The genomic basis of host and vector specificity in non-pathogenic trypanosomatids

Guy R. Oldrieve*, Beatrice Malacart, Javier López-Vidal and Keith R. Matthews

ABSTRACT

Trypanosoma theileri, a non-pathogenic parasite of bovines, has a predicted surface protein architecture that likely aids survival in its mammalian host. Their surface proteins are encoded by genes which account for ~10% of their genome. A non-pathogenic parasite of sheep, Trypanosoma melophagium, is transmitted by the sheep ked and is closely related to T. theileri. To explore host and vector specificity between these species, we sequenced the T. melophagium genome and transcriptome and an annotated draft genome was assembled. T. melophagium was compared to 43 kinetoplastid genomes, including T. theileri. T. melophagium and T. theileri have an AT biased genome, the greatest bias of publicly available trypanosomatids. This trend may result from selection acting to decrease the genomic nucleotide cost. The T. melophagium genome is 6.3Mb smaller than T. theileri and large families of proteins, characteristic of the predicted surface of T. theileri, were found to be absent or greatly reduced in T. melophagium. Instead, T. melophagium has modestly expanded protein families associated with the avoidance of complement-mediated lysis. We propose that the contrasting genomic features of these species is linked to their mode of transmission from their insect vector to their mammalian host.

This article has an associated First Person interview with the first author of the paper.

KEY WORDS: Trypanosoma theileri, Trypanosoma melophagium, Non-pathogenic, Host and vector specificity

INTRODUCTION

Trypanosomatidae are a family of single-celled eukaryotes of the class Kinetoplastea, which are characterised by a specialised mitochondrial genome, the kinetoplast. Trypanosomatidae are monoxenous (single host) or dixenous (two host) species. Dixenous trypanosomatids are obligate parasites of a broad diversity of animals and plants whilst monoxenous species are largely restricted to insects (Podlipaev, 2001). However, the taxonomy of trypanosomatids cannot be distilled into these two broad categories, as many monoxenous species opportunistically infect vertebrates (Kaufer et al., 2017) and some dixenous species have subsequently reverted to a monoxenous lifecycle (Schnaufer et al., 2002).

Expansion into vertebrate hosts gave rise to clades of trypanosomatids which represent medical and veterinary threats. Notably, Trypanosoma cruzi and Leishmania spp. cause important diseases in humans, Chagas disease and Leishmaniasis, respectively. Also, Trypanosoma brucei has been subjected to intense molecular and cytological study as two of its subspecies, T. b. gambiense and T. b. rhodesiense, are causative agents of Human African Trypanosomiasis (HAT) (Cayla et al., 2019). In addition to the medical impact of HAT, other African trypanosome species such as T. b. brucei, Trypanosoma vivax and Trypanosoma congolense cause morbidity and mortality in livestock, constraining agricultural development (Mehlitz and Molynieux, 2019).

The lineage including trypanosomatids diverged from other eukaryotes over a billion years ago (Burki, 2014; Cavalier-Smith, 2010) and possess unique adaptations at the genome level (Maslov et al., 2019). As an example, the T. brucei genome comprises 11 megabase chromosomes (Melville et al., 1998) along with ~5 intermediate chromosomes and ~100 mini-chromosomes (Daniels et al., 2010; Wickstead et al., 2004). Nuclear DNA is highly compact, and genes are organised into co-transcribed units, the primary transcripts of which are trans-spliced and polyadenylated to resolve mature mRNA (Clayton, 2019; Parsons et al., 1984). Their distinctive mitochondrial genome, the kinetoplast, consists of mini- (Kleisen and Borst, 1975; Steinert, 1960) and maxicircles (Borst and Fase-Fowler, 1979; Simpson, 1979) that rely upon RNA editing to generate functional mRNA (Maslov et al., 2019). These features are not always retained across trypanosomatid species.

The divergence in trypanosomatid morphology, transmission, and lifecycle development has facilitated an adaption to a diverse range of hosts and vectors. The different surface protein adaptions of T. brucei, T. cruzi and Leishmania spp. exemplify the variant strategies adopted by these parasites to evade the immune systems of their hosts and vectors. For example, T. brucei is extracellular and proliferates in the blood and tissue of mammals. Its cell surface is encoded by a very extensive repertoire of variant surface glycoproteins (VSGs) (Berriman et al., 2005; Wickstead et al., 2004). VSG variation is essential to sustain long-term infections, during which antigenically distinct VSG types dominate at each peak, facilitating host immune evasion (Borst, 2002; Pats et al., 2004). T. cruzi is an intracellular parasite of wild and domestic mammals. The nuclear genome contains an expanded family of mucin genes that represent up to 6% of the genome (Buscaglia et al., 2006) and, along with trans-sialidases (Nardy et al., 2016), these genes enable sustained infections. Leishmania spp. are intracellular trypanosomatids whose cell surface is covered by a thick layer of glycoconjugates, including families of GP63 major surface proteases (MSPs), also known as leishmanolysins (Yao, 2010).

Studies have largely focused on pathogenic dixenous trypanosomatids, which can hold broad host niches, capable of
infecting multiple mammalian species (Funk et al., 2013). In contrast, non-pathogenic dixenous trypanosomes, such as *Trypanosoma theileri* and *Trypanosoma melophagium*, can be highly specific to their host and vector. These species represent an attractive model to study the basis of host and vector specificity. For the remainder of this manuscript, we refer to species that cause no overt pathogenicity in immunocompetent hosts as non-pathogenic. We recognise that ‘non-pathogenic’ species may still cause slight detriment to their hosts, which is often difficult to detect, especially when prevalence is high in the host population.

*Trypanosoma theileri* is a non-pathogenic bovine parasite which has a reported prevalence of 80% in cattle in the US and Europe when screened via culture-based methods (Farrar and Klei, 1990; Matthews et al., 1979; Mott et al., 2011; Schlafer, 1979). *T. theileri* is transmitted by tabanid flies (Böse and Heister, 1993). It can cause lifelong infections but remains at an extremely low parasitemia in immunocompetent animals, indicating the presence of an effective host immune evasion mechanism or strict self-imposed population control, which prevents overt disease (Doherty et al., 1993; Seif, 1995). Experimentally, *T. theileri* can sustain an infection for at least 12 weeks (Mott et al., 2011) which, combined with their non-pathogenic nature, has stimulated development of *T. theileri* as a potential vaccine delivery vehicle (Mott et al., 2011). The genome of *T. theileri* encodes five predicted surface protein families. These are four unique protein families, *T. theileri* putative surface proteins (TTPSPs) and one MSP family (Kelly et al., 2017). Together these genes represent ~9% of the genome of *T. theileri,* comparable to the representation of the VSG gene family in *T. brucei brucei* (TREU927/4) (Berriman et al., 2005). Lastly, the trans-sialidases that characterise *T. cruzi* were also found to be highly expressed in *T. theileri.* These findings led to the suggestion of a novel immune evasion mechanism in *T. theileri,* contrasting with the well-known system in African trypanosomes and distinct from that in *T. cruzi* or *Leishmania* (Kelly et al., 2017). *T. theileri* is at the base of a clade that comprises *Trypanosoma rangeli,* *Trypanosoma cruzi* and *Trypanosoma grati* (Kelly et al., 2014, 2017), distinct from African trypanosomes, such as *T. brucei*. *T. grati* is transmitted via the tsetse fly between African crocodilians via ingestion of tsetse faeces containing infective metacyclic forms (Hoare, 1929, 1931).

