Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera*

Hugh M. Robertson[1,3] and Karl H.J. Gordon[2]

[1]*Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA;* [2]*CSIRO Entomology, Canberra, ACT 2601 Australia*

The draft assembly of the honey bee *Apis mellifera* genome sequence reveals that the 17 centromeric-distal telomeres are of a simple, shared, and canonical structure, with 3–4 kb of a unique subtelomeric sequence, followed by several kilobases of TTAGG or variant telomeric repeats. This simple subtelomeric structure differs from the centromeric-proximal telomeres on the short arms of the 15 acrocentric chromosomes, which are apparently composed primarily of the 176-bp *Alu*I tandem repeat. This dichotomy between the distal and proximal telomeres may involve differential participation of the telomeres of the 15 acrocentric chromosomes in the Rabl configuration after mitosis and the chromosome bouquet in meiotic prophase I. As expected from the presence of canonical TTAGG telomeric repeats, we identified a candidate telomerase gene in the bee, as well as the silkmoth *Bombyx mori* and the flour beetle *Tribolium castaneum*.

[Supplemental material is available online at www.genome.org.]

Telomeres are fundamental structures at the ends of linear eukaryotic chromosomes. They have been the subject of intense study in diverse organisms since their discovery in the ciliate *Tetrahymena thermophila* (Greider and Blackburn 1985). These structures range in size from a short 36 bp to many kilobases in humans (Greider 1996), but in most organisms involve a canonical organization of a subtelomeric region followed by many short tandem telomeric repeats, from TTTGGG in yeasts to TTTAGGG in plants to TTAGGG in vertebrates (Meyne et al. 1989; Vega et al. 2003). The TTAGG repeat found widely in insects of many orders (Frydrychova et al. 2004) is present in other arthropods and is thought to have evolved from the TTAGGG repeat present in many lower-branching eukaryotes (Vitkova et al. 2005).

Telomeres serve at least three major roles (Blackburn 2001; Cowan et al. 2001). First, constant extension of the telomeric repeats by telomerase prevents truncation of the ends of chromosomes during DNA replication. Complex interplay of telomerase and many associated and telomere-binding proteins regulates activity and maintains telomeres at a regular length. Loss of telomerase activity, and hence the telomeric repeats, is associated with cell senescence and aging, while cancer cells typically have elevated telomerase activity involved in their immortality. Second, telomeres have a closed DNA structure at their extreme termini and are bound by various proteins, together serving to prevent recognition of the telomere as a broken chromosome end and inappropriate end-joining to other chromosomes. Third, some of the telomere-binding proteins mediate their location in the eukaryotic nucleus, most prominently in their localization on the nuclear membrane opposite the centromeres in the Rabl configuration after each mitosis and in the bouquet formation involved with synapsis of homologs during prophase I of meiosis. These localizations of telomeres are thought to involve binding of the subtelomeric regions by specific proteins.

Among the insects in which the TTAGG has been identified are many Coleoptera (Sahara et al. 1999) and all species studied from the Hymenoptera (Okazaki et al. 1993; Sahara et al. 1999; Lorite et al. 2002). In contrast, all dipteran species studied lack the TTAGG repeat (Sahara et al. 1999). Characterization of two telomeres of the fruit fly *Drosophila melanogaster* revealed that they consist instead of many full and partial insertions of particular families of non-LTR retrotransposons (Abad et al. 2004; Biessmann et al. 2005; Casacuberta and Pardue 2005; Cenci et al. 2005; Melnikova and Georgiev 2005). The telomeres of *Chironomus* midges, like those of the mosquito *Anopheles gambiae*, present yet a third type of organization, consisting of many complex tandem repeats (Saiga and Edström 1985; Biessmann et al. 1998; Rosen and Edström 2000). The telomeres of the silkmoth *Bombyx mori* have the canonical TTAGG telomeric repeat structure; however, they are regularly interrupted by the insertion of families of non-LTR retrotransposons (Fujiwara et al. 2005), perhaps reflecting the evolutionary situation in the Endopterygota before the clade Antliophora (including the Diptera) lost their telomerase (Frydrychova et al. 2004) and became completely dependent on reliable insertion of non-LTR retrotransposons to maintain their telomeres. The detailed structures of most other insect telomeres remain unknown.

In a number of insects with the TTAGG telomeric repeat, the expected telomerase activity has been shown to be present (Sasaki and Fujiwara 2000); however, no telomerase gene has been reported from an arthropod. As expected for the maintenance of the TTAGG repeats, Fujiwara et al. (2005) state that they have now identified a candidate telomerase gene in the genome of *B. mori* (Mita et al. 2004; Xia et al. 2004). Consistent with the lack of a TTAGG repeat, genes encoding telomerase could not be identified in the dipteran genomes (Adams et al. 2000; Zdobnov et al. 2002).

The chromosomes of the honey bee *A. mellifera* have been studied cytologically for over 100 yr (see Beye and Moritz 1995), although they are not particularly conducive to cytological study, being mostly small and hard to differentiate. Beye and Moritz (1995) provided identifying cytological features for all 16 chromosomes using FISH with probes to the rDNA that labels two chromosomes, the 547-bp *Ava*I tandem repeat that labels the

[3]Corresponding author.
E-mail hughrobe@uiuc.edu; fax (217) 244-3499.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.5085606. Freely available online through the *Genome Research* Open Access option.

centromeres, and the 176-bp *Alu*I tandem repeat that labels the telomeres adjacent to the centromeres of the 15 acrocentric chromosomes, but not the single large metacentric chromosome. It has not been proven possible to unequivocally connect each of these named chromosomes to those identified by linkage mapping (Solignac et al. 2004), to which the genome assembly has been mapped (Honey Bee Genome Sequencing Consortium 2006). The three broad categories are preserved, however: the large metacentric chromosome 1, the medium-sized acrocentric chromosomes 2–12, and the smaller acrocentric chromosomes 13–16.

