DOI: 10.1534/genetics.108.096289

Note

Simple Telomeres in a Simple Animal: Absence of Subtelomeric Repeat Regions in the Placozoan *Trichoplax adhaerens*

Hugh M. Robertson[1]

Department of Entomology, University of Illinois, Urbana, Illinois 61801

Manuscript received September 15, 2008
Accepted for publication November 2, 2008

ABSTRACT

Simple telomeres were identified in the genome assembly of the basal placozoan animal *Trichoplax adhaerens*. They have 1–2 kb of TTAGGG telomeric repeats, which are preceded by a subtelomeric region of 1.5–13 kb. Unlike subtelomeric regions in most animals examined, these subtelomeric regions are unique to each telomere.

TELOMERES are essential features of the linear chromosomes of eukaryotes, where they protect the ends of chromosomes from degradation or undesirable involvement in end-joining DNA repair, and provide a means for replacing the sequence lost during each cycle of DNA replication (*e.g.*, Pryde *et al.* 1997; Mason *et al.* 2008). Most eukaryotes have simple short telomeric repeats that are added by a specialized reverse transcriptase known as telomerase or telomerase reverse transcriptase (TERT), which has an RNA template for the added telomeric repeats. These repeats are TTAGGG in most animals including the deuterostomes and basal animal lineages like *Trichoplax adhaerens* (Traut *et al.* 2007), although this is modified to TTAGG in arthropods (*e.g.*, Vítková *et al.* 2005), except in Drosophila and some other insects (*e.g.*, Mason *et al.* 2008). Immediately proximal to the telomeric repeats in most studied eukaryotes is a subtelomeric or telomere-associated region that contains sequence shared by all or most telomeres, presumably by gene conversion. This sequence is often repeated within each subtelomeric region and sometimes multiple different subtelomeric repeats exist. In some species this subtelomeric repetitive nature extends to suites of genes with similar copies at each telomere. For example, the draft genome sequence of the honeybee *Apis mellifera* provided a first view of an insect with simple TTAGG telomeric repeats and revealed a simple ~3.5-kb subtelomeric repeat region with ~70% DNA identity between all 17 telomeres on the long euchromatic arms (Robertson and Gordon 2006). These subtelomeric regions could serve several roles, including binding sites for proteins that locate the telomeres within the nucleus or orchestrate the bouquet formation during meiosis, as a barrier against chromatin silencing of neighboring genes, as backup for maintenance of telomere length through recombination in the absence of telomerase, or as transcription initiation sites for TERRA/TelRNA transcripts of the TTAGGG repeats (*e.g.*, Pryde *et al.* 1997; Jacob *et al.* 2004; Riethman *et al.* 2005; Siderakis and Tarsounas 2007; Horard and Gilson 2008; Mason *et al.* 2008).

The placozoan *T. adhaerens* has a compact genome of ~100 Mb in six chromosomes, and the newly available draft genome sequence is high quality, consisting of primarily megabase scaffolds (Srivastava *et al.* 2008). The genome sequence allowed these authors to place the Placozoa phylogenetically between the sponges and Cnidaria, providing a glimpse into the genome architecture of the most basal animal genome available to date. The genome encodes a TERT, although a gene model is not present in the "official" gene catalog, perhaps because its reverse transcriptase domain caused it to be filtered out due to resemblance to a retrotransposon. The gene model at the Joint Genome Institute (JGI) of the Department of Energy web site (http://genome.jgi-psf.org/Triad1/Triad1.home.html) is fgeneshTA2_pg.C_scaffold_12000048 [Triad1:59697]. Searches of the *T. adhaerens* assembly v1.0 with a 1000-base TTAGGG repeat query using the BLASTN algorithm with all filters turned off at the JGI web page identified 16 scaffolds/contigs with significant matches. Ten of these matches are at the ends of contigs at the ends of long scaffolds with the TTAGGG repeats oriented outward, and are