*Trypanosoma melophagium* is a non-pathogenic trypansom of the subgenus *megatrypanum,* transmitted between sheep via the sheep ked. This flightless insect vector has been eradicated from much of its original geographic distribution due to widespread pesticide use. However, where the sheep ked persists, it often carries *T. melophagium* (Gibson et al., 2010; Martinkovic et al., 2012). In a study of organic sheep farms, *T. melophagium* was found to be present in 86% of keds, however, blood smears from sheep on the same farms did not detect trypanosomes (Martinkovic et al., 2012). Other surveys via blood culture found 7.8% of sheep to be infected (Gibson et al., 2010; Martinkovic, 2012). Presumably the divergence of *T. theileri* and *T. melophagium* is associated with their discrete host niches (Gibson et al., 2010; Martinkovic et al., 2012). Nonetheless, *T. theileri* and *T. melophagium* undergo a similar transmission cycle where metacyclic forms are produced in the insect hindgut and the infective forms are then believed to be transmitted to their mammalian host via the mouth, by ingestion of insect faeces or the whole insect body. Trypanosomes then invade their mammalian hosts and proliferate in the blood and, potentially, tissues before being taken up as a bloodmeal by their insect vector (Böse and Heister, 1993; Hoare, 1923).

A notable contrast in the biology of these parasites is the divergence in the life history of their vectors. Sheep keds, which transmit *T. melophagium,* spend their entire life attached to either the skin or wool and hair of sheep. Both male and female keds feed on their mammalian host (Underwood et al., 2015). Tabanids, which transmit *T. theileri,* breed and lay their eggs in soil, water, or trees. The larvae and pupae stages live on vegetation and soil. Only female adults feed on mammalian blood, which is essential for egg production. Although adult tabanids show considerable adaptation to blood feeding, they also feed on the sugars of plants (Chainey, 1993).

Here we derive the *T. melophagium* genome using a combination of long and short read technologies. The genome, and its protein encoding genes, was compared to *T. theileri,* to provide insight into the biological specificity exhibited by each parasite in the context of their close phylogenetic relationship.

**RESULTS**

**T. melophagium** genome assembly

An initial assessment of the *T. melophagium* genome, via k-mer counting, predicted that it was smaller than that of *T. theileri* (22.3 Mb and 27.6 Mb, respectively), this variation being observed in its repeat and unique sequence (Table 1). Notably, both genomes are predicted to have extremely low heterozygosity (0.3 and 0.4 for *T. melophagium* and *T. theileri*, respectively) in comparison to other *Trypanosoma* isolates (Oldrieve et al., 2021). Large gene families, such as the TTPSPs, only account for ~10% of the *T. theileri* genome and so are predicted to have a minor effect in the heterozygosity calculation. The k-mer counting prediction was similar in size to the final assembly (Table 1). The *T. melophagium* assembly consisted of 64 contigs in comparison to 253 for *T. theileri*. BUSCO assessments predict that both assemblies are 100% complete, although *T. theileri* is slightly fragmented (Table 1).

The *T. melophagium* and *T. theileri* genomes were aligned to highlight the conservation of collinearity. The conservation is more similar to the conservation between the species *T. brucei* TREU927/4 and *T. congolense* IL3000 2019 than between isolates of the same species, *T. brucei* TREU927/4 and *T. brucei* Lister 427 2018 (Fig. 1, Fig. S3). The *T. melophagium* genome was annotated with 10,057 protein encoding genes, in comparison to 11,312 in *T. theileri* (Fig. 2C, Table 1) (Kelly et al., 2017).

Kinetoplastid genomes, proteomes and transcriptomes were downloaded from TriTrypDB (Aslett et al., 2010). Proteome completeness was assessed with BUSCO. Those above 85% complete were retained, leaving 44 isolates from 9 genus (File S1). *T. theileri* and *T. melophagium* have some of the smallest genomes in this study and have the lowest GC content (40.1% and 41.2%, respectively), contrasting with the kinetoplastid mean of 48.4% (Fig. 2A).

Environmental temperature (Lao and Forsdyke, 2000; Paz et al., 2004), generation time (Shah and Gilchrist, 2011), neutral drift (Eyrewalker, 1991; Rao et al., 2011), tRNAs (Plotkin et al., 2004), translational accuracy/efficiency (Akashi, 1994; Hu et al., 2013;
Shah and Gilchrist, 2011; Sorensen et al., 1989), gene splicing and protein folding (Novoa and de Pouplana, 2012) have all been hypothesised to cause codon bias. However, codon bias can be influenced by nutrient availability (Seward and Kelly, 2016). Species with low nitrogen availability, such as the plant trypanosomatid, *Phytomonas*, have an AT rich genome, potentially to mitigate the lack of nitrogen in their plant hosts (Seward and Kelly, 2016). Selection acting on genome nucleotide cost (Sc) is in competition with selection acting to alter the translational efficiency of the genome (St) (Seward and Kelly, 2016, 2018). Alternative hypotheses for the cause of codon bias can be excluded by analysing closely related species with similar lifestyles (Seward and Kelly, 2016). Based on universal single copy orthologues, selection pressure acting on the translational efficiency is minimal for both *T. theileri* and *T. melophagium* (Fig. 2B). In contrast, they show the greatest predicted selection pressure acting to reduce the nucleotide cost of any kinetoplastid genome, including the closely related *T. grayi*. The AT biased content can be interpreted as a remodelling of the *T. theileri* and *T. melophagium* genomes to reduce the cost of the genome (Fig. 2A,B). This pattern was consistent when every coding sequence (CDS) was compared (Fig. S4A,B) and in universal single copy orthologs which are essential for every life-cycle stage of the assemblies used in this assessment represent the final draft of the *T. melophagium* assembly produced during this study and the *T. theileri* assembly available from TriTrypDB (Aslett et al., 2010; Kelly et al., 2017). The k-mer spectra plots associated with the first section of the table are found in Fig. S1.

