Novel insertions in the mitochondrial maxicircle of *Trypanosoma musculi,* a mouse trypanosome

Ju-Feng Wang [1,#], Ruo-Hong Lin [1,#], Xuan Zhang [1], Geoff Hide [2], Zhao-Rong Lun [1,2*], De-Hua Lai [1,*]

[1] *Guangdong Provincial Key Laboratory of Aquatic Economic Animals, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, The People's Republic of China.*

[2] *Ecosystems and Environment Research Centre and Biomedical Research Centre, School of Science, Engineering and Environment, University of Salford, Salford, M5 4WT, UK.*

*** Author for correspondence:** Zhao-Rong Lun, E-mail: lsslzr@mail.sysu.edu.cn; De-Hua Lai, E-mail: laidehua@mail.sysu.edu.cn

\# Contributed equally.

Abstract

Trypanosoma musculi is a, globally distributed, mouse-specific haemoflagellate, of the family Trypanosomatidae, which shares similar characteristics in morphology with *Trypanosoma lewisi*. The kinetoplast (mitochondrial) DNA of Trypanosomatidae flagellates is comprised of catenated maxicircles and minicircles. However, genetic information on the *T. musculi* kinetoplast remains largely unknown. In this study, the *T. musculi* maxicircle genome was completely assembled, with PacBio and Illumina sequencing, and the size was confirmed at 34,606 bp. It consisted of two distinct parts: the coding region and the divergent regions (DRs, DRI and II). In comparison with other trypanosome maxicircles (*Trypanosoma brucei*, *Trypanosoma cruzi* and *T. lewisi*), the *T. musculi* maxicircle has a syntenic distribution of genes and shares 73.9%, 78.0% and 92.7% sequence identity, respectively, over the whole coding region. Moreover, novel insertions in *MURF2* (630 bp) and in *ND5* (1,278 bp) were found, respectively, which are homologous to minicircles. These findings support an evolutionary scenario similar to the one proposed for insertions in *Trypanosoma cruzi*, the pathogen of American trypanosomiasis. These novel insertions, together with a deletion (281 bp) in *ND4*, questions the role of Complex I in *T. musculi*. A detailed analysis of DRII indicated that it contains numerous repeat motifs and palindromes, the latter of which are highly conservative and contain A_5C elements.

The comprehensively annotated kinetoplast maxicircle of *T. musculi* reveals a high degree of similarity between this parasite and the maxicircle of *T. lewisi* and suggests that the DRII could be a valuable marker for distinguishing these evolutionarily related species.

Key words: *Trypanosoma musculi*, Maxicircle, Indels, Minicircle, Palindrome, Divergent regions, *Trypanosoma lewisi*

Key Findings

- The *Trypanosoma musculi* maxicircle is comprehensively annotated.
- The insertions in *T. musculi* genes may be due to a gene translocation.
- Divergent region (DR) II may be a good bio-barcode for trypanosomes.

Introduction

Trypanosomes are protozoan parasites that are distributed globally which infect humans, vertebrate animals and intermediate invertebrate hosts. Among them, members of the subgenus *Herpetosoma*, such as *Trypanosoma lewisi* and *Trypanosoma musculi*, are commonly found in rodents (Hoare, 1972; Kostygov *et al.*, 2021). These two trypanosomes cannot be easily distinguished due to their high degree of similarity in morphological characteristics and genetic markers such as the SSU rDNA sequences (Hong *et al.*, 2017). However, they do significantly differ in many aspects. In particular, *T. lewisi* infects only rats and sometimes humans (Sarataphan *et al.*, 2007; Verma *et al.*, 2011), while *T. musculi* infects only mice and is unlikely to be pathogenic to humans (Zhang *et al.*, 2018). Some research has indicated that both *T. musculi* and *T. lewisi* can modulate the host immune response in coinfections with various other infectious agents, to potentially cause more harm to the hosts, by altering the infection kinetics and increasing the duration of colonization in the host (Gao *et al.*, 2021; Lowry *et al.*, 2014; Nzoumbou-Boko *et al.*, 2017; Vaux *et al.*, 2016). To gain a better understanding of the biological characteristics, the kinetoplast DNA (kDNA) of *T. lewisi* has been comprehensively analyzed (Li *et al.*, 2020; Lin *et al.*, 2015). However, little is known of the details of the kDNA in *T. musculi*.

Trypanosomes are members of the kinetoplastea group of protozoa, named due to the presence of the kDNA. Trypanosome kDNA is a specific network structure of interlocking mitochondrial DNA circles, which consists of thousands of minicircles with dozens of maxicircles (Lukes *et al.*, 2002). Previous research in *Trypanosoma brucei* has shown that kDNA comprises at least 5% of the total cellular DNA, while most other eukaryotic mitochondrial DNA accounts for no more than 1% (Lukeš *et al.*, 2018). In general, kDNA maxicircles encode functional homologues of mitochondrial genes which are flanked by non-coding regions that diverge significantly in sequence and size amongst trypanosome species (Simpson *et al.*, 1987; Sloof *et al.*, 1992; Westenberger *et al.*, 2006). One of unusual features of the kinetoplast is that most of the maxicircle gene transcripts are not mature and do not encode functional open reading frames. These encrypted transcripts become translatable only after post-transcriptional processing, namely RNA editing, that inserts and deletes uridine residues (Stuart *et al.*, 1997; Stuart *et al.*, 2005). RNA editing was first discovered in cytochrome oxidase subunit 2 (*COII*) gene of *Trypanosoma brucei* and *Crithidia fasciculata*, whose mRNA transcripts have four uridine insertions (Benne *et al.*, 1986). The minicircles, recognized by a conserved motif of twelve nucleotides (GGGGTTGGTGTA) (Ray, 1989), encode guide RNA (gRNA) molecules that accurately position the editing machinery to ensure correct maxicircle transcripts are produced (Blum & Simpson, 1990).

Here, using PacBio and Illumina sequencing reads, the complete maxicircle sequence of *T. musculi* was assembled and annotated, including the repetitive non-coding variable region. Comparative analyses indicate that the gene organization and distribution in *T. musculi* maxicircles are highly conserved with *T. brucei*, *T. cruzi* and *T. lewisi*. The maxicircle kDNA gene organization of *T. musculi* and comparison with its species relatives were also presented. In addition, the genetic information on the divergent region (DR) II reveals that it may provide a good marker for molecular diagnosis and molecular epidemiological investigation of trypanosomes.

Materials and methods

Parasites, ultrastructure, kDNA extraction and restriction endonuclease digestion

Trypanosoma musculi Partinico II strain was gifted by Prof. Philippe Vincendeau of Université de Bordeaux, France, which was originally obtained from the London School of Hygiene and Tropical Medicine (Krampitz, 1969). *T. musculi* Particino 2, Lincicome and CDC strains were purchased from American Type Culture Collection (ATCC). Trypanosomes were harvested from the blood of infected mice and cultured in RPMI-1640 medium at 37°C supplemented with 10% fetal bovine serum (FBS) and a feeder layer of mouse macrophages as modified from Behr (Behr *et al.*, 1990). Protocols for the use of mice were approved by the Institutional Review Board for Animal Care at Sun Yat-Sen University under license of 31672276. For transmission electron microscopy, trypanosome specimens were prepared according to the method of Bozzola (Bozzola, 2014), and observed by using the JEM-100CX-II microscope system. For *T. musculi* DNA preparations, total DNA was purified using a phenol-chloroform method and kDNA was extracted by sucrose gradient ultracentrifugation according to previously published methods (Pérez-Morga & Englund, 1993). The isolated kDNA network was visualized on a 1% agarose gel and analyzed with restriction enzymes

HindIII, *EcoRI*, *BamHI*, *RsaI*, *HaeIII* and *TaqI* (New England Biolabs, USA). A computer-simulated restriction enzyme digestion map of *T. musculi* maxicircle was generated using the Dnaman 9.0 software (Lynnon Corporation, Quebec, Canada) based on the sequence assembled in this study.

Immunofluorescence assay

Trypanosome cells (1×10^7 cells/mL) were centrifuged for 5 minutes at 3,000 × g and washed twice in PBS. The cells were then transferred onto clean slides, which were left to air-dry in a fume hood, following fixation by methanol for 10 minutes. Dried slides were rehydrated and washed twice in PBS for 5 minutes at room temperature. The slides were then incubated with primary mAb-anti-L8C4 (1:800) followed by incubation with Cy3-Conjugated goat anti-mouse IgG (A10521, Thermo Fisher) (1:400) followed by a counterstaining consisting of 1× PBS with 3 μg/mL 4,6-diamidino-2-phenylindole (DAPI) (KOHL *et al.*, 1999). They were then photographed using a Leica fluorescence microscope.

Deep sequencing, sequence assembly and PCR verification

In order to generate a high quality maxicircle assembly, a kDNA Illumina library was constructed and sequenced using Illumina HiSeq2000 technology commercially (Novogene, China). Also, a PacBio Sequel library was constructed using total DNA and sequenced commercially (Annoroad, China). The Illumina reads were checked for quality and trimmed using fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trimmomatic (Bolger *et al.*, 2014).

Canu 2.0 software was used in *de novo* assembly of the *T. musculi* genome with parameter "genomeSize=30m minReadLength=600 minOverlapLength=300 corOutCoverage=100 corMinCoverage=2 correctedErrorRate=0.035" using the PacBio reads (Koren *et al.*, 2017). Then, the genome contigs, assembled from PacBio reads, were polished using the Illumina reads by pilon software to improve genome assembly (Walker *et al.*, 2014). Finally, the assembly sequences were aligned, with BLAST, to a previously obtained *T. lewisi* maxicircle assembly (KR072974) and redundant overlap deleted by MEGA 7.0 to yield the complete maxicircle sequence (Camacho *et al.*, 2009; Kumar *et al.*, 2016). To obtain *T. b. rhodesiense*, *Trypanosoma grayi* and *T. lewisi* complete maxicircle genomes, processed reads were assembled from WGS data (SRX3199071, SRX620256 and SRR11918574, respectively) freely available on NCBI using SPAdes 3.12.0 with parameter "--plasmid --careful -t 16 -m 200" (Antipov *et al.*, 2016). Then, alignment and trimming were also carried out as completed for *T. musculi*. The sequences of the *T. musculi* maxicircle coding region were also corrected by PCR verification using 12 pairs of primers. Meanwhile, *MURF2*, *ND4* and *ND5* genes were also amplified in three additional *T. musculi* strains (*T. musculi* Particino 2, Lincicome and CDC). Primers used are summarized in Table S1.

Gene annotation

Annotation of *T. musculi* maxicircle coding regions was performed by comparison with *T. brucei* (EATRO 427, M94286.1), *T. cruzi* (CL, DQ343645.1) and *T. lewisi* (CPO02, KR072974.1) manually using BLAST. Patterns of RNA editing of *T. musculi* maxicircle genes were predicted according to GC percent and RNA editing pattern of *T. lewisi* (Li *et al.*, 2020).

Data analysis

Dotplot graphs of the *T. musculi* maxicircle sequence plotted against itself and four other Trypanosomatidae species were generated by YASS software with default parameter (allow 10% indels, 25% mutations and e-value $< 1 \times 10^{-5}$ in alignment) (Noé & Kucherov, 2005). GC percentage, assembly coverage and homology search algorithms were drawn using Circos v0.69 (Krzywinski *et al.*, 2009). Regions (>300 bp) with sufficient sequence identity were plotted as colored ribbons, denoting the percentage of sequence identity. BioEdit software was used to create alignments and calculate nucleotide percentage identity matrices among different Trypanosomatidae species (Hall, 1999). Curation of the palindromes and inserted sequence homology analysis was performed using BLAST. MEME software was used to identify motifs and generate LOGO diagrams in the DRI region (Bailey *et al.*, 2015).

Phylogenetic analysis

The entire coding region of the kinetoplast maxicircles were aligned using ClustalO 1.2.4 (Sievers *et al.*, 2011) and the alignment was trimmed using Gblocks 0.91b with option "-t=d -b4=5 -b5=h" (Talavera & Castresana, 2007). Maximum likelihood trees were generated by RAxML 8.2.12 with 1,000 bootstrap replicates (Stamatakis, 2014). Neighbor joining and Minimum evolution trees were performed using MEGA 7.0 including 1000 bootstrap pseudo-replicates. Maxicircle genome sequences used in this work are summarized in Table S2.

Results

Morphology, ultrastructure, kDNA isolation and restriction enzyme digestion

In culture, *T. musculi* cells tend to attach to each other and form a rosette-like pattern (Fig. 1A), via their flagella, as determined using specific antibodies against the paraflagellar rod (Fig. 1C). At this stage, *T. musculi* is at the epimastigote stage in which the kinetoplast lies closely beside the nucleus. Ultrastructural analysis showed that the kinetoplast DNA disk was 660 ± 99 nm in length and 152 ± 26 nm in width ($n = 50$) (Fig. 1B), which is similar to closely related species such as *T. lewisi* (Lin *et al.*, 2015).

A total of 10^{10} *T. musculi* were harvested and high quality kDNA was obtained with a 260/280 absorbance ratio of 1.80. Kinetoplast DNA was found to be intact and free from contamination with nuclear or host DNAs as judged by agarose electrophoresis (Fig. S1A). Meanwhile, kDNA was incubated with endonucleases of *HindIII*, *EcoRI*, *BamHI*, *RsaI*, *HaeIII* and *TaqI* and a computer-simulated restriction enzyme digestion map was generated based on the maxicircle assembly which is described later (Fig. S1B-C). Some bands smaller than 4.0 kb did not correspond to the computed simulated patterns of *T. musculi* maxicircle, which implies the presence of a high number of possible heterogeneous minicircles in the kDNA of *T. musculi*. The bands consistently observed in Fig. S1B at ~1.3 kb suggest the presence of minicircles of a similar size as reported in *T.*

lewisi (Li *et al.*, 2020). Patterns that are free of kDNA in the wells were achieved using *HaeIII* and *TaqI* and indicated a high frequency of cleavage of kDNA minicircles. Most likely, the bands with molecular sizes of >4 kb correlated with the computer simulated patterns (Fig. S1C), are derived from kDNA maxicircles, except a band (~5 kb) in *RsaI* lane, a potential result of incomplete digestion. Moreover, the presence of four large molecular weight bands, in the *EcoRI* digestion, with sizes of >10 kb, ~7 kb ~6 kb and ~4 kb, indicated that the full-size kDNA maxicircle is larger than the sum of 27 kb.

Assembly and annotation of the kDNA maxicircle

Genomic DNA from the *T. musculi* Partinico II strain was sequenced on PacBio Sequel and Illumina platforms and contigs were assembled with the long-read assembler Canu 2.0 and corrected with the Illumina reads in Pilon. Then, a contig in length of 38,603 bp was identified in a BLAST search against the *T. lewisi* maxicircle (KR072974). This had two overlapping regions of 4,002 bp (covering positions from -3,341 or 31,266 to 661 bp) at each end (Fig. S2) confirming completion of the circle. Meanwhile, the maxicircle sequence has also been confirmed using five overlapping raw reads from the PacBio library (Fig. S2) and the maxicircle coding region sequences were also further refined using 12 pairs of primers and Sanger sequencing validation (Fig. S3). Finally, a 34,606 bp-long complete *T. musculi* maxicircle sequence was obtained, with an average coverage of 13.2X from Illumina reads and 268X from PacBio reads, including the coding regions (16,975 bp) and the DRs

(17,631 bp). The overall GC content of the maxicircle was 23.7%, with 27.5% in coding regions and 20.1% in DRs (Fig. 2).

Twenty genes were annotated in the *T. musculi* maxicircle by comparison with known Trypanosomatidae species (*T. brucei*, *T. cruzi* and *T. lewisi*), as listed in Table 1. All genes were found to be syntenic with the maxicircles of the comparator Trypanosomatidae species *T. brucei*, *T. cruzi* and *T. lewisi* (Fig. 2, S4). A sequence homology analysis (Fig. S5), showed that the *T. musculi* maxicircle has 92.7% identity to *T. lewisi* (blue ribbons). The ribbons change to yellow when compared with *T. cruzi* (78.0% identity) and *T. brucei* (73.9% identity), largely due to the low similarity in extensively-edited genes (Table 2). Moreover, three breaks shown as discontinuations of lines or ribbons (Fig. S4, S5) appear in *T. musculi* genes *MURF2*, *ND5* and *ND4*, with two sections (630 bp and 1,278 bp) inserted and one section deleted (281 bp), respectively (Fig. 3A and Table 1).

The confirmation of mutations in *MURF2*, *ND5* and *ND4* were performed with three other strains (*T. musculi* Particino 2, Lincicome and CDC). PCR results showed that the insertion in *MURF2* is specific to *T. musculi* Partinico II strain and not present in other three strains, while the insertion in *ND5* and the deletion in *ND4* exists in all tested strains (Fig. 3B). Furthermore, these insertions and the deletion have also been confirmed by inspecting alignments of the raw reads mapped back to the maxicircle assembly. Alignment analysis of insertions showed that a fragment (150 bp) at the 5’ end region of *MURF2* insertion sequence is homologous to both the 5’ end and middle regions of *ND5* insertion sequence (Fig. 3C). Moreover, those sequences in *MURF2* and *ND5* respectively share 95.3%, 96% and 95.3% identity with conserved regions of *T. lewisi* minicircles (MN447336.1, MN447339.1 and MN447386.1), and these 150 bp homologous regions cover three conserved sequence blocks (CSBs) of minicircles, indicating a minicircle origin of both insertions in *MURF2* and *ND5*. Together with the data shown in Fig. S1 and Fig. 3C, it seems that, unsurprisingly, *T. musculi* has a similar size and structure to minicircles reported for *T. lewisi* (Li *et al.*, 2020), i.e. ~1.3-kb category I minicircles that have two conserved regions with CSB1-3 motifs (and perhaps also ~1.5 kb category II minicircles with only one conserved region; such a band is also apparent in Fig. S1B). It therefore appears that the ~1.3 kb *ND5* insertion corresponds to a (degenerated) category I minicircle and the 630 bp *MURF2* insertion corresponds to half a category I minicircle.

RNA editing patterns of the maxicircle have been well studied in *T. brucei* and *T. lewisi* (Gerasimov *et al.*, 2018; Li *et al.*, 2020) and they are well correlated with GC%. The GC% pattern in *T. musculi* is fairly similar to *T. brucei* and *T. lewisi.* However, unexpectedly high GC contents were noticed in *MURF2* and *ND5* (Fig. S6), and they are precisely attributed to the insertions in both genes. In another region, *COII* and its cis-acting gRNA were identified (Table 1).

The whole coding region of the maxicircle is considered as a valuable marker for phylogenetic relationships of Trypanosomatidae species (Kaufer *et al.*, 2019). In order to further confirm the evolutionary relationship of *T. musculi* and other Trypanosomatidae species, sequences corresponding to whole coding region of *T. musculi* were aligned with the sequences from other Trypanosomatidae species to infer phylogenetic relationships. In the tree, *T. musculi* and *T. lewisi* are identified as belonging to the same subgenus *Herpetosoma*, clustered with the sister groups of *Schizotrypanum* and *Aneza* (Fig. 4).

Sequence analysis of the maxicircle divergent regions

A common theme of the maxicircle DR is the presence of various repeat arrays, which is also the case for *T. musculi*. The full map of DR of *T. musculi* was built by the YASS and Circos packages to identify homologous regions and to show global patterns of DR organization (Fig. 5A, 5B). Dotplot analyses of the DR showed two typical sections (I, II), flanking either the 12S rRNA or *ND5*. DRI is in a length of about 1.6 kb which is composed of short and highly repetitive units of about 107 bp, with two motifs being found (Fig. 5C). While DRII is in a length of about 14 kb, it consists of a series of tandem elements, namely α, β, γ, σ, and short version α', β', γ' (Fig. S7).

Palindromes are a typical structure already found in *T. cruzi*, *T. lewisi* and *T. rangeli*. Based on identifying homologues using BLAST, four AT-rich conserved palindromes were shown up in the DRII (Fig. 5B, 5D). Palindromes I and IV have the same perfect palindrome structure, 34 bp long, and are located at 19,898 bp and 28,055 bp. While palindromes II and III have one T-to-A substitution, they are located at 23,648 bp and 26,061 bp. A further BLAST analysis with the maxicircles of *T. b. brucei* (Lister 427, MN904526.1), *T. b. equiperdum* (STIB 818, EU185799.1), *Trypanosoma congolense* (IL3000, GCA_003013265.1) and *Trypanosoma vivax* (Y486, MT090068.1) enabled the identification of similar palindromes in these species (Fig. 5D), only one of each species is shown for illustrative purposes. These palindromes are highly conserved and contain an A_5C element.