[1]*Address for correspondence:* Department of Entomology, University of Illinois, 505 S. Goodwin Ave., Urbana, IL 61801.
E-mail: hughrobe@uiuc.edu

TABLE 1

Locations of the 11 identified telomeres in the *Trichoplax adhaerens* genome assembly (ordered by decreasing size of the scaffold bearing them) and the distance to the nearest gene model

GenBank accession	Scaffold and contig no.	Orientation	Nearest gene	Distance (kb)
ABGP01000462.1	Scaffold_6_Cont462	5′ end in RC	Homeobox protein	1.5
ABGP01000617.1	Scaffold_9_Cont617	3′ end	STAM binding protein	5.5
ABGP01000721.1	Scaffold_11_Cont721	3′ end	Protein phosphatase 2	6.5
ABGP01000722.1	Scaffold_12_Cont722	5′ end in RC	Unique protein	4.5
ABGP01000796.1	Scaffold_15_Cont796	3′ end	FUN14 domain-containing	1.5
ABGP01001000.1	Scaffold_28_Cont1000	3′ end	Heatshock 70 protein	11.0
ABGP01001021.1	Scaffold_32_Cont1021	5′ end in RC	Huntington-interacting	13.0
ABGP01001031.1	Scaffold_33_Cont1031	5′ end in RC	Fatty-acid desaturase	7.5
ABGP01001073.1	Scaffold_36_Cont1073	5′ end in RC	Unknown protein	7.5
ABGP01001092.1	Scaffold_38_Cont1092	5′ end in RC	Kinesin	1.5
ABGP01001179.1	Scaffold_47_Cont1179	5′ end in RC	Unknown protein	5.5

RC, reverse complement.

assumed to be telomeres. Six of the poorer matches are to short stretches of TTAGGG repeats buried deeply within scaffolds, and are assumed not to be telomeres. Such short stretches of internal telomeric repeats are known in other genomes (*e.g.*, Riethman *et al.* 2005; Traut *et al.* 2007).

An 11th telomere was identified by a different strategy I employed previously to identify the complex telomeres of the red flour beetle *Tribolium castaneum* (Tribolium Genome Sequencing Consortium 2008). I searched the reads generated for the placozoan genome sequence at the Trace Archive at the National Center for Biotechnology Information using a 1000-bp TTAGGG repeat query and no filters. These reads consist of ~120,000 fosmid mate pairs (or paired end reads) from clones with an average insert size of ~36 kb, as well as ~5–600,000 reads each from plasmid clones with inserts of 6.4 and 2.3 kb average size. The first 200 matches in this search were all fosmid reads, because for unclear reasons these have better quality and longer stretches of TTAGGG repeats than most of the plasmid reads. All of these first 200 matches are in the plus/minus orientation, implying that these reads are at the ends of chromosomes (otherwise one would anticipate a 50/50 ratio of plus/plus and plus/minus matches). BLASTN searches of the genome assembly with their mate pairs almost always indicated they were 1 of the 10 identified telomeres, but 5 mate pairs identified an 11th telomere at the end of scaffold_15. The current assembly stops just before the TTAGGG telomeric repeats, and was extended into them using the somewhat low-quality 3′ end of a read from the Trace Archive.

