Diverse telomeres in trypanosomatids

Ester Poláková1, Kristína Záhonová2,3, Amanda T. S. Albanaz1, Anzhelika Butenko1,2,4, Julius Luke5,6, Vyacheslav Yurchenko1,5

1Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic; 2Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 370 05 České Budějovice (Budweis), Czech Republic; 3Faculty of Science, Charles University, BIOCEV, 252 50 Vestec, Czech Republic; 4Faculty of Sciences, University of South Bohemia, 370 05 České Budějovice (Budweis), Czech Republic and 5Martinsovsky Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov University, Moscow 119435, Russia

Abstract
Telomeres are the ends of linear eukaryotic chromosomes facilitating the resolution of the ‘end replication and protection’ problems, associated with linearity. At the nucleotide level, telomeres typically represent stretches of tandemly arranged telomeric repeats, which vary in length and sequence among different groups of organisms. Recently, a composition of the telomere-associated protein complex has been scrutinized in Trypanosoma brucei. In this work, we subjected proteins from that list to a more detailed bioinformatic analysis and delineated a core set of 20 conserved proteins putatively associated with telomeres in trypanosomatids. Out of these, two proteins (Ku70 and Ku80) are conspicuously missing in representatives of the genus Blastocritidia, yet telomeres in these species do not appear to be affected. In this work, based on the analysis of a large set of trypanosomatids widely different in their phylogenetic position and life strategies, we demonstrated that telomeres of trypanosomatids are diverse in length, even within groups of closely related species. Our analysis showed that the expression of two proteins predicted to be associated with telomeres (those encoding telomerase and telomere-associated hypothetical protein orthologous to Tb927.6.4330) may directly affect and account for the differences in telomere length within the species of the Leishmania mexicana complex.

Introduction
Trypanosomatidae is a family of protozoan parasites possessing a single large mitochondrion, which encompasses a network of catenated circular DNA molecules, the so-called kinetoplast. These species have been attracting research attention because of numerous unique or rare biochemical and molecular traits, such as trans-splicing and polycistronic transcription (Clayton, 2019; Michaeli, 2011), mitochondrial RNA editing (Aphasizheva et al., 2020), presence of modified nucleotides (van Luenen et al., 2012) and unusual organelles (Szöör et al., 2014; Docampo, 2016), or a bizarre variation of the nuclear genetic code (Záhonová et al., 2016). Most of these flagellates are monoxenous (with one host in their life cycle) parasites restricted to invertebrates (Maslov et al., 2013), while members of the genera Endotrypanum, Leishmania, Phytomonas, Porcisia and Trypanosoma have switched to dixeny (two-host life cycle) and infect vertebrates or plants in addition to invertebrates (Luke et al., 2018). It is established beyond a reasonable doubt that the dixenous species have evolved from the monoxenous ancestor(s) independently several times (Luke et al., 2014). Notably, several Leishmania and Trypanosoma spp. are of medical importance, as they cause severe human diseases, and are fairly well-studied (Stuart et al., 2008; Nussbaum et al., 2010).

Telomeres typically represent repetitive physical ends of linear eukaryotic chromosomes, variable in length and sequence in different groups of organisms (Fulnečková et al., 2013). Their main role is to protect chromosome ends from being recognized and processed as DNA double-strand breaks by the cellular repair machinery in order to prevent chromosomal end-to-end fusions (Pfeiffer and Lingner, 2013). Such shielding is provided by the telomere-associated protein complex (Lewis and Wuttke, 2012) or specific complementary DNA structures, such as telomere loops (l-loops) facilitating the protection of chromosome ends (Tomáška et al., 2019). It is generally assumed that telomeres undergo gradual shortening with each round of cell division because of incomplete lagging strand synthesis of linear DNA templates by DNA polymerases, known as the ‘end replication problem’ (Olovnikov, 1973; Greider, 1990; Hackett and Greider, 2002). In order to overcome this problem and, thus, prevent telomere shortening, cells engage a dedicated enzyme called telomerase (Greider and Blackburn, 1985).

Telomeres of kinetoplastids share many traits with those of other eukaryotes. They have the canonical sequence (5’-ttaggg-3’) found in vertebrates, end with a t-loop, are associated with capping protein complexes and maintained by telomerasers (Muñoz-Jordán et al., 2001; Conte and Cano, 2005; Fulnečková et al., 2013). Similar to the situation in other eukaryotic pathogens, genes encoding trypanosomatid virulence factors are often located in the sub-telomeric regions and their expression may be co-regulated with telomeres (Chiurillo et al., 1999;
### Table: Predicted telomere-associated protein complex composition in *T. brucei*

| Protein ID     | Annotation                      | Protein function                                                                                                                                                                                                 | References                                      |
|----------------|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|
| Tb927.2.6100   | Hypothetical protein            | Essential for cell growth and kinetoplast (k)DNA maintenance; kDNA was reduced in size or lost upon RNAi-mediated knock-down of the coding gene                                                                 | Beck et al. (2013)                              |
| Tb927.3.1560   | TRF-interacting factor 2, TIF-2 | Interacts with the ttaagg binding factor (TRF), protecting it from degradation. Its transient depletion decreases level of TRF and increases frequency of variant surface glycoprotein (VSG) switching and sub-telomeric double-strand breaks (DSB) | Jehi et al. (2016), Jehi et al. (2014b)          |
| Tb927.5.1700   | Replication factor A 28 kDa subunit, RPA-2 | Accumulates at DSB sites, where it forms RPA foci, stabilizing resected DNA and triggering cell cycle arrest, RAD51 accumulation and damage repair. The protein was shown to persist throughout the cell cycle in *T. brucei* and regulate metacyclogenesis in *T. cruzi* | Glover et al. (2019), Pavanai et al. (2016)      |
| Tb927.6.4330   | hypothetical protein            | Affects VSG allelic exclusion                                                                                                                                                                                   | Glover et al. (2016)                            |
| Tb927.9.10770  | Polyadenylate-binding protein 2, PABP-2 | An abundant mRNA binding protein involved in translation initiation and general mRNA metabolism                                                                                                                    | Kramer et al. (2013), Zoltner et al. (2018)      |
| Tb927.9.15360  | 40S ribosomal protein 5S        | Regulates numerous cellular processes in eukaryotes                                                                                                                                                               | Ruvinsky and Meyuhas (2006)                     |
| Tb927.9.5020   | HMG-box domain-containing protein | Generally, these small proteins bind DNA and regulate transcription, replication and DNA repair                                                                                                                                 | Hock et al. (2007)                              |
| Tb927.9.8740   | Double-stranded RNA Binding Domain protein 3, DRBD3 | One of RNA-binding proteins (RBPs) that regulate abundance of the specific subset of mRNAs. Its depletion results in a growth arrest followed by the cell death                                                                 | Estévez (2008)                                  |
| Tb927.10.12850 | ttaagg binding factor, TRF      | Essential for telomere end protection. Its ablation caused drastic reduction of G overhangs and chromosome end fusions without affecting the overall telomere length. Expression of TRF with reduced DNA binding affinity leads to increased VSG switching | Jehi et al. (2014a), Li et al. (2005)            |
| Tb927.10.2520  | PrimPol-like protein 2, PPL-2   | A translesion polymerase accumulating in G2 phase of trypanosome cell cycle and involved in postreplication tolerance of endogenous DNA damage. Its knock-down leads to the cell cycle arrest prior to mitosis in late S/G2 and activation of the DNA damage response | Rudd et al. (2013)                              |
| Tb927.10.6030  | Proteasome Subunit Alpha type-1, PSA-1 | A part of a eukaryotic proteasome 20S catalytic core complex. In parasites, proteosomes are involved in cell differentiation and replication                                                                                             | Paugam et al. (2003)                            |
| Tb927.10.6220  | 5'-3' exoribonuclease D, XRND   | A member of the XRN family of 5'-3' exoribonucleases critical for ensuring the fidelity of cellular RNA turnover in eukaryotes. Its knock-down in *T. brucei* inhibited cell growth, but did not affect 5’ processing of several small RNAs | Li et al. (2006)                                |
| Tb927.11.370   | Repressor Activator Protein 1, RAP-1 | A telomeric protein recruited by TRF. Its depletion led to a de-repression of all VSGs in silent expression sites, without affecting telomere length, and resulted in the increased frequencies of the non-coding telomeric repeat-containing RNA (TERRA) and RNA:DNA hybrids and, subsequently, DSBs in telomeric and subtelomeric loci | Nanavaty et al. (2017), Yang et al. (2009)       |
| Tb927.11.5550  | DNA polymerase θ, Pol θ         | A translesion DNA polymerase involved in the repair of DSBs via microhomology-mediated end joining. Its RNAi-mediated depletion resulted in reduced growth rate without a specific cell cycle arrest, accumulation of DNA damage and chromosome segregation defects, and substantial de-regulation of telomeric VSG genes. Orthologues in *T. cruzi* and *L. infantum* control DNA replication and resistance to oxidative damage | de Lima et al. (2019), Fernández-Orgiler et al. (2016), Leal et al. (2020) |
Table 1. (Continued.)