Here, we describe the 17 centromeric-distal telomeres from the 15 acrocentric chromosomes and both telomeres of the metacentric chromosome 1, and show that their organization differs from the apparent *Alu*I tandem repeat structures of the centromeric-proximal telomeres visualized by Beye and Moritz (1995). Instead, they have the simplest canonical structure known in insects, which has implications for how the chromosomes are organized in the nucleus. We also identify a candidate insect telomerase reverse transcriptase (TERT) from bee, *B. mori*, and the flour beetle *Tribolium castaneum*. These genome-facilitated discoveries should allow studies of the possible involvement of telomere length and telomerase activity in the remarkable age differences of worker, drone, and queen bees. They also provide the first insect model for study of canonical telomeres and telomerase.

Results

The 17 distal telomeres

The first distal bee telomere was identified as a result of an effort to "superscaffold" the smallest chromosome, number 16, using manual methods and all available sources of evidence to evaluate the quality of the honey bee genome assembly v2 (Honey Bee Genome Sequencing Consortium 2006). Given the expectation that the distal bee telomeres would contain the canonical insect telomeric TTAGG tandem repeat (Sahara et al. 1999), but not the 176-bp tandem *Alu*I repeat associated with the proximal telomeres (Beye and Moritz 1995), a candidate telomere was identified at the superscaffolded telomeric end of chromosome 16. It is now mapped at the telomeric end of chromosome 16 as the 3′ end of the 93-kb scaffold Group16.21 in Assembly v4. This consists of 6 kb of unique sequence beyond the last annotated gene (GB17231), followed by an AT-rich subtelomeric region of 3.5 kb (70% AT) that has many matches elsewhere in the genome, followed by 430 bp of tandem repeats of TTAGG, or variants thereof. Furthermore, the only two available mate pairs from the 2–4-kb insert clones that should be distal to this assembled sequence consist entirely of these short tandem telomeric repeats.

The subtelomeric 3.5-kb region of chromosome 16 was used in BLASTN searches to identify the partially or fully assembled telomeric ends of 14 other chromosomes in Assembly v4 (Fig. 1).