Table 1. Genome assessment

|                      | *T. melophagium* | *T. theileri* (32) |
|----------------------|------------------|-------------------|
| k-mer based genome survey |                  |                   |
| Heterozygosity (%)   | 0.30             | 0.41              |
| Genome length (Mbp)  | 22.28 / 4.16 / 18.13 | 27.62 / 7.97 / 19.65 |
| Genome assembly statistics (for scaffolds longer than 200 bp) |                  |                   |
| Number of contigs    | 64               | 253               |
| Length (Mbp)         | 23.3             | 29.6              |
| Minimum / maximum / mean (bp) | 5,186 / 1,230,212 / 364,120 | 791 / 1,635,300 / 117,005 |
| GC (%)               | 41.2             | 40.1              |
| Genome BUSCO assessment |                |                   |
| Complete/ single copy/ duplicated/ fragmented | 100 / 100 / 0 / 0 | 99.2 / 99.2 / 0 / 0.8 |
| Annotation           |                  |                   |
| Number of genes      | 10,057           | 11,312            |
| Annotation BUSCO assessment |        |                   |
| Complete/ single copy/ duplicated/ fragmented | 100 / 100 / 0 / 0 | 99.2 / 99.2 / 0 / 0.8 |

The assemblies used in this assessment represent the final draft of the *T. melophagium* assembly produced during this study and the *T. theileri* assembly available from TriTrypDB (Aslett et al., 2010; Kelly et al., 2017). The k-mer spectra plots associated with the first section of the table are found in Fig. S1.

Fig. 1. Synteny of the *T. melophagium* and *T. theileri* genome sequences highlights conservation and identity between the two species. The legend refers to the percentage identity between the sequences.
T. brucei (Fig. S4C,D). T. rangeli displays the highest level of selection pressure acting to increase translational efficiency (Fig. 2B), potentially linked to its reduced genome length (Fig. 2C).

Orthologous protein clustering and phylogenetic inference
Orthologous clustering identified genes that descended from a gene in the last common ancestor of the 44 kinetoplastid proteomes used in this study (Fig. 2, File S1). From the 44 proteomes, 18,274 orthogroups were identified (96.5% of the proteins used in this study were included in one of these orthogroups), 992 orthogroups were single copy and contained all isolates.

A species tree was generated as part of the orthologous protein clustering. Using genetic markers, previous studies have noted the similarity between T. melophagium and T. theileri (Gibson et al., 2010; Martinˇkovi´c et al., 2012). Based on 2,312 gene trees, T. theileri is the closest isolate to T. melophagium and groups with the stercorarian trypanosomes, which include T. cruzi, T. grayi and T. rangeli, rather than with salivarian trypanosomes such as T. brucei (Fig. 3). T. grayi is the closest isolate to the T. melophagium and T. theileri clade and is closer in size to the genome length of T. melophagium than to T. theileri (Fig. 2C). T. melophagium and T. theileri are closely related species with relatively short branch lengths (0.093 and 0.088 substitutions per site, respectively). In comparison, T. congolense is more divergent from T. brucei (0.249) whilst the isolates within T. brucei (T. brucei brucei Lister 427 2018: 0.0009, T. brucei evansi STIB805: 0.003, T. brucei brucei TREU927/4: 0.003 and T. brucei gambiense DAL972: 0.004) show less divergence than seen between T. theileri and T. melophagium.

T. theileri has a greater number of species-specific orthogroups than T. melophagium and 12.9% of its genes are assigned to one of these orthogroups, while T. melophagium has only 2.7% of its genes in a species specific orthogroup (Table 2). Terminal branch length is correlated with specific orthogroup counts, which could account for the discrepancy. However, T. melophagium and T. theileri are each other’s most recent common ancestor and have been evolving at a roughly similar rate since this time, with similar terminal branch lengths (Fig. 3). To visualise these differences, the number of genes in each orthogroup was compared between T. melophagium and T. theileri. Orthogroups associated with host interaction protein families were highlighted based on their identification as a putative

Fig. 2. Kinetoplastid genome comparison. (A) GC content across the whole genome and GC content bias in the CDS of kinetoplastid universal single copy orthologues (n=992). GC content (GC)>0=GC content bias. GC <0=AT content bias. (B) Selection acting on translational efficiency (St) and selection acting on nucleotide cost (Sc) in kinetoplastid universal single copy orthologues. Sc >0=Selection acting to increase codon nucleotide cost. Sc <0=Selection is acting to decrease codon nucleotide cost. St >0=Selection is acting to increase codon translational efficiency. St <0=Selection acting to decrease codon translational efficiency. (C) Counts of annotated protein sequences of publicly available kinetoplastids compared by genome size.
cell surface protein by Kelly et al. (2017). Many of the *T. theileri* species specific orthogroups expansions belong to a cell surface family (Fig. 4A).

**Interaction with the mammalian host and predicted cell surface proteins**

Firstly, and as expected, *T. melophagium* was found to lack any genes in orthogroups which contained VSGs, characteristic of African trypanosomes (File S4). To validate this, a blast search was performed using a relaxed cut off (1e-5) using the *T. melophagium* genome as the database and *T. brucei* TREU927/4 VSGs as the query. No hits were identified.

To enable comparison to the *T. theileri* genome analysis, the genes and orthogroups from this study were annotated with the host interaction genes from Kelly et al. (2017). Across the entire genome, *T. theileri* is predicted to contain 1,265 more genes than *T. melophagium* (Table 2). Examination of orthogroups which were associated with a *T. theileri* putative surface protein (TTPSP) revealed a large expansion in *T. theileri* (1,251 genes) compared to *T. melophagium* (10 genes) which could equate to much of the disparity in genome size (Fig. 4). To confirm the difference in TTPSPs, the *T. theileri* transcripts were subjected to a blastn search against a database consisting of the *T. melophagium* transcripts (1e-25 cut-off). The *T. melophagium* transcripts were derived from the genome annotation analysis. Only nine *T. theileri* TTPSPs aligned to *T. melophagium*.

TTPSPs were split between four orthogroups in the original *T. theileri* genome analysis but were split between 66 orthogroups in this analysis (Table 3) (Kelly et al., 2017). TTPSPs share conserved
C terminal GPI addition and N terminal signal sequences and contain regions of high divergence in the remainder of the sequence. TTPSPs are highly expressed as a family at the population level and are largely contained within tandem arrays, highlighting a similarity to the VSGs of *T. brucei* (Kelly et al., 2017). Excluding *T. melophagium*, TTPSPs are absent from all other kinetoplastid species analysed (File S4) revealing their specific innovation in these related trypanosomatids.