Unlike highly conserved coding regions, DRs show species specificity among trypanosomes (Fig. S5). It displays about 70% sequence identity in DRI between the *T. musculi* and *T. lewisi* maxicircle, while there are only some similar sequences (~400 bp) in DRII. Moreover, there are no other homologous sequences between *T. musculi* and the other two species (*T. cruzi* and *T. brucei*) in the DRs (Fig. S5). Therefore, these results suggest that DRII is highly divergent among trypanosomes, which may have the potential to be a good molecular marker for distinguishing *T. musculi* from related species.

Discussion

In this study, a sequence of the 34,606 bp kDNA maxicircle genome from *T. musculi* was reported and an in-depth investigation of *T. musculi* maxicircle sequences and comparative analysis with other Trypanosomatidae species were also undertaken. The size of the total coding region of the *T. musculi* kDNA maxicircle is 16,975 bp with two pronounced insertions in *T. musculi MURF2* (630 bp), *ND5* (1,278 bp) and one deleted fragment of *ND4* (281 bp) (Fig. 3A). It is different from *T. brucei*, *T. cruzi* and *T. lewisi*, in which their sizes are around 15,000 bp length. The two insertions in *T. musculi* maxicircle genes correspond to either to a partial minicircle (630 bp) containing the one of the CSBs or to a complete minicircle (1,278 bp) containing two CSBs. Such insertions have not been observed in other Trypanosomatidae species except *Leishmania donovani* (1S LdBob strain) and *T. cruzi* (TcV strain) where the insertions were considered to be derived from minicircles due to the conserved CSBs. Therefore, the insertions were also thought to be a consequence of gene translocation, from minicircles to maxicircles (Berná *et al.*, 2021; Nebohácová *et al.*, 2009). Mostly gRNA genes are encoded in minicircles, but some gRNA genes, such as gMurf2 (30-79) and gNd7 (216-252)

(Koslowsky *et al.*, 2014; Li *et al.*, 2020), were reported to be encoding in maxicircles in *T. brucei* and *T. lewisi*, respectively. Moreover, 7 maxicircle-encoded gRNAs were identical in *L. tarentolae* LEM125 and UC strains, which mediate the editing of *Cyb*, *MURF2*, *A6* and *ND7* transcripts (Simpson *et al.*, 2015). It can be assumed that the insertions deriving from minicircles may also possibly encode gRNA genes for RNA editing, therefore these may be an intermediate stage indicating that maxicircle encoding gRNA genes have originated from minicircles.

Maxicircle gene deletions are only rarely found in Trypanosomatidae species, such as the similar deletions seen in *ND4* of the *T. cruzi* Esmeraldo strain (Westenberger *et al.*, 2006), and *ND7* gene from asymptomatic *T. cruzi* isolates (Baptista *et al.*, 2006). The effect of these insertions and deletions on the parasite life cycle is still unclear. *ND5* and *ND4* are known as non-edited genes in other known Trypanosomatidae species and it is inconceivable that these large insertions/deletions could be corrected by U-insertion/deletion editing of the mRNAs. Nevertheless, all of the above insertions/deletions are found in *ND4*, *ND5* as well as *ND7* genes, and these genes all encode subunits of the mitochondrial respiratory chain NADH-dehydrogenase (Complex I). Since the presence of a functional Complex I in Trypanosomatidae species has long been debated (Duarte & Tomás, 2014; Opperdoes & Michels, 2008). Deletions in kDNA encoding Complex I subunits were identified in some strains of *T. cruzi* that seems no impact in mitochondrial bioenergetics, ROS production or redox state in this parasite (César Carranza *et al.*, 2009). Although the presence of Complex I and its involvement in respiration has been clearly demonstrated in *T. brucei*, it appears to be non-essential for procyclic forms (Beattie & Howton, 1996; Surve *et al.*, 2012; Verner *et al.*, 2011). The lack of editing in several Complex I subunits in *L. tarentolae* UC strain also suggests that it may not be essential (Simpson *et al.*, 2015). Therefore, the possibility that the role of Complex I subunits is less important in *T. musculi* was favored with the presence of insertions/deletions in *ND5* and *ND4*. In addition, another insertion occurs in *MURF2*, whose function remains uncertain but hypothesis could be risen. *MURF2* might be a new component in Complex I. The insertion in *MURF2* may be a recent event as it only found in Partinico II strain, but not other three strains of *T. musculi*. The loss of conservation in *MURF2* could probably be attributed to the loss of function of Complex I components and consequently on selection pressures on the gene. To verify this hypothesis, a highly sensitive and accurate identification of the functioning of Complex I in Trypanosomatidae species would be interesting to investigate.

The DR of the kinetoplast maxicircle was initially described as a variable and non-coding region and the DR structure seemed to be drastically different in various species (Borst *et al.*, 1980; Borst *et al.*, 1982; Maslov *et al.*, 1984; Muhich *et al.*, 1983; Stuart & Gelvin, 1982). Therefore, the function of the DRs remains as an enigma. Studies on *T. brucei*, *Crithidia oncopelti*, *Leptomonas collosoma* and *Leishmania seymouri* revealed some CSBs-like sequences in their maxicircle DRs. As CSBs are essential for minicircle replication (Ryan *et al.*, 1988), CSBs-like sequences may play a similar role in maxicircle replication (Flegontov *et al.*, 2006; Gorbat *et al.*, 1990; Myler *et al.*, 1993; Sloof *et al.*, 1992). However, CSB-I or III-like regions were not identified in *T. musculi* DRs, instead, only a CSB-II-like region (CCCGTGT) is located at 19,817 bp. CSB-I or III-like regions were also not found in DRs of the closely related *T. lewisi*, suggesting a CSB-independent maxicircle replication mechanism exists in these species. Therefore, although CSB-like sequences were present in the insertions of *T. musculi MURF2* and *ND5*, it is not clear whether they are also involved in maxicircle replication.

It has been demonstrated that hairpins or cruciform structures (palindromes) are frequently associated with promoters and may also act as protein-binding sites (Wadkins, 2000). Palindromes with an A_5C-element in DRs are suggested as recognition sites for binding of transcription factors or transcription initiation (Flegontov *et al.*, 2006; Vasil'eva *et al.*, 2004). Some palindromes also have been identified in *T. musculi* as well as in a variety of other Trypanosomatidae species, which each consists of one A_5C-element. It may be speculated that these palindromes play a significant role in Trypanosomatidae species maxicircles, judged by their high degree of sequence conservation in the evolution of Kinetoplastida species.

The trypanosome maxicircle presents itself as a complex evolutionary system, and it may be an excellent taxonomic marker in phylogenetic analysis. The coding region of the maxicircle in phylogenetic analyses provides a robust evolutionary insight into the relationships within Trypanosomatidae species (Kaufer *et al.*, 2019; Kay *et al.*, 2020; Lin *et al.*, 2015). A close affinity between *T. musculi* and *T. lewisi* in *Herpetosoma* was also supported, which clustered with the sister groups of *Schizotrypanum* and *Aneza*. Unlike the highly conserved coding region, the DRs of maxicircle, especially DRII sequence, was found to be significantly divergent and species-specific (Kay *et al.*, 2020). The homologies in DRII between closely related species, e.g., *T. musculi* and *T. lewisi*, phylogenetic clades of *T. cruzi*, are limited (Fig. S5, S8). Such a characteristic of DRII provides an opportunity for developing a valuable molecular marker for distinguishing closely related species and subspecies. Actually, a preliminary test on three *T. musculi* strains and six *T. lewisi* strains revealed a consistent amplification of DRII fragments, which could enable them to be distinguished from each other and 13 strains of other trypanosomes (Hong *et al.*, 2017).

In general, this study reports the first detailed description and analysis of the kDNA maxicircle genome of *T. musculi* and reveals a relatively high overall conservation of gene content and synteny with other trypanosome species. Furthermore, the divergence of DRII suggests its potential as a valuable marker for distinguishing these evolutionarily related species.

Supplementary materials

Data

Nucleotide sequence data reported in this paper are available in GenBank databases under accession numbers: *Trypanosoma musculi* maxicircle sequence (OM000218), *Trypanosoma lewisi* maxicircle sequence (OM000219), *Trypanosoma grayi* maxicircle sequence (OM049542), and *Trypanosoma brucei rhodesiense* maxicircle sequence (OM049543). PacBio and Illumina sequencing data have been deposited in NCBI's Sequence Read Archive (SRA) with BioProject ID PRJNA792722.

Acknowledgements

The authors would like to thank Dr. Ling-Ling Zheng for help with bioinformatics analysis. Thanks to all the members in the laboratories who provided useful help during the field and laboratory work. We would also like to appreciate the critical comments from the anonymous reviewers which have greatly helped to improve our paper.

Authors contributions

JF Wang, RH Lin, DH Lai and ZR Lun designed the study. JF Wang, RH Lin and X Zhang conducted data gathering and performed statistical analyses. JF Wang, RH Lin and DH Lai drafted the manuscript and undertook data extraction and screening. G Hide, ZR Lun and DH Lai critically reviewed the paper. All authors approved the final version and agree to be accountable for all aspects of the work.

Financial support

The project was supported by grants from the National Natural Science Foundation of China (31672276, 31720103918) to ZRL.

Conflict of interest

The authors declare there are no conflicts of interest.

Ethical standards

The animal ethical approval of sample collection was obtained from the Institutional Review Board of Animal Care at Sun Yat-Sen University (License no. 31672276).

References

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A. and Pevzner, P. A. (2016). plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics,* **32**, 3380-3387. doi: 10.1093/bioinformatics/btw493.

Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research,* **43**, W39-W49. doi: 10.1093/nar/gkv416.

Baptista, C. S., Vêncio, R. Z. N., Abdala, S., Carranza, J. C., Westenberger, S. J., Silva, M. N., Pereira, C. A. d. B., Galvão, L. M. C., Gontijo, E. D., Chiari, E., Sturm, N. R. and Zingales, B. (2006). Differential transcription profiles in *Trypanosoma cruzi* associated with clinical forms of Chagas disease: Maxicircle NADH dehydrogenase subunit 7 gene truncation in asymptomatic patient isolates. *Molecular and Biochemical Parasitology,* **150**, 236-248. doi: 10.1016/j.molbiopara.2006.08.008.

Beattie, D. S. and Howton, M. M. (1996). The presence of rotenone-sensitive NADH dehydrogenase in the long slender bloodstream and the procyclic forms of *Trypanosoma brucei brucei*. *European Journal of Biochemistry,* **241**, 888-894. doi: 10.1111/j.1432-1033.1996.00888.x.

Behr, M. A., Mathews, S. A. and D'Alesandro, P. A. (1990). A medium for the continuous cultivation of bloodstream forms of *Trypanosoma lewisi* at 37 C. *Journal of Parasitology,* **76**, 711-716. doi: 10.2307/3282987.

Benne, R., Van den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H. and Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell,* **46**, 819-826. doi: 10.1016/0092-8674(86)90063-2.

Berná, L., Greif, G., Pita, S., Faral-Tello, P., Díaz-Viraqué, F., Souza, R. D. C. M. D., Vallejo, G. A., Alvarez-Valin, F. and Robello, C. (2021). Maxicircle architecture and evolutionary insights into *Trypanosoma cruzi* complex. *PLoS Neglected Tropical Diseases,* **15**, e0009719. doi: 10.1371/journal.pntd.0009719.

Blum, B. and Simpson, L. (1990). Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. *Cell,* **62**, 391-397. doi: 10.1016/0092-8674(90)90375-o.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30**, 2114-2120. doi: 10.1093/bioinformatics/btu170.

Borst, P., Fase-Fowler, F., Hoeijmakers, J. H. and Frasch, A. C. (1980). Variations in maxi-circle and mini-circle sequences in kinetoplast DNAs from different *Trypanosoma brucei* strains. *Biochimica et Biophysica Acta (BBA),* **610**, 197-210. doi: 0.1016/0005-2787(80)90001-5.

Borst, P., Weijers, P. J. and Brakenhoff, G. J. (1982). Analysis by electron microscopy of the variable segment in the maxi-circle of kinetoplast DNA from *Trypanosoma brucei*. *Biochimica et Biophysica Acta (BBA),* **699**, 272-280. doi: 10.1016/0167-4781(82)90117-8.

Bozzola, J. J. (2014). Conventional specimen preparation techniques for transmission electron microscopy of cultured cells. In *Electron Microscopy: Methods and Protocols* (ed. Kuo, J.), pp. 1-19. Humana Press, Totowa, NJ.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* **10**, 421. doi: 10.1186/1471-2105-10-421.

César Carranza, J., Kowaltowski, A. J., Mendonça, M. A. G., de Oliveira, T. C., Gadelha, F. R. and Zingales, B. (2009). Mitochondrial bioenergetics and redox state are unaltered in *Trypanosoma cruzi* isolates with compromised mitochondrial complex I subunit genes. *Journal of Bioenergetics and Biomembranes,* **41**, 299-308. doi: 10.1007/s10863-009-9228-4.

Duarte, M. and Tomás, A. M. (2014). The mitochondrial complex I of trypanosomatids - an overview of current knowledge. *Journal of Bioenergetics and Biomembranes,* **46**, 299-311. doi: 10.1007/s10863-014-9556-x.

Flegontov, P. N., Guo, Q., Ren, L., Strelkova, M. V. and Kolesnikov, A. A. (2006). Conserved repeats in the kinetoplast maxicircle divergent region of *Leishmania* sp. and *Leptomonas seymouri*. *Molecular Genetics and Genomics,* **276**, 322-333. doi: 10.1007/s00438-006-0145-5.

Gao, J. M., Yi, S. Q., Geng, G. Q., Xu, Z. S., Hide, G., Lun, Z. R. and Lai, D. H. (2021). Infection with *Trypanosoma lewisi* or *Trypanosoma musculi* may promote the spread of *Toxoplasma gondii*. *Parasitology,* **148**, 703-711. doi: 10.1017/s0031182021000196.

Gerasimov, E. S., Gasparyan, A. A., Kaurov, I., Tichý, B., Logacheva, M. D., Kolesnikov, A. A., Lukeš, J., Yurchenko, V., Zimmer, S. L. and Flegontov, P. (2018). Trypanosomatid mitochondrial RNA editing: dramatically complex transcript repertoires revealed with a dedicated mapping tool. *Nucleic Acids Research,* **46**, 765-781. doi: 10.1093/nar/gkx1202.

Gorbat, A., Maslov, D. A., Peters, L. S., Gaviernik, P., Viustenkhagen, T. and Kolesnikov, A. A. (1990). Analysis of the sequence of repeats in divergent regions of maxi-circular DNA from kinetoplasts of *Crithidia oncopelti*. *Molekuliarnaia Biologiia,* **24**, 1539-1548.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* **41**, 95-98. doi: 10.1021/bk-1999-0734.ch008.

Hoare, C. A. (1972). *The trypanosomes of mammals. A zoological monograph,* Blackwell Scientific Publications, Oxford, U.K.

Hong, X. K., Zhang, X., Fusco, O. A., Lan, Y. G., Lun, Z. R. and Lai, D. H. (2017). PCR-based identification of *Trypanosoma lewisi* and *Trypanosoma musculi* using maxicircle kinetoplast DNA. *Acta Tropica,* **171**, 207-212. doi: 10.1016/j.actatropica.2017.04.007.

Kaufer, A., Stark, D. and Ellis, J. (2019). Evolutionary insight into the Trypanosomatidae using alignment-free phylogenomics of the kinetoplast. *Pathogens,* **8**. doi: 10.3390/pathogens8030157.

Kay, C., Williams, T. A. and Gibson, W. (2020). Mitochondrial DNAs provide insight into trypanosome phylogeny and molecular evolution. *BMC Evolutionary Biology,* **20**, 161. doi: 10.1186/s12862-020-01701-9.

KOHL, L., SHERWIN, T. and GULL, K. (1999). Assembly of the paraflagellar rod and the flagellum attachment zone complex during the *Trypanosoma brucei* cell cycle. *Journal of Eukaryotic Microbiology,* **46**, 105-109. doi: 10.1111/j.1550-7408.1999.tb04592.x.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. *Genome Research,* **27**, 722-736. doi: 10.1101/gr.215087.116.

Koslowsky, D., Sun, Y., Hindenach, J., Theisen, T. and Lucas, J. (2014). The insect-phase gRNA transcriptome in *Trypanosoma brucei*. *Nucleic Acids Research,* **42**, 1873-1886. doi: 10.1093/nar/gkt973.

Kostygov, A. Y., Karnkowska, A., Votýpka, J., Tashyreva, D., Maciszewski, K., Yurchenko, V. and Lukeš, J. (2021). Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biology,* **11**, 200407. doi: doi:10.1098/rsob.200407.

Krampitz, H. E. (1969). Geographical distribution, host-parasite relationship and multiplication of Sicilian strains of *Trypanosoma (Herpetosoma) duttoni* Thiroux 1950 (Protozoa, Trypanosomatidae). *Zeitschrift fur Parasitenkunde (Berlin, Germany),* **32**, 297-315. doi: 10.1016/0166-6851(84)90091-4.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research,* **19**, 1639-1645. doi: 10.1101/gr.092759.109.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution,* **33**, 1870-1874. doi: 10.1093/molbev/msw054.

Li, S. J., Zhang, X., Lukeš, J., Li, B. Q., Wang, J. F., Qu, L. H., Hide, G., Lai, D. H. and Lun, Z. R. (2020). Novel organization of mitochondrial minicircles and guide RNAs in the zoonotic pathogen *Trypanosoma lewisi*. *Nucleic Acids Research,* **48**, 9747-9761. doi: 10.1093/nar/gkaa700.

Lin, R. H., Lai, D. H., Zheng, L. L., Wu, J., Lukeš, J., Hide, G. and Lun, Z. R. (2015). Analysis of the mitochondrial maxicircle of *Trypanosoma lewisi*, a neglected human pathogen. *Parasites & Vectors,* **8**, 665. doi: 10.1186/s13071-015-1281-8.

Lowry, J. E., Leonhardt, J. A., Yao, C., Belden, E. L. and Andrews, G. P. (2014). Infection of C57BL/6 mice by *Trypanosoma musculi* modulates host immune responses during Brucella abortus cocolonization. *Journal of Wildlife Diseases,* **50**, 11-20. doi: 10.7589/2013-03-080.

Lukes, J., Guilbride, D. L., Votýpka, J., Zíková, A., Benne, R. and Englund, P. T. (2002). Kinetoplast DNA network: evolution of an improbable structure. *Eukaryotic Cell,* **1**, 495-502. doi: 10.1128/EC.1.4.495-502.2002.

Lukeš, J., Wheeler, R., Jirsová, D., David, V. and Archibald, J. M. (2018). Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. *IUBMB Life,* **70**, 1267-1274. doi: 10.1002/iub.1894.

Maslov, D. A., Kolesnikov, A. A. and Zaitseva, G. N. (1984). Conservative and divergent base sequence regions in the maxicircle kinetoplast DNA of several trypanosomatid flagellates. *Molecular and Biochemical Parasitology,* **12**, 351-364. doi: 10.1016/0166-6851(84)90091-4.

Muhich, M. L., Simpson, L. and Simpson, A. M. (1983). Comparison of maxicircle DNAs of *Leishmania tarentolae* and *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America,* **80**, 4060-4064. doi: 10.1073/pnas.80.13.4060.

Myler, P. J., Glick, D., Feagin, J. E., Morales, T. H. and Stuart, K. D. (1993). Structural organization of the maxicircle variable region of *Trypanosoma brucei*: identification of potential replication origins and topoisomerase II binding sites. *Nucleic Acids Research,* **21**, 687-694. doi: 10.1093/nar/21.3.687.

Nebohácová, M., Kim, C. E., Simpson, L. and Maslov, D. A. (2009). RNA editing and mitochondrial activity in promastigotes and amastigotes of *Leishmania donovani*. *International Journal for Parasitology,* **39**, 635-644. doi: 10.1016/j.ijpara.2008.10.015.

Noé, L. and Kucherov, G. (2005). YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Research,* **33**, W540-W543. doi: 10.1093/nar/gki478.

Nzoumbou-Boko, R., De Muylder, G., Semballa, S., Lecordier, L., Dauchy, F. A., Gobert, A. P., Holzmuller, P., Lemesre, J. L., Bras-Gonçalves, R., Barnabé, C., Courtois, P., Daulouède,

S., Beschin, A., Pays, E. and Vincendeau, P. (2017). *Trypanosoma musculi* infection in mice critically relies on mannose receptor-mediated arginase induction by a TbKHC1 kinesin H chain homolog. *Journal of Immunology,* **199**, 1762-1771. doi: 10.4049/jimmunol.1700179.

Opperdoes, F. R. and Michels, P. A. M. (2008). Complex I of Trypanosomatidae: does it exist? *Trends in Parasitology,* **24**, 310-317. doi: 10.1016/j.pt.2008.03.013.

Pérez-Morga, D. and Englund, P. T. (1993). The structure of replicating kinetoplast DNA networks. *Journal of Cell Biology,* **123**, 1069-1079. doi: 10.1083/jcb.123.5.1069.