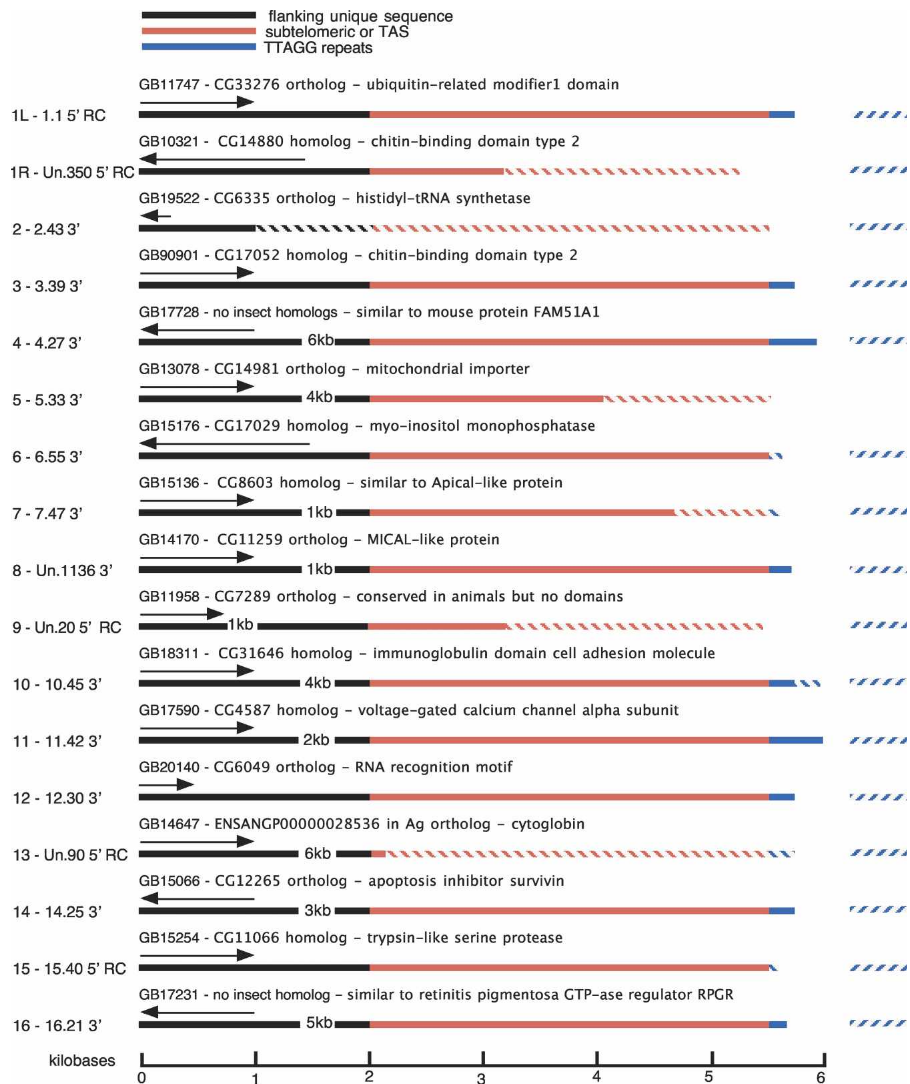

Figure 1. Schematic diagram of the 17 distal honey bee telomeres. Flanking unique DNA sequence is shown as black bars, the subtelomeric sequence is red, and the TTAGG or variant tandem telomeric repeats are blue. Length variation in the subtelomeric sequence (from 3–4 kb) is not shown. Hatched lines indicate regions that were manually assembled. The existence of clone end mate pairs that indicate extensive TTAGG repeats beyond the assembled ends of the chromosomes are indicated on the *right* as separate hatched blue bars. Each chromosome is identified on the *left*, together with the identifier of the terminal scaffold in the format GroupX.Y, where X is the chromosome number or Un for unmapped scaffolds, and Y is the scaffold number on that chromosome or in the unmapped grouping. *Above* each chromosome bar the approximate location and orientation of the ultimate gene is shown by an arrow (if the gene model starts or ends more than 1.5 kb from the subtelomeric sequence the extra length is shown interupting the black bar). The genes are identified by the GB number in the "official" and "ab initio" gene prediction sets (Honey Bee Genome Sequencing Consortium 2006), as well as their best BLASTP match in *Drosophila melanogaster*, or other organisms if the *Drosophila* ortholog has been lost (three genes) and their likely function is shown by similarity to mammalian proteins or any conserved domains shown by the NCBI Conserved Domain search.

Twelve of the 17 telomeres have their subtelomeric regions partially or fully assembled at the ends of the terminally mapped scaffold. The scaffold containing the telomere of chromosome 15, Group15.40, was already mapped to near the chromosome end, but reoriented and placed unequivocally at the terminus using superscaffolding (genetic mapping of several short scaffolds near the telomere of this chromosome has no resolution, and Groups 15.41 and 15.42 are actually internal to both 15.39 and 15.40). The telomeric regions of four chromosomes were already assembled as unoriented and unmapped GroupUn scaffolds, and these were associated with their appropriate chromosomes (1R, 8, 9, and 13) by manual superscaffolding. The telomeres of chromosomes 2 and 13 were assembled manually off the existing ends of the chromosomes using mate pair information from the flanking unique sequences. Only three other short 1–2-kb contigs in assembly v4 match this subtelomeric region, Groups Un.2223, Un.8960, and Un.9020, which appear to be partial separately assembled haplotypes for chromosomes 3, 16, and 12, respectively, so we believe we have identified all 17 distal telomeres and that the 15 proximal telomeres of the acrocentric chromosomes do not contain this sequence.

The subtelomeric sequences of mammals and some other organisms consist of internally repeated segments, but there is little sign of such repeats in the bee subtelomeric regions. A region near the end of the subtelomere has a repeated structure, but it is a simple sequence repeat identified by dotplot analysis (data not shown).

The start of the TTAGG tandem telomeric repeats is evident at the end of the subtelomeric sequence for 13 of the chromosomes, but confidently alignable high-quality reads do not quite extend this far for the other four (Fig. 1). In addition, for all but chromosome 5, at least one and commonly several confident mate pairs of the 2–4-kb insert clones extending off the ends of the assembled subtelomeres, or the flanking unique sequences, consist entirely of TTAGG or variant tandem telomeric repeats, demonstrating that each telomere terminal consists of at least several kilobases of this telomeric repeat. The most common variants are TCAGG and TTGGG. All available mate pairs off the ends of the subtelomeres consist entirely of these repeats, so there is no evidence for additional complexity to the telomere structures.

The similarity of the subtelomeric sequence of 3–4 kb that allowed recognition of all of these telomeres also means that assembling them as unique sequences is sometimes difficult, and this was compounded by the presence of two haplotypes from the original queen bee DH4 (the genomic DNA of hundreds of drone sons was sequenced). The automated assembly therefore succeeded in subjugating this haplotype variation, at least for part of the subtelomeric sequence, for 14 of the 17 telomeres. Those subtelomeres that were partially assembled ended at a point of major haplotype divergence, usually a large indel. We attempted to assemble both haplotypes separately for these and the two subtelomeres that were not assembled (2 and 13); however, it is not always possible to assemble both haplotypes completely due to the relatively low-sequence coverage depth. We were most successful with chromosome 7, where the two haplotypes differ by 15% (counting all nucleotides in numerous indels). Differences between chromosomes range from 8% for 3 and 6%–30% for 3 and 10. This range of sequence divergence is shown graphically in a distance phylogenetic tree of the 17 telomeres plus both chromosome 7 haplotypes (Fig. 2). Visual inspection of the CLUSTALX alignments suggested that the telomeres fell into two roughly equal size groups with considerable sequence divergence, and the tree reflects this, with the position of the telomere from chromosome 9 being somewhat ambiguous. Presumably, these groupings reflect a history of gene conversion between telomeres on different chromosomes, followed by private mutation histories for each. A detailed study of this gene conversion would require separate assembly of all 34 haplotypes, which is not possible with the current sequence read depth.

Figure 2. Phylogenetic relationships of the subtelomeric region for each telomere. This is an uncorrected distance tree rooted at the midpoint. Maximum parsimony and maximum likelihood trees have similar topologies. Bootstrap support values are shown on each branch as percentage of 1000 distance replicates and of 1000 parsimony replicates.