Following the TTPSPs, the largest expanded gene family in *T. theileri* are the MSPs. This protein family belongs to the peptidase M8 family of metalloproteinases. MSPs are likely to have contrasting roles in different life stages but their best understood function is to bind and cleave members of the complement system and for evasion of other cellular and antimicrobial immune defences (Yao, 2010). This role presumably allows for evasion of complement-mediated proteolysis and therefore assists survival in the insect and mammalian hosts (Yao, 2010). Orthogroups which annotated as MSPs were expanded in *T. melophagium*, such as OG0008865. *T. melophagium* also has a species specific orthogroup (OG0009903) consisting of 11 proteins and annotated as MSP. Combined, these results suggest that the surface protein environment of *T. melophagium* and *T. theileri* are distinct and that genes encoding these proteins account for most of the discrepancy in the genome sizes.

**Clade and species specific orthogroups**
*T. grayi, T. melophagium* and *T. theileri* contain 42 clade-specific orthogroups including the MSP orthogroups OG0000590 and

| Table 3. Cell surface orthogroup counts from Kelly et al. (2017) (32) and this study along with counts of genes present in each category |
|-----------------|-----------------|-----------------|-----------------|
| **Cell surface conservation** | **Annotation** | **Orthogroup count** | **Gene count** |
| | | **Kelly et al., 2017** | **This study** | **T. theileri** | **T. melophagium** |
| Conserved | Amastin | 1 | 2 | 7 | 4 |
| Conserved | MASP | 1 | 1 | 1 | 1 |
| Conserved | MSP | 1 | 18 | 229 | 52 |
| Conserved | PSSA-2 | 1 | 1 | 5 | 3 |
| Unique | TTPSP1 | 1 | 42 | 720 | 1 |
| Unique | TTPSP2 | 1 | 10 | 301 | 0 |
| Unique | TTPSP3 | 1 | 11 | 157 | 9 |
| Unique | TTPSP4 | 1 | 3 | 73 | 0 |
OG0008095. Both MSP orthogroups were expanded in *T. theileri* in comparison to *T. grayi* and *T. melophagium*. *T. melophagium* and *T. theileri* share 81 orthogroups specific to the two species. Twelve of the orthogroups are putative cell surface protein families, including five representatives from TTPSP and seven MSP of the orthogroups are putative cell surface protein families, *T. melophagium* has seven orthogroups associated with protein binding (OG0011868) and a repeat family associated with protein binding (OG0011868) and a repeat family associated with protein binding (OG0018131) and a leucine-rich repeat family (OG0018098). Other annotated orthogroups remained, which included a leucine-rich repeat family associated with protein binding (OG0018131) and a leucine-rich repeat family (OG0018098).

**Cell surface modifying enzymes**

Trans-sialidases are differentially expanded in *T. theileri* and *T. melophagium*, with 38 and eight genes, respectively. Of the four orthogroups containing trans-sialidase, two do not contain proteins from *T. melophagium* (OG0011818 and OG0015016). These two orthogroups contain proteins from *T. cruzi* and are likely to have been lost by *T. melophagium* and *T. grayi*. *T. cruzi*, trans-sialidases are involved in host immune evasion (Nardy et al., 2016).

Invertases are typical of the cell surface of *Leishmania* and are thought to transform sucrose into hexose in the gut of the vector. The orthogroups (OG0000150 and OG0000409) that include 20 *T. theileri* invertase genes only have three members in *T. melophagium* and three members in *T. grayi*. This expansion might indicate an adaptation of *T. theileri* to its vector, the tabanid fly, which can feed on sugary flower nectar (Kniepert, 1980). In contrast, the *T. melophagium* vector, the sheep ked, exclusively feeds on mammalian blood. The orthogroups which contain the *T. theileri* invertase genes (OG0000150 and OG0000409) contain only one gene from the plant parasite *Phytoponas* (Jaskowska et al., 2015; Sanchez-Moreno et al., 1992), suggesting a different mechanism for sucrose metabolism in these parasites.

Other putative cell surface modifying molecules, such as UDP-galactose/UDP-N-acetylglucosamine transferases (OG0000001) were expanded in *T. theileri* (n=60) compared to *T. melophagium* (n=11).

**Glycolysis**

Kelly et al. (2017) compared *T. brucei* and *T. theileri* transcriptomes which revealed differences in the abundance of glycolosomal enzyme mRNAs. Particularly, pyruvate orthophosphate dikinase, phosphoenolpyruvate carboxykinase and malate dehydrogenase were found to be >10 fold more abundant in *T. theileri* than in *T. brucei* (Kelly et al., 2017). We confirm that enzymes associated with the glycolytic pathway are present in *T. melophagium* (File S4) and found that *T. melophagium* has expanded the orthogroups associated with three glycolytic enzymes. These included pyruvate orthophosphate dikinase (OG0000570), phosphoenolpyruvate carboxykinase (OG0000120) and malate dehydrogenase (OG0000332), whilst *T. melophagium* has a reduced number of genes in the fumarate reductase orthogroup (OG0000078).

Orthogroups associated with peroxisome targeting were found to be in equal numbers (OG0004108 PEX5 and OG0003998 – PEX7). All other glycolytic enzymes are present in equal numbers in the two species. Therefore, it is likely that the glycolysis pathway is conserved in *T. melophagium*.

**Life cycle**

Extensive studies of *T. brucei* have identified genes which are associated with key stages of the *T. brucei* life cycle. These studies tracked genes associated with stumpy formation in the blood stream form (Cayla et al., 2020; Liu et al., 2011; Lueong et al., 2020; Mony et al., 2014) and regulators of metacyclogenesis (Toh et al., 2021). These genes were combined with a list of validated development associated genes such as the RNA binding proteins RBP6, RBP7, RBP10 and ZFP2 and ZFP3 along with developmental regulators NRK A, NRK B, RDK1, RDK2, MAPK2 and phosphatases such as PTP1 and PIP39 (Domingo-Sananes et al., 2015; Gale and Parsons, 1993; Jones et al., 2014 a; Müller et al., 2002; Szöör et al., 2010, 2006; Walrad et al., 2009). Most of the orthogroups containing these genes were represented with a similar number of genes in each species, indicating the presence of an environmental sensing ability and developmental competence (Fig. 5). However, there were notable differences. There is an expansion in the orthogroups containing KRIP14, which is a mitochondrial SSU component (Mony et al., 2014), in *T. theileri*. *T. melophagium* has expanded its orthogroups containing the kinases NRK (Domingo-Sananes et al., 2015; Gale et al., 1994), NEK and (Gale and Parsons, 1993) ADKF (Mony et al., 2014) along with a dual specificity phosphatase (DsPho) and protein phosphatases 1 (PP1) (Mony et al., 2014; Mony and Matthews, 2015). Both *T. theileri* and *T. melophagium* are missing metacaspase (MCA1) which is associated with the later stages of progression towards metacyclic forms in *T. brucei* (Toh et al., 2021) and Hyp12, which upregulates bound mRNAs during development based on tethering assays in *T. brucei* (Erben et al., 2014; Lueong et al., 2016; Mony et al., 2014; Mony and Matthews, 2015). Puf11, an effector molecule required for kinetoplast repositioning in epimastigotes (Toh et al., 2021), is also absent in *T. melophagium*.