Ray, D. S. (1989). Conserved sequence blocks in kinetoplast minicircles from diverse species of trypanosomes. *Molecular and Cellular Biology,* **9**, 1365-1367. doi: 10.1128/mcb.9.3.1365-1367.1989.

Ryan, K. A., Shapiro, T. A., Rauch, C. A. and Englund, P. T. (1988). Replication of kinetoplast DNA in trypanosomes. *Annual Review of Microbiology,* **42**, 339-358. doi: 10.1146/annurev.mi.42.100188.002011.

Sarataphan, N., Vongpakorn, M., Nuansrichay, B., Autarkool, N., Keowkarnkah, T., Rodtian, P., Stich, R. W. and Jittapalapong, S. (2007). Diagnosis of a *Trypanosoma lewisi*-like (*Herpetosoma*) infection in a sick infant from Thailand. *Journal of Medical Microbiology,* **56**, 1118-1121. doi: 10.1099/jmm.0.47222-0.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* **7**, 539. doi: 10.1038/msb.2011.75.

Simpson, L., Douglass, S. M., Lake, J. A., Pellegrini, M. and Li, F. (2015). Comparison of the mitochondrial genomes and steady state transcriptomes of two strains of the trypanosomatid parasite, *Leishmania tarentolae*. *PLoS Neglected Tropical Diseases,* **9**, e0003841. doi: 10.1371/journal.pntd.0003841.

Simpson, L., Neckelmann, N., de la Cruz, V. F., Simpson, A. M., Feagin, J. E., Jasmer, D. P. and Stuart, K. (1987). Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. *Journal of Biological Chemistry,* **262**, 6182-6196. doi: 10.1016/S0021-9258(18)45555-X.

Sloof, P., de Haan, A., Eier, W., van Iersel, M., Boel, E., van Steeg, H. and Benne, R. (1992). The nucleotide sequence of the variable region in *Trypanosoma brucei* completes the sequence analysis of the maxicircle component of mitochondrial kinetoplast DNA. *Molecular and Biochemical Parasitology,* **56**, 289-299. doi: 10.1016/0166-6851(92)90178-m.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* **30**, 1312-1313. doi: 10.1093/bioinformatics/btu033.

Stuart, K., Allen, T. E., Heidmann, S. and Seiwert, S. D. (1997). RNA editing in kinetoplastid protozoa. *Microbiology and Molecular Biology Reviews,* **61**, 105-120. doi: 10.1128/mmbr.61.1.105-120.1997.

Stuart, K. D. and Gelvin, S. B. (1982). Localization of kinetoplast DNA maxicircle transcripts in bloodstream and procyclic form *Trypanosoma brucei*. *Molecular and Cellular Biology,* **2**, 845-852. doi: 10.1128/mcb.2.7.845-852.1982.

Stuart, K. D., Schnaufer, A., Ernst, N. L. and Panigrahi, A. K. (2005). Complex management: RNA editing in trypanosomes. *Trends in Biochemical Sciences,* **30**. doi: 10.1016/j.tibs.2004.12.006.

Surve, S., Heestand, M., Panicucci, B., Schnaufer, A. and Parsons, M. (2012). Enigmatic presence of mitochondrial complex I in *Trypanosoma brucei* bloodstream forms. *Eukaryotic Cell,* **11**, 183-193. doi: doi:10.1128/EC.05282-11.

Talavera, G. and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology,* **56**, 564-577. doi: 10.1080/10635150701472164.

Vasil'eva, M. A., Bessolitsina, E. A., Merzlyak, E. M. and Kolesnikov, A. A. (2004). Identification of the 12S rRNA gene promoter in *Leptomonas seymouri* mitochondrial DNA. *Molecular Biology,* **38**, 839-843. doi: 10.1023/B:MBIL.0000049859.42603.47.

Vaux, R., Schnoeller, C., Berkachy, R., Roberts, L. B., Hagen, J., Gounaris, K. and Selkirk, M. E. (2016). Modulation of the immune response by nematode secreted acetylcholinesterase revealed by heterologous expression in *Trypanosoma musculi*. *PLoS Pathogens,* **12**, e1005998. doi: 10.1371/journal.ppat.1005998.

Verma, A., Manchanda, S., Kumar, N., Sharma, A., Goel, M., Banerjee, P. S., Garg, R., Singh, B. P., Balharbi, F., Lejon, V., Deborggraeve, S., Singh Rana, U. V. and Puliyel, J. (2011). *Trypanosoma lewisi* or *T. lewisi*-like infection in a 37-day-old Indian infant. *American Journal of Tropical Medicine and Hygiene,* **85**, 221-224. doi: 10.4269/ajtmh.2011.11-0002.

Verner, Z., Čermáková, P., Škodová, I., Kriegová, E., Horváth, A. and Lukeš, J. (2011). Complex I (NADH:ubiquinone oxidoreductase) is active in but non-essential for procyclic *Trypanosoma brucei*. *Molecular and Biochemical Parasitology,* **175**, 196-200. doi: 10.1016/j.molbiopara.2010.11.003.

Wadkins, R. M. (2000). Targeting DNA secondary structures. *Current Medicinal Chemistry,* **7**. doi: 10.2174/0929867003375461.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K. and Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One,* **9**, e112963. doi: 10.1371/journal.pone.0112963.

Westenberger, S. J., Cerqueira, G. C., El-Sayed, N. M., Zingales, B., Campbell, D. A. and Sturm, N. R. (2006). *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics,* **7**, 60. doi: 10.1186/1471-2164-7-60.

Zhang, X., Hong, X. K., Li, S. J., Lai, D. H., Hide, G., Lun, Z. R. and Wen, Y. Z. (2018). The effect of normal human serum on the mouse trypanosome *Trypanosoma musculi in vitro* and *in vivo*. *Experimental Parasitology,* **184**, 115-120. doi: 10.1016/j.exppara.2017.12.005.

Tables

Table 1. Gene positions and lengths in the maxicircle of *T. musculi*.

Gene	**RNA editing**	***T. musculi*** **position**	***T. musculi*** **length**	***T. lewisi*** **length**	***T. cruzi*** **length**	***T. brucei*** **length**
12S rRNA	-	1-1164	1164	1168	1161	1149
9S rRNA	-	1210-1818	609	608	608	611
ND8	extensive	1859-2130	272	285	279	266
ND9#	extensive	2195-2536	342	350	338	321
uS3m#	none	2575-2823	249	241	264	234
ND7	extensive	2874-3647	774	771	755	702
COIII	extensive	3721-4135	415	414	424	439
Cyb	minor	4222-5301	1080	1080	1080	1080
ATPase6	extensive	5342-5644	303	304	336	369
ND2	none	5687-7038	1352	1341	1341	1237
CR3*	extensive	6979-7111	~133	~123	~119	/
ND1 **#**	none	7129-8112	984	942	942	957
COII	minor	8111-8739	629	629	629	626
COII **gRNA**	-	8746-8758	-	-	-	-
MURF2	minor	8767-10439	1673	1053	1056	1041
(Insertion)	-	8897-9526	630	-	-	-
COI **#**	none	10429-12081	1653	1650	1650	1734
CR4#	extensive	12102-12319	218	212	207	185
ND4 Δ	none	12428-13464	1037	1314	1314	1311
ND3 **#**	extensive	13453-13636	184	187	193	256
RPS12	extensive	13711-13908	198	190	191	172
ND5	none	13929-16975	3047	1773	1770	1770
(Insertion)	-	14923-16200	1278	-	-	-

Gene positions are shown relative to the start of the *12S rRNA*

These genes are encoded on the reverse strand

**CR3* two end positions from *T. musculi*, *T. lewisi*, *T. cruzi* and *T. brucei* are uncertain

Δ A fragment deletion is found in the *T. musculi ND4* gene

Table 2. Comparison of the average percentage identities of *T. musculi* kDNA maxicircle with those from the other three Trypanosomatidae species.

Comparison of *T. musculi*	Whole coding region	5’-edited genes	Extensively-edited genes	rRNAs	Non-edited genes
vs. *T. lewisi*	92.7%	95.8%	92.4%	95.4%	92.2%
vs. *T. cruzi*	78.0%	84.5%	73.7%	84.0%	78.4%
vs. *T. brucei*	73.9%	84.4%	60.2%	80.5%	74.8%

Entire coding region: start from 5’end of *12S rRNA* to 3’ end of *ND5*

5’-edited genes: *Cyb*, *COII*

Extensively-edited genes: *ND8*, *ND9*, *ND7*, *COIII*, *ATPase6*, *CR3*, *CR4*, *ND3*, *RPS12*

Non-edited genes: *uS3m*, *ND2*, *ND1*, *COI*

MURF2, *ND4* and *ND5* genes are not calculated in *T. musculi* (5’-edited genes or non-edited genes) due to insertions/deletion

Legends to Figures

Fig. 1. Morphology and ultrastructure of *T. musculi*. (A) Giemsa staining of the epimastigote form of *T. musculi* from *in vitro* culture (scale bar 5 μm). Flagellum (F), nucleus (N) and kinetoplast (K) are indicated. (B) Electron micrograph of the *T. musculi* trypomastigote form (scale bar 200 nm). Basal body (BB), flagellum (F), flagellar pocket (FP) and kinetoplast (K) are indicated. (C) Immunofluorescence analysis (IFA) shows epimastigote-like forms of *T. musculi* from *in vitro* cultivation with antibody L8C4 detecting paraflagellar rods (green) and DAPI detecting kinetoplast DNA and nuclear DNA (blue) (scale bars 5 μm).

Fig. 2. Circos plot of the *T. musculi* maxicircle. The outer track indicates the gene arrangements and gene distribution; the middle track represents the GC content (orange for above-average and green for below-average) and the inner track is a histogram of assembly coverage.

Fig. 3. Analysis of insertions or deletions in the *T. musculi* maxicircle genes *MURF2*, *ND5* and *ND4*. (A) A diagram of insertions or deletions in the *T. musculi* maxicircle genes *MURF2* (630 bp insertion), *ND5* (1278 bp insertion) and *ND4* (281 bp deletion). Insertion regions are labeled as grey boxes. (B) PCR amplification of the *T. musculi MURF2*, *ND5* and *ND4* genes from *T. musculi* Partinico II (1), *T. musculi* Particino 2 (2), *T. musculi* Lincicome (3), *T. musculi* CDC (4), and fragments are analyzed on a 1.0 % agarose gel. M, DL2000 marker (Takara, China). N, negative control. The positions of primers are showed in (A) and Table S1. (C) Alignment of

the conserved regions from the *T. lewisi* minicircle (MN447335.1) and insertion sequences of *MURF2* and *ND5*. Conserved sequence blocks (CSB-I, II, III) are highlighted in grey.

Fig. 4. Phylogenetic analysis of the Trypanosomatidae species using the maxicircle coding regions. Phylogenetic tree is performed based on Maximum likelihood /Neighbor joining/Minimum evolution methods with 1,000 bootstrap replicates with the respective bootstrap confidences indicated at branch points. Branch lengths are indicated by the black line and the scale bar represents the number of nucleotide substitutions per site.

Fig. 5. Sequence analyses of the divergent region of the *T. musculi* maxicircle. (A) Dot plot analysis of the *T. musculi* maxicircle, the main diagonal line represents the sequence's alignment with itself and the lines about the main diagonal represent repetitive patterns within the maxicircle sequence. (B) Circos plot of the divergent region of *T. musculi* maxicircle, the outer track indicates gene arrangement and gene distribution. Ribbons inside the circle connect homologous regions, color represents percent of sequence identity in range [70%; 100%] in the order yellow, green, and blue. Four palindromes are showed with red arrows. (C) The repetitive sequences from DR I are aligned and two motifs predicted by MEME are indicated with black line. LOGO diagrams show nucleotides at a given position of each motif and relative frequency represented by height. (D) The palindromes from the divergent region of the Trypanosomatidae species are shown with the palindromic bases highlighted.

Legends to Supplementary Figures

Fig. S1. Restriction endonuclease analysis of *T. musculi* kDNA. (A) Kinetoplast DNA from *T. musculi* is resolved on a 1.0% agarose gel. (B) Restriction endonuclease digestion of *T. musculi* kDNA as analyzed on a 1.0% agarose gel. (C) Computer simulated virtual restriction patterns derived from the complete *T. musculi* maxicircle with the same set of restriction enzymes from (B). Asterisks indicate >4.0 kb-long bands identified both in (B) and (C), and a triangle indicates the band (~5 kb) only present in (B). The marker is DL10000 (Takara, China).

Fig. S2. The assembly contigs in *T. musculi.* The maxicircle is drawn in a linear form, starting from the *12S rRNA* gene. Regions covered by raw reads and the assembly contigs are shown below in green and overlaps at each end are colored in blue.

Fig. S3. PCR amplification of *T. musculi* maxicircle DNA using twelve pairs of primers, fragments are analyzed on a 1.0 % agarose gel. The marker is DL10000 (Takara, China). The information on the primers has been summarized in Table S1.

Fig. S4. Dot plot comparative analysis of the coding region of *T. musculi* compared with the coding region of *T. lewisi* (A), *T. cruzi* (B) and *T. brucei* (C), respectively. Diagonal lines indicate that the DNA sequences of two compared species are homologous in the corresponding regions (allow 10% indels, 25% mutations and e-value $< 1 \times 10^{-5}$ in alignment). The break regions in seen in the *T. musculi* vs. other Trypanosomatidae species are due to the insertions and deletions in *T. musculi* which are, respectively, marked by blue and orange ellipses.

Fig. S5. Circos plots that compare the maxicircles of two species, (A) *T. musculi* (left sequence) and *T. lewisi* (right sequence); (B) *T. musculi* (left sequence) and *T. cruzi* (right sequence); (C) *T. musculi* (left sequence) and *T. brucei* (right sequence). Ribbons inside the circle connect homologous regions (>300 bp), and the colors represent the percent of sequence identity in the range [70%; 100%] in the order yellow, green, and blue. Sequence identity was calculated using BLAST. Recently developed primers (TM1-2 and TL1-3) for distinguishing *T. musculi* and *T. lewisi* were indicated (Hong *et al.*, 2017).

Fig. S6. GC percentage graphs of the maxicircle coding regions of *T. brucei* (A) and *T. musculi* (B). The regions where percentage GC content value lie above the dashed lines may likely have RNA editing. Window size is 100 bp.

Fig. S7. Diagram of the repeated elements (α, α’, β, β’, γ, γ’, σ) identified by dot plot analysis of DRII.

Fig. S8. Circos plots that compare the maxicircles from three *T. cruzi* strains. (A) *T. cruzi* TCC (left sequence) and *T. cruzi* Dm28c (right sequence); (B) *T. cruzi* TCC (left sequence) and *T. cruzi* Y (right sequence); (C) *T. cruzi* Dm28c (left sequence) and *T. cruzi* Y (right sequence).

Ribbons inside the circle connect homologous regions (>300 bp), and the colors represent the percent of sequence identity in the range [70%; 100%] in the order yellow, green, and blue. Sequence identity was calculated using BLAST.

Figures

Figure 1

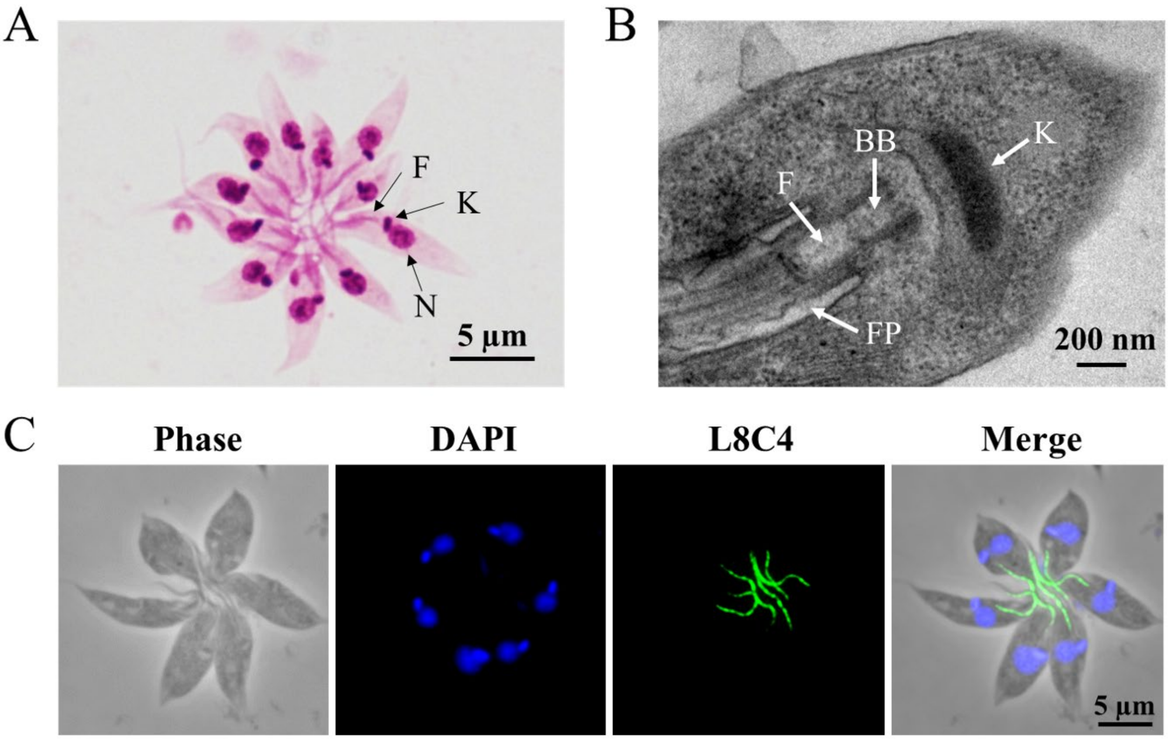


Figure 2

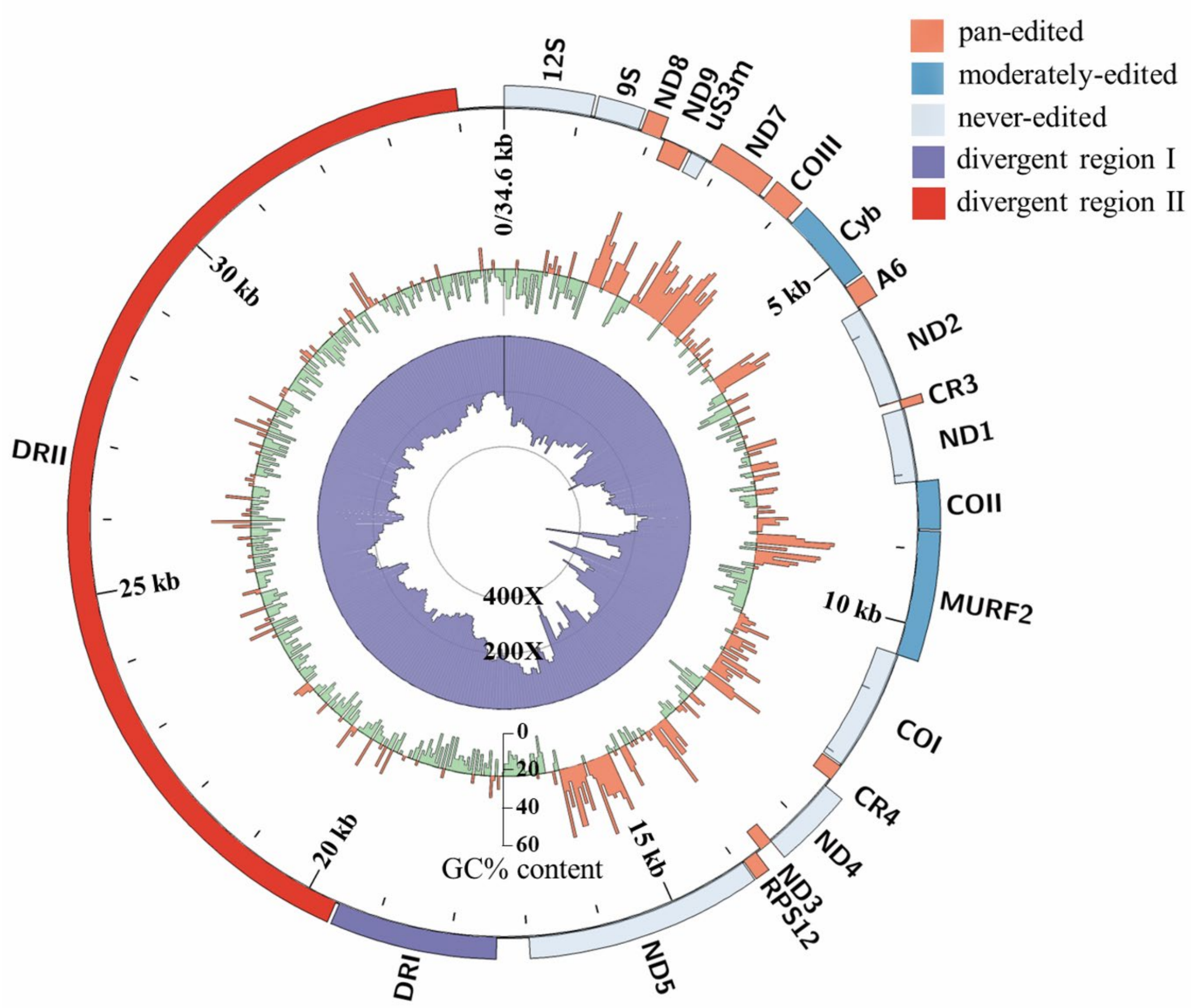

pan-edited
moderately-edited
never-edited
divergent region I
divergent region II
12S
9S
ND8
ND9
uS3m
ND7
COIII
Cyb
A6
ND2
CR3
ND1
COII
MURF2
COI
CR4
ND4
ND3
RPS12
ND5
DRI
DRII
0/34.6 kb
5 kb
10 kb
15 kb
20 kb
25 kb
30 kb
400X
200X
0
20
40
60
GC% content