| Protein ID          | Annotation                              | Protein function                                                                 |
|---------------------|-----------------------------------------|----------------------------------------------------------------------------------|
| Tg027.11.9670       | Telomere-associated protein 1, TelAP-1  | Significantly upregulated in the bloodstream compared to the procyclid forms of T. brucei. Depletion of this protein mis-regulated developmental silencing of VSG silent expression sites. |
| Tg027.23.5030       | Ku70 protein                            | Ku70 proteins play a central role in the classical non-homologous end joining (NHEJ) pathway. In addition, they bind telomeres and facilitate recruitment of telomerase. Ku proteins are conserved across different organisms. |
| Tg027.28.10150      | Ku40 protein                            | Ku40 proteins are involved in telomere maintenance, telomere integrity, and resistance to oxidative stress. |

To the best of our knowledge, there has been very little systematic effort to analyze telomeres in trypanosomatids outside the medically relevant Trypanosoma and Leishmania spp. (Fu et al., 1998; Fu and Barker, 1998a, 1998b; Chiuillo et al., 1999, 2002; Muñoz-Jordán et al., 2001; Janzen et al., 2004; Conte and Campo, 2005; Genest and Borst, 2007). Therefore, we decided to do that for a wide range of trypanosomatids with a special emphasis on largely neglected parasites of insects, which are not pathogenic to humans. We selected more than 20 proteins from a set of recently defined putative trypanosomatid telomere-associated proteins (Reis et al., 2018) for more detailed in silico analyses. For most of these proteins, some functional information is available [Table 1; the TriTrypDB (Aslett et al., 2010) gene IDs are used throughout the text]. The predicted telomere-associated complex appears to be a cohort of proteins with widely variable functions, from ribosome and proteosome subunits to telomerase and even DNA repair proteins (Boulton and Jackson, 1998; Patgum et al., 2003; Riha and Shippen, 2003; Janzen et al., 2004; Dreesen et al., 2005; Ruvinisky and Meyuhas, 2006; Chico et al., 2011; Sandhu et al., 2013; Nenarokova et al., 2019). In this work, we analysed the evolutionary history of telomere-associated proteins in Kinetoplastea, performed a systematic analysis of telomere length variation among trypanosomatids on the dataset, which incorporates a wide range of understudied monoxenous members of the family Trypanosomatidae, and established a correlation between the level of transcription for several analysed telomere-associated proteins and the telomere length.

Materials and methods

In silico analyses

A putative set of telomere-associated proteins of T. brucei brucei TREU927 (Reis et al., 2018) were used as queries for BLAST searches (Altschul et al., 1990) against a dataset of annotated proteins of 64 trypanosomatids and the eubodonid Bodo saltans. First, BLASTp searches were performed with an E-value set to 1 and all the hits with an E-value not exceeding 10^{-15} were retained. If the respective sequence was not identified among annotated proteins, the searches were repeated with the tblASTn algorithm against a database of genome sequences. In case no protein was identified in the genome, HMMER v.3.3 (Eddy, 2009), a more sensitive method for the identification of divergent homologues based on hidden Markov models was employed. Annotated proteins and assembled genome sequences were downloaded from the NCBI Genome (Sayers et al., 2019) and TriTrypDB v. 45/46 (Aslett et al., 2010) databases. The validity of the hits was confirmed using reciprocal BLAST searches against T. brucei proteins Dobson et al., 2006; Hovels-Miner et al., 2012). Moreover, transposable elements are often found in association with telomeres (Pardue et al., 1997; Rahnama et al., 2020). In agreement with this, a sub-telomeric region of Leptomonas pyrrocoris chromosome contains an integrated copy of an RNA-dependent RNA polymerase putatively originating from an RNA virus of the family Tombusviridae infecting this flagellate (Grybchuk et al., 2018), and possibly contributing to the retrotransponson translocation within the trypanosomatid genome. Telomeric regions of kinetoplastid chromosomes also possess several features distinguishing them from their counterpart in most of the other eukaryotes. For example, the telomeres of Trypanosoma brucei increase in length (by approximately 10 bp per generation) until they reach an equilibrium (Bernards et al., 1983; Pays et al., 1983; Horn et al., 2000). In trypanosomatids, a modified nucleobase, base J (β-D-gluco-pyranosyl-oxy-methyl-uracil) is involved in RNA polymerase II transcription termination and is preferentially localized to telomeres (Borst and van Leeuwen, 1997; Genest et al., 2007; van Luenen et al., 2012).

https://doi.org/10.1017/S0031182021000378 Published online by Cambridge University Press
| Species                  | Tb927.2.6100: hypothetical protein | Tb927.3.1560: TRF-interacting factor 2 | Tb927.3.5150: exonuclease, putative | Tb927.5.1700: replication factor A 28 kDa subunit | Tb927.6.4330: telomere-associated protein | Tb927.9.10770: polyadenylate binding protein 2 | Tb927.9.15360: 40S ribosomal protein S6 | Tb927.9.3930: hypothetical protein | Tb927.9.5020: HMG-box domain-containing protein |
|-------------------------|-----------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|
| Crithidia bombi          | +                                 | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Crithidia expoeki        | +                                 | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Crithidia fasciculata CF-Cl | +                               | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Leptomonas seymouri      | +                                 | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Lotmaria passim SF       | +                                 | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Endotrypanum monterogei ATCC30501 | + | + | + | + | + | + | + | + | + |
| Portzia deanei TCC258    | +                                 | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Portzia hertigi TCC260   | +                                 | +                                     | +                                   | +                                             | +                                           | +                                           | +                                           | +                                           | +                                           |
| Leishmania (M.) enrietti LEM3045 | + | + | + | + | + | + | + | + | + |
| Leishmania (M.) macropodum LV756 | + | + | + | + | + | + | + | + | + |
| Leishmania (M.) martinguensis LEM2494 | + | + | + | + | + | + | + | + | + |
| Leishmania (S.) aethiopica HO174 | + | + | + | + | + | + | + | + | + |
| Leishmania (S.) tarentolae ParrotTarII | + | + | + | + | + | + | + | + | + |
| Leishmania (L.) aethiopica L147 | + | + | + | + | + | + | + | + | + |
| Leishmania (L.) tropica L590 | + | + | + | + | + | + | + | + | + |
| Leishmania (L.) arabica LEM1108 | + | + | + | + | + | + | + | + | + |
| Sequence | Description                                                                 |
|----------|-----------------------------------------------------------------------------|
| Tb927.2.6100: | hypothetical protein            |
| Tb927.3.1560: | TRF-interacting factor 2                  |
| Tb927.3.5030: | exounuclease, putative                                    |
| Tb927.5.1700: | replication factor A 28 kDa subunit                          |
| Tb927.6.1760: | telomere-associated protein                                |
| Tb927.6.4330: | telomere-associated protein 2                               |
| Tb927.9.10770: | polyadenylate-binding protein 40S ribosomal S6              |
| Tb927.9.3930: | hypothetical protein                                                   |
| Tb927.9.5020: | hypothetical protein                                                   |
| Tb927.9.5500: | HMG-box domain-containing protein                                |