The expected 12th telomere was not identified from the first 200 fosmid mate pairs, so I assumed it contained sequences that did not clone well in fosmids and therefore instead sought plasmids that might contain it by working backward from the 500th best match in the above search. Most of these reads are from plasmids rather than fosmids, but most were still in the plus/minus orientation, indicating they were at the ends of chromosomes (a few in plus/plus orientation appear to represent the kinds of internal short sets of TTAGGG repeats also found earlier). Mate pairs for these plasmid reads mostly matched within the 11 already-identified telomeric contigs/scaffolds, but a few did not have matches in the current assembly. Some of these might map to some of the gaps in telomeric scaffolds, but most appear to be from regions of the genome that did not clone well, because they also have few matches in the Trace Archive. One 6-kb mate pair of a read that is entirely TTAGGG repeats maps 5 kb inside the 5′ end of 40-kb scaffold_106, so this might represent the 12th telomere; however, the sequence of this end of this scaffold is repeated elsewhere in the genome, so a

```
scaffold_6_Cont462 RC   GTATTCGTATTTAAGAAGCCGAAATATCGCATTAAGGGATTATGATAAATATAACTAATCGATGTTACTAAACTGACACTTCTTGATCATATAAAAGGGGTTAGGGTTAGGG
scaffold_9_Cont617      TTAATGCAAAAATAAGGAATCATCAAGTTAAATTTGCCAATTCTGAATATACTACTAAATATTCAAATAACAAGTCAAGATGCAAATGTTCTGTAATGGGTTAGGGTTAGGG
scaffold_11_Cont721     ATATACTTCAGATGTATTTTTTGTGTTTGGTATCTAATTTATTAATGGAAAGTCTTACAAATTCCAATAATAAATTATTGATGATATTATCGTAGTAAGGTTAGGGTTAGGG
scaffold_12_Cont722 RC  CTAATTTTGTTTGACCCCCCAATATAAGCATACAGTAATTACATTTCAAGTAAACGAGTACTATTAGATTTATGTCTTGTAACATAAATTCTATAAAGGGTTAGGGTTAGGG
scaffold_15_Cont796 EXT ACATTTTAGGTGAACATAAAATGTCTGCAATTGTGTACAGCAAGCGTAATAGTACATGTAAGCGTTAGAGGCAGAATACGGGCTAGCGAATCGGAAAGGATTAGGGTTACCG
scaffold_28_Cont1000    TTATCTCATATCCTAACATGGATACGTGACCTAGTTCCTTACTAGCCCAATCTAGTAAACTAATTAAGCTAATTCCTTATGAATGAAACCTTAATCAAGGTTAGGGTTAGGG
scaffold_32_Cont1021 RC TGGCAAGTAAACTTCCGTCTTTTATTTTCCTACATATTATTCAGAATAAAGAATCAAGGTTAAATCATGATTAGATAGGTTAATTCAAAAATAAAATGCCTTAAGGTTAGGG
scaffold_33_Cont1031 RC ACCCTAATATTGCAAACATTGAAGCAGATAAAGTATTCACAATGCAAAGTAGAGGCAGTAATTGGTGTTGTCGCAATTAAAATATGTTTTATTAATAAGGTTAGGGTTAGGG
scaffold_36_Cont1073 RC ATACGGATATGAATTCGGATTCCGTCTACCTGTTTCGTAATATCCGTATCCTTGCATAATTGAAAGTTTAATTGTAGAAATATTCATCTTAGAGACAGGGTTAGGGTTAGGG
scaffold_38_Cont1092 RC TTAGTAATACAATCCTTAATTTCTCTAGATATACACTCAATTATTTTTTAGATACCTTTTCAGTAATGGATATTAGATATGAAATCCGTGCTAAATGGGGTTAGGGTTAGGG
scaffold_47_Cont1179 RC GCACTTTTTACTGGTTAATTCCAATATCAAGGATTTACTACGGTAATTATTAAAACACTAAATAAACGAAAATAAATGCCTTAAAAGATATTATCAGATGTTAAGGTTAGGG
```

Figure 1.—Last 100 bases of the subtelomeric regions of the 11 identified telomeres in the *Trichoplax adhaerens* genome assembly. The first two TTAGGG repeats are shown at the right end for alignment. The scaffold_15_Cont796 telomeric repeats were added from the low quality 3′ end of a single read. RC, reverse complement.

confident manual assembly out to the potential telomere is not possible, and it is also possible this 6-kb plasmid clone is chimeric.