Figure 3

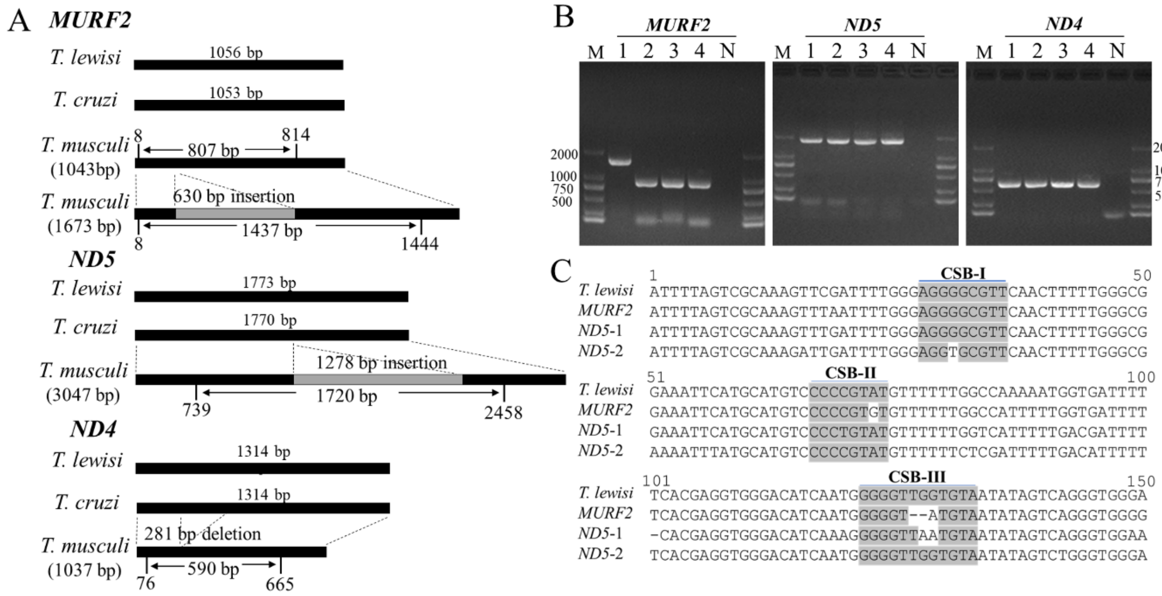


Figure 4

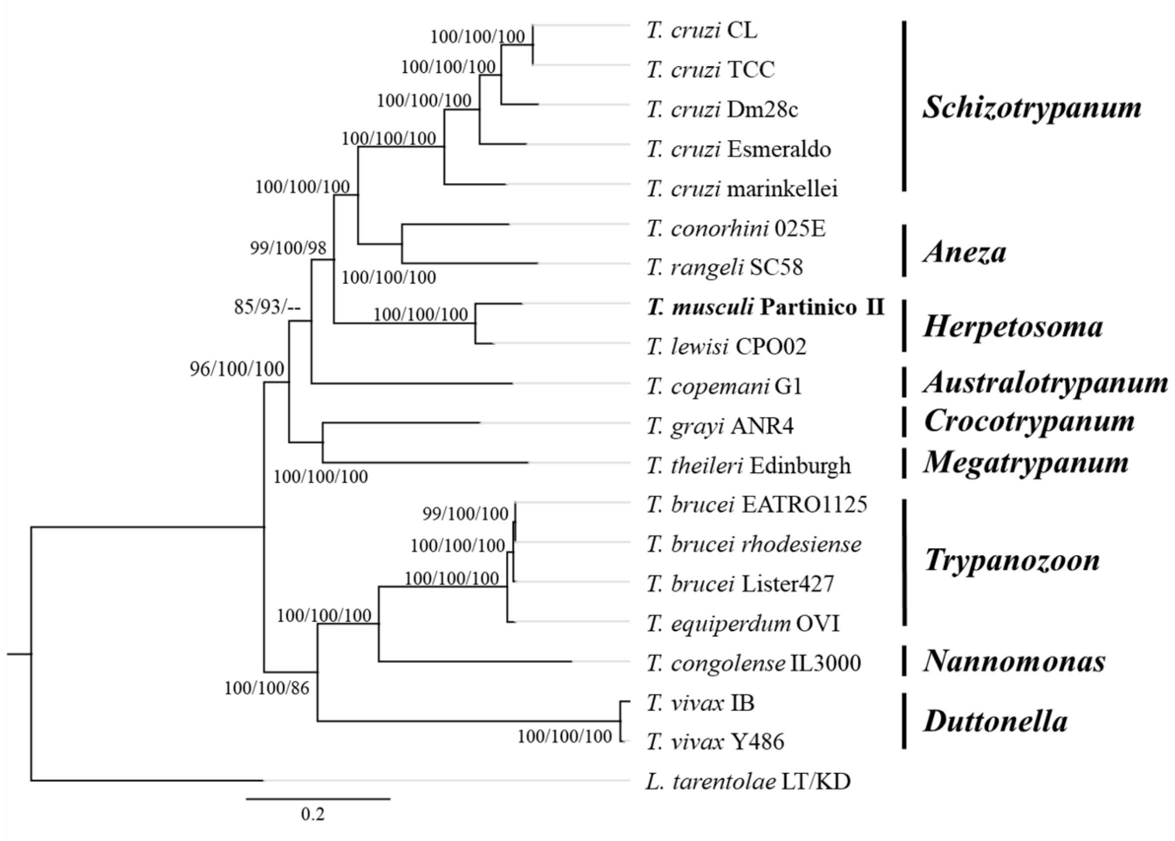
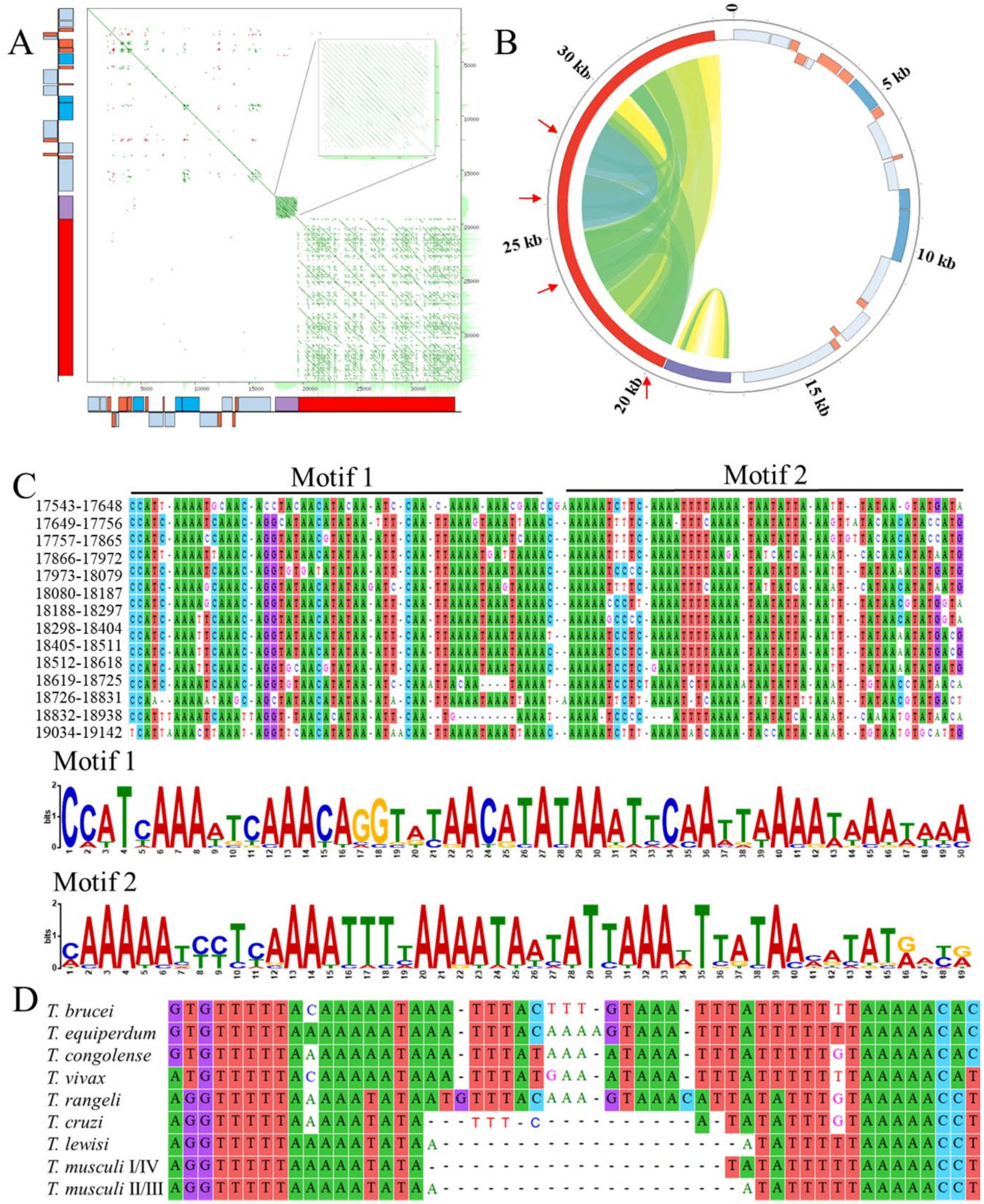


Figure 5



T. brucei GTGTTTTTACAAAAATAAA-TTTACTTT-GTAAA-TTTATTTTTTTAAAAACAC
T. equiperdum GTGTTTTTAAAAAAATAAA-TTTACAAAAGTAAA-TTTATTTTTTTAAAAACAC
T. congolense GTGTTTTTAAAAAAATAAA-TTTATAAA-ATAAA-TTTATTTTTGTAAAAACAC
T. vivax ATGTTTTTACAAAAATAAA-TTTATGAA-ATAAA-TTTATTTTTTTAAAAACAT
T. rangeli AGGTTTTTAAAAATATAATGTTTACAAA-GTAAACATTATATTTGTAAAAACCT
T. cruzi AGGTTTTTAAAAATATA---TTT-C----------A-TATATTTGTAAAAACCT
T. lewisi AGGTTTTTAAAAATATAA--------------------ATATTTTTAAAAACCT
T. musculi I/IV AGGTTTTTAAAAATATA--------------------TATATTTTTAAAAACCT
T. musculi II/III AGGTTTTTAAAAATATAA--------------------ATATTTTTAAAAACCT

Figure S1

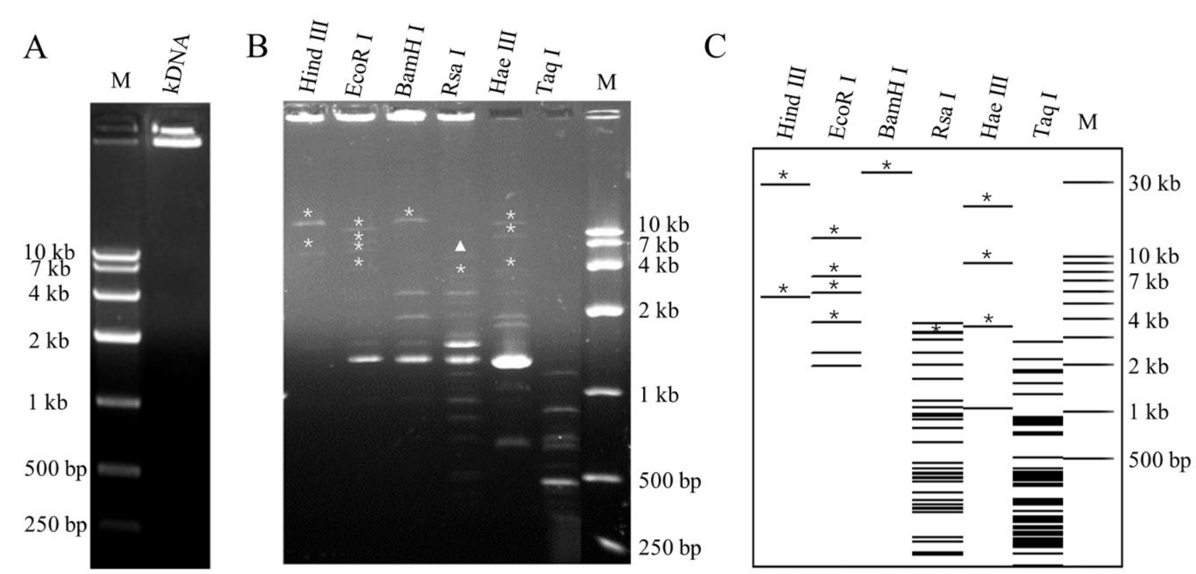


Figure S2

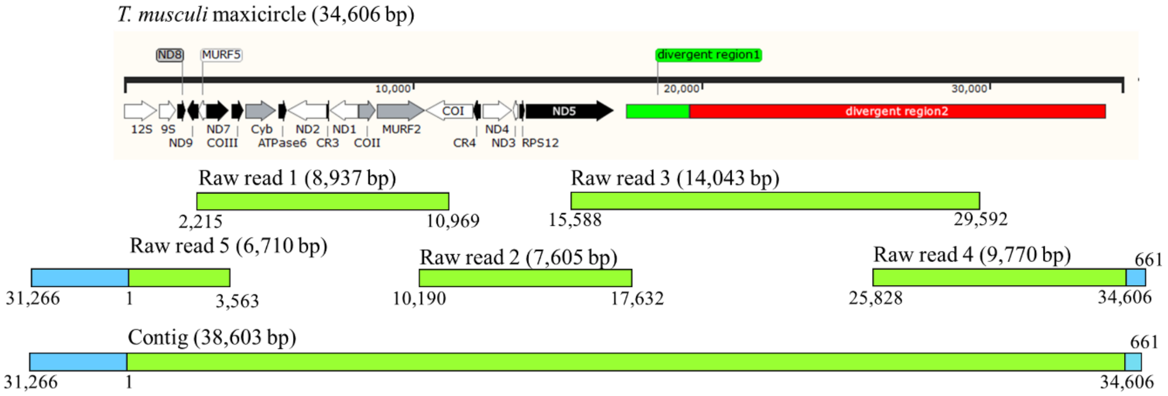


Figure S3

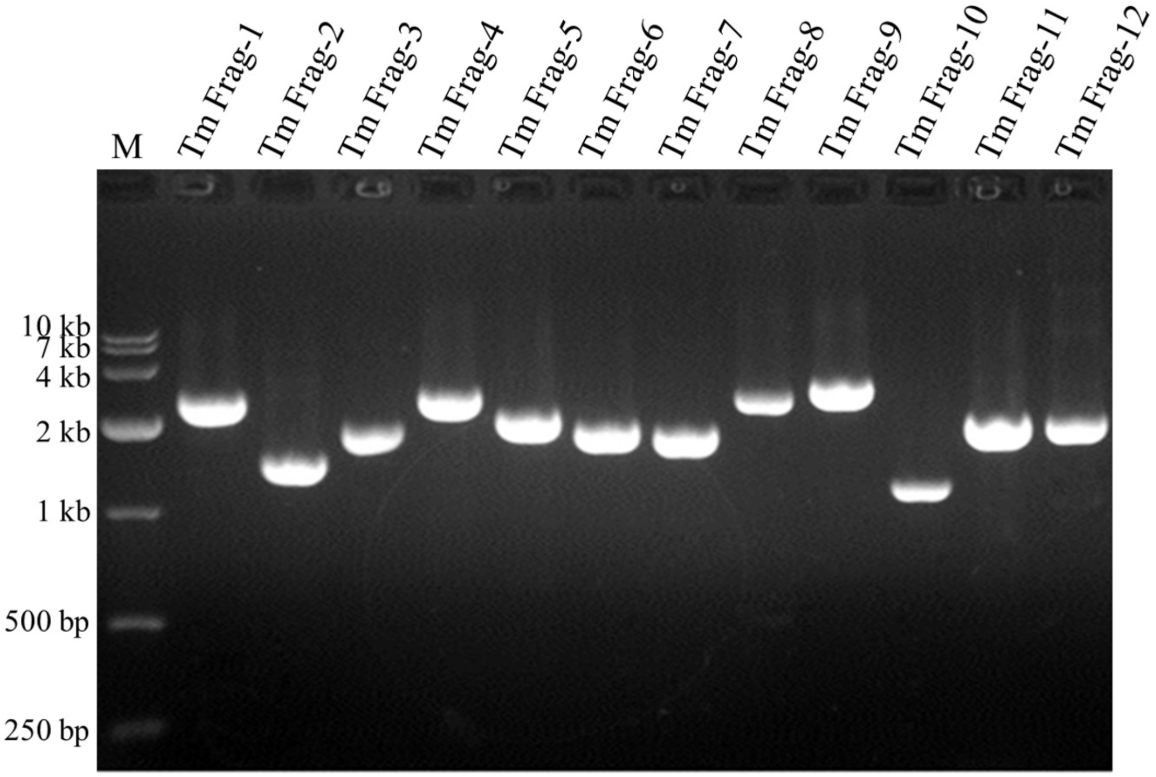

M
Tm Frag-1
Tm Frag-2
Tm Frag-3
Tm Frag-4
Tm Frag-5
Tm Frag-6
Tm Frag-7
Tm Frag-8
Tm Frag-9
Tm Frag-10
Tm Frag-11
Tm Frag-12
10 kb
7 kb
4 kb
2 kb
1 kb
500 bp
250 bp

