THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use: This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Developmental Incompetence in Selected and Naturally Occurring *Trypanosoma* Isolates

THE UNIVERSITY *of* EDINBURGH

Guy R. Oldrieve

Submitted for the degree of Doctor of Philosophy, Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh

2023

Declaration

I declare that this thesis was composed by myself and that this work has not been submitted for any other degree or professional qualification. The work presented in the thesis (including work that has been published elsewhere) is my own, except where indicated.

Guy Oldrieve

Lay summary

Trypanosoma brucei is a single-celled parasite which infects mammals, causing fatal diseases in humans, wild animals and livestock. The parasite is typically transmitted by a tsetse fly, requiring the parasite to undergo complex development to adapt to life in mammal and insect hosts. Developmentally competent *T. brucei* are restricted to the range of the tsetse fly in the African tsetse belt. Some *T. brucei* isolates are developmentally incompetent and have foregone tsetse transmission, expanding their geographic range beyond tsetse endemic regions. These isolates are unable to develop into the insect form of the parasite and lack a density control mechanism, which increases their virulence. We highlight multiple evolutionary origins of these isolates and predict the basis of their developmental incompetence, which we validate in the laboratory. Our findings provide insight into possible diagnostic tools to anticipate increased virulence in the field.

Abstract

Trypanosoma brucei has two distinct life stages in its mammalian host. The proliferative ‘slender’ form develops into a cell-cycle arrested ‘stumpy’ form at high parasite density. This transition occurs in a density-dependent quorum sensing (QS) like process, for which critical molecular regulators have been identified. Naturally occurring *T. brucei* subspecies (*T. b. evansi* and *T. b. equiperdum*) have reduced ability to generate the stumpy form and are described as ‘monomorphic’. Utilising whole genome sequences of 41 naturally occurring monomorphic isolates, we corroborate previous studies in identifying at least four independent monomorphic *T. brucei* clades. Mutations in six genes were then explored for their contribution to monomorphism. The orthologous gene sequences were synthesised and used to replace wild-type alleles, via CRISPR-Cas9, in developmentally competent *T. brucei*. The replacement of two targets with the monomorphic mutant sequence reduced the ability to generate stumpy forms in developmentally competent cells. Furthermore, we identified mutations associated with cell proliferation and motility phenotypes. We also selected monomorphic cell lines from a pleomorphic population and confirmed significant downregulation of transcripts of a developmental regulator, ZC3H20, during the progression to monomorphism. *In vitro* overexpression of ZC3H20 in the selected monomorphic cells restored pleomorphism. Independently selected monomorphic lines generated *in vitro* were also found to show consistently altered regulation of several transcripts, hinting that the initial steps to monomorphism may share similarities in discrete populations. We suggest that, in the field, monomorphism develops on a spectrum via modifications to the regulation of key QS genes, which can be reversed in the first instance. As the scale tips towards developmental incompetence, mutations accrue in key QS genes which lock the parasites in a monomorphic phenotype. This provides insight into the molecular control of the QS process and possible diagnostic tools to anticipate increased virulence in the field.

Acknowledgements

I would like to thank Professor Keith Matthews for his supervision and guidance throughout my PhD and Dr Mathieu Cayla for helping me through the day-to-day of my project. I would also like to thank my thesis committee, Dr Alasdair Ivens and Professor Amy Buck for their insight. The Matthews lab has been an invaluable source of advice and support, especially Dr Frank Venter who kindly performed the *in vivo* work, Dr Emma Briggs for generating the GPR89 sc-RNAseq expression profile, Dr Mabel Tettey Deladem, Dr Stephen Larcombe, Dr Javier López-Vidal, Beatrice Malacart, Dr Kirsty McWilliam, Dr Federico Rojas, Dr Eleanor Silvester, Dr Balázs Szöőr and Julie Young.

Within Ashworth, I am grateful for the insightful discussions, especially with Dr Achim Schnaufer and Dr Kamil Jaron. I was lucky to collaborate with Dr Frederik Van den Broeck (KU Leuven) who played an invaluable role in my project by providing unpublished genome sequencing data. Dr Mylène Verney and Dr Laurent Hébert (Unité PhEED, ANSES) provided insight into the implications of the manuscript in Chapter 2. Professor Richard McCulloch and Craig Lapsley (University of Glasgow) assisted me in sequencing the *T. melophagium* genome in Chapter 7. The Wellcome Trust (220058/Z/19/Z) funded this research.

I thank my partner, friends and family, especially Amarachi, Pip, Mike, Emily, Will, Charlie, Ella and Bobby for their encouragement and support.

List of abbreviations

8-pCPTcAMP 8-(4-Chlorophenylthio)adenosine 3′,5′-cyclic monophosphate

AAT Animal African trypanosomiasis

ATF Adipose tissue form

BLAST Basic local alignment search tool

bp Base pair

BSF Bloodstream form

cAMP cyclic 3', 5'-adenosine monophosphate

CCA Cis-aconitate

cDNA Complementary DNA

CDS Coding sequence

CNS Central nervous system

DAPI 4′,6-diamidino-2-phenylindole

DSB Double-stranded break

DYRK Dual-specificity tyrosine phosphorylation-regulated kinase

EATRO East African Trypanosomiasis Research Organisation

ES Expression site

ESAG Expression site-associated gene

ESB Expression site body

FAZ Flagellar attachment zone

GAPDH Glyceraldehyde 3-phosphate dehydrogenase

GC Gene conversion

gDNA Genomic DNA

GFP Green fluorescent protein

GPI Glycosylphosphatidylinositol

GRESAG Genes related to ESAG

HAT Human African trypanosomiasis

HYP2 Hypothetical protein 2

IFA Immunofluorescence analysis

INDEL Insertion/deletion

IV Intravenous

K Kinetoplast

kb Kilobase

kDNA Kinetoplast DNA

Log2FC Log2(fold change)

mRNA Messenger RNA

N Nucleus

NCBI National Centre for Biotechnology Information

ORF Open reading frame

PAD Protein associated with differentiation

PCF Procyclic form

PKC Protein kinase C

PP1 Protein phosphatase 1

RITseq RNA interference target sequencing

RNAi RNA interference

RNAseq RNA sequencing

RRM RNA recognition motif

rRNA Ribosomal RNA

RT Reverse transcriptase

RT-PCR Reverse transcription polymerase chain reaction

SIF Stumpy induction factor

SNP Single nucleotide polymorphism

SRA Serum resistance associated

TREU Trypanosomiasis Research Edinburgh University

TS Trans-sialidase

UTR Untranslated region

VSG Variant surface glycoprotein

WGS Whole genome sequencing

WHO World Health Organization

WT Wild type

Publications

The following manuscripts have been published and form the basis of chapters in this thesis:

Oldrieve, G. R. (2022). First person – Guy Oldrieve. Biol. Open **11**, bio059369.

Oldrieve, G. R, Verney, M., Jaron, K. S., Hébert, L. and Matthews, K. R. (2021). Monomorphic Trypanozoon: towards reconciling phylogeny and pathologies. Microb Genom **7**, 000632.

Oldrieve, G. R., Malacart, B., López-Vidal, J. and Matthews, K. R. (2022). The genomic basis of host and vector specificity in non-pathogenic trypanosomatids. Biol. Open **11**, bio059237.

I contributed to the following published manuscript during my PhD, although it does not form part of this thesis:

Briggs, E. M., Marques, C. A., Oldrieve, G. R., Hu, J., Otto, T. D. and Matthews, K. R. (2023). Profiling the bloodstream form and procyclic form Trypanosoma brucei cell cycle using single cell transcriptomics. bioRxiv 2023.01.09.523263.

Thesis structure

Chapters 1-6 represent my main PhD project “Developmental Incompetence in Selected and Naturally Occurring Trypanosoma Isolates”. Following the general discussion in Chapter 6, Chapter 7 consists of a manuscript on “The genomic basis of host and vector specificity in non-pathogenic trypanosomatids”.

Contents

1 Introduction

1.1. Background

1.1.1. Trypanosomatid overview

The order kinetoplastida is a group of single-celled eukaryotes which possesses cytological and genome biology that often diverges from the eukaryotic norm and shares a distant common ancestor with metazoa (Clade Discoba, Fig. 1.1). Within kinetoplastida, the family trypanosomatidae consists of a diverse group of eukaryotic protozoan parasites. Although the exact proportion is unknown due to a lack of sampling in non-mammalian hosts, it is assumed that trypanosomatidae predominantly occupy one host (monoxenous). The most basal species of trypanosomatid, which has been analysed at the genome level, *Paratrypanosoma confusum*, is monoxenous and infect insects (Flegontov et al., 2013). Therefore, it is predicted that the ancestor of trypanosomatids displayed a similar life history, further supported by the numerous and diverse species of monoxenous trypanosomatid parasites of insects (Podlipaev, 2001; Santos et al., 2006).

Species of trypanosomatid invested in a complex dixenous life cycle multiple times (Frolov et al., 2016; Tyler and Engman, 2001; Vickerman, 1985; Votýpka et al., 2015). Dixenous trypanosomatids infect mammals and include notorious species that are capable of infecting humans, livestock and wild animals. *Leishmania* spp. and *Trypanosoma cruzi* cause the diseases Leishmaniasis and Chagas disease, respectively, whilst *Trypanosoma brucei* subspecies are the causative agents of human African Trypanosomiasis (HAT) and Animal African Trypanosomiasis (AAT). Monoxenous and dixenous species are polyphyletic, as monoxenous species that commonly infect invertebrates can opportunistically infect vertebrate hosts (Kaufer et al., 2017) whilst some dixenous species have lost their ability to infect the invertebrate host (Frolov et al., 2016; Schnaufer et al., 2002).

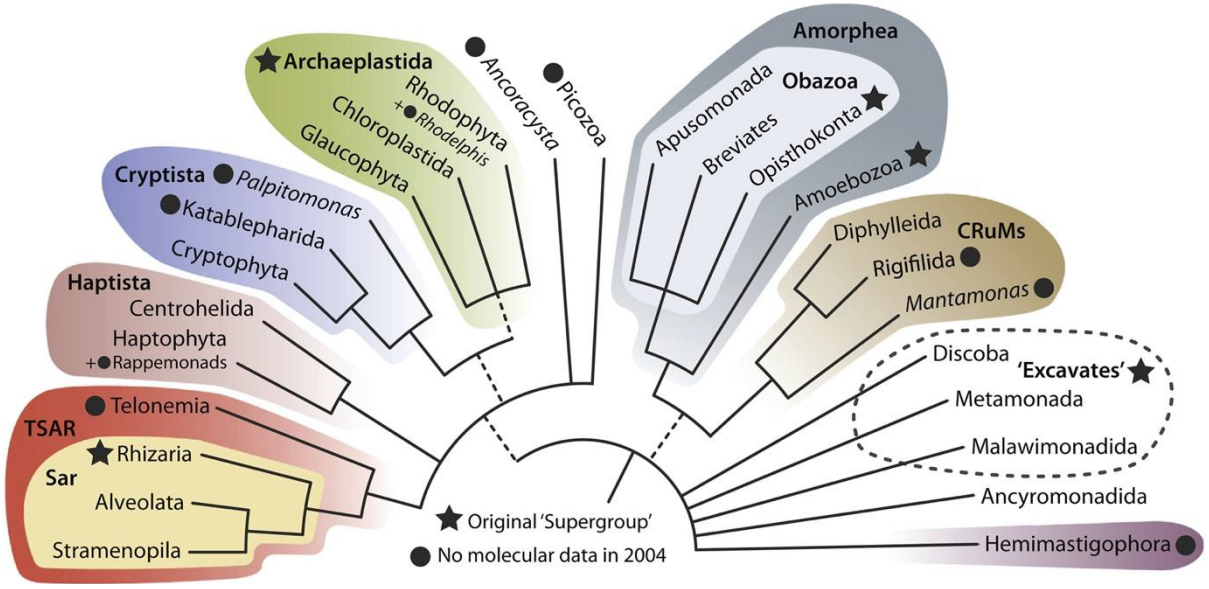


Figure 1.1: Trypanosomatids, represented by Excavates as part of the Clade Discoba, are an early branching group of eukaryotes. Adapted from (Burki et al., 2020).

1.1.2. *Trypanosoma brucei*

T. brucei is part of the salivarian trypanosome group which, if developmentally competent, is transmitted via the Tsetse fly (*Glossina sp.*) between mammals. To facilitate the transition from the mammalian bloodstream through their insect vector, *T. brucei* undergoes a complex life cycle which involves altering its intracellular organisation and cell shape. In the mammalian bloodstream, long slender forms progress through poorly characterised intermediate forms to short stumpy forms, preadapted for uptake by the tsetse fly (Fig. 1.2) (Bruce et al., 1910; Gull, 1999; Sharma et al., 2009; Vickerman, 1969). As such, the geographic range of developmentally competent *T. brucei* is restricted to the African Tsetse belt.

Some subspecies of *T. brucei* are unable to undergo developmental changes in the mammalian bloodstream, and due to this, have simplified their complex life cycle and

escaped the geographic range of their vector. These subspecies provide a chance to study multiple independent reversions from a complex to a simple life cycle, and the life history changes that are associated with this transition.

1.1.3. Human African trypanosomiasis

T. b. gambiense and *T. b. rhodesiense* are developmentally competent subspecies that cause diseases in mammals, including humans. Animal infective subspecies, such as *T. b. brucei*, are unable to infect humans, baboons and gorillas as they are susceptible to innate immune response trypanosome lytic factors (TLF) 1 and 2 (Hajduk et al., 1989; Lugli et al., 2004; Raper et al., 2001; Seed et al., 1990). TLF1 and TLF2 are complexes of high-density lipoprotein containing apolipoprotein L-1 (ApoL1) (Vanhamme et al., 2003) and haptoglobin-related protein (Smith et al., 1995). *T. b. gambiense* and *T. b. rhodesiense* have independently evolved mechanisms to circumvent these trypanolytic factors (Wheeler, 2010). *T. b. rhodesiense* utilises the serum resistance-associated (SRA) gene which is predicted to interfere with lytic factor action within the cell and/or uptake of the lytic factors (Van Xong et al., 1998). *T. b. gambiense* uses a modified VSG, TgsGP to interfere with lytic factors alongside reduced ApoL1 uptake via a haptoglobin haemoglobin receptor (HbHpR) caused by a mutation (Capewell et al., 2013; Uzureau et al., 2013) and lower transcript levels than *T. b. brucei* and *T. b. rhodesiense* (Kieft et al., 2010). However, the expression of HpHbR is beneficial to the fitness of *T. b. brucei* in its mammalian host (Higgins et al., 2017). A recent study highlighted that the deletion of HpHbR prevents the uptake of heme and reduces the ability of the parasite to develop from the slender form to the stumpy form. Expression of a *T. b. brucei* HpHbR in poorly developmentally competent *T. b. gambiense* improves the ability to generate stumpy forms, providing insight into the evolutionary tradeoffs of evolving human infectivity (Horáková et al., 2022).

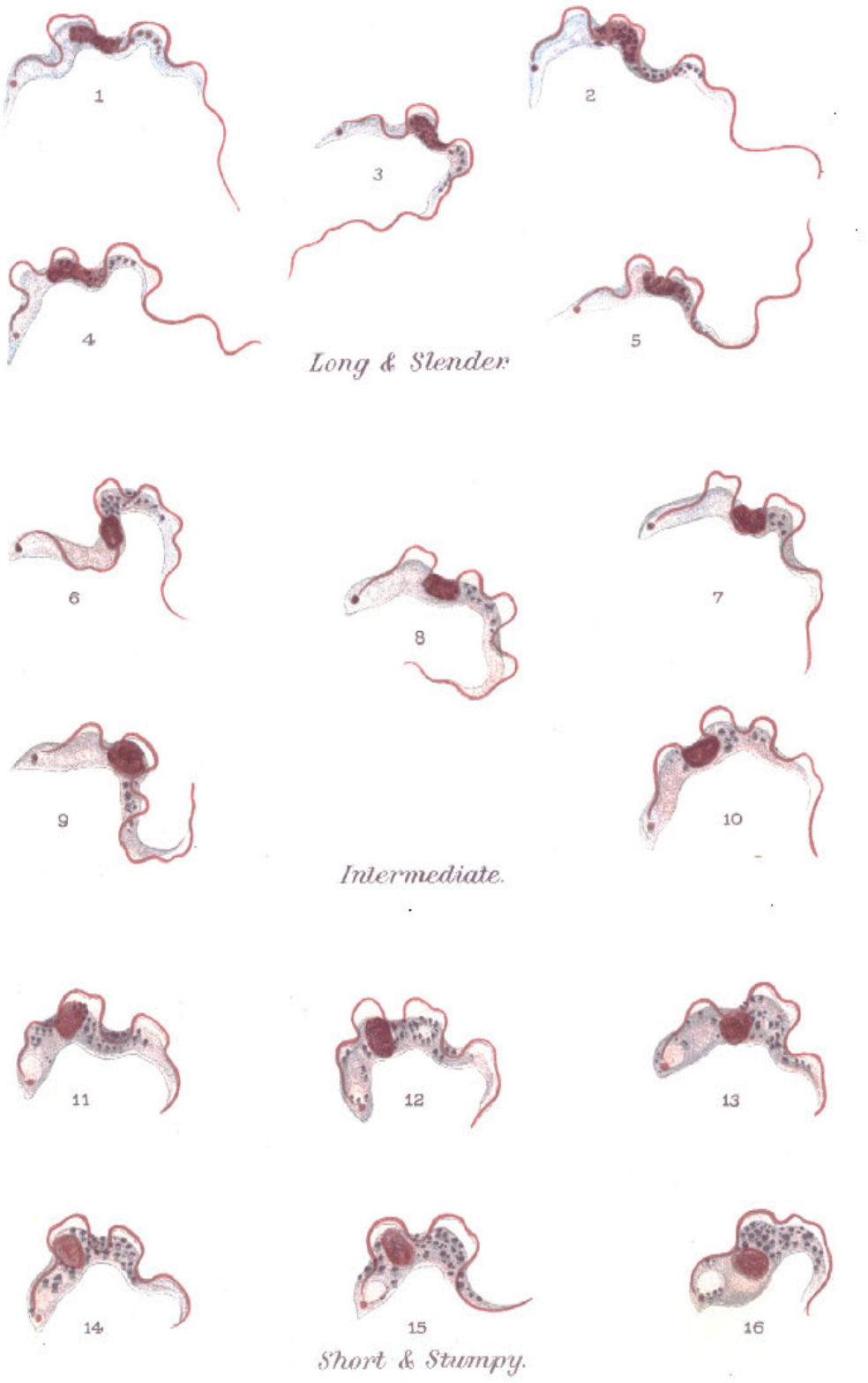


Figure 1.2: In the mammalian bloodstream, *Trypanosoma brucei* develop through intermediate stages during the transition from slender to stumpy form. The transition includes a diverse range of intermediate stages. Adapted from (Bruce et al., 1910).

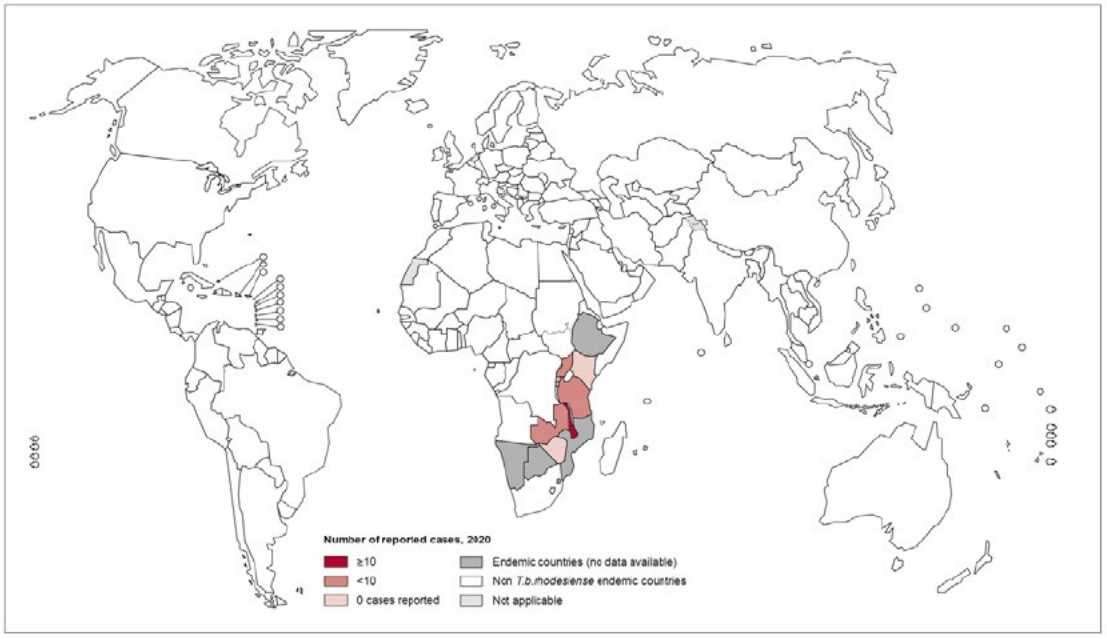
T. b. rhodesiense is transmitted by the Morsitans group of tsetse flies and is found in Eastern and Southern Africa (Fig. 1.3a). *T. b. rhodesiense* is zoonotic, spilling over from reservoirs of livestock and wildlife (Fèvre et al., 2001) and often producing acute symptoms leading to rapid death (Kennedy, 2013). In contrast, *T. b. gambiense* is found in Western and Central Africa and is transmitted by the Papalis clade of the tsetse fly (Fig. 1.3b). *T. b. gambiense* is largely anthroponotic, accounting for 93% of human infections in 2021 (Franco et al., 2022), which are often chronic (Büscher et al., 2018). *T. b. gambiense* and *T. b. rhodesiense* are the causative agents of human African trypanosomiasis (HAT).

In general, HAT begins as an early hemolymphatic stage with parasite replication in the blood, subcutaneous tissue, and lymph, accompanied by mild symptoms such as headaches and malaise. As the disease progresses to the neurological stage, the parasites cross the blood-brain barrier and infect the central nervous system. Symptoms progress to severe sensory and neurological disturbances such as a loss of coordination and a disruption of sleeping patterns. Left to progress, *T. b. rhodesiense* HAT is fatal (Kennedy, 2013). Likewise, *T. b. gambiense* can be fatal, although a study of untreated human *T. b. gambiense* infected patients discovered that 12/53 patients maintained asymptomatic infections (Jamonneau et al., 2012).

HAT has seen epidemics in the 20^{h} century, with 35,000 cases reported between 1997-1998. Cases have severely reduced since then due to vector control efforts and the development of fixed health facilities, with HAT diagnosis and treatment capacity, near populations which are at risk. HAT is on the brink of elimination as a public health problem, emphasised by just 992 and 663 reported cases of HAT in 2019 and 2020, respectively (Franco et al., 2022). Elimination is now in sight, with the development of the drug acoziborole, which displays an efficacy of 95% in patients with *T. b. gambiense* HAT from a single oral dose (Betu Kumeso et al., 2022). However, reservoirs of asymptomatic human carriers and the zoonotic potential from livestock and wild animals threaten the final

elimination of HAT (Büscher et al., 2018). Likewise, global climate change is altering the geographic distribution of the tsetse vector, humans, wildlife, and livestock. In 2019-2020, the Vwaza Marsh Game Reserve and Nkhotakota Wildlife Reserve Park (Malawi) experienced the most severe drought in 24 years. The drought has been linked to an outbreak of *T. b. rhodesiense* cases due to contact pattern alterations of vectors, humans, wildlife, and livestock (Hulsman et al., 2021).

(a)



(b)

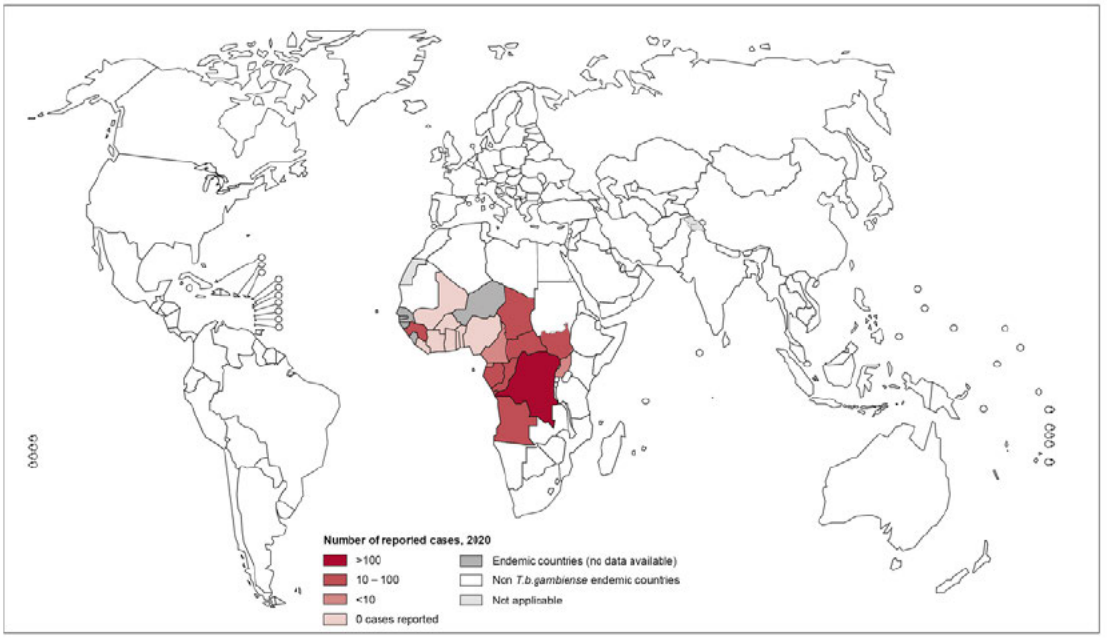


Figure 1.3: The agents of human African trypanosomiasis (HAT) are restricted to tsetse endemic countries. (a) *T. b. rhodesiense* and **(b)** *T. b. gambiense* global distribution in 2020. Adapted from the Control of Neglected Tropical Diseases (NTD) World Health Organization (WHO) (https://bit.ly/3Y95Txh).

1.1.4. Animal African trypanosomiasis (AAT)

Whilst the human infective subspecies have largely been eradicated in tsetse endemic regions, animal infective subspecies continue to cause devastating damage to agricultural production, with 120 million livestock at risk of AAT and over 3 million livestock deaths annually (Auty et al., 2015; Ebhodaghe et al., 2018; Giordani et al., 2016; Shaw et al., 2014). Losses caused by AAT equate to ~7% of agricultural GDP in Nigeria, with the impact predicted to rise unless AAT intervention strategies are employed (Odeniran et al., 2021). AAT exacerbates food insecurity and poverty, leading to widespread health implications that stretch further than the impacts of the human infective subspecies (Giordani et al., 2016). AAT is caused by *T. congolense*, *T. vivax* and to a lesser extent *T. brucei* subspecies, *T. b. brucei*, *T. b. evansi* and *T. b. equiperdum*. Whilst *T. b. evansi* and *T. b. equiperdum* are found in African countries, mechanical transmission (without tsetse involvement) has enabled them to spread outside of tsetse endemic regions and will be discussed in detail below. *T. congolense* and *T. vivax* present the greatest burden to agriculture but are often understudied by researchers in favour of *T. brucei (Morrison et al., 2016)*.

AAT clinical presentation often converges on similar symptoms, even though the disease is caused by different subspecies or species. Animals display weight loss, pyrexia and ataxia, which can lead to death within weeks in acute infections (Maudlin et al., 2004). The vast majority of cases are chronic, characterised by weakness and infertility. Control of AAT largely involves vector control and drug administration. The diverse species which cause AAT (Morrison et al., 2016) exhibit divergent life cycles (Matthews, 2021; Rotureau and Van Den Abbeele, 2013), metabolism-related sensitivity to drugs (Morrison et al., 2016; Steketee et al., 2021) and contrasting mechanisms of antigenic variation (Morrison et al., 2016), which complicates control of the diseases. Even within the species *T. brucei* and *T. congolense*, different isolates display contrasting virulence levels (*T. congolense* Savannah and Forest

types, and the *T. brucei* subspecies *T. b. evansi* and *T. b. brucei*) (Auty et al., 2015; Gibson, 2007). Different trypanosome species and subspecies are also capable of coinfecting their hosts. Concomitant infections with virulent and less virulent strains of *T. congolense* suppressed the pathology of the virulent strain (Morrison et al., 1982) whilst field-based surveys predict that there is an increase in the risk of severe disease when animals are coinfected (Savage et al., 2021), complicating host-parasite interactions (Silvester et al., 2017b; Venter et al., 2022).

1.1.5. Adaptations to parasitism

Mammalian immune system evasion by *T. brucei* has been the focus of intense cytological and molecular study. Each bloodstream-form cell is covered with a coat of variant surface glycoproteins (VSGs). VSGs prevent the adhesion of host antibodies to non-variant surface antigens buried beneath the densely packed VSG coat (Cross, 1975; Vickerman, 1969). VSG switching allows *T. brucei* to maintain infections in its mammalian hosts. Once the VSG has been recognised and the host mounts an effective antibody response to a specific VSG, a sub-population of *T. brucei* expressing a different VSG will outgrow to drive the next wave of parasitaemia (Vickerman, 1985). However, it is not as simple as a single VSG being expressed in each wave of parasitaemia; several VSGs are expressed by subpopulations at the same time during infection, highlighting that a switched VSG does not always dominate the next wave of infection (Mugnier et al., 2015). RNA Pol1 is responsible for the transcription of VSGs from a single expression site (ES) per cell, producing mono-allelic expression (Navarro and Gull, 2001). The ES is in telomeric regions and is accompanied by expression site-associated genes (ESAGs) (Cully et al., 1985). VSG switching occurs via alteration of the active ES or through homologous recombination, or gene conversion, to insert a new VSG into the active ES.

The extensive repertoire of VSGs (Berriman et al., 2005; De Lange and Borst, 1982; Wickstead et al., 2004) along with the ability to assemble mosaic VSG sequences, whereby the combination of multiple VSG sequences generates an entirely novel sequence, creates an almost infinite repertoire of VSG coats (Hall et al., 2013; Mugnier et al., 2015). The VSG repertoire helps the parasite to establish chronic infections. In addition to the bloodstream, VSG diversity could be generated in the tissues of the mammalian host, which are predicted to act as a reservoir to seed the bloodstream (Beaver et al., 2022). Developmentally competent *T. brucei* found in gonadal adipose tissue have a lower percentage of stumpy forms and divide at roughly half the rate of parasites in the bloodstream (Trindade et al., 2022). Separate experiments with a developmentally incompetent *T. brucei* found that lower proliferation is associated with a reduction in protein synthesis (Trindade et al., 2022). Nests of parasites have also been found to colonise lung tissue. Infected mice do not display reduced pulmonary dysfunction, however, the increased stress placed on the immune system did leave mice more vulnerable to coinfection with a respiratory virus (Mabille et al., 2022).

T. brucei VSG diversity is not shared amongst other trypanosome species. For instance, *T. vivax* does not employ the same diverse repertoire of VSG genes, resulting in geographically distinct isolates often expressing more similar VSGs. The reduction of antigenic diversity could cause exhaustion of VSG variants, limiting chronic infections, as has been observed in animals that self-cure *T. vivax* infections (Barry, 1986; Silva Pereira et al., 2020).

Based on subcellular protein localisation data, kinetoplastids have a disproportionate gain in the complexity of their mitochondrion (Billington et al., 2022) (Fig. 1.4). The increase in complexity is likely due to the divergent evolution of the unique mitochondrial genome, the kinetoplast (kDNA). However, the complexity of the mitochondrion is not associated with parasitism as the complexity is seen for all kinetoplastids, including free-living species (Lukeš et al., 2014). A *T. brucei* cell contains a single mitochondrion and, alongside its

well-documented function in energy generation through oxidative phosphorylation, the mitochondrion is associated with cell death, proliferation and signalling (McBride et al., 2006). The size and shape of the mitochondrion depend on the life cycle stage (Priest and Hajduk, 1994; Vickerman, 1965).

In the mammalian bloodstream form, the mitochondrion stretches the length of the cell body and lacks the cristae required for oxidative phosphorylation. Development into the insect stage promotes the enlargement of the organelle to produce a mitochondrial network throughout the cell. The inner mitochondrial membrane has enlarged cristae required for oxidative phosphorylation (Priest and Hajduk, 1994). As the cell contains one mitochondrion harbouring a single unit of kDNA (Jensen and Englund, 2012; Verner et al., 2015), trypanosomes require precise machinery for the segregation of kDNA to guarantee binary fission leaves two daughter cells with one kDNA copy each (Chanez et al., 2006). The precise segregation is achieved by coupling kDNA to the segregation of the basal body of the flagellum (Robinson and Gull, 1991). The structures are physically connected via the tripartite attachment complex (TAC). Disruption of the TAC is lethal, causing over-replicated kDNA which cannot be segregated (Povelones, 2014; Schneider and Ochsenreiter, 2018). Precise mitochondrial genome segregation via the TAC is unique to trypanosomes and their relatives, however, the TAC shares features with the mitotic spindle which segregates chromosomes during cell division in mammals (Blanco-Ameijeiras et al., 2022).

Alongside morphogenesis, the flagellum is vital for motility which is essential for sustaining infections in mammalian hosts and development in the insect host (Griffiths et al., 2007; Rotureau et al., 2014). Each cell contains a single flagellum that moves the cell body in a bihelical motion (Rodríguez et al., 2009). The flagellum is composed of a paraflagellar rod (PFR) and an axoneme, enclosed in a membrane and attached to the cell body via the flagellum attachment zone (FAZ). The distal FAZ corresponds to the localisation of key cell

cycle regulators (Billington et al., 2022; McAllaster et al., 2015; Zhou et al., 2016; Zhou et al., 2018).

Other organelles with a *T. brucei* specific gain in complexity, and therefore linked with parasitism, include the plasma membrane domains and mitotic spindle (Billington et al., 2022), supporting the suggestion that the cytoskeleton is a mediator of generating parasitism (Engstler et al., 2007; Lacomble et al., 2009; Olego-Fernandez et al., 2009; Portman and Gull, 2014; Sunter et al., 2019; Wheeler et al., 2013a).

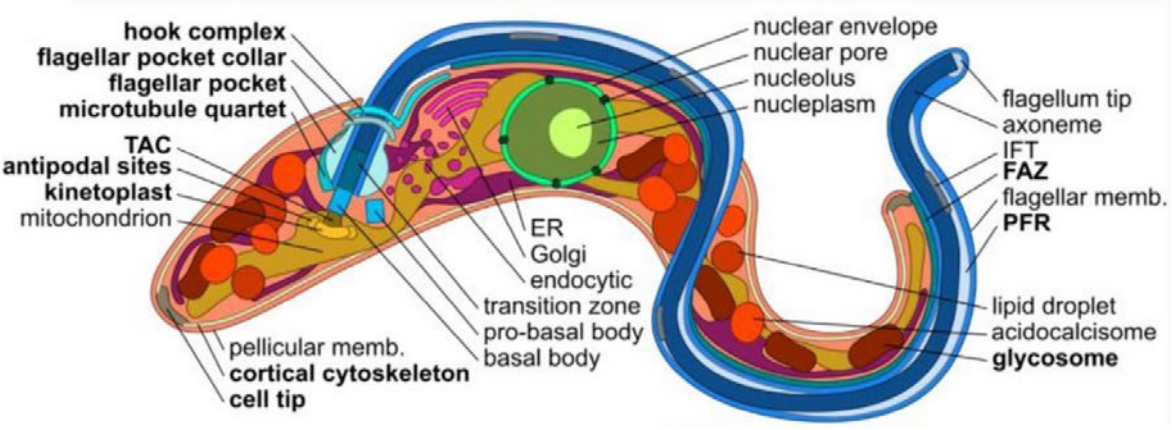


Figure 1.4: *T. brucei* organelles can be unique or display species specific elaboration. Unique organelles or organelles which display considerable elaborations in *T. brucei* are highlighted in bold. The figure was adapted from Billington et al., (2022). Abbreviations: tripartite attachment complex (TAC), endoplasmic reticulum (ER), intraflagellar transport (IFT), flagellum attachment zone (FAZ) and paraflagellar rod (PFR).

1.1.6. Genome architecture

kDNA consists of a network of concatenated circular DNA, minicircles and maxicircles (Jensen and Englund, 2012). There are 20-50 copies of ~23kb maxicircles encoding genes

for the respiratory chain and ribosomal RNAs (Borst and Fase-Fowler, 1979; Maslov et al., 2019; Schnaufer et al., 2002; Simpson, 1979). Minicircles are ~1kb and have 1000s of copies which encode guide RNAs required for editing mitochondrial mRNA transcripts (Hajduk and Ochsenreiter, 2010; Jensen and Englund, 2012; Kleisen and Borst, 1975; Steinert, 1960; Vickerman, 1965). The kinetoplast is essential for the insect stage of the parasite's life cycle as the encoded respiratory components are required for viability in the tsetse fly (Schnaufer et al., 2002; Vickerman, 1965). Some *T. brucei* subspecies have reduced or removed their kDNA, rendering them unviable in the insect vector (Schnaufer et al., 2002). These are called dyskinetoplastic or akinetopastic forms, respectively.

The *T. brucei* nuclear genome encompasses 11 megabase chromosomes along with ~5 intermediate and ~100 mini chromosomes (Daniels et al., 2010; Wickstead et al., 2004). Chromosomal fusion led to a reduced chromosome number in *T. brucei*, compared to other trypanosomatids (El-Sayed et al., 2005). The nuclear genome is tightly packed with genes organised into co-transcribed polycistronic units. Mature mRNA is resolved through trans-splicing and polyadenylation of primary transcripts (Clayton, 2019; Parsons et al., 1984). Polycistronic transcription is universal amongst kinetoplastids (Jackson et al., 2008).

Genome expansion events have been well documented amongst diverse *Trypanosoma*, such as the expansion of *T. theileri* specific surface proteins (TTPSPs) in *T. theileri*, in comparison to the closely related *T. melophagium* (Kelly et al., 2017; Oldrieve et al., 2022). Likewise, *T. brucei* has facilitated the expansion of hexose transporters (Bringaud and Baltz, 1994), a gene array of phosphoglycerate kinases (Le Blancq et al., 1988) and elaborated the VSG repertoire (Taylor and Rudenko, 2006) which arose via duplication and gene conversion (Jackson, 2007a; Jackson, 2007b). Sexual reproduction between subspecies of *T. brucei, T. b. rhodesiense* and *T. b. brucei*, caused alterations in chromosome size, and trisomy has been observed in laboratory studies (Gibson et al., 1992). The last common ancestor of the *T. brucei* clade underwent a segmental duplication event which gave rise to

parts of chromosomes 4 and 8. The duplication is absent from *T. congolense* but observed in subspecies of *T. brucei*, such as *T. b. rhodesiense*, *T. b. gambiense* and *T. b. evansi*. Post duplication, the chromosomes selectively conserved 47% of duplicated genes, enriching glycosyltransferases, amino acid transporters and adenylate cyclases, along with genes containing transmembrane helicases and signal peptides. The duplication provided a "double dose" of some genes and there is evidence of functional innovation facilitated by the duplication (Jackson, 2007b). A protein serine-threonine kinase (NRK) was part of this duplication event. NRK A/B are associated with the control of development from the bloodstream form to the insect midgut procyclic forms (Domingo Sananes et al, 2015). It is not known if the duplication of NRK confers functional innovation along with a double dose, but highlights the potential role of genome architecture in the developmental competence of *T. brucei* (Gale et al., 1994; Jackson, 2007b).

1.2. *T. brucei* developmental competence (pleomorphism)

1.2.1. *T. brucei* life cycle

T. brucei undergoes a complex life cycle, including transmission between mammals via an insect vector, the tsetse fly (MacGregor et al., 2012). To facilitate the transition between the stages of its life cycle, *T. brucei* undergoes morphological and cytological changes. In its mammalian stage, developmentally competent *T. brucei* proliferates in its host as a slender form. Slender forms are characterised by a slim cell which contains a posterior, terminal, kDNA and a central nucleus (Vickerman, 1985). Slender forms have a flagellum which stretches the length of the cell body and extends beyond it. The flagellum facilitates the highly mobile cells to move in a spiral pattern (Hill, 2003). Due to the abundance of blood glucose, the parasite mitochondrion is tubular, lacks cristae and does not have an operational electron transport chain, instead relying on the glycolysis of blood glucose (Smith et al., 2017).

Chronic mammalian infections with *T. brucei* entail waves of parasitaemia. The proliferative slender forms increase in population density via binary fission (Benz et al., 2017; Wheeler et al., 2013b). Similarly to other eukaryotes, they progress through the cell cycle from G1, S, G2 and M and cytokinesis. The timing of the nucleus and kinetoplast division occurs at different rates. Therefore, cells can be observed with 1 kinetoplast and 1 nucleus (1K1N (G1)), 2K1N (S) and 2K2N (G2/M) (Woodward and Gull, 1990). The transcriptional profile of the cell cycle in the bloodstream and the procyclic form of *T. brucei* have recently been profiled using single-cell RNA-seq (scRNA-seq) (Briggs et al., 2023). In the bloodstream, as the population density increases, the cells transition through morphologically intermediate forms to a cell cycle-arrested stumpy form (Matthews and Gull, 1994; Ziegelbauer et al., 1990). The stumpy form contains one kinetoplast and one nucleus (1K1N) and is arrested in

G0/G1. The *T. brucei* stumpy cell is morphologically distinct from the slender form, being broader, along with a defined undulating membrane between the cell body and the flagellum (Fig. 1.2) (Bruce et al., 1910). In comparison to the slender form, the stumpy form has a rounder nucleus which moves to the posterior of the cell. The kinetoplast can be displaced, due to an enlargement of the flagellar pocket (Engstler et al., 2007). Stumpy forms have an elaborated mitochondrial structure to facilitate the increase of mitochondrial activity and utilisation of α-ketoglutarate (Dewar et al., 2018). Using scRNA-seq, the population-wide transcriptome pattern during the transition from slender to stumpy forms has been mapped. Despite their description from morphological data (Fig. 1.2), intermediate-form clusters of cells were not identified (Briggs et al., 2021).

Stumpy forms are preadapted for uptake by the tsetse fly (Fig. 1.5), via resistance to antibody-mediated lysis (McLintock et al., 1993), resistance to the complement pathway via the expression of a factor H receptor (Macleod et al., 2020) and tolerance of pH stresses and proteases (Nolan et al., 2000). While slender forms can establish tsetse fly infections in a laboratory setting, stumpy form preadaptation suggests they produce the majority of tsetse infections in the field (Matthews and Larcombe, 2022; Schuster et al., 2021). Stumpy forms express two members of the PAD family (proteins associated with differentiation). These membrane transporters, PAD1 and PAD2, sensitise the parasites to physiologically relevant levels of citrate in the blood meal. In turn, this stimulates the development of the stumpy forms into procyclic forms, upon reduction in temperature associated with transmission from the mammalian bloodstream to the insect vector (Dean et al., 2009), via a phosphatase signalling cascade (Szöőr et al., 2006; Szöor et al., 2010; Szöőr et al., 2019). The transition from the mammal to the insect involves changes in the temperature, tissues, host defence mechanism and nutrient availability. Procyclic forms re-enter the cell cycle, after arrest as a stumpy form which involves global gene expression changes (Walsh and Hill, 2021).

Early procyclic forms in the gut lumen and proventriculus express a surface coat of GPEET procyclin and EP procyclin. Late procyclic forms repress GPEET expression and cross the peritrophic matrix into the ectoperitrophic space (Acosta-Serrano et al., 2001; Rose et al., 2020; Vassella et al., 1997). Entry into the ectoperitrophic space was thought to involve penetration of a mature peritrophic matrix (Ellis and Evans, 1977; Evans and Ellis, 1978; Gibson and Bailey, 2003). However, recent work suggests this is achieved when the new peritrophic matrix is made by the proventriculus (Rose et al., 2020). Procyclic, long procyclic and mesocyclic forms are found in the ectoperitrophic space. Parasites cross the peritrophic membrane a second time, into the proventriculus lumen, where they progress from long-dividing epimastigotes into short-nondividing epimastigotes (Rotureau and Van Den Abbeele, 2013; Shaw et al., 2019). Short epimastigotes finish their journey through the insect vector in the salivary gland, where they attach to the epithelium (Rotureau and Van Den Abbeele, 2013; Rotureau et al., 2012).

The epimastigotes re-enter the cell cycle and differentiate into metacyclic trypomastigotes, capable of infecting mammals and developing into slender bloodstream forms. The process of metacyclogenesis in the salivary gland has been characterised by scRNA-Seq which highlighted three distinct clusters of cells that represent epimastigote, pre and mature metacyclic parasites. The transcriptomic and metabolic profile reflected preadaptation for survival in the mammalian host (Vigneron et al., 2020). A successful infection establishment in the tsetse salivary gland modifies the saliva composition and alters the feeding behaviour of the fly to enhance transmission (Van Den Abbeele et al., 2010). RNA binding proteins are essential regulators of developmental progression through the insect stages. For example, RBP6 overexpression induces metacyclic generation *in vitro* (Kolev et al., 2012). Progression through the insect is thought to be assisted via social motility (SOMO). The behaviour has been observed *in vitro (Imhof et al., 2015; Lopez et al., 2015; Oberholzer et al., 2010)* and, potentially, *in vivo* (Shaw et al., 2019) in early procyclic form cells which express GPEET (Imhof et al., 2014).

The time taken to produce mammalian infectious metacyclic forms varies by parasite strain, but takes at least 12 days, often considerably longer, and entails passing major bottlenecks (Dyer et al., 2013). Single trypanosomes were tagged with synthetic DNA, and then a population containing a mixture of tagged cells was used for infections. Transmission from mouse to tsetse fly or tsetse fly to mouse allowed the survival of hundreds of cells which contributed to the infection. However, differentiation and migration through the tsetse fly produced major bottlenecks, with a single cell often producing 99% of the subsequent population (Oberle et al., 2010). The major bottlenecks highlight the importance of clonal expansion from a few genotypes within a region, and the importance of recombination between these strains.

T. brucei sexual reproduction occurs in the salivary gland, involving meiotic division and the production of haploid gametes (Peacock et al., 2011; Peacock et al., 2014). Sexual reproduction between different strains of *T. brucei* can occur if a tsetse fly is co-infected with multiple species at the same time (Kay et al., 2022). Tsetse flies produce a robust immune response after the first infection (Jannin, 1999; Weiss et al., 2013; Weiss et al., 2014) and infection chance decreases with age (Walshe et al., 2011; Welburn and Maudlin, 1992; Wijers, 1958), reducing the chance of a secondary infection in the flies. However, tsetse flies can become co-infected through multiple feeds which are 18 days apart (Peacock et al., 2016) and hybrids have been observed 13 days post-infection (Gibson et al., 2008). Analysis of field samples showed that hybridisation events have occurred between *T. brucei* subspecies. Critically, these events were between human and non-human infective subspecies (Kay et al., 2022) supporting experimental evidence of recombination in a laboratory setting (Gibson et al., 1992). Therefore, sexual reproduction facilitates genetic exchange, which maintains the genetic diversity required to withstand environmental and host-based pressures, presenting the threat of new human infective subspecies.

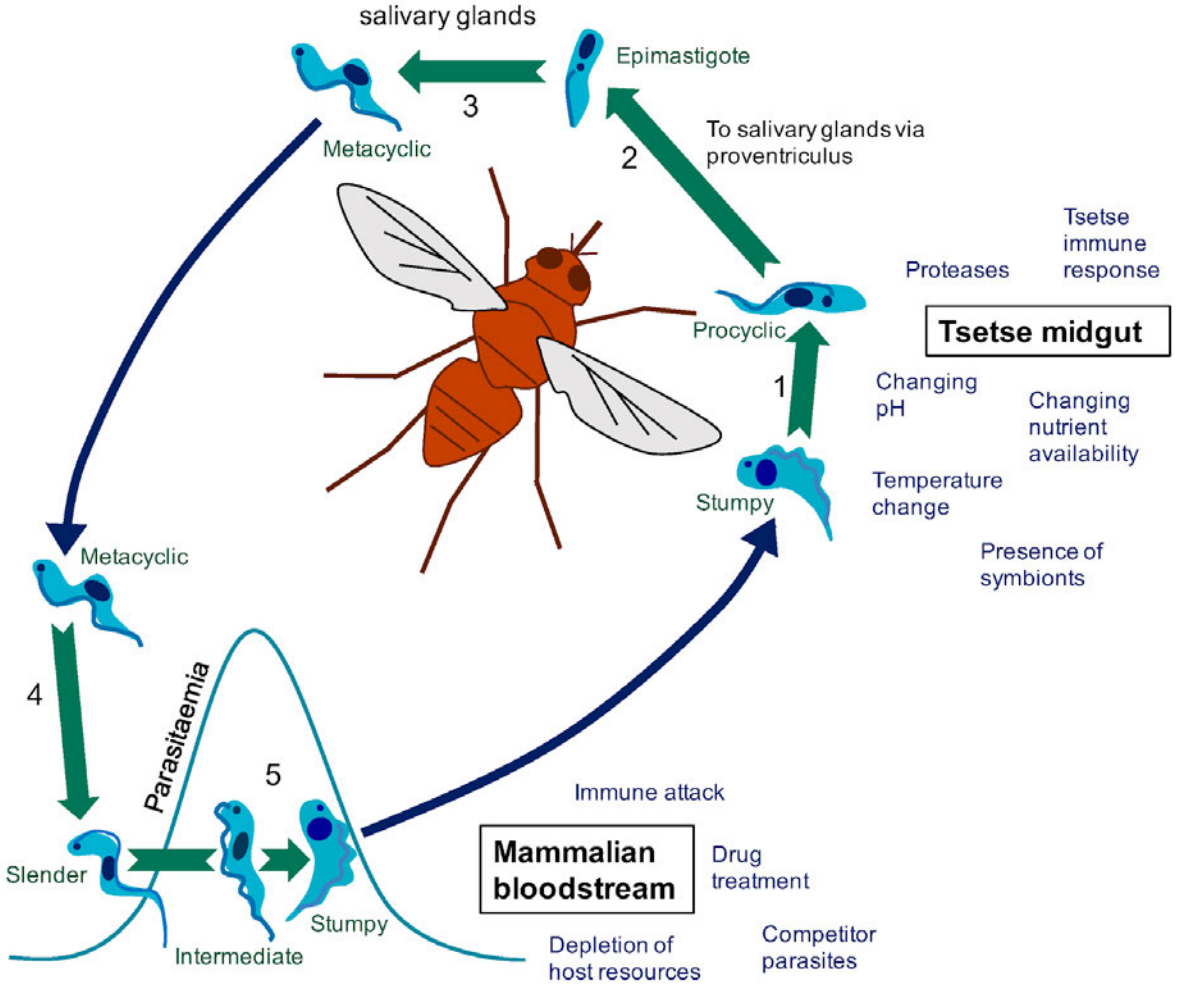


Figure 1.5: The complex *T. brucei* life cycle involves transitions between multiple cell morphologies, each adapted for life within either the insect or mammal host. Adapted from (Silvester et al., 2017a).

1.2.2. Quorum sensing

Quorum sensing (QS), the ability of a microbial population to communicate via secreted signalling molecules to synchronise the behaviour of individuals, is often associated with bacteria (Waters and Bassler, 2005). In the mammalian host, the slender form of *T. brucei* proliferates in waves of parasitaemia (Matthews and Gull, 1994; Ziegelbauer et al., 1990).

The waves of parasitaemia represent an initial increase in the population density of the slender proliferative form (Reuner et al., 1997; Waters and Bassler, 2005). *T. brucei* releases peptidases which hydrolyse host proteins to produce oligopeptides, acting as the QS signal when the parasite density reaches a specific threshold (Tettey et al., 2022). Oligopeptides are transported via GPR89 (Tb927.8.1530), which initiates a signalling cascade, ending in the initiation of stumpy induction and PAD1 expression via a complicated molecular pathway (Mony et al., 2014). The QS pathway causes the transformation of the slender form into a stumpy form (Rojas et al., 2019; Vassella et al., 1997). Stumpy forms are cell cycle arrested and so arrest the increase in parasitaemia in the mammalian host, controlling virulence and promoting transmission to the insect vector (Matthews, 2021; Rojas and Matthews, 2019). QS can be mirrored *in vitro*, suggesting independence from the mammalian immune system (Balber, 1972; Rojas et al., 2019; Rojas et al., 2021). However, the immune system could still influence stumpy formation, for example, the release of proteases by macrophages may accelerate stumpy formation.

Whilst oligopeptide-based QS accounts for most of the generation of stumpy-form cells, parasites which are naturally unable to perform QS can sporadically generate stumpy-form parasites (Hoare, 1972; Yorke and Blacklock, 1912). Modelling of the transition from slender to stumpy form fits a model whereby QS-independent differentiation is possible (Dewar et al., 2018). In support of these findings, (Zimmermann et al., 2017) showed that artificial ectopic expression of a second VSG gene promoted stumpy formation through transcriptional attenuation of the VSG expression site.

1.2.3. The molecular basis of quorum sensing

Combining RITseq technology, a high-throughput genome-wide gene phenotyping screen, with selection using an *in vitro* mimic of the QS signal, cell-permeable cAMP analogue

8-pCPTcAMP/AMP, allowed identification of positive regulators of the QS pathway in a monomorphic cell line (Mony et al., 2014; Vassella et al., 1997). 8-pCPTcAMP/AMP in uninduced monomorphic populations caused cell cycle arrest and eventual cell death. With RNAi induced, cells which silenced genes involved in the QS pathway (and so unresponsive to 8-pCPTcAMP/AMP) continue to grow and become enriched. Sequencing of the selected population highlighted genes involved in the quorum sensing pathway (Mony et al., 2014).

The screen identified genes linked to the signalling cascade, which highlighted the likely role of phosphatases and kinases in regulating the transition from slender to stumpy form, both as effector and signalling molecules. Protein phosphorylation has a major post-translational modification role in eukaryotes. This modification acts on diverse cellular functions and in *T. brucei* a dual specificity phosphatase (DsPhos) and a protein phosphatase type 1 (PP1) were implicated in the transition from slender to stumpy form (Mony et al., 2014). Also, protein kinases contribute, with a MEK kinase (MEKK1), a 5' adenosine monophosphate-activated protein kinase (AMPKα2), NEK17 and DYRK identified (Mony et al., 2014). Subsequent work demonstrated that DYRK acts to phosphorylate and inhibit molecules that retain trypanosomes in the slender form (‘slender retainers’), such as NOT5 whilst also phosphorylating, and activating, stumpy inducers, such as the zinc finger protein, ZC3H20 (Cayla et al., 2020). scRNA-seq of a ZC3H20 null mutant cell line identified early transcripts which began to alter before the pathway was ‘blocked’ by the absence of ZC3H20 (Briggs et al., 2021). These genes provide interesting targets to identify early molecular alterations which initiate the progression towards stumpy forms.

Other post-translational modifiers were identified, such as the NEDD8 activating enzyme APPBP1 (Mony et al., 2014). Ubiquitin and ubiquitin-like proteins, such as NEDD8, are essential cellular regulators from eukaryotes to bacteria. NEDD8, UBA3 and APPBP1 initiate a neddylation cascade (Walden et al., 2003) which is essential for signal transduction (Amir et al., 2002; Read et al., 2000), cytokinesis (Kurz et al., 2002), cell division (Osaka et al.,

2000), and development (Pozo et al., 1998). The protein neddylation pathway has been dissected in *T. brucei* which highlighted that another QS gene, KRIPP14 (HYP13) (Tb927.11.11470) is Nedd8-conjugated or associated (Liao et al., 2017). KRIPP14 is a putative mitochondrial SSU ribosomal protein (Mony and Matthews, 2015).

Gene regulators in the QS cascade ultimately cause changes in transcript abundance or translational changes, which enact the transition from slender to stumpy form. RNA binding proteins, such as ZC3H20 (Briggs et al., 2021; Cayla et al., 2020; Liu et al., 2020) and RBP7A/B are essential for the transition between life stages in *T. brucei*. Knockdown of RBP7A/B causes unresponsiveness to oligopeptides and overexpression of RBP7B drives stumpy formation (Mony et al., 2014). Other RNA binding proteins, such as RBP6 are essential for development in the insect vector (Kolev et al., 2012) whereas RBP10 is essential in promoting the differentiation state of bloodstream-form slender cells to the procyclic form. Depletion of RBP10 causes the transition of slender forms to procyclic forms, although the role of RBP10 in transitioning directly to the stumpy form has not been directly investigated (Mugo and Clayton, 2017). Whilst downstream validation of the QS pathway has focussed on RNA binding proteins, a transcriptional silencer imitation switch (ISWI) protein was also identified as a putative stumpy inducer (Hughes et al., 2007; Mony et al., 2014; Stanne et al., 2011).

A screen for post-transcriptional regulators correlated with the identification of stumpy formation regulators to highlight the importance of several hypothetical proteins in gene regulation, for example, HYP1 (Tb927.11.6600), HYP2 (Tb927.9.4080) and HYP12 (Tb927.11.2250) (Erben et al., 2014; Mony and Matthews, 2015; Mony et al., 2014). HYP1 is proposed to destabilise molecules that act as slender retainers and inhibitors of stumpy formation whilst HYP2 likely stabilises genes, by binding to mRNAs, which drives stumpy formation (Mony and Matthews, 2015). A long non-coding RNA regulator, SnoGRUMPY, has also been implicated in differentiation promotion (Guegan et al., 2022).

Oligopeptide-based quorum sensing initiates a cascade which ultimately ends in the generation of stumpy form parasites (Vassella et al., 1997). The dependency relationships between components of the QS pathway have been dissected, which highlighted that the pathway is non-linear (McDonald et al., 2018). However, the early events which link the reception of the oligopeptide signal to the QS pathway are currently unknown as the initial study used a monomorphic strain and bypassed the signal reception with the use of a cell-permeable cAMP analogue (Mony et al., 2014).

1.2.4. Developmental competence in *T. congolense* and *T. vivax*

The stumpy form is unique to *T. brucei*, with closely related species *T. congolense* and *T. vivax* displaying morphological uniformity in the mammalian bloodstream. However, orthologues of the QS regulators exist in both *T. congolense* and *T. vivax* genomes, and the cells arrest their cell cycle in G1 at high population parasitaemia (Shapiro et al., 1984; Silvester et al., 2017a). Critically, tsetse flies can become infected with *T. congolense* at any stage during an infection. However, maturation of the parasite in the tsetse is more efficient when the blood meal is taken later in the infection, suggesting development into a preadapted form as the mammalian infection progresses (Akoda et al., 2008). Replacing HYP2 with the *T. congolense* (TcIL3000.0.19510) restored stumpy formation in an HYP2 null mutant, suggesting at least one of the QS pathway genes displays the same function in the two species (Silvester et al., 2017a).

Developmentally incompetent isolates of *T. brucei* exist which are unable to generate the stumpy form, and therefore, likely have disrupted QS pathways. Utilising genomic data from these clades might provide further insight into the molecular basis of QS, and how it can be disrupted.

1.3. *T. brucei* developmental incompetence (monomorphism)

1.3.1. Dourine

Dourine is a sexually transmitted disease of equids. Conventionally, it is described as being caused by *T. equiperdum*. Typical clinical presentations include but are not limited to, pyrexia, anaemia, urticarial plaques, and keratitis. Ultimately, the disease enters the neurological phase, leading to ataxia and paralysis which precede death (Büscher et al., 2019). The incubation period can vary from weeks to years, depending on the strain of the parasite and the host's immune status (Constable et al., 2017).

As the causative agents of Dourine are no longer restricted to tsetse endemic regions, the prevalence and presence of the causative agents are often underreported by the World Organisation for Animal Health (WOAH) (Aregawi et al., 2019). Although a systematic approach has not been performed (Desquesnes et al., 2022a), descriptive studies have noted the presence of *T. equiperdum* in Ethiopia (Gizaw et al., 2021), Iran (Sazmand et al., 2020), and Mongolia (Suganuma et al., 2016). In total, 16 countries reported cases of Dourine to WOAH between 2005-2022 (Fig. 1.6a). The distribution highlights outbreak events, such as the spread of Dourine in Italy in 2011 (Calistri et al., 2013; Pascucci et al., 2013) (Fig. 1.6a).

Similarly to AAT, vaccines are not available for Dourine, however, chemotherapeutic treatment with suramin, quinapyramine chloride, melarsomine hydrochloride, isometamidium chloride and diminazene diaceturate can be effective (Büscher et al., 2019). However, strains can display innate or acquired resistance to chemotherapies (Mekonnen et al., 2018).

Chemotherapy can be ineffective during the acute phase of infection since the drugs are unable to penetrate tissues, and therefore unable to kill parasites which have crossed the blood-brain barrier during the neurological stage (Auty et al., 2008; Cauchard et al., 2016; Hébert et al., 2018; Ranjithkumar et al., 2014). Outbreaks are caused by negligence when screening animals that are imported from a Dourine endemic country such as the import of Dromedary camels from Gran Canaria into Spain (Gutierrez et al., 2010; Tamarit et al., 2010), the importation of a horse into Italy (Calistri et al., 2013) and artificial insemination with contaminated semen (Ahmed et al., 2018). The most successful control of Dourine remains to screen animals from endemic countries.

Diagnosing Dourine is difficult due to an absence of specific clinical symptoms, and an often extremely low parasitaemia, below that of even PCR-based assays (Calistri et al., 2013). Diagnosis with Dourine, therefore, often relies upon a combination of epidemiological context, serological, molecular, parasitological and clinical evidence (Büscher et al., 2019; Desquesnes et al., 2022a).

1.3.2. Surra

Surra, historically thought to be caused by *T. evansi*, is found in a diverse repertoire of domestic animals such as bovines, camels, dogs and goats alongside wild animals such as gazelle and elephants. It is mechanically transmitted, largely via tabanids, stomoxys, which are globally distributed, and tsetse (Gingrich et al., 1983; Mihok et al., 1995; Roberts et al., 1989). Whilst Surra is thought of as a blood-borne disease, the parasite can invade tissues such as the central nervous system (Holland et al., 2003). The disease presentation varies between the host, geographic region and genotype of infecting parasite. However, general signs include anaemia, weight loss, fever, chronic wasting and abortion. Surra also reduces milk production, increasing the economic burden of the disease (Desquesnes et al., 2013).

Aregawi et al., (2019) highlighted the distribution of *T. evansi* across the globe, far beyond the tsetse endemic regions of *T. brucei*. For comparison to the range of *T. b. equiperdum*, we have plotted the presence of reported Surra cases reported to WAHIS between 2005-2022. Out of 241 countries, Surra is more widespread than Dourine, with 44 countries reporting the disease, corroborating the global distribution of the causative agents of Surra (Fig. 1.6b).

Treatment of Surra with chemotherapy largely relies on diminazene aceturate, whilst isometamidium chloride, cymelarsan, suramin and quinapyramine are used infrequently (Desquesnes et al., 2013). However, there are inefficiencies with each of these drugs in treating Surra (Desquesnes et al., 2012; Peregrine et al., 1995; Toro et al., 1983; Tuntasuvan et al., 2003). Preventative measures are also used to control the vector. However, unlike a parasite which relies upon a single vector for cyclical transmission, monomorphic *T. brucei* are mechanically transmitted via biting insects, or even larger animals such as vampire bats (Hoare, 1965). Therefore, whilst nets and insect control strategies could be employed, it has been proposed that it is likely futile as another vector will replace the one under control (Desquesnes et al., 2013). Instead, it is preferable to attempt to control the importation of Surra-causing agents into non-endemic regions via quarantine and screening of animals that are transported from endemic regions (Desquesnes et al., 2013).

