTAGG)n - TRAS	OU343121.1-OU343152.1
Noctuidae	*Amphipyra tragopoginis*	**TTAGG**	SART - (TTAGG)n - TRAS	HG991991.1-HG992022.1
Noctuidae	*Apamea monoglypha*	**TTAGG**	SART - (TTAGG)n - TRAS	OU426914.1-OU426945.1
Noctuidae	*Atethmia centrago*	**TTAGG**	SART - (TTAGG)n - TRAS	HG995366.1-HG995396.1
Noctuidae	*Autographa gamma*	**TTAGG**	SART - (TTAGG)n - TRAS	LR989849.1-LR989881.1
Noctuidae	*Autographa pulchrina*	**TTAGG**	SART - (TTAGG)n - TRAS	FR997761.1-FR997793.1
Noctuidae	*Cosmia trapezina*	**TTAGG**	SART - (TTAGG)n - TRAS	LR991019.1-LR991051.1
Noctuidae	*Craniophora ligustri*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990954.1-LR990986.1
Noctuidae	*Dryobotodes eremita*	**TTAGG**	SART - (TTAGG)n - TRAS	OU823241.1-OU823273.1

Noctuidae	*Eupsilia transversa*	**TTAGG**	SART - (TTAGG)n - TRAS	OU611871.1-OU611903.1
Noctuidae	*Griposia aprilina*	**TTAGG**	SART - (TTAGG)n - TRAS	OU744283.1-OU744315.1
Noctuidae	*Hecatera dysodea*	**TTAGG**	SART - (TTAGG)n - TRAS	HG995286.1-HG995318.1
Noctuidae	*Hydraecia micacea*	**TTAGG**	SART - (TTAGG)n - TRAS	OU611774.1-OU611806.1
Noctuidae	*Mamestra brassicae*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990987.1-LR991018.1
Noctuidae	*Mesoligia furuncula*	**TTAGG**	SART - (TTAGG)n - TRAS	OU744790.1-OU744821.1
Noctuidae	*Mythimna ferrago*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342673.1-OU342705.1
Noctuidae	*Mythimna impura*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990339.1-LR990371.1
Noctuidae	*Noctua fimbriata*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990921.1-LR990953.1
Noctuidae	*Noctua janthe*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342552.1-OU342583.1
Noctuidae	*Noctua pronuba*	**TTAGG**	SART - (TTAGG)n - TRAS	LR999891.1-LR999923.1
Noctuidae	*Ochropleura plecta*	**TTAGG**	SART - (TTAGG)n - TRAS	FR997721.1-FR997753.1
Noctuidae	*Omphaloscelis lunosa*	**TTAGG**	SART - (TTAGG)n - TRAS	OU744243.1-OU744275.1
Noctuidae	*Phlogophora meticulosa*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990517.1-LR990548.1
Noctuidae	*Xestia c-nigrum*	**TTAGG**	SART - (TTAGG)n - TRAS	OU745243.1-OU745274.1
Noctuidae	*Xestia xanthographa*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990641.1-LR990672.1
Notodontidae	*Clostera curtula*	**TTAGG**	SART - (TTAGG)n - TRAS	FR997794.1-FR997824.1
Notodontidae	*Furcula furcula*	**TTAGG**	SART - (TTAGG)n - TRAS	OU452242.1-OU452271.1
Notodontidae	*Notodonta dromedarius*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990159.1-LR990190.1
Notodontidae	*Phalera bucephala*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990609.1-LR990640.1
Notodontidae	*Pheosia gnoma*	**TTAGG**	SART - (TTAGG)n - TRAS	FR989894.1-FR989925.1
Notodontidae	*Pheosia tremula*	**TTAGG**	SART - (TTAGG)n - TRAS	HG995397.1-HG995428.1
Notodontidae	*Ptilodon capucinus*	**TTAGG**	SART - (TTAGG)n - TRAS	OU611807.1-OU611838.1