Figure S4

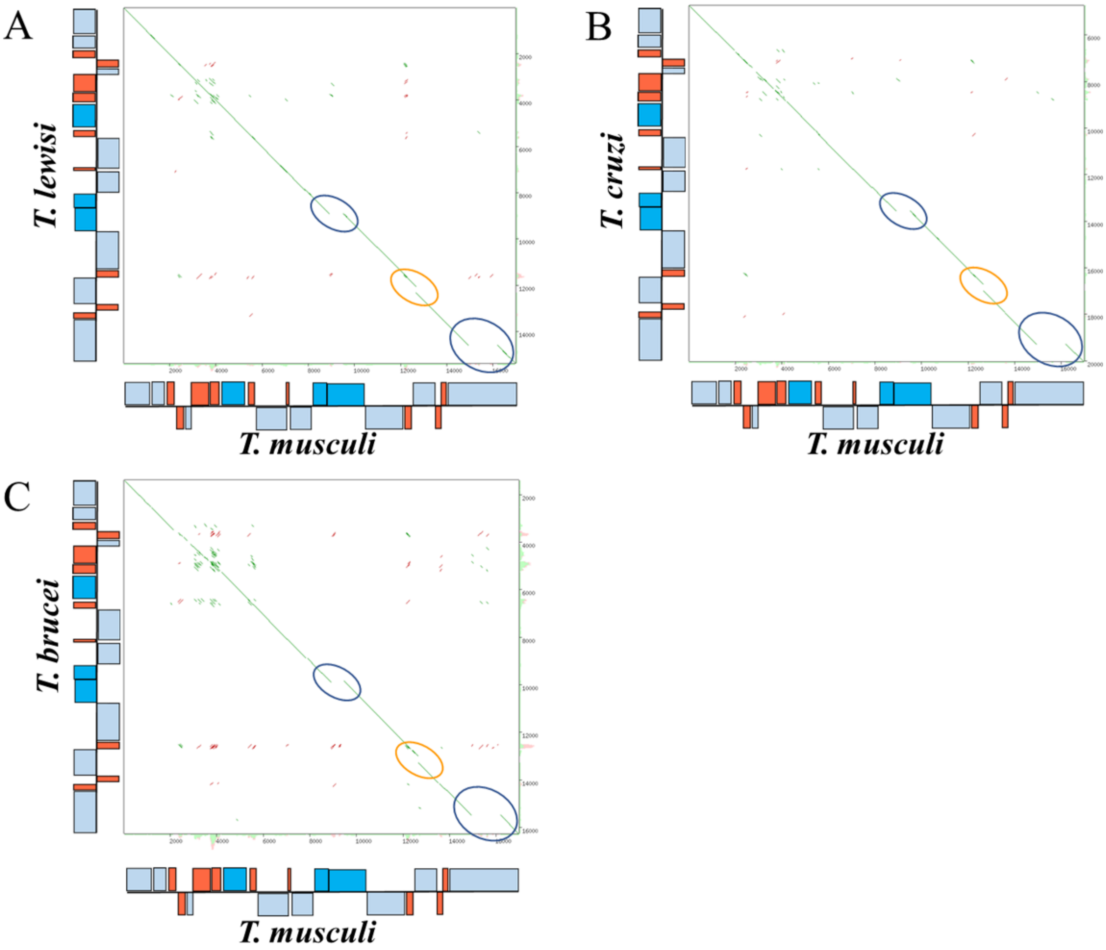


Figure S5

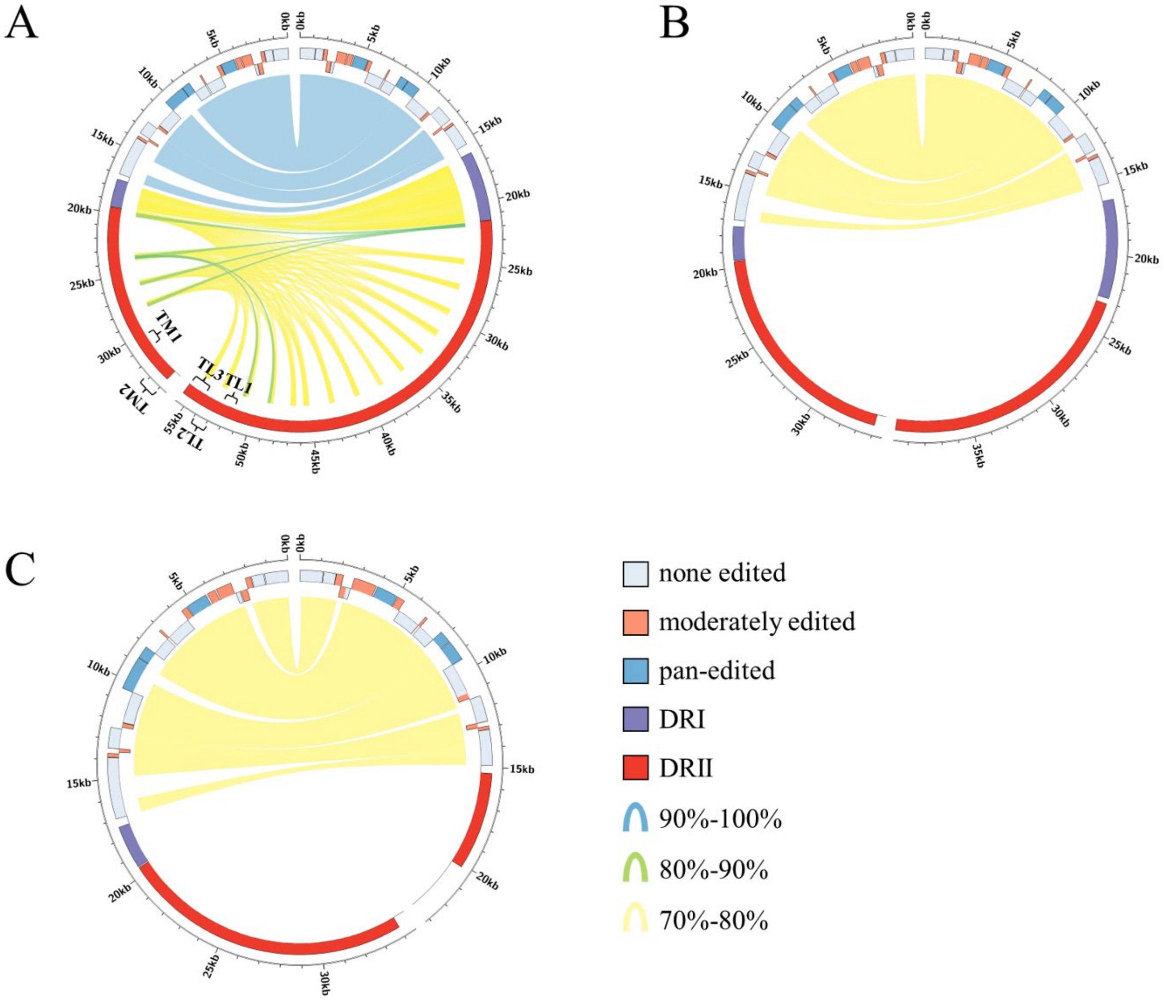

A
B
C
0kb
5kb
10kb
15kb
20kb
25kb
30kb
35kb
40kb
45kb
50kb
55kb
TM1
TM2
TL1
TL2
TL3
none edited
moderately edited
pan-edited
DRI
DRII
90%-100%
80%-90%
70%-80%

Figure S6

Figure S7

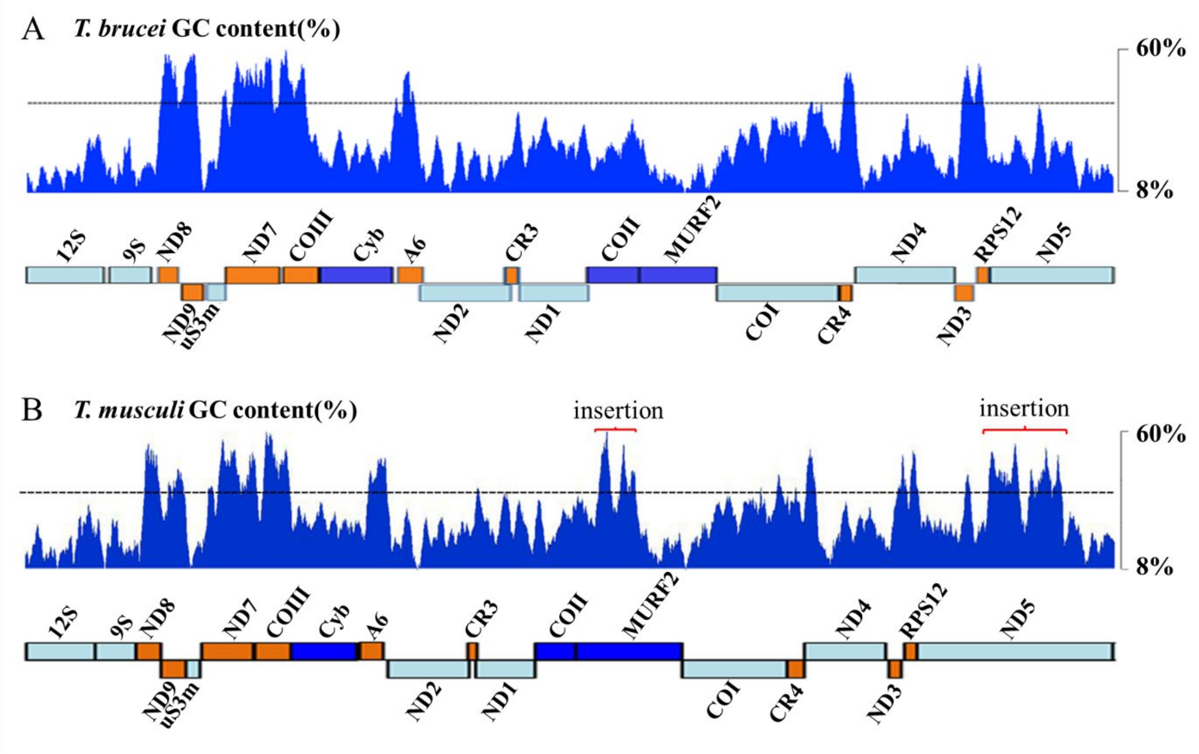


Figure S7

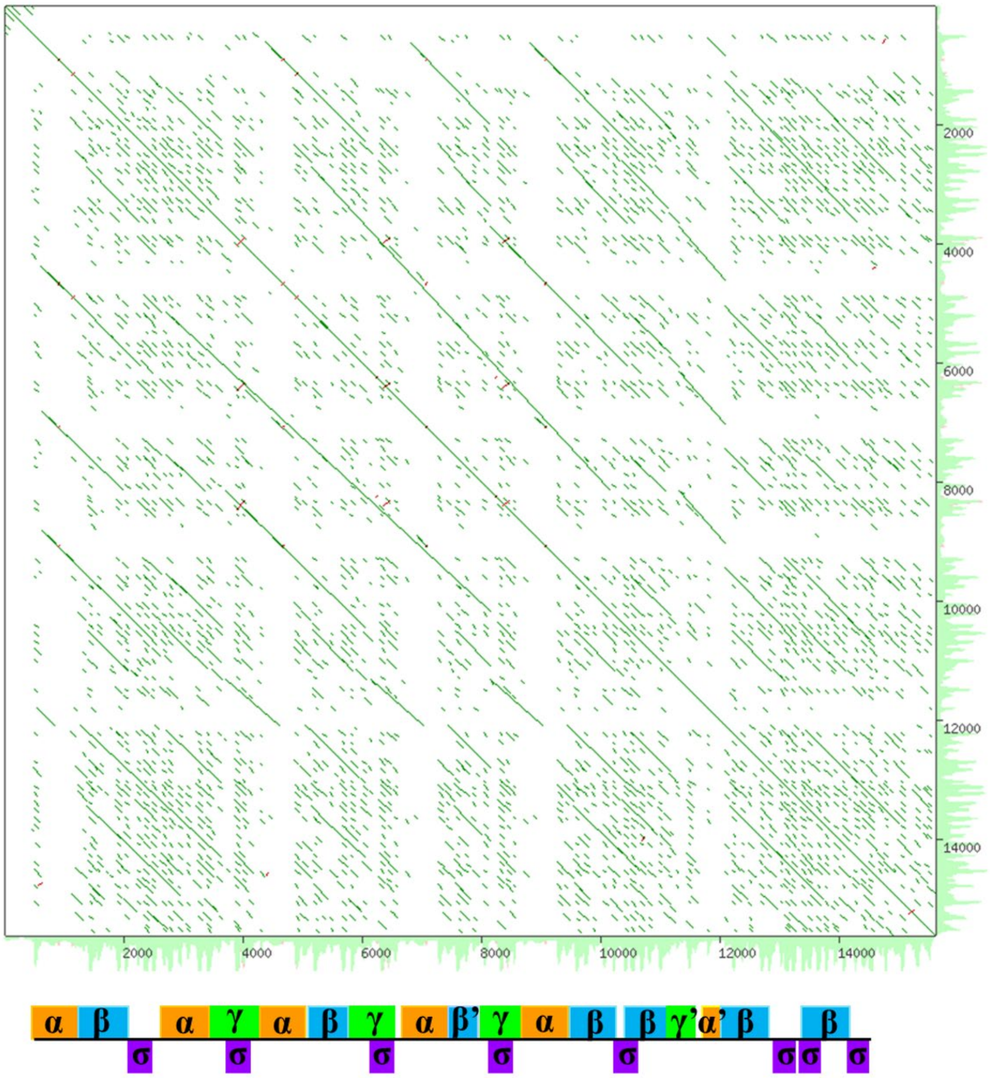
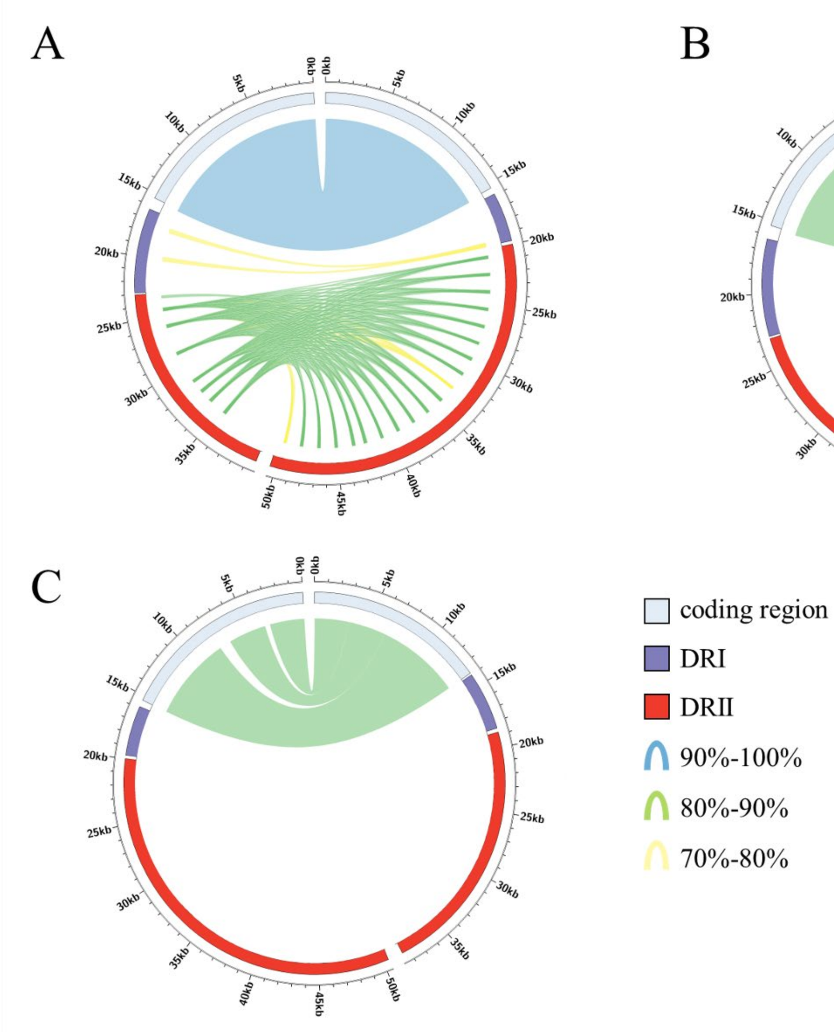

2000
4000
6000
8000
10000
12000
14000
α β α γ α β γ α β' γ α β β γ' α' β β
σ σ σ σ σ σ σ σ

Figure S8


A
B
C
0kb
5kb
10kb
15kb
20kb
25kb
30kb
35kb
40kb
45kb
50kb
coding region
DRI
DRII
90%-100%
80%-90%
70%-80%

Supplementary Table S1. Primers for PCR amplification of *T. musculi* (Tm) maxicircle

Primer name	Forward Primer (5’ →3’)	Reserve Primer (5’ →3’)	Region	Length (bp)
Tm Frag-1	AAACTATTCTTCCCATCTTACAA	CTGGCATCCATTTCTGACTA	33470-1438	2575
Tm Frag-2	AGTACAATTCGTTAGTTGGGTT	GGGGTAGAGAAGGAGGAGAT	1060-2493	1434
Tm Frag-3	AGGTTGCCCTCTTGTCGTCA	TCCCGTGTTCCTCCTTTCG	2216-4089	1874
Tm Frag-4	TTACCCGTTTCAGCACAGG	TCCTACTCAATGGGATGTCTTC	2909-5561	2653
Tm Frag-5	TGGACCAGAGGAGGTTGAAG	AGCGTTATGTTAGCATGTTTGT	3892-5999	2108
Tm Frag-6	GAAGACATCCCATTGAGTAGG	GTGAGAGTGAATTAGTTGCAGG	5540-7436	1897
Tm Frag-7	AAAGGTGAGGAAACCCATAAC	TCCCACCTATTTCACATACACC	7171-8986	1816
Tm Frag-8	CCAGGAAGGTGTAATGAGGT	ACTGAAGAGGGAATGGGTGT	8620-11718	3099
Tm Frag-9	GGTAGGAGGGTGATGTATTATTG	GATTTGTTATCTGTTAGTGGCG	9046-12028	2983
Tm Frag-10	GCACACCGCATCCTACTAAC	TGTAGCCTATGTATGGTTATGAA	11931-13093	1163
Tm Frag-11	TGTTCGCAATAAAGTATCCG	ACTACTAAGGTAGCCGCATGT	12694-14685	1992
Tm Frag-12	ACAGCATACTGCGGTTGAG	AGTGGATTCATACACCCATAAC	14327-16247	1921
MURF2	TAAGAAGTTTTGATTGAGTCG	CAGATGGCATTAAAACGAAAC	8775-10211 (8-1444)	1437
ND5	TGCGGCTACCTTAGTAGTGT	CAAGTCTTTACAATAAAACCCA	14668-16387 (739-2458)	1720
ND4	GTTTTTGTTTTGCGATTGAT	TGTAGCCTATGTATGGTTATGAA	12504-13093 (76-665)	590

Supplementary Table S2. Accession numbers of trypanosomatidae species sequences used in the work.

Species (Strain)	Accession numbers
Trypanosoma cruzi (CL)	DQ343645.1
Trypanosoma cruzi (TCC)	MN904528.1
Trypanosoma cruzi (Dm28c)	MN904527.1
Trypanosoma cruzi (Esmeraldo)	DQ343646.1
Trypanosoma cruzi (marinkellei)	KC427240.1
Trypanosoma cruzi (Y)	MW421591.1
Trypanosoma conorhini (025E)	MKKU01000412.1
Trypanosoma rangeli (SC58)	KJ803830.1
Trypanosoma musculi (Partinico II)	OM000218
Trypanosoma lewisi (CPO02)	KR072974, OM000219
Trypanosoma copemani (G1)	MG948557.1
Trypanosoma grayi (ANR4)	OM049542
Trypanosoma theileri (Edinburgh)	GCF_002087225.1
Trypanosoma brucei (EATRO 1125)	MK584625.1
Trypanosoma brucei (EATRO 427)	M94286.1
Trypanosoma brucei rhodesiense	OM049543
Trypanosoma brucei (Lister 427)	MN904526.1
Trypanosoma brucei equiperdum (OVI)	CZPT02000280.1
Trypanosoma brucei equiperdum (STIB	EU185799.1
Trypanosoma congolense (IL3000)	GCA_003013265.1
Trypanosoma vivax (IB)	MT090069.1
Trypanosoma vivax (Y486)	MT090068.1
Leishmania tarentolae (LT/KD)	MK514114.1