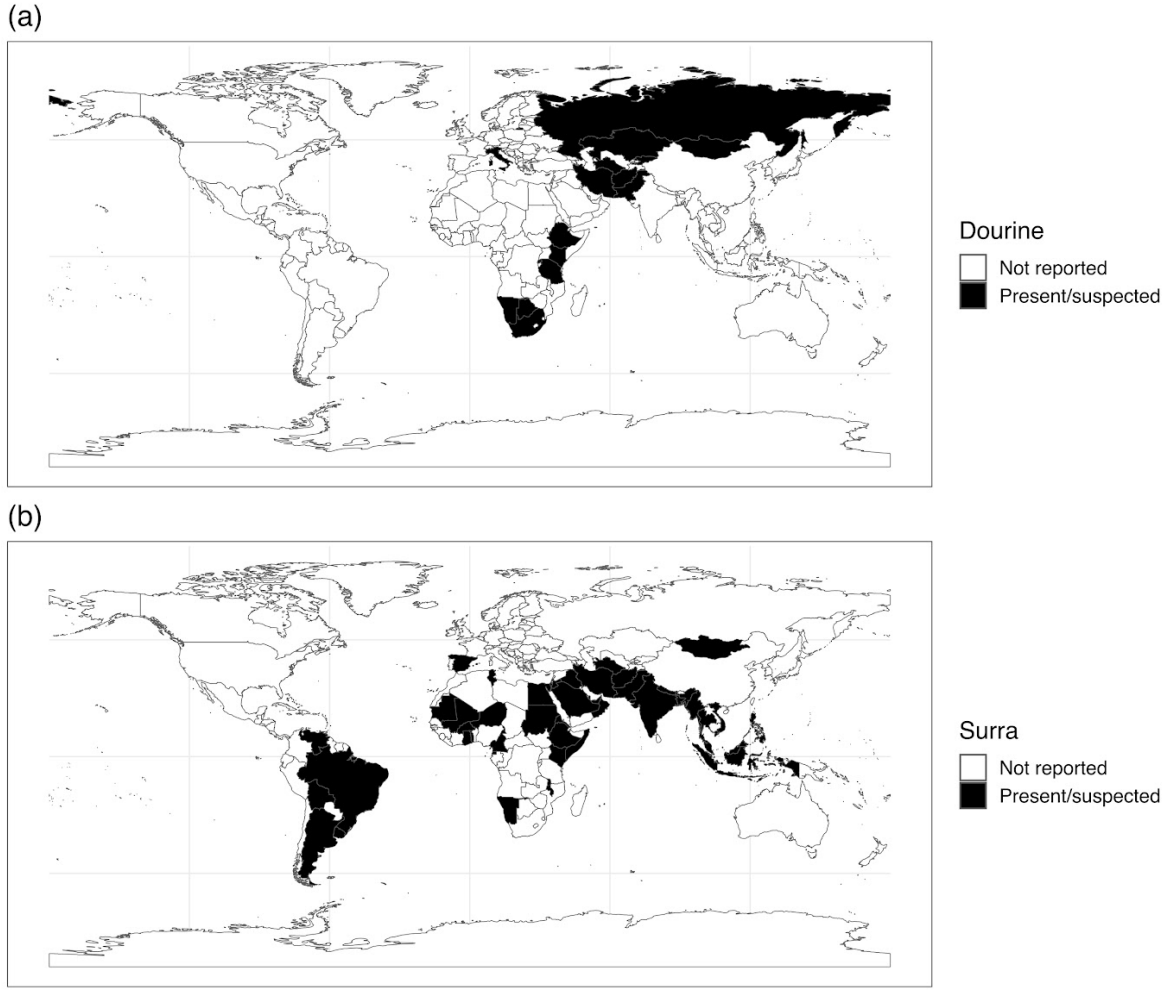


Figure 1.6: The causative agents of Dourine and Surra display a geographical distribution beyond tsetse endemic regions. Presence of **(a)** Dourine and **(b)** Surra cases submitted to OIE-WAHIS between 2005-2022. The presence of each disease was compared to a list of 241 countries in the rnaturalearth package (available at https://github.com/ropensci/rnaturalearth) plotted on a world map using ggplot, sf and rnaturalearth (Pebesma, 2018; Wickham, 2016). Presence data were extracted from the WAHIS disease situation dashboard (https://wahis.woah.org/#/dashboards/country-or-disease-dashboard).

1.3.3. Phylogeny of developmentally incompetent *T. brucei*

Developmentally incompetent *T. brucei* subspecies are less able to transition between the slender proliferative form and the cell cycle-arrested stumpy form. Instead, they rely upon mechanical transmission between mammalian hosts, either via an infected blood meal or venereal transmission between equids (Schnaufer et al., 2002). Monomorphic *T. brucei* have escaped the tsetse fly's geographic range, expanding to most continents. As they no longer undergo cyclical development in their insect host, they do not require the respiratory components required for metabolism in the arthropod vector (Schnaufer et al., 2002; Vickerman, 1965). This allowed for a secondary reduction or removal of their mitochondrial genome, the kinetoplast (Dewar et al., 2018). They infect wild animals and livestock, causing the diseases Surra and Dourine (Büscher et al., 2019). Developmentally incompetent *T. brucei* can also be generated in the laboratory (McWilliam et al., 2019; Turner, 1990).

Historically, monomorphic *T. brucei* were defined as two separate species, *T. evansi* and *T. equiperdum* (Hoare, 1972). They were largely defined by morphology and disease presentation. *T. evansi* was associated with mechanical transmission via biting flies between cattle, causing Surra, and *T. equiperdum* was defined as a sexually transmitted parasite of equines, causing Dourine (Büscher et al., 2019). Historical papers describe a 3rd species of monomorphic *T. brucei, T. equinum*. *T. equinum* was first described in 1901 in South America. It was described as a parasite of equines and appeared to lack a kinetoplast but otherwise, it was morphologically indistinguishable from *T. evansi*. It was hypothesised that *T. equinum* evolved from *T. evansi* upon arrival in South America (Hoare, 1954). However, without molecular data, this was not confirmed. Subsequently, *T. equinum* was synonymised with *T. equiperdum* (Gibson, 2007; Ventura et al., 2000). *T. equinum* highlights the difficulty in defining species, subspecies and types of protozoa without molecular data.

Since 1972, Hoare acknowledged the phenotype associated with monomorphism is likely unassociated with a specific clade or genotype. “If animals infected with *T. b. brucei* had been introduced from an enzootic area of Nagana to a new locality free of tsetse flies, where the trypanosome was subsequently transmitted by mechanical inoculators, especially tabanid flies loss of contact with Glossina would eventually have brought about the transformation of *T. b. brucei* into a monomorphic race indistinguishable from the species now known as *T. evansi*” *(Hoare, 1972)*.

More recent genomic analysis of diverse isolates described as *T. evansi* and *T. equiperdum* highlighted that not only were these isolates subspecies of *T. brucei*, but the subspecies are polyphyletic, with each clade separated by a pleomorphic subspecies (Carnes et al., 2015; Cuypers et al., 2017; Lai et al., 2008). It was suggested that monomorphic clades are most accurately described as ‘types’; *T. b. equiperdum* type BoTat, *T. b. equiperdum* type OVI, *T. b. evansi* type A and *T. b. evansi* type B (Cuypers et al., 2017). Out of species, subspecies and type, type is the lowest rank and, informed by genotypic relationship, illustrates trait(s) that distinguish it. A recent paper has built on the arguments surrounding the phylogeny of *T. brucei* to propose a new definition based on ecotypes (Lukeš et al., 2022). Under the ecotype proposition, the clades of *T. brucei* would not be defined by genotypic data. They would be split into five ecotypes, *T. brucei f. brucei, T. b. f. gambiense, T. b. f. rhodesiense, T. b. f. equiperdum*, and *T. b. f. evansi.*

1.3.4. Surra, Dourine and parasite genotype

Due to the polyphyletic origins of monomorphic *T. brucei* isolates currently described under the *T. evansi* and *T. equiperdum* species, it is impossible to designate a causative species to the diseases Surra and Dourine. The situation is complicated further as isolates which are genotypically related to *T. b. evansi* isolates cause a disease which was diagnosed as

Dourine, historically associated with *T. b. equiperdum* (Oldrieve et al., 2021; Suganuma et al., 2016). Whilst whole genome sequencing can provide enough detail to disentangle the phylogenetic origins of an infecting monomorphic isolate, it is currently not viable in a field setting. Therefore, we suggested that the diseases Surra and Dourine be described as the clinical presentation of a disease with monomorphic *T. brucei* subspecies, without the use of *T. evansi* or *T. equiperdum* (Oldrieve et al., 2021).

The parasite genotype descriptors may become an ever more complicated topic as climate change continues to alter the range of the tsetse fly vector (de Clare Bronsvoort et al., 2013; Lord et al., 2018; Pagabeleguem et al., 2016). Once areas have a reduced, or absent, tsetse presence, there will be evolutionary pressure for pleomorphic *T. brucei* to adopt adaptations for mechanical transmission, such as the increased parasitaemia often associated with monomorphic *T. brucei*. Hence, climate change will likely lead to the emergence of new monomorphic clades, which will not fit in the current taxonomy and which have distinct origins from the currently identified clades.

1.3.5. Virulence

In the context of trypanosomes, virulence, the ability to cause disease in a host, is influenced by diverse factors from the genotype of the parasite to the host's immune status (Morrison et al., 2022). Virulence is often measured by the parasite proliferation rate in a host, but can also focus on host infectivity and transmissibility in the insect vector. In the mammalian host, *Trypanosoma* species undergo cell cycle arrest upon reaching a high parasitaemia, preventing the overproduction of the parasite population. This directly influences virulence, as unlimited parasite proliferation would rapidly prove lethal to the host. Therefore, there is a fine balance between increasing the chance of transmission, through increased population size and maintaining the host's viability (Black et al., 1983). As monomorphic *T. brucei*

isolates are unable to generate stumpy forms, they presumably have uncontrolled population growth to facilitate transmission, which increases their virulence. This description is usually made after studying *T. b. evansi* type A but is often generalised to monomorphic isolates as a whole (Auty et al., 2015; Schnaufer, 2010). *T. b. equiperdum* clades have also lost the ability to generate stumpy forms, however, the infection displays tropism for the genital mucosa, where it can persist for years before developing a neurological stage, and the associated lethal symptoms (Constable et al., 2017).

1.3.6. Life without a kinetoplast and the cause of monomorphism

Monomorphic *T. brucei* are often associated with a reduced or absent kinetoplast. *T. b, equiperdum* type OVI and *T. b, equiperdum* type BoTat isolates often lack minicircles, whilst retaining some maxicircles. *T. b. evansi* type A and *T. b. evansi* type B have generally lost their maxicircle DNA and have either completely removed their minicircles, or retained homogenous sequences (Lai et al., 2008; Schnaufer et al., 2002). Moreover, the presence of kDNA in monomorphic isolates can be extremely unstable (Suganuma et al., 2016). Due to the absence, or reduction, of kDNA in monomorphic isolates, they are unable to develop in the tsetse fly insect vector. ATP production requires oxidative phosphorylation in the tsetse fly, whilst bloodstream forms produce ATP via glycolysis, due to the abundance of blood glucose (Bringaud et al., 2006). However, the mitochondrion is still required in all life cycle stages to facilitate other cellular processes, such as fatty acid synthesis (Stephens et al., 2007).

Mitochondrial ATP synthesis is maintained via an electrochemical potential across the inner mitochondrial membrane (Chacinska et al., 2009). Bloodstream form *T. brucei* generates membrane potential via the mitochondrial F_1F_O–ATP synthase complex, which functions in reverse to pump protons to the intermembrane space (Nolan and Voorheis, 1992; Schnaufer

et al., 2005; Vercesi et al., 1992). This process is normally completed by the proton pumping complexes III and IV and proton translocation is achieved via subunit *a* of the F_O part of this complex. This subunit is kDNA encoded and requires RNA editing (Bhat et al., 1990). Monomorphic *T. brucei* cannot express the *a* subunit, as they lack either the gene or the guide RNAs required to edit the mRNA. However, the loss of kDNA can be entirely compensated for by a single point mutation in the γ subunit (Dean et al., 2013). Akinetoplastic parasites are unable to infect the tsetse fly (Dewar et al., 2022). However, monomorphism and the loss of kDNA have been uncoupled. In the laboratory, akinetoplastic parasites were able to differentiate from slender to stumpy forms at the same rate as wild-type parasites (Dewar et al., 2018; Timms et al., 2002).

T. b. evansi (unknown type) secrets orthologues of peptidases, Tb927.11.12850 (TevSTIB805.11_01.13260) Tb927.11.2500 (TevSTIB805.11_01.2570), which generate the oligopeptide signal (Moreira et al., 2022; Tettey et al., 2022). This suggests that in at least one clade of *T. b. evansi*, the parasites maintain the production of oligopeptides but there is a breakdown in the reception, or signal transmission in the QS pathway, which renders the parasites monomorphic. The occurrence of multiple monomorphic clades allows us to probe the consequences of a transition from a complex to a simple life cycle, along with the change in its life history. The transition from sexual reproduction to asexual reproduction leaves a clade-specific impact on genomic features (Jaron et al., 2021) which have been observed in asexually reproducing *T. brucei* (Oldrieve et al., 2021). A recent study highlighted one of these impacts. Wen et al., (2022) found that procyclin-associated gene 3 (PAG3) is absent or contains a frameshift in multiple naturally occurring monomorphic clades. Deletion or suppression of PAG3 produced developmentally incompetent *T. brucei* (Wen et al., 2022).

1.3.7. Laboratory-selected developmentally incompetent *T. brucei*

Laboratory strains of *T. brucei* have been maintained in culture, along with intermittent infections in mice and other mammals (Gibson, 2012). It is hard to track down the true origin of many commonly used laboratory strains (discussed here https://tryps.rockefeller.edu/trypsru2_pedigrees.html). Whilst it is possible to freeze parasite cultures, stopping the clock on the accumulation of molecular changes, the isolates studied today have undoubtedly changed from their wild ancestors (Mulindwa et al., 2018). The absence of transmission through a tsetse fly facilitates rapid proliferation and increased virulence in bloodstream form parasites (McWilliam et al., 2019). Ultimately, continual serial passage prevents the generation of the cell cycle-arrested stumpy forms (Ashcroft, 1960; Matthews et al., 2004; Turner, 1990). Commonly used laboratory lines, such as Lister 427, display these laboratory-adapted monomorphic features. Laboratory-generated monomorphs are unable to produce the stumpy form but the conditioned medium from these monomorphs is capable of inducing stumpy formation in pleomorphic cell lines (Vassella et al., 1997). Similarly, selected monomorphic *T. brucei* can be forced to differentiate upon exposure to the cell-permeable cAMP analogue pCPTcAMP (Breidbach et al., 2002) suggesting under normal circumstances mutations or gene regulation prevents signal reception or transduction through the QS signalling pathway, rather than a complete inability to differentiate (Silvester et al., 2017a; Vassella et al., 1997).

Cell cycle-arrested stumpy form parasites stop switching VSGs and laboratory-adapted cell lines are reported to be more antigenically stable than pleomorphic parasites (Turner and Barry, 1989). The low VSG switch rate in laboratory-adapted parasites is far lower than in pleomorphic isolates, potentially linking antigenic variation with transmissibility (Hovel-Miner et al., 2012; Lamont et al., 1986; Turner, 1997). However, through the generation of inducible and serially passaged monomorphism, switch rate frequency and developmental competence were uncoupled (McWilliam et al., 2019). Sequencing RNA from selected monomorphs highlighted genes associated with the transition from slender to stumpy and stumpy to procyclic form. Notably, ZC3H20 was significantly downregulated in selected

monomorphic cells (Nilsson et al., 2010). ZC3H20 is a key developmental regulator in the quorum sensing pathway, essential in stumpy formation (Cayla et al., 2020; Ling et al., 2011; Liu et al., 2020).

1.4. Aims

1.4.1. Significance and justification

We aim to highlight potential early and late events in the evolution of monomorphism, and how the complex QS pathway is constructed and can be perturbed and broken. The study also has the potential to provide molecular predictors for increased virulence associated with monomorphic *T. brucei* that could be used diagnostically in the field. Finally, we will discuss the evolutionary costs and benefits of monomorphism.

1.4.2. Aims

1) Confirm the evolutionary relationships between naturally occurring monomorphic isolates and highlight features of monomorphism at the genome level.
2) Identify mutations in naturally occurring monomorphic *T. brucei* isolate, which could give rise to developmental incompetence, and validate the role of these mutations in the laboratory.
3) Identify the early transcriptional events in the generation of monomorphism, by selecting for monomorphism with a pleomorphic cell line. Confirm the role of these transcriptional modifications via manipulation of the target transcripts.

2 Monomorphic Trypanozoon: towards reconciling phylogeny and pathologies

2.1. Preface

Identifying the genetic relationship between monomorphic *T. brucei* isolates is essential to identify heritable traits. Specifically, to identify the genetic basis of monomorphism, we needed to first categorise the relationship between each isolate of monomorphic *T. brucei* for which we had genome sequencing data (Oldrieve et al., 2021). This work led to the following publication.

Our paper has been cited in publications which reviewed the diagnosis of animal trypanosomes (Desquesnes et al., 2022a; Desquesnes et al., 2022b), designed an immunoassay for equine trypanosomiasis (Verney et al., 2022), studied the impact of ivermectin (Gupta et al., 2022), compared the risk factors of coinfection with *T. congolense* and *T. brucei* (Savage et al., 2021), identified a gene (PAG3) potentially involved in the generation of monomorphism (Wen et al., 2022), and discussed a proposal to rename the phylogeny of *T. brucei*, utilising ecotypes (Lukeš et al., 2022). The ecotype proposal will be discussed in detail in Chapter 3.

2.2. Monomorphic Trypanozoon: towards reconciling phylogeny and pathologies

MICROBIAL GENOMICS

SHORT COMMUNICATION

Oldrieve *et al.*, *Microbial Genomics* 2021;7:000632

DOI 10.1099/mgen.0.000632

Monomorphic *Trypanozoon*: towards reconciling phylogeny and pathologies

Guy Oldrieve[1], Mylène Verney[2], Kamil S. Jaron[3], Laurent Hébert[2] and Keith R. Matthews[1,*]

Abstract

Trypanosoma brucei evansi and *T. brucei equiperdum* are animal infective trypanosomes conventionally classified by their clinical disease presentation, mode of transmission, host range, kinetoplast DNA (kDNA) composition and geographical distribution. Unlike other members of the subgenus *Trypanozoon*, they are non-tsetse transmitted and predominantly morphologically uniform (monomorphic) in their mammalian host. Their classification as independent species or subspecies has been long debated and genomic studies have found that isolates within *T. brucei evansi* and *T. brucei equiperdum* have polyphyletic origins. Since current taxonomy does not fully acknowledge these polyphyletic relationships, we re-analysed publicly available genomic data to carefully define each clade of monomorphic trypanosome. This allowed us to identify, and account for, lineage-specific variation. We included a recently published isolate, IVM-t1, which was originally isolated from the genital mucosa of a horse with dourine and typed as *T. equiperdum*. Our analyses corroborate previous studies in identifying at least four distinct monomorphic *T. brucei* clades. We also found clear lineage-specific variation in the selection efficacy and heterozygosity of the monomorphic lineages, supporting their distinct evolutionary histories. The inferred evolutionary position of IVM-t1 suggests its reassignment to the *T. brucei evansi* type B clade, challenging the relationship between the *Trypanozoon* species, the infected host, mode of transmission and the associated pathological phenotype. The analysis of IVM-t1 also provides, to our knowledge, the first evidence of the expansion of *T. brucei evansi* type B, or a fifth monomorphic lineage represented by IVM-t1, outside of Africa, with important possible implications for disease diagnosis.

DATA SUMMARY

The data used in this study are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and the Wellcome Sanger Institute. The accession numbers can be found in Table S1 (available with the online version of this article).

INTRODUCTION

The subgenus *Trypanozoon* (*Trypanosoma brucei* spp.) contains parasites of medical, veterinary and economic significance. In their mammalian form, developmentally competent (pleomorphic) trypanosomes transition from a proliferative 'slender' form to a cell-cycle arrested 'stumpy' form, adapted for survival in the midgut of its vector, the tsetse fly (*Glossina* spp.). Progression to the stumpy form occurs in a density-dependent manner, mediated by a stumpy induction factor [1–4]. Some *Trypanozoon* have a reduced ability to transition from the slender to stumpy morphotype and so are described as 'monomorphic'. In the field, monomorphic *Trypanozoon* were historically classified as independent species, *T. equiperdum* and *T. evansi*, due to their distinct modes of transmission, geographical distribution, disease phenotype and host range [5]. These monomorphic trypanosomes can infect livestock, and are currently implicated in causing dourine and surra, respectively [6].

Received 14 April 2021; Accepted 11 June 2021; Published 16 August 2021

Author affiliations: [1]Institute for Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, UK; [2]Unité PhEED, Laboratoire de Santé Animale, Site de Normandie, ANSES, RD675, 1443012 Goustranville, France; [3]Institute of Evolutionary Biology, Ashworth Laboratories, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK.

***Correspondence:** Keith R. Matthews, keith.matthews@ed.ac.uk

Keywords: asexual; monomorphism; selection efficacy; *Trypanosoma brucei*.

Abbreviations: CDS, coding sequence; kDNA, kinetoplast DNA; NCBI, National Center for Biotechnology Information; SRA, Sequence Read Archive.

Raw reads generated as part of this project are available from the NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/PRJNA720808) (accession no. PRJNA720808).

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Three supplementary figures and three supplementary tables are available with the online version of this article.

000632 © 2021 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

More recently, it was proposed that *T. evansi* and *T. equiperdum* are subspecies of *T. brucei* (*T. brucei evansi* and *T. brucei equiperdum*) which had lost part or all of their kinetoplast DNA (kDNA), the parasites' mitochondrial genome that encodes respiratory components required for viability in the tsetse fly vector [7]. Whole-genome comparisons found that monomorphism arose independently on at least four separate occasions, and further monomorphic isolates could be continuously emerging from pleomorphic *T. brucei* in the field. However, *T. brucei evansi* and *T. brucei equiperdum* are polyphyletic and can be assigned into at least four independently derived lineages, such that their subspecific names do not describe the evolutionary relationships between the different monomorphic *Trypanozoon* [8–10]. *T. brucei evansi* type A and *T. brucei evansi* type B originate from *T. brucei* in Western and Central Africa, whilst *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat evolved from *T. brucei* in Eastern Africa [9]. Whilst many naming conventions exist, for the remainder of this article we will use the proposition by Cuypers *et al.* (2017) [9], which currently most accurately describes the polyphyletic nature of monomorphic trypanosomes (i.e. *T. brucei evansi* type A, *T. brucei evansi* type B, *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat).

The four *T. brucei* lineages converged on a monomorphic phenotype accompanied by a switch from cyclical to mechanical transmission. All of the monomorphic subspecies display a reduction or removal of their kDNA alongside an inability to complete their life cycle in their vector, locking these parasites into a tsetse-independent transmission mode [11, 12]. Current evidence suggests that *T. brucei evansi* type A and *T. brucei evansi* type B predominantly rely on transmission via biting flies (e.g. tabanids and *Stomoxys*), whilst *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat are sexually transmitted between Equidae. Neither are cyclically transmitted via the tsetse vector [6] and, as sexual reproduction occurs in the tsetse salivary gland [13], monomorphic trypanosomes are obligately asexual and proliferate via mitosis [14]. Escape from transmission by the tsetse fly, whose range is restricted to sub-Saharan Africa, has facilitated the expansion of monomorphic trypanosomes to other regions of Africa, Asia, Europe and the Americas, although they have subsequently been eradicated from North America and limited to local outbreaks in Europe [15].

The use of disease pathology, host species, geographical range and kDNA composition can complicate species classification where distinct lineages have converged on a phenotype. Research that treats polyphyletic lineages as a single group may miss more subtle, but important, differences between lineages. Here, we have re-analysed existing genomic data from publicly available monomorphic *Trypanozoon* isolates to confirm their evolutionary relationships. This analysis included a recent isolate from Mongolia, IVM-t1, derived from the genital mucosa of a horse and classified as *T. equiperdum* based on its clinical disease symptoms and host species [16, 17].

Impact Statement

Trypanosoma brucei are unicellular parasites typically transmitted by tsetse flies. Subspecies of *T. brucei* cause human African trypanosomiasis and the animal diseases nagana, surra and dourine. *T.brucei evansi* and *T. brucei equiperdum* have branched from *T. brucei* and, by foregoing tsetse transmission, expanded their geographical range beyond the sub-Saharan tsetse belt. These species can only reproduce asexually and exhibit morphological uniformity in their host ('monomorphism'). *T. brucei evansi* and *T. brucei equiperdum* have historically been classified based on fragmentary information on the parasites' transmission routes, geographical distribution, kinetoplast DNA (kDNA) composition and disease phenotypes. Our analysis of genome sequencing data from monomorphic *T. brucei* supports at least four independent origins with distinct evolutionary histories. One isolate, IVM-t1, typed as *T. equiperdum*, is a closer relative to *T. brucei evansi*, highlighting the risk of using pathognomonic descriptors for subspecies assignment. We show clear lineage-specific variation in the selection efficacy in monomorphic *T. brucei*. Using the evolutionary relationships between lineages, we suggest it would be beneficial to reconcile phylogeny and pathology in monomorphic trypanosomes.

Through re-analysing publicly available whole genome data, we found at least four groups of monomorphic *T. brucei* with independent origins, consistent with previously published phylogenies [8, 9]. Our analysis concludes that currently IVM-t1 forms a clade with *T. brucei evansi* type B, despite its clinical presentation being more typical of the conventional description of *T. equiperdum*. The presence of IVM-t1 in the genital mucosa of a horse with signs of dourine supports the hypothesis that there is considerable plasticity in the mode of transmission, host range and clinical presentation of monomorphic *Trypanozoon* strains of distinct origins, as was suggested by Brun *et al.* and Carnes *et al.* [8, 18]. These findings exemplify the need to classify monomorphic trypanosomes based on genetic information, unbiased by the mode of transmission, disease presentation, kDNA composition and host range. We also identified lineage-specific variation in the heterozygosity and efficacy of selection of the four independent monomorphic lineages, highlighting the importance of characterizing phylogeny-informed lineages. Finally, we note that the ancestor of the *T. brucei evansi* type B clade, or a fifth monomorphic lineage represented by IVM-t1, extended its geographical range outside of Africa.

METHODS

Variant calling

Publicly available genome data were accessed from the National Center for Biotechnology Information (NCBI)

Sequence Read Archive (SRA) [19] and the Wellcome Sanger Institute. Samples sequenced with older technologies, such as solid-state ABI, were excluded. When PacBio and Illumina data were available for the same sample, Illumina data were used preferentially to standardize the comparison.

The *T. brucei* EATRO 1125 AnTat 1.1 90:13 [20] genome was sequenced as part of this study. DNA was extracted using a DNeasy blood and tissue kit with an RNAse A step (Qiagen), following the manufacturer's instructions. The DNA was sequenced (HiSeq 4000) and cleaned by BGI (Hong Kong) (7723274 reads at 150 bp length). The *T. brucei* EATRO 1125 AnTat 1.1 90:13 raw data have been submitted to the NCBI SRA (PRJNA720808). The complete list of genomes analysed in this study, including their accession IDs, are summarized in the Table S1.

The quality of the raw reads was analysed with fastqc (v. 0.11.9) and subjected to quality trimming with trimmomatic (v0.39) [21]. The reads were trimmed with the following filters: SLIDINGWINDOW 4:20; ILLUMINACLIP adapters.fa 2:40:15; MINLEN 25. The trimmed reads were aligned to the *T. brucei* TREU927/4 V5 reference genome [22] with BWA-MEM (v. 0.7.17-r1188) [23]. The reads were prepared for variant calling by following the GATK4 (v. 4.1.4.1) best practices pipeline, which included marking duplicate reads [24, 25]. The read recalibration step was performed by initially calling variants on un-calibrated reads with GATK4's haplotype caller [26]. The top 20% highest confidence calls from the first round were used as the confident call set to re-calibrate the raw BAM files. Variants were then re-called on the recalibrated BAM files with GATK4's haplotype caller [26].

The variants were combined and filtered with the following stringent cut-offs, in keeping with GATK's best practices pipeline and previous studies [9]. SNPs were filtered by quality by depth (<2.0), quality score (<500.0), depth (<5.0), strand odds ratio (>3.0), Fisher's exact (>60.0), mapping quality (<40.0), mapping quality rank sum (<−12.5), read position rank sum (<−8.0), window size (10) and cluster size (3). Indels were filtered on their quality by depth (<2.0), quality (<500.0), Fisher's exact (>200.0) and read position rank sum (<−20.0).

Phylogenetic analysis

The filtered variants, described above, were filtered again to retain sites where a genotype had been called in every sample, using VCFtools (v. 0.1.16) [27]. These sites were analysed in two ways. The first was based on SNPs that occurred across all of the sites in the *T. brucei* TREU927/4 V5 reference genome and the second from SNPs that occurred in a coding sequence (CDS) of a gene (excluding pseudogenes) found on one of the 11 Mb chromosomes [22]. For both of these analyses, a concatenated alignment of each variant was extracted using VCF-kit (v. 0.1.6) [28]. IQ-TREE (v. 2.0.3) [29] was used to create a maximum-likelihood tree from homozygous variant sites. Within the IQ-TREE analysis, a best-fit substitution model was chosen by ModelFinder using models that included ascertainment bias correction (MFP+ASC) [30]. ModelFinder identified TVM+F+ASC+R2 as the best fit for both of the alignments which were subjected to 1000 ultrafast bootstraps generated by UFBoot2 [31]. The consensus trees were visualized and annotated using iTOL (v. 5) [32].

Genome content

Raw reads were used to predict the heterozygosity, G+C content and genome size of each isolate. This analysis was performed using a *k*-mer counting based approach with Jellyfish (v. 2.3.0) [33] and Genomescope (v. 1). A *k*-mer size of 21 was used [34].

Non-synonymous and synonymous SNP ratios (dN/dS) can be used to estimate the selection pressure upon an organism. The filtered SNPs identified for the phylogenetic analysis were split by isolate using bcftools (v. 1.9) [35], and these were then filtered to remove all non-variant sites using GATK4 SelectVariants (v. 4.1.9.0). The reverse complement of each variant call format (VCF) file was generated using SNPGenie (v. 2019.10.31) [36]. SNPGenie within-pool analysis was then completed on each isolate for SNPs found in the longest CDS site of each gene, excluding pseudogenes, on each of the 11 Mb chromosomes of the *T. brucei* TREU927/4 V5 reference genome [22]. The mean was calculated for each isolate. The full dN/dS results are available in Table S2.

Molecular markers

Molecular marker sequences were downloaded from the NCBI: SRA (Z37159.2), RoTat1.2 (AF317914), *T. evansi* VSG JN2118HU (AJ870487), cytochrome oxidase subunit 1 (CO1) (M94286.1:10712–12445) and NADH dehydrogenase subunit 4 (NADH4) (M94286.1:12780–14090). BWA-MEM (v. 0.7.17-r1188) [23] was used to align raw reads to the molecular markers. A minimum overlap of 50 bp between the read and target sequence was used. The molecular marker presence was confirmed by counting the number of bases in the target sequence covered by reads, representing the breadth of coverage, with SAMtools mpileup [37]. Orphan reads were counted and read pair overlap detection was disabled. The breadth of coverage percentage was visualized with pheatmap (v. 1.0.12) [38]. The ATP synthase γ subunit (Tb927.10.180) was screened for lineage-specific variants identified in the variant calling step. Unless stated otherwise, all figures were plotted using ggplot2 (v 3.3.0) [39] and ggrepel (v. 0.8.1) [40] in R (v. 3.6.1) [41].

RESULTS

Monomorphism arose independently at least four times

To determine the evolutionary relationships between monomorphic *Trypanozoon*, genomic data from 17 isolates were aligned to the *T. brucei* TREU 927/4 reference genome. This included publicly available data from monomorphic (*n*=9) and pleomorphic isolates (*n*=8). Across the *T. brucei* TREU 927/4 v5 reference genome, 472794 variant sites (574775 unique variant alleles, 370154 SNPs and 204621 indels) passed the strict quality filtering steps. These sites were filtered further to identify 244013 homozygous variant SNPs across

the genome. A total of 91853 of these SNPs were present in a CDS on one of the 11 Mb chromosomes of the *T. brucei* TREU 927/4 v5 reference genome. The SNPs across the whole genome and those in a CDS were used to generate two unrooted phylogenetic trees.

Monomorphic *T. brucei* form at least four independent clades (Fig. 1). Our results corroborate previous findings [8, 9] which identified that the *T. brucei equiperdum* type OVI clade arose in Eastern Africa and displays minimal variation between the isolates. *T. brucei equiperdum* type BoTat is separated from other *T. brucei equiperdum* isolates and represents an Eastern African isolate of distinct origin. Trypanosomes designated *T. brucei evansi* also form two discrete clades of Western and Central African origin, with STIB805, RoTat1.2 and MU09 (*T. brucei evansi* type A) displaying low genetic diversity and being distinct from MU10 (*T. brucei evansi* type B).

Interestingly, the Mongolian isolate IVM-t1, with infection and disease characteristics similar to *T. brucei equiperdum*, branched with MU10, which is of West African origin. This contrasts with its previous designation as a *T. equiperdum* isolate, whose ancestors originated from Eastern Africa [9] (Fig. 1). Nonetheless, whilst IVM-t1 and MU10 shared a more recent common ancestor with each other than with any other strain in this analysis, MU10 and IVM-t1 have diverged considerably. Should pleomorphic *T. brucei* isolates be identified that divide the clade composed of MU10 and IVM-t1, the isolates would represent independent clades. The results generated from SNPs found across the whole genome were similar to a tree built from SNPs found in only the CDS, with the former displaying a slight reduction in bootstrap confidence (Fig. S1).

There is quantifiable variation in the genetic diversity and efficacy of selection between the four asexual monomorphic clades

Asexuality is expected to reduce the heterozygosity of a lineage, although this can be lineage specific [42]. Asexuality can also reduce the efficacy with which selection can act, as reviewed by Otto in 2021 [43]. Efficacy of selection can be estimated by calculating the ratio of nonsynonymous to synonymous variants (dN/dS) compared to a reference genome.

Our analysis demonstrated that monomorphic genomes generally have a lower heterozygosity compared to the genomes derived from pleomorphic isolates (Figs 2 and S2). Notable outliers in this analysis include MU10 (*T. brucei evansi* type B), which has a heterozygosity value closer to that of pleomorphic genomes. The higher dN/dS ratio observed in *T. brucei equiperdum* type OVI is in contrast to the other monomorphic lineages, highlighting the lineage-specific variation, and evolutionary histories, of monomorphic *Trypanozoon*.

Interestingly, IVM-t1 has the second lowest heterozygosity and one of the lowest dN/dS ratios in this analysis, more similar to that of *T. brucei evansi* type A and *T. brucei evansi* type B, and in contrast to the levels observed in *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat (Fig. 2). Raw reads from *T. brucei brucei* TREU 927/4 were included in the analysis. The variants called against its own genome assembly represent heterozygous loci and misaligned reads. *T. brucei equiperdum* BoTat has the longest branch length in this analysis, but has a far lower dN/dS ratio than the *T. brucei equiperdum* type OVI clade (Fig. S3). The low dN/dS of *T. brucei brucei* TREU 927/4 called against its own genome and the pattern of branch length and dN/dS ratio highlights that the high dN/dS of *T. brucei equiperdum* type OVI is not the result of the evolutionary distance between the clade and the reference genome.

Genetic markers corroborate phylogenetic and genome content analysis

The occurrence of established molecular markers for different *T. brucei* subspecies was compared between the isolates to further validate whole-genome analysis. The markers' occurrence in each isolate was visualized in a hierarchical clustered heatmap. The clustering highlights a clear distinction between the monomorphic lineages and, furthermore, with just five molecular markers, it is possible to recreate a similar pattern to the phylogenetic tree that is based on whole-genome sequencing data (Figs 1 and 3).

As our phylogenetic and genome content analysis identified a potential discrepancy between the disease description of IVM-t1 and its genotype, particular interest was paid to this isolate. Firstly, being akinetoplastic at the point of sequencing, IVM-t1 lacks coverage of the mitochondrial maxicircle genes cytochrome oxidase subunit 1 (CO1) (M94286.1:10712–12445) and NADH dehydrogenase subunit 4 (NADH4) (M94286.1 : 12780-14090), as do *T. brucei evansi* type A and *T. brucei evansi* type B. However, when IVM-t1 was initially isolated, a PCR for NADH4 found that it contained the gene, albeit with a faint signal [16]. Therefore, maxicircle presence in IVM-t1 appears to have been unstable, as is the case in kDNA independent isolates. IVM-t1 reportedly became akinetoplastic after long-term-culture adaptation, as has been observed in other monomorphic isolates [44, 45]. In contrast to IVM-t1, CO1 and NADH4 are present in *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat genomes.

Secondly, IVM-t1 lacks the RoTat 1.2 VSG (AF317914.1), which is diagnostic for *T. brucei evansi* type A [46], but does have sequence coverage of VSG JN2118HU (AJ870487.1) that is present in *T. brucei evansi* type B, along with some *T. brucei brucei* strains [47, 48]. In contrast, JN2118HU is absent in *T. brucei evansi* type A, *T. brucei equiperdum* type OVI *and T. brucei equiperdum* type BoTat (Fig. 3).

Thirdly, IVM-t1 does not have the M282L ATP synthase γ subunit mutation that has been characterized in other *T. brucei evansi* type B genomes, such as MU10 (Table S3) [48, 49]. This mutation is unable to compensate for kDNA loss [49]. However, IVM-t1 does have two homozygous mutations within that gene that are absent from all other isolates in this

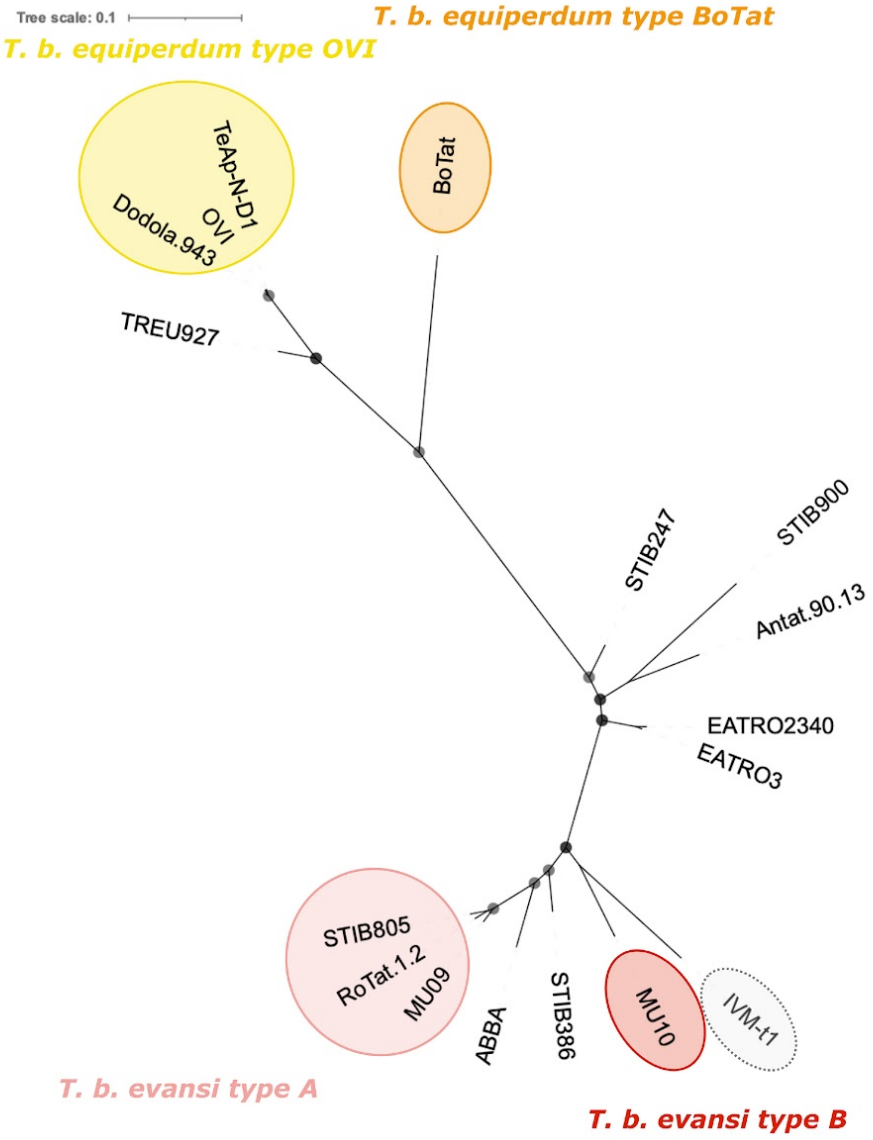


Fig. 1. An unrooted phylogenetic tree created with 244013 homozygous variant SNPs found across the *T. brucei* TREU 927/4 reference genome. The tree was built using a TVM+F+ASC+R2 model. Bootstrap confidence is reported by the size of the grey circles; all bootstrap values were 100 and so each circle is the same size. The branch length scale represents the number of substitutions per site. Monomorphic genomes form four distinct lineages that have expanded from Eastern (*T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat) and Western/Central Africa (*T. brucei evansi* type A and *T. brucei evansi* type B) [9]. IVM-t1 was originally typed as *T. equiperdum*, but groups here with *T. brucei evansi* type B.

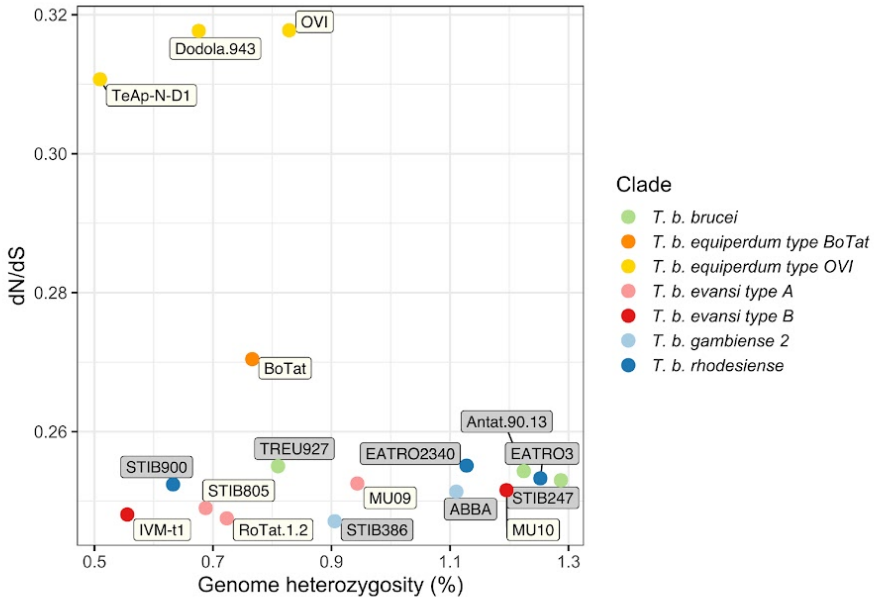


Fig. 2. Whole-genome heterozygosity and dN/dS ratio of SNPs present in the longest CDS of every annotated gene, excluding pseudogenes, found on one of the 11 Mb chromosomes of the *T. brucei* TREU 927/4 reference genome. The values were calculated for all publicly available monomorphic isolates and representative pleomorphic isolates. Each point is coloured by clade and the label colour represents a pleomorphic (grey) or monomorphic (white) isolate.

analysis (C-terminal genomic codons G817T and A898G), although both are synonymous and so predicted not to influence protein function. In combination, these analyses further support the separation of IVM-t1 from the *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat lineages, and also highlights differences to *T. brucei evansi* MU10 lineage despite the phylogenetic relationship between the isolates.

DISCUSSION

Trypanozoon phylogeny has historically been based on clinical disease pathology, mode of transmission, geographical range, host species range and kDNA composition, which has complicated the classification of monomorphic trypanosomes and fuelled a long-standing debate in the literature [5]. Here, we re-analyse the molecular phylogeny of monomorphic trypanosomes, providing support for the complete separation of these isolates into at least four clades based on their evolutionary relationships. Our results are consistent with previously published phylogenies of monomorphic *T. brucei* subspecies (Fig. 1) [8, 9]. We consider it is important to classify the monomorphic lineages through their evolutionary relationships because, although they have converged upon a monomorphic phenotype, lineage-specific variation could be missed if their polyphyletic origin is not fully acknowledged. Expanding the repertoire of publicly available genomic data for monomorphic *T. brucei* will negate the limitations of this study.

Variation in heterozygosity can be associated with a transition to asexuality. Asexual taxa can present high heterozygosity if the lineage arose from a hybrid origin, but other types of origin usually lead to a reduction in heterozygosity [42]. Monomorphic strains have lost their tsetse transmission ability. Given that meiotic events occur in tsetse salivary glands, monomorphic strains are obligately asexual and proliferate via mitosis [13, 14]. This asexuality is apparent in the reduction of heterozygosity observed in the majority of monomorphic isolates (Figs 2 and S2). Such reduction in heterozygosity can occur via mitotic recombination and/or gene conversion, with gene conversion having been observed in another asexual *T. brucei* subspecies, *T. brucei gambiense* group 1 [50]. Gene conversion is proposed to reduce or even completely stop the mutational attrition associated with asexuality, and it can occur at a faster rate than the accumulation of spontaneous mutations [51] (Figs 2 and S2).

Notably, *T. brucei equiperdum* type OVI and, to an extent, *T. brucei equiperdum* type BoTat isolates have low heterozygosity

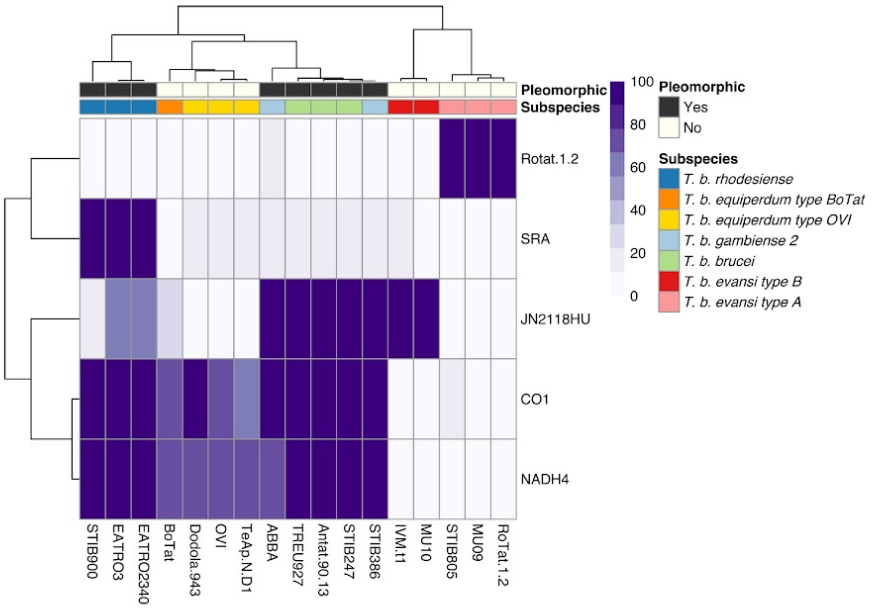


Fig. 3. Occurrence of individual genetic markers corroborates the phylogenetic and whole-genome analysis, which highlights at least four independent monomorphic lineages. Genetic markers: SRA (Z37159.2), RoTat1.2 VSG (AF317914.1), JN2118HU VSG (AJ870487.1), cytochrome oxidase subunit 1 (M94286.1: 10712–12445) and NADH4 (M94286.1: 12780–14090). The scale represents the percentage of the marker covered by sequencing reads.

and a higher dN/dS ratio than other clades, indicative of a smaller efficacy of purifying selection in removing deleterious alleles. In contrast, the efficacy of selection in *T. brucei evansi* type A and *T. brucei evansi* type B are closer to that observed in pleomorphic lineages. Further, IVM-t1, originally typed as *T. equiperdum*, has one of the lowest dN/dS ratios, more indicative of *T. brucei evansi* type A or *T. brucei evansi* type B. The efficacy of selection can be influenced by the effective population size, which is linked to events during the evolutionary history of a lineage, such as population bottlenecks and variation in the mode of inheritance [52]. Asexuality is predicted to reduce the efficacy with which selection can act [43]. Hence, the observed lineage-specific variation between selection efficacy could be associated with a different length of time as an asexual lineage as the predicted build-up of deleterious mutations is a gradual process, if not completely counteracted by processes like gene conversion. Overall, the analyses of heterozygosity and dN/dS ratio support the different monomorphic lineages displaying contrasting evolutionary histories.

The extremely low heterozygosity of IVM-t1 highlights that when the sample was sequenced, it did not comprise a mixed infection. Furthermore, its low heterozygosity does not support a hybridization-based event at the emergence of the IVM-t1 lineage [42]. Therefore, the long branch and discrepancy in heterozygosity between IVM-t1 and MU10 could be due to an expansion in diversity of the MU10 branch or independent origins of monomorphism within the MU10/IVM-t1 clade. This divergence could have facilitated the distinct transmission mechanism and host range displayed by IVM-t1 with respect to *T. brucei evansi* type B such as MU10. However, at present IVM-t1 and MU10 group as a separate clade and share the monomorphic phenotype. As such, IVM-t1 is currently most accurately described as *T. brucei evansi* type B. As more *T. brucei* are isolated and sequenced, it may be more accurate to define IVM-t1 as a separate clade. In this case, IVM-t1 would represent a fifth independent emergence of monomorphism in *T. brucei*.

The potential expansion of monomorphic lineages, along with the isolation of IVM-t1 from the genital mucosa of a horse with signs of dourine, suggest it could be beneficial to reconcile phylogeny and disease [16]. To fully uncouple the link between phylogeny and disease, studies will be required on the direct mode of transmission of these isolates. For instance, although the presence of IVM-t1 in the genital mucosa of a horse with signs of dourine points towards sexual transmission, it cannot be ruled out that the initial infection was a coinfection of IVM-t1 with an independent

T. brucei equiperdum type OVI or *T. brucei equiperdum* type BoTat isolate which caused the dourine symptoms but was not recovered after culture. If plasticity in the mode of transmission is established, dourine and surra would best refer to the disease presentation and not the causative agent, particularly where limited clinical information is available for an isolate, precluding an understanding of any variable disease manifestation between individual animals [6].

The emergence of at least four independent monomorphic lineages suggests there could be a selective advantage to monomorphism, at least in the short term. Since monomorphic lineages lose the growth control inherent in the generation of stumpy forms, they display an increase in parasitaemia that improves the chance of non-tsetse transmission when tsetse vectorial capacity is reduced. This adaption to a loss of cyclical transmission could also be the only option for those *Trypanozoon* isolates that become isolated from the tsetse fly through environmental change or geographical relocation of the host [7, 53, 54]. Regardless of the potential adaptive advantage to monomorphism, monomorphic lineages could be constantly emerging from pleomorphic *T. brucei* populations across Africa that remain unidentified due to a lack of sampling. Indeed, as climate change rapidly alters the tsetse flies' range [55], creating a potential selective advantage for mechanical transmission in areas where tsetse flies are no longer found, the rate of monomorphic *T. brucei* subspecies emergence could increase. Since monomorphic lineages have lost or reduced their growth control mechanism, they can be highly virulent, posing a threat to livestock in their country of origin and with the risk of escape outside traditional disease boundaries.

The analysis of IVM-t1 provides. to our knowledge, the first evidence that *T. brucei evansi* type B, or a fifth monomorphic lineage, has expanded its geographical range outside of Africa. Previously *T. brucei evansi* type B has only been found in Kenya and Ethiopia [47, 48, 56]. The presence of IVM-t1 outside of Africa could complicate the existing screening of animals exhibiting signs of dourine or surra. Currently, the molecular diagnosis of surra and dourine remains limited by the parasitaemia in infected hosts, which can be below the detection limit of parasitological tests and can even be below the detection limit of DNA tests, especially in Equidae and African cattle. Therefore, serological methods are prescribed by the World Organisation for Animal Health for surra and dourine diagnosis. In some regions and for some hosts, for which *T. brucei evansi* type A strains are widely present, the use of serological tests based on the recognition of the specific RoTat 1.2 VSG can provide good sensitivity and specificity [46]. However, this gene is absent in *T. brucei evansi* type B (Fig. 3). Alternatively, surveillance of kDNA integrity remains a useful method of identification, which often aligns with the disease presentation despite these phenomena not being biologically linked. It is important to remember, however, that kDNA integrity comes with the risk of the independent appearance of dyskinetoplasty in multiple lineages or the spontaneous or progressive dyskinetoplasty observed after maintenance *in vitro* culture or in isolates indistinguishable at the level of the nuclear genome.

To avoid these complications, it is possible to rely on markers that are generic for all trypanosome subspecies. Indeed, given that treatment success depends mainly on the stage of the disease rather than the specific *Trypanozoon* lineage [6], we suggest that currently it remains preferable not to aim for a distinction between taxa within the subgenus *Trypanozoon* for first line diagnosis. However, genome sequencing is rapidly reducing in cost whilst improving in portability. Therefore, adaptive genome sequencing represents a promising method for screening of animals infected with monomorphic *Trypanozoon* [57].

We note that although the use of four clades (*T. brucei evansi* type A, *T. brucei evansi* type B, *T. brucei equiperdum* type OVI and *T. brucei equiperdum* type BoTat [9]) attempts to acknowledge the polyphyletic origin of monomorphic *Trypanozoon*, the use of *evansi* or *equiperdum* at the subspecific level is not based on the genetic relationship of the strains. This is the case even when phylotypes are utilized after *evansi* or *equiperdum*. Therefore, we suggest it would be beneficial to break the links between the fragmentary information available for taxonomy, disease phenotype, host range, mode of transmission and the extent of dyskinetoplasty in monomorphic *Trypanozoon*.

Instead, to fully acknowledge their polyphyletic origin and distinct evolutionary histories, we suggest the taxonomy of monomorphic *Trypanozoon* should be based on whole-genome analysis alone, which is quantitative, non-subjective and can be assessed from samples lacking detailed case records. A taxonomy based on the evolutionary relationships between isolates will assist future research by identifying lineage-specific variation in monomorphic *Trypanozoon*. For example, we consider IVM-t1 is currently more accurately classified as a branch of *T. brucei evansi* type B, rather than *T. equiperdum*. Similarly, isolates such as STIB818 (isolated China in 1979) and ATCC 30023 (isolated in France in 1903) were initially classified as *T. equiperdum* but cluster with *T. brucei evansi* [8, 58]. The disease manifestation and tissue specificity of IVM-t1 also suggests that *T. brucei evansi* type B, or a fifth monomorphic lineage, can manifest dourine symptoms with sexual transmission, although direct transmission evidence is needed to confirm this.

The incongruity between the parasites' evolutionary position and the induced pathology, mode of transmission and tissue tropism highlights the potential for the host physiology and immune response to contribute to clinical disease manifestation, rather than being solely parasite driven [6]. This provides an important exemplar of the potential distinction between the taxonomic position of monomorphic trypanosomes and the diseases they cause.

Funding information
This research was funded in part by the Wellcome Trust (grant no. 103740/Z/14/Z and no. 108905/B/15/Z). For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. K. R. M. was supported by a Wellcome Trust Investigator award grant (103740/Z/14/Z) and G. O. by a Wellcome Trust PhD studentship (108905/B/15/Z). L. H. and M. V. were supported by ANSES, the

European Commission through DG SANTE funding for the Reference Laboratory for Equine Diseases Other than African Horse Sickness, the Regional Council of Normandy, and the GIS Centaure Recherche Equine. K. S. J. was supported by funding from the European Research Council Starting Grant (PGErepo) awarded to Laura Ross.

Acknowledgements
We thank Achim Schnaufer for his valuable insights and comments on the manuscript.

Author contributions
Conceptualization: G. O., K. R. M. Methodology: G. O., K. S. J. Software: G. O. Validation: G. O. Formal analysis: G. O. Investigation: G. O. Resources: G. O., K. R. M. Data Curation: G. O. Writing – original draft preparation: G. O. Writing – review and editing: G. O., K. R. M., M. V., L. H., K. S. J. Visualization: G. O., K. R. M., M. V., L. H., K. S. J. Supervision: K. R. M. Project administration: G. O., K. R. M. Funding: K. R. M.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. **Vassella E, Reuner B, Yutzy B, Boshart M**. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci* 1997;110:2661–2671.
2. **Mony BM, MacGregor P, Ivens A, Rojas F, Cowton A, *et al.*** Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. *Nature* 2014;505:681–685.
3. **McDonald L, Cayla M, Ivens A, Mony BM, MacGregor P, *et al.*** Non-linear hierarchy of the quorum sensing signalling pathway in bloodstream form African trypanosomes. *PLoS Pathog* 2018;14:e1007145.
4. **Rojas F, Silvester E, Young J, Milne R, Tettey M, *et al.*** Oligopeptide signaling through TbGPR89 drives *Trypanosome quorum* sensing. *Cell* 2019;176:306-317.
5. **Hoare CA**. *The Trypanosomes of Mammals: a Zoological Monograph*. Oxford: Blackwells; 1972.
6. **Büscher P, Gonzatti MI, Hébert L, Inoue N, Pascucci I, *et al.*** Equine trypanosomosis: enigmas and diagnostic challenges. *Parasit Vectors* 2019;12:234.
7. **Lai DH, Hashimi H, Lun ZR, Ayala FJ, Lukes J**. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. *Proc Natl Acad Sci U S A* 2008;105:1999–2004.
8. **Carnes J, Anupama A, Balmer O, Jackson A, Lewis M, *et al.*** Genome and phylogenetic analyses of *Trypanosoma evansi* reveal extensive similarity to *T. brucei* and multiple independent origins for dyskinetoplasty. *PLoS Negl Trop Dis* 2015;9:e3404.
9. **Cuypers B, Van den Broeck F, Van Reet N, Meehan CJ, Cauchard J, *et al.*** Genome-wide SNP analysis reveals distinct origins of *Trypanosoma evansi* and *Trypanosoma equiperdum*. *Genome Biol Evol* 2017;9:1990–1997.
10. **Kay C, Williams TA, Gibson W**. Mitochondrial DNAs provide insight into trypanosome phylogeny and molecular evolution. *BMC Evol Biol* 2020;20:161.
11. **Schnaufer A, Domingo GJ, Stuart K**. Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *Int J Parasitol* 2002;32:1071–1084.
12. **Dewar CE, MacGregor P, Cooper S, Gould MK, Matthews KR, *et al.*** Mitochondrial DNA is critical for longevity and metabolism of transmission stage *Trypanosoma brucei*. *PLoS Pathog* 2018;14:e1007195.
13. **Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, *et al.*** Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. *Proc Natl Acad Sci USA* 2011;108:3671–3676.
14. **Wheeler RJ, Scheumann N, Wickstead B, Gull K, Vaughan S**. Cytokinesis in *Trypanosoma brucei* differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis. *Mol Microbiol* 2013;90:1339–1355.
15. **Radwanska M, Vereecke N, Deleeuw V, Pinto J, Magez S**. Salivarian trypanosomosis: a review of parasites involved, their global distribution and their interaction with the innate and adaptive mammalian host immune system. *Front Immunol* 2018;9:2253.
16. **Suganuma K, Narantsatsral S, Battur B, Yamasaki S, Otgonsuren D, *et al.*** Isolation, cultivation and molecular characterization of a new *Trypanosoma equiperdum* strain in Mongolia. *Parasit Vectors* 2016;9:481.
17. **Davaasuren B, Yamagishi J, Mizushima D, Narantsatsral S, Otgonsuren D, *et al.*** Draft genome sequence of *Trypanosoma equiperdum* strain IVM-t1. *Microbiol Resour Announc* 2019;8:e01119-18.
18. **Brun R, Hecker H, Lun ZR**. *Trypanosoma evansi* and *T. equiperdum*: distribution, biology, treatment and phylogenetic relationship (a review). *Vet Parasitol* 1998;79:95–107.
19. **Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration**. The Sequence Read Archive. *Nucleic Acids Res* 2011;39:D19–D21.
20. **Engstler M, Boshart M**. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. *Genes Dev* 2004;18:2798–2811.
21. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
22. **Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, *et al.*** The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005;309:416–422.
23. **Li H**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013.
24. **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, *et al.*** A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–498.
25. **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, *et al.*** From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;11:11.
26. **Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, *et al.*** Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv* 2017.
27. **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al.*** The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–2158.
28. **Cook DE, Andersen EC**. VCF-kit: assorted utilities for the variant call format. *Bioinformatics* 2017;33:1581–1582.
29. **Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ**. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
30. **Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
31. **Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS**. Ufboot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
32. **Letunic I, Bork P**. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007;23:127–128.
33. **Marçais G, Kingsford C**. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 2011;27:764–770.
34. **Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, *et al.*** GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 2017;33:2202–2204.
35. **Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, *et al.*** BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 2016;32:1749–1751.
36. **Nelson CW, Moncla LH, Hughes AL**. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* 2015;31:3709–3711.

37. **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
38. **Kolde R**. Pheatmap: Pretty Heatmaps. R package; 2019.
39. **Wickham H**. ggplot2: Elegant Graphics for Data Analysis; 2016.
40. **Slowikowski K**. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2', version 08. R Package; 2018.
41. **R Core Team**. R: A language and environment for statistical computing. 2019.
42. **Jaron KS**, **Bast J**, **Nowell RW**, **Ranallo-Benavidez TR**, **Robinson-Rechavi M**, *et al*. Genomic features of parthenogenetic animals. *J Hered* 2020;112:19–33.
43. **Otto SP**. Selective interference and the evolution of sex. *J Hered* 2021;112:9–18.
44. **Zweygarth E**, **Kaminsky R**, **Webster P**. *Trypanosoma brucei* evansi: dyskinetoplasia and loss of infectivity after long-term *in vitro* cultivation. *Acta Trop* 1990;48:95–99.
45. **Kaminsky R**, **Schmid C**, **Lun ZR**. Susceptibility of dyskinetoplastic *Trypanosoma evansi* and *T. equiperdum* to isometamidium chloride. *Parasitol Res* 1997;83:816–818.
46. **Claes F**, **Radwanska M**, **Urakawa T**, **Majiwa PA**, **Goddeeris B**, *et al*. Variable surface glycoprotein RoTat 1.2 PCR as a specific diagnostic tool for the detection of *Trypanosoma evansi* infections. *Kinetoplastid Biol Dis* 2004;3:3.
47. **Ngaira JM**, **Olembo NK**, **Njagi ENM**, **Ngeranwa JJN**. The detection of non-RoTat 1.2 *Trypanosoma evansi*. *Exp Parasitol* 2005;110:30–38.
48. **Birhanu H**, **Gebrehiwot T**, **Goddeeris BM**, **Büscher P**, **Van Reet N**. New *Trypanosoma evansi* type B isolates from Ethiopian dromedary camels. *PLoS Negl Trop Dis* 2016;10:e0004556.
49. **Dean S**, **Gould MK**, **Dewar CE**, **Schnaufer AC**. Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes. *Proc Natl Acad Sci USA* 2013;110:14741–14746.
50. **Weir W**, **Capewell P**, **Foth B**, **Clucas C**, **Pountain A**, *et al*. Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *Elife* 2016;5:e11473.
51. **Marais GAB**, **Campos PRA**, **Gordo I**. Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol Evol* 2010;2:347–357.
52. **Charlesworth B**. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 2009;10:195–205.
53. **Jensen RE**, **Simpson L**, **Englund PT**. What happens when *Trypanosoma brucei* leaves Africa. *Trends Parasitol* 2008;24:428–431.
54. **Schnaufer A**. Evolution of dyskinetoplastic trypanosomes: how, and how often? *Trends Parasitol* 2010;26:557–558.
55. **Lord JS**, **Hargrove JW**, **Torr SJ**, **Vale GA**. Climate change and African trypanosomiasis vector populations in Zimbabwe's Zambezi Valley: a mathematical modelling study. *PLoS Med* 2018;15:e1002675.
56. **Aregawi WG**, **Agga GE**, **Abdi RD**, **Büscher P**. Systematic review and meta-analysis on the global distribution, host range, and prevalence of *Trypanosoma evansi*. *Parasit Vectors* 2019;12:67.
57. **Payne A**, **Holmes N**, **Clarke T**, **Munro R**, **Debebe B**, *et al*. Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. *BioRxiv* 2020;Feb 3.
58. **Claes F**, **Büscher P**, **Touratier L**, **Goddeeris BM**. *Trypanosoma equiperdum*: master of disguise or historical mistake? *Trends Parasitol* 2005;21:316–321.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.