Nymphalidae	*Boloria selene*	**TTAGG**	SART - (TTAGG)n - TRAS	HG993131.1-HG993161.1
Nymphalidae	*Erebia ligea*	**TTAGG**	SART - (TTAGG)n - TRAS	OU785219.1-OU785248.1
Nymphalidae	*Fabriciana adippe*	**TTAGG**	SART - (TTAGG)n - TRAS	FR989982.1-FR990011.1
Nymphalidae	*Limenitis camilla*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990224.1-LR990255.1
Nymphalidae	*Maniola jurtina*	**TTAGG**	(TTAGG)n - TRAS	HG995207.1-HG995237.1
Nymphalidae	*Melitaea cinxia*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992209.1-HG992240.1
Nymphalidae	*Mellicta athalia*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992176.1-HG992208.1
Nymphalidae	*Nymphalis io*	**TTAGG**	SART - (TTAGG)n	LR989895.1-LR989926.1
Nymphalidae	*Nymphalis polychloros*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992241.1-HG992273.1
Nymphalidae	*Nymphalis urticae*	**TTAGG**	SART - (TTAGG)n - TRAS	LR989982.1-LR990014.1
Nymphalidae	*Pararge aegeria*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990891.1-LR990920.1
Nymphalidae	*Vanessa atalanta*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990549.1-LR990581.1
Nymphalidae	*Vanessa cardui*	**TTAGG**	SART - (TTAGG)n - TRAS	LR999924.1-LR999956.1
Peleopodidae	*Carcina quercana*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342426.1-OU342456.1
Pieridae	*Anthocharis cardamines*	**TTAGG**	SART - (TTAGG)n - TRAS	FR989950.1-FR989981.1
Pieridae	*Aporia crataegi*	**TTAGG**	SART - (TTAGG)n - TRAS	OU538729.1-OU538755.1
Pieridae	*Colias croceus*	**TTAGG**	SART - (TTAGG)n - TRAS	HG991958.1-HG991990.1
Pieridae	*Pieris brassicae*	**TTAGG**	SART - (TTAGG)n - TRAS	LR989932.1-LR989948.1
Pieridae	*Pieris napi*	**TTAGG**	SART - (TTAGG)n - TRAS	HG993162.1-HG993187.1
Pterophoridae	*Emmelina monodactyla*	**TTAGG**	SART - (TTAGG)n - TRAS	OU745296.1-OU745326.1
Pyralidae	*Endotricha flammealis*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990852.1-LR990884.1

Sesiidae	*Bembecia ichneumoniformis*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342520.1-OU342551.1
Sesiidae	*Sesia apiformis*	**TTAGG**	SART - (TTAGG)n - TRAS	OU611947.1-OU611978.1
Sesiidae	*Synanthedon vespiformis*	**TTAGG**	SART - (TTAGG)n - TRAS	OU906945.1-OU906976.1
Sphingidae	*Deilephila porcellus*	**TTAGG**	SART - (TTAGG)n - TRAS	LR999970.1-LR999999.1
Sphingidae	*Hemaris fuciformis*	**TTAGG**	SART - (TTAGG)n - TRAS	OU015431.1-OU015461.1
Sphingidae	*Laothoe populi*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992146.1-HG992175.1
Sphingidae	*Mimas tiliae*	**TTAGG**	SART - (TTAGG)n - TRAS	HG995238.1-HG995267.1
Tineidae	*Tinea semifulvella*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342584.1-OU342629.1
Tineidae	*Tinea trinotella*	**TTAGG**	SART - (TTAGG)n - TRAS	HG992305.1-HG992335.1
Tortricidae	*Apotomis turbidana*	**TTAGG**	SART - (TTAGG)n - TRAS	LR990280.1-LR990308.1