LOCUS Seq1 34606 bp DNA circular INV 22-DEC-2021

DEFINITION Trypanosoma musculi strain Partinico II maxicircle kinetoplast, complete sequence.

```
ACCESSION    Seq1
VERSION
KEYWORDS      .
SOURCE          Trypanosoma musculi
  ORGANISM    Trypanosoma musculi
              Eukaryota; Discoba; Euglenozoa; Kinetoplastea; Metakinetoplastina;
              Trypanosomatida; Trypanosomatidae; Trypanosoma.
REFERENCE    1    (bases 1 to 34606)
  AUTHORS     Wang,J.F., Lin,R.H., Zhang,X., Zheng,L.L., Hide,G., Lun,Z.R. and
              Lai,D.H.
  TITLE         Mitochondrial maxicircle of Trypanosoma musculi, a mouse
              trypanosome
  JOURNAL     unpublished
REFERENCE    2    (bases 1 to 34606)
  AUTHORS     Wang,J.F., Lin,R.H., Zhang,X., Zheng,L.L., Hide,G., Lun,Z.R. and
              Lai,D.H.
  TITLE         Direct Submission
  JOURNAL     Submitted (22-DEC-2021) School of Life Sciences, Sun Yat-Sen
              University, No. 135, Xingang Xi Road, Guangzhou, Guangdong 510275,
              China
```

```
COMMENT     Bankit Comment: ALT EMAIL:lanse0812@gmail.com
            Bankit Comment: TOTAL # OF SEQS:1

            ##Assembly-Data-START##
            Assembly Method       :: Canu v. 2.0
            Sequencing Technology :: PacBio
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..34606
                     /organism="Trypanosoma musculi"
                     /mol_type="genomic DNA"
                     /strain="Partinico II"
                     /db_xref="taxon:71806"
     rRNA            1..1164
                     /product="12S ribosomal RNA"
     rRNA            1210..1818
                     /product="9S ribosomal RNA"
     gene            1859..2130
                     /gene="ND8"
                     /note="NADH dehydrogenase subunit 8; RNA editing required
```

for translation"

gene **complement(2195..2536)**
/gene="ND9"
/note="NADH dehydrogenase subunit 9; RNA editing required for translation"
gene **complement(2575..2823)**
/gene="uS3m"
/note="ribosomal protein S3"
gene **2874..3647**
/gene="ND7"
/note="NADH dehydrogenase subunit 7; RNA editing required for translation"
gene **3721..4135**
/gene="COIII"
/note="cytochrome oxidase subunit 3; RNA editing required for translation"
gene **4222..5301**
/gene="Cyb"
/note="apocytochrome b; RNA editing required for translation"

gene 5342..5644
/gene="ATPase6"
/note="ATP synthase subunit 6; RNA editing required for translation"
gene complement(5687..7038)
/gene="ND2"
/note="NADH dehydrogenase subunit 2"
gene 6979..7111
/gene="CR3"
/note="C-rich region 3; RNA editing required for translation"
gene complement(7129..8112)
/gene="ND1"
/note="NADH dehydrogenase subunit 1"
gene 8111..8739
/gene="COII"
/note="cytochrome oxidase subunit 2; RNA editing required for translation"
gene 8767..10439
/gene="MURF2"

```
misc_feature     8897..9526
                  /gene="MURF2"
                  /note="insertion"
gene              complement(10429..12081)
                  /gene="COI"
                  /note="cytochrome oxidase subunit 1"
gene              complement(12102..12319)
                  /gene="CR4"
                  /note="C-rich region 4; RNA editing required for
                  translation"
gene              12428..13464
                  /gene="ND4"
                  /note="NADH dehydrogenase subunit 4"
gene              complement(13453..13636)
                  /gene="ND3"
                  /note="NADH dehydrogenase subunit 3; RNA editing required
                  for translation"
gene              13711..13908
                  /gene="RPS12"
                  /note="ribosomal protein S12; RNA editing required for
```

```
                     translation"
     gene            13929..16975
                     /gene="ND5"
                     /note="NADH dehydrogenase subunit 5"
     misc_feature    14923..16200
                     /gene="ND5"
                     /note="insertion"
     repeat_region   17400..19550
                     /note="divergent region I"
     repeat_region   19600..34000
                     /note="divergent region II"
     misc_feature    19898..19931
                     /note="palindromes I"
     misc_feature    23648..23681
                     /note="palindromes II"
     misc_feature    26061..26094
                     /note="palindromes III"
     misc_feature    28055..28088
                     /note="palindromes IV"
BASE COUNT     13756 a    3989 c    4229 g   12632 t
```

```
ORIGIN
        1 tttgtatcaa atttatttaa atttataata atgaagccaa ctttgttagt actaaaagtt
       61 atttaaattt tgtatatata tattattatt tgtattgtat aattagtata tgtattatat
      121 ttaaatgtta aatttgttgt tttatattta gataaagatt tgtgtattat cattgtaaat
      181 tataatttta aagtttattg aactgtgatg tgtagtttaa tattttttgt ttgatgttta
      241 aatatttaac tataggaggc acagttgtta tatatgtacc aagtaaaaat agtaagatta
      301 ttttatctaa attaataaat aaatattcaa actacaaaaa aatgaatatt atgaagtatt
      361 taaaaattaa actatttaac tgatatgaat atttaaaatt atgcagattt atttgaatat
      421 taaaattaca tatataattt gtaatttatt ttttatgaca taaagattca gtaagtaaat
      481 taccgcaatt aatatttaaa atttaatatt taatatttgt ttaattaaaa gtaacctata
      541 tcaaaatcat aagtattatt ataaaatact atttataaaa atataaatat attgttaata
      601 aaattatcgt attttaaatg cgtttattaa atgcgtttgt ctaagaaaaa tatttaagat
      661 tattcttgta tatatattta taaattaata attctatttt aaaaataaaa tcctcaattt
      721 acactttcaa agtagcatag taatttgtta actaattatt aaagcgttcc atagaaattt
      781 tttaaattaa aacaaattaa ataagttatg aaataaaata aaaattttaa taaaaattaa
      841 ataattaaaa tagggcaagt cctactctcc tttacaaaga gaatgtcaat aataaaatac
      901 ataattgtat gtttgattgg ggcaatacta tatttattta tatagcataa gaactataat
      961 tattgaaata ataaaaggtt cgagcaggtt aacaagcatt aaaactaaat gtgtttcatc
     1021 gtctacttat tgttgaaatc aaattgattg ttcatcaaaa gtacaattcg ttagttgggt
     1081 taaaatcgtt gtaaagcaga tttgtttata tatttaatta ttataaaatg taaaaagtaa
```