These observations are consistent with the expectation of six chromosomes for *T. adhaerens* (Srivastava *et al.* 2008), as well as the FISH hybridization with a TTAGGG probe by Traut *et al.* (2007), which revealed ~12 spots in interphase chromosome spreads and ~24 spots in a mitotic spread. Unfortunately the chromosomes are not well resolved in their spreads, so it is not possible to determine if all their FISH spots are at telomeres, and there are some additional faint spots that might be the ~6 shorter internal sets of TTAGGG repeats noted above.

The locations of the 11 identified telomeres are listed in Table 1. The TTAGGG telomeric repeats appear not to be very long, because fosmid mate pairs of reads with TTAGGG repeats are generally the expected ~36 kb internal to these scaffold ends. Searches of the Trace Archive with the first 1 kb of each subtelomeric region revealed only a few 2.3-kb plasmid reads with plus/plus matches, the mate pairs of which either are the same subtelomeric sequence (indicating that the inserts were short) or are TTAGGG repeats, suggesting that the telomeric repeats are only a couple of kb long. The subtelomeric sequence immediately internal to the telomeric repeats is essentially unique to each telomere, as shown for the proximal 100 bp for each telomere in Figure 1. A few weak matches between subtelomeric regions were noted, but involve divergent copies of repeats also present elsewhere in the genome. The distance between the start of the TTAGGG telomeric repeats and the end of the nearest predicted gene model varies from 1.5 to 13 kb (Table 1), and these 11 genes share no obvious similarities. There is therefore no indication of any subtelomeric sequence shared between telomeres, and there are no repeated sequences within individual subtelomeric regions.

Not all subtelomeric regions are as uniformly repeated at each telomere as those of the honeybee (Robertson and Gordon 2006); for example, human subtelomeric regions, while commonly sharing subtelomeric repeats or Srpts, are quite distinctive overall (*e.g.*, Riethman *et al.* 2005), and there are some eukaryotes known to have no shared subtelomeric repeats, such as Tetrahymena (*e.g.*, Jacob *et al.* 2004). Nevertheless, the absence of subtelomeric repeats in *T. adhaerens* suggests that if these have important roles in other organisms, they are either dispensed with in placozoans or substituted for by other means.

LITERATURE CITED

Horard, B., and E. Gilson, 2008 Telomeric RNA enters the game. Nat. Cell. Biol. **10:** 113–115.

Jacob, N. K., A. R. Stout and C. M. Price, 2004 Modulation of telomere length dynamics by the subtelomeric region of Tetrahymena telomeres. Mol. Biol. Cell **15:** 3719–3728.

Mason, J. M., R. C. Frydrychova and H. Biessmann, 2008 Drosophila telomeres: an exception providing new insights. BioEssays **30:** 25–37.

Pryde, F. E., H. C. Gorham and E. J. Louis, 1997 Chromosome ends: all the same under their caps. Curr. Opin. Genet. Dev. **7:** 822–828.

Riethman, H., A. Ambrosini and S. Paul, 2005 Human subtelomere structure and variation. Chromosome Res. **13:** 505–515.

Robertson, H. M., and K. H. Gordon, 2006 Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera*. Genome Res. **16:** 1345–1351.

Siderakis, M., and M. Tarsounas, 2007 Telomere regulation and function during meiosis. Chromosome Res. **15:** 667–679.

Srivastava, M., E. Begovic, J. Chapman, N. H. Putnam, U. Hellsten *et al.*, 2008 The Trichoplax genome and the nature of placozoans. Nature **454:** 955–960.

Traut, W., M. Szczepanowski, M. Vítková, C. Opitz, F. Marec *et al.*, 2007 The telomere repeat motif of basal Metazoa. Chromosome Res. **15:** 371–382.

Tribolium Genome Sequencing Consortium, 2008 The genome of the model beetle and pest *Tribolium castaneum*. Nature **452:** 949–955.

Vítková, M., J. Král, W. Traut, J. Zrzavý and F. Marec, 2005 The evolutionary origin of insect telomeric repeats, (TTAGG)n. Chromosome Res. **13:** 145–156.

Communicating editor: R. S. Hawley