The chromosome 16 subtelomeric sequence began 6 kb from the nearest confident gene model, and no other candidate genes could be identified in this 6 kb. Remarkably, this is true for the other 16 telomeres, with confident gene models for proteins conserved in other insects and/or other animals starting or ending between 0.5 and 7 kb of the subtelomeric sequence (Fig. 1). None of these genes is closely related to each other, so the gene-conversion processes responsible for maintaining the homogeneity of the subtelomeric sequences do not extend beyond them.

A candidate insect telomerase

Identification of extensive insect consensus TTAGG tandem telomeric repeats at the ends of these 17 distal telomeres, as well as their apparent presence at the ends of the proximal telomeres (Sahara et al. 1999), implies that the bee genome should encode telomerase. Telomerase is a highly divergent reverse transcriptase, and a single lineage of this protein is responsible for telomere maintenance in all organisms studied with TTAGG-style tandem repeats (Malik et al. 2000). Using sequences of other telomerases known from vertebrates, fungi, plants, and protozoans (Nakamura et al. 1997; Malik et al. 2000; Xia et al. 2000; Kuramoto et al. 2001; Oguchi et al. 2004; Yap et al. 2005), we identified a candidate gene encoding bee telomerase as Fgenesh-predicted gene model GB91203 in the 3.5-kb AT-rich contig GroupUn.8105. The N-terminal exon of this model was replaced with a more proximal one indicated both bioinformatically and by 5′ RACE experiments (GenBank DQ494417) (the gene model has been updated as GB31000). The C terminus of the model was truncated by the end of the contig, so it was obtained by 3′ RACE (GenBank DQ494418) and the genic region by PCR amplification and sequencing across the gap to the next scaffold, which is 9 kb GroupUn.1069 (GenBank DQ494419). The resultant nine exon gene model encodes an 801 amino acid bee telomerase (GenBank

RefSeq NM_001040681) with 21% identity and 43% similarity over a ±580 amino acid alignment with the 1132 amino acid human TERT. We failed to connect these scaffolds to already mapped scaffolds using manual superscaffolding, but they have recently been mapped to the centromeric end of chromosome 2 (M. Solignac, pers. comm.).

Identification of a candidate bee telomerase allowed us to identify orthologs in the available silk moth *B. mori* and flour beetle *T. castaneum* genomes. Most of the *B. mori* gene is in the 6-kb Ctg021552 (Accession AADK01021552) of the Chinese 6X assembly (Xia et al. 2004), and a shorter 630-bp contig542579 (Accession BAAB01136182.1) from the Japanese 3X assembly (Mita et al. 2004) was used to correct a frameshift in contig Ctg021552. The result is a long ORF encoding 511 amino acids comprising the conserved C terminus of the TERT. The C-terminal half is also contained in two ESTs (GenBank CK562856.1 and CK518004.1; Xia et al. 2004). We obtained the N terminus by 5′ RACE (GenBank DQ494420) and it is encoded by a single exon in the 18-kb contig Ctg004553 (Accession AADK01004553). The resultant full-length protein is 703 amino acids (see online FASTA file) with 20% identity and 39% similarity to the bee telomerase. The *T. castaneum* gene was identified in the 30-kb Contig2849 (Accession AAJJ01001013, http://www.hgsc.bcm.tmc.edu/projects/tribolium/, v2), which similarly contains a single long ORF encoding 596 amino acids. A 5′ RACE identified an upstream exon (GenBank DQ494421), but it has no coding capacity in frame with this long ORF, and appears to be only a 5′ UTR. This candidate beetle telomerase (GenBank RefSeq NM_001040706) has 20% identity and 41% similarity to the bee telomerase.

The three candidate insect telomerases contain most, but not all, of the motifs identified in TERT proteins from other eukaryotes, and therefore differ from other known TERTs. The conserved motifs present in the insect enzymes include the seven motifs (1, 2, A–E) defining the well-conserved core RT domain as shown in Figure 3. The insect TERTs also include the TERT-specific T motif that is located immediately upstream from the core RT domain (Fig. 3). This T motif is not found in non-TERT RTs; it is also absent from the *Giardia lamblia* TERT (Malik et al. 2000). Upstream from the T domain, the insect TERTs include some, but not all, of the sequence motifs identified in other TERTs that are not found in RTs (Xia et al. 2000). These include the CP and QFP domains. Missing from the insect enzymes is the most N-terminal of the TERT domains, the GQ domain. These motifs are almost completely absent from the *Caenorhabditis elegans* and *G. lamblia* TERTs (Malik et al. 2000). Vertebrate TERTs have been shown to include these three N-terminal motifs in longer stretches of sequence homology, termed the vertebrate domains v-I to v-IV, as well as including three further conserved domains C-terminal to the RT core (v-V to v-II) (Yap et al. 2005; see Fig. 3). The extent of sequence conservation in the motifs defining the key enzyme domains is visible in a complete alignment of the sequences of these motifs from insect and selected other eukaryote TERTs (Supplemental data). The sequences do not show any further outlying conserved domains specific to the insect TERTs.

Amino acid sequences comprising the core RT domains (Malik et al. 2000) were used for analysis of the phylogenetic relationship between these three insect TERTs and those from other animals, fungi, and protozoans. The tree shown in Figure 4 is based on the alignment covering these motifs in the Supplemental data, using the *G. lamblia* TERT as an outgroup, and was obtained using maximum likelihood analysis. All of the insect TERTs clustered together, consistent with a common evolutionary origin. Other clusters obtained comprised the ciliate/yeast TERTs and those from vertebrates/plants, in agreement with Oguchi et al. (2004). The high levels of TERT sequence divergence mean that how these clusters are related could not be resolved.