1141 tattagtacg caaggattga ttgtttgata tcaaaaaaat ttacctccac aatgtttgtt

1201 tgtaaacttg agtctgttgt tattattcat attaattttt ttaaaagttt ttattttta

1261 ttttagttta tttaatttta aaagatgtaa taaattttag gaatagataa taataattta

1321 taattttgat tagatgtatt tgttaatgct ataaaatggg tgtggaaata tttgtaaaat

1381 ttataacata ataataataa taaatcacat taaattatta gtcagaaatg gatgccagcc

1441 gttgcggtaa tttctatgct tttaaatatt atacatttat caaataaaat tgttaatatc

1501 aaattttagc caaaaaaaat tattgttaac ctttatttgt ttttaaacac catatggtat

1561 atgcaaatat aaaattatag taattatgta ttatattata ttatatttat tcacatattt

1621 tacataaaaa tatttaatga ttttgacacc atgataaggc tttaaatatt gaaattgtaa

1681 attttataat caaaacttat tatttatatg aaatattaaa atttagataa aacaataaat

1741 taaaaaggta ttgttgccca ccaattttta taataaaaat aacgtgcagt aattaatata

1801 tttataaaaa tgtattttta acatctaatt ttataataaa aatattttaa taaataaaag

1861 gagttttttt ggaaggggggg atcggttatt ttcccagaaa acaagagagg tccgggaatt

1921 ttttcagcga tttttgaggg agaacggagc gcgagagccc cagggcaaga gttttttttt

1981 cgggggggtca tcggcggagt tttttaacaa agtatttttt cctttgtcct tggaaggggg

2041 gcaggcccga cttttatgcc aacacatttt caggaggggg gccatgaagt ttttgcgctt

2101 tttttttaat taagggggag agaaggggag gggttttaga ggttttgttt tttatgtttg

2161 atgttttaga ttatgtcagc cacaaaatta atttaaaact ataatttaaa aacaaaggtt

2221 gccctcttgt cgtcaaaaaa atctatcaaa atcctctctt cctcacaata aaacactaaa

2281 atgaaaaggt ctttccccct ctccctgaat gtctcttcca agcataacct aaatccttcc

2341 aaaaaagaaa caaataaaaa aacatcaaaa caatgttccc cctctccccc ctaaaaacct

2401 cgccaaaact cttaaatccc tcgccaaccc tctctcctaa tccacaaacc aaaattcaaa

2461 aactaaaatc caaatctcct ccttctctac ccccaaacct ctccccaaaa tcaaaaattc

2521 cccgcaatct ctttctcttt atatttacac aaatttaatt ttatacatgt aactttataa

2581 aatgaaaata attaaatatt ctatattatt taataatatt aagaaggaat attttatcaa

2641 aataatattt atatgtaata tgtttaatca aattaatatt ccatttaagc aaaaaaagat

2701 catattgtat ttatcaatat aacaaacaag tgtatcaaaa aaaaaaaaat tatttcattt

2761 aaaccgtcaa gatataaaat aattataaac aaacatccat tttattttta gaataaacaa

2821 aattataatt atatgccaaa aatcgactac atgataaaat gtaaaaagaa ttaagaggga

2881 caaatcgttt acagacttca cagctttttt acccgtttca gcacaggggg ggaattgttg

2941 ggataagatt tattgatttt tgaaaggatt ttacatcgtg gaacagaaaa gttatgtgaa

3001 tataagtcaa tagaacagtg tttaccaatt tgatagatta gaagattttg gggaagaaca

3061 ggcattggag aaggcggggg caacgggcaa gcgaaagaga gactttttcg agaaggggga

3121 acattttagg gaaagggact ccggagacag aggaatttcg tgataaggag acggatttta

3181 ttttgggggc gagcaggttt ttttttagga gggaccttga gagggggaaa aggttaaacg

3241 ggaaggagat ttttgaggtg gggaatttcg gaacggcggg gcggatttga gatttgtgaa

3301 ttttaggagg ggatccgcaa ggaggggaga tgttcttcga aggatttttg aaattttata

3361 ggagaaggaa ttttgggagc ggactcgaat gggaattttt gcgagagaaa tttttttgaa

3421 gttttcaagg gggtttttttg gggggaaaag gattttttga accagtttta aaccaaagaa

3481 agggagggac ctggaatttc ttttttagct ttgcggaaca ccccaagttg ggagatttgg

3541 aatttgggtt ggagtttttt tcccggcgac gggcggaaca tttttatttg aagaggacgg

3601 ggacagaagg ctttcgagtg aaatttgagg ggaccgagat cgatttgtta agtatttgtt

3661 catcatcatt attaaaatta taacaagtaa aaaataataa gacatacaaa aaccagaaat

3721 aggctttttg gaaggggatt tttgggggac acccgccaga ttttggagga gaggaggaag

3781 ttttgggggg aggttgtttt aggggagggg gaagagggaa cgggagaggg acgtttgatc

3841 ttcaaagaga ggattggaag gaggcaaccg gggggggagg gatttagctt ttggaccaga

3901 ggaggttgaa gggaggagga aaaggaaaaa caattttaga gtttggggaa gggtttttga

3961 aggggttttt ttgggaggag aggagggggg aattttcaag ttgggaaaca gagacattta

4021 aggggggca gaattttgaa tgcaacacat ttgggggagg agttttggtt cgaaaggagg

4081 aacacgggag ggtttaagac aggattttgg ggggcgagag agagtgagag gggaattaga

4141 atggtttgtt tttaagattt taaaggtttg ggttttagga gtagaataac taaaataaaa

4201 aattatattt taaaaataaa aagcggagaa aaaagaaagg gtcttttaat gtcaggttgt

4261 ttgtatagaa tttatggagt agggtttagt ttaggttttt ttatagcgtt acagattata

4321 tgtggtgttt gtttagcatg actttttttt agttgtttta tatgtgctaa ttgatatttt

4381 gtgttatttt tgtgggattt tgatttaggt tttgtagtgc gtagtgtaca tatatgtttt

4441 acatcattat tatatttttt actatatgta catatcttta agtgtattgt tcttgtaata

4501 ttattcgaca cccatataat ggtgtgattt gtaggattta ttctatttat attcataata

4561 gtaattgcat ttataggtta tgttttacca tgtacgatga tgtcctattg aggtctaact

4621 gtatttagta atatcatagc aacagttcca attttaggtc agtgattatg ttattgaata

4681 tgaggtagtg aatttataaa cgactttaca ttattaaaac tacatgtatt acatgtgtta

4741 ttaccatttg ttttaatttt ggtattgttt ttacatctat tttgtttaca ctattttatg

4801 agttcagatg cattttgtga taggtttgca ttttattgtg aacgattatg tttttgtatg

4861 tgattttatt tgagggatat gtttttagcc tttttgattt tattttgtat gatgtatata

4921 atatttataa attgatattt tgtctttcac gaagagtcat gaattatagt ggatacactt

4981 aaaacatctg ataagatttt accagaatgg ttttttttat ttttatttgg ttttttaaaa

5041 gcagttccag acaaatttat gggattgttt ttaatggtta ttttattgtt tgcattgttt

5101 ttgtttatat taaactgcat attgtggttt gtttattgta gaagttcatt attatgacta

5161 acatattcat taatactatt ttatagtgtc tgaatgagtg gttttttagc tttatatgta

5221 gttttagctt atccaatatg aatggaatta caatattggg tattattaat atttttatta

5281 attgtgtgta gattagatta aataaaaatc aaactcaaag agaagtgtta aattaatata

5341 tatgaggagt ttggacggat tagagaaggg gatctaggag aggaaattga agaagaggga

5401 gaaattgggg agaggaggag gaagaaggat gatatttttg aacttatttg aagttagatg

5461 aggaatagca agggggaaaa agtgttaggg gaattgatgg agatcgggga gaagcaggct

5521 ttggacaaca gcttggtttg aagacatccc attgagtagg atttttgaga gagaatggga

5581 ggggggaatt tgggaaagag gccttttgcc gagcttagaa gaaagaacaa gaaaggagag

5641 ggaggtttgg aaagggggag gttttagttt agttttgtgt taaatttata aaatcaaatc

5701 aaataagtta ataatatcaa aaaaacctat ataaaaaaga caacaacaac aaaaaaataa

5761 attattaaga catactaatt gtaatctaat tacccaaaaa ccaataaatc gataagaaaa

5821 aataaaaata ttaactataa gtctaaagta aaaaatatag attataataa gtcataaaat

5881 aatataacac atatacctaa atagcttaat tgtagcagta aacaaaacac ttttaagaag

5941 aatccaaata caataggaat tgcaccaata cacaaaaaca aacatgctaa cataacgcta

6001 agtataaaat tattaaaaaa cgaaaaacga agttaataag aaaataattc tcattgtatt

6061 taatatcaaa tatattatat ataataaaaa aatagtataa aatcattaaa atacataaac

6121 ttatactata aatcaaatga aaatttaaat ataaataaaa aatataattt ttggttgata

6181 taaatattaa aaataaaaaa cagaaattat taacgctaag tacgaaaagc aaacttacct

6241 atatagtcca attaacacac aactacatga ttgtataagt actaatttta taagtagtat

6301 ataaaagtaa ataaagcata taagaataat aataataggt ataaaacaga ataaaacata

6361 aaaggtaagc atataaataa agataaaaat tcctaatcgt atgtaaagtt gttccataaa

6421 taaaattaaa aatgcaagat ataatttaaa taggattagt aatgtataga ataagcataa

6481 aacacccatc cattcttcac caaagaccaa aaaacaaaag tcacataaat aaaaacaaaa

6541 gtgtaaatta agtaatataa taacaaaaaa agtagtataa aattaattaa tcatgtaata

6601 ataccacaaa atcagtattt tataaccaaa acgactagaa attaaaagta agaaaagtaa

6661 aaggctaaaa agttcaaata atatcgatat aaaaattata tgtgttataa aaaggaatag

6721 gaatataaca atatattgaa taataattat cataaaaaga aatccagtat cgaatgtgag

6781 aaagcaaaac attaaaataa aacagcataa aaataaaact aaaaagacaa agcatacata

6841 tggaattcat aaacaaaaga atgaaaccat taatgagtgt taatcaaaat aaatcaaata

6901 ttattagtga catattgatg tagtcacaca agataacaaa actataaaat accaaaaaaa

6961 taaaaatgta aaggaagtat aaaaaacaaa aaaggatttt ttttttggag tttgttttgg

7021 ggacagtttg aggttttatt gggaaggaga aaggaagttt tgggattttg agggacagaa

7081 ggacgttttt agaagaagtt ttcgttatta ataccacaaa ttctattatt aaagaagtaa

7141 acaaatcgtt tttgataggg ctaatattaa aaaggtgagg aaacccataa caaaaagaaa

7201 aagtaaaata aaagtttgtg ctgtagtaat ttttaatcta aaacagataa ctcttggtac

7261 taaagataat aaagctatta ttaaaataga tttaaaacaa acaaacaaac caccaaaaca

7321 taaacttaaa aataaaatac tacataaaag taaatgacta atttctaaaa tagaataaat

7381 aacaaaaaaa gttcctgata gttctgtaat taatcctgca actaattcac tctcacattc

7441 catataatca aagggtaatc ttaatccatc tgcaagtaac cctattcaaa atataccgac

7501 gaacaatccc cctaagtaaa aacaattgta caaagttagt tgaccaacac atatgtcttt

7561 aataccaaaa aaactaaaaa agtctaataa ataaatcgaa taaaataaaa ttaataagga

7621 actttcagaa ataataccaa agaaaagtgt acgcatagtt gcaagataaa caaagcaact

7681 agaaaataag aagcaaccca ctaaaaatac actaaacaca ttacacacta aatgtatccc

7741 tagtaaaatc aaaatagaaa atcctttatc aaataaaatg aaccctagcg aaaaaaaaaa

7801 tcaaggaaaa aatatacaga acaatgtaat tgtcatagct attaataagt aaaaaatatc

7861 aaaaaatgta attatcatag ctattaataa gtaaaaaata tcaaaactaa taacaaataa

7921 tacaaattta agaaaaagtt taattccatc agtaataggt gttagtaaac caaataaaaa

7981 aagcgcaggt ccaattcgaa gttgtacaat agccaaaact tttcgttcac ataaactaac

8041 aaaaccagat aagactaaca aaactagaat aaaaataaaa ataataacaa taacatcaaa

8101 atgtattaac atgagttata tattaacttt ttgaatgata tttttaatag attcaattat

8161 agtcttatta tcttttgtcg tttttatatg tatttgaata atagcattat tattatctac

8221 aattttattt gtaactaagg taaatgatgt atattgtaca tgagatttta tatcatcaaa

8281 atttatagat acttattgat ttgttatagg ggtaatgttt attatgtgtt tattgttgcg

8341 tctatgtttg ttactatatt ttggttgttt aaattttgtg agtttcgatt tatgtaaggt

8401 tgtaggtttc cagtgatact gagtatattt tttatttggt gagacaacta tatttagtaa

8461 tttaatatta gaaagtgatt atttagtggg agatatgcgt ttattacagt gtaaccatgt

8521 gcttacatta ttaagtttag taatatataa gttatgggta tctgcagtag acgttattca

8581 ttcattcacg ttacctagtt taggtataaa agtagagaac caggaaggtg taatgaggta

8641 attttatttg catcaaataa tgcaacaatt tatggtcagt gtagtgaatt atgtggtgtg

8701 ttacatggat ttatgccagt agtaatatgt tttatataga taaatggtgt ataatctaca

8761 ataaaagggg attttaagaa gttttgattg agtcgaattt ttgatttgtt atgtgttaga

8821 acgtatgatt ttatattatg gtgatttgac ttggatttca tattatatga ttttgttttt

8881 gattttgtag tatgtatatg taatatagtc agggtgggaa ttgacgtgga atttgggggg

8941 tgatatcttt tgtggaggtg gggtggtgta tgtgaaatag gtgggattat tttgtcgatg

9001 ttgggagggg agggtgagag tggtaggggt gagaatgttg ggaggggtag gagggtgatg

9061 tattattgtg ggtggggtgg gtggggtggg agggggaaat taattatata tgatatgtat

9121 tatagatgtg atatagagtg ttcttattat attgtacatc aataacagag agtgtggtga

9181 tgatgagttg cgaaatttgg tgtgatataa ttatactaac atagatatac tcttaaatga

9241 attattattt gtgtttgggt gggggtgggt gtagaagatg atgatgatgg gtggggtgtt

9301 gggtgggagg gcgtggatgt tgtttgttaa aatagtgggg gtgggaaaca taggaggggt

9361 cgaagataaa aatttataaa ttatcgataa ttttggaccc aattttagtc gcaaagttcg

9421 attttgggag gggcgttcaa ctttttgggc ggaaattcat gcatgtcccc cgtgtgtttt

9481 ttgtcgattt ttggtgattt tttacgagat gggacatcaa tgggggtata acatttatat

9541 ttatatttat tttgggtttt tttatacgaa ttttttttag ttttgtattt gtattgttat

9601 ttataacgtt tttcggatca tatgctttat ctatgttgta tactgatata ttttatatct

9661 atatactcta taattgtttt ttcttcgcat ttagcataaa ttatatatta tactatctag

9721 aattttttac atttataata ttttatattt tttttgattt tgtaagtttt tcatgttata

9781 cctacaattt ttttggatta ttttatttat ttaatatcat attttgttct tatttatttt

9841 gtttatttta ttttataata tatttccttt tttgttttat attttttgta atacgatgtt

9901 tatttattgt aattttagat tttttgtttt ttaattttga tatctttact tcgttaacat

9961 tatgtgatat gagctatata gattttgtat gttttttaat tttatatgca aattttatct

10021 caaattacat atatggattt tattgtttta tatatatata ttaggtctta tatttcttat

10081 agtatatatg gtaataagtc tttttttgg ttttatattt ataatatatg gtttagtaac

10141 tcctctaata aattatttgt attgattata tgtaatttat tgtagaagtt gtttcgtttt

10201 aatgccatct gtattaatat tttttaaatt tatatatttt gatatatttt ttatatttgt

10261 ttttatatta atactattta taattagttt ttttagtttt tttcttaaag attttatgtt

10321 tttatcgtta ttttttgata tatttggttc attatttaat tatgatatac aagtatttaa

10381 tttattttac ttccaagagc agttttattt aacacaatta acatcaattt atatgtaaaa

10441 aaacaaaaac aatgaatgtc aaaaatatat aaagaaaaca aagcaaaaac aaagaattaa

10501 taataaataa tctaatacaa tatgtgtaaa gtctactaca agtaaatata gtgccattac

10561 acaaggtaat cacgaataaa aaaaaaaaaa acaaaataaa ctgtatacgt ataaattaac

10621 aaaaaataaa cagtaatccc aaaaaaataa aatactaaac aaacaacagc aaaacaaaac

10681 taaaaaagct ataagaacca taccgtataa tgtaaacgaa cttcaatata agaaactgat

10741 tggataatca gaaattcttc ttggaaaacc aaacatacca caactatgta gtggaaaaaa

10801 caccatatta gatccacatc acaaacaaca aataaagaaa aattctcaga ataaataaag

10861 ttcaatgggt aaccatttca tcaaaaaatg tataaaaccc ccaaaaaagc ctatagtagc

10921 ccctaacgac aaaacataat gaaagtgtgc tacaacaaag taagtatcat gtagtaaaat

10981 atcaagccca acatttgaca aaaataaacc tgtaactcca ccaagtaaaa acataatcac

11041 aaatatataa acgaagtaaa tttcaaagca tatacataag tcagtgtata aaaaactgta

11101 taaccaatta aataatttaa tacatgttgg taatccaatt aaaactgtaa ctgtaccaaa

11161 ataagtcctg gtatcaacat ccataccaac tacaaacata tgatgtgctc acacaaaaaa

11221 tcctaaaata cctataagca acatagaata aatcatagca acagcactaa atacacatct

11281 aaaaccaacg acctcgacaa ttgtagatac taatccaaac acaggcataa taataatata

11341 aacttcaggg tgaccaaaaa atcaaaataa atgttgaaat agaactaaat caccccctcc

11401 cacaacatca taaaatgatg tattaaaatt tctatcacat agaagtagtg taacaccacc

11461 agctaaaacg ggtagagtca ctattagtaa aatagaagta actaaaacac cccaaataaa

11521 taacgttcaa ataaaaaaac taaagtattt ccttctacaa caaaaaatag ttccaactac

11581 attaatggag ttcaaaatgc tagatacacc taaaaaatgt acagaaaaaa taacaaaatc

11641 acaagctaaa ctggaatgaa aatcaacaca aatcaatgtg ggatataacg ttcaacccac

11701 acccattccc tcttcagtca aaaacccact aactaaacat ccaaaccccc ccataaacat

11761 ccaaaaactc atattattta atctaggaaa taccatatca ggaaatccaa ccataacagg

11821 tgcaaaatag ttaactaatc cacctagtac aacaggcata ataaaagcaa ataccataat

11881 aagtccatgt gcagtaataa ttacattata aaattgataa tccccaaata gcacaccgca

11941 tcctactaac gcaagttcta atctaataaa caatgagtat acatatccaa caaatccgca

12001 taaaatcgcc actaacagat aacaaatccc aatcatctta tgtgatacac tcaaacaaac

12061 aaggcaaata aaaaacataa ttaaatataa caccattaaa actttcttat caacttaaac

12121 taattttccc gctttatcat tccctccccc aactaccccc aacccaacaa aaacccccac

12181 tttcttccac ctcccccaca aaaaaaaacc tccccatcat ttcgaaacct tcttccccaa

12241 ctctccaaaa ccccctccat tcttctcctt cccctttctt aaaaaaatcc tctatcaaaa

12301 tcctaccaaa actaaaaata aaagccaaaa aaaaaaatta aacctagttt tttttaaaac

12361 aataaatatt tataataaca cagatatcaa tcaaaatgta aagatgaaat taggttaaag

12421 taattaaatg ttaaaattaa tgttggtatt tataaatttt ataatattaa taacatccat

12481 aatatatata tatataaatt atagtttttg ttttgcgatt gatataaatt atatatacat

12541 aaatttatat attaatcata tatcactttg atttatattt ttaatggggc tcgtgttata

12601 ttttttaata tttttgttgt tcaaaaaatg tgtatcttat ataaaattct tctatttatt

12661 tagcgatttt catttgaatt ctattgttta taatgttcgc aataaagtat ccgatatgac

12721 ctttacatgt atgattacca gaaatgcatg tagaagtaaa tacagagatg agtgttctat

12781 tagcaagtgt agtactaaaa ataggttttt ttggaatgtt taaattttta tttatctcat

12841 ttaataatgt ttcattatga tttttgggat ttatagacag tataataata ttaggtataa

12901 catttttgtc aatgttatta atctttttgt ctgattacaa aaaaatagta gcacattggt

12961 caataatcca tacaggtatc ggattaatat tgttatgaca taatgatctt ttgtttatag

13021 ggttattatt aatgtgtaat ttagcacata tattaagttt atcatttatg ttcataacca

13081 tacataggct acatgtacga taattatggt gtgagaatat tccttttaat gatttcattt

13141 ttggagttag tgcatgaagt tcactttttt aggtcttttt ttatttaata tagactttcc

13201 ttttatgtta ctactttatg ttgacatttt tatactatat ggtttaattt cagtgtcata

13261 catatatata tattatgttt ttatgttgtt tcattatcag tgtttatatc atccttatac

13321 atttatatgt gtttaagttt ttattctttt gtatgattag ataagtattt aagattagat

13381 ttaactataa atgatatata tttttatatt ttaacctcgt ttttagtcat aacattttat

13441 tacgtcatat atttattatt ttaatttcta acaaatctaa caaaataccc tctaaaaacc

13501 tactactacc aaaactacca aaatggtcaa tctccccgaa tccttccttt aaaataaact

13561 aaaaaactct ctcccaaaac acacaacaaa cccttggttc ttctccaatc cttcttgatt

13621 taaaacccgt gattcttctc cccaaatgcc ccccgctctg ccccctaacc tttttcccca

13681 tatttttatg aatcaaaaat aaaataaaaa aggcgaggat ttttttttg agttttgggt

13741 tactttttgt ttgggagaaa gagcccactt tgaacccagt ccggaaccga cggtgaagct

13801 tcttggatta aaaggttggg gcgttggggg gagtagttca aaaagatttt attgtggggg

13861 agaacccttt tgtttttgat aaagaatttt tacatcgtag taagtaaatt ggtgggtgta

13921 aatgatagat gtttttgatt ttttttatag tattttttct atttggtttt atatcgggaa

13981 gtttactttt aggaagacat tttttgagtt tttgagtaac tacagtatta gtagtatctg

14041 tagcattatg tattatgttc acattttata tattgtcagt atgtatatat ggatattgct

14101 attatgattt ttgtataata ctaatgctag acttttgttt tgtatgttta actttttcat

14161 gtacaggctt ctatttattc tattatttat agatctggta ttttgtttta tactgtttta

14221 ctcattttat tacatgtatt ccgacatctt tttgggacgt tttttaaata tattttgatt

14281 atttgtaata tgcatgaatt tttttatatt atcttatgat ttctaacag catactgcgg

14341 ttgagaatta ttaggtttat tttcatattt tcttatatca tatttttgat ttcgatatta

14401 tgcactaaaa ttcggtttta aggctttttt tattagtaaa ataggagata ttttattaat

14461 atcaaccttt tctataatat tatctatgaa tggtttttgt gtaataacct ttcattatat

14521 aaatttttta tgtgtagact attattatat atgcttttcc atatgaatac taatgtgctg

14581 tgctttcaca aaaagtacac agtttggatt gcatatctga ttgccagatg caatggaagg

14641 cccaattcca gtttcagcat taatacatgc ggctacctta gtagtgtgtg gaataatatt

14701 agtttcattt atatattgat gtttcgattt ttgatttagt tatttttata acttaatctt

14761 gtgatcaact ttaataatag tattaatgag tttatgtgtt ttttataatt ttgatgcgaa

14821 acgttatata gcttttagta cgatatgtca aataagtttt tcattatttt gctgtatttg

14881 tttggattta tatatcggtt gtttattttt ttgctatcat acacctatag gaaataatca

14941 tatattaatt atttatttgg gtggaagatg tatgggaggg gttcgaagag ggggaggggt

15001 aagaggtagg ggtgcaaggg ggtcgatcat aaaatccacg aaaaaatgga cccaatttta

15061 gtcgcaaagt ttgattttgg gaggggcgtt caactttttg ggcggaaatt catgcatgtc

15121 ccctgtatgt tttttggtca tttttgacga ttttcacgag gtgggacatc aaagggggtt

15181 aatgtaatat agtcagggtg gaatcgacgg ggaaatttaa taaaattgtg ggaggaggag

15241 gattggtgtt tttgtggttg gaataatttt tggtagatgg tgggtgttcg aggttaggtg

15301 tttgggtata agaatgtttg tagaatttgg ttatgggggt ggagagtaat ttgtgtggag

15361 gggtgtagga taatatgggg agggaggggt gggaggggtg ggaggggtgt cgaagagaaa

15421 attaatatta tggctatata tagttattga ttattgatgt gatatgatat gtacataaaa

15481 caaacagaac gagatgtgag cggtgattat ataaaattta attatatcta tcctataggt

15541 aatatcttaa atttaaattt tatttattta ttgggggatg ggaaagaatt ttggtgggag

15601 gggtagacga gagggtagag ggttttgata aaatattttg gggaggggta gaagagggta

15661 taagagatta ttaaaattca ccaaaaaatg gccccaattt tagtcgcaaa gattgatttt

15721 gggaggtgcg ttcaactttt tgggcgaaaa tttatgcatg tcccccgtat gtttttctc

15781 gatttttgac attttttcac gaggtgggac atcaatgggg gttggtgtaa tatagtctgg

15841 gtgggatcga cgaggatttt agggtatgta gtcgaggggt gggagggtgg ggagaatatt

15901 atatggtgtg ataatgtatg ttggtgtttt tgtttatgtg atgtgtatgt tatcggaact

15961 aggtggttgt aggtttgtgt agtgatatga agtggaagtg gattggtggt atagatagag

16021 gtagaatatc ttgtagttgg ttactatctt ggaggggtgg gagggggtgg gggtgggggt

16081 gatataatta atagtgtatg atatttatta gagttgatat attgttatag attatataaa

16141 acaaacataa aatgatatct aaattataat agaggtagaa attatacctt aaagtgatat

16201 gttttctaaa gctactcttt ttatagttat gggtgtatga atccactttt tctttggttt

16261 acaagatata cgttgttatt tttttctata tttttgtggt tgtatattag ctaggttatt

16321 gcttatattt gcaattttaa attcatgttc aatatgattt ttatgtgggt tttattgtaa

16381 agacttgtta ttagcaatgt tattattaat atcttttcat aatatagtgg aatttgtatt

16441 tataagtata aatatttata ttttttacaa taatatacaa ttattttttg ttattttttt

16501 tattttatga aatggttttg tttagttgat tgtttatttt tattatttga ctttgaatgt

16561 tgtcttgtat attgtttaat ttgcgtctat atgtgtttta taactatatt tttcatattg

16621 gattatatat atatatttat atttgcaagt aattgtacat tttggatatt ttatctatat

16681 ttttacaatt tttttgatat agcaatattt acaatattta taataactat gtcagcattt

16741 ttgtattatg gttgcatatt tttttatttt tttaatgtag attgcatcat gttattttgg

16801 agaataataa taataatagt tacaatttgt gtgtttgcta tgttttgttg ttgatatttt

16861 atatgtttaa ttatttatgt atcgttgttt gtatgaaatt ttgttatata ttttagatat

16921 aacatgaagt attgtctatt tttttgcata ctatgattaa tctatatata aataacacgt

16981 gagtatcaat aatgtaaaaa gttatctaat tttaaagaat ataaaaaaac actaatataa

17041 tttggtttat tgaaatatat acgttaatct cataccaata tatataaaca aaaatatata

17101 tcaaactatt ttaaaattaa aatcaaaaac atatctaaaa ataaaatacg tataaatata

17161 caaaaataca taaacattaa actaaacaaa ttttcaaatt tctatttcaa taacataaat

17221 aaaaaccata aaaaatacta ataattgagc atggttaatt aaaagtaaat gcaaaaatgt

17281 ctaaaagtat aagatttata tcaaactact ttaaaatatc ttaagatccc tataattaaa

17341 agcaaaacaa aacaaaacaa aacaaaacca aaaaattaaa tataaatgca cagccaatgt

17401 ttaaaaccta acacacacac acacacaaaa aaaattccaa aatatggcaa caatatgtat

17461 atataaccaa caaaattaag cattaatcaa aaattttaaa atcttgaaat atcatcaaac

17521 ttctaagtgc actaacaaaa tccattaaaa tgcaacacct acaacataca aatccaacaa

17581 aaaaacgaac cgaaaaaatc ttcaaaattt taaaataata ttaaatttat aagtatgata

17641 acaaaatcca tcaaaatcaa acaggcataa catataattt caattaaagt aaattaaaca

17701 aaaattttca aatttcaaaa taatattaaa gttatacaac ataccatgca aaaacccatc

17761 aaaaccaaac aggtataacg tataaattca attaaaataa atcaaacaaa aattttcaaa

17821 attttaaaat aatattaaag tgttacaaca taccatgata aaacccatta aaattaaaca

17881 ggtataacat ataaattcaa ttaaaatgat taaaacaaaa attttcaaaa ttttaagata