3 Genomic features of monomorphic *T. brucei*

3.1. Introduction

Categorising *T. brucei* isolates by their genetic relationship is essential for downstream functional validation, exemplified by the misidentified *T. b. evansi* type IVM-t1 isolate (Oldrieve et al., 2021; Suganuma et al., 2016). We were kindly gifted whole genome sequencing data of monomorphic isolates by Frederik Van Den Broeuk (KMU, Leuven) which we used to extend our analysis from Chapter 2.

Recent work sequenced the transcriptomes of several naturally occurring monomorphic *T. brucei* isolates and found that “transcriptional patterns in *T. evansi* and *T. equiperdum* were seen to be diverse” (Cai et al., 2021). In the analysis, the isolates were typed as “*T. evansi*” or “*T. equiperdum*”, without a distinction between the polyphyletic monomorphic clades. Alongside this, a recent proposal suggested that the *T. brucei* subspecies could be redefined to utilise ‘ecotypes’ (Lukeš et al., 2022). In the context of *T. brucei* phylogeny, they define ecotypes as “a group of organisms within a species that is adapted to particular environmental conditions and therefore exhibit genetic and often behavioural, structural, or physiological differences from other species members” (Lukeš et al., 2022). Under this proposition, *T. b. evansi* type A and *T. b. evansi* type B would be defined as *T. b. f. evansi,* whereas *T. b. evansi* type IVM-t1, *T. b. equiperdum* type BoTat and *T. b. equiperdum* type OVI would be defined as *T. b. f. equiperdum*.

Here, we aimed to identify the relationship between the monomorphic isolates for which we were gifted genome data and the monomorphic isolates which had diverse mRNA profiles (Cai et al., 2021). Monomorphism causes a reversion to an obligate monoxenous asexual life cycle which is predicted to reduce the efficacy of selection and genetic diversity. Informed by phylogeny, we aimed to identify genome-wide features of monomorphism. Our results are placed in the context of the ecotype proposition, challenging its use in the context of

polyphyletic clades. Combining these insights, we highlight the presence of mutations in monomorphic clades and prioritise these mutations to explore the putative causes of monomorphism for each independent clade by experimental analysis.

3.2. Methods

Each tool used in the following analysis, plus the version and its optional flags, can be found in Table 3.1. Unless stated otherwise, all figures were plotted using ggplot2 (Wickham, 2016) and ggrepel (available at https://cran.r-project.org/web/packages/ggrepel/index.html) in R (R Core Team, 2022).

3.2.1. Variant calling

Genome data from 31 isolates were gifted by Frederik Van Den Broeuk (KMU, Leuven). These were added to the genome data used in Chapter 2. Mutations were called across the *T. b. brucei* TREU927/4 reference genome, as described in Chapter 2. Briefly, the quality of the raw reads was analysed with fastqc (Andrews and Others, 2010) and subjected to quality trimming with trimmomatic (Bolger et al., 2014). The reads were trimmed with the following filters: SLIDINGWINDOW 4:20; ILLUMINACLIP adapters.fa 2:40:15; MINLEN 25. The trimmed reads were aligned to the *T. brucei* TREU927/4 V5 reference genome with bwa-mem (Li and Durbin, 2009). The reads were prepared for variant calling by following the gatk4 best practices pipeline, which included marking duplicate reads (DePristo et al., 2011; Van der Auwera et al., 2013). The read recalibration step was performed by initially calling variants on un-calibrated reads with gatk4's haplotype caller (Poplin et al., 2018). The top 20% highest confidence calls from the first round were used as the confident call set to re-calibrate the raw bam files. Variants were then re-called on the recalibrated bam files with gatk4's haplotype caller (Poplin et al., 2018).

The variants were combined and filtered with the following stringent cut-offs, in keeping with gatk's best practices pipeline and previous studies (Cuypers et al., 2017). SNPs were filtered

by quality by depth (<2.0), quality score (<500.0), depth (<5.0), strand odds ratio (>3.0), Fisher's exact (>60.0), mapping quality (<40.0), mapping quality rank sum (<−12.5), read position rank sum (<−8.0), window size (10) and cluster size (3). Indels were filtered on their quality by depth (<2.0), quality (<500.0), Fisher's exact (>200.0) and read position rank sum (<−20.0). The complete list of genomes analysed in this study, including their accession IDs, is summarised in Table S3.1.

3.2.2. Phylogenetic analysis

A phylogenetic tree was generated as described in Chapter 2. The filtered variants, described above, were filtered again to retain sites where a genotype had been called in every sample, using VCFtools (Danecek et al., 2011). SNPs were identified across all of the sites in the *T. brucei* TREU927/4 V5 reference genome (Berriman et al., 2005). A concatenated alignment of each variant was extracted using VCF-kit (Cook and Andersen, 2017). IQ-TREE (Nguyen et al., 2015) was used to create a maximum-likelihood tree from homozygous variant sites. Within the IQ-TREE analysis, a best-fit substitution model was chosen by ModelFinder using models that included ascertainment bias correction (MFP+ASC) (Kalyaanamoorthy et al., 2017). ModelFinder identified TVM+F+ASC+R4 as the best fit for both of the alignments which were subjected to 1000 ultrafast bootstraps generated by UFBoot2 (Hoang et al., 2018). The consensus trees were visualized and annotated using iTOL (Letunic and Bork, 2007).

3.2.3. Genome content

Genome summary statistics, including heterozygosity, genome repeat length and genome unique length, were calculated by k-mer counting with jellyfish (Marçais and Kingsford, 2011)

and GenomeScope2 (Vurture et al., 2017). A k-mer length of 21 was used. The inbreeding coefficient (F) was calculated for each isolate using vcftools (Danecek et al., 2011) and the mutations mentioned above. F was calculated using the equation, $F = (O - E) / (N - E)$, where O is the observed number of homozygous sites, E is the expected number of homozygous sites (given population allele frequency), and N is the total number of genotyped loci.

3.2.4. Phylogenetic analysis of published monomorphic *T. brucei* RNAseq data

Raw RNA reads from monomorphic *T. brucei* were downloaded from NCBI GEO GSE184265 (Cai et al., 2021). The data were combined with the raw reads from 3.2.1 and Read2Tree (Dylus et al., 2022) was used to align reads to trypanosomatid-specific orthologs identified by OMA which were used to create a phylogenetic tree with IQ-TREE (Nguyen et al., 2015). The tree was visualised with iTOL (Letunic and Bork, 2007).

3.2.5. Variant calling and dN/dS calculation

Bcftools was used to create a variant call file (VCF) file for each isolate from the filtered SNPs in section 3.2.1. Non-variant sites were removed using GATK4 SelectVariants (DePristo et al., 2011; Van der Auwera et al., 2013). The reverse complement of each VCF file and sequence was generated using SNPGenie (Nelson et al., 2015). The diversity of synonymous to non-synonymous mutations (dN/dS) was calculated using SNPGenie (Nelson et al., 2015). Within-pool analysis was performed for each isolate using SNPs found in the longest CDS of each gene, excluding pseudogenes, on the 11 megabase

chromosomes of the TREU927/4 reference genome (Berriman et al., 2005). Four gene categories were generated based on their identification as part of (a) the QS pathway (McDonald et al., 2018; Mony and Matthews, 2015; Mony et al., 2014; Rojas et al., 2019; Tettey et al., 2022), (b) transition from the midgut/proventriculus (MG-PV) stage (Naguleswaran et al., 2021), (c) transition from the proventriculus/salivary gland stage (PV-SG) (Naguleswaran et al., 2021)and (d) as a control, a random subset of genes, generated by picking 100 genes from a list of the 11,832 annotated transcripts in the TREU927/4 reference genome using the sample function in R.

3.2.6. Target prediction

Mutations from section 3.2.1 were assigned to annotated genes in the TRUE927/4 reference genome using snpEFF (Cingolani et al., 2012). We created a snpEFF database from the TREU927/4 genome using the methods described here https://pcingola.github.io/SnpEff/se_buildingdb/. Protein domains and putative phosphorylation sites were identified in target genes using InterProScan accessed through TriTrypDB (Aslett et al., 2010; Blum et al., 2021; Jones et al., 2014) and NetPhos (Blom et al., 1999; Blom et al., 2004), respectively. snpSIFT was then used to identify mutations which occurred in a monomorphic clade but not a pleomorphic clade using the case-control function (Cingolani et al., 2012). Target mutations putatively associated with monomorphism were then filtered and prioritised to create a target list of genes with ‘monomorphism-associated mutations. See section 3.3.5 for full details on filtering and prioritisation.

Samtools faidx (Li et al., 2009) and bcftools consensus were used to generate a fasta file for each of the target genes for each isolate (described above). The alternative allele was used if the mutation was heterozygous (Narasimhan et al., 2016). The sequences were aligned

with Clustal Omega and the alignments were translated from nucleotide to amino acid in Geneious Prime for visualisation.

3.2.7. Cellular localisation and protein modelling

Representative monomorphic mutant and pleomorphic control amino acid sequences were used to generate a protein structure model with ColabFold (Mirdita et al., 2022), which uses AlphaFold2 (Jumper et al., 2021) and Alphafold2-multimer (Evans et al., 2022). The protein models were created using default settings and the best model for each sequence, based on a predicted local-distance difference test (pLDDT), was aligned and visualised with PyMOL. Protein structure similarity searches were performed with Foldseek using the TREU927/4 AlphaFold model (van Kempen et al., 2022). Cellular localisation data for each target was accessed from the TrypTag (Dean et al., 2017) database (http://tryptag.org/) and visualised with FIJI (Schindelin et al., 2012). The transmembrane domains in Tb927.5.2580 were visualised with TOPCONS (Tsirigos et al., 2015)

3.2.8. Computational tools

Table 3.1: Computational tools and optional flags used in this chapter.

Tool	Version	Optional flags	
jellyfish	2.3.0	count -C -m 21 -s 1000000000	(Marçais and Kingsford, 2011)
jellyfish	2.3.0	histo	(Marçais and Kingsford, 2011)

GenomeScope2	2.0	-k 21	(Ranallo-Benavidez et al., 2020)
vcftools	0.1.16	–het	(Danecek et al., 2011)
Read2Tree	0.1.2	--tree	(Dylus et al., 2022)
IQ-TREE	2.2.0.3	-T auto -m MFP -B 1000	(Nguyen et al., 2015)
iTOL	6	Online tool	(Letunic and Bork, 2007)
snpEFF	5.0	-V Tb927	(Cingolani et al., 2012)
snpSIFT	4.3	filter	(Cingolani et al., 2012)
bcftools	1.9	view	(Danecek et al., 2021)
GATK4	4.1.4.1	SelectVariants --exclude-non-variants	(DePristo et al., 2011; Van der Auwera et al., 2013)
bcftools	1.9	+split	(Danecek et al., 2021)
SNPGenie	2019.10.31	vcf2revcom.pl	(Nelson et al., 2015)
SNPGenie	2019.10.31	--vcfformat=1	(Nelson et al., 2015)
NetPhos	3.1	Residues to predict - all three	(Blom et al., 1999; Blom et al., 2004)
samtools	1.12	faidx	(Li et al., 2009)
bcftools	1.9	consensus -H A	(Danecek et al., 2021)
Clustal Omega	1.2.3	-i -o --full --distmat-out	(Sievers et al., 2011)
Geneious	2022 1.1		https://www.geneious.com/

ColabFold	1.3.0	num_relax: 0 template_mode: none	(Mirdita et al., 2022)
PyMOL	2.5.2	align	https://pymol.org/2/
Foldseek	3-915ef7d		(van Kempen et al., 2022)
FIJI	2.3.0/1.53f		(Schindelin et al., 2012)
ggplot2	3.3.0	geom_density_ridges(rel_min_height = 0.000005	(Wickham, 2016)
ggrepel	0.8.1		ttps://cran.r-project.org/web/packages/ggrepel/index.html
R	3.6.1		(R Core Team, 2022)
TOPCONS	2.0		(Tsirigos et al., 2015)
fastqc	0.11.9		(Andrews and Others, 2010)
trimmomatic	0.39	SLIDINGWINDOW 4:20; ILLUMINACLIP adapters.fa 2:40:15; MINLEN 25	(Bolger et al., 2014)
bwa-mem	0.7.17-r1188	-t	(Li and Durbin, 2009)

3.3. Results

3.3.1. Extended phylogenetic analysis of monomorphic *T. brucei* identified at least four independent clades

Monomorphic *T. brucei* isolates form at least four independent clades

As in Chapter 2, analysis of 42 monomorphic genome sequences demonstrated that there are at least 4 independent clades, each separated by a pleomorphic isolate (Fig. 3.1). The vast majority of samples incorporated into this extended analysis are from the *T. b. evansi* type A clade. However, at least one additional sample has been incorporated from each of the four monomorphic clades, increasing our confidence in their independent origin. In Chapter 2, we hypothesised that the *T. b. evansi* type IVM-t1 isolate could represent a 5th clade of monomorphic *T. brucei*. In this analysis with an increased number of isolates, *T. b. evansi* type IVM-t1 diverges with *T. b. evansi* type B, but retains a long phylogenetic branch length (representing the number of nucleotide substitutions per nucleotide site), highlighting a relatively distant evolutionary relationship between *T. b. evansi* type IVM-t1 and *T. b. evansi* type B. The IVM-t1 isolate branch length (0.07109) is far greater than the mean branch length between isolates within the other monomorphic clades *T. b. evansi* type A (0.002581), *T. b. equiperdum* type OVI (0.0011816) and *T. b. equiperdum* type BoTat (0.001922). Furthermore, the distance between the two *T. b. evansi* type B (0.003144) isolates is smaller than the distance between these isolates and IVM-t1. Hence, we suggest that IVM-t1 could represent a 5th independent monomorphic clade.

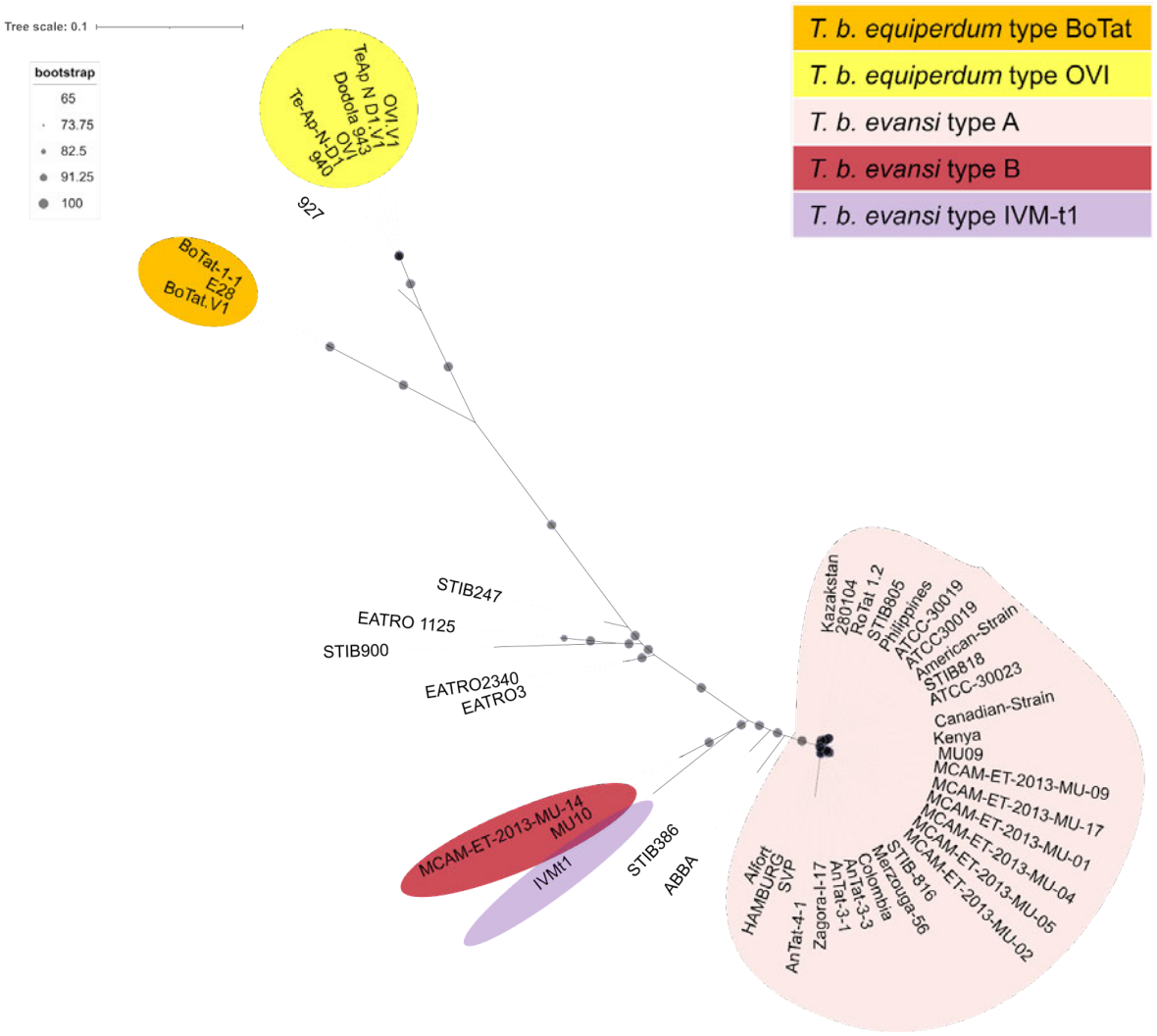


Figure 3.1: Monomorphic *T. brucei* are polyphyletic and form at least four independent clades. An unrooted phylogenetic tree was created with 235,571 homozygous SNPs (195,521 parsimony-informative SNPs) found across the *T. brucei* TREU927/4 reference genome. Bootstrap confidence is reported by the size of grey circles. Monomorphic genomes form at least four distinct clades which have expanded from Eastern (*T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat) and Western/Central Africa (*T. b. evansi* type A, *T. b. evansi* type B and *T. b. evansi* type IVM-t1). Monomorphic isolates are highlighted as one of the monomorphic clades. Pleomorphic isolates are not highlighted. A full list of the isolates used in this tree, and their novelty in this analysis, is found in Table S3.1.

Orthologous protein clustering highlights misidentified monomorphic isolates

An analysis of selected and naturally occurring monomorphic *T. brucei* has recently been published which reported “transcriptional patterns in *T. evansi* and *T. equiperdum* were seen to be diverse” (Cai et al., 2021) (Fig. 3.2). However, the analysis did not consider the genetic basis of the monomorphic isolates and relied upon clinical presentation to assign a (sub)species to each isolate. Therefore, we used Read2Tree to build a phylogenetic tree which incorporated the published monomorphic RNAseq data with the isolates in Fig. 3.1. Read2Tree bypasses traditional pipelines for orthogroup comparisons, which typically include cleaning raw reads, genome assembly, annotation, and orthologous protein clustering, by aligning raw reads to a group of OMA orthologous proteins (Altenhoff et al., 2021). The resultant phylogenetic tree (Fig. 3.3) largely recreated the results of Fig. 3.1 and previous studies (Cuypers et al., 2017; Oldrieve et al., 2021). However, in this version, *T. b. evansi* type IVM-t1 branches with the pleomorphic isolate ABBA (*T. b. gambiense* type 1) rather than *T. b. evansi* type B, supporting our suspicion that the isolate represents a 5th independent monomorphic clade.

The naturally occurring monomorphic *T. brucei* isolates which showed diverse mRNA profiles (Cai et al., 2021), are distantly related to one another (Fig. 3.3). The isolates typed as *T. equiperdum* are genotypically *T. b. equiperdum* type OVI (STIB841*), T. b. equiperdum* type BoTat (STIB842) and *T. b. evansi* type A (STIB818). The isolates typed as *T. evansi* are all *T. b. evansi* type A (CPOgz, STIB805 and STIB816). The two genetically closest related isolates, CPOgz and STIB816 (Fig. 3.3) have the closest mRNA profile (Fig. 3.2). STIB841 (*T. b. equiperdum* type OVI) is closest to *T. b. evansi* type A isolates STIB816 and CPOgz (Fig. 3.3). STIB805 (*T. b. evansi* type A) shows a similar pattern to STIB818 (*T. b. evansi* type A) and STIB842 (*T. b. equiperdum* type BoTat). Interestingly, the *T. b. equiperdum* type BoTat (STIB842) and OVI (STIB841) isolates display disparate mRNA profiles.

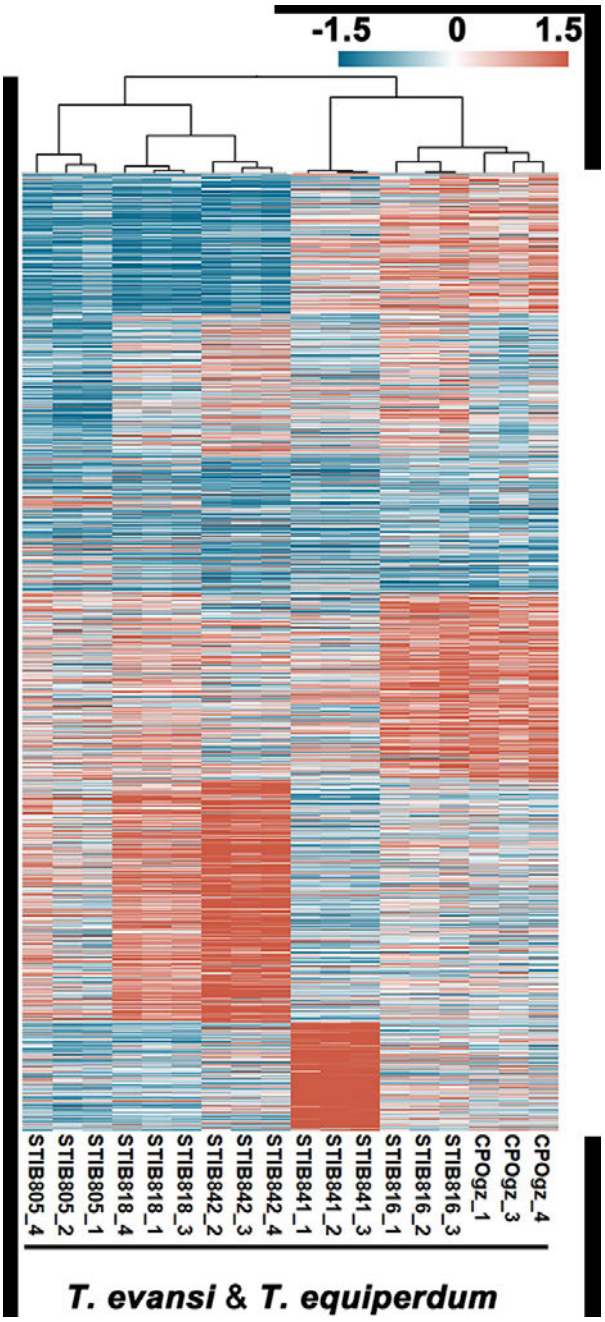


Figure 3.2: The RNAseq profiles of isolates identified by phenotype as "*T. evansi*" and "*T. equiperdum*" do not cluster together. The figure was adapted from Cai *et al* (2022) and represents the clustering of 9,483 genes using mRNA profiles from different monomorphic samples. Although not explicitly stated in the original paper, we presume the scale bar represents the rlog of each gene. Rlog normalises count data to the log2 scale to minimise differences between samples for rows with small counts and normalises with respect to library size.

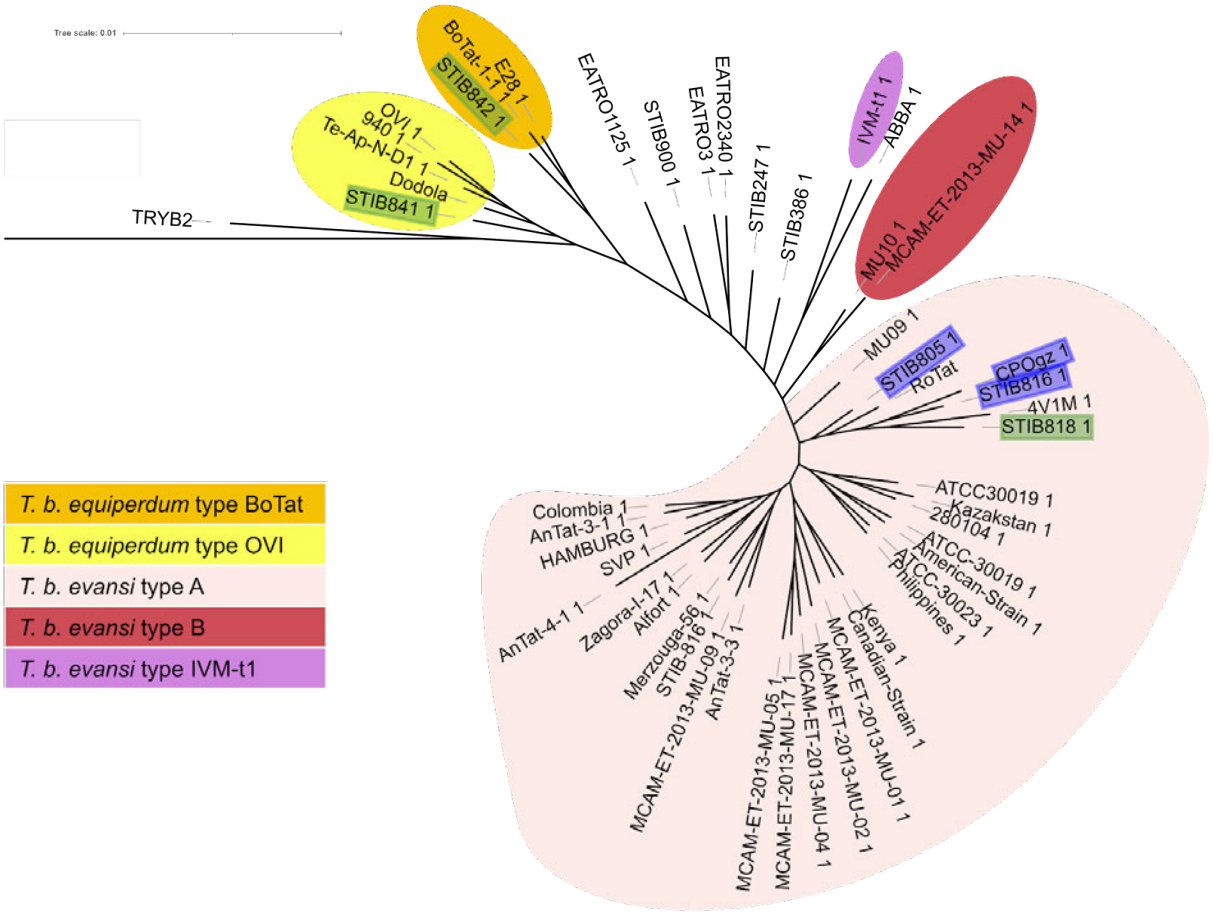


Figure 3.3: Phylogenetic analysis of isolates from Cai et al., (2021) highlights discrepancies between their genetic relationship and their species assignment. Orthologous kinetoplastid proteins were downloaded from OMA and Read2Tree was used to extract orthologues from the monomorphic isolates used in Fig. 3.1 along with RNAseq data from 6 monomorphic isolates (Cai et al., 2021). The isolates sequenced by Cai et al., (2022) have been highlighted in this tree. Green represents isolates typed as "*T. equiperdum"* and blue indicates isolates typed as "*T. evansi"* by Cai *et al.,* (2021). The tree was initially rooted with *Bodo saltans* and then this leaf was pruned from the tree for visualisation purposes. See Fig. S3.1 for the untrimmed version.

3.3.2. Monomorphic *T. brucei* display clade-specific genome content variation

To identify common features of monomorphic genomes, we utilised a k-mer (a string of nucleotides of any given length) counting-based tool to predict the unique and repeat genome length and the genomic heterozygosity. K-mer counting allowed us to rapidly and reliably compare genomic features without any prior processing of the raw data, such as genome assembly and annotation. The pitfall of the method is that it is not possible to identify the regions of the genome which account for the heterozygosity or variation in the genome repeat/unique length. For the following analysis, we used isolates in Fig. 3.1.

The five monomorphic *T. brucei* clades have a similar unique genome length, with no obvious divergence from the three pleomorphic *T. brucei* clades analysed in this study (Fig. 3.4a). There is variation in the *T. b. rhodesiense* isolates, with one isolate (EATRO3) presenting a unique genome length of 11Mb compared to 26Mb (STIB900) and 27Mb (EATRO2340). EATRO3 also had the lowest model fit and highest read error count, therefore, this discrepancy could be caused by sequencing errors. EATRO3 was treated with caution in the following analysis. Regardless, monomorphic isolates appear to have relatively uniform unique genome content, especially in comparison to pleomorphic clades such as *T. b. brucei.* The genome of *T. brucei* is predicted to contain ~20% repetitive DNA (Pita et al., 2019). Monomorphic *T. brucei* clades have the lowest repeat genome length, especially *T. b. evansi* type C, *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat. The repeat length is lower than pleomorphic *T. b. rhodesiense*, *T. b. gambiense* and *T. b. brucei*. TREU927/4 (*T. b. brucei*) has the lowest *T. b. brucei* repeat genome length, 7.2Mb, which creates a shoulder in the *T. b. brucei* curve (Fig. 3.4b). Whilst the read error is not low, the model fit was 87%, below the *T. b. brucei* model fit mean of 96%, potentially producing the lower repeat genome content prediction for this isolate.

K-mer counting predicted that heterozygosity is reduced in monomorphic *T. brucei* clades. The pattern is especially clear in *T. b. evansi* type C, although *T. b. evansi* type B is an exception (Fig. 3.4c). Finally, we calculated the inbreeding coefficient (F) for each isolate whilst taking population parameters into account. Monomorphic *T. brucei* are the most inbred clades, especially *T. b. evansi* type C and *T. b. equiperdum* type OVI. Isolates within *T. b. gambiense* and *T. b. rhodesiense* also appear to be relatively inbred, potentially indicating clonal isolates within these clades or a variable length of time in culture (Fig. 3.4d). In Fig. 3.4 there are multiple peaks within clades, highlighting divergence between isolates within the clade.

Combined, this data highlights discrepancies in the genome content of monomorphic and pleomorphic clades. The clade-specific variation could easily be missed if monomorphic isolates were analysed without first considering the phylogenetic relationships between the isolates. Therefore, it is essential to analyse each monomorphic clade independently in future analysis.

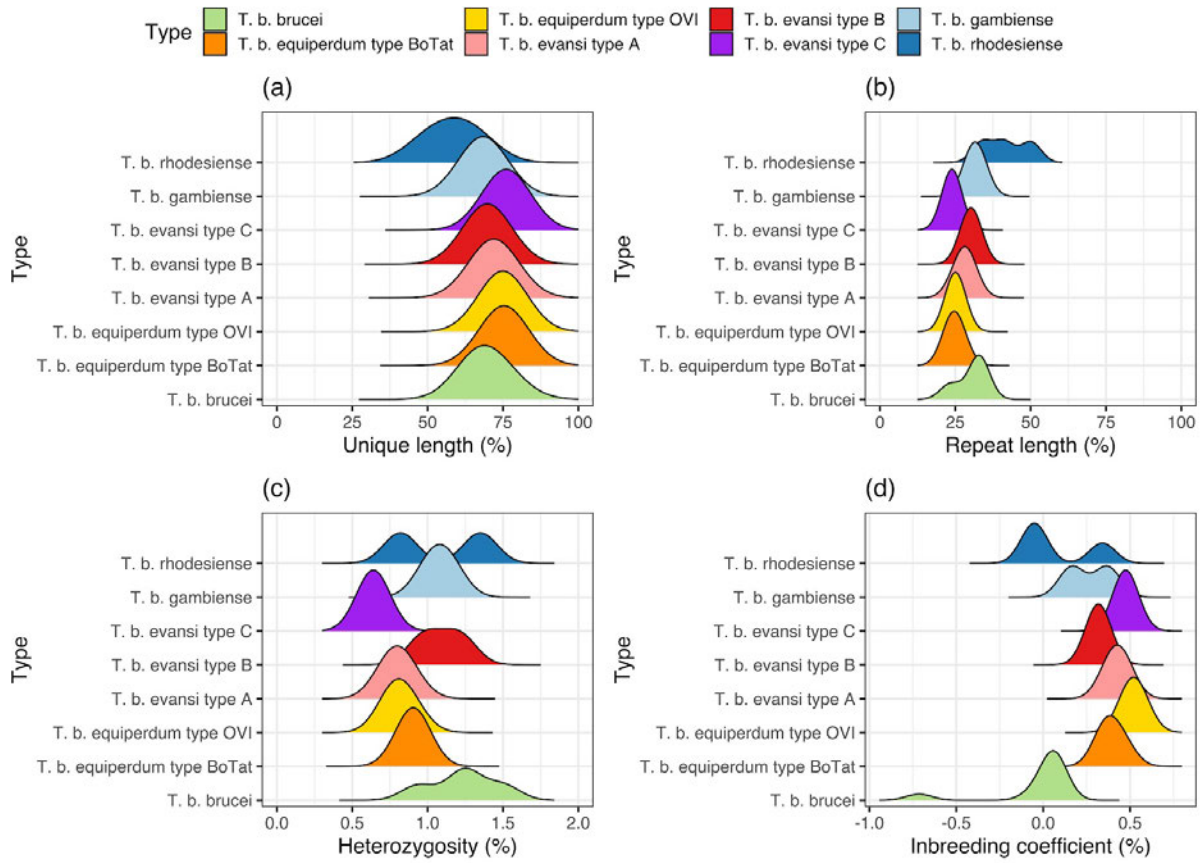


Figure 3.4: Monomorphic *T. brucei* display clade specific reductions in heterozygosity and genome repeat content. *T. brucei* genome-wide **(a)** percentage of unique nucleotides across the genome, **(b)** percentage of nucleotides that are repetitive across the genome, **(c)** genome-wide heterozygosity and **(d)** inbreeding coefficient (F). Genome summary statistics were calculated for each clade of *T. brucei*. **(a-c)** Were calculated based on k-mer counting using a k-mer length of 21 and **(d)** was calculated with SNP mutations called against the TREU927/4 reference genome. The y-axis represents the count for each clade, a summary of isolates used for each clade can be found in Table S3.1. The figure was created using ggridges which first estimates data densities and then draws the ridge profile (available at https://cran.r-project.org/web/packages/ggridges/vignettes/gallery.html).

3.3.3. Fixed mutations are not overrepresented in monomorphic *T. brucei* subspecies

Throughout, we use mutations to describe fixed non-synonymous and synonymous mutations unless explicitly stated otherwise. We searched for the presence of synonymous and non-synonymous mutations in each clade of *T. brucei* in groups of genes in the quorum sensing pathway (McDonald et al., 2018; Mony et al., 2014; Rojas et al., 2019; Tettey et al., 2022) and insect stages of the parasite's life cycle (Naguleswaran et al., 2021). These genes are presumably not used by monomorphic *T. brucei*, as they cannot transition from the slender to stumpy form or undergo cyclical development in the insect vector. As a control, the number of mutations was counted for a random subset of genes, generated by picking 100 genes from a list of the 11,832 annotated transcripts in the TREU927/4 reference genome using the sample function in R. The number of mutations was also counted across every gene in the TREU927/4 reference genome.

Mutations are not overrepresented in monomorphic *T. brucei* clades; instead, the overall presence of mutations mirrors the phylogenetic divergence of each isolate from the reference genome, TREU927/4. For instance, the monomorphic clade *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat are the closest isolates to the reference genome (Fig. 3.1). They also have the least mutations across the whole genome, as well as the insect stage of the parasite's life cycle, QS genes and a random group of genes, in comparison to all other clades in this analysis (Fig. 3.5).

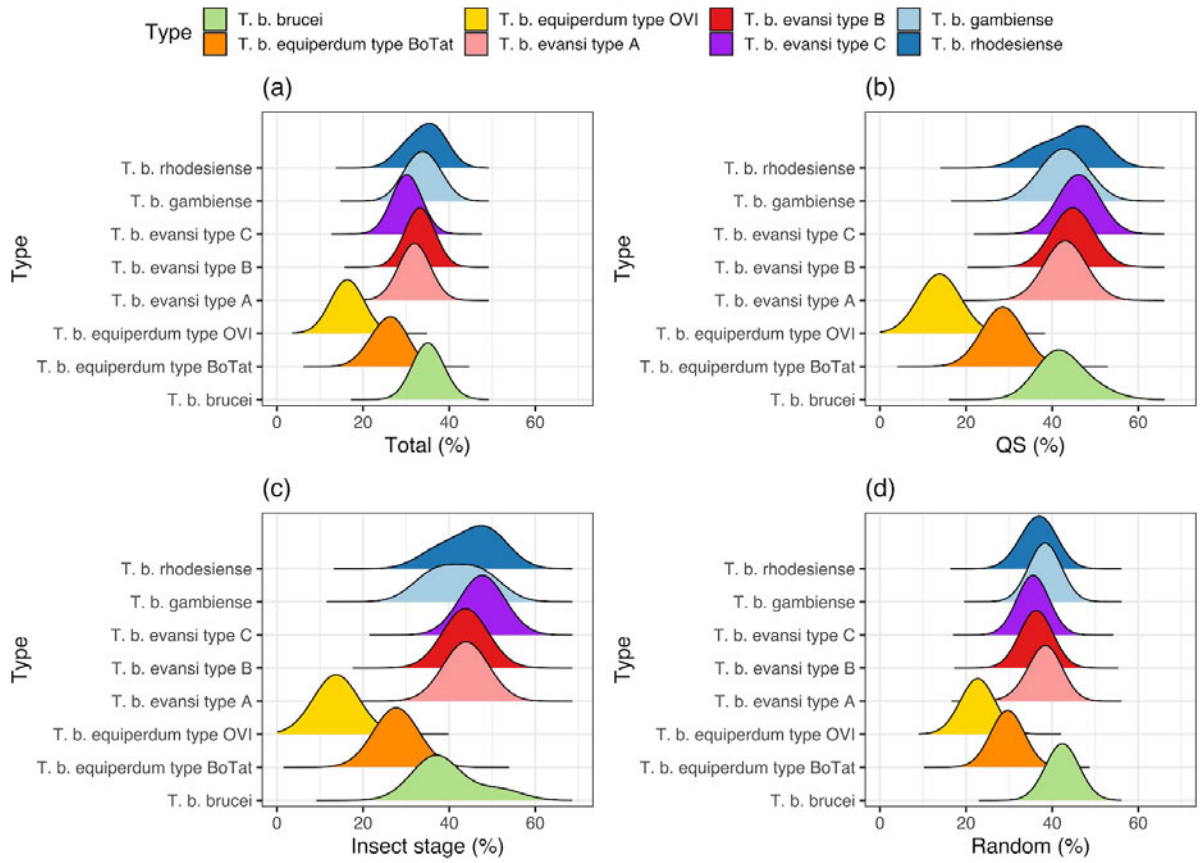


Figure 3.5: Mutations are not overrepresented in monomorphic *T. brucei* clades. Heterozygous and homozygous mutations found in each clade as a percentage of the total number of mutations found across all isolates used in this analysis **(a)** for the whole genome, **(b)** genes in the QS pathway, **(c)** genes associated with the insect stage of the parasite's life cycle and **(d)** a random group of genes. The number of mutations was counted for each clade of *T. brucei*. The y-axis represents the count for each clade of *T. brucei,* a summary of isolates used for each clade can be found in Table S3.1. The figure was created using ggridges which first estimates data densities and then draws the ridge profile (available at https://cran.r-project.org/web/packages/ggridges/vignettes/gallery.html).

3.3.4. Monomorphic *T. brucei* display clade-specific variation in selection efficacy

Whilst we do not see an overrepresentation of mutations in monomorphic *T. brucei* clades, asexuality is expected to reduce heterozygosity (Jaron et al., 2021). Asexuality can also reduce the efficacy with which selection can act (Otto, 2021). The efficacy of selection can be estimated by calculating the ratio of nonsynonymous to synonymous mutations (dN/dS) compared to a reference genome.

Across the whole genome, the dS value was well below one for all isolates, indicating the absence of inflation which can saturate and inflate dN/dS calculations (minimum = 0.002460, mean = 0.009701 and maximum = 0.011523). Focusing on groups of genes associated with the QS pathway, the insect stage of the parasite's life cycle and a random subset of genes, each category of genes has a dN/dS ratio below one for every *T. brucei* clade, indicative of restrictive selection acting to maintain the integrity of gene sequence (Fig. 3.6). This pattern was seen for all of the monomorphic and pleomorphic clades. Genes associated with the QS pathway and the midgut/proventriculus stage have the lowest dN/dS ratio. The value is almost identical for each clade. The random group of genes, and genes associated with expression in the proventriculus and the salivary gland, have a higher dN/dS ratio, indicating that there is less selection pressure acting to maintain the coding sequence of these genes. *T. b. equiperdum* type OVI isolates have the lowest dN/dS ratio in the random set of genes compared to all other clades in this analysis. The same clade displays the highest dN/dS ratio in genes associated with the QS pathway and is elevated in genes expressed in the midgut, proventriculus and salivary gland, indicative of a reduction in the efficacy of selection acting to maintain the functionality of these gene sequences in *T. b. equiperdum* type OVI.

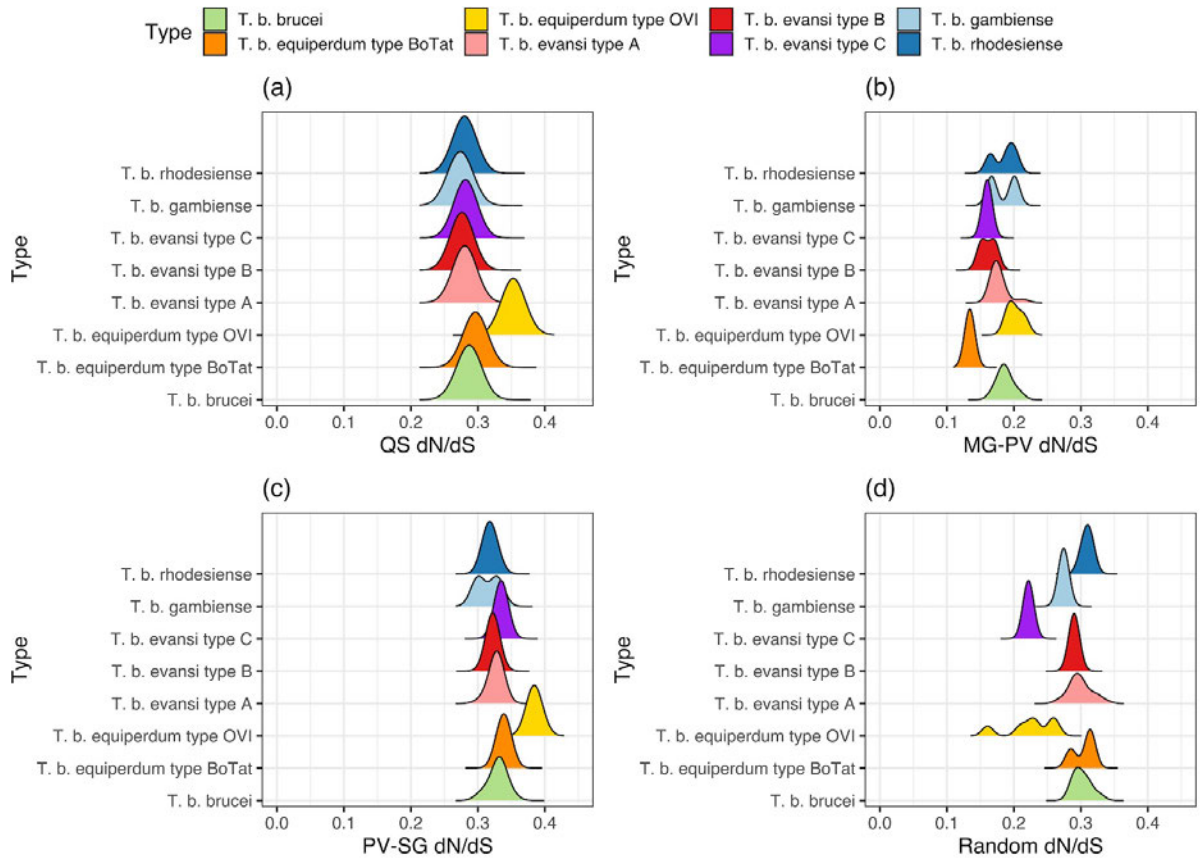


Figure 3.6: **Monomorphic *T. brucei* have clade-specific variation in the efficacy of selection.** dN/dS ratio for groups of genes involved in **(a)** the QS pathway, **(b)** the midgut/proventriculus (MG-PV) stage, **(c)** the proventriculus/salivary gland (PV-SG) and **(d)** a random group of genes. The dN/dS ratio was calculated for each clade of *T. brucei*, highlighting clade-specific variation in the efficacy of selection. The y-axis represents the count for each clade of *T. brucei,* the summary of isolates used for each clade can be found in Table S3.1. The figure was created using ggridges which first estimates data densities and then draws the ridge profile (available at https://cran.r-project.org/web/packages/ggridges/vignettes/gallery.html).

3.3.5. Mutations in 6 genes have been identified for experimental analysis of monomorphism

Genomic mutations shared by isolates within monomorphic clades and absent in pleomorphic isolates were identified to create a list of potential mutations contributing to monomorphism for each monomorphic clade. The impact of each mutation was categorised with SnpEff into high impact (e.g. frameshift), moderate impact (e.g. codon insertion) or modifier (e.g. non-synonymous) mutations. Full details of the SnpEff categories can be found here (https://pcingola.github.io/SnpEff/se_inputoutput/#eff-field-vcf-output-files). Two categories were then created based on the following criteria:

1. Any high-impact, moderate-impact or modifier mutation in a monomorphic clade in a known QS pathway gene (McDonald et al., 2018; Mony et al., 2014; Rojas et al., 2019; Tettey et al., 2022) or a high impact mutation in a gene with the following pattern in the RNA Interference Target Sequencing (RITseq) dataset: smaller log fold change in D3, D6 and PF than in the DIF category whilst displaying a log fold change in the DIF category greater than 1.5. The RITseq dataset consists of the outputs from a genome-wide phenotypic screen for genes which have roles in the fitness of the parasite at different stages of development. The screen compared a baseline uninduced library to RNAi-induced libraries of bloodstream-form cells at day 3 (D3), day 6 (D6), procyclic form parasites (PF) and cells induced throughout growth as bloodstream form, differentiation and growth as procyclic forms (DIF) (Alsford et al., 2011).
2. Novel genes potentially associated with the monomorphic phenotype were identified by extracting genes which had a positive selection pressure in a monomorphic clade (dN/dS >1) and a restrictive selection pressure in pleomorphic clades (dN/dS <1).

Gene targets that had a frameshift in a pleomorphic clade were exMutationsnd any gene annotated as a pseudogene or VSG was excluded. After filtering, 48 mutations in Category 1 and 45 mutations in Category 2 (Table 3.2) were prioritised based on the highest score using 3 criteria:

1. A mutation was found in the same gene in another monomorphic clade (+1).
2. The mutation hits a predicted protein domain (+1).
3. Unique mutation found in a pleomorphic clade (-1).

Based on prioritisation following these three criteria, six genes were taken forward for validation (four genes from Category 1 and two genes from Category 2 (Table 3.3)). Alongside the orthologous monomorphic genes, a replacement of each gene with a pleomorphic *T. b. brucei* sequence was generated as a control.

Table 3.2: Presence of putative monomorphism causing mutations in each monomorphic *T. brucei* clade.

clade	Category 1	Category 2
T. b. evansi type A	2	1
T. b. evansi type B	4	1
T. b. evansi type IVM-t1	15	3
T. b. equiperdum type OVI	6	14
T. b. equiperdum type BoTat	21	25
Total	48	44

Table 3.3: Target genes chosen for validation. Yellow: *T. b. brucei* control sequence generated for each target. Green: a sequence was generated for the monomorphic clade. Grey: the gene does not contain a clade-specific mutation.

Gene	Code	brucei	A	B	IVM-t1	OVI	BoTat
(1) Known QS pathway							
APPBP1	Tb927.2.4020						
GPR89	Tb927.8.1530						
HYP1	Tb927.11.6600						
PP1	Tb927.4.3650						
(2) Novel genes identified via selection pressure							
HYP	Tb927.5.2580						
FAZ41	Tb927.11.3400						

3.3.6. Summary of putative monomorphism causing targets

Category 1: Four putative targets are known QS pathway genes

Tb927.2.4020 – NEDD8-activating enzyme E1 (APPBP1)

Along with UBA3, APPBP1 binds NEDD8 and activates the neddylation pathway which is analogous to ubiquitination. APPBP1 is highly conserved amongst eukaryotes, performing post-translational protein modifications associated with a diverse repertoire of cellular processes (Amir et al., 2002; Kurz et al., 2002; Osaka et al., 2000; Pozo et al., 1998; Read et al., 2000). APPBP1 is a member of the QS pathway alongside another neddylation pathway member, KRIPP13 (Liao et al., 2017; Mony et al., 2014). The human crystal structure of the APPBP1-NEDD8-UBA3 complex has been resolved (Fig. 3.7a) (Walden et al., 2003). Whilst alignment of the *T. b. brucei* and human amino acid sequence revealed varying divergence in NEDD8 (44.2%), UBA3 (32.5%) and APPBP1 (25.2%), the human and modelled *T. brucei* APPBP1-NEDD8-UBA3 complex structure is highly conserved (RMSD = 1.629) (Fig. 3.7b). According to TrypTag, APPBP1 localises to the cytoplasm, flagellar cytoplasm and nuclear lumen (Fig. 3.7c).

Clade-specific non-synonymous mutations in APPBP1, including the addition of putative phosphorylation sites, are found in the monomorphic clades *T. b. evansi* type A (4), *T. b. evansi* type IVM-t1 (5) and *T. b. equiperdum* type BoTat (4) in comparison to TREU927/4 (Fig. 3.8). The amino acid sequence is identical for all isolates within each clade. The predicted structures have minimal alteration generated by the mutations in *T. b. evansi* type

IVM-t1 (RMSD = 0.210), *T. b. evansi* type A (RMSD = 0.541) and *T. b. equiperdum* type BoTat protein structure (RMSD = 0.672) in comparison to TREU927/4 (Fig. S3.2-S3.4).

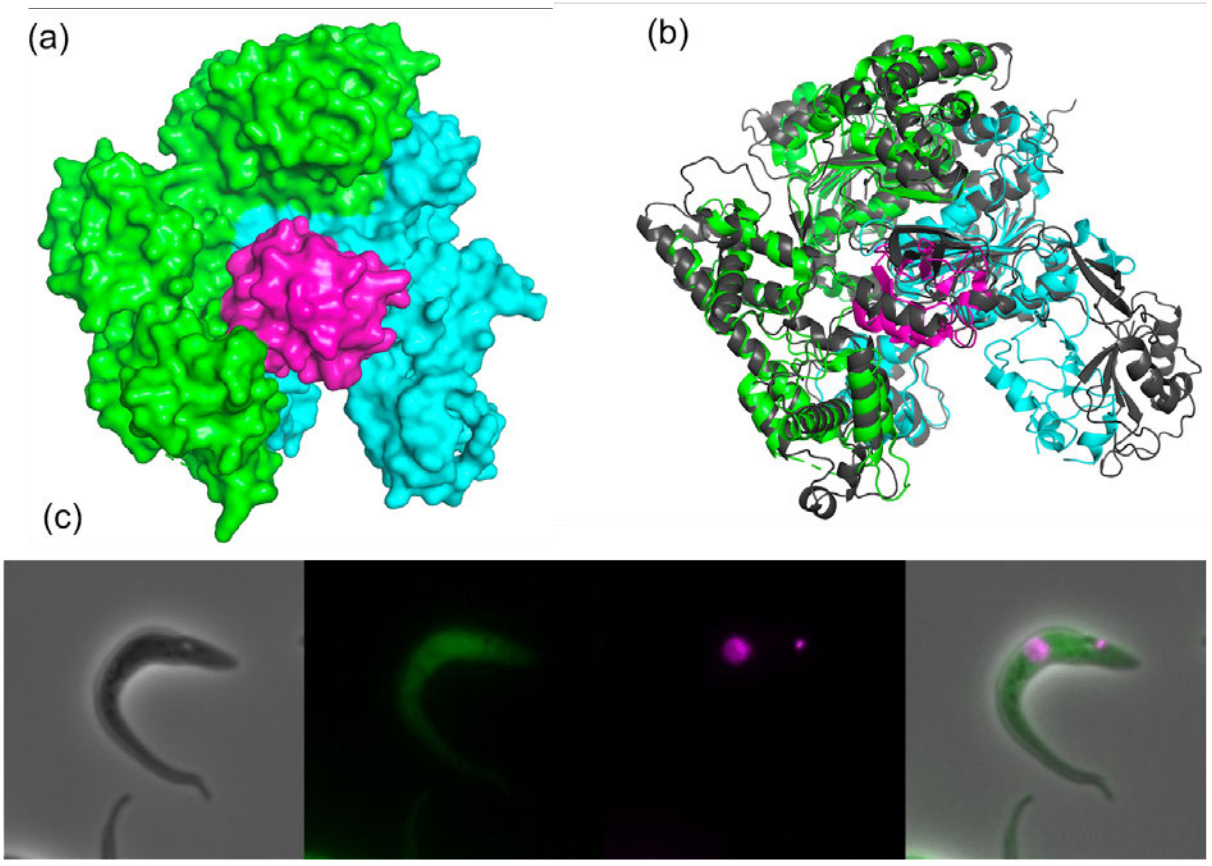


Figure 3.7: The APPBP1-NEDD8-UBA3 complex is highly conserved and localises to the cytoplasm. (a) APPBP1-NEDD8-UBA3 human protein crystal structure, **(b)** alignment of APPBP1-NEDD8-UBA3 human protein crystal structure and *T. b. brucei* APPBP1-NEDD8-UBA3 complex generated with CollabFold. A surface version of Fig. 3.7b can be found in Fig. S3.5. The human crystal structure was visualised in PyMol using the 1R4M Biological Assembly 1 model (Walden et al., 2003). The model is coloured by chain: Green = APPBP1, Blue =UBA3, Pink =NEDD8, Grey = *T. b. brucei* TREU927/4 overlay. **(c)** Cellular localisation of APPBP1 (Tb927.2.4020) in *T. brucei*. APPBP1 localises to the cytoplasm, flagellar cytoplasm and nuclear lumen based on N-terminal TrypTag data (Dean et al., 2017).

```
TREU927/4                   KLMSNEKYDRQLRLWGHAGQMALAESHVVVLGATATAVEMLKNMILPGLGFFTLVDGSFV   60
T. b. evansi type A         ............................................................   60
T. b. evansi type IVM-t1    ............................................................   60
T. b. equiperdum type BoTat ............................................................   60

TREU927/4                   DADTLGNNYFVELSDYAARKPLSEVLVKRLCELNPHSSGTACVQSCVEWSDSFISSTSGG  120
T. b. evansi type A         ............................................................  120
T. b. evansi type IVM-t1    ............................................................  120
T. b. equiperdum type BoTat ............................................................  120

TREU927/4                   RDFVGRFPTLIVATPRLPATHLRRLADHLKSASLRSIPLMYVQTCGLSGIIQIQDRERLV  180
T. b. evansi type A         ............T......A............P..N.......................  180
T. b. evansi type IVM-t1    ............T......A............P..N.......................  180
T. b. equiperdum type BoTat ..................A.....................................  180

TREU927/4                   VHAEPKQEMRVADLRLFNPFPELRSWLDAHDPRDEVLYRTDPAGYGHLPWIAMIYHALQC  240
T. b. evansi type A         ............................................................  240
T. b. evansi type IVM-t1    ..............................................S.............  240
T. b. equiperdum type BoTat ............................................................  240

TREU927/4                   LRCDKENAEFIPCSKADYDTLRTAVSSLPYLSDRPPDGVHEAMENCRMILNRPMCLSESL  300
T. b. evansi type A         ............................................................  300
T. b. evansi type IVM-t1    ............................................................  300
T. b. equiperdum type BoTat ............R.....................................H........  300

TREU927/4                   EQLLRDPRANGHFIDWSGGFPVSGVPPITWAVLYGINRFIAENNGVPPFCGFVPDINTTT  360
T. b. evansi type A         ............................................................  360
T. b. evansi type IVM-t1    ............................................................  360
T. b. equiperdum type BoTat ........................T...................................  360

TREU927/4                   QWYRELRSIYNNKMEDDCCVVQGYAMEAINASNVMGGKFTSDTAASTALLVGEKTGLETE  420
T. b. evansi type A         ............................................................  420
T. b. evansi type IVM-t1    ............................................................  420
T. b. equiperdum type BoTat ............................................................  420

TREU927/4                   VSALTRSLVQNIWTLQLTTFSPSLDIDSHELLQRESFERYLSDCVDASEQFTVMLYAALL  480
T. b. evansi type A         ............................................................  480
T. b. evansi type IVM-t1    ............................................................  480
T. b. equiperdum type BoTat ............................................................  480

TREU927/4                   GARCFEEMFGRWPGLLDSLNGRVHYGDKDDTLLWMSDGEELIRIVGERILNESGRSDKVV  540
T. b. evansi type A         ............................................................  540
T. b. evansi type IVM-t1    ............................................................  540
T. b. equiperdum type BoTat ............................................................  540

TREU927/4                   DECLMRKACMEVARCGCGEPFPTAGVMGAAAAQECIKLLQHRRVPIQRPLVFDGYRSCFW  600
T. b. evansi type A         ............................................................  600
T. b. evansi type IVM-t1    ............................................................  600
T. b. equiperdum type BoTat ............................................................  600

TREU927/4                   VAEAQ*GS                                                    608
T. b. evansi type A         ........                                                    608
T. b. evansi type IVM-t1    ........                                                    608
T. b. equiperdum type BoTat ........                                                    608
```

Figure 3.8: Amino acid sequence alignment of monomorphic *T. b. evansi* type A, *T. b. evansi* type IVM-t1, *T. b. equiperdum* type BoTat and pleomorphic *T. b. brucei* (TREU927/4) APPBP1 (Tb927.2.4020). HindIII and BamHI sites are shown as part of the alignment. Full stops represent a shared amino acid.

Tb927.8.1530 - Golgi pH regulator (GPR89)

GPR89 transports the oligopeptide-based quorum sensing signal, initiating the transition from slender to stumpy form. The protein is essential in the mammalian bloodstream form (Rojas et al., 2019). *T. b. equiperdum* type BoTat contains five non-synonymous mutations in comparison to *T. b. brucei.* Two mutations are clade-specific (C112Y, which creates a putative phosphorylation site, and D183V) (Fig. 3.10). The sequence is identical in all *T. b. equiperdum* type BoTat isolates. C112Y is part of one of the 9 transmembrane domains and D183V is part of the intracellular loop (Fig. 3.9a). We created models of the GPR89 structure using the *T. b. brucei* (TREU927/4) and *T. b. equiperdum* type BoTat sequences, which aligned well (RMSD = 1.756). C112Y has no impact on the protein structure but the D183V mutation is predicted to cause a change in the angle of the intracellular loop. After the deviation caused by D183V, the model re-aligns until the end of the sequence (Fig. 3.9a). Overexpression of GPR89 induced early stumpy form generation whilst overexpression of a protein truncated of the intracellular loop did not produce the same phenotype (Rojas et al., 2019). TrypTag indicates that GPR89 localises to the cytoplasm, flagellar cytoplasm and nuclear lumen in procyclic form cells (Fig. 3.9b), in bloodstream forms Rojas et al., (2019) report a surface location. Both protein localisation studies used a C terminus tag.