Life-cycle regulatory genes and genes controlling meiosis SP011, MN1, HOP1 and DMC1, along with the cell fusion protein HAP2/GCS1 (Peacock et al., 2021), were found in *T. melophagium* and *T. theileri* (Fig. 5), suggesting maintenance of a sexual stage.

**RNA interference and transposable elements**

All five core genes that represent the trypanosome RNAi machinery (AGO1, DCL1, DCL2, RIF4 and RIF5) were present in *T. melophagium*, with an extra gene in the orthogroup containing DLC2 (File S4). Therefore, a functional gene silencing pathway is likely to be present in *T. melophagium*, matching the prediction in *T. theileri*.

Retrotransposon counts highlighted an expansion in *T. cruzi* and *T. vivax* isolates, along with *T. brucei* Lister 427 2018 (Fig. S5A, File S1). *T. melophagium* and *T. theileri* have not expanded their retrotransposon repertoire. A similar pattern was observed for long terminal repeat (LTR) retrotransposon counts, which show a positive correlation with genome size (Fig. S5B).

**DISCUSSION**

*T. melophagium* and *T. theileri* are closely related trypanosomes that have distinct hosts and vectors (Gibson et al., 2010; Martiniković et al., 2012). Here, the genome of *T. melophagium* was sequenced...
quantified in orthogroups associated with developmental regulation have been sequenced technologies. Comparing assemblies obtained using the tools used are specific for the sequencing technologies whilst there has been substantial development in assembly methods between the two studies. However, based on the convergence of the k-mer counting based prediction with the assembly sizes, along with 100% complete BUSCO scores, we were reassured by the quality of the draft assemblies and the subsequent comparisons of their genome content.

Using k-mer counting based predictions, *T. melophagium* was anticipated to have a smaller genome than *T. theileri*. This held true when the data were assembled (Table 1). The *T. melophagium* genome is more similar in size to *T. grayi*, the closest relative to *T. melophagium* and *T. theileri*, than to *T. theileri* (Figs 2 and 3). *T. theileri* has likely expanded its genome size since speciation occurred. A peculiarity of the *T. theileri* and *T. melophagium* isolates analysed in this study is their highly reduced heterozygosity, in contrast to African trypanosomes (Oldrieve et al., 2021). Whilst only one genome is available for both species, should these isolates represent the species as a whole, the reduced heterozygosity could be linked to a founder effect (Pool and Nielsen, 2007). As *T. theileri* and *T. melophagium* have specific host and vector niches, the small population that initially expanded into the niches possibly underwent a significant population bottleneck, especially as host domestication caused eradication of wild progenitors and wild relatives, which could have facilitated a reduction in heterozygosity, induced by genetic drift. Alternatively, the absence of a sexual cycle could contribute to the reduced heterozygosity. Although *T. melophagium* and *T. theileri* contain genes required for meiosis, this does not confirm the species undergo sexual reproduction.

Selection appears to be acting to reduce the genome wide nucleotide biosynthesis cost in both *T. theileri* and *T. melophagium* (Fig. 2B) which has remodelled their genomes toward an AT bias, contrasting with all other kinetoplastid genomes analysed in this study (Fig. 2A,C). The predicted selection pressure acting to reduce nucleotide cost is at the expense of translational efficiency (Fig. 2C) and is greater than for the free-living *Bodo saltans*, or monoxenous insect parasites such as the early-branching *Paratrypanosoma confluens* and *Phytomonas EM1*. *Phytomonas* has limited access to nitrogen as it infects nitrogen deficient plants and has been highlighted as an example where diet can cause selection to reduce the species genome nucleotide cost, through a reduction in GC nucleotide biosynthesis cost in both *T. theileri* and *T. melophagium* (Fig. 2B) which has remodelled their genomes toward an AT bias, contrasting with all other kinetoplastid genomes analysed in this study (Fig. 2A,C). The predicted selection pressure acting to reduce nucleotide cost is at the expense of translational efficiency (Fig. 2C) and is greater than for the free-living *Bodo saltans*, or monoxenous insect parasites such as the early-branching *Paratrypanosoma confluens* and *Phytomonas EM1*. *Phytomonas* has limited access to nitrogen as it infects nitrogen deficient plants and has been highlighted as an example where diet can cause selection to reduce the species genome nucleotide cost, through a reduction in GC content (Seward and Kelly, 2016). We propose that the reduction in the selection cost of *T. theileri* and *T. melophagium* may be related to their non-pathogenic nature. By remodelling their genome to an AT bias, they may have reduced their cost to their host, facilitating reduced pathogenicity. Closely related species of bacteria exist on a spectrum from pathogen to symbiont, highlighting how a selective advantage can arise from a parasite reducing the cost to its host (Toft and Andersson, 2010). It is possible that *T. melophagium* and *T. theileri* are part of a similar spectrum amongst trypanosomatids.

We acknowledge that alternative hypotheses exist for the reduced nucleotide cost associated with the AT rich genome, such as tissue niche adaptation in their mammalian host or selection primarily operating within the arthropod vector rather than mammalian host.

This clade-specific genome remodelling provides an example of the similarity between *T. theileri* and *T. melophagium*. However, the species have contrasting hosts, vectors, and genome sizes. Genome annotation and orthology inference identified candidates for their discrepancy in genome size. When species specific orthogroups

and a draft assembly was produced and annotated. The annotated proteome was incorporated into a comparison with *T. theileri*, and other publicly available kinetoplastid proteomes, to determine their phylogenetic relationship and to explore the genomic basis of the host and vector specificity of these non-pathogenic trypanosomatids.

Although the two genomes compared in this analysis are predicted to be complete (Table 1), their assembly and annotation were performed 5 years apart using different assembly pipelines and sequencing technologies. Comparing assemblies obtained using the same sequencing technologies and assembly pipelines would allow for greater confidence in the observed variations in their genome content as updated methods and long-read sequencing continue to improve the quality of genome assemblies and, therefore, completeness. This was not possible since many of the assembly tools used are specific for the sequencing technologies whilst there has been substantial development in assembly methods between the two studies. However, based on the convergence of the k-mer counting based prediction with the assembly sizes, along with 100% complete BUSCO scores, we were reassured by the quality of the draft assemblies and the subsequent comparisons of their genome content.