**Leishmania (L.) turanica** LEM423

- + + + + + + + + + + + +

**Leishmania (L.) gerbilli** LEM452

- + + + + + + + + + + + +

**Leishmania (L.) major Friedlin**

- + + + + + + + + + + + +

**Leishmania (L.) major UV39**

- + + + + + + + + + + + +

**Leishmania (L.) major SD75**

- + + + + + + + + + + + +

**Leishmania (L.) donovani** BPK282A1

- + + + + + + + + + + + +

**Leishmania (L.) infantum JPCM5**

- + + + + + + + + + + + +

**Leishmania (L.) amazonensis** M2269

- + + + + + + + + + + + +

**Leishmania (L.) mexicana** M379

- + + + + + + + + + + + +

**Leishmania (V.) braziliensis** M2903

- + + + + + + + + + + + +

**Leishmania (V.) braziliensis** M2904

- + + + + + + + + + + + +

**Leishmania (V.) peruviana** PAB-4377

- + + + + + + + + + + + +

**Leishmania (V.) panamensis** L13

- + + + + + + + + + + + +

**Novymonas esmeraldas** E262AT

- + + + + + + + + + + + +

**Blastocrithidia sp.** p57

- + + + + + + + + + + + +

**Vickermania ingensplastis** CP21

- + + + + + + + + + + + +

**Phytomonas francoi** TCC064

- + + + + + + + + + + + +

**Phytomonas serpens** 9T

- + + + + + + + + + + + +

**Phytomonas sp.** HART1

- + + + + + + + + + + + +

(Continued)
| Protein Name | Phytomonas sp. EM1 | Wallacemonas collosoma ATCC30261 | Wallacemonas rigidus Sld | Wallacemonas sp. MB04 | Wallacemonas sp. 1955L | Wallacemonas sp. Trypx | Wallacemonas sp. Wsd | Angomonas deanei TCC036E | Angomonas desouzai TCC079E | Strigomonas culica TCC012E | Strigomonas galatii TCC219 | Strigomonas oncophelii TCC290E | Blechomonas ayala BI08-376 | Trypanosoma brucei gambiense DAL972 | Trypanosoma brucei Lister 427 | Trypanosoma evansi STIB_805 | Trypanosoma equiperdum OH_V2 | Trypanosoma congolense IL3000 | Trypanosoma vivax Y486 |
|--------------|-------------------|-----------------------------------|--------------------------|-----------------------|-----------------------|----------------------|---------------------|----------------------|----------------------|----------------------|----------------------|----------------------|---------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Tb927.2.6100: hypothetical protein | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.3.1560: TRF-interacting factor 2 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.3.5150: exonuclease, putative | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.5.1700: replication factor A 28 kDa subunit | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.6.1760: telomere-associated protein | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.9.10770: polyadenylate-binding protein 2 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.9.15360: 40S ribosomal protein S6 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.9.3930: hypothetical protein | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Tb927.9.5020: HMG-box domain-containing protein | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |

(Continued)
|                  | Tb927.2.6100: hypothetical protein | Tb927.3.1560: TRF-interacting factor 2 | Tb927.3.35150: exonuclease, putative | Tb927.5.1700: replication factor A 28 kDa subunit | Tb927.6.1760: telomere-associated protein | Tb927.6.4330b: polyadenylate-binding protein 2 | Tb927.9.10770: 40S ribosomal protein S6 | Tb927.9.3930: hypothetical protein | Tb927.9.4000: hypothetical protein | Tb927.9.5020: HMG-box domain-containing protein |
|------------------|----------------------------------|----------------------------------------|--------------------------------------|-----------------------------------------------|-------------------------------------------|---------------------------------------------|------------------------------------------|-------------------------------------------|------------------------------------------|-----------------------------------------------|
| Trypanosoma cruzi CL-EL | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Trypanosoma cruzi CL-Br NEL | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Trypanosoma cruzi marinkellei BT | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Trypanosoma rangeli SC58 | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Trypanosoma grayi ANR4 | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Trypanosoma theileri Edinburgh | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Paratrypanosoma confluens Cul13 | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |
| Bodotria saltans Lake_Konstanz | +                                | +                                      | +                                    | +                                             | +                                         | +                                           | +                                       | +                                         | +                                       | +                                             |

|                  | Tb927.9.8740: double RNA binding domain protein 3 | Tb927.10.12850: ttagg binding factor | Tb927.10.15200: PrimPol-like protein 2 | Tb927.10.24002: hypothetical protein | Tb927.10.42200: telomerase reverse transcriptase | Tb927.11.10190: hypothetical protein | Tb927.11.16120: repressor activator protein 1 | Tb927.11.16150: DNA polymerase theta | Tb927.11.9870: telomere-associated protein 1 |
|------------------|------------------------------------|--------------------------------------|----------------------------------------|-----------------------------------------------|---------------------------------------------|-------------------------------------------|------------------------------------------|------------------------------------------|-----------------------------------------------|
| Crithidia bombi 08.076 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Crithidia expoeki BJ80.175 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Crithidia fasciculata CF-C1 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Leptomonas pyrrhocoris H10 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Leptomonas seymouri ATCC30220 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Lotmaria passim SF | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Endotrypanum monterogenei ATCC30507 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Endotrypanum monterogenei LV88 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |
| Porcisia deanei TCC258 | +                                | +                                    | +                                    | +                                             | +                                           | +                                         | +                                       | +                                         | +                                             |

(Continued)
| Protein Name | Accession | Description |
|--------------|-----------|-------------|
| Tb927.9.8740: double RNA binding domain protein 3 | | |
| Tb927.10.12850: ttagg binding factor | | |
| Tb927.10.2200: hypothetical protein | | |
| Tb927.10.2520: PrimPol-like protein 2 | | |
| Tb927.10.4220: hypothetical protein | | |
| Tb927.10.6030: proteasome subunit alpha type-1 | | |
| Tb927.10.6220: 5'-3' exoribonuclease D | | |
| Tb927.11.10190: telomerase reverse transcriptase | | |
| Tb927.11.16120: hypothetical protein | | |
| Tb927.11.370: repressor activator protein 1 | | |
| Tb927.11.5550: DNA polymerase theta | | |
| Tb927.11.9870: telomere-associated protein 1 | | |