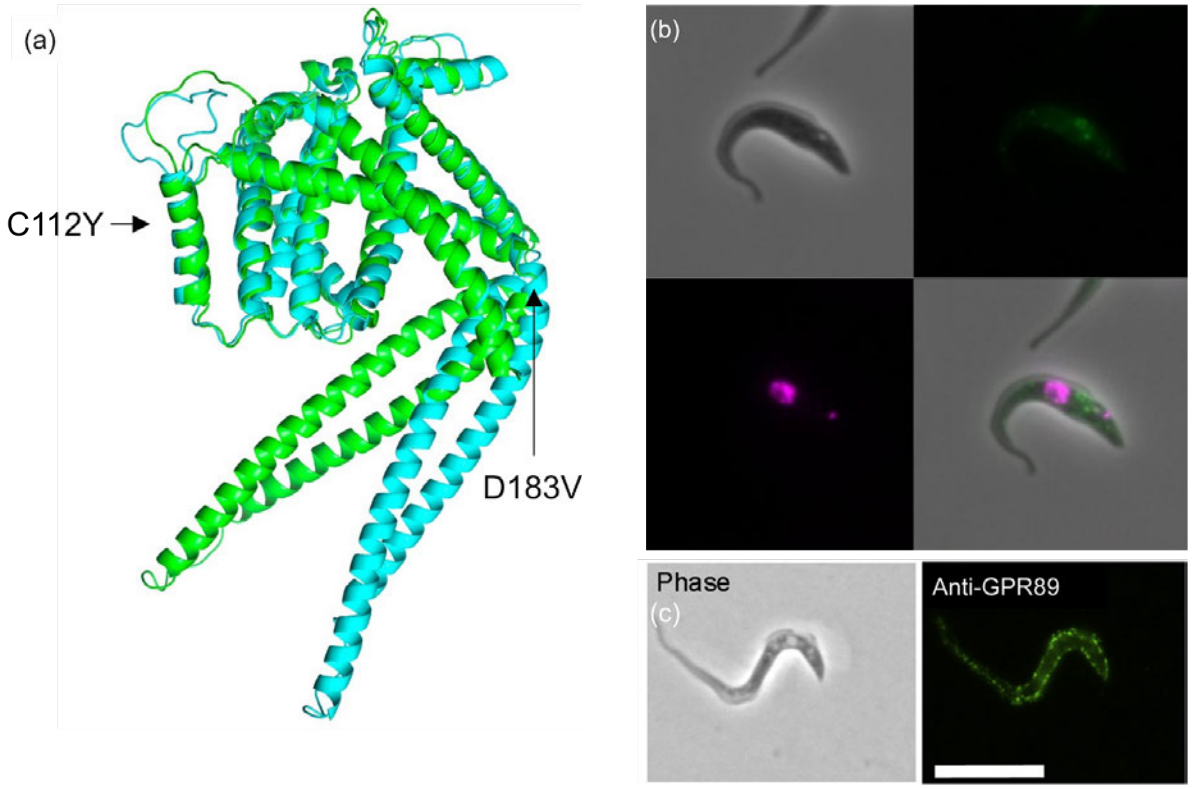


Figure 3.9: Mutations in *T. b. equiperdum* type BoTat GPR89 are predicted to alter the angle of its intracellular loop. (a) Structure of GPR89 (Tb927.8.1530) predicted with AlphaFold using the *T. b. brucei* (blue) and *T. b. equiperdum* type BoTat (green) sequence (RMSD = 1.756). **(b)** C-terminal tagged GPR89 localises to the cytoplasm, flagellar cytoplasm and nuclear lumen. Images were taken from TrypTag (Dean et al., 2017). (c) GPR89 is reported to localise to the cell surface in the bloodstream form. Images were taken from (Rojas et al., 2019).

```
TREU927/4                  KLMLLKLGAVLLLLVLYHVGSLFSRLVFVYTIQQGTTKFCFRFVFALSTYAFILLLLDAS   60
T. b. equiperdum type BoTat ............................................................   60

TREU927/4                  QVSERIKSNFSIQTLRCVLAVDLVLIALFCPFLVVRSLLPNVRLYYFFLVLLACGLVKML  120
T. b. equiperdum type BoTat .L.............................................Y......  120

TREU927/4                  FGASVAAWSWTLSSTQGIPTLFSWDVVSVAVSTTGVVVVGLLSGYAAVTTPLAFVEPLFE  180
T. b. equiperdum type BoTat ............................................................  180

TREU927/4                  HGSGDQARLAVGVLAKRQKHLLELWMLKRQQIARAYGAVSRADGGRDGSNRGSIAAGRRM  240
T. b. equiperdum type BoTat ....V....................................................  240

TREU927/4                  WNWVANSIYSSVRASGADIAAMETESDGIRAVSMAVFLQMSEMDTLVRSAGSGETWRRRF  300
T. b. equiperdum type BoTat ............................Q...........................  300

TREU927/4                  NALFGLILFTHALVKFLSTVISLLRWGVLAAENAAPHREDTATKVMNFLEAYGLATPHGD  360
T. b. equiperdum type BoTat ............................................................  360

TREU927/4                  GAEQRVVWVSIALNAWMIASAIRGFFLLVFRLTTHIAFISLDTTVTILTAGMGAFFVGQL  420
T. b. equiperdum type BoTat ................................M...........................  420

TREU927/4                  VLLRLTPSLERESVLYTALREQLPQHGAYCHLNDLVFVVAAALVAMMRRCMFSSSATALC  480
T. b. equiperdum type BoTat ............................................................  480

TREU927/4                  AAAD*GS                                                     487
T. b. equiperdum type BoTat .......                                                    487
```

Figure 3.10: Amino acid sequence alignment of monomorphic *T. b. equiperdum* type BoTat and pleomorphic *T. b. brucei* (TREU927/4) GPR89 (Tb927.8.1530). HindIII and BamHI sites are shown as part of the alignment. Full stops represent a shared amino acid.

Tb927.11.6600 – Hyp1

Hyp1 was identified in a screen for posttranscriptional regulators (Erben et al., 2014). When artificially tethered to transcripts, HYP1 downregulates mRNAs (Erben et al., 2014; Mony and Matthews, 2015; Mony et al., 2014). RNAi-mediated knockdown of HYP1 highlighted its role in stumpy formation, potentially by destabilising transcripts of stumpy form inhibitors or slender form retainers (Mony and Matthews, 2015; Mony et al., 2014). HYP1 localises to the cytoplasm and nuclear lumen (Dean et al., 2017) (Fig. 3.11).

Hyp1 contains five non-synonymous mutations in *T. b. evansi* type B and one in *T. b. equiperdum* type OVI compared to TREU927/4 (Fig. 3.12). The sequence is identical between all isolates for both clades. Predicting the protein structure of Hyp1 produced low-confidence models and alignment of both *T. b. evansi* type B (Fig. S3.6, RMSD = 14.778) and *T. b. equiperdum* type OVI (Fig. S3.7, RMSD = 12.469) to the *T. b. brucei* model was poor.

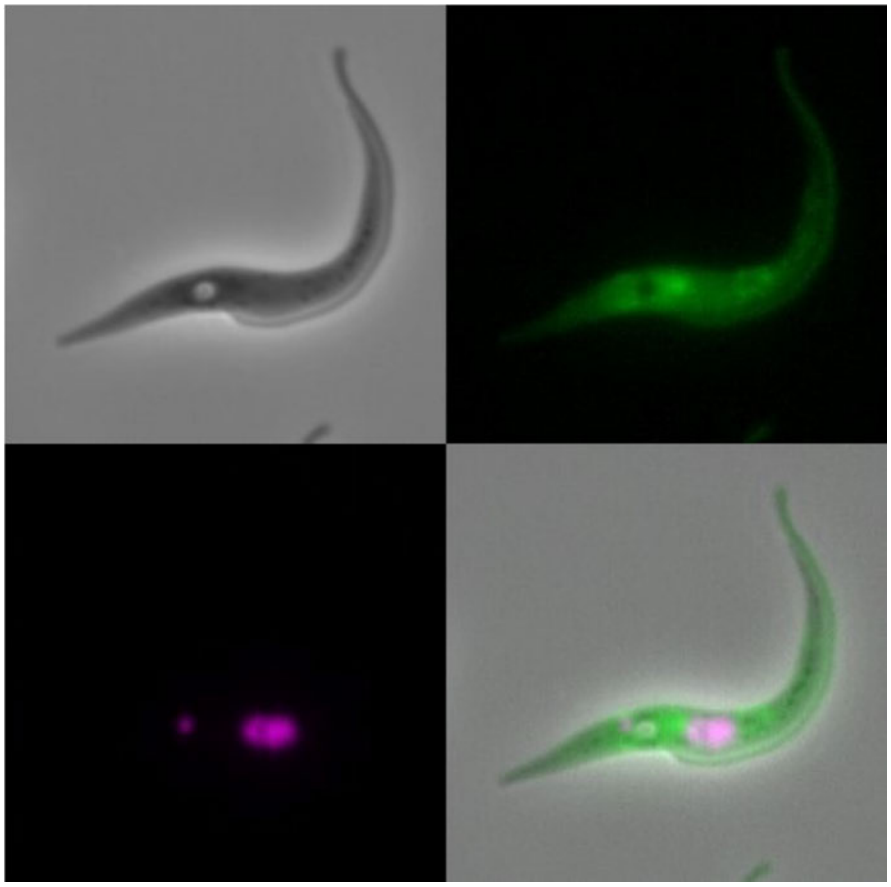


Figure 3.11: N-terminal tagged Hyp1 (Tb927.11.6600) localises to the cytoplasm, flagellar cytoplasm and nuclear lumen. Images were taken from TrypTag (Dean et al., 2017).

```
TREU927/4                KLMDPRNPALTEHSLDNAAGQPCDGWISHGVWLSGPPSESITDRMLTGPSDMRNDNAAPG   60
T. b. evansi type B      ............................................S.............   60
T. b. equiperdum type OVI ..............................................................   60

TREU927/4                MRRAPPHLFISTARQPGCDEVCHNMLQDGSDPLSMKTRGYSPLMESNVHSVPPSFGGSSA  120
T. b. evansi type B      ..............................................................  120
T. b. equiperdum type OVI ..............................................................  120

TREU927/4                SSPTKSSMTYSPYAGAMTASPFLSISTPVISPGVSEGNGVSFQRSTPPSLPVGLSTFVAS  180
T. b. evansi type B      ..............................................................  180
T. b. equiperdum type OVI ..............................................................  180

TREU927/4                QGNSNFVMVTQPNTPHTLYVQGSCPAPGGVPHYVVAQPQLSMDVCGGMTYVPMTPAQQTF  240
T. b. evansi type B      ............................S.............................  240
T. b. equiperdum type OVI ..............................................................  240

TREU927/4                GTVMGDQLTAVTPGPNSAAPFSPTVTCFPVSVSSSSPLGNGLLQPSTPMNSSPLAAACAA  300
T. b. evansi type B      ..........................................G.................  300
T. b. equiperdum type OVI ......................................F.......................  300

TREU927/4                AGVGSPPSYDSHQVQQSGRKAAAGHQRSSVNVHGTLQGFCLYHSYNISNTISRYGQTDTV  360
T. b. evansi type B      ...V..........................................................  360
T. b. equiperdum type OVI ..............................................................  360

TREU927/4                SKEDAGRTLLVFFQMFPCELRQRAVVVLNRVLEVVCGDGMAVAESVEWRSETSFIARVRT  420
T. b. evansi type B      ..............................................................  420
T. b. equiperdum type OVI ..............................................................  420

TREU927/4                KDIWTLIYQVRCRVLMDRHGFWYAADREQYSHLRKYCERVRNLPQQTRHSETDGLPCMPL  480
T. b. evansi type B      ................................R.............................  480
T. b. equiperdum type OVI ..............................................................  480

TREU927/4                IVELSRGELNPPPRPPTVPDSFDCSEPMMTVDRRSRTAVPTAAEFGRCGISA*GS       535
T. b. evansi type B      ...................................................            535
T. b. equiperdum type OVI ...................................................            535
```

Figure 3.12: Amino acid sequence alignment of monomorphic *T. b. evansi* type B and *T. b. brucei* (TREU927/4) Hyp1 (Tb927.11.6600). HindIII and BamHI sites are shown as part of the alignment. Full stops represent a shared amino acid.

Protein phosphatase type 1 (PP1) was highlighted in the genome-wide screen for quorum-sensing genes (Mony et al., 2014). It is part of an eight-gene family; four of these PP1 genes are found in a tandem array. Tb927.4.3650 is separated from the other three by the insertion of two unrelated genes. It also has unique UTRs and there is divergence in its coding sequence, compared to the other PP1 proteins in close proximity (Jackson, 2007a). In *T. cruzi*, inhibition of PP1 by protein phosphorylation inhibitors caused disruption of the cell cycle and cell shape (Orr et al., 2000), supporting findings in higher eukaryotes (Ceulemans and Bollen, 2004). In *T. brucei*, simultaneous RNAi knockdown of the 7 PP1 family members was not lethal but did reduce the growth rate (Li et al., 2006) whilst treatment with an inhibitor of PP1 family activity caused defective genome segregation and cytokinesis (Das et al., 1994). Tb927.4.3650 is reported to localise to the nucleolus (Fig. 3.13) (Dean et al., 2017).

T. b. evansi type B has three mutations in PP1 compared to *T. b. brucei* (Fig. 3.14). The protein sequence is uniform in all of the *T. b. evansi* type B isolates. Protein modelling produced a low-confidence model, preventing alignment of the *T. b. evansi* type B and *T. b. brucei* structures (Fig. S3.8, RMSD = 7.925).

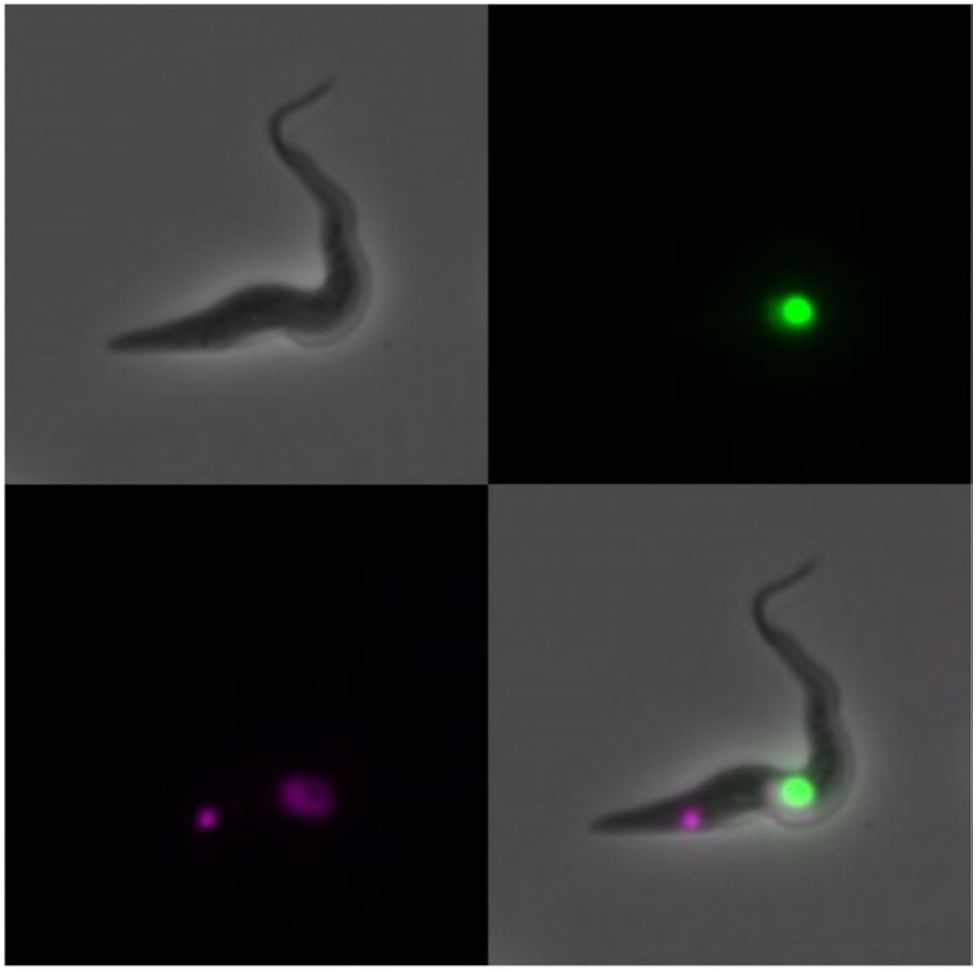


Figure 3.13: N-terminal tagged PP1 (Tb927.4.3650) localises to the nucleolus. The images were taken from TrypTag (Dean et al., 2017).

```
TREU927/4           KLMSVLVKGWGRGTCKINHIRDDDISSSESEGDDANVAVPLEELNVGTQHMITEEMRFRL   60
T. b. evansi type B .................V.....................................L....   60

TREU927/4           RRKEKLSVQDDELGRLREAMRRQVFASGLLQPGQSAGVAPRRAEIIFSCAVEIQSSGLME  120
T. b. evansi type B ................................V.......................  120

TREU927/4           EAPTASPRVGDTSAWLKQLRKFILDRVPFQRSVRVMVPRTWLARAQETDVV*GS        174
T. b. evansi type B ................................................           174
```

Figure 3.14: Amino acid sequence alignment of monomorphic *T. b. evansi* type B and *T. b. brucei* (TREU927/4) PP1 (Tb927.4.3650). HindIII and BamHI sites are shown as part of the alignment. Full stops represent a shared amino acid.

Category 2: Two targets are novel genes identified via selection pressure

Tb927.5.2580 – hypothetical protein

Tb927.5.2580 is a mitochondrially located hypothetical protein (Fig. 3.15). It is tightly associated with the cytochrome c oxidase complex (complex IV) (Zíková et al., 2008). The cytochrome c oxidase complex is stage regulated, being expressed in procyclic forms but repressed in bloodstream forms which do not have an intact respiratory chain and do not directly utilise the mitochondrion for ATP production (Mayho et al., 2006). Tb927.5.2580 has a transmembrane domain (amino acids 281-302, Fig. 3.16b) and shares structural similarity to a bacterial SH3 domain-containing protein (I4CAI4_DESTA).

The protein has a positive dN/dS ratio in the monomorphic clade *T. b. evansi* type A, in contrast to pleomorphic and other monomorphic clades. Protein modelling created a low-quality model and the models did not align well (Fig. S3.9, RMSD = 8.299). The sequence is uniform across all available *T. b. evansi* type A isolates and contains four non-synonymous mutations compared to TREU927/4 (Fig. 3.16a). Only one of these is unique to this clade (A149P). The mutations are contained within the predicted extracellular portion of the protein (Fig. 3.16b).

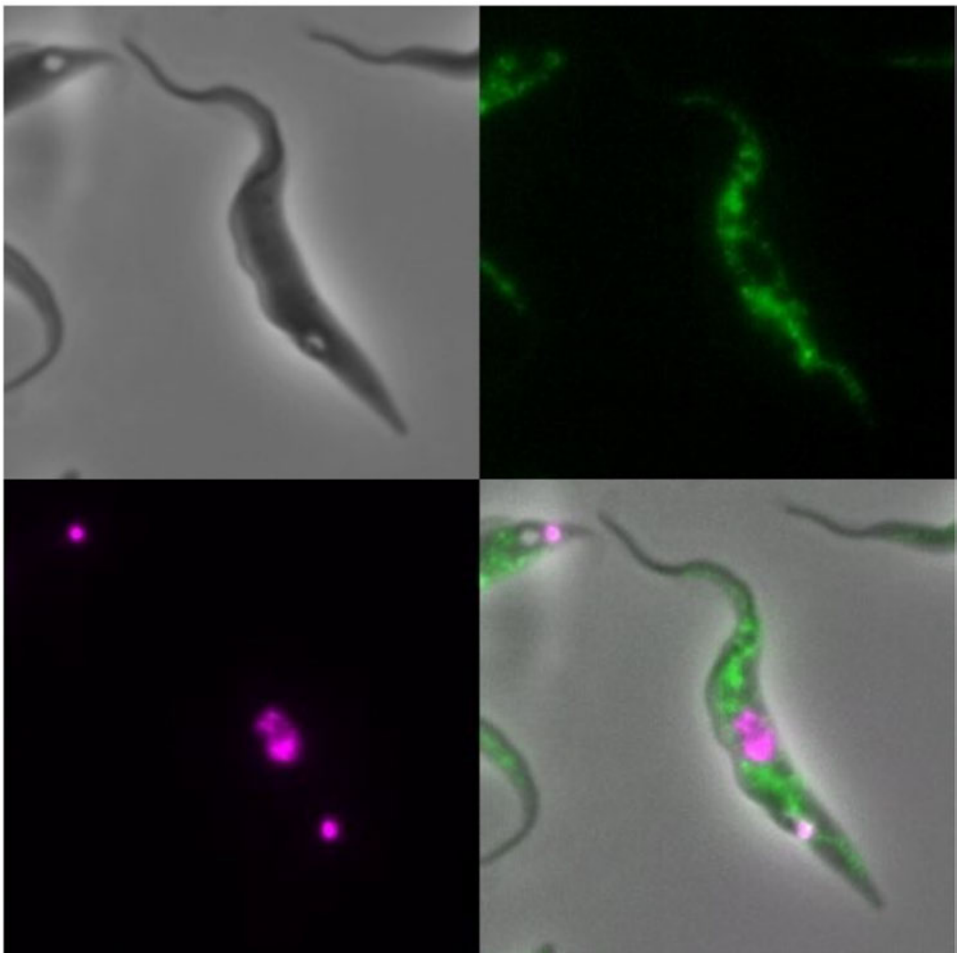


Figure 3.15: C-terminal tagged hypothetical protein (Tb927.5.2580) localises to the mitochondrion. The images were taken from TrypTag (Dean et al., 2017).

(a)

```
TREU927/4           KLMAVMRCILRIHVTDGISTRCLAYRASFLCVKRAHSNFRTGQEDQYTTKRSEDESFGEG   60
T. b. evansi type A .....................Y.....................................G.   60

TREU927/4           DSHYPSKDRTTTSPTSSQGSSDTHEHTAASSAEEKLSSASRGARDDISHQELEPMRWPQE  120
T. b. evansi type A ..............A.....................................................  120

TREU927/4           APGSPALTEPIIDEKGQFIVSRIQWPTGELAYATPPPPDTRVAPRFGYNVVQVKKHVSWW  180
T. b. evansi type A ............................P.....................................  180

TREU927/4           KHYQQHPRISVAYINIQLLFLLGAAWLMAFLVEEYRRVTDELRTPGAMVGEHRGRGPVEK  240
T. b. evansi type A .....................................................................  240

TREU927/4           GKQKISFTNDEMTSLIGRAQDNWMDARAEANYIGSKDYTMKKIPRPKEFSVDDFRKR*GS  300
T. b. evansi type A .....................................................................  300
```

(b)

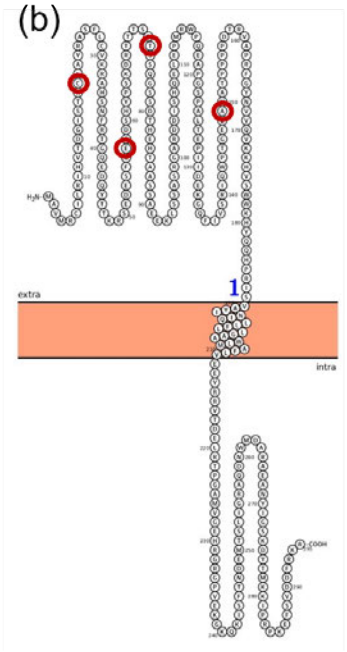


Figure 3.16: Mutations in *T. b. evansi* type A Tb927.5.2580 are found within the predicted extracellular portion of the protein. (a) Amino acid sequence alignment of monomorphic *T. b. evansi* type A and *T. b. brucei* (TREU927/4) hypothetical protein (Tb927.5.2580). HindIII and BamHI sites are shown as part of the alignment. Full stops represent a shared amino acid. **(b)** Transmembrane domain model of *T. b. brucei* hypothetical protein (Tb927.5.2580) generated with TOPCONS (Tsirigos et al., 2015). Mutation sites in *T. b. evansi* type A are highlighted in red.

The flagellum attachment zone (FAZ) is a cytoskeletal structure which connects the cell body cytoskeleton to the flagellum skeleton. The FAZ influences differentiation and cell division by manipulating organelle position and cell length (Sunter and Gull, 2016). FAZ41 stretches along the majority of the cell (Dean et al., 2017) (Fig. 3.17) and is unique to the *Trypanosoma* genus (Oldrieve et al., 2022).

T. b. equiperdum type OVI and *T. b. equiperdum* type BoTat both have a positive dN/dS ratio for FAZ41, compared to pleomorphic and other monomorphic clades. Although the two monomorphic clades are currently described as a single species or subspecies (*T. b. equiperdum*) they are separated by a pleomorphic isolate *T. b. brucei* TREU 927/4 (Fig. 3.1). They share five identical non-synonymous mutations in FAZ41 compared to TREU927/4 (Fig. 3.18). FAZ41 protein models for *T. b. equiperdum* type OVI (Fig. S3.10, RMSD = 20.508) and *T. b. equiperdum* type BoTat (Fig. S3.11, RMSD = 20.508) align poorly with *T. b. brucei*, likely due to the low confidence models.

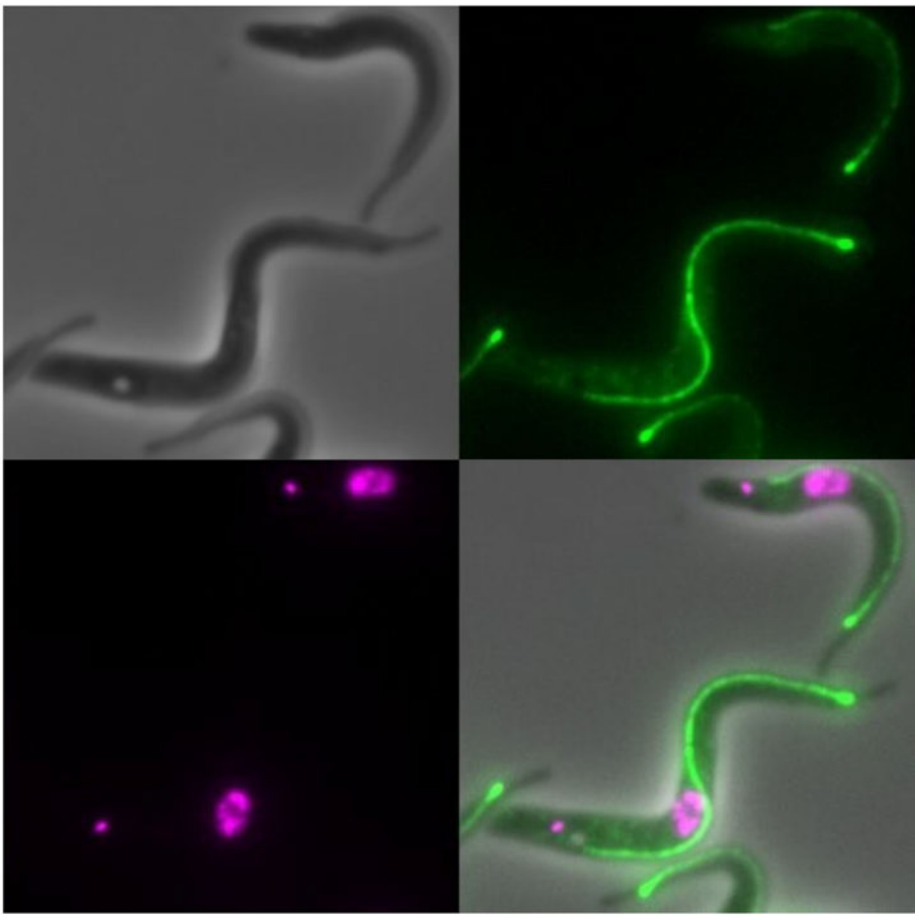


Figure 3.17: C-terminal tagged FAZ41 (Tb927.11.3400) localises to the flagellum attachment zone. Images were taken from TrypTag (Dean et al., 2017).

```
TREU927/4                  KLMEGRGGEDPYMPPLRRRPGTRRDYQPLYASDASRGPRSRPNTARCHSQRPAPPSVRWR   60
T. b. equiperdum type OVI  ...............................V..........................   60
T. b. equiperdum type BoTat ..............................V..........................   60

TREU927/4                  SPDHLNTSYRETDTSSIMRATPRTGRTHVAEERTSRHQQRRCQSATPPVTRSVLSWDLQV  120
T. b. equiperdum type OVI  ...............................M..........................  120
T. b. equiperdum type BoTat ..............................M..........................  120

TREU927/4                  LDGVELGFSDPGSLRLLELQVPCSTSANLFGTHERSKTFIPGDPVGLKLRRDAMSIHPSR  180
T. b. equiperdum type OVI  .........................T..................................  180
T. b. equiperdum type BoTat .........................T..................................  180

TREU927/4                  PRDLCFLSPAAAKRNNIPLSLGNIGDGTIATIPYYRVSRISRCRFASTDRRIVVRLTGTT  240
T. b. equiperdum type OVI  .......................V..................................  240
T. b. equiperdum type BoTat .......................V..................................  240

TREU927/4                  GLVLTLAFTDKALYTTAMKVITGYCSNCDRRPDGVRCSSPARNMSHPEGEGRYVCFAEPE  300
T. b. equiperdum type OVI  ..............................T............................  300
T. b. equiperdum type BoTat ..............................T............................  300

TREU927/4                  VRTLRSSRANGYTY*GS                                          317
T. b. equiperdum type OVI  ................                                           317
T. b. equiperdum type BoTat ................                                           317
```

Figure 3.18: Amino acid sequence alignment of monomorphic *T. b. equiperdum* type OVI, *T. b. equiperdum* type BoTat and pleomorphic *T. b. brucei* (TREU927/4) FAZ41 (Tb927.11.3400). HindIII and BamHI sites are shown as part of the alignment. Full stops represent a shared amino acid.

3.4. Discussion

We have extended the analysis in Chapter 2 which supported there being at least four independent monomorphic clades, corroborating previous studies (Cuypers et al., 2017). Additionally, *T. b. evansi* type IVM-t1 represents a long phylogenetic branch from *T. b. evansi* type B, potentially indicative of a 5th independent monomorphic clade. We also placed isolates used for published transcriptome comparisons in the context of our phylogenetic analysis. As for *T. b. evansi* type IVM-t1, these isolates represent an example of the common misassignment of monomorphic isolates, due to assumptions based on their clinical presentation and host species. The misassignment of these isolates has confused the analysis of their mRNA profiles (Cai et al., 2021) and highlights why it is essential to treat monomorphic *T. brucei* clades as independent, based on genotypically assigned types.

Monomorphic genomes generally displayed reduced repeat genome content, potentially indicative of selection for genome size reduction. Whilst genomic features of asexual organisms are often clade-specific, repetitive transposable element (TE) numbers are generally lower in the genomes of naturally occurring asexual organisms (Jaron et al., 2021). Experimental evolution of asexual yeast from a sexual ancestor highlighted that transposable elements were reduced over 1000 generations (Bast et al., 2019) although the reproducibility of the results has been questioned (Chen and Zhang, 2021). TE account for 2-5% of the trypanosomatid genome content (Bringaud et al., 2008), potentially accounting for some of the genome repeat content reduction in monomorphic genomes. The experimental evolution of bacteria highlighted large, and often consistent, losses of accessory genes associated with adaptation to a specific niche (Lee and Marx, 2012). Accessory genes are those which are not essential for the organism but assist with an adaptation, such as pathogenesis. Adaptation to monomorphism could have facilitated a similar loss of duplicated accessory genes required for developmental competence.

Similarly, repeat content could have been lost from the 70bp repeats associated with recombination homology during VSG switches (Hovel-Miner et al., 2016), from telomeric repeats, which can shorten with continual passage (Dreesen et al., 2005), or extensive centromeric repeats (Echeverry et al., 2012). Our analysis is unable to detect the region of the repeat loss.

As well as a reduction in repeat sequences, monomorphic *T. brucei* display reduced heterozygosity and an increased inbreeding coefficient. Variation in heterozygosity can be associated with a transition to asexuality (Jaron et al., 2021). Monomorphic strains have lost their tsetse transmission ability and therefore are obligately asexual (Peacock et al., 2011; Wheeler et al., 2013b). Asexuality is apparent in the reduction of heterozygosity and the inbreeding coefficient observed in the majority of monomorphic isolates. Such reduction in heterozygosity can occur via mitotic recombination and/or gene conversion (Weir et al., 2016). Gene conversion is the modification of one of two alleles by the other, involving the nonreciprocal correction of the “acceptor” sequence by a donor sequence, which remains unchanged. The mutational attrition observed in asexually reproducing organisms, which have chromosomes under independent evolution, can be reduced or even completely stopped by gene conversion. Through gene conversion, mutations of the acceptor sequence can be removed in favour of the donor sequence at a faster rate than the accumulation of spontaneous mutations (Marais et al., 2010).

Our analysis was unable to detect variation in the number of mutations found in each *T. brucei* clade, due to a bias in the number of mutations based on phylogenetic distance from the reference genome. However, using the ratio of mutation types (synonymous and non-synonymous) we found that *T. b. equiperdum* type OVI had the greatest reduction in its selection efficacy in genes associated with the insect stage of the parasite's life cycle as well as the quorum sensing pathway (Fig. 3.6). This is in contrast to monomorphic and

pleomorphic clades. Importantly, the change in selection pressure was not mirrored in a random subset of genes, highlighting the specificity of the alteration in selection efficacy for these developmentally relevant genes. The efficacy of selection can be influenced by the evolutionary history of a clade, such as population bottlenecks and alteration to the mode of inheritance, such as the transition to obligate asexual reproduction (Charlesworth, 2009). Such changes in the efficacy of selection can lead to an accumulation of deleterious mutations, which could eventually lead to the extinction of an asexual organism as mutations accumulate, often in genes associated with sexual traits (Jaron et al., 2021; Schwander et al., 2013). However, the variation in selection efficacy was only observed in one monomorphic clade, potentially indicative that other monomorphic clades were more recently derived. Alternatively, since the *T. b. equiperdum* type OVI clade is largely thought of as sexually transmitted between equines and invades the genital mucosa, its variation in genome content could be associated with a complete dissociation from the insect host or its novel tissue tropism. Even though *T. b. evansi* type A and *T. b. evansi* type B are mechanically transmitted, they do still briefly come into contact with an insect when they are taken up in a blood meal and are found in the mammalian bloodstream. In future studies, a branch and site model analysis with this data would provide a more detailed view of positive selection acting on genes and sites associated with the transition to monomorphism (Yang and dos Reis, 2011).

Using the recent ‘ecotype’ proposition to define *T. brucei* isolates (Lukeš et al., 2022), the IVM-t1 isolate would be assigned to *T. b. f. equiperdum*. Likewise, STIB818, STIB841 and STIB842 (Cai et al., 2021) would be classified as *T. b. f. equiperdum* (Cai et al., 2021). Here, we have shown clade-specific variation in genomic content, and mRNA profiles, of each of these clades, highlighting the need to treat monomorphic clades as independent. The use of ecotypes would preclude any analysis of the genetic basis of monomorphism and further confuse the analysis of these isolates. Furthermore, the ecotype definition suggests that with a few simple mutations, it would be possible to generate a different ecotype. We suggest this

is an oversimplification. Indeed, if we were to mutate a *T. b. evansi* type A laboratory strain to encode the mutations required to create the *T. b. f. equiperdum* ecotype, sequence its genome and perform phylogenetic analyses, it would still group with other *T. b. evansi* type A isolates. The ecotype definition also does not resolve how we would deal with the independent origins of monomorphic *T. brucei*. Rather, our phylogenetic and genome content analysis highlights dissimilarities between the independent monomorphic clades, which we have accounted for when identifying clade-specific mutations which could be associated with developmental incompetence.

Utilising publicly available protein localisation and protein modelling allowed us to predict the effect that mutations found in monomorphic clades might have on the protein structure, and where these alterations may take effect in the cell. Based on the AlphFold pLDDT score, the structure of GPR89 and APPBP1 were accurately modelled. These models predict that the *T. b. equiperdum* type BoTat mutations alter the structure of GPR89, bending the loop of the intracellular portion of the molecule (Rojas et al., 2019). In contrast, the protein models for APPBP1 aligned almost perfectly, suggesting that the mutations would not alter the structure. It remains possible, however, that the mutations could be associated with changes in the potential for post-translational protein modifications, such as the creation of the phosphorylation site G224S in the *T. b. evansi* type IVM-t1 sequence. Furthermore, dynamic conformation alterations would not be detected with the fixed structural models we created, and they do not take interaction partners into account.

When protein models were unable to create an accurate representation of the molecular structure, we used additional information, such as the presence of transmembrane domains, to identify important domains in the protein sequences. In the hypothetical mitochondrial protein (Tb927.5.2580) we highlight a transmembrane domain. *T. b. evansi* type A has an accumulation of mutations in this extracellular portion of the protein. This protein is associated with the cytochrome c complex, which is stage regulated in *T. brucei* being

expressed in procyclic forms and repressed in bloodstream forms (Mayho et al., 2006; Zíková et al., 2008).

Although we picked targets from known QS pathway genes, the studies that isolated these genes were carried out in laboratory monomorphs using a cell-permeable mimic of the QS signal. This prevented the identification of the early stages of the reception of the oligopeptide signal (Mony et al., 2014). Therefore, screening for genes with a positive selection pressure in a monomorphic clade, alongside genes known to be associated with QS, had the potential to highlight genes associated with other aspects of the monomorphic phenotype aside from developmental incompetence. Using an experimental gene replacement approach we have tested these computational predictions in Chapter 4.

3.5. Supplementary material

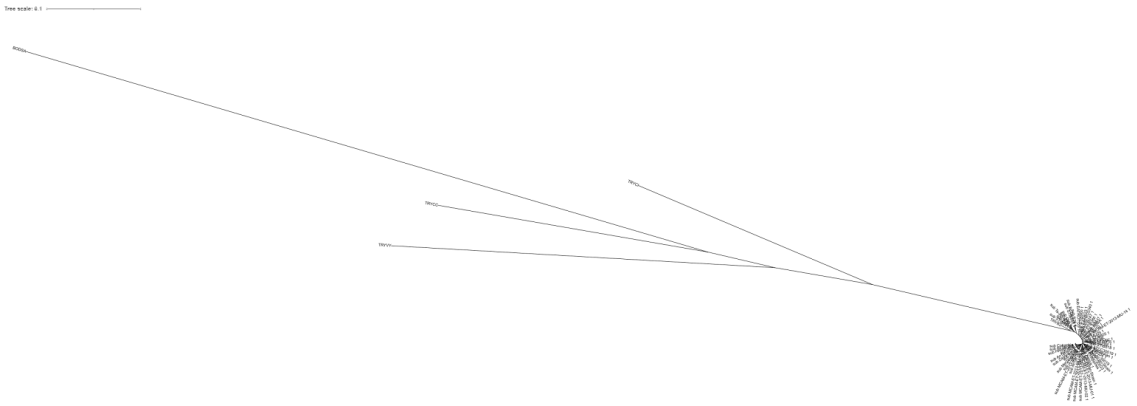


Figure S3.1: A phylogenetic tree based on orthologous protein sequences. Orthologous kinetoplastid proteins were downloaded from OMA and Read2Tree was used to extract orthologues from the monomorphic isolates used in Fig. 3.1 along with RNAseq data from 6 monomorphic isolates (Cai et al., 2021). The tree was rooted with *Bodo saltans* and represents the untrimmed version of Fig. 3.3.

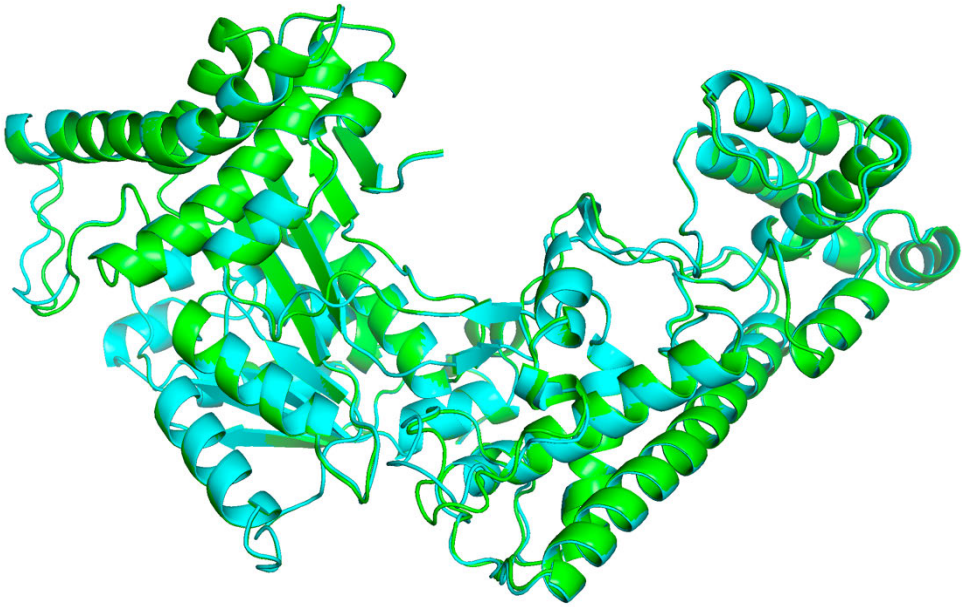


Figure S3.2: Alignment of Tb927.2.4020 APPBP1 *T. b. brucei* (blue) and *T. b. evansi* type IVM-t1 (green) protein structures. RMSD = 0.210.

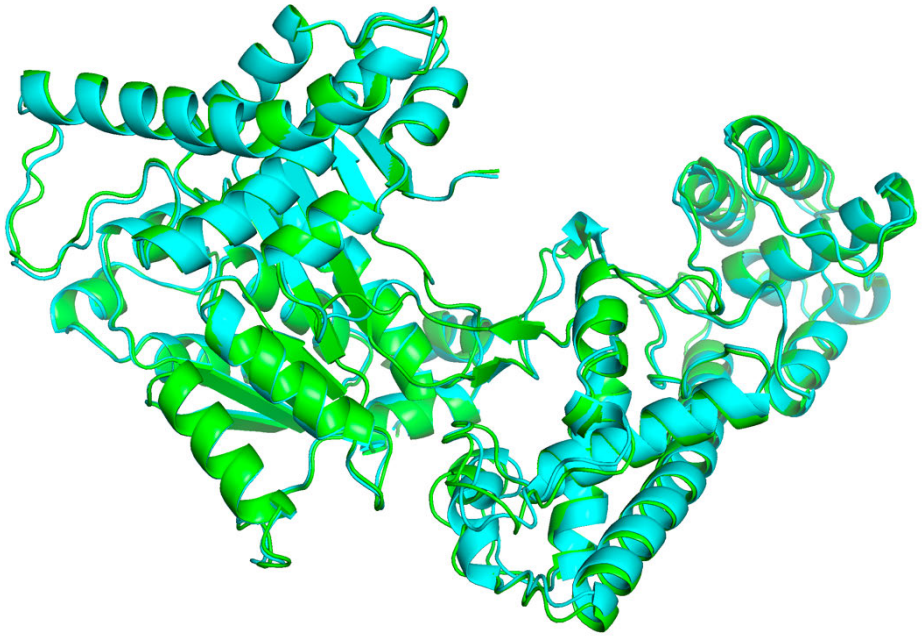


Figure S3.3: Alignment of Tb927.2.4020 APPBP1 *T. b. brucei* (blue) and *T. b. evansi* type A (green) protein structures. RMSD = 0.541.

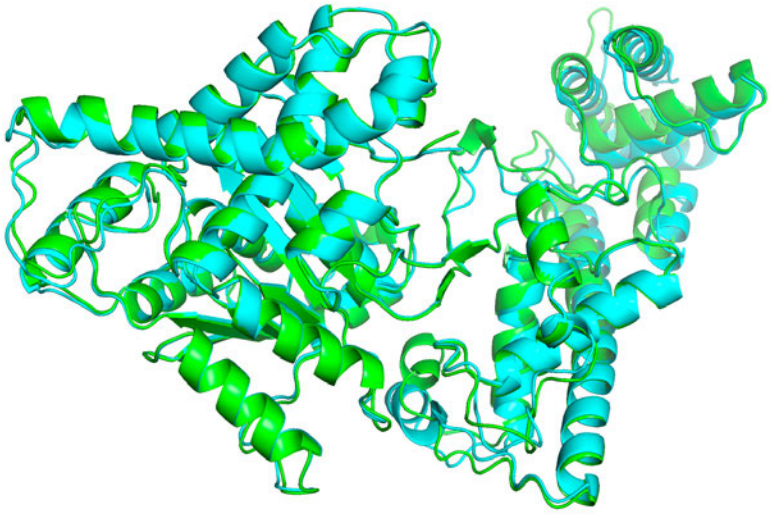


Figure S3.4: Alignment of Tb927.2.4020 APPBP1 *T. b. brucei* (blue) *T. b. equiperdum* type BoTat (green) protein structures. RMSD = 0.672.

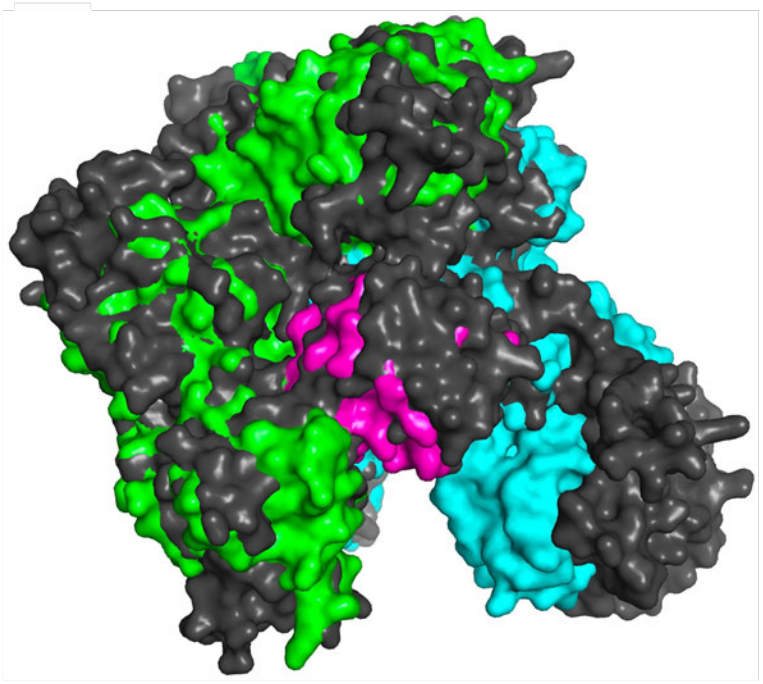


Figure S3.5: Alignment of APPBP1-NEDD8-UBA3 human protein crystal structure and *T. b. brucei* APPBP1-NEDD8-UBA3 complex generated with CollabFold. The human crystal structure was visualised in PyMol using the 1R4M Biological Assembly 1 structure (Walden et al., 2003). The structure is coloured by chain: Green = APPBP1, Blue =UBA3, Pink =NEDD8, Grey = *T. b. brucei* TREU927/4 overlay.

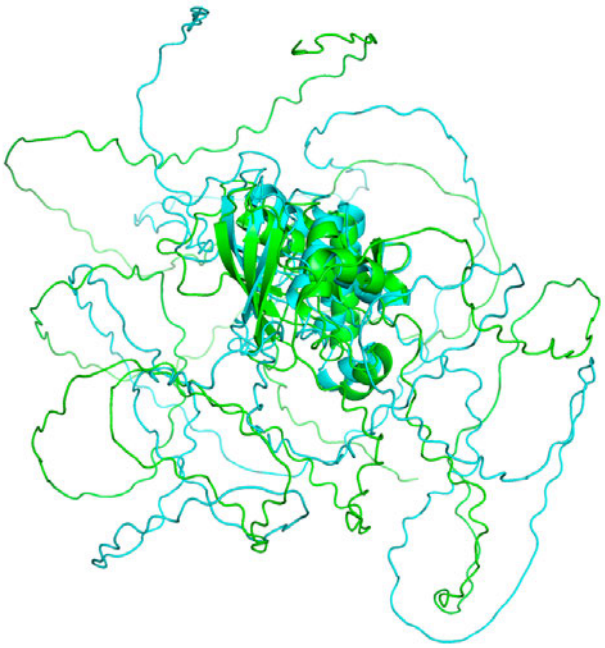


Figure S3.6: Alignment of Tb927.11.6600 Hyp1 *T. b. brucei* (blue) and *T. b. evansi* type B (green) protein structures. RMSD = 14.778.

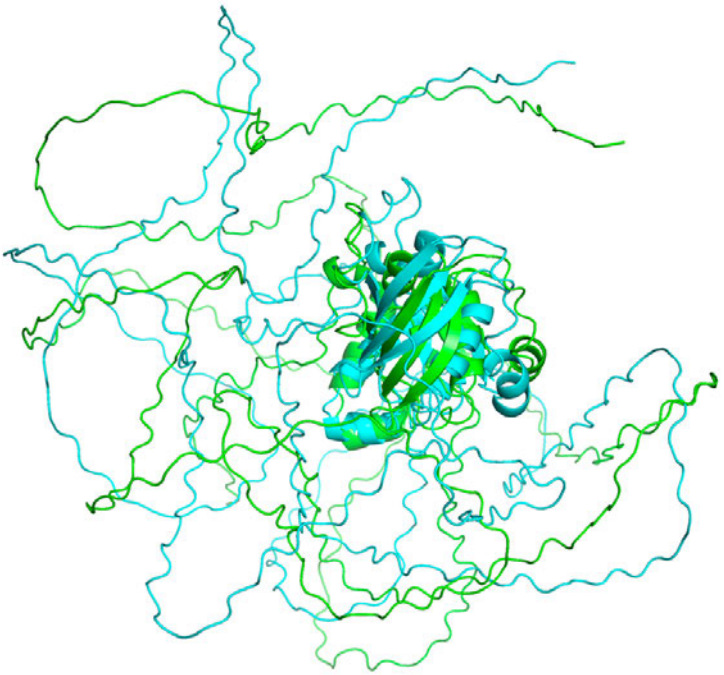


Figure S3.7: Alignment of Tb927.11.6600 Hyp1 *T. b. brucei* (blue) and *T. b. equiperdum* type OVI (green) protein structures. RMSD = 12.469.

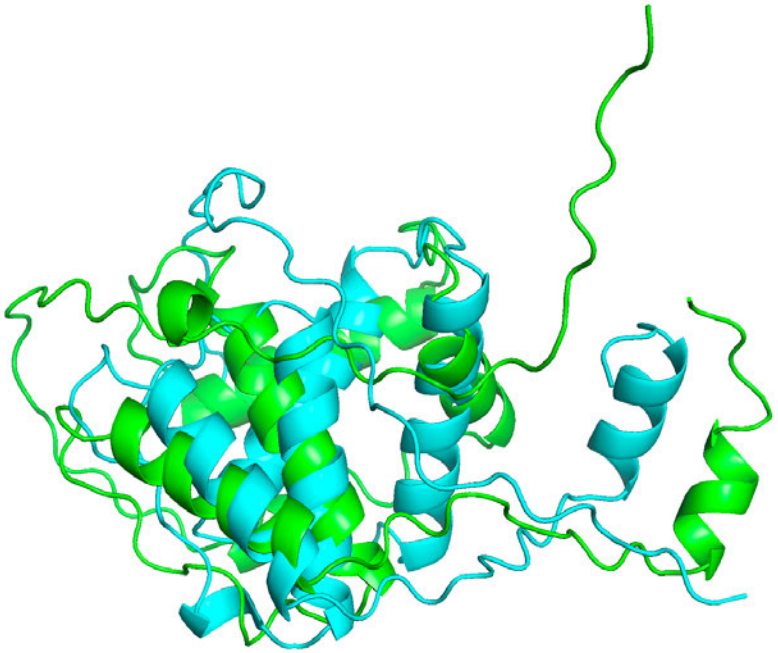


Figure S3.8: Alignment of Tb927.4.3650 PP1 *T. b. brucei* (blue) and *T. b. evansi* type B (green) protein structures. RMSD = 7.925.

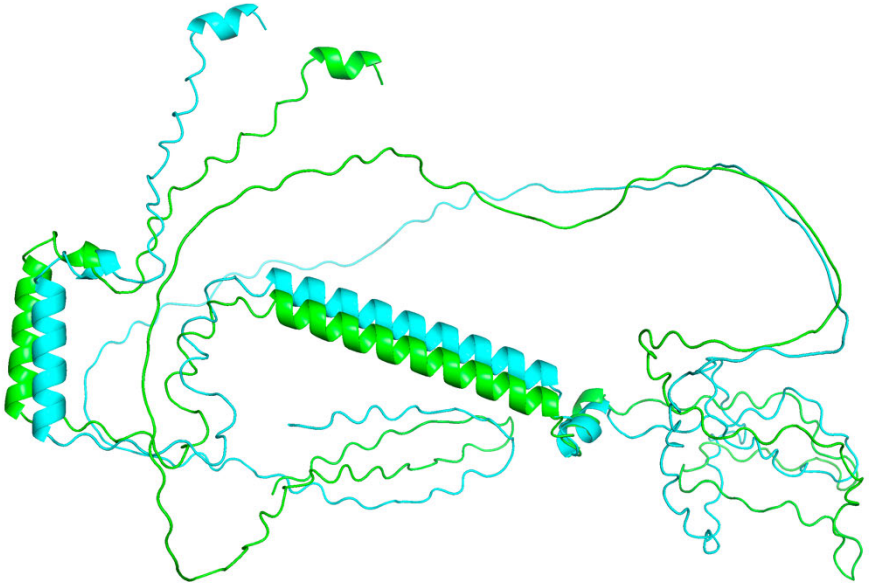


Figure S3.9. Alignment of Tb927.5.2580 hypothetical protein *T. b. brucei* (blue) and *T. b. evansi* type A (green) protein structures. RMSD = 8.299.

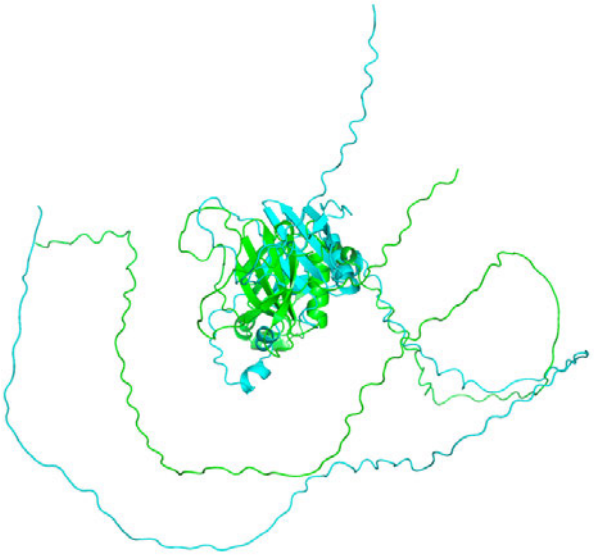


Figure S3.10. Alignment of Tb927.11.3400 FAZ41 *T. b. brucei* (blue) and *T. b. equiperdum* type OVI (green) protein structures. RMSD = 20.508.

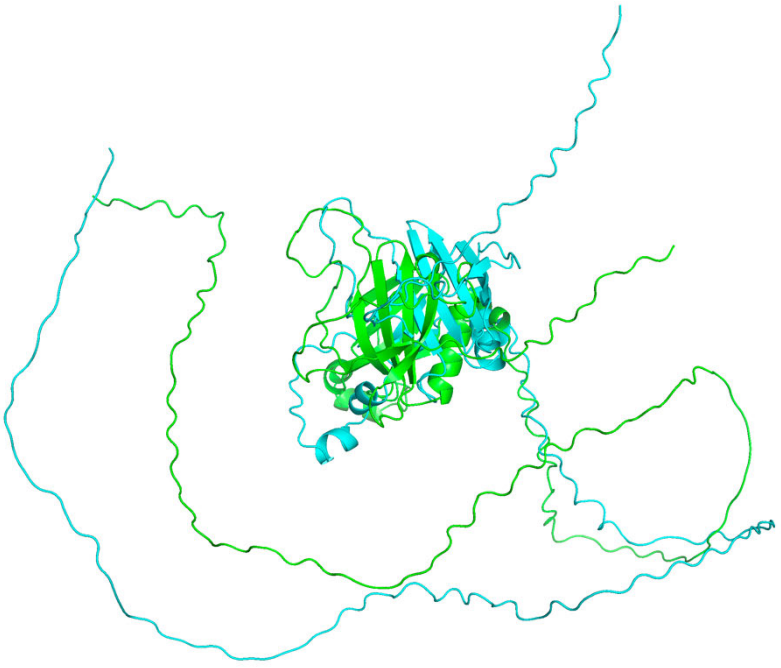


Figure S3.11. Alignment of Tb927.11.3400 FAZ41 *T. b. brucei* (blue) and *T. b. equiperdum* type BoTat (green) protein structures. RMSD = 20.508.

Table S3.1: Isolate assignment to *T. brucei* clade.

Name	Type	Pleomorphic	New to chapter 3
927	*T. b. brucei*	Y	N
940	*T. b. equiperdum* type OVI	N	Y
280104	*T. b. evansi* type A	N	Y
ABBA	*T. b. gambiense*	Y	N
Alfort	*T. b. evansi* type A	N	Y
American-Strain	*T. b. evansi* type A	N	Y
AnTat-3-1	*T. b. evansi type A*	N	Y
AnTat-3-3	*T. b. evansi type A*	N	Y
ATCC-30019	*T. b. evansi type A*	N	Y
ATCC-30023	*T. b. evansi type A*	N	Y
ATCC30019	*T. b. evansi type A*	N	Y
BoTat-1-1	*T. b. equiperdum type BoTat*	N	Y
BoTat.V1	*T. b. equiperdum type BoTat*	N	N
Canadian-Strain	*T. b. evansi type A*	N	Y
Colombia	*T. b. evansi type A*	N	Y
Dodola_943	*T. b. equiperdum type OVI*	N	N
E28	*T. b. equiperdum type BoTat*	N	Y
EATRO2340	*T. b. rhodesiense*	Y	N
EATRO3	*T. b. rhodesiense*	Y	N
HAMBURG	*T. b. evansi type A*	N	Y
IVMt1	*T. b. evansi type C*	N	N
Kazakhstan	*T. b. evansi type A*	N	Y
Kenya	*T. b. evansi type A*	N	Y
MCAM-ET-2013-MU-01	*T. b. evansi type A*	N	Y
MCAM-ET-2013-MU-02	*T. b. evansi type A*	N	Y
MCAM-ET-2013-MU-04	*T. b. evansi type A*	N	Y
MCAM-ET-2013-MU-05	*T. b. evansi type A*	N	Y
MCAM-ET-2013-MU-09	*T. b. evansi type A*	N	Y
MCAM-ET-2013-MU-14	*T. b. evansi type B*	N	Y
MCAM-ET-2013-MU-17	*T. b. evansi type A*	N	Y

Merzouga-56	*T. b. evansi type A*	N	Y
MU09	*T. b. evansi type A*	N	N
MU10	*T. b. evansi type B*	N	N
EATRO 1125	*T. b. brucei*	Y	Y
OVI	*T. b. equiperdum type OVI*	N	Y
OVI.V1	*T. b. equiperdum type OVI*	N	N
Philippines	*T. b. evansi type A*	N	Y
RoTat_1.2	*T. b. evansi type A*	N	N
STIB-816	*T. b. evansi type A*	N	Y
STIB247	*T. b. brucei*	Y	N
STIB386	*T. b. gambiense*	Y	N
STIB805	*T. b. evansi type A*	N	N
STIB818	*T. b. evansi type A*	N	Y
STIB900	*T. b. rhodesiense*	Y	N
SVP	*T. b. evansi type A*	N	Y
Te-Ap-N-D1	*T. b. equiperdum type OVI*	N	Y
TeAp_N_D1.V1	*T. b. equiperdum type OVI*	N	N
Zagora-I-17	*T. b. evansi type A*	N	Y

4 Linking genotype to phenotype in naturally occurring monomorphic *T. brucei*

4.1. Introduction

Clade-specific mutations were identified and prioritised to predict putative causes of monomorphism for each of the five independent monomorphic *T. brucei* clades (Chapter 3). Here we replaced the wild-type alleles of genes in pleomorphic cells with sequences containing mutations present in monomorphic clades, to assess the contribution of these mutations to the monomorphic phenotype. For some of the gene targets, there were multiple monomorphic clades with mutations of interest. Therefore, 16 replacement cell lines were generated (10 monomorphic and 6 pleomorphic controls) making use of a previously generated pleomorphic CRISPR-Cas9 compatible cell line (Rojas et al., 2019).

Further, brain heart infusion broth (BHI) oligopeptides can be used to mimic the oligopeptide signal *in vitro*. BHI causes pleomorphic, developmentally competent, parasites to arrest their cell cycle in 1K1N and express the stumpy stage-specific protein, PAD1 (Rojas et al., 2019). BHI can be added to the growth media (HMI-9) and the developmental competence phenotype of the cell line can then be rapidly screened *in vitro* before the ultimate test of developmental competence *in vivo* (Rojas et al., 2019). A similar method has been used to validate the role of PAG3 in the monomorphic phenotype, and the transition from slender to stumpy form (Wen et al., 2022)

The targets we chose were four known QS genes and two genes which were previously not identified in the QS pathway but have a positive selection pressure in a monomorphic clade (Summarised in Chapter 3). As a control, the target genes were also replaced with the pleomorphic *T. b. brucei* TREU927/4 sequence. These replacement assays controlled for any observed phenotypes in the monomorphic replacement cell line being generated by the processing of the cells during multiple rounds of transfection and selection since these manipulations have the potential to select for monomorphism.

4.2. Methods

Each tool used in the following analysis, plus the version and optional flags used to run the tool, can be found in Table 4.1.

4.2.1. Synthesis of target sequences

In Chapter 3, we identified a list of genes which contained mutations in monomorphic clades which we putatively associated with the monomorphic phenotype. We confirmed that all of the target protein sequences were identical within the clade of interest and the sequence from a representative isolate from each clade was then chosen to synthesise a nucleotide sequence (Genewiz Gene Synthesis). As a control for each of these targets, the pleomorphic TREU927/4 sequence was also synthesised. To aid the downstream processing of the synthesised sequences, we added the enzyme sites HindIII and BamHI at the 5' and 3' terminus, respectively. HindIII and BamHI sites were identified within the coding sequence of Tb927.8.1530 and Tb927.2.4020, respectively. The sites were re-coded to remove the enzyme site whilst maintaining the synonymous amino acid sequence.

4.2.2. Generation of plasmids required for CRISPR/ Cas9 transfections

The synthesised monomorphic sequences described above were generated in the pUC-GW-Kan plasmid (Genewiz) and were transformed into XL1-competent cells. A small-scale plasmid preparation was performed using the GeneJET Plasmid Miniprep Kit (ThermoFisher), following the manufacturer's instructions. The plasmid was then digested using the HindIII and BamHI high-fidelity enzymes (NEB), following the manufacturer's

instructions. The digested gene was excised from an agarose gel using the Monarch DNA Gel Extraction Kit (NEB), following the manufacturer's instructions. The gene was inserted into a pPot plasmid using T4 DNA Ligase (NEB) following the manufacturer's instructions. Blasticidin, hygromycin, phleomycin and G418 resistance pPOT plasmids (Beneke et al., 2017) were used at different stages throughout the following experiments, however, the same drug resistance plasmid was used for each set of experiments on a given gene. Hygromycin and blasticidin were used to replace the wild-type alleles of every target gene apart from GPR89, for which we used Hygromycin and Phleomycin. Add-back of wild-type sequences to the initial replacement cell lines was performed using Phleomycin and G418. The following concentrations of each drug were used: G418 (2.5 µg/ml), Hygromycin (0.5 µg/ml), Puromycin (0.05 µg/ml), Blasticidin (2 µg/ml) and Phleomycin (2.5 µg/ml).

4.2.3. Generation of products required for CRISPR/ Cas9 transfections

The Bar-seq primer design from LeishGEdit (Beneke et al., 2017; Beneke et al., 2019) was used to amplify a repair template from the pPOT plasmid. The LeishGEdit primers allow tagging of the wild-type allele or knock-out of the gene. "Upstream" and "Downstream" primers contain primer binding sites compatible with pPOT plasmids and a 30 nucleotide homology arms for recombination."sgRNA" primers consist of a T7 RNA polymerase promotor (for *in vivo* transcription of RNA), a 20 nucleotide sgRNA target sequence to introduce the double-strand break at a locus of interest and a 20 nucleotide overlap to the sgRNA backbone sequence allowing the generation of sgRNA templates by PCR (Bassett and Liu, 2014; Beneke and Gluenz, 2020; Beneke et al., 2017) (Fig. 4.1). We modified the primers in Fig. 4.1 to allow us to replace the wild-type gene with a tagged monomorphic or pleomorphic sequence we had previously synthesised (4.2.1). We used the upstream forward primer 1 as per the LeishGEdit protocol and, for the downstream reverse primer, we used the start of primer 2 which binds the GS-linker of the pPOT plasmid and the tail of

primer 7 which binds the 3' UTR of the target gene (Fig. 4.1). All of the replacements were N terminally Ty tagged (Bastin et al., 1996) and introduced under 3' UTR endogenous expression. The 5' sgRNA and 3' sgRNA primers were maintained as per the LeishGEdit protocol (Fig. 4.1). The Leishgedit protocol uses the TREU927/4 genome to create primer binding sites. The binding sites of the primer sequences to the endogenous locus were visualised in Geneious (available at https://www.geneious.com/) and the binding sites of each were screened for a mutation in the *T. b. brucei* EATRO 1125 genome, based on SNP calls (Chapter 3). If necessary, primers were recoded to match the EATRO 1125 sequence.