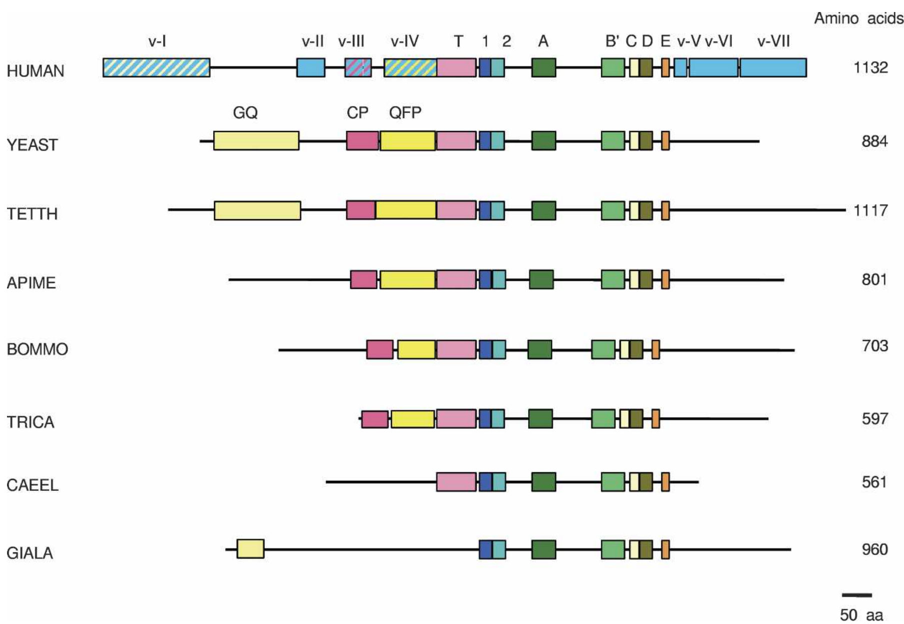

Figure 3. Schematic showing conserved domains in selected TERTs from vertebrates, fungi, ciliates, insects, nematodes, and *Giardia*. The domains are as previously identified for these TERTs (Malik et al. 2000; Xia et al. 2000; Kuramoto et al. 2001; Yap et al. 2005). See Supplemental data for the details of accessions. To avoid repetition, not all the TERTs listed there are shown in this figure; those not shown closely resemble one of those depicted.

Discussion

Our identification of the 17 centromere-distal telomeres of honey bee as having the relatively simple canonical structure of a shared short 3–4-kb nonrepetitive subtelomeric sequence, followed by extensive TTAGG or variant tandem telomeric repeats, and the discovery of a candidate bee, moth, and beetle telomerase have several implications.

Fifteen of these subtelomeric regions were already at least partially assembled by the automated assembly program ATLAS (Honey Bee Genome Sequencing Consortium 2006), with nine of them reaching all the way to the TTAGG tandem telomeric repeats, and 12 were already correctly mapped and oriented at the ends of chromosomes. This shows that the assembly succeeded in subjugating most of the haplotype differences in this genome and that this draft assembly v4 is remarkably complete. That we were able to assemble all 17 telomeres to near completeness for

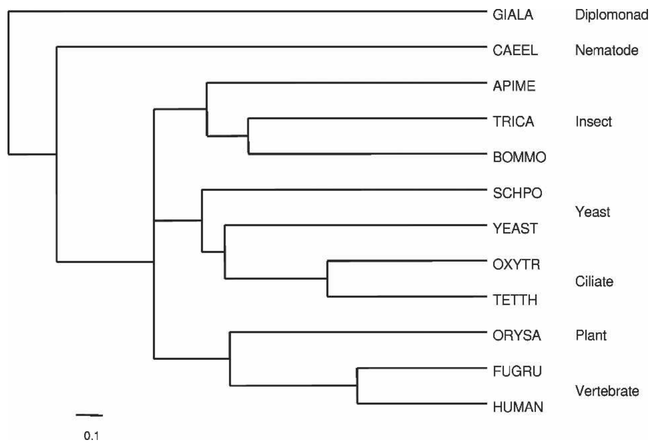

Figure 4. Phylogenetic analysis of the telomerase proteins of insects and diverse other organisms. The maximum likelihood tree (PHYLIP) is based on the alignments of core RT domains 1, 2, A-E for all of the TERTs shown in the Supplemental data.

the subtelomeric region shows the completeness of coverage of the plasmid libraries.

These 17 distal telomeres show no evidence of insertions of non-LTR or any other transposons. This is in keeping with observations from the genome assembly that non-LTR transposons are essentially absent (Honey Bee Genome Sequencing Consortium 2006). Bee telomeres have therefore been free of the kind of "parasitization" that typifies *Bombyx* telomeres (Fujiwara et al. 2005) and presumably led to the loss of telomerase and canonical telomeres in flies (e.g., Biessmann et al. 2005; Casacuberta and Pardue 2005; Melnikova and Georgiev 2005).

The subtelomeric regions of these distal bee telomeres are also quite different from those of *Drosophila* telomeres, not only as expected in sequence, but in their simple organization. Characterization of the 2L and X subtelomeric regions showed them to consist of many kilobases of at least three kinds of tandem repeats and, in fact, to be quite different from each other (Karpen and Spradling 1992; Walter et al. 1995). Further differences between the subtelomere sequences of *Drosophila* chromosomes have emerged, with that of 2R closely resembling 3R, whereas 2L and 3L show similar sequences (Abad et al. 2004). Few other insect subtelomeric regions are known, e.g., for *Chironomus* flies (Saiga and Edström 1985; Rosen and Edström 2000) and a cricket (Kojima et al. 2002), and they are highly repeated structures reminiscent of the apparent organization of the bee proximal telomeres with their extensive *Alu*I repeats (Beye and Moritz 1995).