Fig. 5. Genes associated with development, and related proteins, at various stages throughout the *T. brucei* life cycle. The number of genes in orthogroups associated with developmental regulation have been quantified in *T. theileri* and *T. melophagium*. More genes are seen to be associated with development in *T. theileri* compared to *T. melophagium*.
were compared, the greatest contrast was between orthogroups associated with the putative cell surface, with the largest expansions detected in *T. theileri* being of TTPSP and MSP surface protein families (Fig. 4). Although both species undergo a cyclical transmission cycle, which includes mammalian and insect stages, we hypothesise the respective prevalence in their mammalian hosts, and the contrasting life history of their respective vectors could explain the genome expansion in *T. theileri*. *T. theileri*, spread by tabanids, are found in over 80% of livestock (Farrar and Klei, 1990; Matthews et al., 1979; Mott et al., 2011; Schlafer, 1979). In comparison, *T. melophagium* exhibits lower detected prevalence, being rarely identified in its mammalian host via blood smears (Martinković et al., 2012) or after blood culture (Serpel and Zafer, 2008). Moreover, sheep keds, which transmit *T. melophagium*, are intimately associated with their mammalian host, spending their entire life either on the sheep’s skin or wool. Here, males and females feed solely on mammalian blood providing many opportunities for transmission of *T. melophagium* from the sheep to the sheep ked (Hoare, 1923). Therefore, there is potentially less advantage for *T. melophagium* to invest in mammalian immune evasion mechanisms required to extend the length of its infection in sheep, since it has many transmission opportunities. The limited investment in *T. melophagium* is emphasised by their relatively unsophisticated putative TTPSP-related repertoire alongside modestly expanded species-specific MSP families (Fig. 4). Instead, *T. melophagium* could rely on its ancestral ability to sustain infections in invertebrate hosts, which, although able to be primed to defend against a specific pathogen, rely upon an innate immune response (Cooper and Eleftherianos, 2017).

In contrast, *T. theileri* has a transient host-vector interaction. Tabanid flies of either gender survive on plant sugars, while adult females occasionally feed on mammalian blood (Chainey, 1993). Therefore, potentially *T. theileri* requires extended survival in its mammalian host to sustain transmission between cattle, compared to the intimate long-term association of sheep keds with *T. melophagium*. The investment from *T. theileri* in an expanded surface protein repertoire is likely to support adaptive immune evasion and prolonged survival in the mammalian stage of its life cycle. Alternatively, or additionally, differences between the bovine and ovine immune responses could contribute (Wang et al., 2013). It should be noted that both *T. melophagium* and *T. theileri* prevalence was surveyed via blood smear or blood culture. Although this is a standard approach, studies have highlighted the prevalence of *T. brucei* in adipose tissue (Trindade et al., 2016) and we cannot exclude one species preferentially infecting these tissues, rather than the bloodstream.

Many of the gene families identified in *T. theileri* (Kelly et al., 2017) were divided into multiple orthogroups in this study. The discrepancy is likely to be explained by evolution of the methods used by OrthoFinder. At the time of publication of the *T. theileri* study, OrthoFinder v.1 was available, while our analysis used version v.2.5. For instance, OrthoFinder v.2.5 uses updated sequence alignment tools, such as DIAMOND ultra-sensitive. For this reason, we can speculate that the clustering in this study is more refined, such that the TTPSP families should be divided into smaller protein families. However, large paralogous orthogroups remain the toughest challenge for orthogroup clustering software and so the relationships between this set of proteins will likely continue to evolve alongside the software (Emms and Kelly, 2020).

Genes involved in the trypanosome life cycle, cellular quiescence and meiosis were all detected in *T. melophagium*, suggesting a competent developmental cycle along with the machinery for sexual recombination. There is an expansion of the *T. theileri* invertase orthogroup which was not present in *T. melophagium*. This is potentially associated with the use of sucrose in the tabanid fly’s diet. Although the glycolysis pathway is present in *T. melophagium*, there was an expansion in the pyruvate orthophosphate dikinase, phosphoenolpyruvate carboxykinase and malate dehydrogenase orthogroups. These genes are associated with the branch of the glycolytic pathway that converts pyruvate to succinate to facilitate the recovery of NAD⁺ (Kelly et al., 2017). This branch of the glycolytic pathway was upregulated in *T. theileri* in contrast to *T. brucei* (Kelly et al., 2017). Interestingly, the core RNAi genes were detected in *T. melophagium*, consistent with *T. theileri* but distinct from *T. cruzi*, which is a stercorarian trypanosome, but which lacks the requisite molecular machinery (Ulfu et al., 2004).

In summary, we have found that *T. theileri* and *T. melophagium* are closely related species that display substantial remodelling of their genomes to facilitate a reduction in their nucleotide costs, which might reduce the costs they impose on their hosts. *T. theileri* displays a considerable genome expansion, which is associated with a large repertoire of unique proteins that characterise its cell-surface and host-interaction gene repertoire. These genes could facilitate a lifelong infection in its mammalian host. In contrast, the comparatively unsophisticated immune evasion repertoire displayed by *T. melophagium* suggests more limited adaptation to its mammalian host.

### MATERIALS AND METHODS

#### Trypanosome culture, DNA/RNA extraction and sequencing

The full list of tools used in this study, and the options used to run those tools, can be found in Table S1.

*T. melophagium* was isolated from sheep blood collected on the island of St Kilda, Scotland, UK (kindly provided by Professor Josephine Pemberton, University of Edinburgh). Whole blood was collected into heparinized vacutainers and used within 2 days. 1 ml of blood was diluted with 5 volumes of a 50% mix of HMI9 supplemented with 20% fetal bovine serum (FBS) and Madin-Darby bovine kidney (MDBK) conditioned medium. All cultures were kept at 37°C and were examined microscopically every 3 days for 6 weeks. After propagation of *T. melophagium* by culturing of the blood sample, the specimens were transferred and co-cultured with fibroblast-like primary cells as feeder cells, isolated from the same blood sample. Due to the short lifespan of these primary cultured cells, *T. melophagium* was subsequently co-cultivated with MDBK cells and then progressively adapted to axenic conditions with a 50% mix of HMI9 and MDBK conditioned medium.