The repair and sgRNA products were amplified for transfection using Phire Hot Start II DNA Polymerase (ThermoFisher), following the manufacturer's instructions. The PCR was run for 30 cycles and an annealing temperature of 65°C and 60°C was used for the sgRNA and repair products, respectively.

4.2.4. CRISPR/ Cas9 transfections

5µg of each product was used for transfections into the *T. brucei* EATRO 1125 AnTat1.1 J1339 (J1339) pleomorphic cell line (Rojas et al., 2019). The J1339 cell line contains the pJ1339 plasmid derived from pJ1173 that carries a puromycin resistance marker, the tet repressor, T7 RNA polymerase and Cas9, which is expressed constitutively (Beneke et al., 2017).

To transfect the parasites, we adapted a protocol described by (MacGregor et al., 2013). $1x10^7$ parasites per transfection were centrifuged at 1,600g for five minutes and washed with 1 ml of 37°C transfection buffer, please see page 178 for the recipe (Schumann Burkard et al., 2011). The cells were then resuspended in 100µl of transfection buffer and 5-10µl DNA (5µg) was added to the positive transfection culture. The same volume of dH_2O was added

to mock-transfected cells. The cells were transferred to an Amaxa cuvette and transformed using a Nucleofector II (Amaxa) with the CD4 T cells Z-001 program. The cells were immediately transferred into 3.5mL of pre-warmed HMI-9. 125µL of this media was then added to each well of rows A and E in a flat bottom 96 well plate. In empty wells (rows B, C, D, F, G and H) 200µL of completed HMI-9 containing 1x concentration of the drug selection was added. The cells were grown for 24 hours at 37ºC. 125µL of pre-warmed completed HMI-9 containing 2X drug selection was then added to rows A and D. These wells were then serially diluted by taking 50µL to the following row (Dilution: 1:5, 1:25, 1:125, 1:625) to isolate clonal transfections. The cells were then grown at 37ºC for 5-7 days. The control plate, with cells mock transfected using dH_2O, was monitored to confirm all the cells had died. Wells which contained live cells in the positive plate were treated as successful transfectants, with those at the highest dilutions considered clonal lines provided less than 50% of wells contained growing parasites. The parasites were split into different wells straight after transfection, therefore, each well is treated as an independent transfection event.

4.2.5. gDNA PCR to confirm successful transfection

Six clones from each transfection were grown in 5 ml of HMI-9 and gDNA was extracted using the DNeasy Blood & Tissue Kit with an RNAse A step (Qiagen). Successful replacements were then validated using the section of primers 1 and 7 which bind the genomic DNA of the target gene. Once a double allele replacement had been confirmed, the PCR product was sequenced to confirm the correct sequence had been used to replace the original alleles (Genewiz Sanger Sequencing).

4.2.6. Western blot to confirm expression of the epitope-tagged replacement gene

Expression of the replacement gene was then confirmed via Western blot using the Ty1 epitope-specific BB2 antibody (Bastin et al., 1996). Cells were suspended in PEME lysis buffer (Höög et al., 2010) and Laemmli loading buffer (Laemmli, 1970) (see supplementary material for recipe), boiled for 5 minutes and then sonicated for 10 minutes using a Bandelin Sonorex. For the western blot of the GPR89 replacement cell line, the samples were incubated at 37°C for 15 minutes, instead of boiling, following the method of (Rojas et al., 2019). The protein sample was then run on a 10-12% polyacrylamide gel, dependent on the predicted protein size, for 10 minutes at 100V, followed by 60 minutes at 120V. The protein was transferred to a nitrocellulose membrane using the BioRad Trans-Blot Turbo Transfer System**.** The membrane was washed with dH_2O and incubated in Ponceau S (Thermo Scientific) to visualise the protein transfer. The membrane was then washed with dH_2O and incubated in LI-COR Intercept (TBS) Blocking Buffer for a minimum of 30 minutes at room temperature. The membrane was stained with the primary BB2 anti-Ty1 antibody (Bastin et al., 1996) diluted (1:1000) in 2% BSA in TBS-T (0.1% Tween in TBS) and then the secondary LI-COR IRDye 800CW Goat anti-Rabbit IgG antibody (1:500). Each step was performed at room temperature for 1 hour. Between each antibody, the membrane was washed 3X for 5 minutes with TBS-T. During each incubation, the membrane was agitated throughout. The membrane was imaged with the LI-COR Odyssey imager system.

FIJI (Schindelin et al., 2012) was used to measure the BB2 anti-Ty1 and Ponceau S (Thermo Scientific) intensity. A background reading was taken from each of these bands. Each band was normalised by the background reading and the ratio of the BB2 anti-Ty1 intensity over the Ponceau S stain value was calculated (BB2 normalised value/Ponceau S normalised value). The clone with the highest ratio for each gene was used to screen the cell

line for developmental competence. When the expressed protein could not be detected, up to 3 clones were used for the phenotype screen, depending on the number of successfully generated clones.

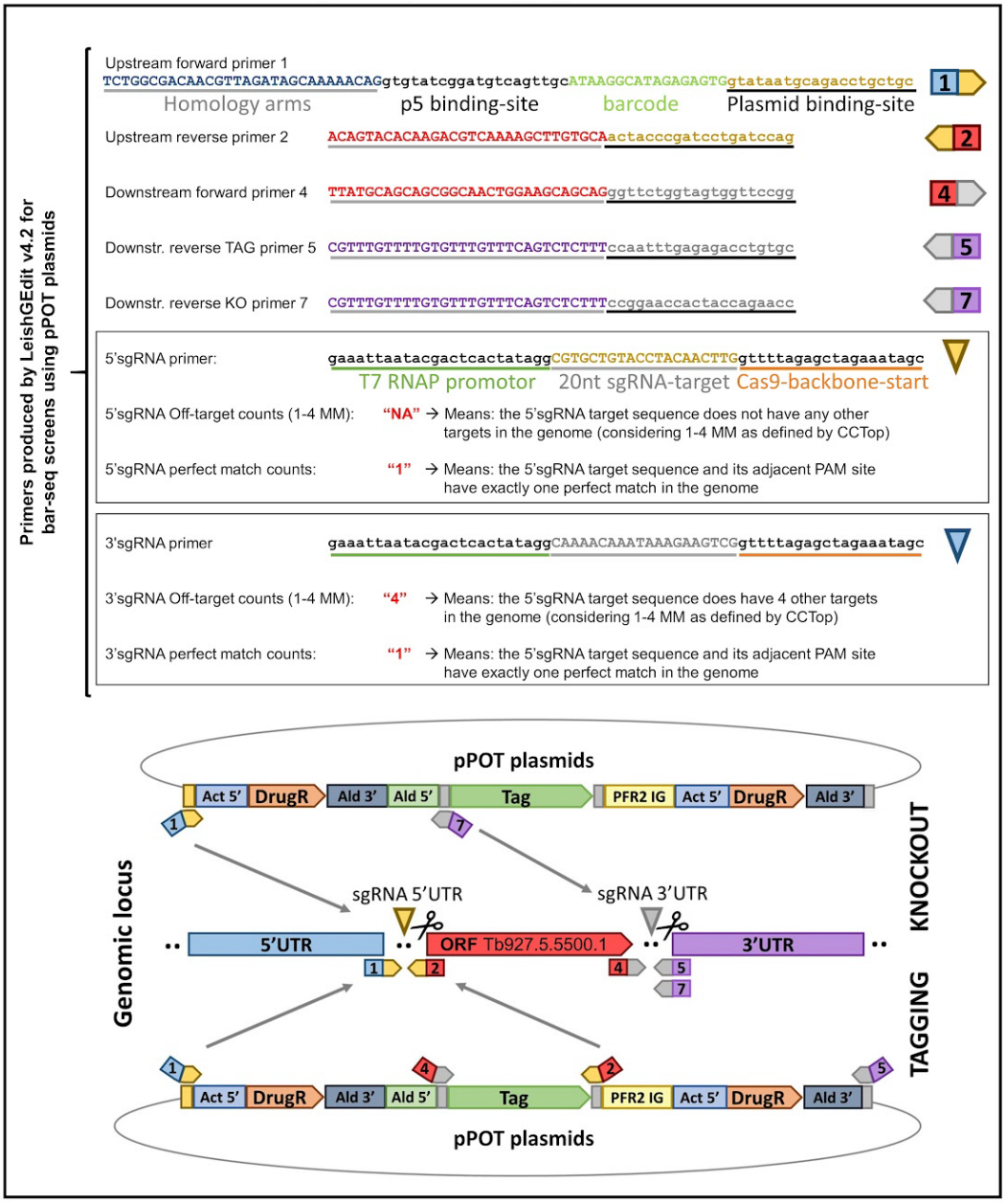


Figure 4.1: Schematic of the LeishGedit tagging and knockout approach. The image is taken from http://www.leishgedit.net/Home.html. "Upstream" and "Downstream" primers contain primer binding sites compatible with pPOT plasmids and a 30 nucleotide homology arms for recombination. "sgRNA" primers consist of a T7 RNA polymerase promotor (for *in vivo* transcription of RNA), a 20 nucleotide sgRNA target sequence to introduce the double-strand break at a locus of interest and a 20 nucleotide overlap to the sgRNA backbone sequence allowing the generation of sgRNA templates by PCR (Bassett and Liu, 2014; Beneke and Gluenz, 2020; Beneke et al., 2017).

4.2.7. *In vitro* developmental competence phenotype screen

Replacement clones were seeded at $1x10^5$ cells/ml and grown in either HMI-9 or HMI-9 supplemented with 15% BHI in triplicate. The parasite population density was counted at 0, 24, 48 and 72 hours using the Beckman Coulter Z2 Cell and Particle Counter.

A separate culture was used to generate immunofluorescence images of the gene replacement cell lines. The gene replacement clones were seeded at $1x10^5$ cells/ml and left for 48 hours. Cell cycle status and PAD1 expression were analysed by 4',6-diamidino-2-phenylindole (DAPI) (100 ng/ml) and an anti-PAD1 antibody (MacGregor and Matthews, 2012). Briefly, 10ml of the culture was centrifuged at 1,600g for eight minutes at room temperature. The cells were washed with vPBS and then resuspended in vPBS and 8% PFA (1:1) and incubated for 10 minutes at room temperature. 2µL of IGEPAL® CA-630 (Sigma-Aldrich) 10% in PBS was added and the cells were incubated for a further 10 minutes. The cells were centrifuged at 900g for eight minutes and resuspended in 150µL 0.1% glycine in vPBS and incubated for 10 minutes. The cells were then centrifuged at 900g for eight minutes and resuspended in 50µL of vPBS. 25µL of the fixed cells were added to a polylysine cell, prewashed with 70% ethanol, and allowed to dry for one hour. In an attempt to visualise the surface localisation of GPR89, the experiment was repeated without the addition of IGEPAL® CA-630, which prevented the permeabilisation of the cells.

Next, the cells were rehydrated in PBS for five minutes and blocked by incubation of the cells in PBS 2% BSA for one hour. The blocking buffer was removed and anti-PAD1 (1:1000) PBS 0.2% BSA was added to the cells and incubated for one hour. For the GPR89 replacement cell lines, BB2 anti-Ty1 (Bastin et al., 1996) (1:5) PBS 0.2% BSA was also added at the same time. The cells were washed 5x with PBS and then incubated in the

Alexa Fluor 568 Goat anti-Rabbit secondary antibody (1:500) (Invitrogen) PBS 0.2% BSA for one hour. For the GPR89 replacement, the cells were also incubated in the Alexa Fluor 488 Goat anti-mouse secondary antibody (1:500) (Invitrogen) PBS 0.2% BSA for one hour. 25µL of DAPI (10µg/ml),) in PBS was added to the cells and incubated for five minutes. The cells were then washed five times with PBS. Finally, the slides were mounted with Fluoromount-G Mounting Medium (Invitrogen) following the manufacturer's instructions. All steps were performed at room temperature. The slides were then imaged with a ZEISS Axio Imager 2 using a 40x objective with a Prime BSI Camera. Images were generated with Micro-Manager 2.0 (Edelstein et al., 2014) and analysed with FIJI (Schindelin et al., 2012). The number of 1K1N, 2K1N and 2K2N (Woodward and Gull, 1990) parasites were counted along with the number of PAD1-positive cells.

The GPR89 gene replacement images were also taken with a ZEISS Axio Imager 2 100x objective using the same method described above. These images were used to quantify the anti-Ty1 BB2 fluorescence. Images were imported into FIJI and the outline of the cell was drawn with the freehand selection tool. The area and mean grey value (signal intensity of anti-Ty1 BB2 fluorescence) were then measured for 50 cells from the cell culture of each replacement cell line in HMI-9 or HMI-9 supplemented with 15% BHI.

4.2.8. Motility assay

Cells were grown to a density of $5x10^5$ cells/ml. 1 ml of these cells was centrifuged at 800g for five minutes and resuspended in 40µl of HMI-9. The resuspended cells were then placed onto a slide and a cover slip was mounted using Vaseline. The slides were then incubated for 5 minutes (37°C, 5% CO_2). Each slide was visualised at 40x magnification using the Olympus CKX53 and a time-lapse image was generated with a QImaging Retiga-2000R digital camera. 100 images were then taken, one image every 0.25 seconds. One slide was

made for each cell line on three consecutive days. Each slide was imaged 3 times per day (3 biological replicates and 3 technical replicates = 9 time-lapse series per cell line). The images were imported into FIJI (Schindelin et al., 2012) and the background was subtracted from the images. The image was then inverted to improve the contrast between the cell and the background. TrackMate (Ershov et al., 2021) was used to identify cells using the LoG detector. An initial threshold was applied to filter the detected 'spots'. The Kalman tracker was used to track cell motility. A summary of each track was then exported for each cell line.

4.2.9. *In vivo* infections

In vivo infections were performed by Dr Frank Venter. Three female MF1 mice were infected for each replacement cell line. Immunocompromised mice were generated via cyclophosphamide treatment (25 mg/ml) 24 hours before infection. Each mouse was infected with 1,000 parasites via intraperitoneal injection. Parasitaemia was monitored from 3 days post-infection by tail snip. Blood-smeared slides were air dried at room temperature for five minutes and then fixed via submersion in methanol at -20°C for 10 minutes or stored in methanol at -20°C. The slides were then processed for immunofluorescence analysis as in section 4.2.2.

4.2.10. Computational tools

Unless stated otherwise, figures and statistical analysis were created with ggplot2 (Wickham, 2016) and ggrepel (available at https://cran.r-project.org/web/packages/ggrepel/index.html) in R (R Core Team, 2022). For the comparison of growth curves, a t-test was performed between the test and control replacement population density at each time point. A Wilcoxon rank-sum test was performed

to compare the anti-Ty1 fluorescence of GPR89 and the motility of the FAZ41 replacement cell line.

Table 4.1: Computational tools and optional flags used for the analysis in this chapter.

Tool	Version	Optional flags	Citation
Geneious	2022 1.1		Available at http://www.geneious.com
Micro-Manager	2.0	100 images taken, one image every 0.25 seconds	(Edelstein et al., 2014)
FIJI	2.3.0/1.53f		(Schindelin et al., 2012)
TrackMate	7.6.1	LoG detector (estimated object diameter = 15, quality threshold = 0). Quality (> 0.7). Kalman tracker (initial search radius = 20, search radius = 15, max frame gap = 15).	(Ershov et al., 2021)
ggplot2	3.3.0		(Wickham, 2016)
ggrepel	0.8.1		Available at CRAN - Package ggrepel
R	3.6.1		(R Core Team, 2022)

4.3. Results

4.3.1. Known QS pathway genes contain mutations in monomorphic clades which cause insensitivity to BHI oligopeptides

Tb927.2.4020 – NEDD8-activating enzyme E1 (APPBP1)

Clade specific non-synonymous mutations in APPBP1, including the addition of putative phosphorylation sites, are found in the monomorphic clades *T. b. evansi* type A (4), *T. b. evansi* type IVM-t1 (5) and *T. b. equiperdum* type BoTat (4) in comparison to TREU927/4 (Fig. 3.8). The amino acid sequence is identical for all isolates within each clade.

Replacement of APPBP1

Homozygous replacement of the wild-type alleles of APPBP1 was confirmed via amplification of genomic DNA surrounding the target gene from transfected cell lines. Transfected parasites have longer amplicons caused by the insertion of the drug resistance and TY-tag, along with the replacement gene (blasticidin resistance 2,992 and hygromycin resistance 3,194) compared to amplification of the region in untransfected parasites (1,992 bases) (Fig. S4.1a). The expression of the replaced gene was confirmed by western blotting using the anti-Ty1 antibody BB2 (Fig. S4.1e-f). The protein expression was quantified and the PCR product of the target gene amplification was sequenced to ensure the correct sequence had been inserted.

The replacement cell lines all grew well in standard growth media (HMI-9), with no significant difference between monomorphic and control replacement cell lines at 72 hours (Fig. 4.2). The growth of cells expressing the *T. b. evansi* type IVM-t1 sequence (clone 3) lagged at 48 hours but eventually reached the same population density as cells expressing the wild type *T. b. brucei* sequence (clone 1) (Fig. 4.2c). Cells expressing the *T. b. equiperdum* type BoTat (clone 4) and *T. b. evansi* type A (clone 1) sequences arrested their growth once exposed to HMI-9 media supplemented with BHI, a mimic of the oligopeptide signal, similar to cells expressing the *T. b. brucei* sequence (Fig. 4.2a&b). In contrast, the cells expressing the *T. b. evansi* type IVM-t1 sequence continued to grow in BHI-supplemented HMI-9 (Fig. 4.2c). The continued growth in BHI indicates an insensitivity to the oligopeptides-based QS signal, potential disruption to the QS pathway, and the development of monomorphism.

Immunofluorescence confirmed the majority of cells arrest their cell cycle in 1K1N configuration upon exposure to BHI for the cell lines which expressed the *T. b. brucei* (92%), *T. b. evansi* type A (88%) and *T. b. equiperdum* type BoTat (88%) APPBP1 sequence. At 48 hours, the parasites also expressed the stumpy stage-specific protein PAD1 when exposed to BHI (*T. b. brucei* 46%, *T. b. evansi* type A 40% and *T. b. equiperdum* type BoTat 55%). In contrast, cells expressing the *T. b. evansi* type IVM-t1 APPBP1 sequence presented resistance to growth arrest in BHI oligopeptides with fewer cells in 1K1N configuration (64%) and no PAD1 expression (Fig. 4.3).

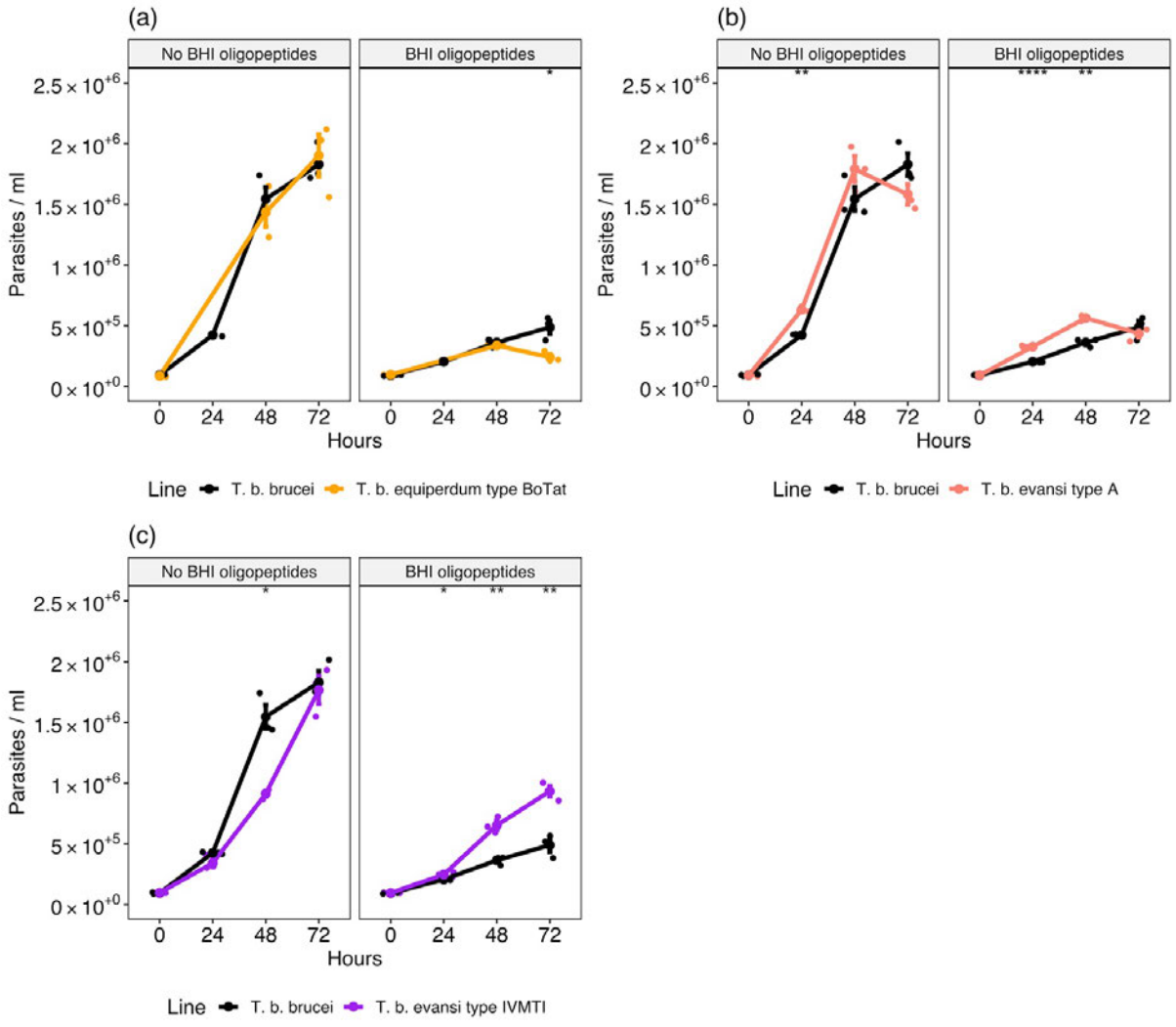


Figure 4.2: Replacement of APPBP1 with the orthologous *T. b. evansi* type IVM-t1 sequence causes insensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) cells expressing the pleomorphic *T. b. brucei* APPBP1 sequence compared to cells expressing the monomorphic **(a)** *T. b. equiperdum* type BoTat sequence, **(b)** *T. b. evansi* type A sequence or **(c)** *T. b. evansi* type IVM-t1 sequence. The cells were either grown in HMI-9 (No BHI oligopeptides) or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

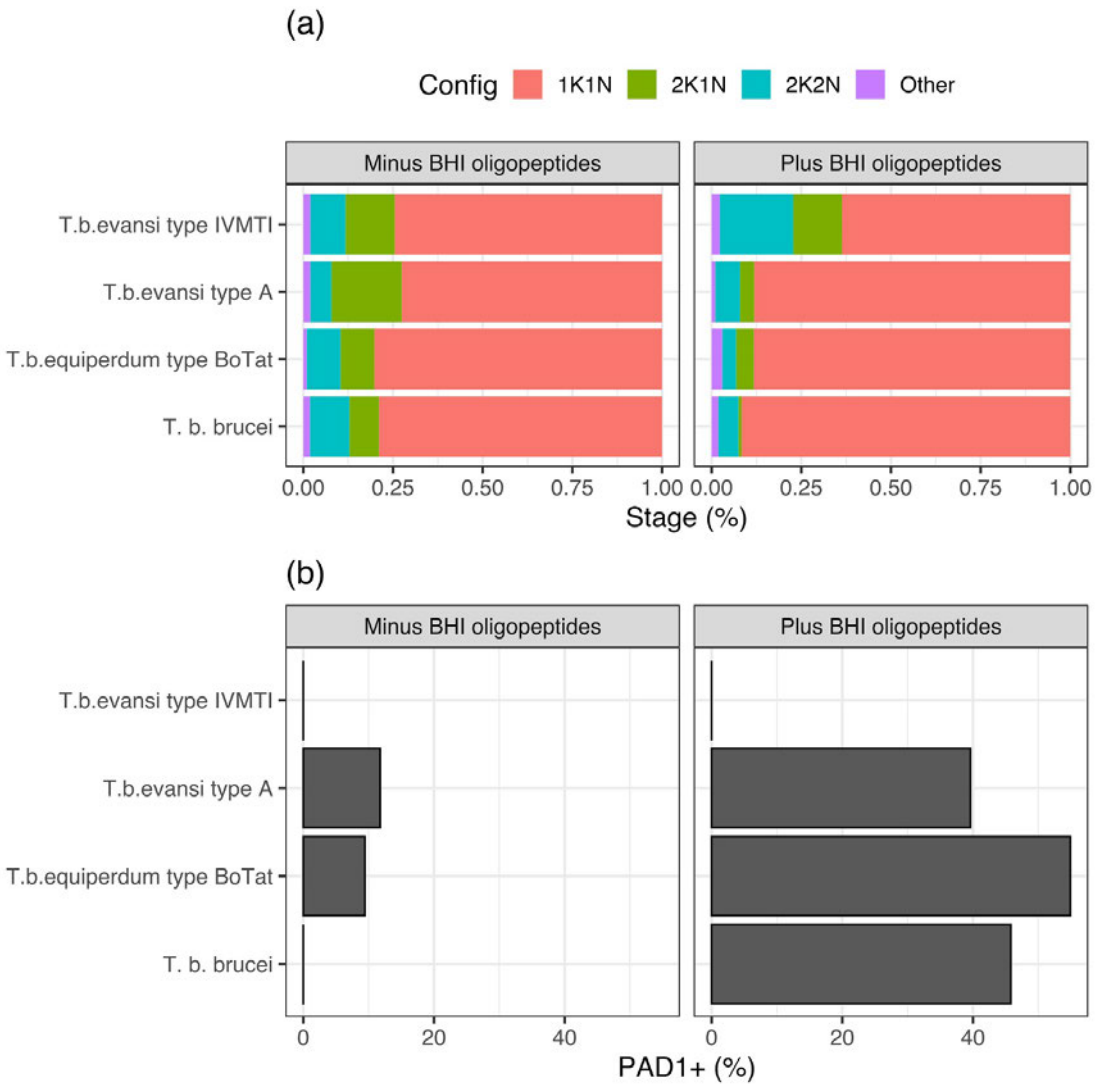


Figure 4.3: Replacement of APPBP1 with the orthologous *T. b. evansi* type IVM-t1 sequence induces insensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the APPBP1 monomorphic *T. b. equiperdum* type OVI sequence, *T. b. evansi* type A sequence, *T. b. evansi* type IVM-t1 sequence or a pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

Mutations in the *T. b. evansi* type IVM-t1 APPBP1 sequence cause insensitivity to BHI oligopeptides

To confirm relative insensitivity to the BHI oligopeptides, the growth profile in Fig. 4.2c was repeated. In addition, two separate clones (clones 4 and 5) expressing the orthologous APPBP1 sequence from *T. b. evansi* type IVM-t1 were tested for reduced developmental competence. All three clones displayed the same reduced sensitivity to BHI oligopeptides (Fig. 4.4).

Clone 3, expressing the *T. b. evansi* type IVM-t1 sequence (Fig. 4.4b), was subjected to single allele add-back with the pleomorphic *T. b. brucei* APPBP1 sequence. Two separate heterozygous add-back clones (add-back brucei/IVMT1 clones G1 and P1) were generated (Fig. S4.1b) and we confirmed the expression of the add-back with the anti-Ty1 antibody BB2 (Fig. S.4.e-f). Both heterozygous add-back clones retained reduced sensitivity to BHI, although they did display slower growth in unsupplemented HMI-9 (Fig. 4.5a&b). The two single allele add-back clones were then subjected to a second transfection to generate homozygous add-backs with the wild-type sequence. Two clones were generated from each of the initial heterozygous add-back cell lines (Fig. S4.1c) and expression of the replacement sequence was confirmed (Fig. S4.1g-h. All four homozygous add-back clones (add-back clones P+G2&3 and clones G+P2&3) were found to arrest their growth once exposed to BHI oligopeptides, demonstrating restoration of the sensitivity to BHI-induced arrest after pleomorphic sequence add-back (Fig. 4.5c&d). Homozygous add-back cells arrested their cell cycle in 1K1N when grown in BHI, in contrast to cells expressing the *T. b. evansi* type IVM-t1 sequence (Fig. 4.6a). Further, the add-back clones also expressed PAD1, although Clone G+P2 did not express PAD1 to the level of cells expressing the pleomorphic *T. b. brucei* sequence (Fig. 4.6b). Nonetheless, the add-back of the *T. b. brucei* sequence proved that the reduced developmental competence of cells expressing the *T. b. evansi* type IVM-t1 APPBP1 sequence is not caused simply by the long-term passage of the parasites.

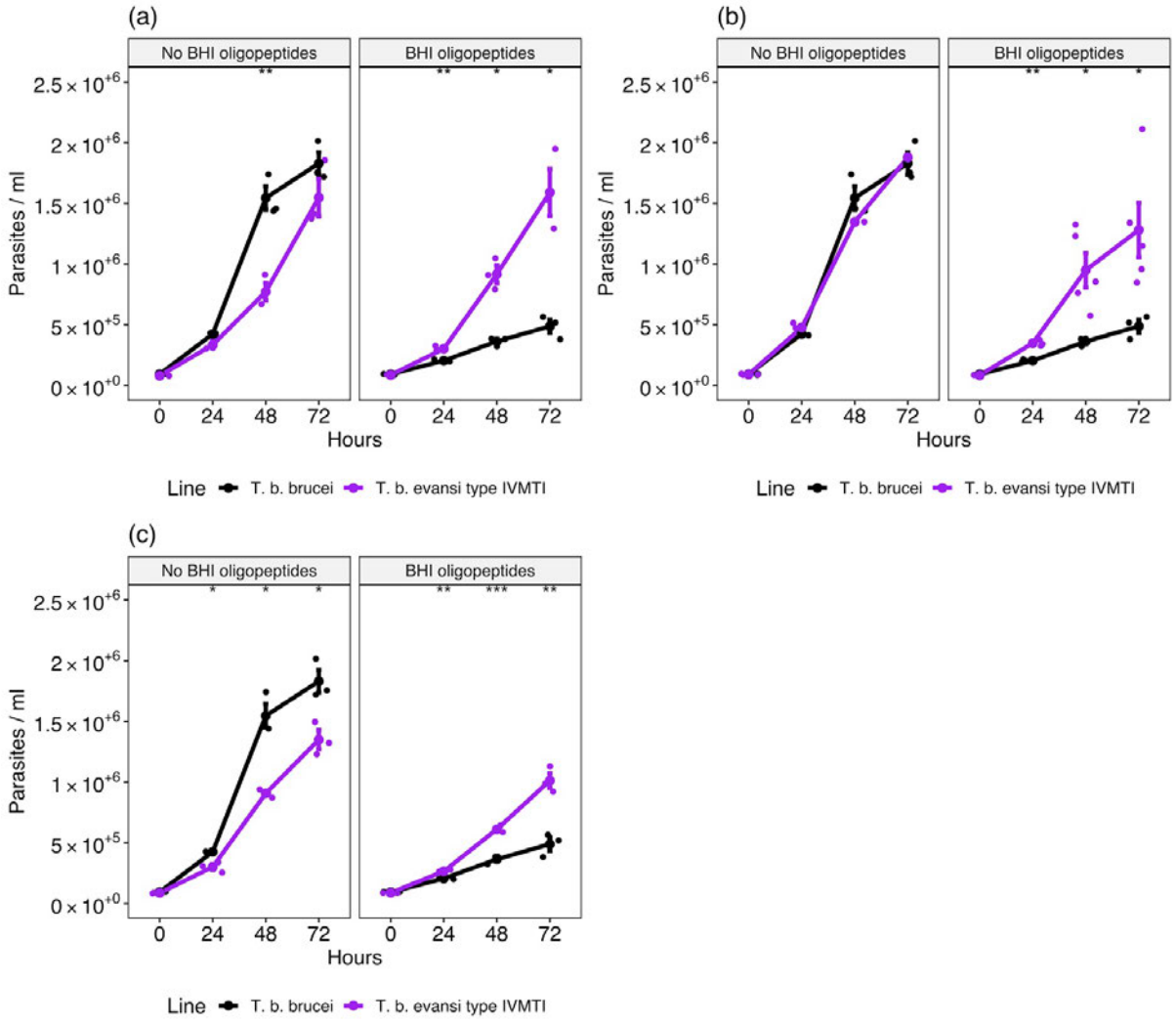


Figure 4.4: Insensitivity to BHI oligopeptides was replicated in three independent clones. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) (J1339 (Rojas et al., 2019)) expressing the pleomorphic *T. b. brucei* APPBP1 sequence compared to three clones **(a-c)** expressing the *T. b. evansi* type IVM-t1 sequence. The cells were either grown in HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

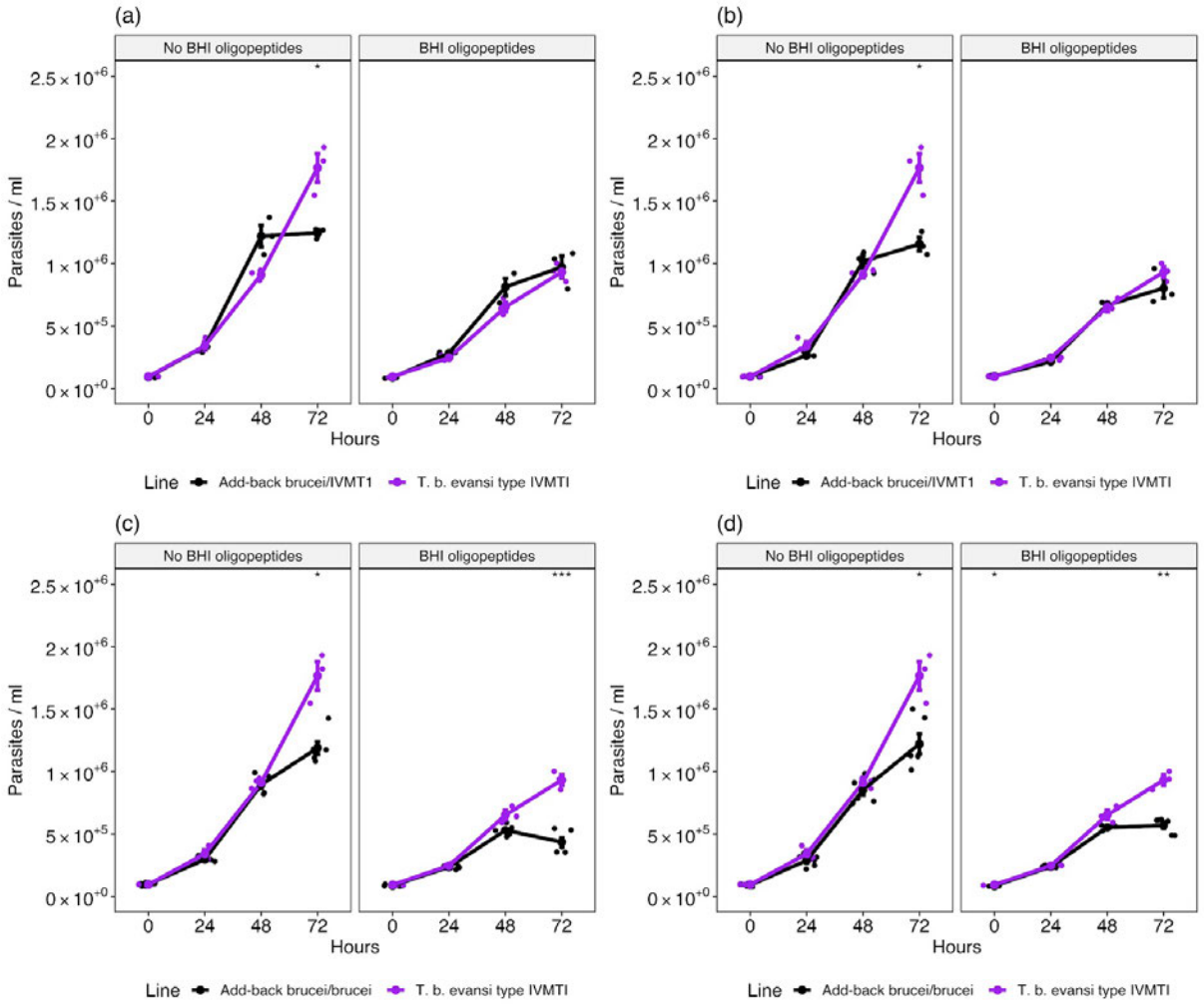


Figure 4.5: Add-back of the *T. b. brucei* APPBP1 sequence to the *T. b. evansi* type IVM-t1 replacement cell line restores pleomorphism. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the monomorphic APPBP1 *T. b. evansi* type IVM-t1 sequence and then subjected to **(a&b)** heterozygous and **(c&d)** homozygous add-back of the pleomorphic *T. b. brucei* APPBP1 sequence. The cells were either grown in HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

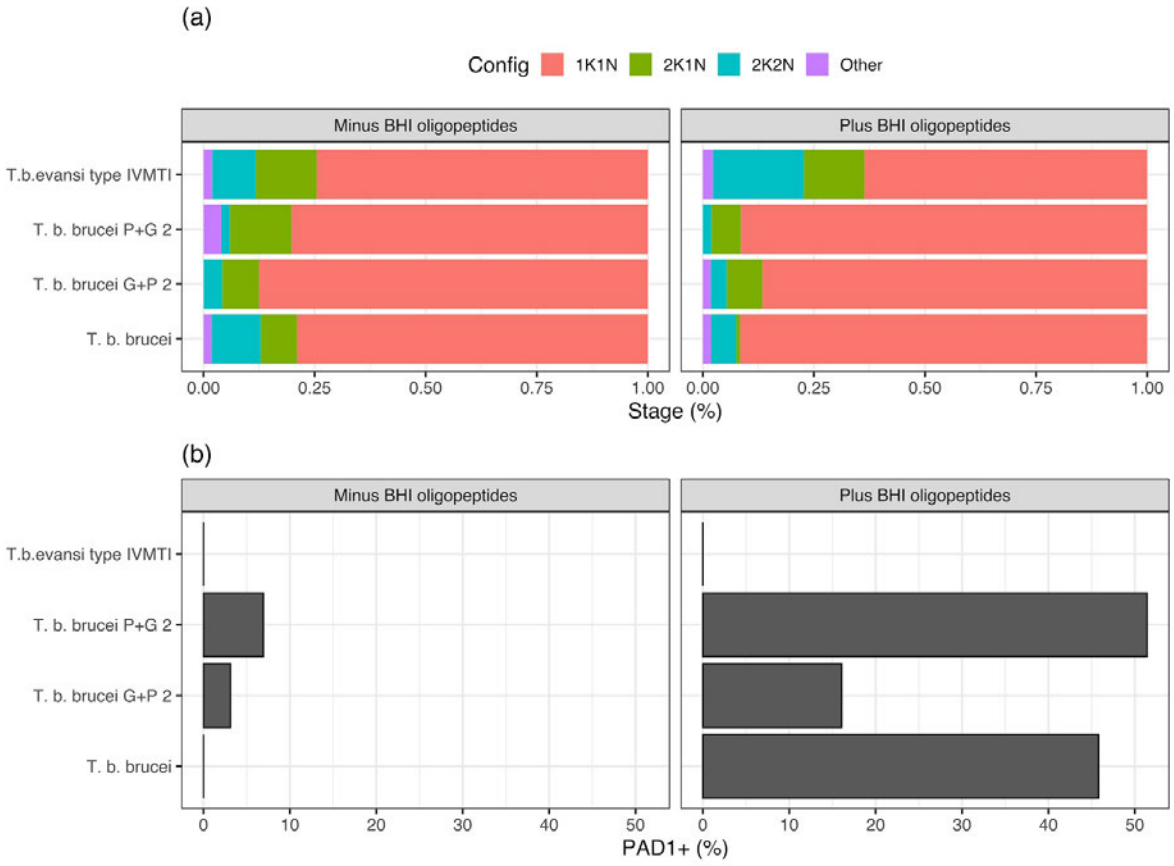


Figure 4.6: Add-back of the *T. b. brucei* APPBP1 sequence to the *T. b. evansi* type IVM-t1 replacement cell line restores pleomorphism. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the APPBP1 pleomorphic *T. b. brucei* sequence, the monomorphic *T. b. evansi* type IVM-t1 sequence, and the *T. b. evansi* type IVM-t1 replacement cell line subjected to homozygous add-back of the pleomorphic *T. b. brucei* sequence (*T. b. brucei* P+G2; *T. b. brucei* G+P2). The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

The *T. b. evansi* type IVM-t1 APPBP1 G224S mutation alone does not cause developmental incompetence

We noted that the APPBP1 sequences of *T. b. evansi* type A and *T. b. evansi* type IVM-t1 differ from one another by a single non-synonymous mutation, G224S. Replacement of APPBP1 with the *T. b. evansi* type A sequence did not impact developmental competence (Fig. 4.2b). To gain evolutionary insight into this amino acid, the orthogroup that contained APPBP1 was extracted from a comparison of kinetoplastid proteomes (Oldrieve et al., 2022). Sequence alignment highlighted 39.9% sequence identity across kinetoplastids. At position 224, glycine is common across subspecies of *T. brucei* and is also present in the basal species, *Bodo saltans*. Therefore, this mutation appears to be a reversion to the ancestral sequence. However, serine is the most common amino acid at this site, including the most basal trypanosomatid species included in this analysis, *P. confusum* and closely related species in the *Trypanosoma* genus, such as *T. congolense* (Fig. 4.7). The G224S mutation is found in a region of high sequence divergence amongst trypanosomatids, with low sequence similarities between the *Leishmania* and *Trypanosoma* isolates. G224S creates a potential cell-division cycle 2 (cdc2) phosphorylation site (Blom et al., 1999; Blom et al., 2004) in the monomorphic *T. b. evansi* type IVM-t1 sequence, which is absent in pleomorphic *T. b. brucei.* Phosphorylation by cdc2 kinases is an evolutionarily conserved form of cell cycle regulation (Mottram and Smith, 1995).

A sequence was generated which was identical to the *T. b. brucei* APPBP1 sequence, with just the single G224S mutation that differed between the sequences present in *T. b. evansi* type IVM-t1 and *T. b. evansi* type A. Cell lines expressing the G224S APPBP1 sequence as either heterozygous (clone 1) or homozygous (clone 2) replacements were generated (Fig. S4.1d, g-h). The cells expressing the G224S sequence arrested their growth upon exposure to BHI oligopeptides equivalently to those expressing the wild-type *T. brucei* sequence (Fig.

4.8). Further, the G224S homozygous replacement cell line arrested its cell cycle (Fig. 4.9a) and expressed PAD1 upon exposure to BHI (Fig. 4.9b). Therefore, the G224S mutation alone is not sufficient to generate a developmentally incompetent phenotype.

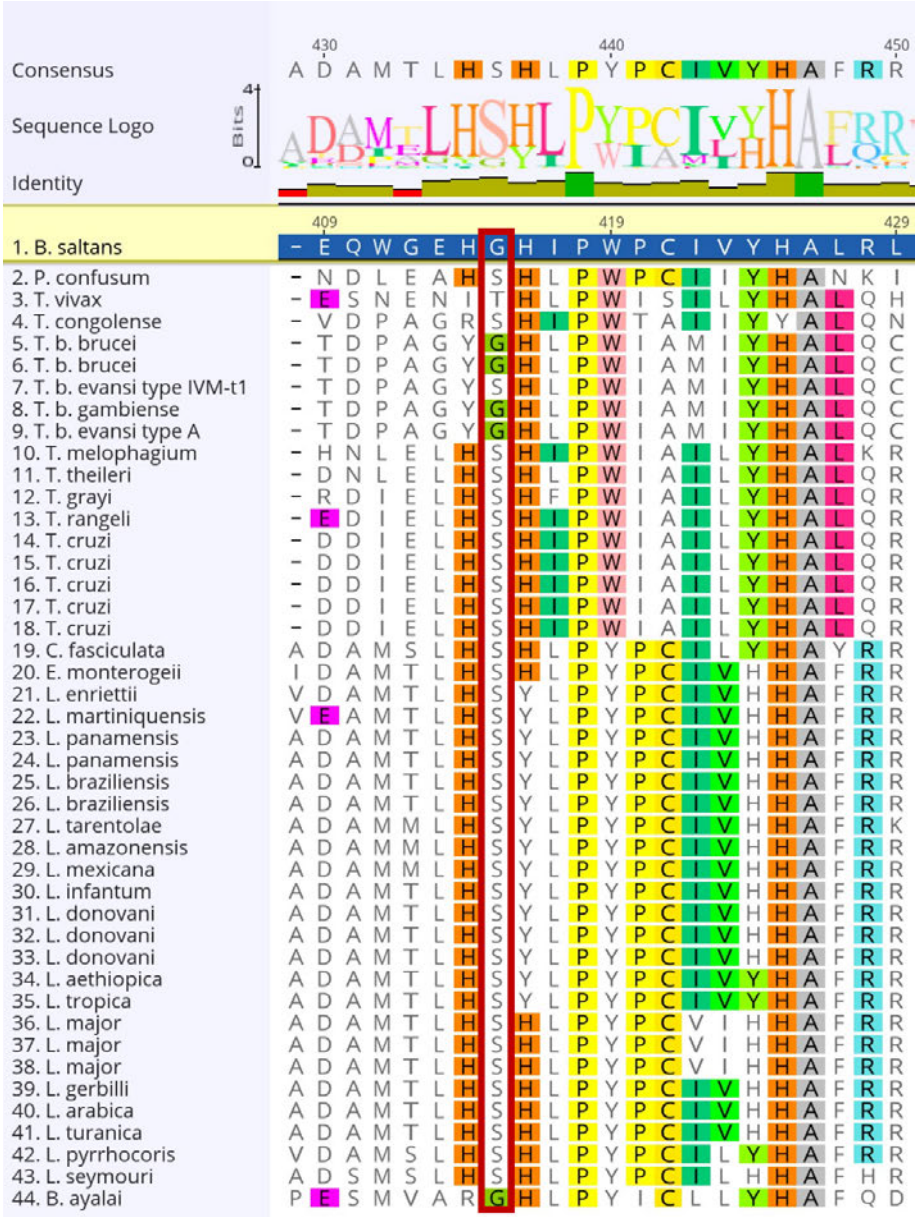


Figure 4.7: The G224S mutation is a reversion to the ancestral state. Amino acid sequence alignment of kinetoplastid APPBP1 sequences (Tb927.2.4020) surrounding the G224S mutation (position 436 in the consensus sequence, outlined in red).

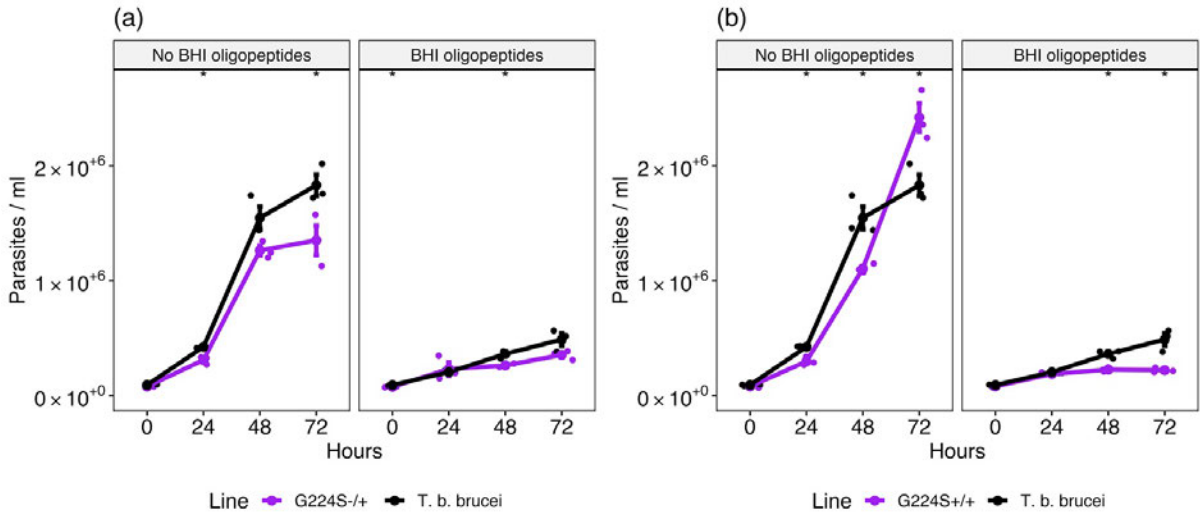


Figure 4.8: The G224S mutation alone does not cause insensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* APPBP1 sequence compared to **(a)** heterozygous or **(b)** homozygous expression of the *T. b. brucei* APPBP1 sequence with the G224S mutation. The cells were either grown in HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

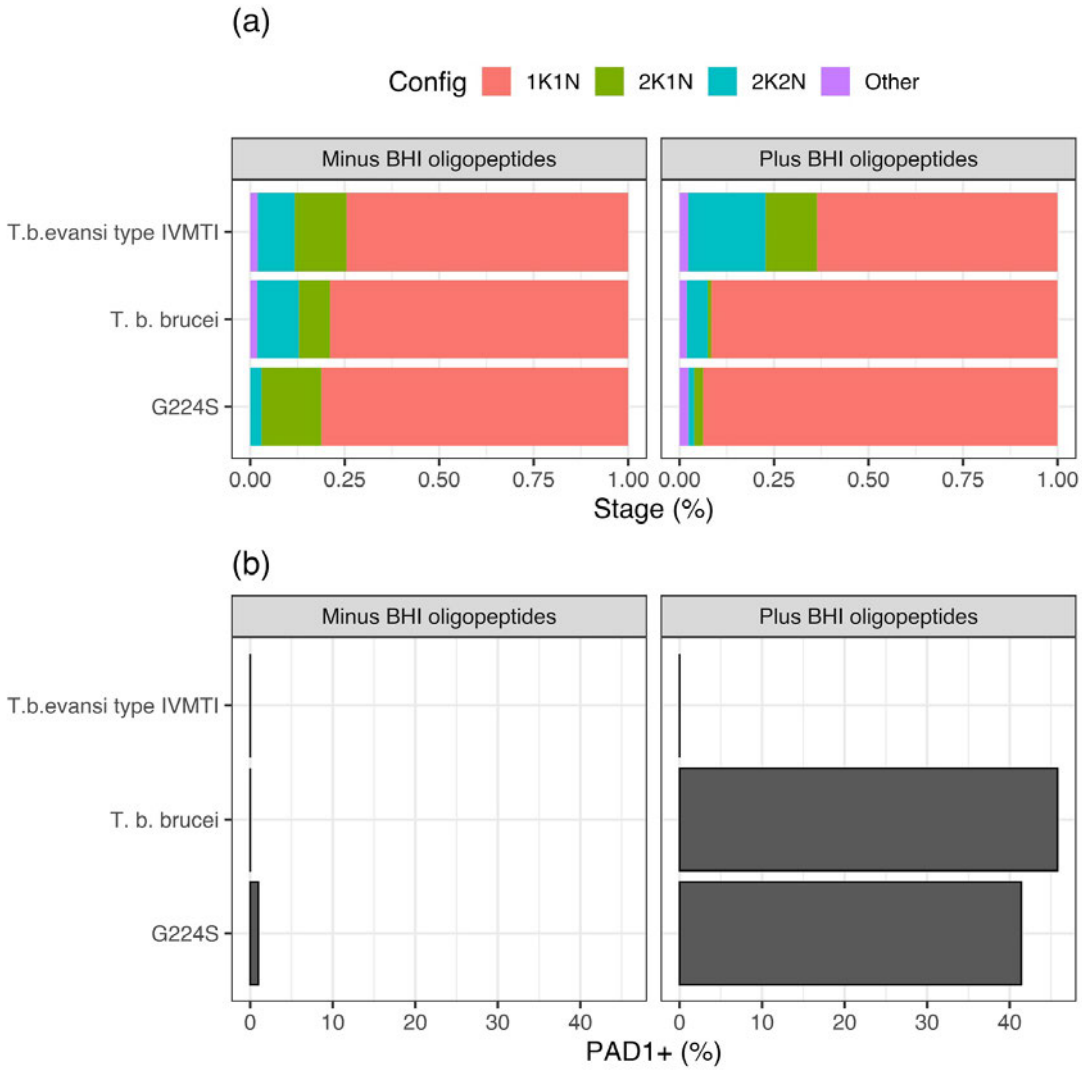


Figure 4.9: The G224S mutation alone does not cause insensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the APPBP1 monomorphic *T. b. evansi* type IVM-t1 sequence, the *T. b. brucei* sequence containing the G224S mutation and the pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

In vivo infections with cells expressing *T. b. evansi* type IVM-t1 APPBP1 sequence display delayed developmental progression

After validating cells expressing the *T. b. evansi* type IVM-t1 APPBP1 sequence *in vitro*, we tested their developmental competence *in vivo*. Dr Frank Venter performed *in vivo* infections into groups of MF1 mice and monitored parasitaemia by blood smears. Cells expressing the *T. b. brucei* APPBP1 (clone 1), *T. b. evansi* type IVM-t1 sequence (clone 3) and a wild-type gene add-back cell line (clone P+G2) were used to infect three mice from an *in vitro* culture. Of the nine mice which were inoculated, three mice that were infected with parasites expressing the *T. b. brucei* sequence and one mouse that was infected with parasites expressing the *T. b. evansi* type IVM-t1 sequence established a parasitaemia. In contrast, none of the mice infected with the wild-type gene add-back cell line established parasitaemia.

Although this experiment will be repeated in future to analyse *in vivo* infections in triplicate, the initial results have cautiously been analysed. The single mouse that became infected with parasites expressing the *T. b. evansi* type IVM-t1 replacement sequence reached peak parasitaemia two days after those expressing the *T. b. brucei* wild-type sequence (Fig. 4.10a). To assist comparison of the infections, the peak parasitaemia for each line was normalised to ‘days pre/post peak parasitaemia’ (DPP) (Fig. 4.10b).

Using DPP, the parasites expressing the *T. b. evansi* type IVM-t1 sequence delayed generation of stumpy forms, exemplified by a slower arrest in 1K1N (Fig. 4.10c) and delayed expression of PAD1, in comparison to the cells expressing the *T. b. brucei* sequence (Fig. 4.10d).

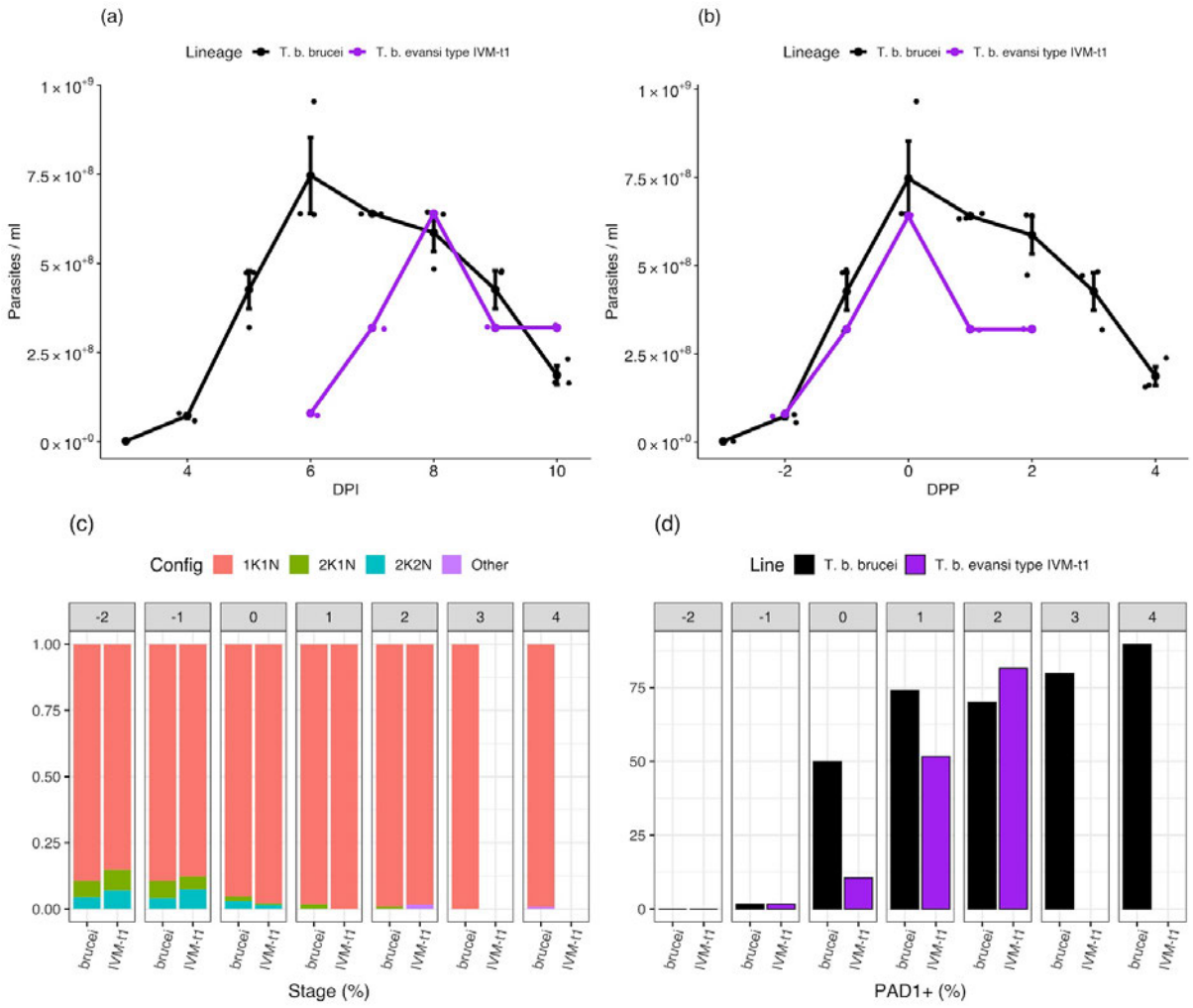


Figure 4.10: Pleomorphic *T. b. brucei* expressing the monomorphic *T. b. evansi* type IVM-t1 APPBP1 sequence display delayed developmental progression. *In vivo* growth of pleomorphic *T. b. brucei* expressing the monomorphic *T. b. evansi* type IVM-t1 APPBP1 sequence or the pleomorphic *T. b. brucei* sequence using **(a)** days post-infection (DPI) and **(b)** *in vivo* growth normalised to days pre/post peak parasitaemia (DPP). **(c)** Cell cycle stage and **(d)** the percentage of cells expressing PAD1 as assessed by immunofluorescence based on DPP. Each cell line was used to infect three mice (although only the cell line expressing the *T. b. brucei* sequence generated infections in all three mice), represented by the dots at each time point. Error bars represent mean standard error.

Tb927.8.1530 - Golgi pH regulator (GPR89)

T. b. equiperdum type BoTat contains five non-synonymous mutations in comparison to *T. b. brucei.* Two mutations are clade-specific (C112Y, which creates a putative phosphorylation site, and D183V) (Fig. 3.10). The sequence is identical in all *T. b. equiperdum* type BoTat isolates. C112Y is part of one of the 9 transmembrane domains and D183V is part of the intracellular loop (Fig. 3.9a).

Replacement of wild-type GPR89 causes a cell proliferation defect

Replacement of both alleles of GPR89 was achieved (Fig. S4.2a-b) and the TY1-epitope tagged protein was detected in three clonal cell lines expressing the *T. b. equiperdum* type BoTat (clones 1, 2 and 6) and one clonal cell line expressing the *T. b. brucei* (clone 9) GPR89 sequence (Fig. S4.2c-d). Multiple bands were detected at a similar size, as has previously been described for this protein (Rojas et al., 2019). Interestingly, the protein in cells expressing the *T. b. equiperdum* type BoTat sequence appears to have a greater molecular weight compared to the pleomorphic *T. b. brucei* protein (Fig. S4.2c). *T. b. equiperdum* type BoTat contains a clade-specific mutation, C112Y, with the potential to be phosphorylated, potentially causing the discrepancy in the band sizes. Quantification of the western blot signal intensity indicated that cells expressing the *T. b. equiperdum* type BoTat sequence had a lower BB2 anti-Ty1 intensity compared to cells expressing the *T. b. brucei* sequence, although this difference was not significant (Fig. S4.2e).

The cells expressing the monomorphic *T. b. equiperdum* type BoTat sequence displayed a significantly impaired growth profile, compared to those expressing the *T. b. brucei* sequence. All three clones expressing the *T. b. equiperdum* type BoTat sequence arrested

their growth (Fig. 4.11), arrested their cell cycle (Fig. 4.12a) and expressed PAD1 upon exposure to BHI (Fig. 4.12b).

We wondered whether the alteration in the protein size of GPR89 detected by western blot could be accompanied by an alternative protein localisation caused by protein post-translational modifications. Therefore, alongside PAD1 and DAPI analysis, we stained cell lines with an anti-Ty1 tag antibody and monitored the expression of GPR89 at the cellular level. GPR89 expression was significantly weaker in permeabilised (Fig. 4.13a) and non-permeabilised cells (Fig. 4.13b) expressing the *T. b. equiperdum* type BoTat GPR89 sequence when grown in BHI. Non-permeabilised cells expressing the *T. b. equiperdum* type BoTat sequence and grown in unsupplemented BHI also expressed a significantly lower level of GPR89. There was no significant difference between GPR89 expression when the cells were permeabilised and grown in unsupplemented HMI-9 (Fig. 4.13a). There was no obvious change in localisation between cells expressing the monomorphic and wild-type sequences (Fig. S4.3-4).

We detected GPR89 expression in cells induced to differentiate to the stumpy form with BHI oligopeptides. The cells display the stumpy form morphology and express PAD1, a stumpy stage-specific marker. These cells also express GPR89, identified via the anti-Ty1 tag BB2 antibody. scRNA-seq data (Briggs et al., 2021) shows that GPR89 transcript level expression does not alter during the slender to stumpy transition (Fig. S4.5), however, Rojas et al. (2019) did not detect GPR89 in stumpy forms via western blot.