Porcisia hertigi TCC260
Leishmania (M.) enriettii LEM3045
Leishmania (M.) macropodum LV756
Leishmania (M.) martinsiquensis LEM2494
Leishmania (S.) aedleri HO174
Leishmania (S.) tarentolae ParrotTarli
Leishmania (L.) aethiopica L147
Leishmania (L.) tropica L590
Leishmania (L.) arabica LEM1108
Leishmania (L.) turonica LEM423
Leishmania (L.) gerbilli LEM452
Leishmania (L.) major Friedlin
Leishmania (L.) major LV39
Leishmania (L.) major SD75
Leishmania (L.) donovani BPK382A1
Leishmania (L.) infantum JPCM5
Leishmania (L.) amazonensis M2269
Leishmania (L.) mexicana M379

(Continued)
| **Table 2.** (Continued.) | Tb927.9.8740: double RNA binding domain protein 3 | Tb927.10.12850: taggg binding factor | Tb927.10.2200: hypothetical protein | Tb927.10.2520: PrimPol-like protein 2 | Tb927.10.4220: hypothetical protein | Tb927.10.6030: proteasome subunit alpha type-1 | Tb927.10.6220: 5'-3' exoribonuclease D | Tb927.11.10190: telomerase reverse transcriptase | Tb927.11.16120: hypothetical protein 1 | Tb927.11.370: repressor activator protein 1 | Tb927.11.5550: DNA polymerase theta | Tb927.11.9870: telomere-associated protein 1 |
|-------------------------------|-------------------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| **Leishmania (V.) braziliensis M2903** | + | + | - | + | + | + | + | + | + | + |
| **Leishmania (V.) braziliensis M2904** | + | + | - | + | + | + | + | + | + | + |
| **Leishmania (V.) peruviana PAB4377** | + | + | - | + | + | + | + | + | + | + |
| **Leishmania (V.) panamensis L13** | + | + | - | + | + | + | + | + | + | + |
| **Novymonas esmeraldas E262AT** | + | + | - | + | + | + | + | + | + | + |
| **Blastocrithidia sp. p57** | + | + | - | + | + | + | + | + | + | + |
| **Vickermania ingensoplastis CP21** | + | + | - | + | + | + | + | + | + | + |
| **Phytomonas francai TCC064** | + | + | - | + | + | + | + | + | + | + |
| **Phytomonas serpens E2** | + | + | - | + | + | + | + | + | + | + |
| **Phytomonas sp. HART1** | + | + | - | + | + | + | + | + | + | + |
| **Phytomonas sp. EM1** | + | + | - | + | + | + | + | + | + | + |
| **Wallacemonas collosoma ATCC30261** | + | + | - | + | + | + | + | + | + | + |
| **Wallacemonas rigida Sld** | + | + | - | + | + | + | + | + | + | + |
| **Wallacemonas sp. MBr04** | + | + | - | + | + | + | + | + | + | + |
| **Wallacemonas sp. 1895L** | + | + | - | + | + | + | + | + | + | + |
| **Wallacemonas sp. Trypx** | + | + | - | + | + | + | + | + | + | + |
| **Wallacemonas sp. Wsd** | + | + | - | + | + | + | + | + | + | + |
| **Angomonas deanei TCC036E** | + | + | - | + | + | + | + | + | + | + |
| **Angomonas desouzai TCC079E** | + | + | - | + | + | + | + | + | + | + |

(Continued)
| Protein | Strigomonas culicis TCC012E | Strigomonas galati TCC219 | Strigomonas ancapelli TCC290E | Blechomonas gyalai B08-376 | Trypanosoma brucei gambiense DL1972 | Trypanosoma brucei Lister 427 | Trypanosoma evansi STIB_805 | Trypanosoma equiperdum QM_V2 | Trypanosoma congolense IL3000 | Trypanosoma vivax Y486 | Trypanosoma cruzi CL-EL | Trypanosoma cruzi CL-BR NEL | Trypanosoma cruzi marinkellei B7 | Trypanosoma rangeli SC58 | Trypanosoma groi ANR4 | Trypanosoma lehedi Edinburgh | Paratrypanosoma confusum CUL13 | Bodo saltans Lake_Konstanz |
|---------|-----------------------------|--------------------------|-------------------------------|-----------------------------|--------------------------------|--|------------------|--------------------------|-----------------------------|-------------------|-----------------|-----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Tb927.9.8740: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| double RNA binding domain protein | Tb927.10.12850: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| ttaggg binding factor | Tb927.10.2200: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| hypothetical protein | Tb927.10.2520: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| PrimPol-like protein | Tb927.10.4220: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| hypothetical protein | Tb927.10.6030: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| proteasome subunit alpha type-1 | Tb927.10.6220: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| 5′-3′ exoribonuclease D | Tb927.11.10190: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| telomerase reverse transcriptase | Tb927.11.16120: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| hypothetical protein | Tb927.11.370: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| repressor activator protein 1 | Tb927.11.5550: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| DNA polymerase theta | Tb927.11.9870: | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| telomere-associated protein 1 | Species analysed by Southern blotting are shaded. | +, identified; empty, not identified; *, identified in strain UA301. |
and alignments including the query and all identified proteins, if necessary. The resulting gene presence/absence table and a cladogram manually written in a Newick format based on recent publications (Butenko et al., 2019; Kostygov et al., 2014, 2020; Kostygov and Yurchenko, 2017; Lukeš et al., 2018; Frolov et al., 2019; Kato et al., 2019) were used for Dollo parsimony analysis in the Count software (Csúrós, 2010) and results were visualized in a graphical editor.

Fig. 1. Gains and losses of genes encoding putative telomere-associated proteins in kinetoplastids.
Trypanosomatid isolates and cultivation

Cultures of Crithidia exoekii (BJ08.175), C. fasciculata (Cf-Cl), C. fasciculata (ATCC20358), C. termophila (ATCC30817), L. pyrrhocoris (H10), L. seymouri (ATCC30220), Novymonas esmeraldas (E262AT), Strigomonas oncopelti (TCC290E) and Zelonia costarcensis (15EC) were grown in BHI medium (Oxoid/Thermo Fisher Scientific, Basingstoke, UK) supplemented with 2 μg mL⁻¹ Hemin (Sigma-Aldrich, St. Louis, USA) and 50 units mL⁻¹ of Penicillin/Streptomycin (BioSera, Nuaillé, France) at 23°C. Cultures of Endotrypanum monterogeii (ATCC30507), Herpetomonas samuelpessoai, Leishmania (L. m. major) (LV39), L. (L.) amazonensis (Josefa, LV78, LV79 and PH8), L. (L.) mexicana (Josefa, LV78, LV79 and PH8), L. (L.) martiniensis (LEM2494), L. (M. orientalis) (PCM2), L. (Sauroleishmania) tarentolae (ParrotTarII), L. (V. braziliensis) (human2017), L. (V.) guyanensis (M4147), Phytomonas sp. (EM1), Porcisia deanei (TCC258) and P. hertigi (TCC260) were grown in M199 medium supplemented with 2 μg mL⁻¹ Bioterin, 2 μg mL⁻¹ Hemin (all Sigma-Aldrich), 25 mM HEPES (Lonza, Basel, Switzerland), 50 units mL⁻¹ of Penicillin/Streptomycin (BioSera), and 10% fetal bovine serum (BioWest, Nuaillé, France) at 23°C. Cultures of Angomonas deanei (CT-I0C 044), A. desouzai (CT-I0C 109), Blastocrithidia sp. (p57), Blastocrithidia triatomae, Blechomonas ayalai (B08-376), Blechomonas pulvixeans (ATCC50186), Herpetomonas muscarum (MOM-01), Jaenimonas sp. (Finn-01.02), Kentomonas sorgosonicus (MF-08.02), Lafontella sp. (GMO-01), Vickermania ingenoplastis (CP21), Wallacomonas colosoma (ATCC30261) and W. rigidus (PL11) were maintained in SDM medium (BioWest) supplemented with 10% fetal bovine serum (BioWest) and 50 units mL⁻¹ of Penicillin/Streptomycin (BioSera) at 23°C. In the cases of Lafontella and Endotrypanum, cultures were grown in a bi-phasic medium, overlaying blood agar. All species were validated by amplifying and sequencing the 18S rRNA gene as described previously (Kostygov et al., 2014).