Tortricidae	*Cydia splendana*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342871.1-OU342899.1
Tortricidae	*Hedya salicella*	**TTAGG**	SART - (TTAGG)n - TRAS	FR990096.1-FR990121.1
Tortricidae	*Notocelia uddmanniana*	**TTAGG**	SART - (TTAGG)n - TRAS	LR991052.1-LR991080.1
Tortricidae	*Pammene fasciana*	**TTAGG**	SART - (TTAGG)n - TRAS	OU452272.1-OU452300.1
Ypsolophidae	*Ypsolopha scabrella*	**TTAGG**	SART - (TTAGG)n - TRAS	OU342960.1-OU342991.1
Zygaenidae	*Zygaena filipendulae*	**TTAGG**	SART - (TTAGG)n - TRAS	OU015649.1-OU015680.1
TRICHOPTERA				
Limnephilidae	*Limnephilus lunatus*	**TTAGG**	Various sequences with predominance of 166 bp and 22 bp repeats or (TTAGG)n – TRAS structure	OU830592.1-OU830604.1
Limnephilidae	*Limnephilus rhombicus*	**TTAGG**	(TTAGG)n – TRAS	OV815270.1- OV815276.1
Limnephilidae	*Limnephilus marmoratus*	**TTAGG**	(TTAGG)n – TRAS	OU862906.1-OU862920.1
COLEOPTERA				
Attelabidae	*Apoderus coryli*	**TTGGG**	10 bp short variable subtelomeric repeats (TTTGACTCAC, TTTTACTCAC, TTTGACTCAT, TTTGACTCTC, or similar); or short variable subtelomeric repeats TTTAATTA (and similar); or long repeats (chromosome 6, right), or different repeats, or SART (chromosome 6 left). In the chromosomes 6, 7, 11, 12, 14, 16, 18, subtelomeric sequences are found at the very ends of the chromosomes (as telomeric elements)	OU452198.1-OU452217.1
Cantharidae	*Cantharis rustica*	**TTAGG**	SART - (TTAGG)n	OU426877.1-OU426883.1
Cantharidae	*Rhagonycha fulva*	**TTAGG**	SART - (TTAGG)n	HG996554.1-HG996560.1
Carabidae	*Leistus spinibarbis*	**TTAGG**	SART - (TTAGG)n	OW121791.1-OW121792.1
Carabidae	*Pterostichus madidus*	**TTAGG**	SART - (TTAGG)n	OU452301.1-OU452319.1
Coccinellidae	*Adalia bipunctata*	**TTAGG**	SART - (TTAGG)n	OU342948.1-OU342958.1
Coccinellidae	*Coccinella septempunctata*	**TTAGG**	SART - (TTAGG)n (found only in the right telomere of the chromosome 8)	OU015573.1-OU015581.1
Coccinellidae	*Harmonia axyridis*	**TTAGG**	SART - (TTAGG)n	OU611927.1-OU611934.1
Elateridae	*Agrypnus murinus*	**TTAGG**		OV816024.1- OV816023.1
Melyridae	*Malachius bipustulatus*	**TCAGG**	Various repeats	OU342630.1-OU342639.1
Pyrochroidae	*Pyrochroa serraticornis*	**TCAGG**	SART - (TCAGG)n	HG995152.1-HG995161.1
Staphylinidae	*Ocypus olens*	**TTAGG**	SART - (TTAGG)n	OU343047.1-OU343065.1
Staphylinidae	*Philonthus cognatus*	**TTAGG interspersed with TTTAGG**	SART - (TTAGG)	OW052241.1-OW052250.1

HYMENOPTERA				
Andrenidae	*Andrena dorsata*	**TTAGTTTGGG**		OV815487.1- OV815489.1
Andrenidae	*Andrena haemorrhoa*	**TTAGGTCTGGG**	(TTAGGTCTGGG)n - combination of different short and medium-sized repeats - (TTAGGTCTGGG)n, TTAGGTCTGGG - SART - TTAGGTCTGGG, combination of different short and medium-sized repeats	OU342940.1-OU342946.1
Andrenidae	*Andrena minutula*	**TTAGGG**	Various short and medium repeats	OV815993.1- OV815999.1