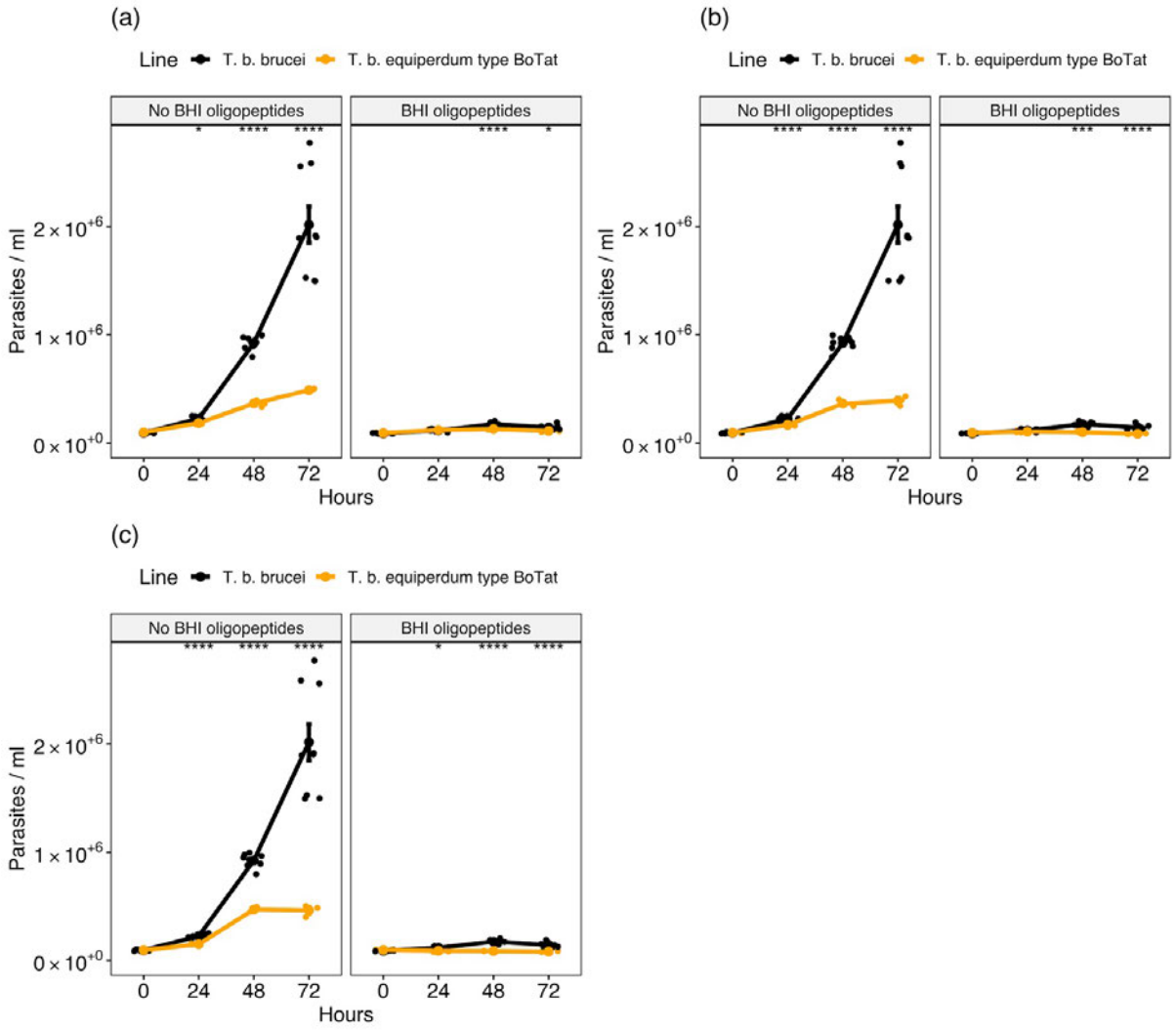


Figure 4.11: Pleomorphic *T. b. brucei* expressing the orthologous *T. b. equiperdum* type BoTat GPR89 sequence display a cell proliferation defect. Growth curves of pleomorphic *T. b. brucei* expressing the *T. b. brucei* GPR89 (Tb927.8.1530) sequence compared to three clones **(a-c)** expressing the monomorphic *T. b. equiperdum* type BoTat sequence. The cells were either grown in HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

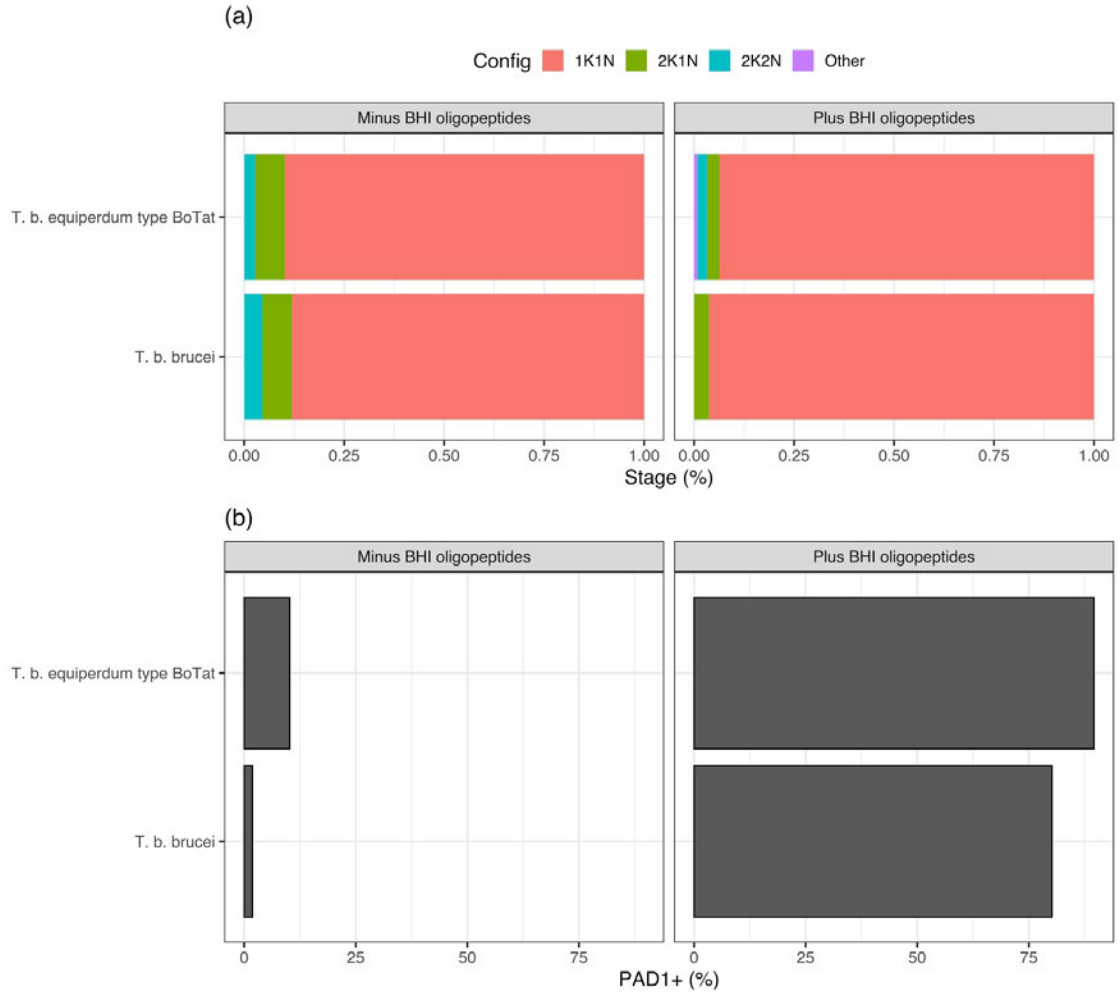


Figure 4.12: Pleomorphic *T. b. brucei* expressing the orthologous *T. b. equiperdum* type BoTat GPR89 sequence maintain developmental competence. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing GPR89 with the monomorphic *T. b. equiperdum* type BoTat sequence or the pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

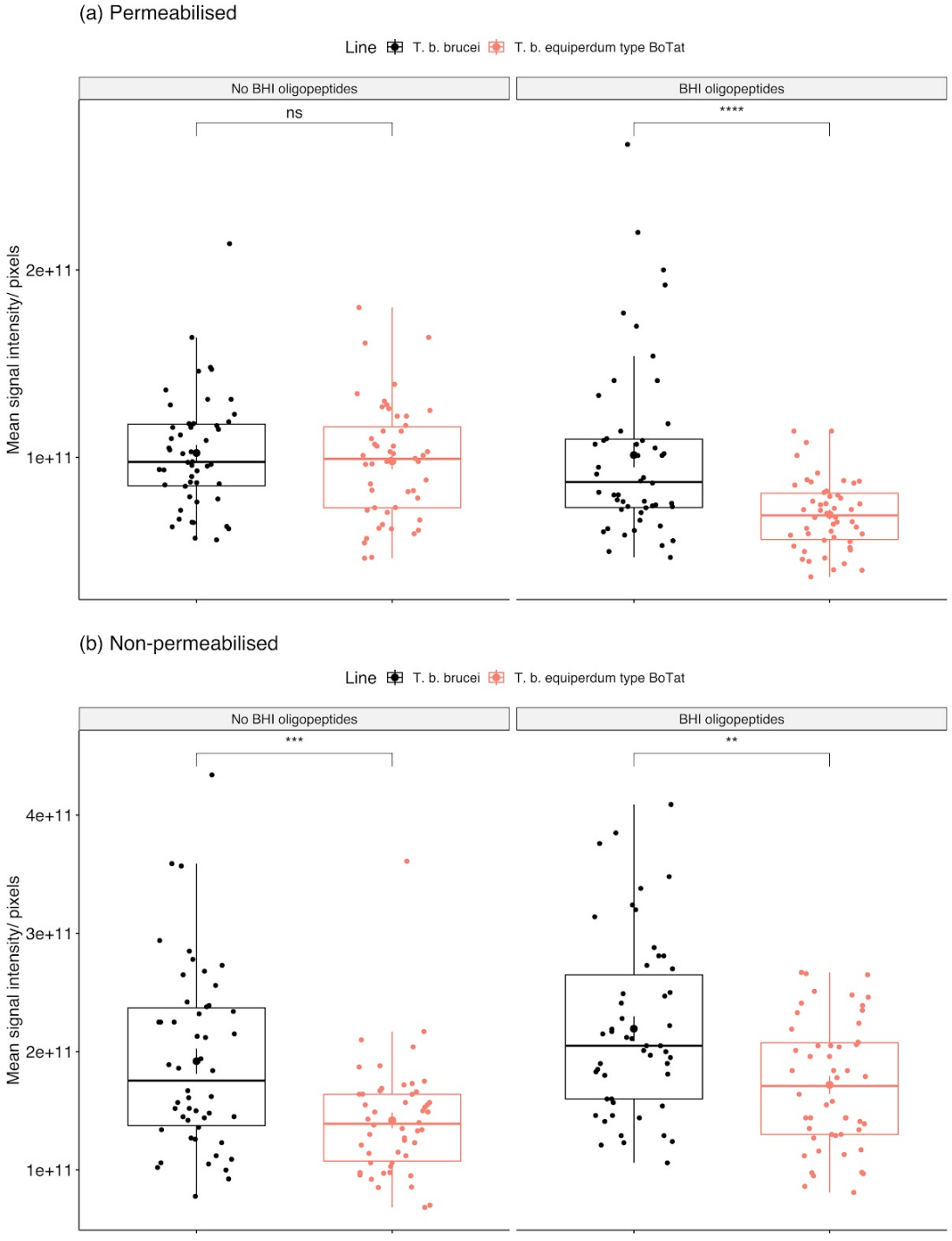


Figure 4.13: GPR89 expression is significantly weaker in permeabilised and non-permeabilised cells expressing the *T. b. equiperdum* type BoTat GPR89 sequence. The intensity of GPR89 anti-Ty1 fluorescence in **(a)** permeabilised and **(b)** non-permeabilised pleomorphic *T. brucei* cells expressing the *T. b. brucei* GPR89 sequence or the monomorphic *T. b. equiperdum* type BoTat sequence. Signal intensity was quantified

as the mean grey value within each cell area. The cell area was measured in pixels and used to normalise the signal intensity per pixel. A Wilcoxon test was used to quantify the significance of the difference between each replacement cell line. See Fig. S4.3-4 for representative images. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

Tb927.11.6600 – Hyp1

Hyp1 contains five non-synonymous mutations in *T. b. evansi* type B and one in *T. b. equiperdum* type OVI compared to TREU927/4 (Fig. 3.12). The sequence is identical between all isolates for both clades.

Replacement of Hyp1 does not cause developmental incompetence

Replacement of Hyp1 with the *T. b. brucei* sequence (clone 4) and *T. b. equiperdum* type OVI (clone 3) sequence created cell lines (Fig. S4.6a) which we confirmed expressed the replacement sequence (Fig. S4.6b-c) and maintained their developmental competence. The cells arrested their growth (Fig. 4.14a), arrested their cell cycle (Fig. 4.15a) and began to express PAD1 (Fig. 4.15b) once exposed to BHI. In contrast, cells expressing the *T. b. evansi* type B (clone 3) sequence appeared to display reduced developmental competence as the cells continued to grow (Fig. 4.14b) and progress through their cell cycle once exposed to BHI (Fig. 4.15a). However, these cells did express PAD1 at a similar level to those expressing the *T. b. brucei* gene (Fig. 4.15b).

To confirm these findings, the developmental competence of clone 3, alongside two additional clones (clones 1 and 4) expressing the *T. b. evansi* type B Hyp1, was tested. All three of the clones arrested their growth, displaying robust developmental competence (Fig. 4.16). Consequently, the *T. b. evansi* type B Hyp1 sequence was not explored further.

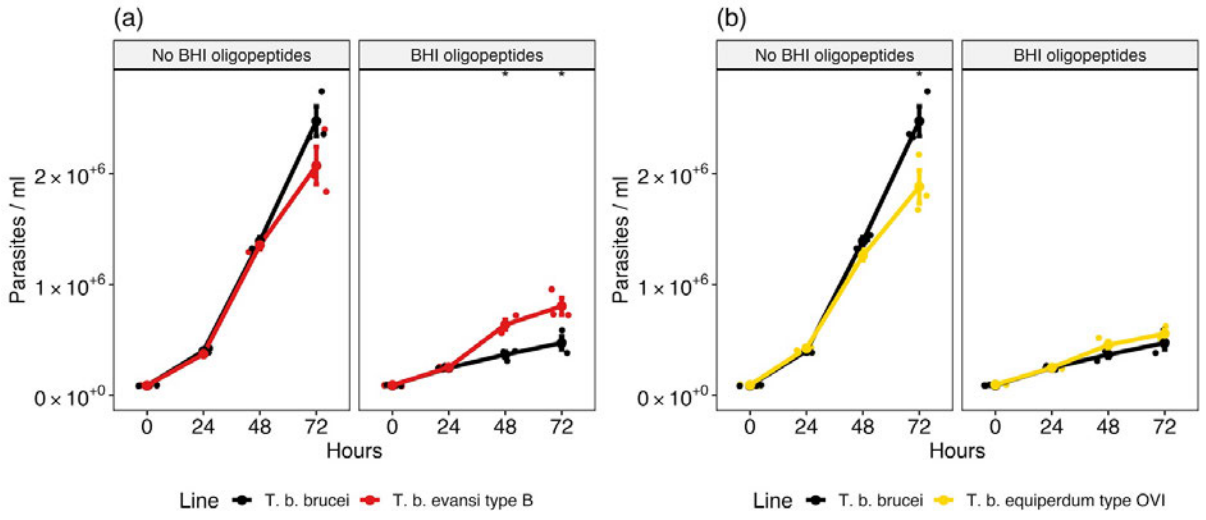


Figure 4.14: Pleomorphic *T. b. brucei* expressing the *T. b. evansi* type B HYP1 sequence display potential insensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* Hyp1 (Tb927.11.3400) sequence compared to cells expressing the monomorphic **(a)** *T. b. evansi* type B sequence or **(b)** *T. b. equiperdum* type OVI sequence. The cells were either grown in HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

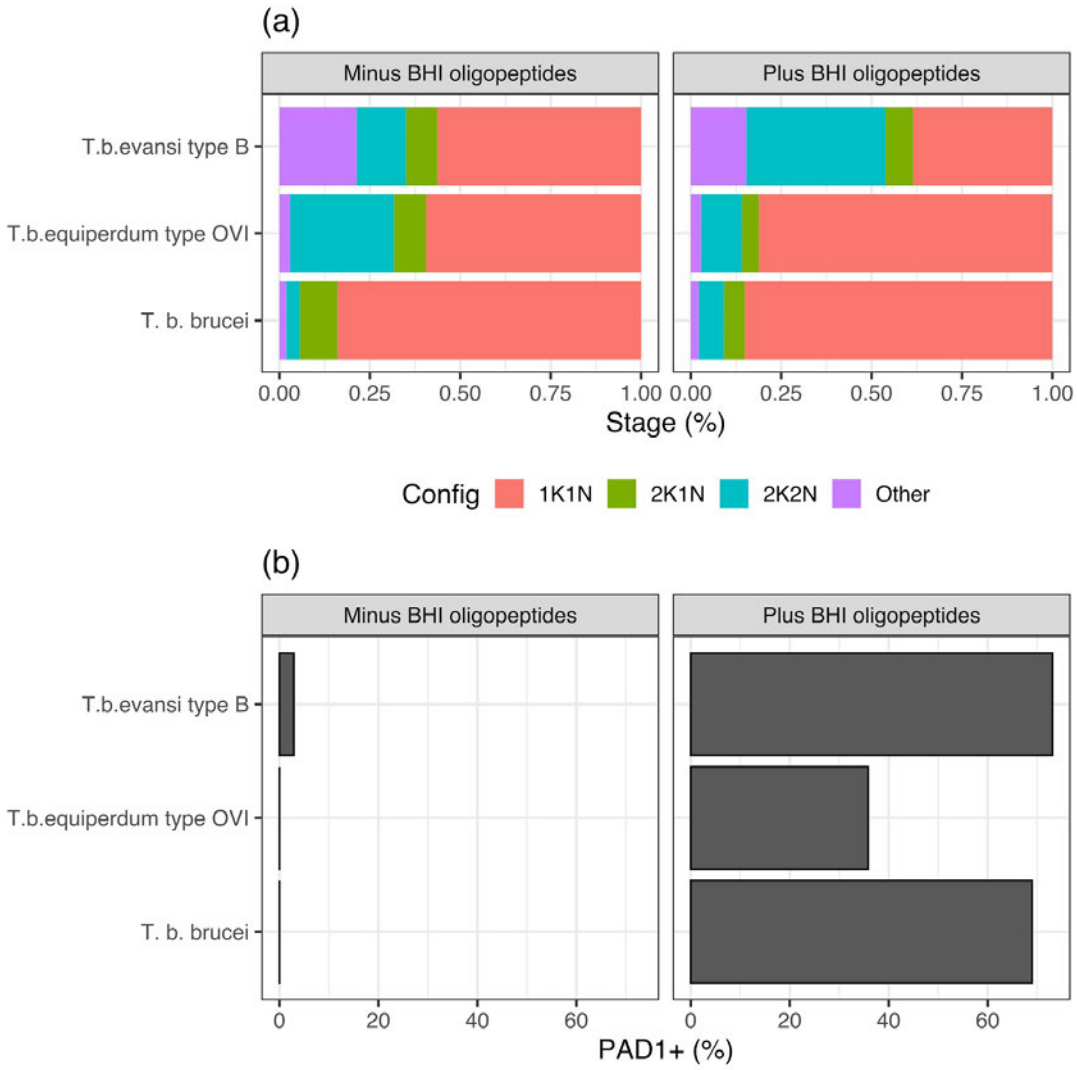


Figure 4.15: Pleomorphic *T. b. brucei* expressing the *T. b. evansi* type B HYP1 sequence display potential insensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the Hyp1 monomorphic *T. b. equiperdum* type OVI sequence, *T. b. evansi* type B sequence and the pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

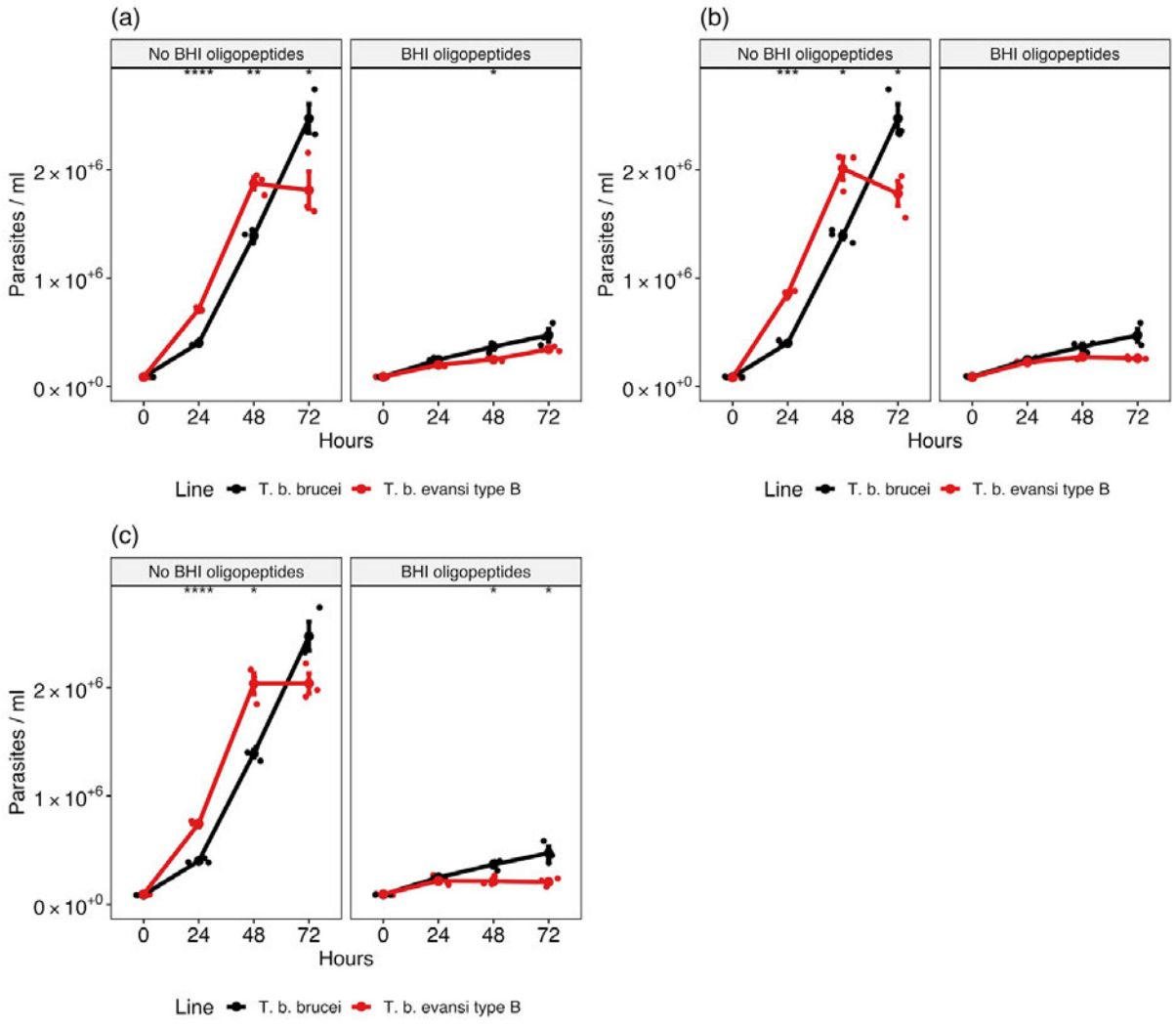


Figure 4.16: Repeat experiments found pleomorphic *T. b. brucei* expressing the *T. b. evansi* type B HYP1 sequence maintain sensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* Hyp1 sequence (Tb927.11.3400) compared to three **(a-c)** clones expressing the monomorphic *T. b. evansi* type B sequence. The cells were either grown in HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

The *T. b. evansi* type B PP1 mutations do not cause developmental incompetence

Replacement of PP1 with the *T. b. brucei* (clone 1) and *T. b. evansi* type B (clones 2 and 3) derived sequence was confirmed (Fig. S4.7a), although we were unable to detect expression of the replacement with the anti-Ty1 tag BB2 antibody (Fig. S4.7b-c). The cell lines expressing the *T. b. evansi* type B sequence did not alter their growth compared to cells expressing the *T. b. brucei* sequence (Fig. 4.17). *T. b. evansi* type B (clones 2) underwent cell cycle arrest (Fig. 4.18a) and expressed PAD1 (Fig. 4.18b) upon exposure to BHI oligopeptides. Consequently, the *T. b. evansi* type B replacement was not validated further.

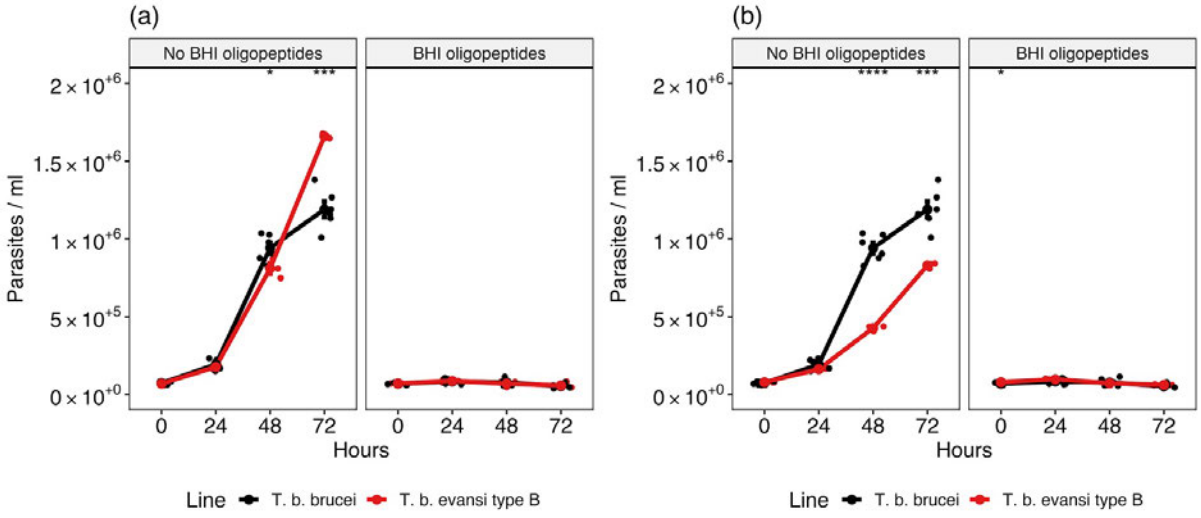


Figure 4.17: Pleomorphic *T. b. brucei* expressing the *T. b. evansi* type B PP1 sequence retain sensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* PP1 (Tb927.4.3650) sequence (clone 1) compared to **(a-b)** clones 2 and 3 expressing the monomorphic *T. b. evansi* type B

sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

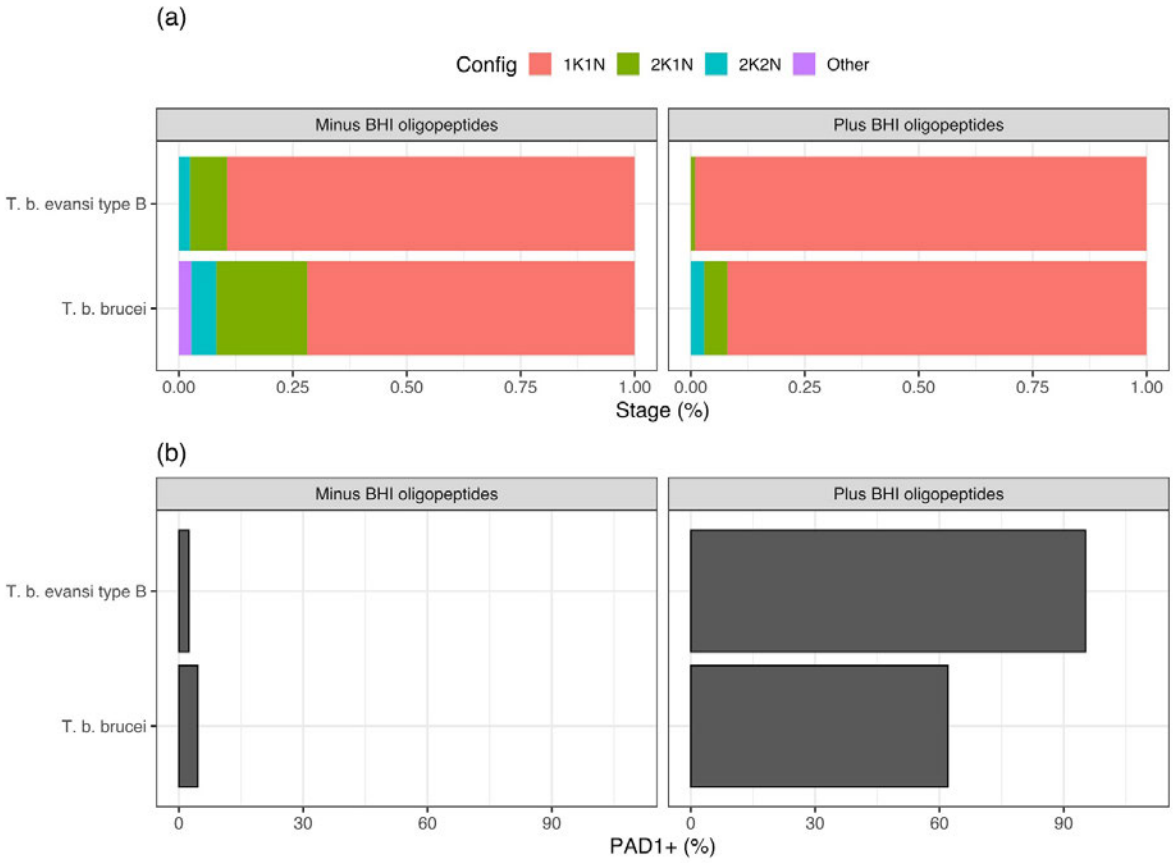


Figure 4.18: Pleomorphic *T. b. brucei* expressing the *T. b. evansi* type B PP1 sequence retain sensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the PP1 (Tb927.4.3650) monomorphic *T. b. evansi* type B sequence (clone 2) or the pleomorphic *T. b. brucei* sequence (clone 1). The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

4.3.2. Novel genes, identified via selection pressure, contain mutations in monomorphic clades which cause insensitivity to BHI oligopeptides

Tb927.5.2580 – hypothetical protein

Tb927.5.2580 has a positive dN/dS ratio in the monomorphic clade *T. b. evansi* type A, in contrast to pleomorphic and other monomorphic clades. The sequence is uniform across all available *T. b. evansi* type A isolates and contains four non-synonymous mutations compared to TREU927/4 (Fig. 3.16a). Only one of these is unique to this clade (A149P). The mutations are contained within the predicted extracellular portion of the protein (Fig. 3.16b).

Replacement of a hypothetical protein (Tb927.5.2580) with the *T. b. evansi* type A sequence causes insensitivity to BHI oligopeptides

Tb927.5.2580 was replaced with either the *T. b. evansi* type A sequence (clones 1 and 2) or the *T. b. brucei* sequence (clone 1 and 4) (Fig. S4.8a). We were unable to detect protein expression via the anti-Ty1 tag BB2 antibody (Fig. S4.8b-c). In contrast to the cells expressing the *T. b. brucei* control sequence, the two clones expressing the *T. b. evansi* type A sequence displayed rapid growth in unsupplemented HMI-9 and continued to grow after exposure to BHI. The effect was more obvious for clone 1 (Fig. 4.19a) however, *T. b. evansi* type A clone 2 grew to a significantly greater population density than when the *T. b. brucei* gene was expressed (Fig. 4.19b). *T. b. evansi* type A clone 1 continued progressing through its cell cycle (Fig. 4.20a) and displayed low levels of PAD1-positive cells (Fig. 4.20b) upon exposure to BHI.

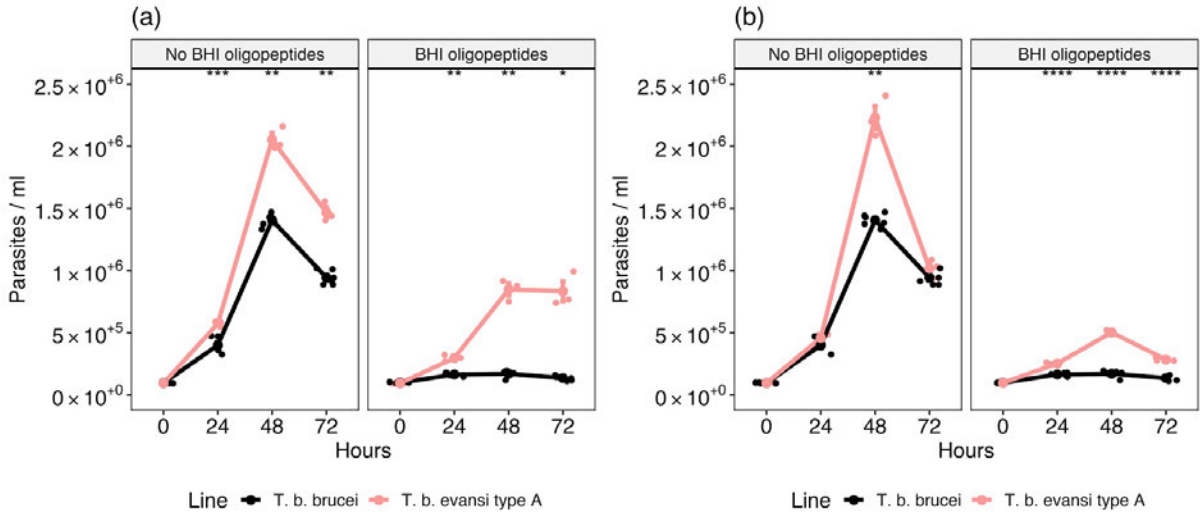


Figure 4.19: Pleomorphic *T. b. brucei* expressing the *T. b. evansi* type A hypothetical protein (Tb927.5.2580) sequence display insensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* sequence of a hypothetical protein (Tb927.5.2580) compared to **(a-b)** two clones expressing the *T. b. evansi* type A sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

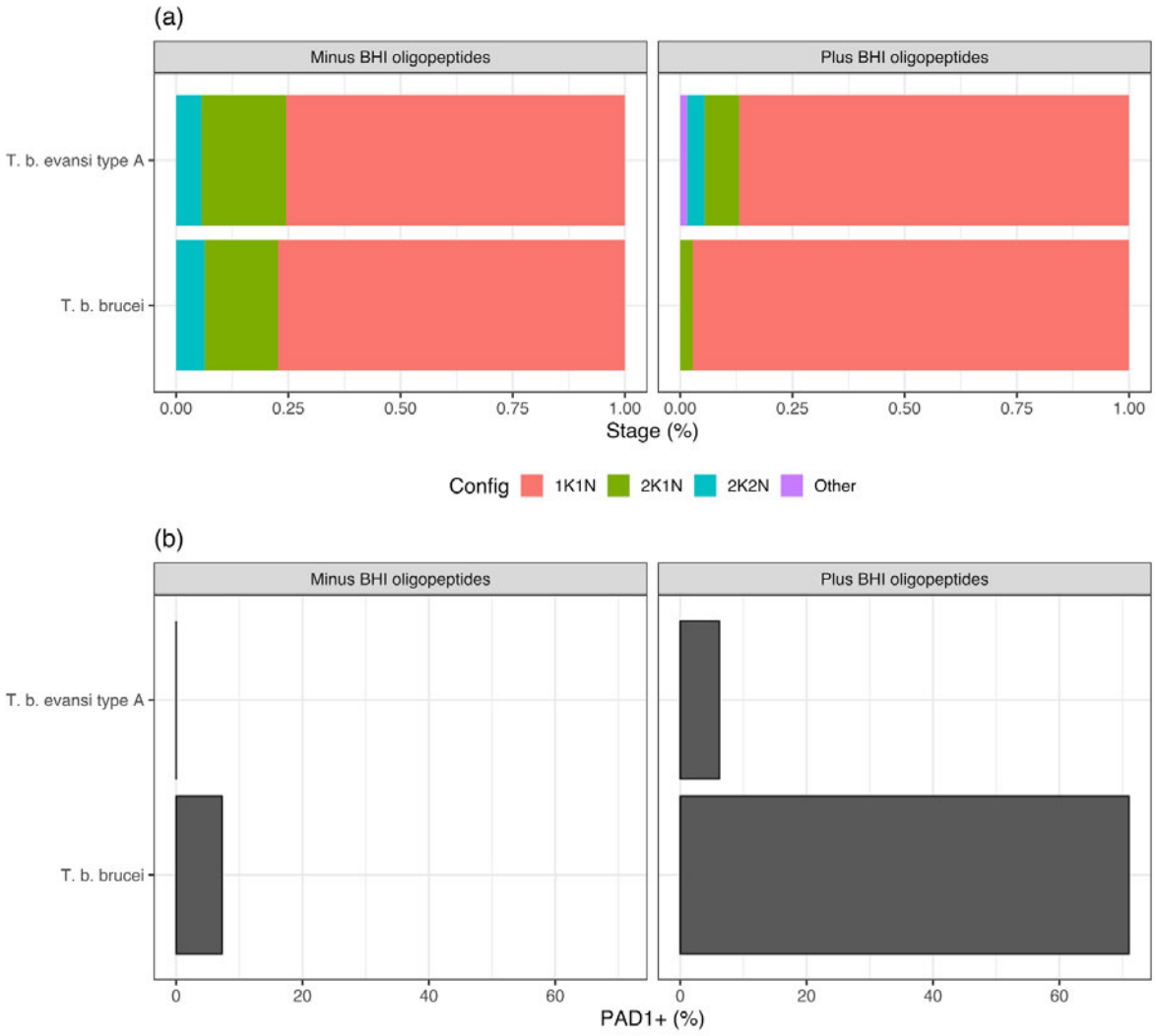


Figure 4.20: Pleomorphic *T. b. brucei* expressing the *T. b. evansi* type A hypothetical protein (Tb927.5.2580) sequence displays insensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing a hypothetical protein (Tb927.5.2580) monomorphic *T. b. evansi* type A sequence or the pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

T. b. evansi type A clone 1 was subjected to heterozygous (clones 1 and 3) and homozygous (clone 6) add-back of the *T. b. brucei* Tb927.5.2580 sequence (Fig. S4.8g-h). Similarly to the original replacement cell lines, we were unable to detect protein expression via the anti-Ty1 tag BB2 antibody (Fig. S4.8i-j). Both of the heterozygous add-back clones displayed partial restoration of developmental competence, with slower growth in unsupplemented HMI-9 and significantly reduced growth in BHI-supplemented media, compared to the original replacement cell line (Fig. 4.21a-b). Upon exposure to BHI, the heterozygous add-back clones arrested their cell cycle and increased the percentage of PAD1-positive cells, compared to the original *T. b. evansi* type A replacement cell line (Fig. 4.22). Homozygous add-back generated an almost complete recovery of the pleomorphic phenotype. The homozygous add-back clone lost its rapid growth in unsupplemented HMI-9 and fully arrested its growth upon exposure to BHI (Fig. 4.21c). Exposure to BHI caused the homozygous add-back clone to arrest its cell cycle (Fig. 4.22a) and increase the percentage of PAD1-positive cells, although this was not at the same level as the *T. b. brucei* replacement cell line (Fig. 4.22b).

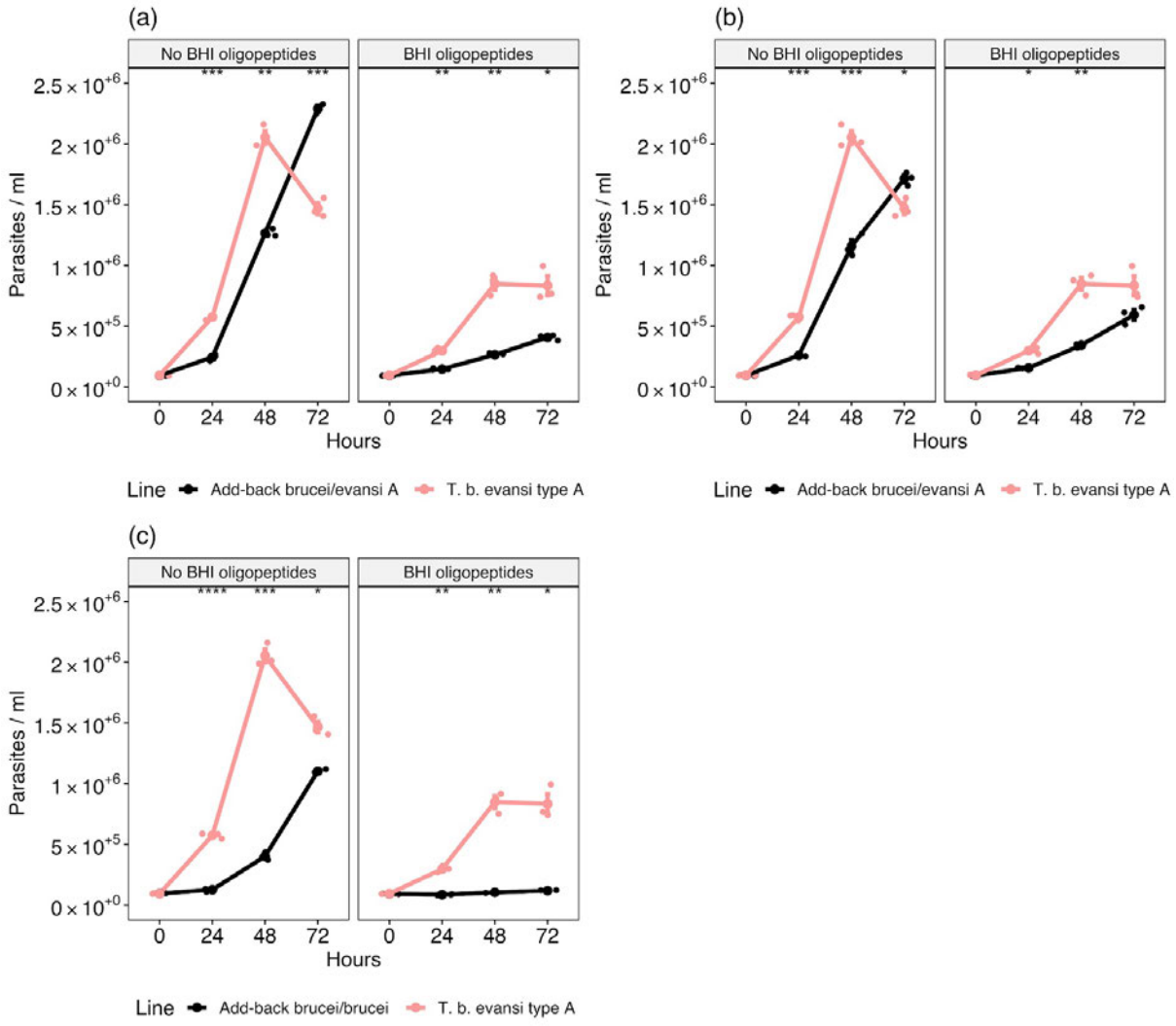


Figure 4.21: Add-back of the *T. b. brucei* hypothetical protein (Tb927.5.2580) sequence to a cell line expressing the *T. b. evansi* type A hypothetical protein (Tb927.5.2580) sequence restores sensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the monomorphic hypothetical protein (Tb927.5.2580) *T. b. evansi* type A sequence compared to the **(a)** heterozygous or **(b)** homozygous add-back of the pleomorphic *T. b. brucei* sequence to the same cell line. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

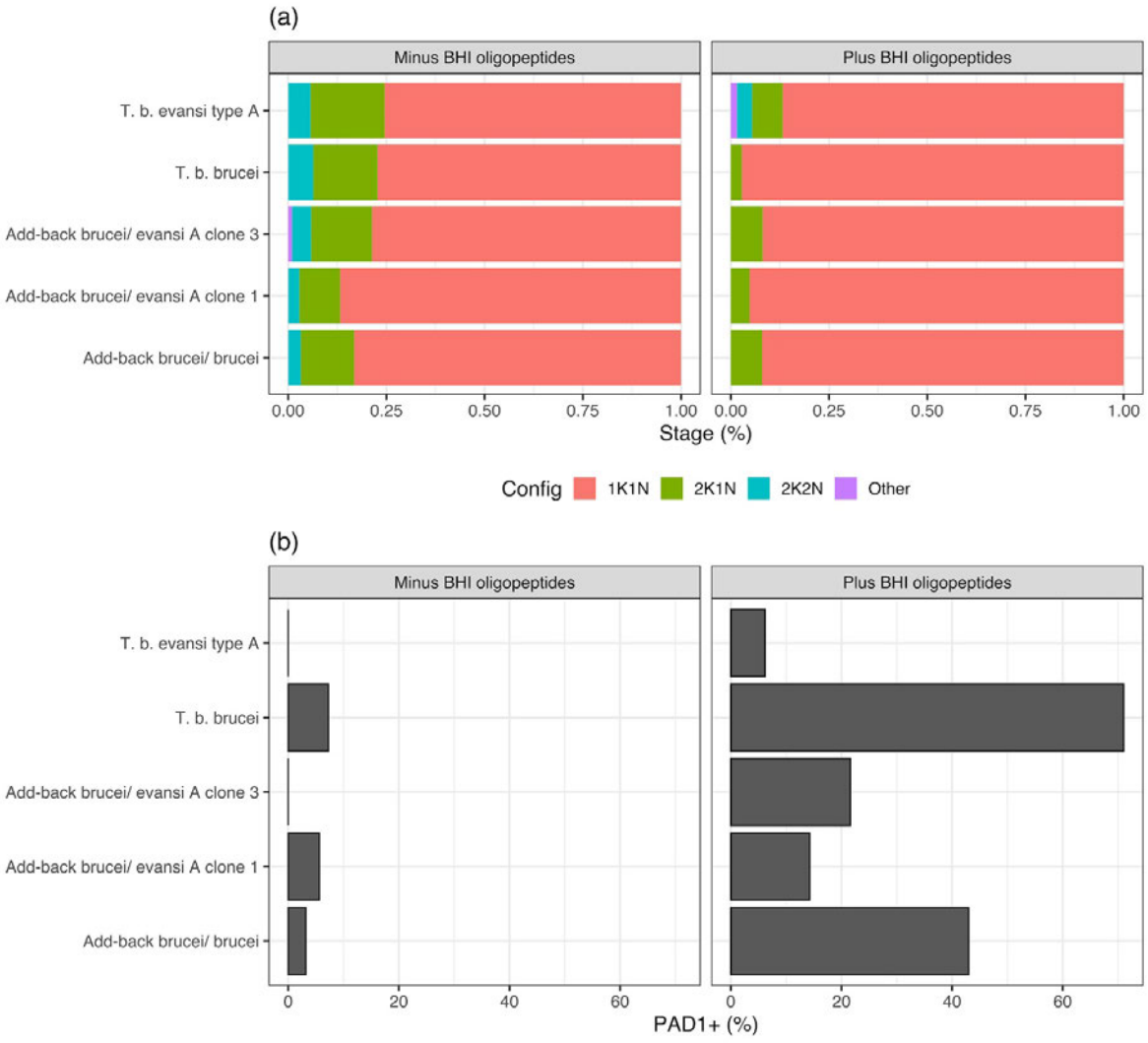


Figure 4.22: Add-back of the *T. b. brucei* hypothetical protein (Tb927.5.2580) sequence to a cell line expressing the *T. b. evansi* type A hypothetical protein (Tb927.5.2580) sequence restores sensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* hypothetical protein (Tb927.5.2580) sequence, the *T. b. evansi* type A sequence, and the *T. b. evansi* type A replacement cell line subjected to heterozygous or homozygous add-back of the pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

The *T. b. evansi* type A A149P mutation alone does not cause developmental incompetence

In the *T. b. evansi* type A clade, Tb927.5.2580 contains one non-synonymous homozygous mutation which is unique, A149P. The alanine residue is highly conserved amongst kinetoplastids at this site, including *B. saltans*. The A149P mutation is shared with the *T. b. brucei* clade Lister427, a commonly used laboratory cell line which has become monomorphic after long-term serial passage (Fig. 4.23). However, Lister427 is heterozygous for this mutation, similarly to other pleomorphic isolates, whilst the *T. b. evansi* type A clade is homozygous. Consequently, a *T. b. brucei* pleomorphic sequence with just the A149P mutation was generated and used to replace the wild-type allele in pleomorphic *T. brucei* (Fig. S4.8d). We were unable to detect protein expression via the anti-Ty1 tag BB2 antibody (Fig. S4.8e-f). Four homozygous A149P clones (clones 1, 2, 4 and 5) were generated and each maintained their developmental competence when exposed to BHI, arresting their growth (Fig. 4.24). Clone 1 arrested its cell cycle (Fig. 4.25a) and expressed PAD1 (Fig. 4.25b) upon exposure to BHI. Consequently, these were not explored further.

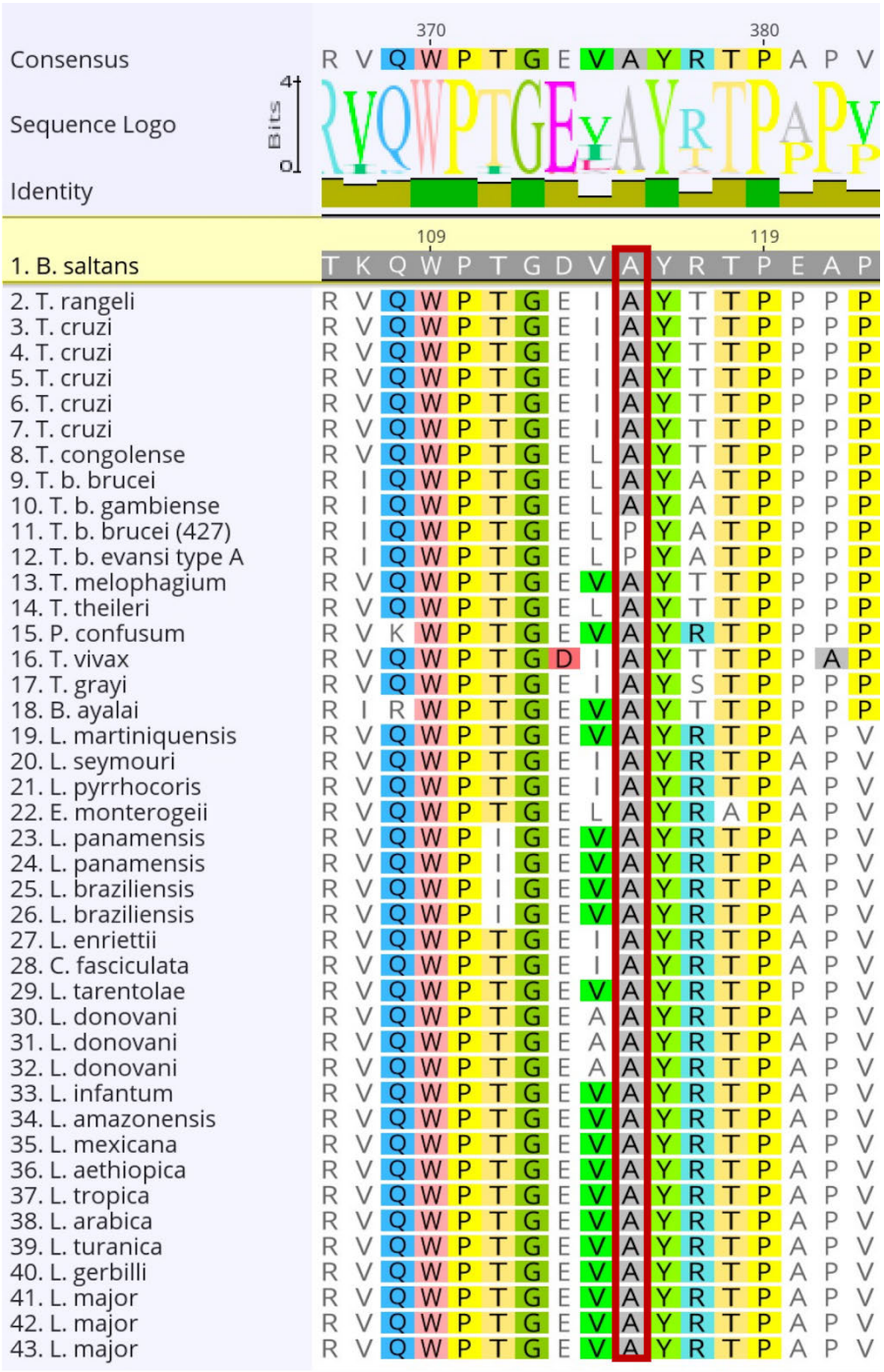


Figure 4.23: A149 is highly conserved in kinetoplastids. Amino acid sequence alignment of kinetoplastid hypothetical protein (Tb927.5.2580) surrounding the A149P mutation (position 115 in consensus sequence, outlined in red).

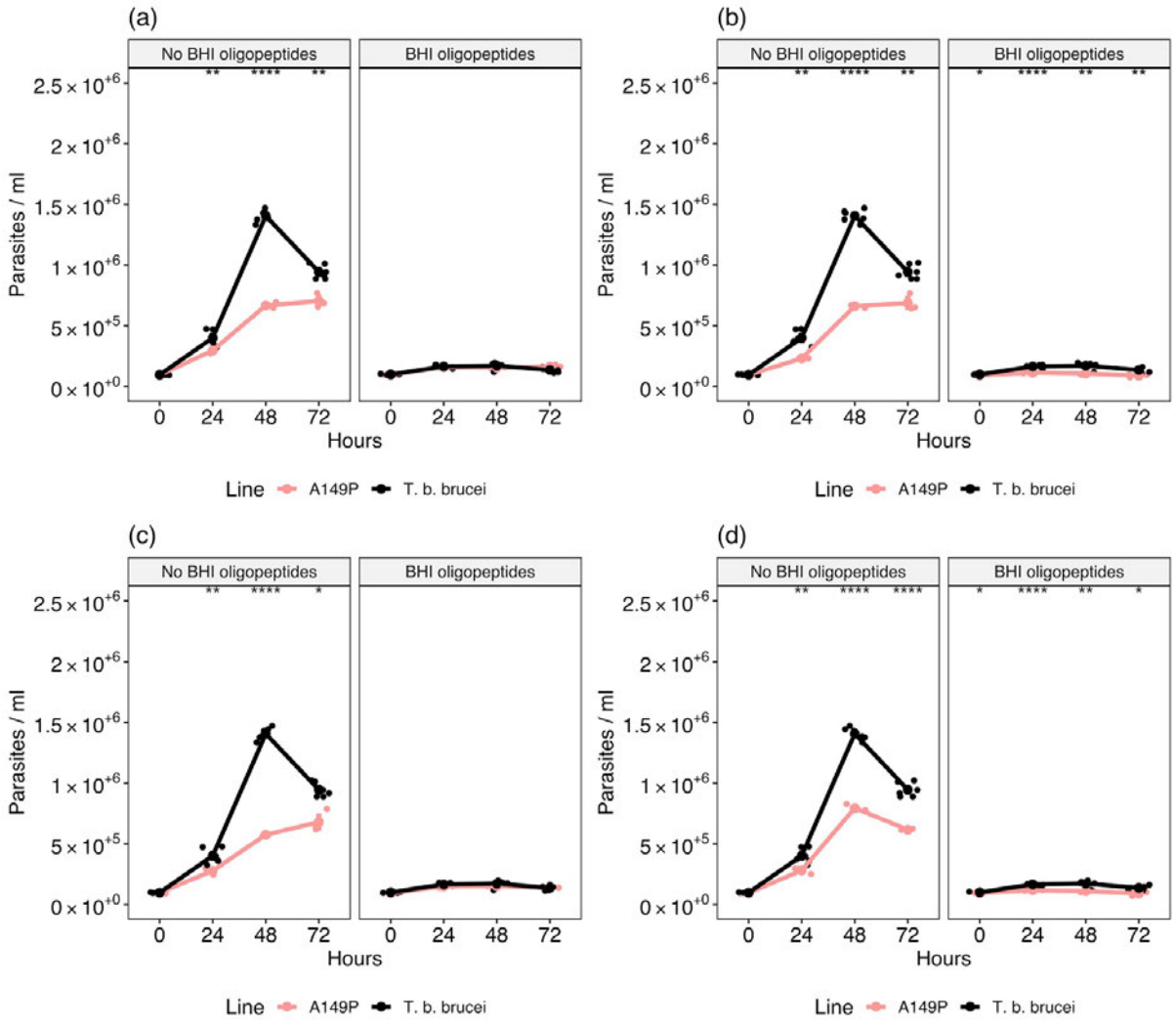


Figure 4.24: The A149P mutation alone does not cause insensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the pleomorphic *T. b. brucei* hypothetical protein (Tb927.5.2580) sequence compared to **(a-d)** four separate clones expressing the *T. b. brucei* sequence with the A149P mutation. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

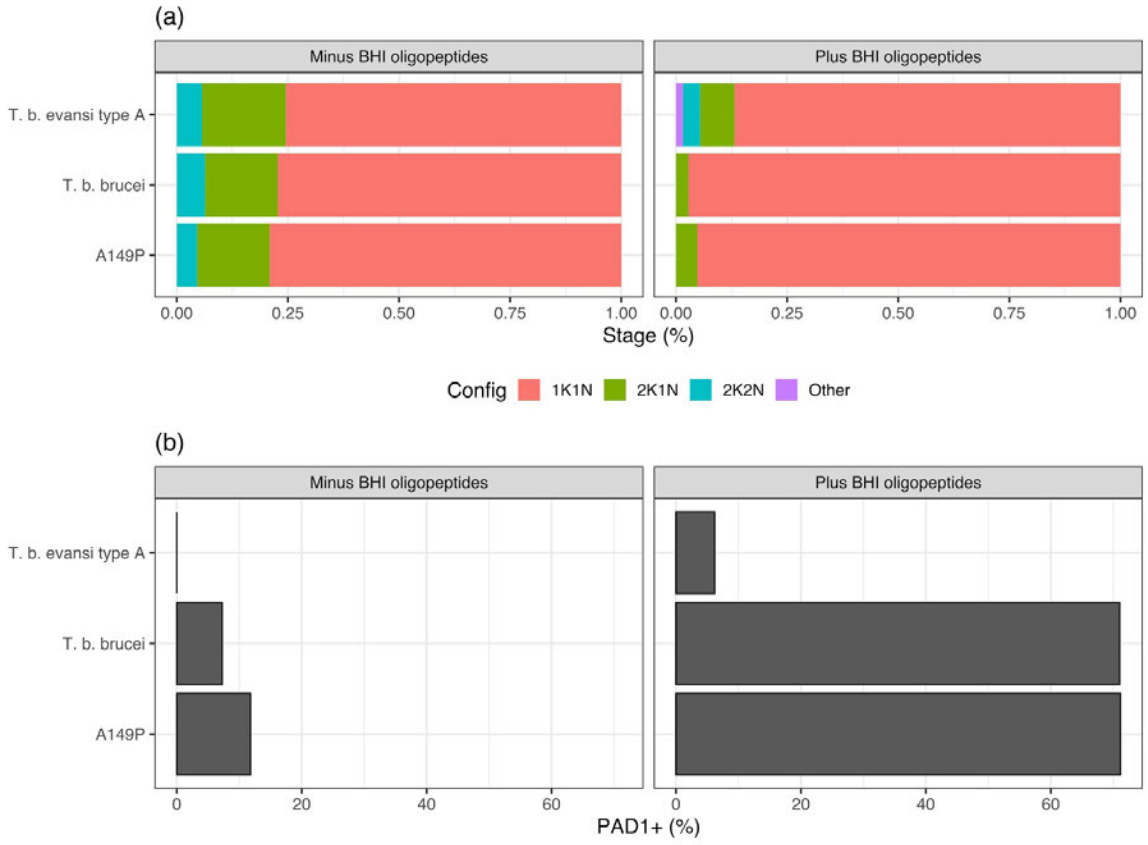


Figure 4.25: The A149P mutation alone does not cause insensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing a hypothetical protein (Tb927.5.2580) monomorphic *T. b. evansi* type A sequence, the pleomorphic *T. b. brucei* sequence or the pleomorphic *T. b. brucei* sequence with the A149P mutation. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

Tb927.11.3400 - Flagellum attachment zone protein 41 (FAZ41)

T. b. equiperdum type OVI and *T. b. equiperdum* type BoTat both have a positive dN/dS ratio for FAZ41, compared to pleomorphic and other monomorphic clades. Although the two monomorphic clades are currently described as a single species or subspecies (*T. b. equiperdum*) they are separated by a pleomorphic isolate *T. b. brucei* TREU 927/4 (Fig. 3.1). They share five identical non-synonymous mutations in FAZ41 compared to TREU927/4 (Fig. 3.18).

Replacement of wild-type FAZ41 does not cause developmental incompetence

Replacement of FAZ41 with *T. brucei brucei* sequence (clone 3), *T. b. equiperdum* type BoTat sequence (clone 2) and *T. b. equiperdum* type OVI sequence (clone 1) (Fig. S4.9a) was detected with anti-Ty1 tag BB2 antibody (Fig. S4.9b-c). The replacement cell lines arrested their growth (Fig. 4.26), arrested their cell cycle (Fig. 4.27a) and expressed PAD1 (Fig. 4.27b) upon exposure to BHI oligopeptides. Thus, the mutations in FAZ41 do not cause developmental incompetence.