The distal telomeres are apparently quite different, at least in their subtelomeric region, from the 15 proximal telomeres. Sahara et al. (1999) reported that all bee telomeres have the TTAGG telomeric repeats by FISH visualization; however, the proximal telomeres appear to consist primarily of large numbers of the tandemly repeated 176-bp *Alu*I repeat (Tares et al. 1993; Beye and Moritz 1995). No *Alu*I repeats were found in or near the distal telomeres; indeed, only a few of these repeats are present in a few short unmapped scaffolds in the draft assembly, while many thousands are in thousands of reads in an unassembled high-copy repeat-read database. Beye and Moritz (1995) observed that their *Alu*I repeat probe did not label the large metacentric chromosome 1, consistent perhaps with this chromosome resulting from centromeric fusion of two acrocentric chromosomes with loss of the proximal telomeres.

Thus, the proximal and distal honey bee telomeres appear to be quite different, at least in their subtelomeric regions. This difference might be important in the context of the Rabl configuration typical of chromosomes immediately after mitosis, in which the centromeres and telomeres are clustered at opposite poles of each new nucleus (for review, see Cowan et al. 2001). Although it is unknown whether bees show the Rabl configuration, if they do, the 15 acrocentric chromosomes would need to have different proximal and distal telomeres so that only the distal telomeres plus the telomeres of the large metacentric chromosome 1 could be clustered at the opposite pole to the centromeres. The different composition of the proximal telomeres would allow them to remain with the centromeres of the acrocentric chromosomes. Presumably there are proteins associated with the nuclear lamin that bind to the subtelomeric regions of the distal telomeres to localize them. The homogenization of these subtelomeric sequences would be important to maintain these protein–telomere interactions. This difference between the distal and proximal telomeres of acrocentric chromosomes might also be important for formation of the chromosome bouquet at the start of prophase I of meiosis, where proximity of the distal telomeres might facilitate synapsis of homologs (see Cowan et al. 2001; Scherthan 2001). The differences between the subtelomeric sequences of the metacentric *Drosophila* chromosomes have already been referred to; different telomeric structures or sequences have also been observed at the proximal and distal telomeres of acrocentric chromosomes before, e.g., the small fourth chromosome of a chironomid fly (Cohn and Edström 1992). An early example of differences between the (sub)telomeres of individual chromosomes came from *Xenopus laevis*, where the 5S ribosomal RNA genes were found at the telomere of only the longer arm of each chromosome (Pardue et al. 1973). Despite these observations, no evidence for differential telomere maintenance on individual chromosomes has been found to date. How telomeres consisting of repeats such as those found in *Chironomus* are maintained remains unclear, with gene conversion or RT-based mechanisms being possible (Rosen and Edström 2000); in other systems recombination-based mechanisms have been implicated (Pardue and DeBaryshe 1999; Fajkus et al. 2005).

The candidate insect telomerase we have identified encoded in the *Apis*, *Bombyx*, and *Tribolium* genomes has roughly 20% sequence identity to the vertebrate, fungal, and plant telomerases across most of its length and most of the known conserved motifs of these telomerases. It is a unique protein in these insect genomes, and in agreement with previous studies, no homolog is encoded by the *Drosophila* or *Anopheles* genomes. The insect telomerases all carry the RT-specific domains and one of the major telomerase-specific domains identified by others. The insect telomerase sequences also contain some, but not all of the amino-terminal telomerase domains identified in the enzymes from vertebrates, fungi, and ciliates, but not nematodes or *Giardia*. Initial results from a genome-wide tiling array and quantitative RT/PCR indicate that the bee telomerase is expressed in most tissues at low levels (Honey Bee Genome Sequencing Consortium 2006), but additional studies will be needed to determine whether it is differentially highly expressed in the long-lived queens or their ovaries.

We have also sought other telomere- or telomerase-associated proteins (e.g., Vega et al. 2003; Smogorzewska and de Lange 2004) in these insect genomes, which might also be missing from fly genomes. However, in some cases, these other proteins are so rapidly evolving in the lineage from which they are

known that homologs are not easily recognized in insects, e.g., the telomere-binding factors 1 and 2 (TRF1/2) proteins in vertebrates and the Est3 telomere elongation protein of yeasts (Smogorzewska and de Lange 2004). Otherwise, they are conserved in all eukaryotes and present in flies, presumably because they are also involved in other cell processes, e.g., the conserved helicase Pif1 in yeast, which is a negative regulator of telomerase (Schulz and Zakian 1994; Zhou et al. 2000; Boule et al. 2005), and the ever-shorter telomeres Est1A protein of vertebrates (Smogorzewska and de Lange 2004). It is therefore possible that only the core telomerase activity and the TTAGG repeats were lost from flies. The presence in the bee of many proteins known to play roles in regulation of telomerase function in other organisms, as well as of the canonical telomerase itself, makes it likely that these proteins may play similar roles in the bee. The other key component of the telomerase catalytic core is the RNA template (Cech 2004); while the structure of this contains conserved features including a hairpin and pseudo-knot (Chen and Greider 2004; Lin et al. 2004), the actual nucleotide sequence evolves too rapidly to find in insects using the vertebrate, ciliate, or yeast versions as queries.