Quantification of transcription level of genes encoding telomeric proteins using RT-qPCR

RNA was isolated and transcript levels of the telomeric proteins were assessed by RT-qPCR as described previously (Záhonová et al., 2014; Kraeva et al., 2019). Sequences of the specific primers for L. mexicana/amazonensis orthologues of T. brucei genes are listed in Supplementary Table 1. Expression values were normalized to those of 18S rRNA.

Southern blotting

The previously established terminal restriction fragment analysis of telomere lengths protocol was followed (Janzen et al., 2004). In brief, total genomic DNA from the log-phase grown cells was isolated and digested with Alul, HinfI and RsaI overnight. Restriction fragments were separated in 0.75% agarose gel, transferred to a ZetaProbe blotting membrane (Bio-Rad, Hercules, USA), probed with the DIG-labelled telomeric probe [CCCTAA]x25 in the DIG Easy Hyb buffer (Roche Diagnostics, Indianapolis, USA), and visualized with the DIG Luminescent Detection Kit (Roche Diagnostics). The probe was labelled by the Dioxigenin NT Labeling Kit (Jena Bioscience GmbH, Jena, Germany). Statistics of the telomere lengths were obtained with an online tool WALTER (Web-based Analyser of the Length of TelomereRes) (Lvčka et al., 2021). For the loading and integrity control in the L. mexicana complex analysis, DNAs were processed as above, and the membrane was probed against a fragment of a gene encoding telomerase (LmxM.36.3930) (Supplementary Table 1).
Table 3. Telomere lengths (weighted median, minimum–maximum) in selected species of Trypanosomatidae

| Species                                      | Median (min-max) of telomere length, bp |
|----------------------------------------------|----------------------------------------|
| Crithidia expoeckii BJ08.175                 | 4,271 (1619–33446)                     |
| C. fasciculata ATCC20358                     | 506 (252–1047)                         |
| C. fasciculata CI-CI                        | 368 (252–983)                          |
| C. termophila ATCC30817                     | 1374 (512–3655)                        |
| Leptomonas pyrrhocoris H10                  | 874 (328–4859)                         |
| L. seymouli ATCC30220                       | 875 (386–2941)                         |
| Endotrypanum monteroei ATCC30507            | 916 (450–2182)                         |
| Percisia deanei TCC258                      | 1875 (705–10469)                       |
| P. hertigi TCC260                           | 4992 (1305–27381)                      |
| Leishmania (Mundinia) martiniquensis LEM2494| 2519 (1009–8231)                       |
| L. (M.) orientalis PC22                     | 1488 (1033–2327)                       |
| L. (Saura)leishmania tarentolae ParrotTaro   | 3938 (1263–27381)                      |
| L. (Leishmania) major LV39                  | 1842 (573–13381)                       |
| L. (L.) amazoneensis LV78                   | 435 (247–779)                          |
| L. (L.) amazoneensis LV79                   | 3363 (252–34459)                       |
| L. (L.) amazoneensis PH8                    | 362 (253–1260)                         |
| L. (L.) amazoneensis Josefa                 | 443 (253–2441)                         |
| L. (L.) mexicana M379                       | 393 (271–840)                          |
| L. (L.) mexicana M379 ΔKu80                | 630 (248–3463)                         |
| L. (L.) mexicana M379 ΔKu70                 | 521 (246–3421)                         |
| L. (Vannia) braziliensis human2017          | 1911 (587–7865)                        |
| L. (V.) guyanensis M4417                   | 5105 (1782–27381)                      |
| Novynomason esmeraldas E262AT              | 1238 (536–15198)                       |
| Zelonia costaricensis 15EC                  | 937 (264–3818)                         |
| Blastocrithidia sp. p57                    | 1500 (478–10944)                       |
| B. triatomae                               | 614 (390–1011)                         |
| Vickermania ingenoplaxis CP21              | 4078 (815–33440)                       |
| Herpetomonas muscarnar MMO-01              | 1282 (388–4675)                        |
| H. samuelpessouai                          | 1989 (862–32804)                       |
| Lofantella sp. GMO-01                      | 1169 (708–3080)                        |
| Phytonomas sp. EM1                         | 2554 (1507–4906)                       |
| Jaenimonas sp. Finn-01.02                 | 2112 (879–5679)                        |
| Wolacemoson collosoma ATCC30261            | 1290 (466–7332)                        |
| W. rigidus PL11                            | 1253 (408–9550)                        |
| Angomonas deanei CT-IOC 044                | 1245 (466–3948)                        |
| A. desouzaï CT-IOC 109                    | 588 (294–1334)                         |
| Strigomonas ancopelti TCC290E              | 420 (282–600)                          |
| Kentomonas soroegonis MF-08.02             | 1017 (384–6178)                        |
| Blechomonas ayalai BO8-376                 | 1693 (1097–3152)                       |
| Ble. pyluscellantis ATCC50186              | 1972 (687–8889)                        |
| Trypanosoma brucei brucei Lister 427 (BF)  | 3422 (474–24711)                       |
| T. b. brucei Lister 427 29-13 (PF)         | 3108 (470–24281)                       |
| T. mego CP029                              | 414 (252–715)                          |

Results and discussion

The core set of proteins putatively involved in telomere maintenance in kinetoplastids is conserved

To study the phylogenetic distribution of proteins predicted to be involved in telomere maintenance (Reis et al., 2018), we analysed the presence/absence of the corresponding 24 genes in the available genomes of trypanosomatids and their close eubodonid relative, B. saltans (Table 2). Most of the studied proteins (20 of 24) are well conserved and we consider them as a core set putatively involved in telomere maintenance in kinetoplastids. It is worth noting that the telomere association and function in telomere maintenance has already been confirmed for some of these proteins, while some others have not been functionally characterized yet. Thus, the composition of the core set of proteins involved in telomere maintenance, as defined previously (Reis et al., 2018) and discussed herein, should be taken with caution. Despite the fact that most of the respective genes are conserved across Kinetoplastea (Fig. 1, Table 2) and, thus appear to be present in the kinetoplastid common ancestor, we came across several interesting exceptions that are discussed in detail below.

A set of three proteins (orthologues of T. brucei Tb927.3.1560, Tb927.9.5020 and Tb927.11.370) was acquired by the common ancestor of trypanosomatids upon the separation from bodonids (Fig. 1). One of them, Tb927.3.1560 [TIF-2, an orthologue of mammalian TINF2 (Jehi et al., 2014a; 2014b)] was suggested to be essential, as it is involved in shelterin (a protein complex implicated in telomere protection) assembly and telomerase-mediated telomere length maintenance in other organisms (Walne et al., 2008; Frank et al., 2015). Yet, it is not present in bodonids and was secondarily lost in all other trypanosomatids outside of the genera Paratrypanosoma, Trypanosoma and Blechomonas (Fig. 1), raising a question of how do they cope with its absence or whether they replaced it with a functional analogue? Tb927.2.6100 is Trypanosoma-specific, confirming previous report (Beck et al., 2013). Surprisingly, this protein was shown to be specifically associated with kDNA, so its role in telomere maintenance, if any, remains to be elucidated by functional genetics approaches. Two proteins (orthologues of T. brucei Tb927.9.3930 and Tb927.9.4000) are present only in four species of the T. brucei group and may determine specific traits of these parasites.