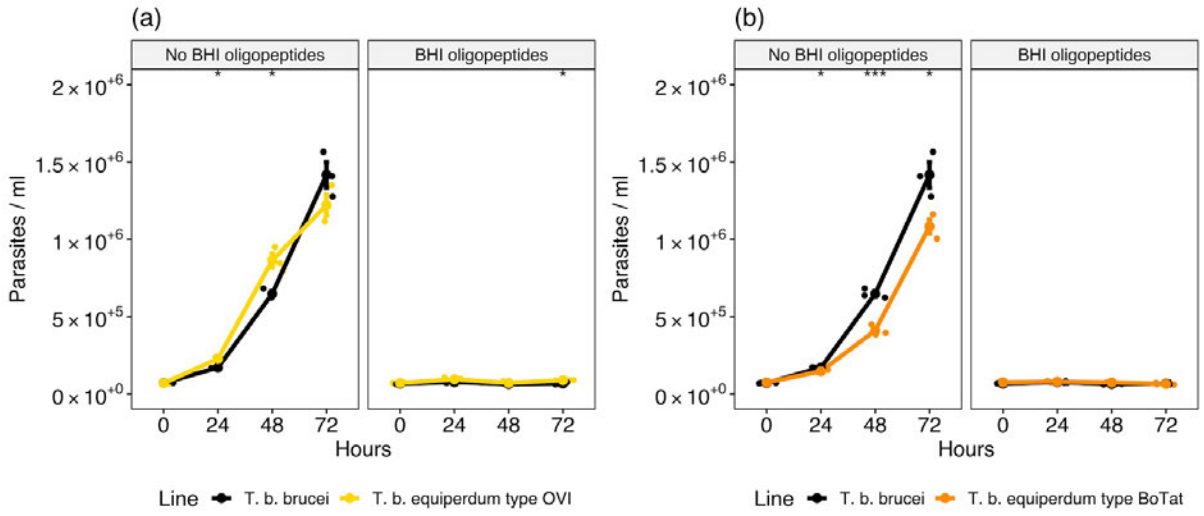


Figure 4.26: Pleomorphic *T. b. brucei* expressing the FAZ41 *T. b. equiperdum* type OVI or *T. b. equiperdum* type BoTat sequence retain sensitivity to BHI oligopeptides. Growth of pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the FAZ41 (Tb927.11.3400) *T. b. brucei* sequence compared to cells expressing the monomorphic **(a)** *T. b. equiperdum* type OVI sequence or the **(b)** *T. b. equiperdum* type BoTat sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI. A t-test was used to compare the population density at each time point. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

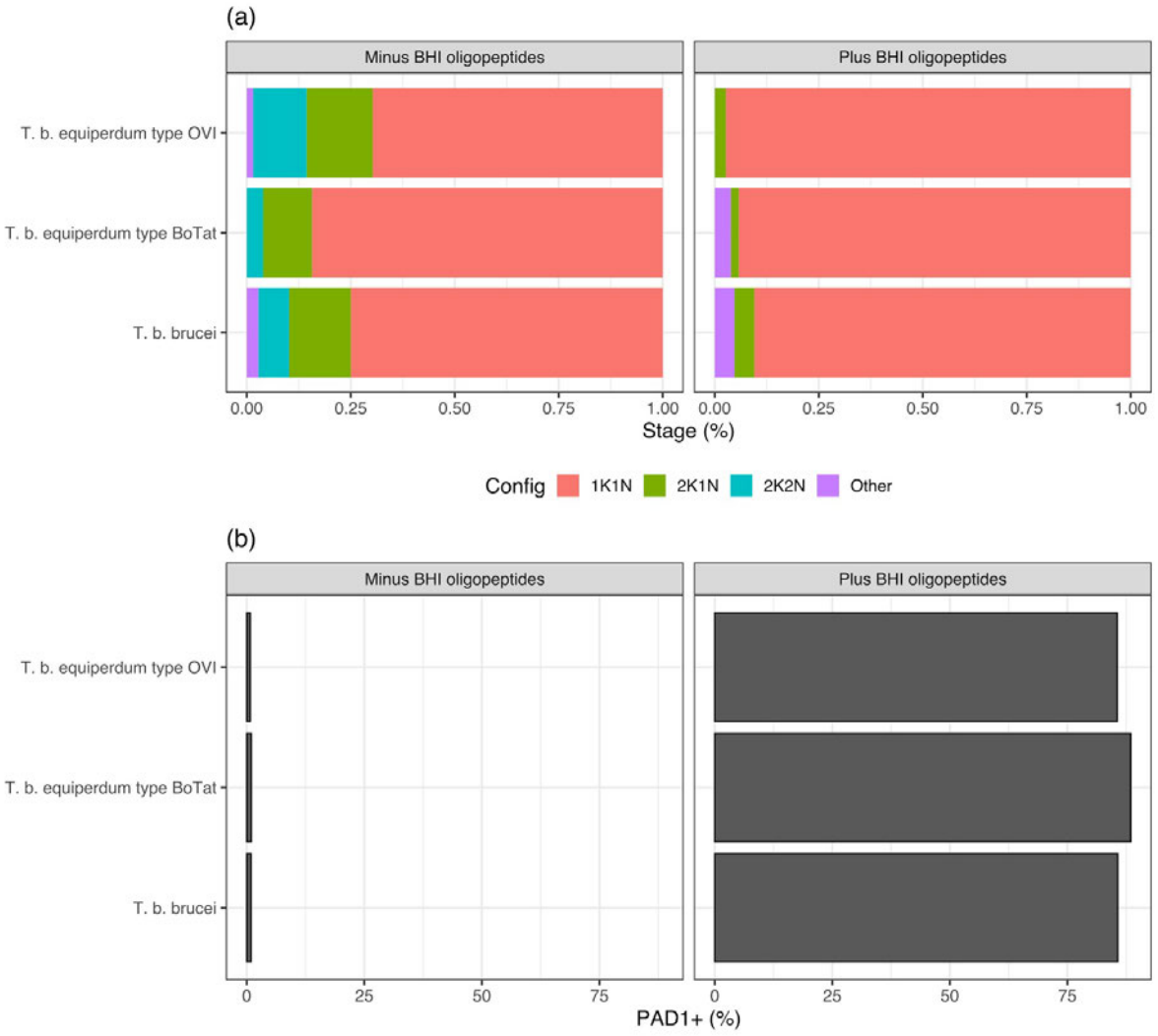


Figure 4.27: Pleomorphic *T. b. brucei* expressing the FAZ41 *T. b. equiperdum* type OVI or *T. b. equiperdum* type BoTat sequence retain sensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence for pleomorphic *T. b. brucei* (J1339 (Rojas et al. 2019)) expressing the FAZ41 (Tb927.11.3400) monomorphic *T. b. equiperdum* type OVI sequence, the *T. b. equiperdum* type BoTat sequence or the pleomorphic *T. b. brucei* sequence. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

Although no developmental phenotype was detected, during the handling of the FAZ41 replacement cell lines, parasites expressing the monomorphic sequence appeared to be less motile than those expressing the pleomorphic *T. b. brucei* sequence. To investigate this, the cells were monitored for their relative motility using FIJI (Schindelin et al., 2012) and TrackMate (Ershov et al., 2021) as detailed in 4.2.3. Quantification showed that cells expressing either the *T. b. equiperdum* type OVI or *T. b. equiperdum* type BoTat sequence showed significantly reduced mean straight-line speed in comparison to those expressing the *T. b. brucei* sequence (2.6, 2.2 and 3.9 microns/s, respectively) (Fig. 4.28). The reduced motility was significant for cells expressing either the sequences derived from *T. b. equiperdum* type OVI or *T. b. equiperdum* type BoTat in three independent replicate experiments (performed on three subsequent days) (Fig. S4.10).

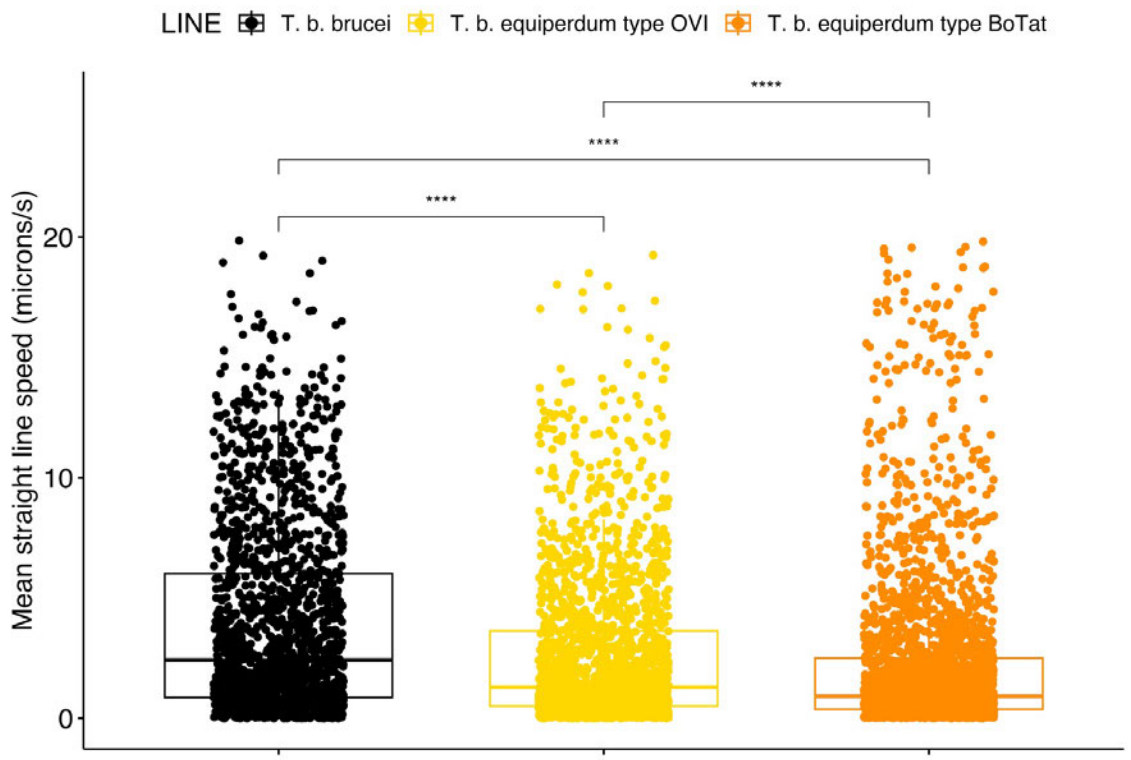


Figure 4.28: Pleomorphic *T. b. brucei* expressing the FAZ41 *T. b. equiperdum* type OVI or *T. b. equiperdum* type BoTat sequence are less motile. Motility assay of pleomorphic *T. b. brucei* expressing the pleomorphic *T. b. brucei* FAZ41 (Tb927.11.3400) sequence, the monomorphic *T. b. equiperdum* type OVI sequence or the *T. b. equiperdum* type BoTat sequence. Motility is displayed as the mean straight speed for each cell track. A Wilcoxon test was performed between each cell line. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. The motility of each cell line was measured on three subsequent days. Error bars represent mean standard error.

4.4. Discussion

The cell line in which the endogenous APPBP1 gene was replaced with the C terminal epitope-tagged version of the orthologous version from *T. b. evansi* type IVM-t1 displays clear reduced developmental competence and a slower growth rate. The add-back of the wild-type gene allowed us to be confident that the reduced competence was caused by the expression of the monomorphic sequence rather than unrelated factors such as the number of rounds of transfection, or *in vitro* passage. These *in vitro* results were partially supported by *in vivo* experiments since the cell line expressing the monomorph derived sequence displayed slowed developmental progression, although further validation is required. Notably, although developmental progression was slowed, the cell line expressing the monomorphic sequence was cleared earlier than the cell line expressing the pleomorphic sequence. We suggest this was caused by culture adaptation and the slow growth rate of the cell line expressing the monomorphic sequence. Future experiments will begin the infection from bloodstocks, ensuring an equivalent infection dose with viable parasites is delivered, and the mice will be cyclophosphamide treated before the infection to counteract any differential sensitivity to the host immune system between the lines. Protein modelling predicted that the mutations in the monomorphic lines should not alter the overall structure of APPBP1 although the predicted creation of a possible phosphorylation site could cause the alteration in the functional ability of APPBP1. The G224S mutation, which was found in the developmentally incompetent cell line, creates a putative phosphorylation site in the catalytic cysteine domain portion of APPBP1. This domain interacts with NEDD8 and in our model, the mutation is at the interaction site between APPBP1 and UBA3. Interestingly, however, the single mutation that distinguishes the monomorphic sequences alone was not enough to cause reduced developmental competence. The remaining four mutations found in *T. b. evansi* type IVM-t1 APPBP1 (A131T, T138A, S151P and S154N) are found in the adenylation domain which helps form a cleft to accommodate protein substrates (Walden et

al., 2003). These four mutations are not predicted to interact with over members of the APPBP1-UBA3-NEDD8 complex. In future, we will assay the possible phosphorylation of the expressed *T. b. evansi* type IVM-t1 APPBP1 mutations via a phos-tag gel (Szöőr and Cayla, 2020). It was not possible to link the developmental phenotype to a specific combination of mutations and therefore future studies will be required to identify the exact suite of mutations required to create the developmental phenotype exhibited by *T. b. evansi* type IVM-t1. Protein modelling could be used further to prioritise mutations which display interactions. We will repeat the *in vivo* experiments to confirm the delay in developmental progression. Overall, however, our results support the role of neddylation in *T. brucei* developmental competence.

Mutations in GPR89 are predicted to alter the structural angle of the molecule's intracellular loop, which has been linked to developmental competence in previous studies (Rojas et al., 2019). Interestingly, the GPR89 mutations in *T. b. equiperdum* type BoTat generated an unexpected phenotype whereby the cell line expressing this molecule has an impaired growth rate but retains its developmental competence. Our results could demonstrate an example of the build-up of deleterious mutations in the monomorphic clade. In future, we will confirm that the reduced growth rate is caused by the expression of the *T. b. equiperdum* type BoTat GPR89 sequence via the add-back of the pleomorphic *T. b. brucei* sequence. Notably, western blot showed that *T. b. equiperdum* type BoTat GPR89 migrated differently compared to the *T. b. brucei* GPR89 protein. Potentially, the clade-specific C112Y mutation could account for this discrepancy in size as it would introduce a putative phosphorylation site that could influence the migration of the protein. Alternatively, the mutations could alter the ubiquitination of the protein, which would target it for destruction.

Expression of the *T. b. equiperdum* type BoTat GPR89 sequence in cells induced to differentiate to stumpy forms was accompanied by a reduction in the protein level of GPR89, quantified by permeabilised and non-permeabilised immunofluorescence intensity at a

single-cell level. Non-permeabilised cells, to detect surface localisation, expressing *T. b. equiperdum* type BoTat GPR89 sequence cells also displayed a significantly lower GPR89 protein level when grown in BHI. Interestingly, there was no significant difference between non-permeabilised cells expressing *T. b. equiperdum* type BoTat when grown in unsupplemented media. Therefore, it seems that the sequence from the monomorphic isolate affects the transport of GPR89 to the cell surface in slender form parasites, rather than the overall production of the protein. GPR89 appears to cover the entire cell surface, with potential enrichment at the flagellar pocket, a site of exocytosis and endocytosis (Field and Carrington, 2009), in slender form (Fig. S4.3-4) and procyclic form (Fig. 3.9b) parasites.

We detected GPR89 expression in cells induced for differentiation to stumpy form *in vitro* using BHI. GPR89 is thought to be absent in stumpy-form parasites, as Rojas et al. (2019) did not detect the protein using a GPR89-specific antibody in stumpy-form enriched parasites via western blot. However, bulk (Naguleswaran et al., 2021) and single-cell (Briggs et al. 2021) transcriptome analysis found that GPR89 is expressed in stumpy forms. In this study, we used a TY-tagged protein, and generated stumpy-form parasites *in vitro*, rather than *in vivo*, potentially causing the discrepancy between our results and Rojas et al., (2019). As GPR89 is essential in slender-form cells (Rojas et al., 2019), the reduced protein level associated with the expression of the *T. b. equiperdum* type BoTat sequence might account for the impaired growth rate observed in this cell line. As for the APPBP1 mutation, we will assess the comparative phosphorylation of each variant of the protein via a phos-tag gel (Szöőr and Cayla, 2020) and we will also block proteasome degradation with MG-132 (Tiengwe et al., 2018) to see if the monomorphic mutation accumulates, linking the mutation to ubiquitination.

Monomorphism is thought to promote uncontrolled proliferation of the population as monomorphic populations are unable to control their population density via the generation of cell cycle-arrested stumpy forms (Desquesnes et al., 2009). However, these observations

are often from infections with *T. b. evansi* type A isolates. Whilst *T. b. evansi* type IVM-t1, *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat can be established and maintained in culture (Baltz et al., 1985; Suganuma et al., 2016), the clades infect equines and are very rarely observed in the bloodstream. Instead, they localise to capillaries of the urogenital tract. Foals have been found with infections, suggesting vertical transmission of monomorphic *T. brucei* is possible (Brun et al., 1998). Comparison of the growth rate of *T. b. evansi* type IVM-t1, *T. b. equiperdum* type BoTat, *T. b. equiperdum* type OVI, or cell lines expressing mutant proteins from one of these clades may not be truly representative unless analysed in an equid host or an *in vitro* mimic of the preferred tissue. Whilst Dourine can be fatal, the infection route and tissue tropism could necessitate a distinct parasite phenotype, compared to *T. b. evansi* type A and *T. b. evansi* type B. Hence selected or observed mutations from genotype analysis may reflect different parameters under selection, these being potentially unrelated to QS. A recent study has shown that *T. brucei* infections have distinct populations in the bloodstream and adipose tissue. Gonadal adipose resident parasites divide at roughly half the rate of those found in the bloodstream and, critically, the parasites have a reduced rate of protein production and a consistently lower proportion of stumpy forms, assessed with a PAD1 UTR reporter cell line, in a population which is often ~6-fold more concentrated than the bloodstream population (Trindade et al., 2022). Therefore, the mutations in GPR89 and APPBP1 which appear to reduce the growth rate of the parasite could be beneficial for sexually transmitted monomorphic forms, which have evolved in a different microenvironment to the other mechanically transmitted bloodstream-dominant monomorphic clades. Their reduced proliferation rate could perform some form of population density control. Still, it would not preclude the ability of the population to reach high density in a tissue where stumpy formation is already reduced (Trindade et al., 2022). This would be especially relevant if vertical transmission is an alternative route of transmission, which would require the parasite to maintain the survival of the host.

The expression of a gene with the mutations in the hypothetical cytochrome c oxidase complex associated protein identified in *T. b. evansi* type A (Tb927.5.2580) generated a phenotype more typically associated with monomorphism. The cell line expressing the monomorphic sequence has a rapid growth rate, along with reduced responsiveness to the oligopeptide signal. The mutations were identified in a gene associated with the mitochondrion and the encoded protein contains a transmembrane domain. Add-back of the *T. b. brucei* sequence confirmed the phenotypes are associated with the dose-dependent expression of the *T. b. evansi* type A mutations, with homozygous add-back fully restoring developmental competence.

Tb927.5.2580 is tightly associated with the cytochrome c oxidase complex (complex IV) (Zíková et al., 2008). The cytochrome c oxidase complex is stage regulated, being expressed in procyclic forms but repressed in bloodstream forms (Mayho et al., 2006). Other mitochondrial proteins have previously been linked with differentiation in *T. brucei*. For example, overexpression of RBP6 induces differentiation to the insect epimastigote and then metacyclic forms (Kolev et al., 2012), which initiates redirection of the flow of electrons from cytochrome to the alternative oxidase pathway. The transition also increases the production of reactive oxygen species (ROS). Overexpression of catalase, a scavenger of ROS, prevented differentiation (Doleželová et al., 2020). Furthermore, ROS induces the activation of AMP-activated kinase alpha, which induces arrest in stumpy form (Saldivia et al., 2016) whilst disruption of the cytochrome c oxidase complex increases the production of ROS (Gnipová et al., 2012). Hence, it appears that the mitochondrion can act as a signalling organelle to release ROS and drive differentiation. Consequently, the mutations found in Tb927.5.2580 could potentially act to decrease the production of ROS, preventing the arrest of the parasites as stumpy forms and producing the rapid growth observed when the mutant protein is expressed. Despite this, the A149P mutation alone is not sufficient to cause the monomorphic phenotype. Importantly, however, the mutations were found in a gene which was not previously associated with the QS pathway by experimental screening, highlighting

the power of the genomic analysis approach for identifying novel genes associated with monomorphism.

Finally, the mutations found in both *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat in FAZ41 highlights the presence of mutations in genes which are not associated with QS but could be associated with the insect stage of the parasite's life cycle. Akinetoplastic parasites display reduced motility, preventing parasite invasion of insect tissues (Dewar et al., 2022). Whilst the reduced motility associated with the parasites expressing the monomorphic FAZ41 sequence is not linked with their ability to generate stumpy forms, the reduction in motility could decrease the ability of the parasites to colonise the insect vector, and therefore, complete its life cycle. Potentially, these mutations accumulated after the generation of monomorphic *T. brucei* as deleterious mutations accrued in genes associated with the redundant insect stage of the life cycle. Alternatively, the positive selection pressure could indicate that it is beneficial to reduce motility in *T. b. equiperdum* type BoTat and *T. b. equiperdum* type OVI, such as a reduction in energy use. A reduction in energy would fit with results that highlight reduced protein and DNA synthesis in monomorphic *T. brucei* which are found in the gonadal adipose tissues (Trindade et al., 2022). Future experiments will add back the *T. b. brucei* sequence to confirm the phenotype association. Ultimately, we would also like to perform *in vivo* infections to test the ability of the replacement cell line to colonise the insect vector. These mutations present another example of contrasting phenotypes produced via replacements with our target genes.

Throughout this chapter, we tested the differentiation capacity of cell lines expressing monomorphic derived sequences with an oligopeptide broth, BHI. *In vitro* differentiation using BHI has been used previously and promotes convincing developmental progression (Rojas et al., 2019). However, by exposing the parasites directly to BHI, we could miss the initial steps in the generation or processing of oligopeptides. BHI was chosen due to the high throughput nature of the experiments. With unlimited time and money, it would have been

ideal to use another *in vitro* method, basement membrane matrix enriched gel (BME), which could be more representative of the entire QS process but necessitates the use of small culture volumes making assays more difficult (Rojas et al., 2021). Nonetheless, our results have begun to be validated *in vivo*, providing the ultimate test of differentiation capacity. Replacement of two of the target genes, Hyp1 and PP1, did not cause a phenotype of interest. PP1 (Tb927.4.3650) is part of an eight-gene family, with four PP1 genes found in a tandem array in *T. brucei*. Although PP1 Tb927.4.3650 is separated from the other three members by the insertion of two unrelated genes (Jackson, 2007a), the mutations we introduced could have been compensated for by other copies of the PP1 gene family.

A recent study identified the absence, or putative frameshift, in a procyclin-associated gene 3 (PAG3, Tb927.6.460) in monomorphic cell lines. The gene is upregulated in stumpy forms and knock-down or knock-out of PAG3 creates a monomorphic phenotype with reduced stumpy formation causing an increase in pathogenicity (Wen et al., 2022). The absence, or frameshift, of PAG3 in naturally occurring monomorphic *T. brucei* could represent another example of the deterioration of genes associated with developmental competence. Whilst their findings support our suggestion that monomorphic *T. brucei* displays an accumulation of deleterious mutations, we disagree with their conclusion that “Both situations in PAG3 (deleted or frameshifted) were found in both *T. equiperdum* and *T. evansi*, suggesting that each *T. equiperdum* strain could be the ancestral source of their corresponding *T. evansi* strains”. Instead, our findings, based on whole genome sequences, corroborate previous studies (Cuypers et al., 2017) which highlight that *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat emerged from distinct ends of the *T. brucei* phylogenetic tree, and are independent in origin. Regardless, the deterioration in the same gene in multiple clades of monomorphic *T. brucei* is intriguing, potentially indicating the key role of the gene in developmental competence.

The limitations of our study mean we cannot discount a contribution of compensatory mutations within the genome of the monomorphic isolates restoring normal developmental, growth or motility of the cells after the mutations we have highlighted occurred, as shown by the loss of kDNA compensated for by a single point mutation in the γ subunit of the nuclearly encoded F_1F_O–ATPase (Dean et al., 2013). Therefore, our findings, and those of Wen et al., (2022), cannot definitively identify the initial causes of monomorphism. We suggest that it is likely that these mutations occurred in organisms which had already developed a 'proto-monomorphic' phenotype, whereby reversible changes in gene expression initiate the process. In the field, the selection of proto-monomorphism could arise after a change in the life history of the parasite, such as the movement of its host to a geographical region absent from tsetse flies. In this scenario, mechanical transmission via a biting fly or sexual transmission would be advantageous. There would be no benefit to maintaining the stumpy form and the complex developmental pathway required to initiate the transition, therefore facilitating a reduction in the use of these genes and the reversible generation of a monomorphic phenotype. These genes would then likely suffer an accumulation of deleterious mutations, such as those in APPBP1, FAZ41, GPR89, Tb927.5.2580 and PAG3 (Wen et al., 2022), which could lock the proto-monomorphic parasites into a monomorphic phenotype. The accumulation of these mutations could have occurred at the same time as, or before or after, the loss of kDNA (Dewar et al., 2018).

In Chapter 5, we select monomorphic *T. brucei* from a pleomorphic cell line in the lab to identify the early events in the generation of proto-monomorphism.

4.5. Supplementary material

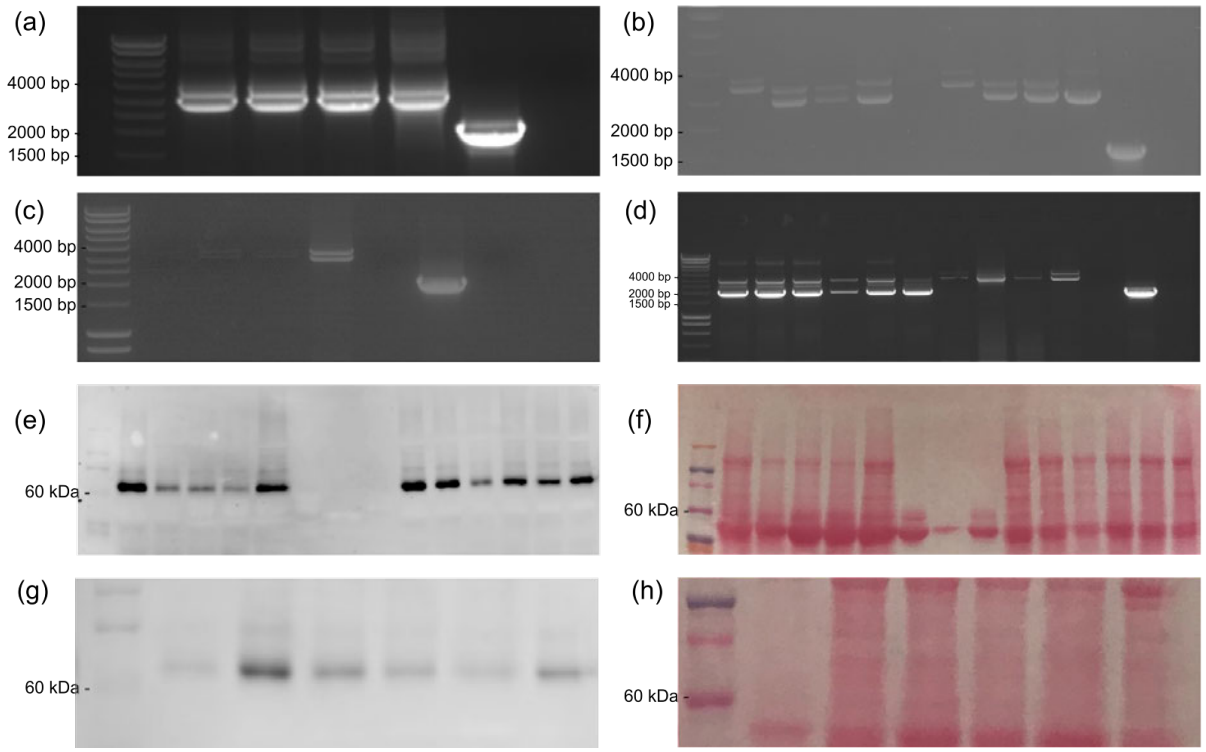


Figure S4.1: Genomic DNA PCR confirmed the **(a)** replacement of wild-type APPBP1 (Tb927.2.4020) with the *T. b. equiperdum* type BoTat sequence (clone 4, lane 1), *T. b. evansi* type A sequence (clone 1, lane 2), *T. b. evansi* type IVM-t1 sequence (clone 3, lane 3) and *T. b. brucei* sequence (clone 1, lane 4), **(b)** Heterozygous phleomycin add-back (lanes (clones) 1-4), Heterozygous G418 add-back (lanes (clones) 5-8), **(c)** Homozygous phleomycin+G418 add-back (lanes 1-2 (clones 2-3)) and homozygous G418+phleomycin add-back (lanes 3-4 (clones 2-3)) and **(d)** G224S heterozygous replacement (lanes (clones) 1-6) and homozygous replacements (lanes 7-10 (clones 1-4). Genomic DNA was amplified using primers that flank the repair primers' homology arms. Wild-type DNA (penultimate lane) and a negative sample (final lane) were used in each PCR. Expected sizes (bp): Wild type Tb927.2.4020 = 1,992, pPOT-Hygromycin-Tb927.2.4020 = 3,425, pPOT-Blasticidin-Tb927.2.4020 = 2,996, pPOT-Phleomycin-Tb927.2.4020 = 2,221, pPOT-Neomycin-Tb927.2.4020 = 3,194. Western blot confirmed the expression of the replacement construct **(e-f)** *T. b. brucei* (lane 1), *T. b. evansi* type A (lanes 2-4), *T. b. evansi*

type IVM-t1 (lanes 5-7), *T. b. equiperdum* type BoTat (lanes 8-12) heterozygous add-back (lanes 13-14), **(g-h)** homozygous add-back (lanes 1-2) G224S -/+ (lanes 3-4), G224S +/+ (5-6). Tb927.2.4020-Ty1 expected size = 67 (kDa). **(e, g)** BB2 anti-Ty1 and **(f, h)** Ponceau S stain loading control.

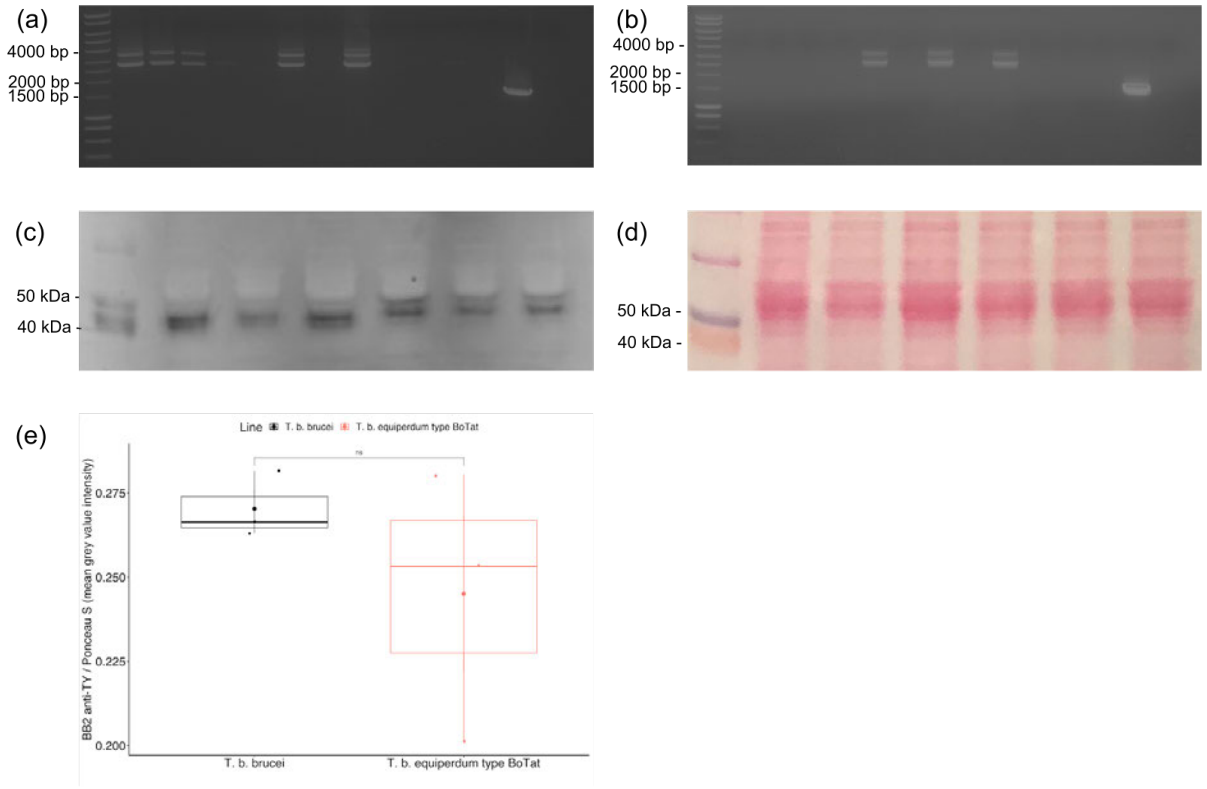


Figure S4.2: Genomic DNA PCR confirmed the replacement of wild-type GPR89 (Tb927.8.1530) with **(a)** the *T. b. equiperdum* type BoTat sequence (lanes (clones) 1-10) and **(b)** *T. b. brucei* (lanes (clones) 1-9) sequence. Genomic DNA was amplified using primers which flank the homology arms of the repair primers. Wild-type DNA (penultimate lane) and a negative sample (last lane) were used in each PCR. Expected sizes (bp): Wild type Tb927.8.1530 = 1,673, pPOT-Hygromycin-Tb927.8.1530 = 3,047, pPOT-Blasticidin-Tb927.8.1530 = 2,618, pPOT-Phleomycin-Tb927.8.1530 = 1,843, pPOT-Neomycin-Tb927.8.1530 = 2,816. Western blot confirmed the expression of the replacement construct with **(c-d)** the *T. b. brucei* sequence (lanes 1-3) and the *T. b. equiperdum* type BoTat sequence(lanes 4-6). **(c)** BB2 anti-Ty1 and **(d)** Ponceau S stain loading control. **(e)** Mean grey value intensity BB2 anti-Ty1 / Ponceau S. Tb927.8.1530-Ty1 expected size = 53 (kDa). Rojas et al. (2019) found that ectopically expressed GPR89 migrates at <40 kDa.

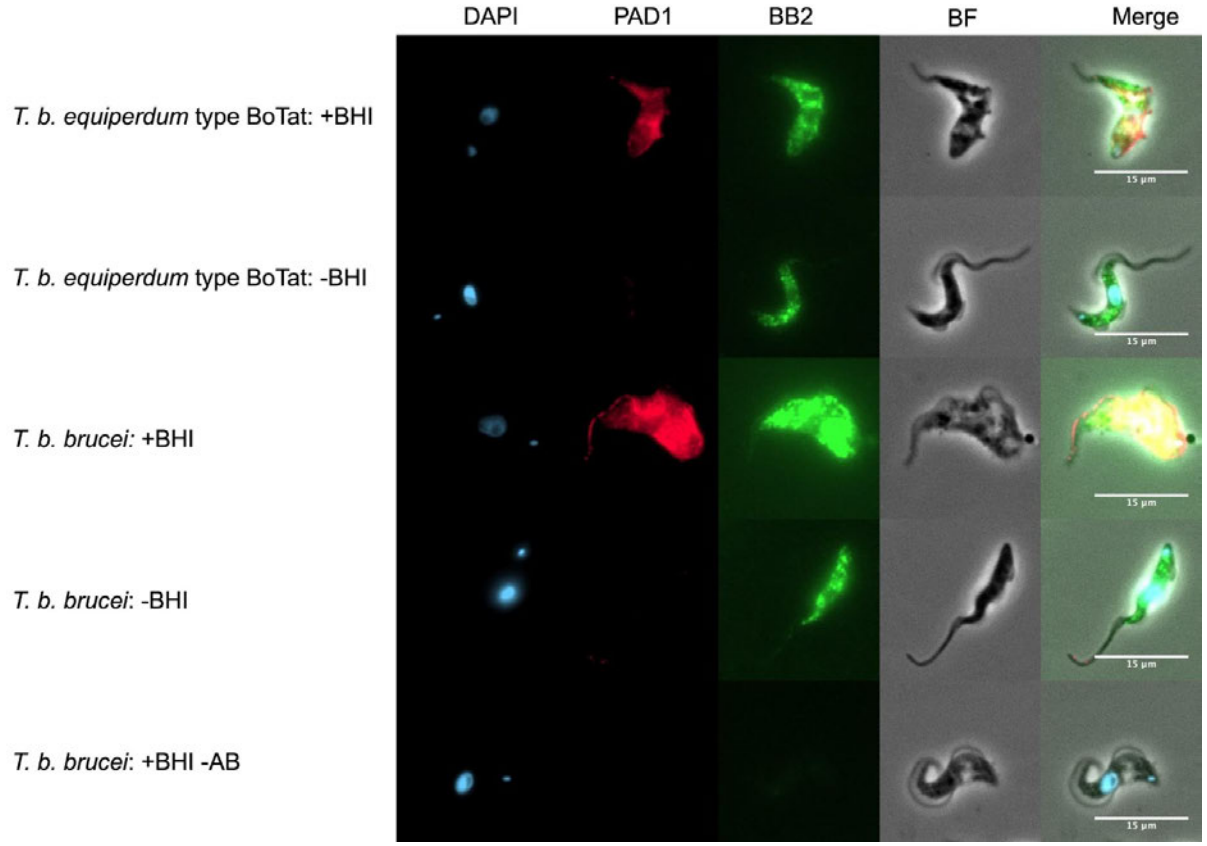


Figure S4.3: Immunofluorescence of permeabilised cell lines subjected to replacement of GPR89 with the *T. b. equiperdum* type BoTat sequence or the *T. b. brucei* control sequence. A control was included using *T. b. brucei* sequence cells that were not incubated with the primary antibody (-AB). The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h. The cells were permeabilised and then DAPI, anti-PAD1, BB2 anti-Ty1 and bright field images were taken and merged.

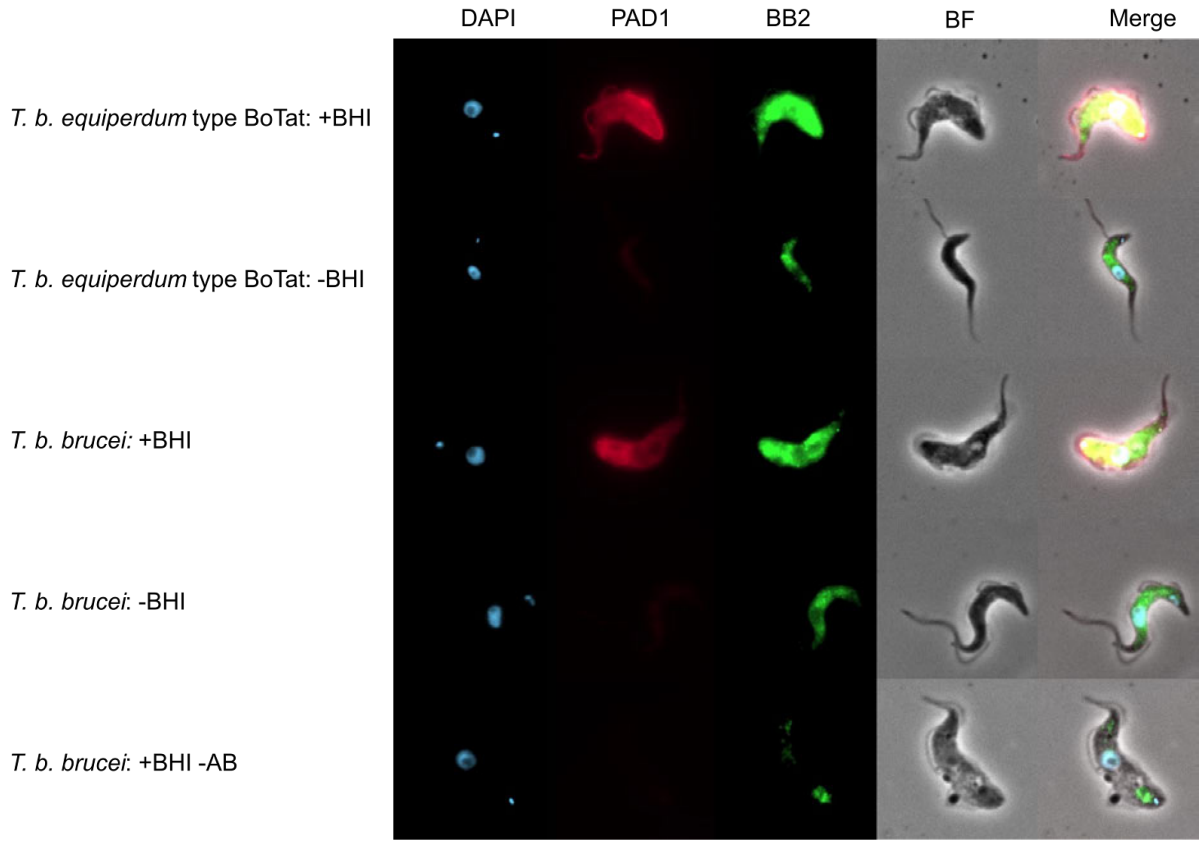


Figure S4.4: Immunofluorescence of non-permeabilised cell lines subjected to replacement of GPR89 with the *T. b. equiperdum* type BoTat sequence or the *T. b. brucei* control sequence. A control was included using *T. b. brucei* sequence cells that were not incubated with the primary antibody (-AB). The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h. The cells were permeabilised and then DAPI, anti-PAD1, BB2 anti-Ty1 and bright field images were taken and merged.

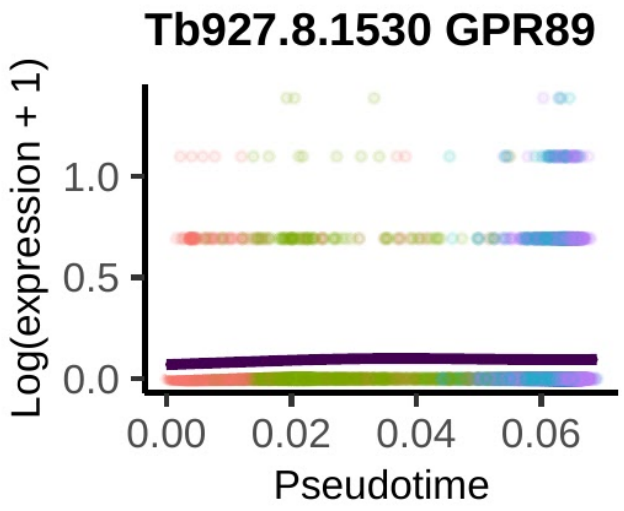


Figure S4.5: GPR89 (Tb927.8.1530) expression (log2(transcript count +1)) across pseudotime during slender to stumpy differentiation. Each point is one cell coloured by cluster. Cluster identity (long slender (LS) A in red, LS B in green, short stumpy (SS) A in blue and SS B in purple). The dark blue line is smoothed average expression across pseudotime. The figure was kindly generated by Dr Emma Briggs using data from Briggs et al. (2021).

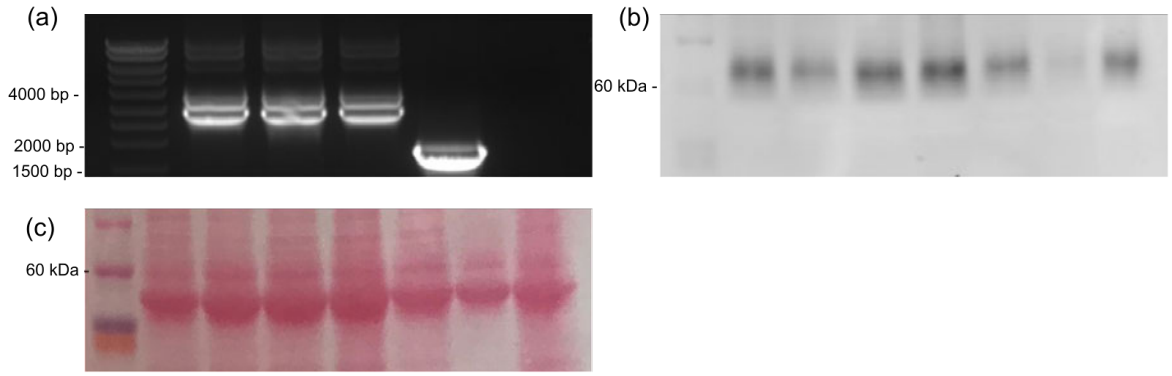


Figure S4.6: Genomic DNA PCR confirmed the replacement of wild-type Hyp1 (Tb927.11.6600) with **(a)** the *T. b. evansi* type B sequence (lane 1, clone 3), *T. b. equiperdum* type OVI (lane 2, clone 3) sequence and the *T. b. brucei* (lane 3, clone 4) sequence. Genomic DNA was amplified using primers which flank the homology arms of the repair primers. Wild-type DNA (penultimate lane) and a negative sample (last lane) were used in each PCR. Expected sizes (bp): Wild type Tb927.11.6600 = 1,802, pPOT-Hygromycin-Tb927.11.6600 = 3,206 and pPOT-Blasticidin-Tb927.11.6600 = 2,777. Western blot confirmed the expression of the replacement construct with **(b-c)** *T. b. brucei* (lane 1), *T. b. equiperdum* type OVI (2-4) and *T. b. evansi* type B (5-7). **(b)**BB2 anti-Ty1 and **(c)** ponceau S stain loading control. Tb927.8.1530-Ty1 expected size = 57 kDa.

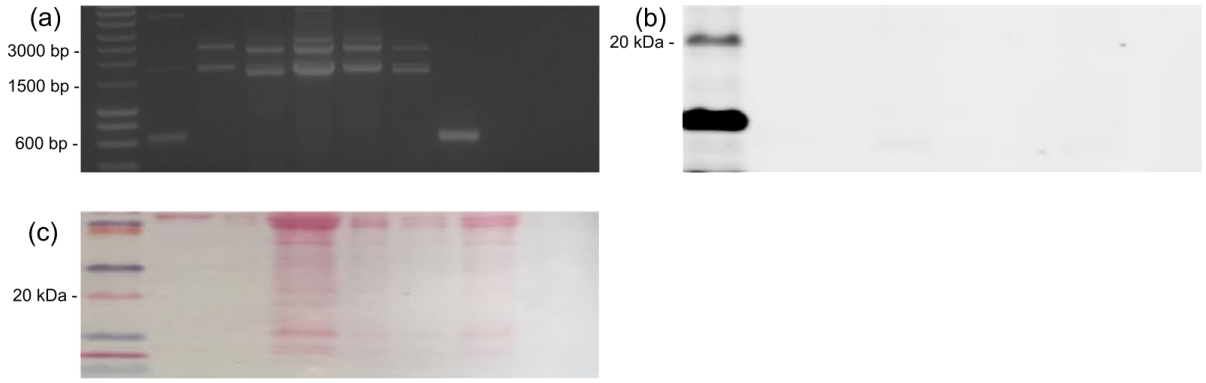


Figure S4.7: Genomic DNA PCR confirmed the replacement of wild-type PP1 (Tb927.4.3650) with **(a)** the *T. b. evansi* type B sequence (lanes (clones) 1-3) and the *T. b. brucei* sequence (lanes 4-6, clones 1-3). Genomic DNA was amplified using primers which flank the homology arms of the repair primers. Wild-type DNA (penultimate lane) and a negative sample (last lane) were used in each PCR. Expected sizes (bp): Wild type Tb927.4.3650 = 648, pPOT-Hygromycin-Tb927.4.3650 = 2,123 and pPOT-Blasticidin-Tb927.4.3650 = 1,694. Western blot confirmed the expression of the replacement construct with the **(b-c)** *T. b. brucei* (lanes 1-3) sequence and *T. b. evansi* type B (lanes 4-6) sequence. **(b)** BB2 anti-Ty1 and **(c)** Ponceau S stain loading control. Tb927.4.3650-Ty1 expected size = 19 kDa.

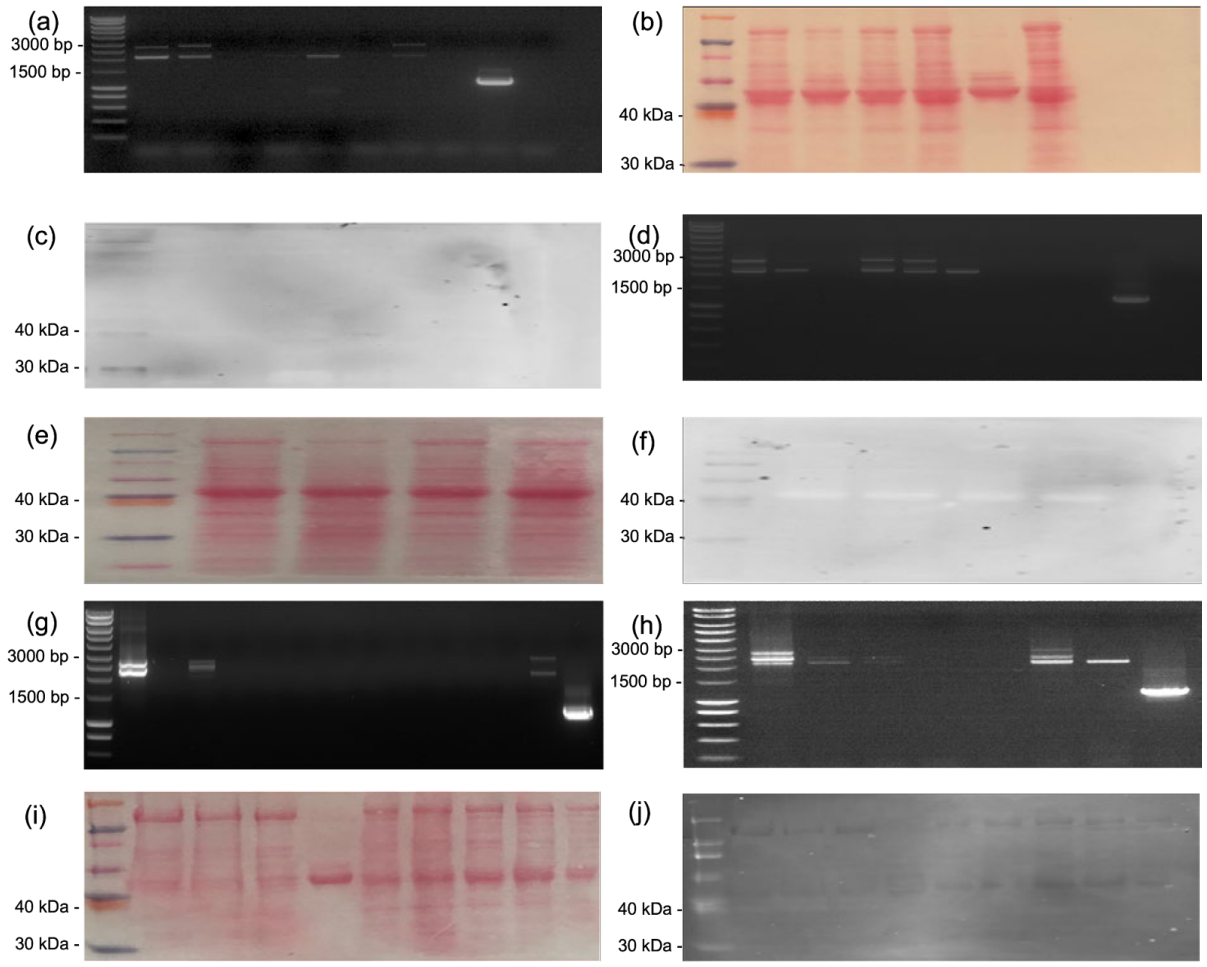


Figure S4.8: Genomic DNA PCR confirmed the replacement of wild-type hypothetical protein (Tb927.5.2580) with **(a)** the *T. b. evansi* type A sequence (lanes (clones) 1-3) and the *T. b. brucei* sequence (lanes 4-8, clones 1-5), **(d)** A149P sequence (clones 1-7), **(g)** heterozygous add-back and **(h)** homozygous add-back. Genomic DNA was amplified using primers which flank the homology arms of the repair primers. Wild-type DNA (penultimate lane) and a negative sample (last lane) were used in each PCR. Expected sizes (bp): Wild type Tb927.5.2580 = 1,134, pPOT-Hygromycin-Tb927.5.2580 = 2,337, pPOT-Blasticidin-Tb927.5.2580 = 1,908, pPOT-Phleomycin-Tb927.5.2580 = 1,893, pPOT-Neomycin-Tb927.5.2580 = 2,106. Western blot confirmed the expression of the replacement construct with **(b-c)** *T. b. brucei* (lanes 1-3) sequence and *T. b. evansi* type B (lanes 4-6) sequence, **(e-f)** A149P replacement (clones (lanes) 1, 2, 4 and 5), **(i-j)** heterozygous add-back (clones 2 and 3, lanes 7-8) and homozygous add-back (clone 6,

lane 9). Tb927.5.2580-Ty1 expected protein size = 34 kDa. **(c, f & j)** BB2 anti-Ty1 and **(b, e & i)** Ponceau S stain loading control.

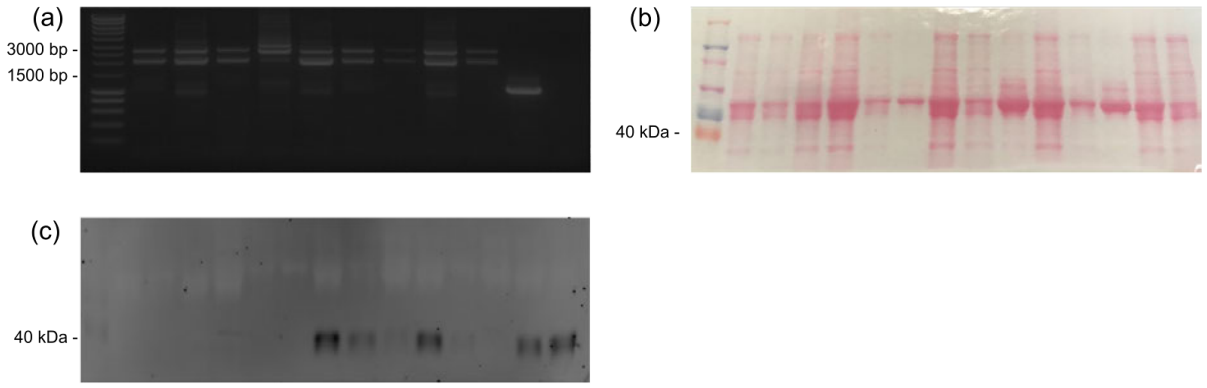


Figure S4.9: Genomic DNA PCR confirmed the replacement of wild-type FAZ41 (Tb927.11.3400) with the **(a)** *T. b. equiperdum* type BoTat (lanes (clones) 1-3) sequence, *T. b. brucei* (lanes 4-6, clones 1-3) sequence and *T. b. equiperdum* type OVI (lanes 7-9, clones 1-3) sequence. Genomic DNA was amplified using primers which flank the homology arms of the repair primers. Wild-type DNA (penultimate lane) and a negative sample (last lane) were used in each PCR. Expected sizes (bp): Wild type Tb927.11.3400 = 1,110, pPOT-Hygromycin-Tb927.11.3400 = 2,552 and pPOT-Blasticidin-Tb927.11.3400 = 2,123. Western blot confirmed the expression of the replacement construct with the **(b-c)** *T. b. equiperdum* type BoTat (lanes 7-8), *T. b. brucei* (lanes 9-11) and *T. b. equiperdum* type OVI (lanes 12-14) sequences. **(b)** BB2 anti-Ty1 and **(c)** Ponceau S stain loading control. Tb927.11.3400-Ty1 expected size = 34 kDa.

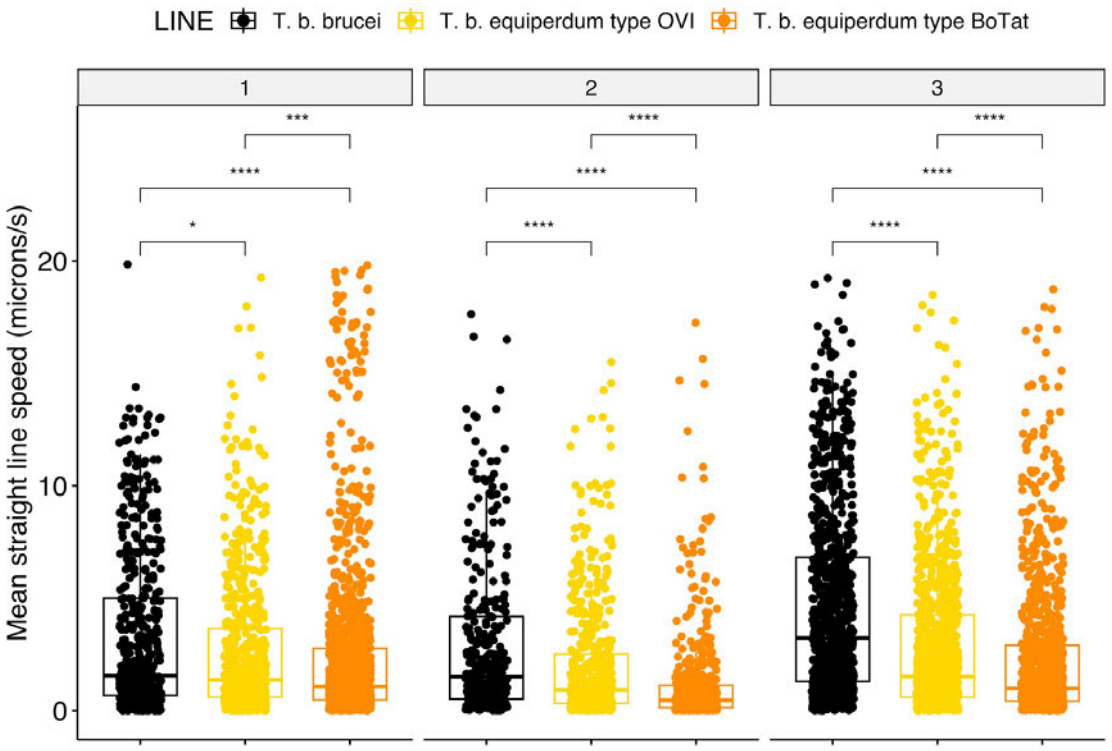


Figure S4.10: Motility assay of pleomorphic *T. b. brucei* expressing the pleomorphic *T. b. brucei* FAZ41 (Tb927.11.3400) sequence, the monomorphic *T. b. equiperdum* type OVI sequence or the monomorphic *T. b. equiperdum* type BoTat sequence. Motility is displayed as the mean straight speed for each cell track on three separate days. A Wilcoxon test was performed between each cell line. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. The motility of each cell line was measured on three subsequent days (1-3). Error bars represent mean standard error.

Recipes

Laemmli 6X Sample Buffer (10ml)

- 60mM Tris-HCL, pH 6.8 1.2Ml of 0.5M stock
- 2% SDS 1.2g
- 10% glycerol 4.7mL
- Bromophenol blue 6mg
- DTT 0.93g

PEME lysis buffer (10ml)

- PIPE 100mM (stock 0.5M pH7) 2ml
- EGTA 2mM (stock 0.1M pH7.6) 200ul
- $MgSO_4 \cdot 7H_2O$ 1mM (stock 1M) 50ul
- EDTA 1mM (stock 0.5M pH8) 20ul
- Igepal 0.5% 50ul
- H_20 up to 10 ml
- Protease inhibitor mini EDTA free (Roche) 1 tablet.

5 Developmental incompetence in selected monomorphic *T. brucei*

5.1. Introduction

T. brucei can lose developmental competence through long-term *in vitro* or *in vivo* passage (Ashcroft, 1960; Cai et al., 2021; Matthews et al., 2004; McWilliam et al., 2019; Turner, 1990). Selected monomorphic *T. brucei* can be forced to undergo many facets of differentiation from slender to stumpy form upon exposure to membrane-permeable cAMP derivative 8-(4-chlorophenylthio)-cAMP (pCPTcAMP) (Breidbach et al., 2002). Whilst pCPTcAMP can initiate analogous differentiation events to QS, these events occur due to the hydrolysed products of 8-pCPTcAMP, such as AMP, rather than cAMP itself (Laxman et al., 2006). The responsiveness of monomorphic *T. brucei* to 8-pCPTcAMP suggests that under normal circumstances mutations or gene regulation prevent reception or transduction through the QS signalling pathway, rather than a complete inability to differentiate in selected monomorphs. Through the creation of an inducible monomorphic cell line, and via serial passage, McWilliam et al., (2019) found that VSG switch rate and monomorphism are not connected. Furthermore, the study identified genes which alter expression during the selection for monomorphism, including the downregulation of ZC3H20 (Tb927.7.2660) (McWilliam et al., 2019). ZC3H20 has subsequently been validated as an important regulator of the QS pathway (Cayla et al., 2020).

In a similar approach, monomorphic *T. brucei* have been selected *in vivo* (Cai et al., 2021). Cai et al., (2021) “identified two DCTs [differentiation characteristic transcripts] that were consistently downregulated in the laboratory-adapted monomorphic trypanosomes, including the telomere DNA binding protein TRF1 (Tb927.11.370) and a zinc-finger domain-containing protein, ZC3H11 (Tb927.5.810). However, whether they are the direct cause of monomorphism still remains to be established”.

However, we have concerns regarding the clustering of the *in vivo* selected RNAseq replicates (Cai et al., 2021), which could have precluded the identification of common differentially expressed transcripts. Moreover, ultimately, these experiments require functional validation. Therefore, we continued the analysis of selected monomorphism to ask if there are common changes in the early stages of the generation of monomorphism which can be manipulated to restore pleomorphism.

5.2. Methods

Each tool used in this chapter, plus optional flags used to run the tool, can be found in Table 5.1. Unless stated otherwise, figures and statistical analysis were created with ggplot2 (Wickham, 2016) and ggrepel (available at https://cran.r-project.org/web/packages/ggrepel/index.html) in R (R Core Team, 2022).

5.2.1. Monomorph selection

Unaltered pleomorphic *T. brucei* EATRO 1125 AnTat 1.1 90:13 (designated UPA for 'Unaltered Pleomorphic AnTat') were grown in HMI-9 at 37°C and 5% CO_2 (Hirumi and Hirumi 1989). In a similar approach to previous studies, the cell line was passaged three times per week for a total of 72 days (McWilliam et al., 2019). The cell density was counted at each passage using a Beckman Z2 Coulter particle counter and size analyser. High cell density was deliberately maintained to select proliferating cells at high parasite population density. Cultures were cryopreserved weekly during the selection period at -80°C in HMI-9 supplemented with 10% glycerol.

A second selection was performed using a clonal *T. brucei* EATRO 1125 AnTat 1.1 90:13 cell line. Clones were generated via serial dilution as described in 4.2.1. The clonal population was then selected in parallel in either HMI-9 or HMI-9 supplemented with BHI oligopeptides. Initially, the BHI-supplemented selection was grown in HMI-9 2.5% BHI. Subsequently, the percentage of BHI was progressively increased until the cells were capable of growing in 15% BHI. The selection was performed over 30 passages and cultures were stored weekly at -80°C in HMI-9 supplemented with 10% glycerol. Prior to the cultures being screened for

their responsiveness to BHI, the cells were washed twice with PBS and grown in HMI-9 for one week to remove any residual BHI from the media.

After confirmation that the selection of monomorphism was more efficient in BHI, six independent clones were generated and HMI-9 supplemented with BHI was used to select for monomorphism as above.

5.2.2. Genomic mutation identification

Genomic DNA was extracted from the ‘start’, ‘intermediate’ and ‘end’ of the UPA selection using DNeasy Blood & Tissue Kit with an RNAse A step (Qiagen), following the manufacturer’s instructions. The DNA was sequenced (HiSeq 4000) and cleaned by BGI (Hong Kong) (~1Gb/ sample). Mutation analysis was performed as in Chapters 2 and 3. Mutations which had arisen during the selection were identified.

5.2.3. Differential expression analysis

RNA was extracted from triplicate cultures of the ‘start’, ‘intermediate’ and ‘end’ from the UPA selection and the ‘start’ and ‘end’ of the clonal selection (clones A1-A7). After the selection had finished, each culture was removed from the storage at -80^{o}C and grown for one week in HMI-9. The cells were grown to a density of between $7.5*10^5$ and $1*10^6$ and RNA was then extracted using the Qiagen RNeasy Mini kit, following the manufacturer's instructions. The RNA was then sent to BGI Hong Kong for quality assessment and sequenced by DNBseq (~4Gb/ sample).

Our analysis largely followed a standard workflow for analysing RNAseq data (Love et al., 2015). Salmon (Patro et al., 2017) was used to index the TREU927/4 reference genome, which was then used to quantify transcript abundance for each sample. Transcript abundance values were imported into R with tximport (Soneson et al., 2015). A database was created using GenomicFeatures (Lawrence et al., 2013). Genes with a count of less than 10 were excluded from the analysis and a VSD transformation was applied to visualise the data as heatmaps and PCA plots. A PCA plot highlighted that the biological replicates of clones A1 and A3 did not cluster (Fig. S5.6). Therefore, clones A2, A4, A5, A6 and A7 (Fig. 5.10) were used for differential expression analysis.

DESeq2 (Love et al., 2014) was used to quantify differentially expressed genes. Differentially expressed genes were identified between the start and end of the selection. For the clonal selection, initially differentially expressed genes were identified by grouping all of the clonal libraries. The contrast function was then used to identify differentially expressed genes for each clonal selection using the same model.

To compare the similarity in differentially expressed genes between the clonal selections, a Venn diagram was created with InteractiVenn to highlight shared differentially expressed genes between the clonal selections (Heberle et al., 2015). A Fisher's Exact Test for count data was used to test for enrichment in the list of differentially expressed genes.