Methods

We used "manual superscaffolding" to connect four partially assembled subtelomeric regions to their appropriate chromosomes ends (1R, 8, 9, and 13), and to orient and position terminally the partially mapped assembled telomere of chromosome 15. The automated assembly was conservative in requiring at least two unconflicted mate pairs to join contigs into scaffolds, so we used single mate pairs to join scaffolds as well as existing unmerged overlaps between scaffolds built from the two alternative haplotypes in this genome sequence. The telomeres of chromosomes 2 and 13, and both haplotypes of chromosome 7 were manually assembled using mate pair information connected from their unique flanking sequences. All reads and mate pair information were from 1–4 kb insert plasmid clones at the Trace Archive at NCBI.

Phylogenetic analysis of the telomeres was performed using uncorrected distances in PAUP4* (Swofford 2001) after aligning the 3–4-kb subtelomeric regions in CLUSTALX (Thomson et al. 1994). One-thousand bootstrap replications were used to evaluate branch confidence. Phylogenetic analysis of the motifs common to all RTs—1, 2, A-E—in the telomerase (TERT) protein sequences after alignment in CLUSTALX was performed using PHYLIP (Felsenstein 1989).

Acknowledgments

We thank Kim Walden for amplification and sequencing of the C-terminal region of the bee telomerase gene and RACE experiments on telomerase cDNAs; Michel Solignac for sharing his 2000-microsatellite-marker linkage map used for mapping genome assembly v4 before publication and for specifically mapping the telomerase gene; the Baylor Human Genome Sequencing Center for access to the *Apis mellifera* and *Tribolium castaneum* draft genome sequences before publication; and Bill Gelbart for suggesting the connection of the different subtelomeric sequences to the Rabl configuration and meiotic bouquet. This work was supported by NIH grant AI56081 to H.M.R. and funding from the Grains Research and Development Corporation and the Cotton Research and Development Corporation in Australia to K.H.J.G.

References

Abad, J.P., de Pablos, B., Osoegawa, K., de Jong, P.J., Martin-Gallardo, A., and Villasante, A. 2004. Genomic analysis of *Drosophila melanogaster* telomeres: Full-length copies of HeT-A and TART elements at the telomeres. *Mol. Biol. Evol.* **21:** 1613–1619.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.

Beye, M. and Moritz, R.F. 1995. Characterization of honeybee (*Apis mellifera* L.) chromosomes using repetitive DNA probes and fluorescence in situ hybridization. *J. Hered.* **86:** 145–150.

Biessmann, H., Kobeski, F., Walter, M.F., Kasravi, A., and Roth, C.W. 1998. DNA organization and length polymorphism at the 2L telomeric region of *Anopheles gambiae*. *Insect Mol. Biol.* **7:** 83–93.

Biessmann, H., Prasad, S., Walter, M.F., and Mason, J.M. 2005. Euchromatic and heterochromatic domains at *Drosophila* telomeres. *Biochem. Cell Biol.* **83:** 477–485.

Blackburn, E.H. 2001. Switching and signaling at the telomere. *Cell* **106:** 661–673.

Boule, J.B., Vega, L.R., and Zakian, V.A. 2005. The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature* **438:** 57–61.

Casacuberta, E. and Pardue, M.L. 2005. HeT-A and TART, two *Drosophila* retrotransposons with a bona fide role in chromosome structure for more than 60 million years. *Cytogenet. Genome Res.* **110:** 152–159.

Cech, T.R. 2004. Beginning to understand the end of the chromosome. *Cell* **116:** 273–279.

Cenci, G., Ciapponi, L., and Gatti, M. 2005. The mechanism of telomere protection: A comparison between *Drosophila* and humans. *Chromosoma* **114:** 135–145.

Chen, J.-L. and Greider, C.W. 2004. An emerging consensus for telomerase RNA structure. *Proc. Natl. Acad. Sci.* **101:** 14683–14684.

Cohn, M. and Edström, J.E. 1992. Chromosome ends in Chironomus pallidivittatus contain different subfamilies of telomere-associated repeats. *Chromosoma* **101:** 634–640.

Cowan, C.R., Carlton, P.M., and Cande, W.Z. 2001. The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol.* **125:** 532–538.

Fajkus, J., Sykorova, E., and Leitch, A.R. 2005. Telomeres in evolution and evolution of telomeres. *Chromosome Res.* **13:** 469–479.

Felsenstein, J. 1989. PHYLIP—Phylogeny inference package (Version 3.2). *Cladistics* **5:** 164–166.

Frydrychova, R., Grossmann, P., Trubac, P., Vitkova, M., and Marec, F. 2004. Phylogenetic distribution of TTAGG telomeric repeats in insects. *Genome* **47:** 163–178.

Fujiwara, H., Osanai, M., Matsumoto, T., and Kojima, K.K. 2005. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. *Chromosome Res.* **13:** 455–467.

Greider, C.W. 1996. Telomere length regulation. *Annu. Rev. Biochem.* **65:** 337–365.

Greider, C.W. and Blackburn, E.H. 1985. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43:** 405–413.

The Honey Bee Genome Sequencing Consortium. 2006. Insights into social insects from the genome of the honey bee *Apis mellifera*. *Nature* (in press).

Karpen, G.H. and Spradling, A.C. 1992. Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single-P element insertional mutagenesis. *Genetics* **132:** 737–753.