An orthologue of Tb927.11.9870 (TelAP-1) is present in most species, but it is conspicuously absent from the representatives of two monoxenous groups (Blastocrithidia and Vickermania spp.) and most Phytomonas spp., plant pathogens with streamlined genomes (Porcel et al., 2014). While we cannot rule out a possibility that the protein is divergent beyond recognition by available bioinformatics tools, there may exist another component fulfilling the role of TelAP-1 in these species.

Of special attention is the absence of Tb927.3.5030 (Ku70) and Tb927.3.5030 (Ku80) orthologues in Blastocrithidia sp., which has a non-canonical nuclear genetic code with all three stop codons reassigned to encode amino acids (Záhonová et al., 2016). It has been recently proposed that such an absence may lead to the accumulation of numerous insertions in many protein-coding genes of these organisms (Nenarokova et al., 2019).

Trypanosomatid telomeres are variable in length

We performed a systematic screen of the telomere length across Trypanosomatidae by Southern blotting (Fig. 2, Table 3). Our analysis revealed that monoxenous Leishmaniniae (Kostygov and Yurchenko, 2017) of the genera Leptomonas, Novyomonas and Zelonia have fairly short telomeres (weighted medians
900–1200 bp; hereafter only rounded weighted median data are compared in the text, see Table 3 for minimum and maximum values), while telomeres in analysed *C. fasciculata* CF-C1 to 4300 bp in *C. expoeki*. These numbers correlate well with previous reports on telomere length in the selected representatives of the genera *Crithidia*, *Leishmania* and *Trypanosoma* (Genest et al., 2007). Of note, the repertoire of genes implicated in telomere maintenance is identical in these flagellates (Table 2), so these differences can be explained by either the presence of other proteins involved in this process, or (more likely) differences in gene expression. Telomeres of *Blechomonas*, *Herpetomonas*, *Jaenimonas* and *Wallacemonas* spp. are 1300–2100 bp long. The endosymbiont-containing *Strigomonadinae* [*Angomonas*, *Kontomonas* and *Strigomonas* spp. (Votýpka et al., 2014)] differ in telomere length, with *S. oncopelti* bearing the shortest chromosome ends of ~400 bp.

Representatives of three genera (*Blastocrithidia*, *Leishmania* and *Trypanosoma*) deserved special attention. Uniquely among trypanosomatids, *Blastocrithidia* spp. lack Ku proteins (Nenarokova et al., 2019), yet their telomeres are of similar length to telomeres of other trypanosomatids (600 and 1500 bp in *B. triatomae* and *Blastocrithidia* sp., respectively), arguing that either Ku proteins are dispensable for the telomere length maintenance in these species, or their loss can be compensated by other factors. Telomere sizes vary in different *Trypanosoma* spp. represented by short telomeres in *T. mega* (400 bp) and substantially longer telomeres in *T. brucei* Lister 427 (3100–3400 bp). In contrast to the previous report (Dreesen and Cross, 2008), we did not document differences in telomeres’ length between the procyclic and

![Fig. 3. Transcript levels of telomere-associated proteins and telomere lengths in the species of *L. mexicana* complex. (A) Quantitative RT-PCR analysis of the core set of proteins implicated in telomere maintenance. Gene expression is presented as normalized means and standard deviations of three replicates. Data are presented in two graphs to account for differences in expression values. (B, C) Southern blotting analysis of telomere repeats (B) and telomerase-encoding gene (C, used as an additional DNA integrity control) in *L. amazonensis* LV78, LV79, PH8, Josefa and *L. mexicana* M379. Marker sizes are indicated on the left. DNA integrity controls are presented in Supplementary Fig. 1 (right panel).](https://doi.org/10.1017/S0031182021000378 Published online by Cambridge University Press)
bloodstream stages of *T. brucei*. However, both strains in our ana-
lysis have the same origin (Lister 427), while the abovementioned study compared Lister 427 and TREU927 strains. Similar to the cases discussed above, despite possessing the same repertoire of telomere-bound proteins, the distribution of telomere sizes in the *Leishmania–Pbrasia–Endotrypanum* clade (Espinosa et al., 2018) is wide, exemplified by two extreme cases of *P. hertigi* (5000 bp) and *L. mexicana* (400 bp, Fig. 2). Variable telomere length in *Leishmania* spp. (and possibly other Leishmaniinae) may be explained by the presence of a stress-sensitive telomere-
proximal replication activity outside S phase of the cell cycle in these species (Damasceno et al., 2020, 2021).

**RNA level of telomerase and several telomere-associated proteins correlates with telomere length in the species of *L. mexicana* complex**

We analysed telomere length and expression of the core set of proteins putatively involved in telomere maintenance in closely related species forming the *L. mexicana* complex (Eresh et al., 1994). Similar to the cases discussed above, telomeres in *L. mexicana* and four isolates of *L. amazonensis* greatly differed in length from ~400 bp in *L. mexicana* M379 to ~3400 bp in *L. amazonen-
sis* LV79 (Fig. 3, Table 3, Supplementary Fig. 1). Such a wide range of telomere lengths correlated well with the expression of the *Leishmania* spp. telomerase (orthologue of Tb927.11.10190) and a telomere-associated hypothetical protein (orthologue of Tb927.6.4330). The higher expression of these proteins correlated with longer telomeres. The specific roles of these and other pro-
teins remain to be further elucidated by functional studies.

**Conclusions**

The genome analysis has allowed us to identify a core set of 20 conserved proteins predicted to be responsible for telomere main-
tenance in trypanosomatids. Several proteins, previously identified in *T. brucei* pull-downs, are trypanosome-specific. Out of 20 proteins conserved in Trypanosomatidae, two (Ku70 and Ku80) are conspicuously missing in *Blastocritidia* spp., yet telo-
meres in these species do not appear to be affected by their loss. We documented that telomeres of trypanosomatids are diverse in length, even within groups of closely related species. One such group is a complex of species, related to *L. mexicana*. Our analysis demonstrated that the expression of several telomere-associated proteins correlates with the documented differences in telomere length within species of the *L. mexicana* complex, which is indi-
cative of a potential role these proteins may play in the telomere length maintenance.

**Supplementary material.** The supplementary material for this article can be found at [https://doi.org/10.1017/S0031182021000378](https://doi.org/10.1017/S0031182021000378).

**Acknowledgements.** We thank Dr A Kostygov (University of Ostrava) for his advice on trypanosomatid cultivation, Dr P. Volf and Dr J. Sádlová (both Charles University) for providing *Leishmania* spp. isolates, Dr C. Janzen (University of Würzburg) for sharing the plasmid pSP73 bearing the telomeric probe, and Dr C. Janzen and Dr R. McCulloch (University of Glasgow) for specific advice on methodology.