DNA was extracted from cultured *T. melophagium* using a MagAttract high molecular weight DNA kit, following the manufacturer’s instructions (Qiagen) and cleaned via ethanol precipitation. The DNA was sequenced with Oxford Nanopore Technology’s (ONT) MinION (R9.4.1), following the ONT Rapid Sequencing protocol. Base-calling was performed in high accuracy mode using Guppy (available at https://community.nanoporetech.com/) which produced 1.059 gigabases (Gb) of data. PycoQC was used to visualise the data (Leger and Leonardi, 2019). The same DNA was sequenced with BGI’s DNaseq (4.201Gb, 150 base pair reads). RNA was extracted with the RNeasy mini kit (Qiagen) including a DNAse step, and sequenced with BGI’s DNaseq (5.019Gb, 100 bp reads). Raw DNA and RNA DNaseq reads were trimmed with Trimmomatic (Bolger et al., 2014).

*T. theileri* sequencing data was downloaded from NCBI. 170 bp genomic reads were used (SRR13482812).

#### T. melophagium genome assembly and annotation

Jellyfish and GenomeScope were used to provide a k-mer based estimate of the genome size and heterozygosity using the short DNA reads described above (Marçais and Kingsford, 2011; Vurture et al., 2017).
ONT long reads were assembled using Wtdbg2 (Ruan and Li, 2020). For the polishing steps, BWA-MEM (Li, 2013 preprint) was used to align short reads and Minimap2 (Li, 2018) was used to align ONT reads. ONT reads were aligned to the draft assembly. Each contig was labelled as ‘VSG’ in the T. brucei TREU927/4 reference genome were downloaded from TriTrypDB. A blastn search was performed using these VSG sequences as the query and the T. melophagium genome as the database using a loose cut-off (e-value=1e-5). A similar search was performed to confirm the reduced TTPSP counts in T. melophagium. For this, the T. theileri transcripts were used to query a database made from the T. melophagium transcripts (e-value=1e-5).

The cell surface orthologs in this study were annotated with the orthogroups from Kelly et al. (2017) based on the orthogroup membership of genes in the two analyses. Unless stated otherwise, all of the figures in this study were plotted in R (Team, 2019) using ggplot2 (available at http://ggplot2.tidyverse.org) and ggrepel (available at http://github.com/slowkow/ggrepel).

Acknowledgements
We thank Steve Kelly (University of Oxford) for his comments and insights on the manuscript.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualisation: G.O., B.M., J.L.V., K.M.; Methodology: G.O., B.M.; Software: G.O.; Validation: G.O.; Formal analysis: G.O., B.M.; Investigation: G.O., B.M.; Resources: G.O., J.L.V., K.M.; Data curation: G.O.; Writing – original draft: G.O., B.M.; Writing – review & editing: G.O., B.M., J.L.V., K.M.; Visualization: G.O., B.M., K.M.; Supervision: G.O., K.M.; Project administration: G.O., K.M.; Funding acquisition: K.M.

Funding
This research was funded in part by the Wellcome Trust [103740/Z/14/Z and 108905/B/15/Z], the ‘Supporting Evidence Based Interventions’, Bill and Melinda Gates foundation, and a Royal Society GCRF Challenge grant CH160034, each awarded to K.M.; K.M. was supported by a Wellcome Trust Investigator award grant [103740/Z/14/Z]; G.O. by a Wellcome Trust PhD studentship [108905/B/15/Z]; and J.L.V. by a Royal Society GCRF Challenge grant [CH160034]. Open Access funding provided by University of Edinburgh. Deposited in PMC for immediate release.

Data availability
T. melophagium DNA and RNA sequencing data, along with the draft genome assembly and its annotation, can be found under the NCBI BioProject PRJNA786535. The T. melophagium genome analysed in this study refers to version GCA_022059095.1.

References
Akashi, H. (1994). Synonymous codon usage in drosophila-melanogaster - natural-selection and translational accuracy. Genetics 136, 927-935. doi:10.1093/genetics/136.3.927
Alsford, S., Turner, D. J., Obado, S. O., Sanchez-Flores, A., Glover, L., Berriman, M., Herz-Fowler, C. and Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res. 21, 915-924. doi:10.1101/gr.115089.110
Alslett, A., Arnebrant, J., Berriman, M., Bredest, J., Brunk, B. P., Carrington, M., Deplege, D. P., Fischer, S., Gajria, B., Gao, X. et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 38, D457-D462. doi:10.1093/nar/gkp851
Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P. and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691-1692. doi:10.1093/bioinformatics/btr473
Berriman, M., Ghanem, E., Herz-Fowler, C., Blandin, G., Renaud, H., Bartholomew, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B. et al. (2005). The genome of the African trypanosome Trypanosoma brucei. Science 309, 416-422. doi:10.1126/science.1112642

Orthology inference
Orthologous proteins from 44 kinetoplastid proteomes were identified with OrthoFinder (Emms and Kelly, 2019) and protein clusters were summarised with KinFin (Laetsch and Blaxter, 2017a), using InterProScan annotations based on the Pfam and signalP databases (Jones et al., 2014b; Mistry et al., 2021; Nielsen, 2017). A minimal cut-off threshold was not applied to the orthogroup annotation. The orthogroup annotation summary can be found in File S2. A species tree was produced by STAG and STRIDE, as part of the OrthoFinder analysis (Emms and Kelly, 2017, 2018), which was visualised with iTOL (Letunic and Bork, 2007). STRIDE identified Bodo saltans as the best root for the consensus species tree.

To confirm the presence of VSGs in T. melophagium, all CDS sequences labelled as ‘VSG’ in the T. brucei TREU927/4 reference genome were downloaded from TriTrypDB. A blastn search was performed using these VSG sequences as the query and the T. melophagium genome as the database using a loose cut-off (e-value=1e-5). A similar search was performed to confirm the reduced TTPSP counts in T. melophagium. For this, the T. theileri transcripts were used to query a database made from the T. melophagium transcripts (e-value=1e-5).