Kojima, K.K., Kubo, Y., and Fujiwara, H. 2002. Complex and tandem repeat structure of subtelomeric regions in the Taiwan cricket, *Teleogryllus taiwanemma*. *J. Mol. Evol.* **54:** 474–485.

Kuramoto, M., Ohsumi, K., Kishimoto, T., and Ishikawa, F. 2001. Identification and analyses of the *Xenopus* TERT gene that encodes the catalytic subunit of telomerase. *Gene* **277:** 101–110.

Lin, J., Ly, H., Hussain, A., Abraham, M., Pearl, S., Tzafati, Y., Parslow, T.G., and Blackburn, E.H. 2004. A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. *Proc. Natl. Acad. Sci.* **101:** 14713–14718.

Lorite, P., Carrillo, J.A., and Palomque, T. 2002. Conservation of (TTAGG)n telomeric sequences among ants (Hymenoptera, Formicidae). *J. Hered.* **93:** 282–285.

Malik, H.S., Burke, W.D., and Eickbush, T.H. 2000. Putative telomerase catalytic subunits from *Giardia lamblia* and *Caenorhabditis elegans*. *Gene* **251:** 101–108.

Melnikova, L. and Georgiev, P. 2005. *Drosophila* telomeres: The non-telomerase alternative. *Chromosome Res.* **13:** 431–441.

Meyne, J., Ratliff, R.L., and Moyzis, R.K. 1989. Conservation of the

human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci.* **86:** 7049–7053.
Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., et al. 2004. The genome sequence of silkworm, *Bombyx mori. DNA Res.* **11:** 27–35.
Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277:** 955–959.
Oguchi, K., Tamura, K., and Takahashi, H. 2004. Characterization of *Oryza sativa* telomerase reverse transcriptase and possible role of its phosphorylation in the control of telomerase activity. *Gene* **342:** 57–66.
Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H., and Fujiwara, H. 1993. Identification of a pentanucleotide telomeric sequence, TTAGG.n, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.* **13:** 1424–1432.
Pardue, M.L. and DeBaryshe, P.G. 1999. Telomeres and telomerase: More than the end of the line. *Chromosoma* **108:** 73–82.
Pardue, M.L., Brown, D.D., and Birnstiel, M.L. 1973. Location of the genes for 5S ribosomal RNA in *Xenopus laevis*. *Chromosoma* **42:** 191–203.
Rosen, M. and Edström, J.E. 2000. DNA structures common for chironomid telomeres terminating with complex repeats. *Insect Mol. Biol.* **9:** 314–347.
Sahara, K., Marec, F., and Traut, W. 1999. TTAGG telomeric repeats in chromosomes of some insects and other arthropods. *Chromosome Res*. **7:** 449–460.
Saiga, H. and Edström, J.-E. 1985. Long tandem arrays of complex repeat units in *Chironomus* telomeres. *EMBO J.* **4:** 799–804.
Sasaki, T. and Fujiwara, H. 2000. Detection and distribution patterns of telomerase activity in insects. *Eur. J. Biochem.* **267:** 3025–3031.
Scherthan, H. 2001. A bouquet makes ends meet. *Nat. Rev. Mol. Cell Biol.* **2:** 621–627.
Schulz, V.P. and Zakian, V.A. 1994. The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. *Cell* **76:** 145–155.
Smogorzewska, A. and de Lange, T. 2004. Regulation of telomerase by telomeric proteins. *Annu. Rev. Biochem.* **73:** 177–208.
Solignac, M., Vautrin, D., Baudry, E., Mougel, F., Loiseau, A., and Cornuet, J.M. 2004. A microsatellite-based linkage map of the honeybee, *Apis mellifera* L. *Genetics* **167:** 253–262.
Swofford, D.L. 2001. PAUP*: *Phylogenetic analysis using parsimony and other methods*, Version 4. Sinauer Press, New York.
Tares, S., Cornuet, J.M., and Abad, P. 1993. Characterization of an unusually conserved *Alu*I highly reiterated DNA sequence family from the honeybee, *Apis mellifera*. *Genetics* **134:** 1195–1204.
Thomson, J.D., Higgins, D.J., and Gibson, T.J. 1994. CLUSTAL W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.
Vega, L.R., Mateyak, M.K., and Zakian, V.A. 2003. Getting to the end: Telomerase access in yeast and humans. *Nat. Rev. Mol. Cell Biol.* **4:** 948–959.
Vitkova, M., Kral, J., Traut, W., Zrzavy, J., and Marec, F. 2005. The evolutionary origin of insect telomeric repeats, (TTAGG)n. *Chromosome Res.* **13:** 145–156.
Walter, M.F., Jang, C., and Kasravi, B. 1995. DNA organization and polymorphism of a wild-type *Drosophila* telomere region. *Chromosoma* **104:** 229–241.
Xia, J., Peng, Y., Mian, I.S., and Lue, N.F. 2000. Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase. *Mol. Cell. Biol.* **20:** 5196–5207.
Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., et al. 2004. A draft sequence for the genome of the domesticated silkworm *Bombyx mori. Science* **306:** 1937–1940.
Yap, W.H., Yeoh, E., Brenner, S., and Venkatesh, B. 2005. Cloning and expression of the reverse transcriptase component of pufferfish (*Fugu rubripes*) telomerase. *Gene* **353:** 207–217.
Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M., et al. 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298:** 149–159.
Zhou, J., Monson, E.K., Teng, S.C., Schulz, V.P., and Zakian, V.A. 2000. Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science* **289:** 771–774.

Received December 21, 2005; accepted in revised form May 31, 2006.