**Financial support.** Computational resources were funded by the project ‘e-Infrastruktura CZ’ (e-INFRA LM2018140) provided within the program ‘Projects of Large Research, Development and Innovations Infrastructures’. This work was supported by the Grant Agency of Czech Republic (18-15962S and 20-07186S) and the European Regional Funds (project ‘Centre for Research of Pathogenicity and Virulence of Parasites’ CZ.02.1.01/ 16_019/0000759) to V.Y. and J.L., and the grant SGS/PrF/2021 from the University of Ostrava (to E.P. and A.T.S.A.). A.T.S.A. and E.P. were also sup-
ported by the Moravskoslezský kraj research initiative (RRC/02/2020) and City of Ostrava (0493/2M1822/8), respectively. A part of this work on telo-
mere diversity in *Leishmania* spp. was funded by the Russian Science Foundation (grant 19-15-00054) to V.Y. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Conflict of interest.** None.

**Ethical standards.** Not applicable.

**References**

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.

Aphasizheva I, Alfonzo J, Carnes J, Cestari I, Cruz-Rayes J, Goringer HU, Hajduk S, Lukei J, Madison-Antenucci S, Maslov DA, McDermott SM, Ochsnerreiter T, Read LK, Salavati R, Schnauffer A, Schneider A, Simpson L, Stuart K, Yurchenko V, Zhou ZH, Ziková A, Zhang L, Zimmer S and Aphasizheva R (2020) Lexas and grammar of mitochondrial RNA processing in trypanosomes. *Trends in Parasitology* 36, 337–355.

Aslett M, Aurecrocechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardiner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissing JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myller PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasanmooorthy G, Stoeckert CJ Jr., Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G and Wang H (2010) TriTrypDB: a func-
tional genomic resource for the Trypanosomatidae. *Nucleic Acids Research* 38, D457–D462.

Beck K, Acestor N, Schuler F, Anupama A, Carnes J, Panighri AK and Stuart K (2013) *Trypanosoma brucei* Tb927.2.6100 is an essential protein associated with kinetoplast DNA. *Eukaryotic Cell* 12, 970–978.

Bernards A,Michals PA, Lincke CR and Borst P (1983) Growth of chromo-
some ends in multiplying trypanosomes. *Nature* 303, 592–597.

Borst P and van Leeuwen F (1997) beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. *Molecular and Biochemical Parasitology* 90, 1–8.

Boulton SJ and Jackson SP (1998) Components of the Ku-dependent non-
homologous end-joining pathway are involved in telomere length mainten-
ance and telomeric silencing. *EMBO Journal* 17, 1819–1828.

Butenko A, Kostygov AY, Sádlová J, Kleschenyov Y, Bečvár T, Podešvová I, Macedo DH, Zihala D, Lukei J, Bates PA, Volf P, Oppedores FR and Yurchenko V (2019) Comparative genomics of *Leishmania* (*Mundinia*). *BMC Genomics* 20, 726.

Campeolo R, Diaz Lozano I, Figarella K, Diaz Lozano I, Figarella K, Osuna A and Ramirez JL (2015) *Leishmania major* telomerase TERT protein has a nuclear/mitochondrial eclipsed distribution that is affected by oxidative stress. *Infection and Immunity* 83, 57–66.

Chico L, Ciudad T, Hua M, Lue NF and Larriba G (2011) The Candida albicans Ku70 modulates telomere length and structure by regulating both telo-
merase and recombination. *PLoS ONE* 6, e23732.

Chiurillo MA, Cano I, Da Silveira JF and Ramirez JL (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* 100, 173–183.

Chiurillo MA, Peralta A and Ramirez JL (2002) Comparative study of *Trypanosoma rangeli* and *Trypanosoma cruzi* telomeres. *Molecular and Biochemical Parasitology* 120, 305–308.

Clayton C (2019) Regulation of gene expression in trypanosomatids: living with polycistrionic transcription. *Open Biology* 9, 190072.

Conte FF and Cano MI (2005) Genomic organization of telomeric and sub-
telomeric sequences of *Leishmania* (*Leishmania*) amazonensis. *International Journal for Parasitology* 35, 1435–1443.

Csúrós M (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics (Oxford, England)* 26, 1910–1912.

Damasceno JD, Marques CA, Beraldi D, Crouch K, Lapsley C, Obonaga R, Tosi LR and McCulloch R (2020) Genome duplication in *Leishmania major* relies on persistent subtelomeric DNA replication. *elife* 9, e58030.

Damasceno JD, Marques CA, Black J, Briggs E and McCulloch R (2021) Read, write, adapt: challenges and opportunities during kinetoplastid gen-
ome replication. *Trends in Genetics* 37, 21–34.
de Lima LP, Calderano SG, da Silva MS, de Araujo CB, Vasoncelos EJR, Iwai LK, Pereira CA, Fragoso SP and Elias MC (1999) Ortholog of the polymerase theta helicase domain modules DNA replication in Trypanosoma cruzi. Scientific Reports 9, 2888.
Dobson DE, Scholtes LD, Myler PJ, Turco SJ and Beverley SM (2006) Genomic organization and expression of the expanded SC versus LR gene family. Proceedings of the National Academy of Sciences of the USA 103, 1506–1515.
Hackett JA and Greider CW (2002) Balancing instability: dual roles for telomerase and telomere dysfunction in tumorigenesis. Oncogene 21, 619–626.
Hock R, Furusawa T, Ueda T and Bustin M (2007) HMG chromosomal proteins in development and disease. Trends in Cell Biology 17, 72–79.
Horn D, Spence C and Ingram AK (2000) Telomere maintenance and length regulation in Trypanosoma brucei. EMBO Journal 19, 2352–2359.
Hovd-Miner GA, Boothroyd CE, Mugnier M, Deesens O, Cross GA and Pappazissis FN (2012) Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. PLoS Pathogens 8, e1002990.
Janzen CJ, Lander F, Deesens O and Cross GA (2004) Telomere length regulation and transcriptional silencing in Kus60-deficient Trypanosoma brucei. Nucleic Acids Research 32, 6575–6584.
Jehi SE, Li X, Sandhu R, Ye F, Benmerzouga I, Zhang M, Zhao Y and Li B (2014a) Suppression of subtelomeric VSG switching by Trypanosoma brucei TRF requires its TTAGGG repeat-binding activity. Nucleic Acids Research 42, 12499–12511.
Jehi SE, Wu F and Li B (2014b) Trypanosoma brucei TIF2 suppresses VSG switching by maintaining subtelomere integrity. Cell Research 24, 870–885.
Jehi SE, Nanavaty V and Li B (2016) Trypanosoma brucei TIF2 and TRF suppress VSG switching using overlapping and independent mechanisms. PLoS ONE 11, e0156746.
Kato H, Caceras AG, Seki C, Silipani Garcia CR, Holguin Maurici C, Castro Martínez SC, Moreno Paico D, Castro Muniz JL, Troyes Rivero LD, Villegas Briones ZI, Guerrero Quincho S, Sulca Jayo GL, Tineo Villafuerte E, Manrique de Lara Estrada G, Arias FR, Passara FS, Ruelas Llenera N, Kubo M, Tabbabi A, Yamamoto DS and Hashiguchi Y (2019) Further insight into the geographic distribution of Leishmania species in Peru by cytobrome b and mannose phosphate isomerase gene analyses. PLoS Neglected Tropical Diseases 13, e0074946.
Kostyov AG and Yurchenko V (2017) Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). Folia Parasitologica 64, 020.
Kostyov AG, Grybchuk-leremenko A, Malysheva MN, Frolov AO and Yurchenko V (2014) Molecular revision of the genus Wallacea. Protist 165, 594–609.
Kostyov AG, Frolov AO, Malysheva MN, Ganyukova AI, Chistyakova LV, Tashyeva D, Tesarová M, Spodareva VV, Yurchenko V and Kostyov AG (2019) Development of Phytomonas lipae sp. n. (Kinetoplastea: Trypanosomatidae) in the true bug Coreus marginatus (Heteroptera: Coreidae) and insights into the evolution of life cycles in the genus Phytomonas. PLoS ONE 14, e0214484.
Fu G and Barker DC (1998a) Characterisation of Leishmania telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence. Nucleic Acids Research 26, 2161–2167.
Fu G and Barker DC (1998b) Rapid cloning of telomere-associated sequence using primer-tagged amplification. Biotechniques 24, 386–390.
Fu G, Perona-Wright G and Barker DC (1998) Leishmania braziliensis: characterisation of a complex specific subtelomeric repeat sequence and its use in the detection of parasites. Experimental Parasitology 90, 236–243.
Fulnečková J, Ševčíková T, Fajkus J, Lukašová A, Lukaš M, Vlček Č, Lang BF, Kim E, Eliáš M and Šýkrová E (2013) A broad phylogenetic survey unravels the diversity and evolution of telomers in eukaryotes. Genome Biology and Evolution 5, 468–483.
Genest PA and Borst P (2007) Analysis of telomere length variation in Leishmania over time. Molecular and Biochemical Parasitology 151, 213–215.
Genest PA, Ter Riet B, Cijouss T, van Luenen HG and Borst P (2007) Telomeric localization of the modified DNA base J in the genome of the protozoan parasite Leishmania. Nucleic Acids Research 35, 2116–2124.
Glover L, Hutchinson S, Alsford S and Horn D (2016) VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. Proceedings of the National Academy of Sciences of the USA 113, 7225–7230.
Glover L, Marques CA, Suska O and Horn D (2019) Persistent DNA damage foci and DNA replication with a broken chromosome in the African trypanosome. MBio 10, e01252–e01219.
Greider CW (1990) Telomeres, telomerase and senescence. Bioessays 12, 363–369.
Greider CW and Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405–413.
Grybchuk D, Akopyants NS, Kostyov AG, Konovalovas A, Lye LF, Dobson DE, Zangger H, Fasel N, Butenko A, Frolov AO, Votýpka J, d'Avila-Leyvi CM, Kulich P, Moravcová J, Plevka P, Rogozin IB, Serva S, Lukej K, Beverley SM and Yurchenko V (2018) Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite Leishmania. Proceedings of the National Academy of Sciences of the USA 115, E506–E515.
Reis H, Schwedes M, Dietz S, Janzen CJ and Butter F (2018) TelAP1 links telomere complexes with developmental expression site silencing in African trypanosomes. Nucleic Acids Research 46, 2820–2833.