17941 tcatcaaaat cacaacatat aatgacaaaa tccatcaaaa tcaaacaggt gtgatatata

18001 aattcaatta aaataaataa aacaaaaatc cccaaaattt taaaataata ttaaatttat

18061 aaaatatgat gacaaaatcc atcaaaagca aacaggtata acatataaga tccaattaaa

18121 ataagtaaaa caaaaatttt caaaatttca aaatattatc aaaatcataa catataatga

18181 caaaatccat caaaagcaaa caggtataac atataaattc aattaaaata aataaaacaa

18241 aaacccttaa aattttaaaa taatattaaa tttataacgt atggtagtaa caaaatccat

18301 caaattcaaa caggtataac atataaattc aattaaaata aataaaacaa aaagccccaa

18361 aattttaaaa taatattaaa tttataacat atggtaacaa aatccatcaa attcaaacag

18421 gtataacata taaattcaat taaaataaat aaaataaaaa tcctcaaaat tttaaaataa

18481 tattaaattt ataaaatatg acgataaaat ccatcaaatt caaacaggta taacatataa

18541 attcaattaa aataaataaa acaaaaatcc tcaaaatttt aaaataatat taaatttata

18601 aaatatgacg ataaaatcca tcaaattcaa acaggtgcaa cgtataaatt caattaaaat

18661 aaataaaaca aaaatcctcg aaattttaaa ataatattaa atttataaaa tatgatgaca

18721 aaatccatca aaatcaaaca ggtgtaacat ataaatccaa attacaataa aataaaaatc

18781 ctctaaaatc ttaaaaataa tattaaaatt gtaacctata acaacaaaat ccaaaaaaat

18841 aagcagctat aacatataaa tacaattaaa ataaattaaa taaaaaattc ttaaaattca

18901 aaatattatt ttaaattata acgtatgact gcaaaatcca tttaaaatca aattaggtta

18961 acacataaat tcaatgaaaa taaaatcccc attttaaaat aatatcaaaa tcaaaatgta

19021 taacaacaaa attcattaaa acttaaatag gttcaacata taaataacaa ttaaaataaa

19081 ttaaacaaaa atctttaaaa tatcaaaata ccattaaaat tgtaatgtgc attgacaaaa

19141 ctttattaaa acttaatagt tataacatat aaatccaatt taaataaaaa tccccaaata

19201 cttaaaatat cattaacgtc ataacaagcg ttaacaaacc tatcaataat taattaggtg

19261 taatttatta aacttagtca aaataaataa aaaaaaattc tcaaagcttt taaattttaa

19321 ttgattttgt taacatatag taacaaagaa agcacattag atcaaatagg tatcatatat

19381 aaataacaat aaaagcgtat agatcaaaaa caaaagtttc taaaattttg aagataatat

19441 caaaactaca ctgacaaaat tcattaaaat ctaattgagt tttcaaatcc tcaatacatt

19501 tatcacccaa aacactaaac aataatttga aattagtcac tatatccaat gcaccctata

19561 ttgtttatac gaatatattc gaattttcta aataatactt gtctgaacaa aatctttatt

19621 ctttacttta aatttttacg ccttaaataa aactaaaaat caatagcgct ttaaaaagta

19681 aaaatatatc acctaaaagc aaataacaaa agattattgg gttatttaaa agtaaaatta

19741 aatttgtaaa agtaaaaatt aataactcta tttcaataaa atacatcaaa acaaagcatt

19801 agcagtccgc acgcgcacac gggattaata ttataatata atagtatata tatatatata

19861 tatatatatt tgtgtacttt ctttatatat taccattagg tttttaaaaa tatatatatt

19921 tttaaaaacc ttctcacgat cgaattattt taaatggagg acagaaaggg ctagaagtca

19981 ttgtatatga agctaaataa tagaaaatat gcgatattgt ctatgtaaaa gtatatttac

20041 agttataata tattgggatg tatttattca aaagcaaata ttgaatgcta gttatatcac

20101 aagtaatagg gtgcaatcaa cttatcattt taagaaatta aataaatttt aaaatttatc

20161 ttagtgcttt tatatgcttt tatatgaatg ataattggaa ttctaaatga aaaatttcga

20221 attcttcata tctcataaaa ataacaaaaa gtactataat aaattaaaaa tactttaaat

20281 atcacatacc ttatattatc ccacataaat gtcgcttagg tccccaaatt tcaattgtgt

20341 aaacaaagtc tttattttct ctttaaaatt acatttaaaa agcaaataaa atattgttag

20401 atttaaccct aattaaccat tattttgaga atttaaataa taatttgaaa ttagttttga

20461 tattttatat tgtttcatat gaataatagt gatatcccaa cagtcaatta tttaaattaa

20521 aacttttatt tcataaagca aatgacaaaa acactagtta tttgaaacta aaaaaataat

20581 ttagaattag ttttgatctt atattacatt ttaacattgt ttaatatgaa ttgtacggaa

20641 gttctaagta gtcttatttt ttccaaacaa aagtcaatca tatttcatta ttcaaaaaag

20701 ttccaaagat agttttgaca tccaacatat tcattccata ttgtttcata ttaaacacta

20761 tcttgagttc tcaaatccta aataacaatt gtgtaaacaa agttttccct ctctttcctc

20821 tcatgcaaac ttctacattt ctcttaaagt atacatacta agacagtatt tcaaataatc

20881 taacttatta ttttaagaaa ctgaacaata atttaaaact agttttaata tttaatatat

20941 gttgtttcat atgaataaaa taatccaaag ttctaagtag tcaataacta tctaaatcaa

21001 aattttactt ctttttattc ataaaacaaa taacaaaaaa actaaataat aatataaaat

21061 tagtttcgat atttaacaca ttatatatat gctttcatta tgtaatatca ccaagttatc

21121 aattaccaaa aatccttgta ttttctttca ttaaaacttc tatctcgttt taaagtaaaa

21181 atacattaaa atttaattca aaaaaataaa gtttaaatta ttctttattt gtgaaattac

21241 ataataattt aagattggtt tcgatattta aaaataattt atgttgttaa gcatgaataa

21301 aaatttatta gaaacttatc taaattaaat caaccaacca ttatttcaaa aatttgaata

21361 ataatttgaa attagttttg atacccaatg tattttatat cgtttcatat aaataatagt

21421 aaagttccaa tcaaccgttt aaatcaaaaa ttccagattc tacctcacaa agtaaataac

21481 aaaaattaaa taataattta aaattagttt tgatattcaa catattttat atgatctcat

21541 aagaatatta atttaggtct ctctctctct ctacttttag ttgtgtaaac taaaatcttt

21601 attttctttt aagattgtat ctaaaacaaa atcaatgaat cacttttgaa aaatcgaata

21661 acaataatcc aagattaatt tcgatatcca atataattta tactgtttta tacaaataaa

21721 aattgttagg attttatgta aatttaatca ttattttgaa aatttgaata ataatttgaa

21781 attagttttg atatctaatg aattttatat tgttttatac gagtaagatg tttaatagtg

21841 atgttctgtc aattatttaa attaaaactc atttccatct tagaaagtaa acaaaaacaa

21901 catcagttgt ctaaaaccaa aaaattaatt tagaattagt tttaatatct attggtattt

21961 tatattgttt catatgaata tgaattatat ggaagttcta aatagttcta aatttttcga

22021 acaaaaatcg ccatattgat tactatccaa aaaagatcca aaagtagttt cgacatccaa

22081 tgtatcctat attgtttcat atcgaacact atattgagtt ttcaaatcct gaaaagcaat

22141 tgtgcaaata aagttctttc ttttgctctt tttctcatcc aaacttctac atctctctta

22201 gagtataaat atcaaaaaat agtatctcaa gtaatctaac ttattatttt aagaagccga

22261 acaataattt aaaactagtt tcgatattta atgtaataca tgttatttca tatgaataat

22321 ctaaagtttt aagactcaat aactatttaa attaaaattc tatttctttt cccctataaa

22381 aattaaataa caaaaactta aataaattta aaattagttt cgatatttaa cacattatat

22441 acgctttcat atgaataata caaaagtttc aattatcaat tacgtaaaca aagtcttttt

22501 tatcctcctt tgtcaaaact tctgtttctc ttttaaaagt aaaaatttat taaatcgtaa

22561 tccaaaaaat agaattaatt acttctattc gtgaaattgt ataataattt agaaattagt

22621 ttcgcaaatt caatgcattt tattaaatac tatttaaaat tcggttaatt gttttaaatc

22681 taaagtttcc aattattttt atcacataaa gtaaaaaaac aaaattatta aatacccaaa

22741 agcaaatatc aaaagtaatt aagttattta aaactaatca acaaaacgtt atcaagtaat

22801 gcaaaagcaa aacattaatc gattattatt tgagaaatta aatacaaagg tattgtatca

22861 aaaaaaacaa aataatcact gttttaaaaa attaaataat aatttgaaat tagttttgat

22921 atccaaaaca atttatgttg ttctacatga ataaagattt attagaactt tattcaaatt

22981 aacccactaa ttattatttt aagaatttga ataataattt gaaattagtt tgatatccaa

23041 tgtactctaa tattgcataa atagtactga agttctaata attgtttaaa ttaaaaactc

23101 tagcatctat aaaataaaac aaataacgaa aaataataat ttaaaactag tttatttaac

23161 ccgttttata taaatattga aaacttggtt gattgttcaa ataaacctcc aatctatttc

23221 ttatcaacac aaacttaaaa caaaaatttt aaattgaaaa caattaagtt attctaaaag

23281 taaccaaaca aaagtttatt ttatattgtt tcatatgaat aattgtttgt atttataata

23341 aattatttga ataaaatttt tatttcaata ttacactttt ctccaaactt tattctaatc

23401 tcaattaaaa ctaaaaacta aatagtatcc aaaagcaaaa acgagtcact caaaagcaaa

23461 taataaaagt actggtatct aaaaataaaa ttaaatttgc aaaaagtaaa aattaataaa

23521 tttattccaa taaaatacat taaaacaaag cattaacagt tcgcgcgcgc acaggggatt

23581 aatattataa tataatagta tatatatata tatatatatt tatgtacttt ctttatatat

23641 tactattagg tttttaaaaa tataaatatt tttaaaaacc tactcacgat cgaattattt

23701 taaatggagg acagaaaggg ctagaagtta ttgtatatga agctaaatat tagaaaatat

23761 gcgatattgt ctatgtaaaa gtatatttac agttatgaat gtattgggat gtatttactc

23821 aaaagcaaac attaaaagct agttatatat cacaagtaat aggtacaatc aacattatta

23881 ttttaagaaa ttaaataaat tttaaaatta gtttgatatc ttagtaaatt ctatatggtt

23941 ttatacgaat aataattgaa attctaaatg aaaaattttg aattcttttt atctcattaa

24001 aaataacaaa aactactaaa ataaattaaa attactttta atatccaacg cgccttatgt

24061 aatcccatat aagtatcgct taagtcccca aattttcatt gtgtaaacta aagtctttat

24121 tttcttttaa gactacatct aaataaaata aaaattgtta gatttaatta atcctaatta

24181 accattattt tgagaattaa ataataattt gaaattagtt ttgatatttt atattgtttc

24241 atatgaataa tagtgatatt tcaacggtta attatttaaa ttaaaattta cttttatctc

24301 ataaagcaaa taaataacaa aaatactagt tacttgaaac taaaaaaata atttagaatt

24361 agttttaatc ttatattata ttttatatga attgcatgga agttctaagt agtcttaatt

24421 tttcgaacaa aaagtcaatc atattgatca ctattcaaaa aagatccaaa ggtagtcttg

24481 acatccaata tgtttcctat tgtttcataa taaacactat cttgaatttt taaatcttaa

24541 atagcaattg tgtaaataaa gttttctctt ttctcttttt cttattcaaa cttttacaat

24601 ttcttttaga gtaaaaatat taaaaatagt atctcaagta atctaactta ttattttaag

24661 acgctgaaca ataatttaaa actagtttcg atatttaata taatacatgt tatttcatat

24721 gaataatcta aagttttaag actcaataac tatctaaata taaaattcta tttcttttca

24781 cttataaaaa acaaacaaca aaaactaaac aatagtaata atttaaaatt agtttcgata

24841 tttaacacat tatatacgct ttcatacgaa tatcacaaaa ttttcaatta ccaattatgt

24901 aaacaaagtt ttttttatcc ccctttgtca aaacttctgt ttctcttttt aaagtaaaaa

24961 tttattaaat tataatccaa aaatagaatt aaatcacttc tatttgtgaa atcgtataat

25021 aatttagaaa ttagtttcgc aattcaatgc attttattta attgattcca ttaaatacta

25081 tcaaaatttg gttaattgtt ttaaacctaa agccttcaat tatttttatc acataaagta

25141 aaaaacaaaa ttattaaata tccaaaagca aatataaaaa gtaattaagt tatttaaaac

25201 taatcaacac aaagttatca agtaatgcaa aagcaaaaca ttaatcgatt attattttga

25261 gaaattaaat acaaaggtat tgtatcaaaa aaaacaaaat aattactgtt ttaaaaaatt

25321 aaataataat ttgaaattag ttttgatatc caaaacgatt tatgttgttc tacatgaata

25381 aagatttatt agaactttat tcaaatataa cccactaatt attattttaa gaatttgaat

25441 aataatttga aattagtttg atacccaatg tattctaata ttgcataaat agtactgaag

25501 ttctaataat tgtttaaatt aaaaactcta gcatctataa aacaaataac gaaaaataat

25561 aatttaaaat tagtttaact taacccgttt tatataaata ttgaaaactt ggctaattgc

25621 tcaaataaac ctccaatcta tttcttatca acacaaactt aaaacaaaaa ttttaaattg

25681 aaaacaatta agttattcta aaagtaacca aacaaaagtt tattttatat tttttcatat

25741 gaatattgtt tgtatttata ataaattatt tgaataaaat ttttatttca atattacact

25801 tttctccaaa ctttattcta atcttaatta aaactaaaaa ctaatagtat ccaaaagcaa

25861 aaacgagtca ctcaaaagca aataataaaa gtactggtat ctaaaaataa aattaaattt

25921 gcaaaaagta aaaatcaata aatttattcc aataaaatac attaaaacaa agcattaaca

25981 gtccgcgcgc gcacagggga ttaatattat aatataatag tatatatata ttttatgtac

26041 tttctttata tattactatt aggtttttaa aaatataaat atttttaaaa accttctcac

26101 gatcgaatta ttttaaatgg aaaacagaaa gggctagaag ttattgtata tgaagttaaa

26161 tagagaaaat atacgatatt gtgtatgtaa aagtatattt acagttatga atatattgcg

26221 atgtatttac tcaaaagcaa gcatttatat cataagtaag tagggtataa tcaacttatt

26281 attttaagaa attaaataaa ttttaaaatt agtttgacat cttagtgcat tctatacgtt

26341 ttatacgaat aataattgaa attctaaatg aaaaattttg aattcctcct atctcataaa

26401 aataacaaaa ggtactataa taaattaaaa ttacttttaa tatccaacat accttatgta

26461 atcccacata aatatcgctt aggtcctcaa atttccattg tgtgcacaaa gtctttattt

26521 ttttaagatt acatctaaaa aacaaataaa aattgttaga tttaacctta attaaccatt

26581 attttgagaa ttaaataata atttgaaatt agttttgata ttttatattg tttcatatga

26641 ataatagtga tgtttcaaca gttaattatt taaattaaaa cttactttta tttcataaag

26701 caaataacaa aaacactagt tacttgaaac taaaaaataa tttagaatta gttattatat

26761 ttcacattgt ttcatatgga attacatgga agttctaagt agtcttaatt tttcgaacaa

26821 aagtcaatca tattgattac tactcaaaaa agatccaaag gtatttttga catccaaaca

26881 tatttcatat tgtttcatac taaactatct tggattctta aatcctaaat agcaattgtg

26941 taaataaagt tttttctttt ctctttttct cattcaaact tctacatttc tcttagagta

27001 tacatactaa ggtcatgaaa taatctaact tattatttta agaattaaat aataaatttg

27061 aaattagttt cgatatttta tataatatat gttacttaat atgaaatata atgcaaagct

27121 ccaagactta ataataaaca aaattattaa atatccaaaa gcaaatataa aaagtaatta

27181 agttatttaa aactaatcaa cacaaagtta tcaagtaatg caaaagcaaa acattaatcg

27241 attattattt tgagaaatta aatacaaagg tattgtatca aaaaaaaaca aaataattac

27301 tgttttaaaa aattaaataa taatttgaaa ttagttttga tatccaaaac gatttatgtt

27361 gttctacatg aataaagatt tattggaact ttattcaaat ataacccact aattattatt

27421 ttaagaattt gaataataat ttgaaattag tttgataccc aatgtattct aatattgcat

27481 aaatagtact gaagttctaa taattgttta aattaaaaac tctagcatct ataaaacaaa

27541 taacgaaaaa taataattta aaattagttt aacttaacct gttttatata aatattgaaa

27601 acttggctaa ttgctcaaat aaacctccaa tttatttctt atcaacacaa acttaaaata

27661 aaaattttaa attgaaaaca attaagttat tctaaaagta accaaacaaa agtttatttt

27721 atattttttc atatgaatat tgtttgtatt tataataaat tatttgaata aaatttttat

27781 ttcaatatta cacttttctc caaactttat tctaatctca attaaaacta aaaactaata

27841 gtatccaaaa gcaaaaacga gtcactcaaa agtaaataat aaaagtactg gtatctaaaa

27901 ataaaattaa atttgcaaaa agtaaaaatc aataaattta ttccaataaa atacattaaa

27961 acaaagcatt aacagtccac gcgcgcacag gggattaata ttataatata atagtatata

28021 tatattttat gtactttctt tatatattac tattaggttt ttaaaaatat atatattttt

28081 aaaaaccttt tcacgatcga attattttaa atggaggaca gaaagggcta gaagttattg

28141 tatatgaagt taaatagaga aaatatacga tattgtgtat gtaaaagtat atttacagtt

28201 atgaatatat tgcgatgtat ttactcaaaa gcaagcattt atatcataag taagtagggt

28261 ataatcaact tattatttta agaaattaaa taaattttaa aattagtttg acatcttagt

28321 gcattctata cgttttatac gaataataat tgaaattcta aatgaaaaat tttgaattcc

28381 tcttatctca taaaaacaac aaaaggtagt ataataaatt aaaattactt ttaatatcga

28441 acatacctta tgtaatccca cataaatatc gcttaggtcc tcaaatttcc attgtgtgca

28501 caaagtcttt attttttaa gattacatct aaaaaacaaa taaaaattgt tagatttaac

28561 cttaattaac cattattttg agaattaaat aataatttga aattagtttt gatattttat

28621 attgtttcat atgaataata gtgatgtttc aacagttaat tatttaaatt aaaacttact

28681 tttatttcat aaagcaaata acaaaaacac tagttacttg aaactaaaaa ataatttaga

28741 attagtttta atcttatatt atatttcaca ttgtttcata tggaattaca tggaagttct

28801 aagtagtctt aatttttcga acaaaagtca atcatattga tcactattca aaaaagatcc

28861 aaaggttatt tttgacatcc aaacatattt catattgttt catactaaac tatcttggat

28921 tcttaaatcc taaatagcaa ttgtgtaaat aaagttttct cttttctctt tttctcattc

28981 aaacttctac atttctctta gagtatacat actaaggtca tgaaataatc taacttatta

29041 ttttaagaat taaataataa atttgaaatt agtttcgata ttttatataa tatatgttac

29101 ttaatatgaa atataattca aagctccaag acttaataat tatttaaatc aaaattctat

29161 ttcttttcat ccataaaata acaaaaaact taaataatag taataattta aaattagttt

29221 cgatatttaa catattatat atgttttcat attactacta agttctccat tatcaattat

29281 gtaaacaaat cctttatttt cctttgttaa aacttctatc tctcttttaa agtaaaaatt

29341 tattaaagca taaattcaaa aataagatta aattatctct tatatgtggc attacataat

29401 aattcgaaat tagtttcgat atttaaaaac aatttatgtt gtttcatatg aataaaaagt

29461 ttattagaaa tgtattcaaa ttaattttat caatcattat tttaaaaatt tgaataataa

29521 tttgaaatta gttttgatat ctaatgaatt ttatattgtt ttatacgaat ataatgttcg

29581 atagtgatgt tctgtcaatt atttaaatta aaacttcttt tcatcttaga aaataaacaa

29641 caaaaacatc agttttctaa aactaaaaaa ataatttaga attagttttg atatctattg

29701 gtttttatat cgtttcatat aaatcgtatg gaagttctta aaataacccc aaatttttcg

29761 aattaaaagt caaccatatt aattattatc caaaaaagat ttaaaagtag tttcgacatc

29821 caacacattc catattgtct tatatcgaat actattttga gttcccaaat agtagttgtg

29881 taaataaagt tctctctttt ttttcactca tttaaatttt tatatttctt ttagactaca

29941 aatatcaaaa agtgtattct aaataatcta attattattt caagaaattg agcaataatt

30001 tgaaattagt ttcgatatat atgttgttcc atatgaataa cgcaaagttc caaatagtca

30061 ataactatct aaattaaaat tctatttctt tcctctcata aatcaaataa caaaaaacta

30121 aataatagta ataatttaaa attagttcca ttatgtaaac aaaactttta ttttccttta

30181 tcaaaactcc tgtttttctt ttaaagtaaa aatttattaa atcttaatcc aaaaaacaaa

30241 attaaattat ctttattata atatttagaa attagttttg taattcaatg cattttattt

30301 aattcttaat tattcaaaat cacatattaa aagtaattaa gttatttaaa actaatcaat

30361 aaaaagttat taagtcattt aaaagcaaaa cattaatcga ttattatttg agaaattaaa

30421 ttaaacacaa aggtattgta tctaaaaaaa acaaaataac tactatttaa aaaaattgaa

30481 taatagtttg aaattagttt cgataccaaa acaattttgt tgttttatat gcataaagga

30541 tttattggaa cttattcaaa ttaactcatt aaccattatt ttaagaataa aacaaataac

30601 aaaaataata acttaaaatt agttaataca tcttacttat ccgttttata taaatattga

30661 aaacttggta attgtttaaa ttaaaccttc aatccttttt catcaataca aacttaaaac

30721 aaaaatttta aattgaaaac aatttagtta tctcaaaagt aaccaaacaa aagtttgttt

30781 tatattgttt catatgaata ttgcttgaat ttataaataa taaattatct gaataaaatt

30841 tttattcaat attacaattt cctccaactc tactcttaca tctcaattaa aactaaaaac

30901 taatagtgtc taaaagcaaa aaatgaacca cataaaagca aacaataaaa gtactggtat

30961 tcaaaagtaa aattaaattt gcaaaaagta aaaattaata attttgttct aataaaatat

31021 atcaaaataa aggattaaca atccgctcaa cacacagggg tttaacatta taatataata

31081 acatacttta tatggtccca cgtaaatatt acttaggtct gtttccaaat tccaattgtg

31141 taaacaatta catctaaaaa aacaaaaaaa attgttagat ctaatcctaa ttaacttaac

31201 ccttatttta agaattaaat aataatttga aattagtttc gatattttat attgttttat

31261 atgaatatca gtgataccca aacagttaat tatttaaagt aaaacttact tttatctcat

31321 aaagtaaaca acaaaaacat tagttacttg aaacttaaaa aataatttag aattagtttt

31381 aatctttaat tacattgttt catatgaatt gtatggaaat cctaagtagt ctttagtttt

31441 ttgaataaaa gtcaccatat tgatccccat tcaaaaaaga tccaaaggta gtttcgacat

31501 ccaacatatt ccatattgtt tcatattaaa cactatttta gattctcaaa ttcctaaata

31561 gcaattgtgt aaacaaagtt ctccctctct tctcccccac tcatgcaaac ttctacattt

31621 ctcttagagt atacatacta aaacagtatt ccaaacaatc taacttatta tttcaagaaa

31681 ctggacaata atttaaaact agtttcgata cttaatatat gttgttccat atgaataatc

31741 caaagttcta agtagtcaat aactatctaa atcaaaattc tatttctttt catccataaa

31801 gcaaacaaca aaaaaactaa acaataatat aaaattagtt tcgatattta acacattata

31861 tatatgtttt tacattaata ttactaagtt tccaattacc aattatgtaa acaaatcctt

31921 ttattatctc ttcaccaaca tttctgtttc caaatacatc aaaacttaat tcaaaaaaaa

31981 aaataaagtt aaattatctc tatttatgaa attgtataat aatttgaaat tagtttcgta

32041 tgaataaaaa tttattagaa acttatccaa actcaaccat tattttaaaa atttgaataa

32101 taatttgaaa ttagttttga tattcaatgt attttatatt gttccatata aatatcaacc

32161 atttaaacta aacattccag attttatctc acaaagtaaa taacaaaaat taaataatga

32221 tttaaaatta gttttgatat ccaatatatt ttatatgatc tcatgagaat actaattcag

32281 gtctttctat ttttagttgt gcaaagtctt tattttcttt taagattgta tttaaaaaca

32341 aaatcaatcg aatcacttat gaaaaattga ataataattc aagattaatt tcgatatcca

32401 atataattta tactgtttta tacaaataaa aattgttagg attttatgta aatttaatca

32461 ccattctgaa aatttgaata ataatttgaa atcagtttta atatcaatga attttatatt

32521 gttttatacg aatatatgat gttcgatagt gatgttctaa attacttaac tcaaaatttc

32581 tttccatctc agaaagtaaa caataaaaca ccagttatct aaaactaaaa aataatttaa

32641 aattagtttt aatatctatt ggcattttat atcgtttcat ataaatccta tggaagtttt

32701 taaaatagtc ttaaattttt cgaattaaaa gtcaaccata ttgattactt ttcaaaaaaa

32761 gatttaaaag tagtttcgat atccaataca ctctatattg tttcatattg aacactatct

32821 tgggtttcca aatcctaaat aacaattata taaacaaagt tctttctttt ctcttttttt

32881 cattcaaact tctgtatttc tcttagagta atataatata aaactagttt cgatatttaa

32941 tataatatat gttctttcat atgaataatc tcaaagtttt aagactcaat aactatctaa

33001 actaaaattc tatttcattt catccataaa aacaaataac aaaaactaaa taatagtaat

33061 aatttaaaat tagtttcgat atttaacaca ttatatatgc ctccatatga atattataaa

33121 attctcaatt accaattatg taaaataaaa caaagccttt attttccttt atcaaaacct

33181 ctgtttctct tttaaagcaa aaatttatta aatcttaatt caaaaaataa aattaaatta

33241 tttctatttg tgaaattgca taataattta aaattagttt cgatatttaa aaatcaattt

33301 atgttgtttt gtatgaacaa aaaaaattat tagaaattta tccaaattaa atctacttaa

33361 tcattatttc aaaaatttga ataataattt gagattattt taatatccaa tgcatttcat

33421 attatttaat atagataata gtgaagttct taataatcgt ttaaatcaaa aactattctt

33481 cccatcttac aaagcaaata acaaaattaa ataataattt aaaattagtt ttgatatcca

```
33541 acatatttta tatattctca taagaatact atttaggttc tacttttagt tgtgtaaaca
33601 aagtttttat tttcttttaa gattgtatct aaaaaataaa agcaatcgaa ctttttgaaa
33661 aattgaataa tacttcaaga ttattttcca atatccaaca tattttagac tgtttttata
33721 taaataaaaa agattttatg taaattaatt tattattttg aaaatttgaa taataatttg
33781 aaattagttt tgatatctaa tgactatacg aataataatg tccaatagtg atgttctatc
33841 aattatttaa attaaaactt tttaattatc ttagaaagta aacaacaaaa atacctagtt
33901 acctaaaact aaaaataatt tagaattggt tttaatgtct tttataatcg ttttatatga
33961 attataaagt tctaaataat cttaaattct tcgaagctat ttcttttcac ccataaaaca
34021 aataacaaaa aactaggtaa taatttaaaa ttagtttcga tatttaacac attatatatg
34081 ctttcataag aatatattac tcaattgtca atcatgaaat ccctttatct ccctttatcc
34141 aaacattctg tctttctttt aaagaaaaaa atttattaaa ttttaattta aaaaagtaaa
34201 tagtagtgaa gttctaatca attgtttaaa ttaaaaactc cagaattttc cccatcccac
34261 aaaacaaata acaaaaatta aaataataat ttaaaaatta gttttaaata ttcaacatat
34321 ttcatatggt ttcataagaa tttacttagg tctctctctc tctaccttta aaattgtatc
34381 taaaaatcaa aatcactcaa attactttta aaaaaaaata aataagatcc attccgatat
34441 ccaatacaat ctacactgtt ttatataaat aaaaattgtt aagatcttac tagttttgat
34501 atctaatgaa ttttatattg tttcctatga attaatagta aatttccaat agttcttcaa
34561 caatttaatt tagcaagtaa aaaatcaaaa aatatatttt attttg
//
```

1466