The cell surface orthologs in this study were annotated with the orthogroups from Kelly et al. (2017) based on the orthogroup membership of genes in the two analyses. Unless stated otherwise, all of the figures in this study were plotted in R (Team, 2019) using ggplot2 (available at http://ggplot2.tidyverse.org) and ggrepel (available at http://github.com/slowkow/ggrepel).
Laetsch, D. R. and Blaxter, M. L. (2017b). BioTools: interrogation of genome assemblies. F1000Research 6, 1287. doi:10.26384/f1000research.12232
Lao, P. J. and Forsdyke, D. R. (2000). Thermophilic bacteria strictly obey Szybalski’s transcription direction rule and politely purine-load RNAs with both adenine and guanine. Genome Res. 10, 228-236. doi:10.1101/gr.6729
Leger, A. and Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. J. Open Source Soft. 4, 1236. doi:10.21105/joss.01236
Leinonen, R., Sugawara, H. and Shumway, M. and International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic Acids Res. 39, D19-D21. doi:10.1093/nar/gkq1019
Letunic, I. and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogeny tree display and annotation. Bioinformatics 23, 127-128. doi:10.1093/bioinformatics/btl529
Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.1303.3997.
Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100. doi:10.1093/bioinformatics/bty191
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. and Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078-2079. doi:10.1093/bioinformatics/btp352
Ling, A. S., Trotter, J. R. and Hendriks, E. F. (2011). A zinc finger protein, TzbCZH20, stabilizes two developmentally regulated mRNAs in trypanosomes. J. Biol. Chem. 286, 20152-20162. doi:10.1074/jbc.M111.139261
Liu, B., Kamanyi Marucha, K. and Clayton, C. (2020). The zinc finger proteins ZC3H20 and ZC3H21 stabilise mRNAs encoding membrane proteins and mitochondrial proteins in insect-form Trypanosoma brucei. Mol. Microbiol. 113, 430-451. doi:10.1111/mmi.14429
Lomsadze, A., Burns, P. D. and Borodovsky, M. (2010). RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Bioinformatics 26, 1203-1209. doi:10.1093/bioinformatics/btq079
Lao, P. J. and Forsdyke, D. R. (2015). Assembling the components of the mitochondrial proteome of Trypanosoma brucei: facts, challenges and mysteries. Curr. Opin. Microbiol. 37, 369-374. doi:10.1016/j.mib.2004.05.001
Paz, A., Mester, D., Baca, I., Nevo, E. and Korol, A. (2004). Adaptive role of increased frequency of polyuridine tracts in mRNA sequences of thermophiles. Proc. Natl. Acad. Sci. U.S.A. 101, 2961-2966. doi:10.1073/pnas.0305894100
Peacock, L., Kay, C., Farrer, C., Bailey, M., Carrington, M. and Gibson, W. (2021). Sequential production of gametes during meiosis in trypanosomes. Commun. Biol. 4, 555. doi:10.1038/s42003-021-02058-5
Plötkin, J. B., Robins, H. and Levine, A. J. (2004). Tissue-specific codon usage and the expression of human genes. Proc. Natl. Acad. Sci. U.S.A. 101, 12558-12561. doi:10.1073/pnas.0404957101
Podlipaev, S. (2001). The more insect trypanosomatids under study-the more diverse Trypanosomatidae appears. Int. J. Parasitol. 31, 648-652. doi:10.1016/S0020-7519(01)00139-4
Pool, J. E. and Nielsen, R. (2007). Population size changes reshape genetic patterns of diversity. Evolution 61, 3001-3006. doi:10.1111/j.1558-5646.2007.00238.x
Porter, A. (1910). The Structure and Life-History of Crithidia Melophagia (Flu), an Endo-parasite of the Sheep-Ked, Melophagus Ovinus. J. Cell Bio. 218, 189-224. doi:10.1242/jeb.18.218
Rao, Y. S., Wu, G. Z., Wang, Z. F., Chai, X. W., Nie, Q. H. and Zhang, X. Q. (2016). Role of eukaryotic splicing in trypanosome gene expression. Proc. Natl. Acad. Sci. USA 113, 1287. doi:10.1073/pnas.1518786113
Seifi, H. A. (2007). Population size changes reshape genetic patterns of diversity. Evolution 61, 3001-3006. doi:10.1111/j.1558-5646.2007.00238.x

Swingle, L. D. (1911). The relation of crithidia melophagia to the sheep’s blood, with remarks upon the controversy between Dr. Porter and Dr. Woodcock. Trans. Am. Microsc. Soc. 30, 275. doi:10.2307/3221144

Szőr, B., Wilson, J., McElhinney, H., Tabernero, L. and Matthews, K. R. (2006). Protein tyrosine phosphatase TbPTP1: A molecular switch controlling life cycle differentiation in trypanosomes. J. Cell Biol. 175, 293-303. doi:10.1083/jcb.200605090

Szőr, B., Ruberto, I., Burchmore, R. and Matthews, K. R. (2010). A novel phosphatase cascade regulates differentiation in Trypanosoma brucei via a glycosomal signaling pathway. Genes Dev. 24, 1306-1316. doi:10.1101/gad.570310

Team, R. C. (2019). R: a language and environment for statistical computing. R Foundation for Statistical Computing.

Toft, C. and Andersson, S. G. E. (2010). Evolutionary microbial genomics: insights into bacterial host adaptation. Nat. Rev. Genet. 11, 465-475. doi:10.1038/nrg2798

Toh, J. Y., Nkouawa, A., Sánchez, S. R., Shi, H., Kolev, N. G. and Tschudi, C. (2021). Identification of positive and negative regulators in the stepwise developmental progression towards infectivity in Trypanosoma brucei. Sci. Rep. 11, 5755. doi:10.1038/s41598-021-85225-2

Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Young, S. A., Pinto, A., Van den Abbeele, J. et al. (2016). Trypanosoma brucei parasites occupy and functionally adapt to the adipose tissue in mice. Cell Host Microbe 19, 837-848. doi:10.1016/j.chom.2016.05.002

Ullu, E., Tschudi, C. and Chakraborty, T. (2004). RNA interference in protozoan parasites. Cell Microbiol. 6, 509-519. doi:10.1111/j.1462-5822.2004.00399.x

Underwood, W. J., Blauwikel, R., Delano, M. L., Gillesby, R., Mischler, S. A. and Schoell, A. (2015). Biology and diseases of ruminants (sheep, goats, and cattle). In Laboratory Animal Medicine, pp. 623-694. Elsevier.

Vaser, R., Sović, I., Nagarajan, N. and Sikic, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737-746. doi:10.1101/gr.214270.116

Vurtle, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J. and Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202-2204. doi:10.1093/bioinformatics/btx153

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

Walrad, P., Paterou, A., Acosta-Serrano, A. and Matthews, K. R. (2009). Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. PLoS Pathog. 5, e1000317. doi:10.1371/journal.ppat.1000317

Wang, F., Ekiert, D. C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., Torkamani, A., Raudsepp, T., Mwangi, W., Criscitiello, M. F. et al. (2013). Reshaping antibody diversity. Cell 153, 1379-1393. doi:10.1016/j.cell.2013.04.049

Wickstead, B., Ersfeld, K. and Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome Res. 14, 1014-1024. doi:10.1101/gr.2227704

Yao, C. Q. (2010). Major surface protease of trypanosomatids: one size fits all? Infect. Immun. 78, 22-31. doi:10.1128/IAI.00776-09