Riha K and Shippen DE (2003) Ku is required for telomeric C-rich strand maintenance but not for end-to-end chromosome fusions in Arabidopsis. Proceedings of the National Academy of Sciences of the USA 100, 611–615.

Rudd SG, Glover L, Jozwiakowski SK, Horn D and Doherty AJ (2013) PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African trypanosome. Molecular Cell 52, 554–565.

Ruvinsky I and Meyuhas O (2006) Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. Trends in Biochemical Sciences 31, 342–348.

Sandhu R, Sanford S, Basu S, Park M, Pandya UM, Li B and Chakrabarti K (2013) A trans-spliced telomerase RNA dictates telomere synthesis in Trypanosoma brucei. Cell Research 23, 537–551.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Caneke K, Clark K, Connor R, Fiorini N, Funk K, Hefferson T, Holmes JH, Kim S, Kimchi A, Kitts PA, Latham S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD and Ostell J (2019) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 47, D23–D28.

Stuart K, Brun R, Croft S, Fairlamb A, Gurtler RE, McKerrow J, Reed S and Tarleton R (2008) Kinetoplastids: related protozoan pathogens, different diseases. Journal of Clinical Investigation 118, 1301–1310.

Szöör B, Haanstra JR, Gualdrón-López M and Michels PA (2014) Evolution, dynamics and specialized functions of glycosomes in metabolism and development of trypanosomatids. Current Opinion in Microbiology 22, 79–87.

Tomáška L, Nosek J, Kar A, Willcox S and Griffith JD (2019) A new view of the t-loop junction: implications for self-primed telomere extension, expansion of disease-related nucleotide repeat blocks, and telomere evolution. Frontiers in Genetics 10, 792.

van Luenen HG, Farris C, Jan S, Genest P, Tripathi A, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramsamy G, Vainio S, Heidebrecht T, Perrakis A, Page I, van Steensel B, Myler PJ and Borst P (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell 150, 909–921.

Votýpka J, Krasilnikova M, Záhonová K, Hadariová L, Vacula R, Yurchenko V, Eliáš M, Krajčovič J and Yurchenko V (2016) Ruvinsky I and Meyuhas O (2006) Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. Trends in Biochemical Sciences 31, 342–348.

Sandhu R, Sanford S, Basu S, Park M, Pandya UM, Li B and Chakrabarti K (2013) A trans-spliced telomerase RNA dictates telomere synthesis in Trypanosoma brucei. Cell Research 23, 537–551.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Caneke K, Clark K, Connor R, Fiorini N, Funk K, Hefferson T, Holmes JH, Kim S, Kimchi A, Kitts PA, Latham S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD and Ostell J (2019) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 47, D23–D28.

Stuart K, Brun R, Croft S, Fairlamb A, Gurtler RE, McKerrow J, Reed S and Tarleton R (2008) Kinetoplastids: related protozoan pathogens, different diseases. Journal of Clinical Investigation 118, 1301–1310.

Szöör B, Haanstra JR, Gualdrón-López M and Michels PA (2014) Evolution, dynamics and specialized functions of glycosomes in metabolism and development of trypanosomatids. Current Opinion in Microbiology 22, 79–87.

Tomáška L, Nosek J, Kar A, Willcox S and Griffith JD (2019) A new view of the t-loop junction: implications for self-primed telomere extension, expansion of disease-related nucleotide repeat blocks, and telomere evolution. Frontiers in Genetics 10, 792.

van Luenen HG, Farris C, Jan S, Genest P, Tripathi A, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramsamy G, Vainio S, Heidebrecht T, Perrakis A, Page I, van Steensel B, Myler PJ and Borst P (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell 150, 909–921.

Votýpka J, Kostygov AY, Kraeva N, Gyrchubk-Ieremenko A, Tesařová M, Grychub D, Lukéj and Yurchenko V (2014) Kinetomonas gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. Protist 165, 825–838.

Walne AJ, Vulliamy T, Beswick R, Kirwan M and Dolk I (2008) TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. Blood 112, 3594–3600.

Yang X, Figueiredo LM, Espinal A, Okubo E and Li B (2009) RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei. Cell 137, 99–109.

Záhonová K, Hadariová L, Vacula R, Yurchenko V, Eliáš M, Krajčovič J and Vester M (2014) A small portion of plastid transcripts is polyadenylated in the flagellate Euglena gracilis. FEBS Letters 588, 783–788.

Záhonová K, Kostygov A, Ševčíková T, Yurchenko V and Eliáš M (2016) An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Current Biology 26, 2364–2369.

Zoltner M, Krienitz N, Field MC and Kramer S (2018) Comparative proteomics of the two T. brucei PABP2s suggests that PABP2 controls bulk mRNA. PLoS Neglected Tropical Diseases 12, e0006679.