Apidae	*Apis mellifera*	The main repeat **TTAGG** interspersed with variant repeats TTAGGGTT, TTATAGG	15 bp repeat $(TTAGGTCAGGCTGGG)_n$, 20 bp repeat $(TTAGGCTAGGTCAGGCTGGG)_n$ (chromosome 1, left telomere)	CM009931.2 - CM009946.2
Apidae	*Bombus campestris*	The telomere consists of the main repeat **TTAGGTTGGGG** interspersed with variant repeats TTAGGTGGGGG, TTAGGTTGGTGGGG, TTAGGTTGGTTAGTGGGG, TTAGGTTGGGTAGGTTTGG, TTAGGTTGGGT, TTTGGGTTGGGG	SART - (TTAGGTTGGGG)n	HG995126.1-HG995150.1
Apidae	*Bombus hortorum*	**TTAGGTTGGGG**	SART - (TTAGGTTGGGG)n	HG995188.1-HG995205.1
Apidae	*Bombus hypnorum*	**TTAGGTTGGGG**	SART - (TTAGGTTGGGG)n	OU427020.1-OU427031.1
Apidae	*Bombus pascuorum*	**TTAGGTTGGGG**	SART - (TTAGGTTGGGG)n	HG995268.1-HG995284.1
Apidae	*Bombus pratorum*	**TTAGGTTGGGG**	SART - (TTAGGTTGGGG)n	OV883983.1-OV884000.1
Apidae	*Bombus sylvestris*	**TTAGGTTCGGG** (majority) interspersed with TTAGGTTAGGTTCGGG	Various repeats with significant inclusion of the 41 bp motif TTCCAAATCTGTTCAGATCCATCCCGGTTCAGTCCAGACCG (and its variations)	OU443141.1-OU443164.1
Apidae	*Nomada fabriciana*	**TTAGG**	Various short, medium and long repeats	OU015688.1-OU015699.1
Bembicidae	*Nysson spinosus*	**TTAGGTCTGGG**	Combination of various short and medium repeats	OU342830.1-OU342855.1
Crabronidae	*Ectemnius continuus*	**TTAGGTCTGGG**	Combination of various short and medium repeats	OU342856.1-OU342869.1
Crabronidae	*Ectemnius lituratus*	**TTAGGTCTGGG**	TTAGGTGTGGG, TTTTCCTGGAA, TTTTCCTGGAATTTCCCGCGG (and other variations)	OU343033.1-OU343045.1
Formicidae	*Formica selysi*	**TTAGG**	TTAGCACGGGGTTACGG, TTAGGGTTAGCATGGGG, TTTGGTCAGGCTAGGCAAGGCAATGTCAGA (and other variations)	CM020805.1-CM020831.1
Formicidae	*Solenopsis invicta*	**TTAGG**	TTTGGTTATGGTTT, TTTGGTTATGGTCC, TTTGGTTATGGTCT (and other variations)	CM028732.1-CM028747.1
Halictidae	*Lasioglossum lativentre*	**TTAGGTCTGGG** (see Table 1)	see Table 1	OU744355.1-OU744368.1
Halictidae	*Lasioglossum morio*	**TTAGG**	22 bp motif TTAGGTTAGCGGATTCCGGG (variant TTAGGTTAGAGCGATTTCCGAG in the chromosome 1); however, the chromosome 7 has another subtelomere organization: combination of various short and medium motifs just before telomere motif (TTAGG)n	OU744323.1-OU744334.1
Halictidae	*Lasioglossum pauxillum*	**TTAGG**	SART - (TTAGG)n	OW121699.1-OW121707.1
Halictidae	*Seladonia tumulorum*	**TTAGGG**	Combination of various short, medium and long repeats with prevalence of TTTTTGCGTCCGTAGAAGG, TTTTTCGAGGTCGGGGTGGT,	OU565266.1-OU565282.1

			TTTTTGCGTCCGTAGAAGGTTG (and similar)	
Halictidae	*Sphecodes monilicornis*	**TTAGGTCTGGG**	TTAGGCCCTGCTCGG, TTAGGCCTGCTCGGT	OU565284.1-OU565302.1
Ichneumonidae	*Amblyteles armatorius*	**T**	18 bp subtelomeric repeat CAGCGCCCACAGCGTCGG and similar repeats	OW121686.1-OW121697.1
Ichneumonidae	*Buathra laborator*	**TTCCTC**	Various sequences	OW203892.1-OW203902.1
Ichneumonidae	*Ichneumon xanthorius*	**TTAAAACGCC**, TTGAAACGCC, TTAAAGCGCC, TGGATGGGA, TTAAGGCGTT, AAAAAATTCTTTG ATGC (see also text)	Various short, medium and long repeats	OU824200.1-OU824211.1
Melittidae	*Macropis europaea*	The motives **TTGGG** and **TTAGG** are mixed. The TTGGG motif prevails. The TTAGG motif almost never occurs twice in a row. TTGGG motif up to 13 times in a row.	121bp repeat	OU744343.1-OU744353.1
Pemphredonidae	*Mimumesa dahlbomi*	**TTAGG**	Combination of various short and medium repeats; SART - (TTAGG)n (chromosome 9)	OU824113.1-OU824130.1
Pemphredonidae	*Pemphredon lugubris*	**TCTGGG**	SART - (TTAGG)n	OW121815.1-OW121819.1
Philanthidae	*Cerceris rybyensis*	**TTAGGTCTGGG**	TTAGGTGGTTGGACATG, TTAGGTGGGTGGACATG, TTAGGTGGTTGGACATG	OU342789.1-OU342802.1
Pompilidae	*Anoplius nigerrimus*	**TTAGGTCTGGG**	Combination of various short, medium and long repeats	OU612077.1-OU612091.1
Pteromalidae	*Nasonia vitripennis*	Within the main motif **TTATTGGG**, which obviously prevails, modified sequences are regularly found - TTATTGGGGG, TTTTTATTGGG, TTTTTTATTGGG, TTATTGG, TTATTGGAGGG, TTATTGGGG, TTATTATTGGG, TTTTGTTATTGGG	Combination of various short, medium and long repeats	CM020934.1-CM020938.1
Tenthredinidae	*Tenthredo notha*	**TTAGG**	Various short, medium and long repeats; TTAGG interspersed by various repeats; SART - (TTAGG)n	OU611906.1-OU611925.1
Vespidae	*Ancistrocerus nigricornis*	**TTGGGTCTGGG**	SART - (TTGGGTCTGGG)n	OU696664.1-OU696669.1
Vespidae	*Dolichovespula saxonica*	**TTGCGTCTGGG**	31 bp motif: TTTTGGACACTACCGCGCTTTCGTA ACAGTA, TTTAGACACTACTACGCCTTCATAA GAGTAG, TTTTGGACACTACCGCGCTTTCGTA ACAGTA, TTTTAGAGACTACCGCGCCTTCGTA ACAGTA and other variants	OU426993.1-OU427018.1
Vespidae	*Dolichovespula media*	**TTGCGTCTGGG**	TTTTAGTTTACGCTCGCGTCTTGGA, TTTTAGTTTACGCTCGCGTCTTGGA, TTTTTGGTTTTCGGTCGCGTCTGGG, CCCCCGCGTATGGGTTGCCTCTG	OU426851.1-OU426875.1
Vespidae	*Dolichovespula sylvestris*	**TTGCGTCTGGG**	The telomeric repeat TTGCGTCTGGG, (TTGCGTCTGGG)2 (TTGCGTCTGGG)3 is often found in the subtelomeric region	OU964961.1- OU964986.1
Vespidae	*Vespa crabro*	**TTGCGTCAGGG**	TTTTTTCTGGTTCCTTCCGGAGTTGT GACTGGTTACGTTTGTAG	OU342400.1-OU342424.1

Vespidae	*Vespa velutina*	**TTGCGTCAGGG**	AAAAGTTGTGTCTGGTTGCGACTGC AGTTGCGTCTGGTTACGTCCGGGTA GCGTCAGCGTAGCGACTGACGTTGC GTCTGGTAGTGTTT	OU525123.1-OU525147.1
Vespidae	*Vespula germanica*	**TTGCGTCTGGG**	TTGCGTCTGAAGG	HG996528.1-HG996552.1
Vespidae	*Vespula vulgaris*	**TTGCGTCTGGG**	TTGGGTCTGAGG, TTGGGTCTCAGG, TTGGGTCTGAAGG	FR997668.1-FR997692.1
DIPTERA				
Conopidae	*Sicus ferrugineus*	**374 bp sequence**		OV277348.1-OV277354.1 -
Sciomyzidae	*Coremacera marginata*	**371 bp sequence**		OU612043.1-OU612093.1
Syrphidae	*Melanostoma mellinum*	**327-329 bp sequence**		OU612058.1-OU612063.1
Syrphidae	*Scaeva pyrastri*	**327 bp sequence**		LR989927.1-LR989931.1
Tachinidae	*Tachina fera*	**173-178 bp sequence**		LR999963.1-LR999969.1