5.2.4. Overexpression of ZC3H20

Since ZC3H20 was the most significantly differentially expressed QS pathway gene in this study (UPA analysis), mirroring previous results (McWilliam et al., 2019), this gene was taken forward to validate its role in monomorphism. The ZC3H20 sequence was amplified from pleomorphic EATRO 1125 genomic DNA using primers designed to add a TY-tag to the

5’ or 3’ terminus (Bastin et al., 1996). These amplicons were ligated into the TOPO plasmid using T4 ligase (NEB), following the manufacturer's instructions. The ligated plasmids were then transformed into competent cells, and grown overnight. A small-scale plasmid preparation was performed on the culture to isolate the plasmid using the GeneJET Plasmid Miniprep Kit (ThermoFisher), following the manufacturer’s instructions. This plasmid was digested using XbaI and BstXI (5’ tag) or HindIII and SpeI (3’ tag) enzymes (NEB) following the manufacturer’s instructions, and the genes were ligated into the doxycycline-inducible pDex-577y plasmid (Kelly et al., 2007). This plasmid integrates into the 177bp repeat sequences of mini chromosomes and encodes for phleomycin resistance and inclusion of a TY-tag into the expressed target protein (Bastin et al., 1996) which allows the quantification of expressed protein expression via western blot. ZC3H20-pDEX-577y was transfected into unaltered pleomorphic *T. brucei* EATRO 1125 AnTat 1.1 90:13 (‘UPA’ cells) passaged 30 times to select for monomorphism (UPA P30), and protein expression was probed using the Ty1 epitope-specific BB2 antibody (Bastin et al., 1996), using the protocol described in 4.2.1.

5.2.5. Validation of differentiation capacity

The ability of a cell line to generate stumpy forms, and therefore developmental competence, was determined *in vitro* using BHI-based oligopeptides. The protocol was identical to that described in section 4.2.2.

5.2.6. *In vivo* infections

In vivo infections were performed by Dr Frank Venter. Three female MF1 mice were infected for each cell line. The mice were not subjected to cyclophosphamide treatment, rendering

them immunocompetent. Aside from this, the protocol was identical to that described in section 4.2.5.

For the comparison of growth profiles, a t-test was performed between the test and control replacement population density at each time point.

Table 5.1: Computational tools and optional flags used for the analysis in this chapter.

Tool	Version	Optional lags	Citation
FIJI	2.3.0/1.53f		(Schindelin et al., 2012)
ggplot2	3.3.0		(Wickham, 2016)
ggrepel	0.8.1		https://cran.r-project.org/web/packages/ggrepel/index.html
R	3.6.1		(R Core Team, 2022)
tximport	1.24.0		(Soneson et al., 2015)
GenomicFeatures	1.48.4		(Lawrence et al., 2013)
DESeq2	1.36.0	DESeqDataSetFromTximport(txi, colData = samples_clonal, design = ~ stage)	(Love et al., 2014)
vsn	3.64.0		(Huber et al., 2002)
RColorBrewer	1.1-3		https://cran.r-project.org/package=RColorBrewer
pheatmap	1.0.12		https://CRAN.R-project.org/package=pheatmap

data.table	1.14.6		https://cran.r-project.org/package=data.table
ggpubr	0.5.0		https://CRAN.R-project.org/package=ggpubr
EnhancedVolcano	1.14.0	pCutoff = 10e-32, FCcutoff = 0.5	https://github.com/kevinblighe/EnhancedVolcano
Radviz	0.9.3		https://cran.r-project.org/web/packages/Radviz

5.3. Results

5.3.1. Serial passage of unaltered pleomorphic *T. brucei* EATRO 1125 AnTat 1.1 90:13 (designated UPA for 'Unaltered Pleomorphic AnTat') generates developmental incompetence

Serial passage of the UPA cell line reduced its sensitivity to BHI oligopeptides

A population of unaltered pleomorphic *T. brucei* EATRO 1125 AnTat 1.1 90:13 ('UPA' cells) was passaged 30 times. The population density was purposefully maintained high, by leaving the cells to grow to a density $>1^6$ cells/ml, to select for cells which continued to divide at a high population density, rather than arresting as stumpy forms (Fig. S5.1). UPA cells displayed reduced sensitivity to the BHI-based oligopeptide signal as the selection for monomorphism progressed from passage 0 (wild type) to passage 3 (P3), passage 15 (P15) and passage 30 (P30) (Fig. 5.1a-d). Compared to UPA P3, UPA P30 displayed a reduced capacity to generate PAD1-positive cells (35% and 7%, respectively) and continued to progress through the cell cycle (cells arrested in 1K1N; 89% and 86%, respectively) at 48 hours after exposure to BHI oligopeptides (Fig. 5.2)

We sequenced DNA and RNA at the beginning (P0), middle (P15) and end (P30) of the UPA selection to identify mutations and transcript changes which occurred during the transition from pleomorphism to monomorphism.

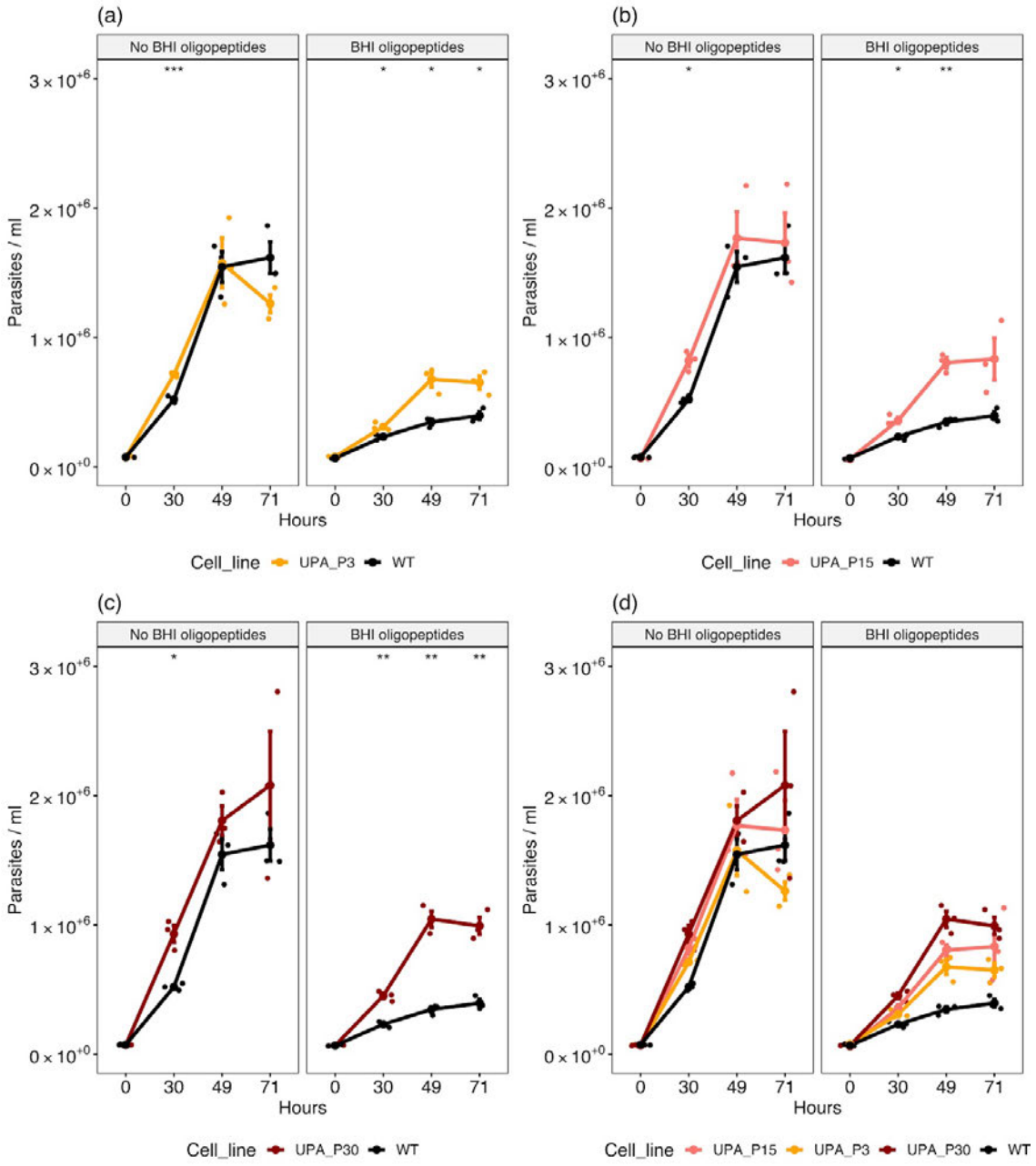


Figure 5.1: Serial passage reduces sensitivity to BHI oligopeptides. Growth of AnTat 1.1 90:13 passaged **(a)** three **(b)** 15 and **(c)** 30 times before their developmental competence was compared to pleomorphic wild-type (WT) parasites. **(d)** The parasites progressively developed insensitivity to BHI oligopeptides. The parasites were grown in HMI-9 or HMI-9 supplemented with BHI oligopeptides. A t-test was used to compare the population density of each stage to the WT. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate,

represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

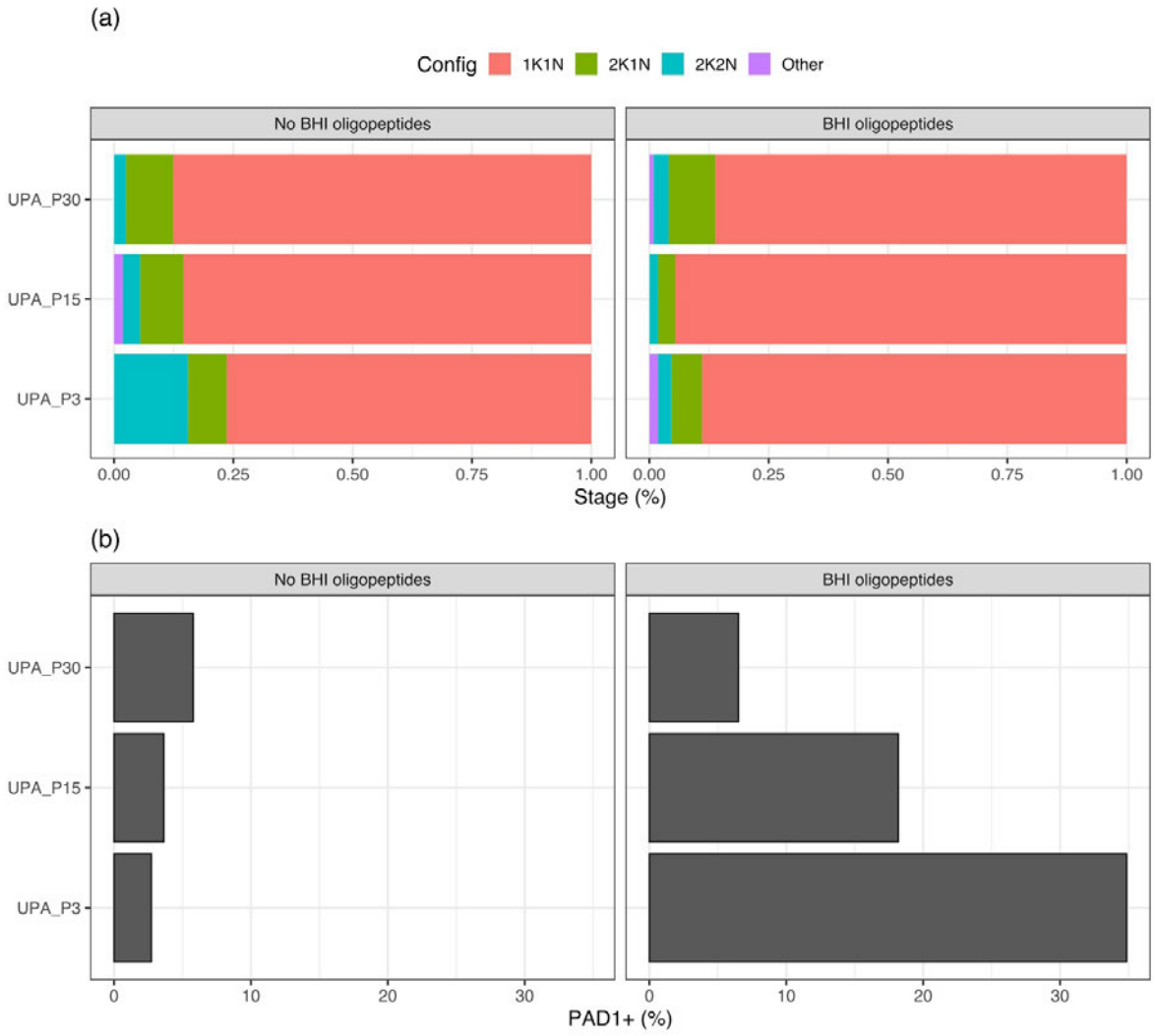


Figure 5.2: Serial passage reduces sensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence of the UPA selection at passages 3, 15 and 30. The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides and samples were prepared for analysis at 48h.

De novo mutations were not found in the serially passaged UPA cell line

Genomic data from the start and end of the UPA selection were included in the analysis performed in Chapter 3. We used the same criteria to identify mutations in two categories of genes putatively associated with QS:

1. Any high-impact, moderate-impact or modifier mutation in the UPA line in a known QS pathway gene (McDonald et al., 2018; Mony et al., 2014; Rojas et al., 2019; Tettey et al., 2022) or a high impact mutation in a gene with the following pattern in the RNA Interference Target Sequencing (RITseq) dataset: smaller log fold change in D3, D6 and PF than in the DIF category whilst displaying a log fold change in the DIF category greater than 1.5 (Alsford et al., 2011).
2. Novel genes potentially associated with the monomorphic phenotype were identified by extracting genes which had a positive selection pressure in a monomorphic clade (dN/dS >1) and a restrictive selection pressure in pleomorphic clades (dN/dS <1).

None of the identified mutations in the UPA P30 genome passed these filtering criteria.

Key quorum sensing genes, ZC3H20 and GPR89, were significantly downregulated in the serially passaged UPA cell line

The greatest transcriptional variance (65%) was detected between the ‘start’ vs ‘intermediate’ and ‘start’ vs ‘end’ of the UPA selection (Fig. 5.3). The difference between the ‘intermediate’ and ‘end’ of the UPA selection accounted for 12% of the variance. UPA P30 displayed the least sensitivity to BHI oligopeptides (Fig. 5.1-2) and exhibited a greater number of significantly differentially expressed genes compared to the start of the selection (Table 5.2). Focussing on the QS pathway genes, only two QS genes, ZC3H20 and GPR89, were significantly differentially expressed in the same direction in both the UPA selection and a previous selection for monomorphism (McWilliam et al., 2019). These genes are well-characterised as components of the quorum-sensing pathway (Cayla et al., 2020; Ling et al., 2011; Liu et al., 2020; Rojas et al., 2019).

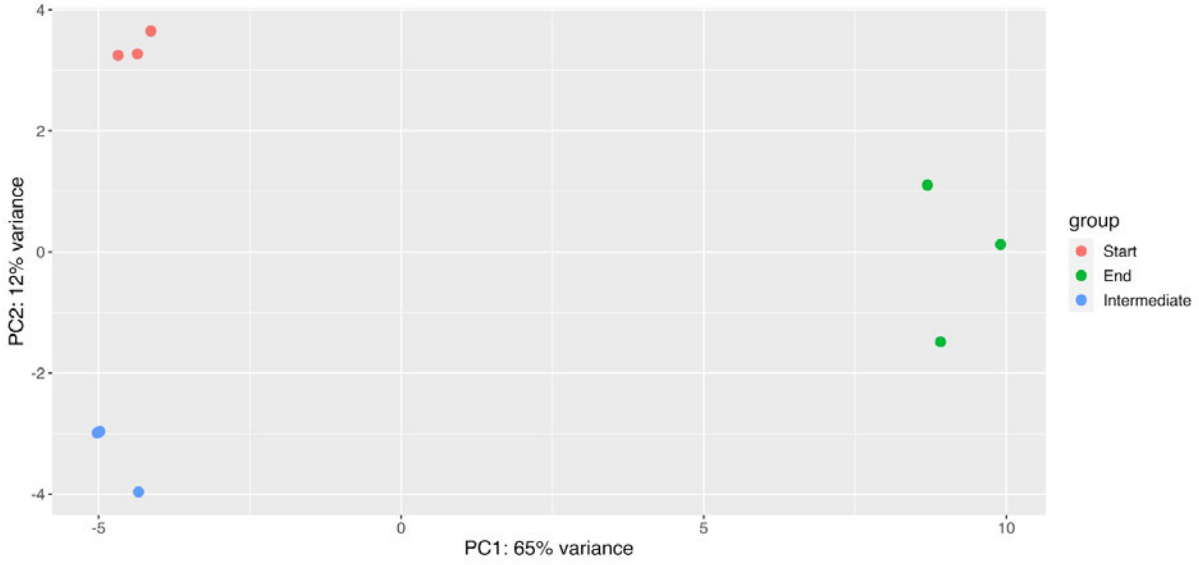


Figure 5.3: Serial passage causes variance between RNAseq libraries generated at the start, intermediate and end of the UPA selection. Principal component analysis of the UPA cell line selected for monomorphism in HMI-9. RNAseq was performed on three replicates at the start, intermediate and end stage of the selection for monomorphism.

Table 5.2: Genes are differentially expressed throughout the UPA selection. Significantly differentially expressed genes during the UPA selection (adjusted P-value < 0.05).

	Start vs intermediate	Start vs end	Intermediate vs end
LFC <0 (down)	175	1,594	1,635
LFC > 0 (up)	201	1,406	1,790

Overexpression of ZC3H20 restored developmental competence in the serially passaged UPA cell line

Given the observed downregulation of the ZCH20 transcripts in the selected UPA lines, we overexpressed ZC3H20 in the selected UPA P30 monomorphic cell line to determine whether this would restore developmental competence. This was achieved by transfecting UPA30 with a plasmid capable of ZC3H20 expression under doxycycline control. When UPA P30 ZC3H20 overexpression cells were grown in normal media, without induction of Z3CH20 overexpression, the parasites grew at a similar rate to wild-type pleomorphic *T. brucei*. The uninduced cells arrested their growth once exposed to BHI (Fig. 5.4b), potentially indicative of leaky expression of ZC3H20 from the overexpression construct. However, induction of ZC3H20 overexpression 24 hours before the start of the growth analysis caused the parasites to arrest their growth in HMI-9 and HMI-9 supplemented with BHI (Fig. 5.4c) suggesting an inducible effect caused by ZCH20 expression. Consistent with this, the induction of ZC3H20 overexpression at the start of the growth analysis generated a cell proliferation defect in HMI-9 and complete arrest in HMI-9 supplemented with BHI (Fig. 5.4d). Overexpression of ZC3H20 in the uninduced and induced cultures also caused an increase in cell cycle arrest (Fig. 5.5a) and the proportion of PAD1 positive cells (Fig. 5.5b), compared to UPA P30.

Although the response to doxycycline induction produced phenotypes indicating ZCH20 expression rescued developmental progression, we were unable to detect the expression of TY-tagged ZC3H20 in this cell line via western blotting (Fig. S5.2). Thus, the ability of ZCH20 to restore differentiation capacity could not be definitively established.

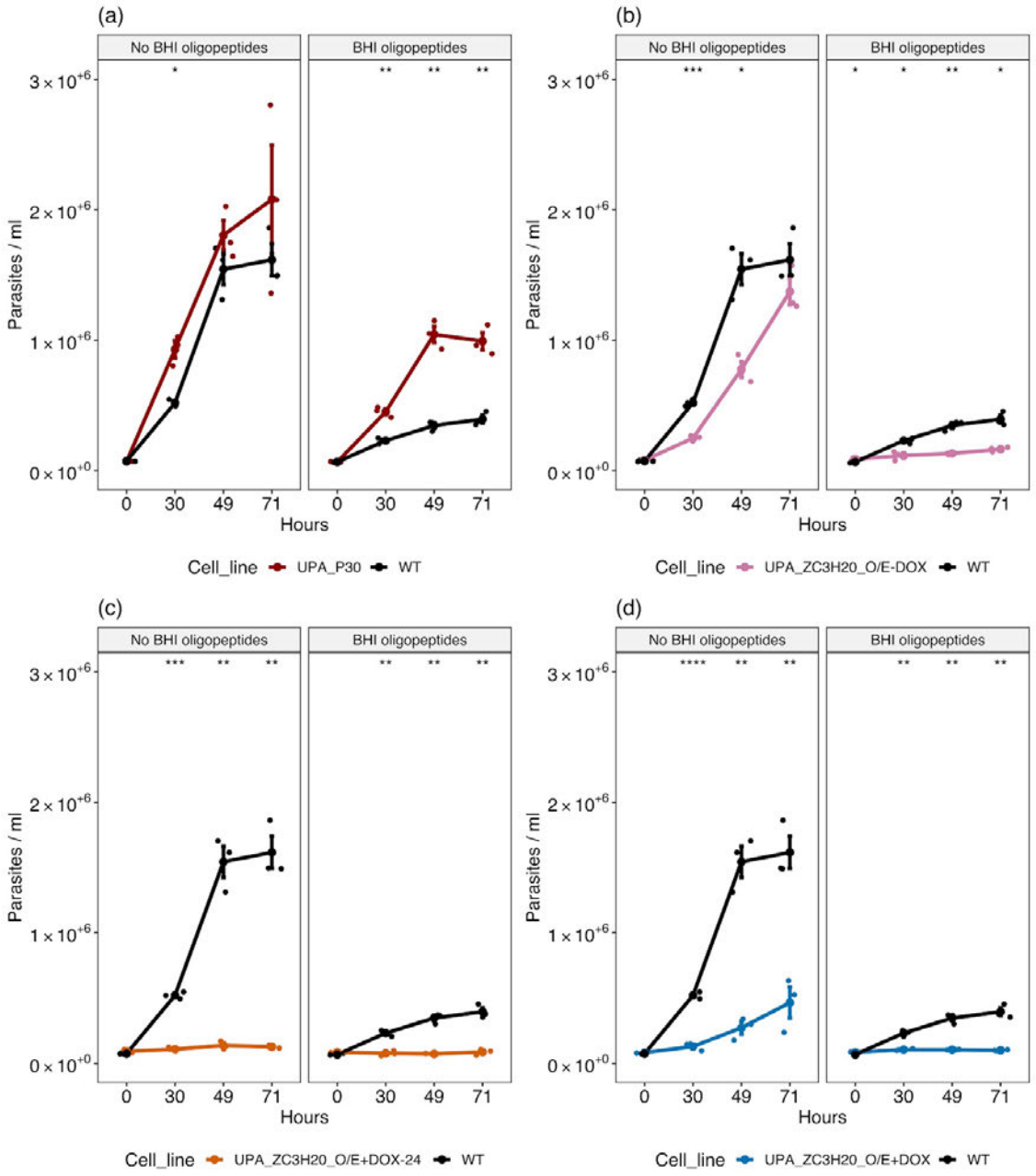


Figure 5.4: Overexpression of ZC3H20 restores sensitivity to BHI oligopeptides. Growth of **(a)** UPA P30, **(b)** UPA P30 without induction of ZC3H20 overexpression, **(c)** UPA P30 ZC3H20 with induction of ZCH20 overexpression 24 hours prior to the growth analysis and **(d)** UPA P30 with induction of ZC3H20 overexpression at the start of the growth analysis. The parasites were grown in HMI-9 or HMI-9 supplemented with BHI oligopeptides. A t-test was used to compare the population density of each stage to wild-type cells (WT). The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the dots at each time

point. A dot also represents the mean for each time point. Error bars represent mean standard error.

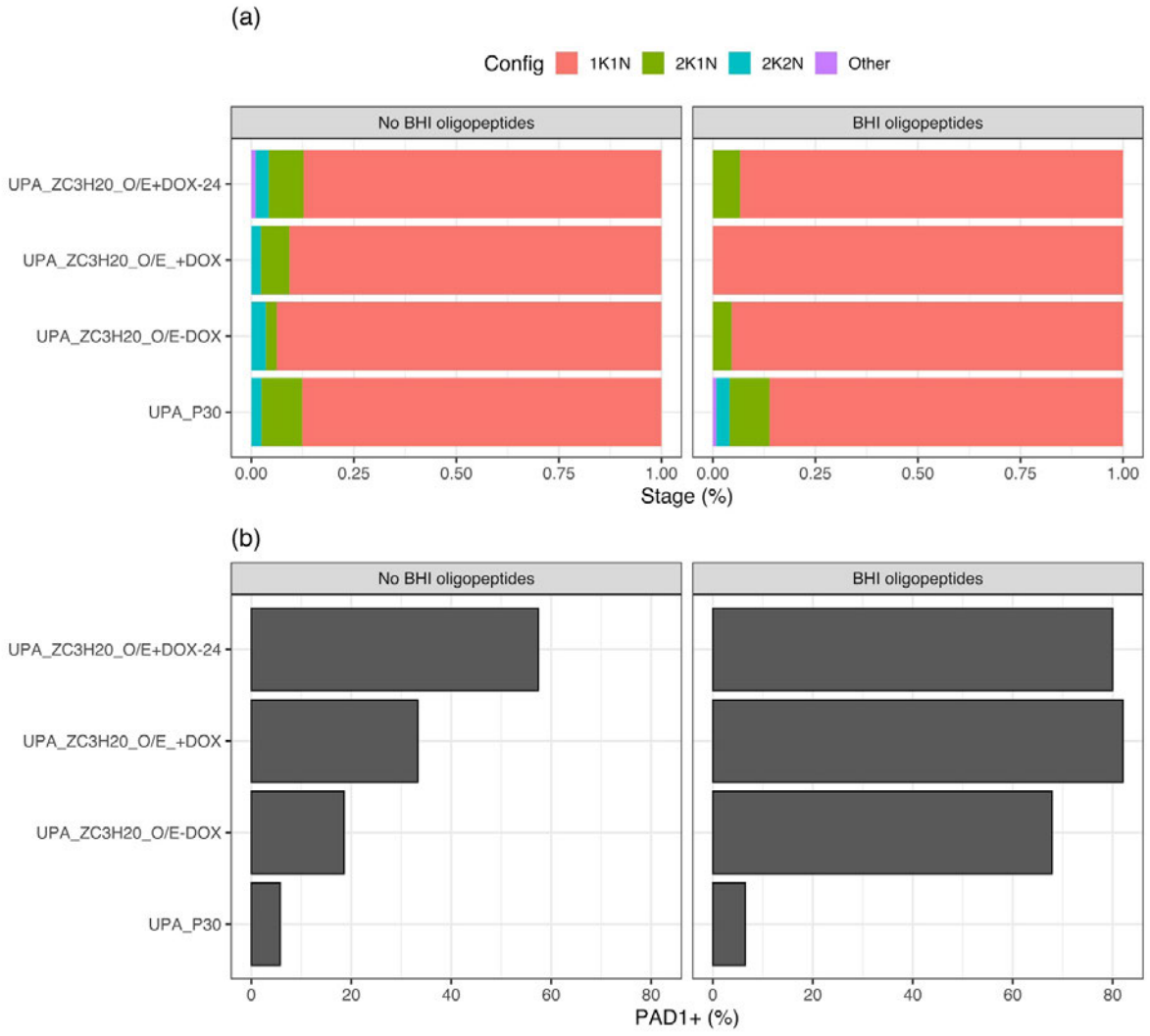


Figure 5.5: Overexpression of ZC3H20 restores sensitivity to BHI oligopeptides. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence of the UPA P30 cell line with doxycycline-inducible ZC3H20 overexpression. The parasites were either uninduced (UPA_ZCH20_O/E-DOX), induced 24 hours prior to the growth analysis UPA_ZCH20_O/E-DOX+24) or induced at the start of the growth analysis (UPA_ZCH20_O/E+DOX). The cells were either grown in normal HMI-9 or HMI-9 supplemented with 15% BHI oligopeptides.

5.3.2. BHI oligopeptides increase the speed at which developmental incompetence is generated in serially passaged pleomorphic EATRO 1125 AnTat 1.1 90:13 cells

Selection for monomorphism using BHI oligopeptides rapidly reduces developmental competence

We attempted to increase the selection pressure for monomorphism by repeating the selective regime used to isolate the UPA line but using HMI-9 media supplemented with BHI oligopeptides. Specifically, a clonal pleomorphic *T. brucei* EATRO 1125 AnTat 1.1 90:13 cell line was generated and the same clone was then serially passaged in either HMI-9 or HMI-9 supplemented with BHI oligopeptides. The amplification of clonal cell lines required three passages, so BHI was first introduced in passage 3 (P3). After passage 30, the selected lines were grown for one week in HMI-9, creating passage 31 (P31) (Fig. S5.3).

The developmental competence of the derived clonal line was compared in passages 6, 12, 19, 24 and 31. The cells selected in HMI-9 and HMI-9 supplemented with BHI both became unresponsive to BHI at the end of the selection (P31), however, selection in BHI increased the rate (Fig. 5.6a) and efficiency (Fig. 5.6e) of the development of reduced sensitivity to BHI. The reduced sensitivity was confirmed by the greater number of proliferative (Fig. 5.7a) and PAD1 negative cells (Fig. 5.7b) as the selection progressed. We observed a relatively high level of PAD1-positive cells in the wild type and HMI-9 passage 6 cell lines grown in HMI-9 without the addition of BHI, indicative of a QS response induced by high parasite density (Fig. 5.7b).

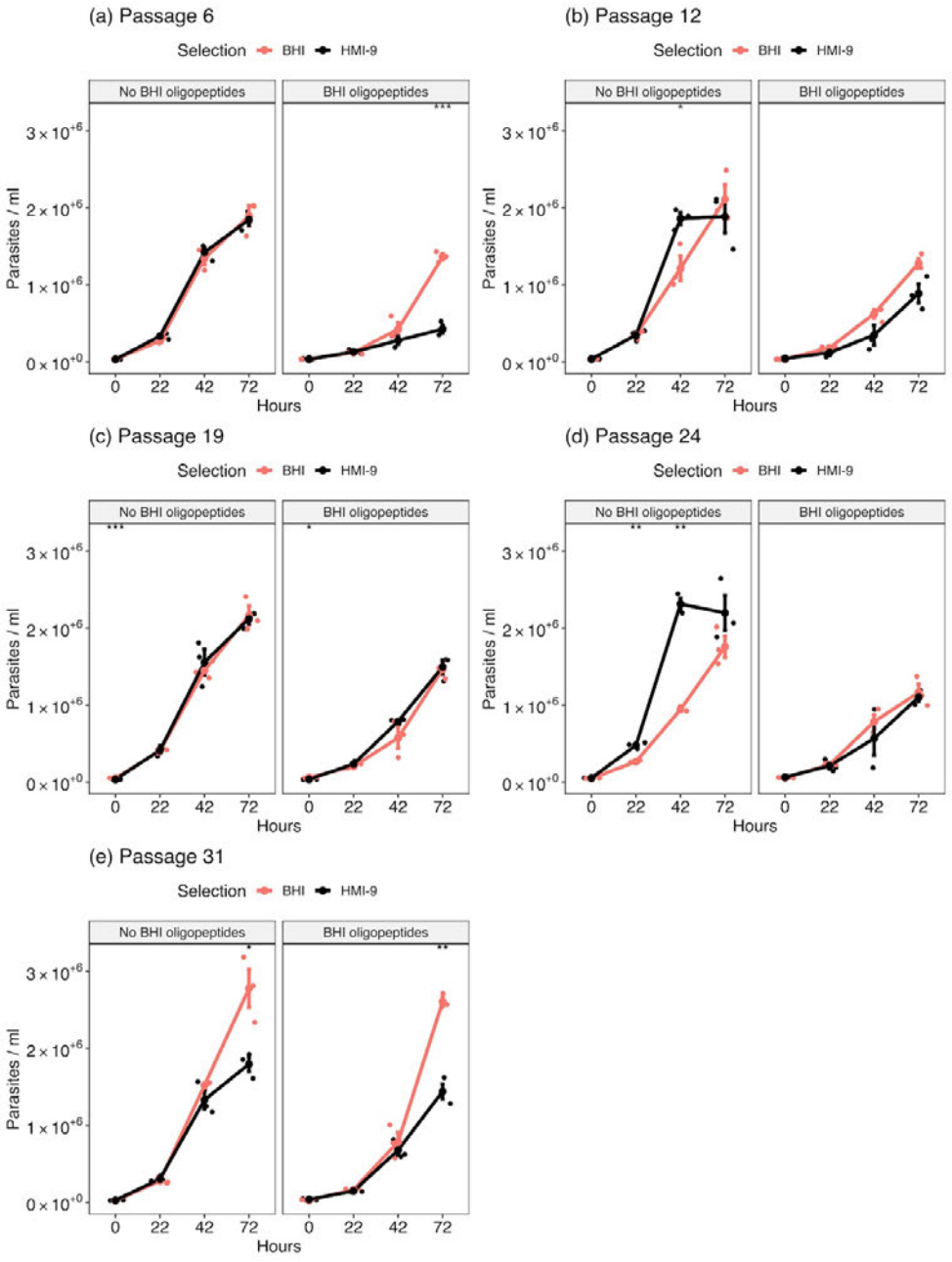


Figure 5.6: BHI oligopeptides increase the speed of selection for monomorphism. Growth of a pleomorphic *T. b. brucei* EATRO 1125 AnTat 1.1 90:13 clone selected for monomorphism in HMI-9 or HMI-9 supplemented with BHI oligopeptides. Passages **(a)** six, **(b)** twelve, **(c)** nineteen, **(d)** twenty-four and **(e)** thirty-one were grown in HMI-9 or HMI-9 supplemented with BHI oligopeptides. A t-test was used to compare the population density at each stage between the selection using HMI-9 or HMI-9 supplemented with BHI oligopeptides. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate, represented by the

dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

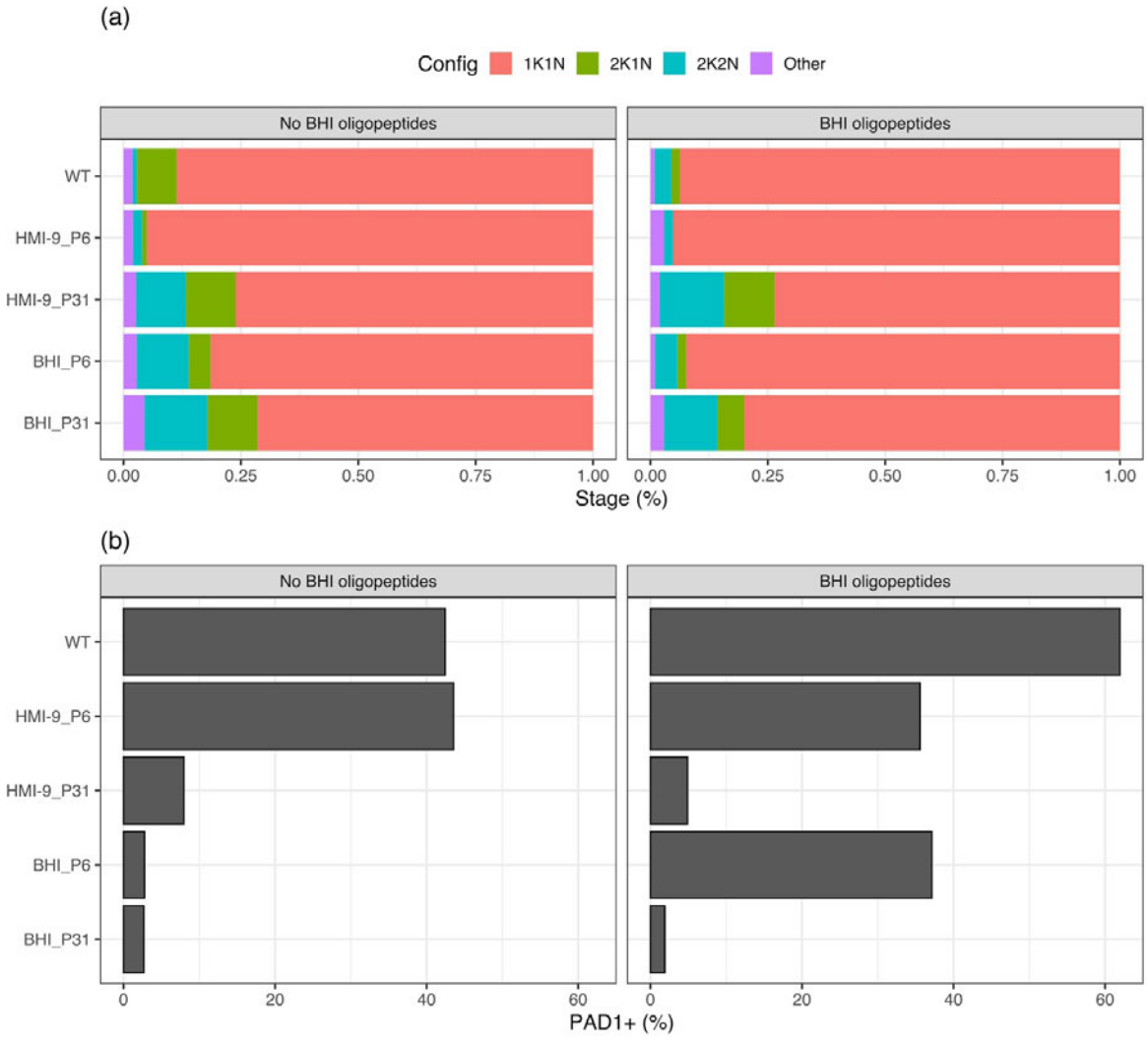


Figure 5.7: BHI oligopeptides increase the speed of selection for monomorphism. (a) Cell cycle stage and **(b)** the percentage of cells expressing PAD1 as assessed by immunofluorescence of pleomorphic cells selected for monomorphism in HMI-9 or HMI-9 supplemented with BHI oligopeptides. Wild-type cells (WT) and passages six and thirty-one were grown in HMI-9 (HMI-9_P6; HMI-9_P31) or HMI-9 supplemented with 15% BHI oligopeptides (BHI_P6; BHI_P31).

Developmentally competent cell lines, generated via *in vitro* serial passage, are developmentally incompetent *in vivo*

To determine the level of monomorphism in the BHI-selected cell line, pleomorphic wild-type (WT) and clonal passages 6 and 31 derived under BHI selection were used to infect mice. The parasitaemia of the P31 line was similar to the WT cell line whilst the P6 line displayed the fastest growth rate in this experiment. In each case, mice cleared each of the infections after eight days (Fig. 5.8a-b). At the peak of parasitaemia (day 6), P31 parasites continued to progress through their cell cycle as assessed by the proportion of 2K1N and 2K2N cells (Fig. 5.8c). The proportion of PAD1-positive cells was low for mice infected with P31 whilst P6 and WT parasites maintained developmental competence, accumulating as 1K1N and expressing PAD1 (Fig. 5.8d). This demonstrated that the parasites had become less able to generate stumpy forms after 31 passages in the presence of BHI. One mouse infected with wild-type cells (WT_1Z) and two mice infected with passage 31 cells (P31_3R and P31_3RR) cleared the infection on day seven. By day eight, all of the mice had cleared the infection, thus although the selected cell line displays developmental incompetence, the mouse is able to control the infection, potentially through immune clearance in these immunocompetent mice.

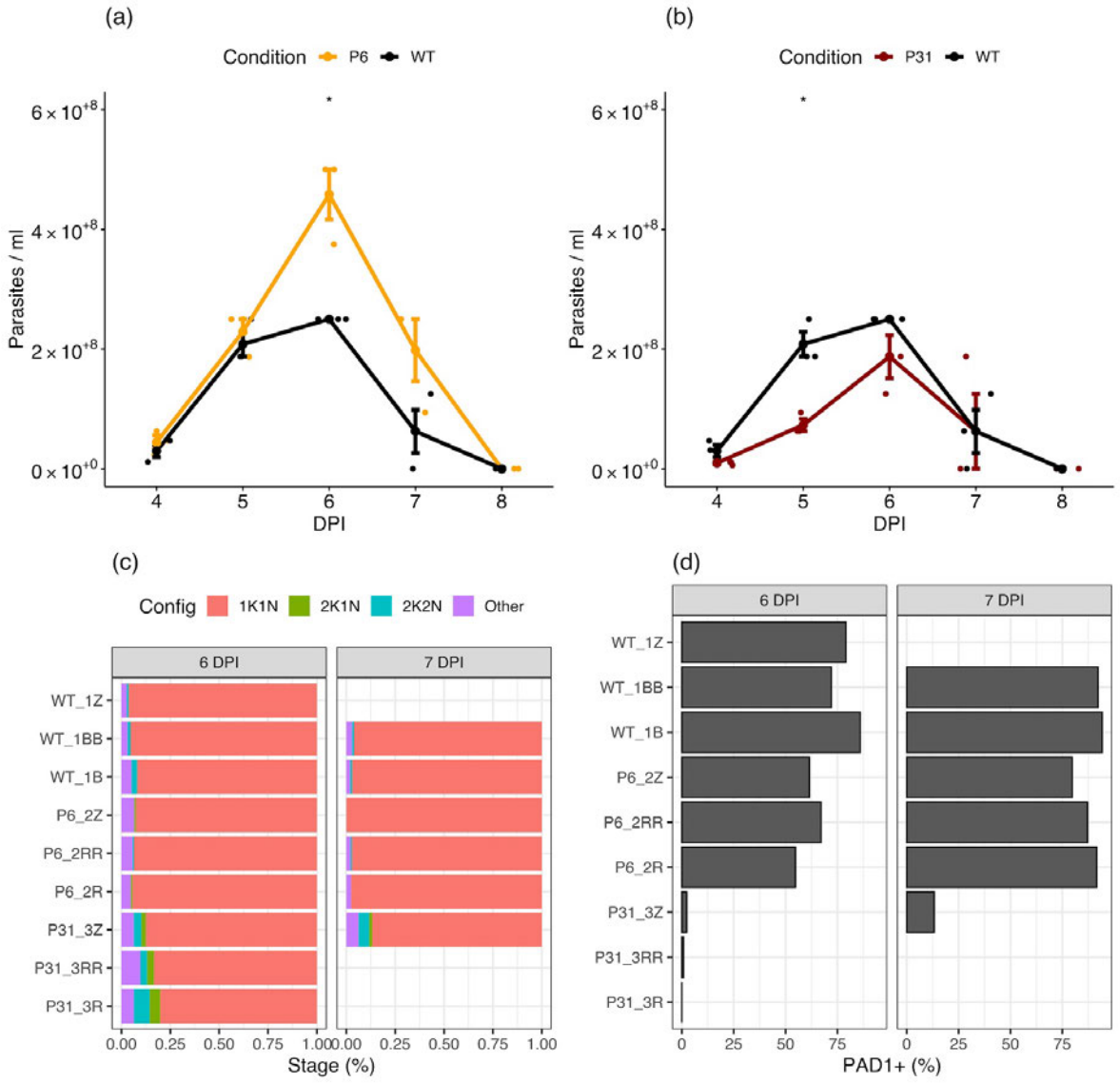


Figure 5.8: Selected monomorphic cells are developmentally incompetent. *In vivo* population growth of wild type (WT) and **(a)** passage six or **(b)** passage 31. **(c)** Cell cycle stage and **(d)** the percentage of cells expressing PAD1 at days six and seven post-infection of pleomorphic wild type (WT_1Z, WT_1B and WT_1BB) and pleomorphic cells selected for monomorphism in HMI-9 supplemented with BHI oligopeptides at passage 6 (P6_2Z, P6_2R and P6_2RR) and passage 31 (P31_3Z, P31_3R and P31_3RR). Three mice were infected with each cell line. The mice successfully cleared the infection by day 7 for WT_1Z, P31_3R and P31_3RR.

5.3.3. Six independent clones, subjected to serial passage, display developmental incompetence and common transcriptomic profiles

Six independent clones, derived by serial passage in increasing concentrations of BHI-based oligopeptides, display developmental incompetence

After confirmation that selection in BHI-supplemented media rapidly produced a monomorphic phenotype, we began the selection of monomorphism from six separate clonal pleomorphic *T. brucei* lines in HMI-9 supplemented with BHI (Fig. S5.4). This was intended to explore whether the generation of monomorphism in independent lines followed the same molecular trajectory. Comparison of the start (P0) and end (P30) of the selections confirmed that *in vitro* passage had generated six clonal monomorphic cell lines which displayed reduced sensitivity to BHI (Fig. 5.9).

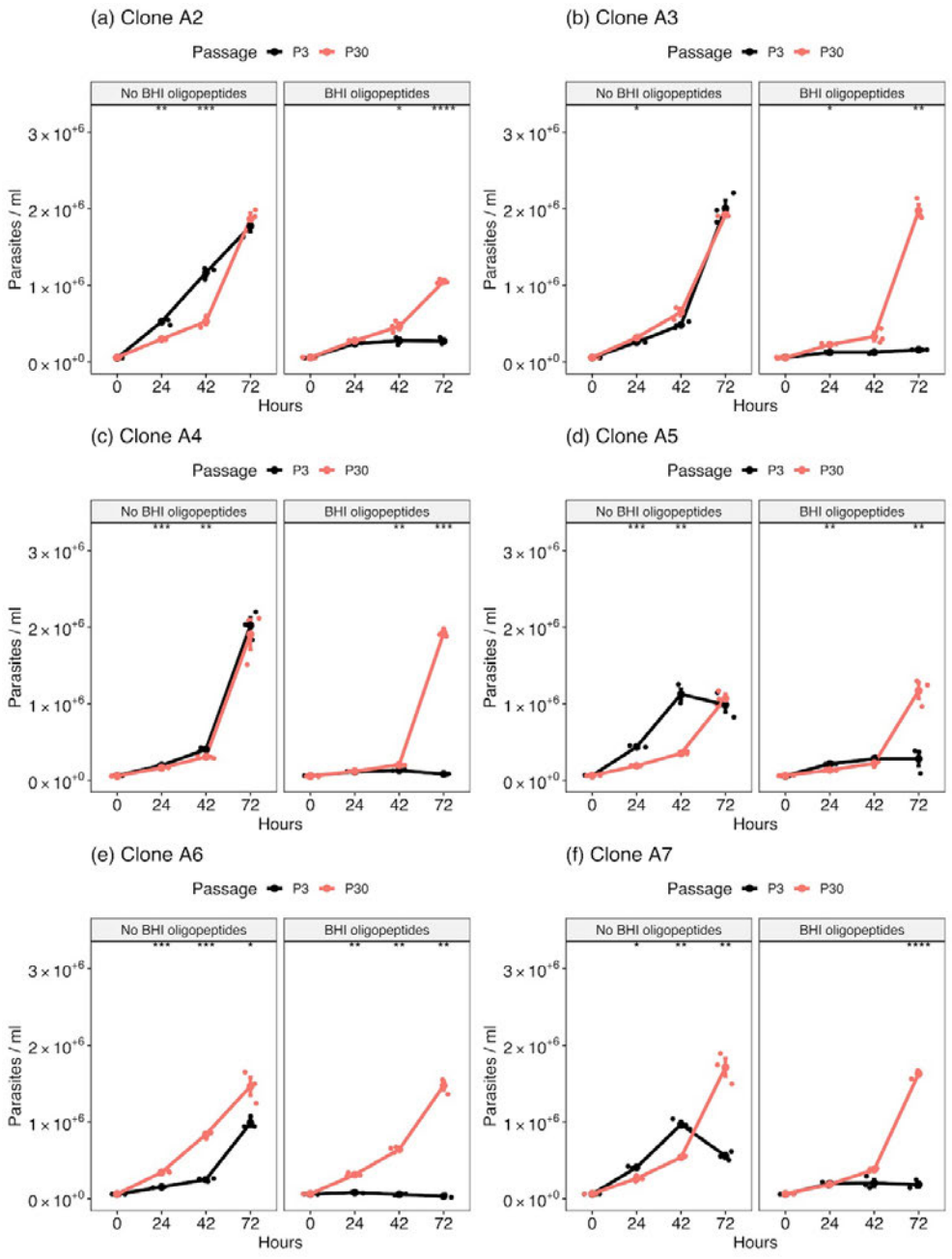


Figure 5.9: Six independent clones (A2-A7), derived by passage in increasing concentrations of BHI-based oligopeptides, display developmental incompetence. Black lines indicate the growth profile of the clones at the start of the selection and salmon represents clones at the end of the selection after 30 passages. The clones were grown in either HMI-9 or HMI-9 supplemented with 15% BHI. A t-test was used to compare the population density of each stage. The significance is represented by asterisks; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001 and ****: p <= 0.0001. Each cell line was grown in triplicate,

represented by the dots at each time point. A dot also represents the mean for each time point. Error bars represent mean standard error.

Independently derived developmentally incompetent clones share 20 significantly differentially expressed genes

In order to assess transcript changes associated with the selection for monomorphism, parasites from the ‘start’ and ‘end’ of each clonal selection series were grown in HMI-9 and RNA was harvested from triplicate cultures of each clone. There was no significant difference between the population density of the ‘start’ and ‘end’ cultures for each clone when RNA was harvested (Fig. S5.5). PCA analysis highlighted that clone A1 (pilot selection) displayed a distinct profile from the other clones and one of the pleomorphic replicate libraries of clone A3 grouped with the monomorphic libraries (Fig. S5.6). These clones were excluded from further analysis.

The variance between the libraries was described by the selection from pleomorphism to monomorphism (41%) (Fig. 5.10). Although the ‘start’ and the ‘end’ transcript profiles consistently grouped, replicates of the same clone within each stage of the selection did not always cluster, suggesting the variance observed within the biological replicates was greater than the variance between the independent clonal selections for monomorphism. There were thousands of differentially expressed genes for each clone during the selection of monomorphism, however, retaining significantly differentially expressed genes with a log fold change greater than one which was shared between each clone, left 20 genes (Table 5.3, Fig. 5.11 and Fig. 5.12).

RBP10 (Tb927.8.2780), a known regulator of trypanosome differentiation, was consistently significantly downregulated in the selected monomorphic lines (Fig. 5.12). The 20

significantly differentially expressed genes identified in each selection were significantly enriched for genes bound by RBP10 in an mRNA tethering assay (8 genes) (Mugo and Clayton, 2017). However, these genes were often expressed in the opposite direction between our selected cell lines exhibiting a downregulation of RBP10 and published experiments with an RNAi-induced suppression of RBP10 (Table 5.4) (Mugo and Clayton, 2017).

Due to the interest in RBP10, we manually analysed mutations in this gene present in naturally occurring monomorphic *T. brucei.* Across all of the monomorphic and pleomorphic isolates, 99.5% of the sites were identical, and monomorphic clade-specific mutations were largely predicted to have a low-impact phenotype. However, *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat share a homozygous non-synonymous mutation (N151S), which creates a putative phosphorylation site (Table 5.5). The mutation is heterozygous in *T. b. brucei* TREU927/4, which is pleomorphic.

Finally, the expression profile of a manually curated list of QS genes was analysed. The libraries cluster into pleomorphic and monomorphic isolates. The greatest differentially expressed gene in this subset was ZC3H20 (Tb927.7.2660), this forming a discrete branch from the remaining QS genes that also includes PKA-R (Tb927.11.4610) and adenylosuccinate synthetase (Tb927.11.3650) (Fig. 5.13).

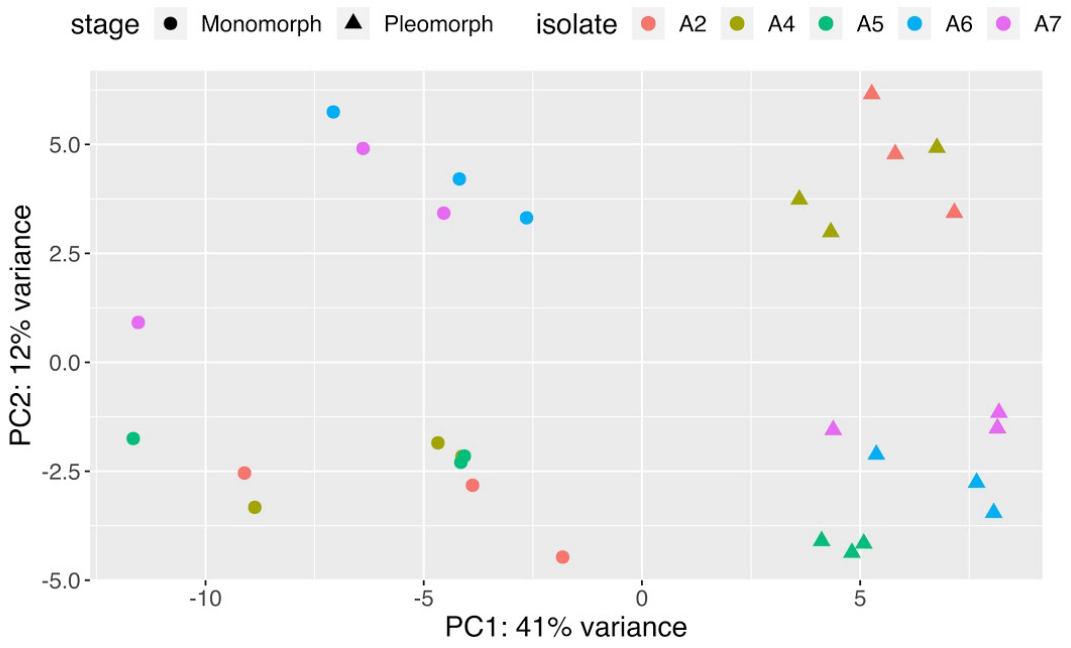


Figure 5.10: Variance between independent clonal RNAseq libraries is described by the selection from pleomorphism to monomorphism. Principal component analysis of five clonal cell lines selected for monomorphism in BHI-supplemented HMI-9. RNAseq was performed on three replicates of each clone at the start and end of the selection for monomorphism.

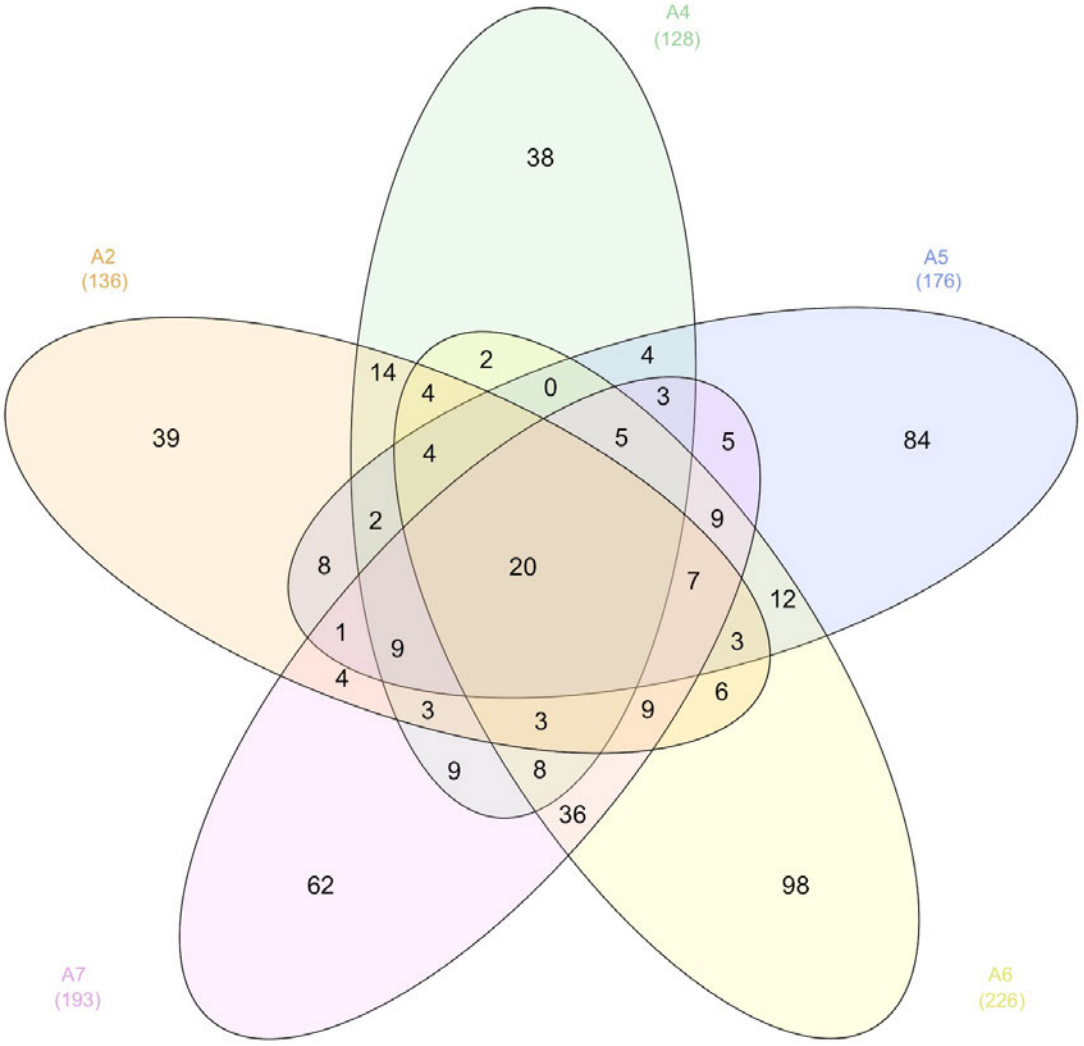


Figure 5.11: 20 significantly differentially expressed genes are shared between clones selected for monomorphism. Counts of significantly differentially expressed upregulated or downregulated genes during the selection of monomorphism in BHI-supplemented HMI-9. The genes were filtered to retain genes which had a greater than 1 log2 fold change and an adjusted p-value below 0.05.

Table 5.3: Common significantly differentially expressed (adjusted p-value <0.05) genes with a log fold change greater than 1 between the 'start' and 'end' of the selection of monomorphism in five independent clones.

Gene ID	Product	Gene	Bound by RBP10?
Tb09.v4.0150	VSG, pseudogene		No
Tb11.v5.0242	dynein heavy chain, putative		No
Tb11.v5.0753	hypothetical protein, conserved		No
Tb11.v5.0833	kinesin, putative		No
Tb927.10.12650	hypothetical protein, conserved		No
Tb927.10.12660	pumilio/PUF RNA binding protein 2	PUF2	No
Tb927.10.14550	ATP-dependent RNA helicase HEL67	HEL67	No
Tb927.10.3410	hypothetical protein		Yes
Tb927.10.5240	cAMP binding protein, putative		No
Tb927.10.8050	TFIIF-stimulated CTD phosphatase, putative		Yes
Tb927.11.1280	hypothetical protein, conserved		Yes
Tb927.11.3440	hypothetical protein, conserved		No
Tb927.2.4200	CMGC/CLK family protein kinase, putative		Yes
Tb927.4.1910	hypothetical protein, conserved		Yes
Tb927.5.293b	hypothetical protein		No
Tb927.5.4575	hypothetical protein		No

Tb927.7.2680	zinc finger protein family member, putative	ZC3H22	Yes
Tb927.8.2780	RNA-binding protein RBP10, putative	RBP10	Yes
Tb927.8.510	hypothetical protein		No
Tb927.8.6490	Basal body protein	BBP122	Yes

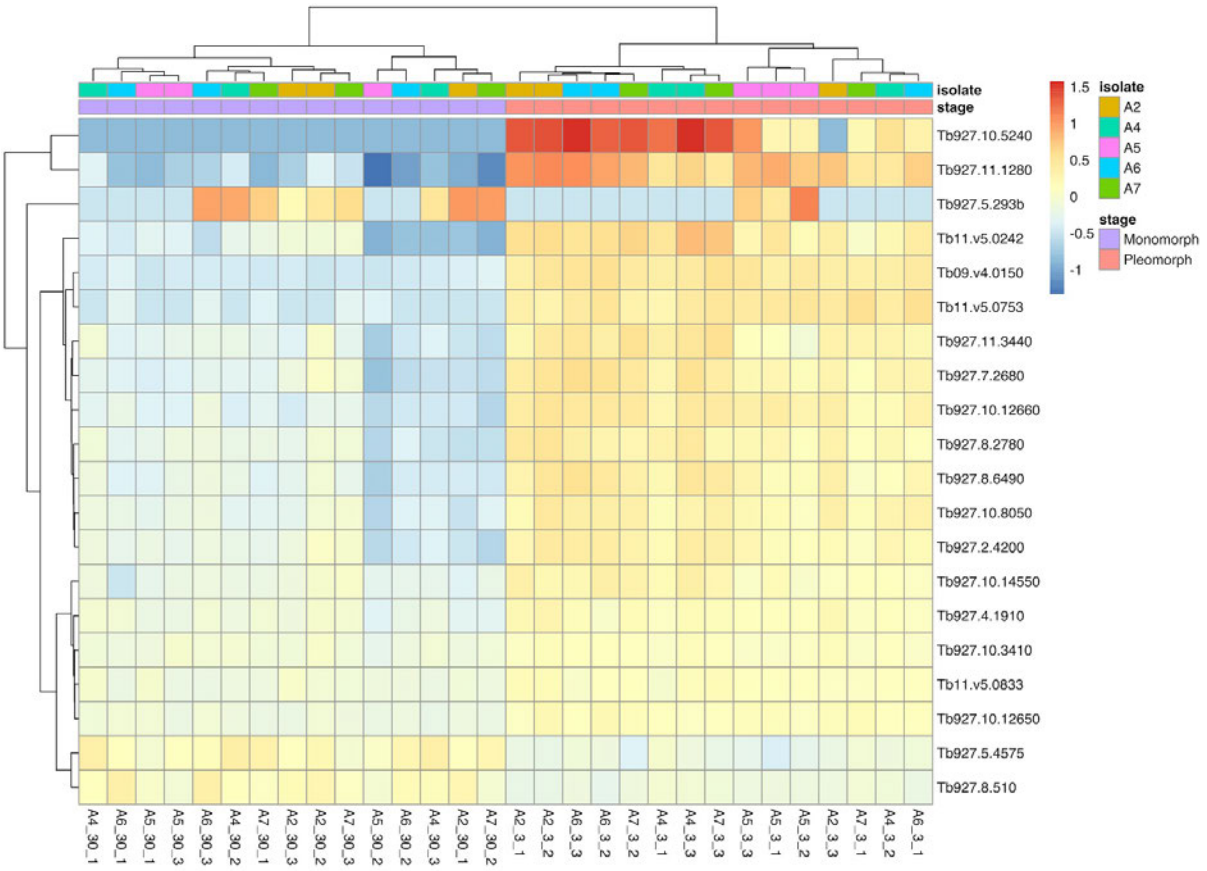


Figure 5.12. 20 significantly differentially expressed genes are shared between clones selected for monomorphism. Normalised rlog of 20 significantly differentially expressed genes between the start (pleomorphic: A2_3, A4_3, A5_3, A6_3 and A7_3) and end (monomorphic: A2_30, A4_30, A5_30, A6_30 and A7_30) of the selection of monomorphism in five clonal cell lines. Three biological replicates were generated for each stage of each clone (AX_X_1, AX_X_2 and AX_X_3). Normalisation was achieved by calculating the deviation of expression from each gene in a specific sample from the gene's mean expression across all samples. Blocks of genes and samples which covary are clustered, indicated by branches. The split between each stage of the selection was not specified.

Table 5.4: Common significantly differentially expressed genes between the 'start' and 'end' of the monomorph selection which are bound by RBP10 in an mRNA tethering assay (Mugo and Clayton, 2017). The list of 20 significantly differentially expressed genes is highly enriched with genes bound by RBP10 (8 genes). Fisher's Exact Test for Count Data for enrichment p-value = 5.435e-07.

Gene ID	Product	Gene Name	RBP10 RNAi (Mugo and Clayton, 2017)	Monomorph (this study)
Tb927.10.3410	hypothetical		Down	Down
Tb927.10.8050	TFIIF-stimulated CTD phosphatase		Up	Down
Tb927.2.4200	CMGC/CLK family kinase		Up	Down
Tb927.4.1910	hypothetical		Up	Down
Tb927.11.1280	hypothetical		Down	Down
Tb927.7.2680	zinc finger protein family	ZC3H22	Up	Down
Tb927.8.2780	RNA-binding protein	RBP10	Down	Down
Tb927.8.6490	Basal body	BBP122	Up	Down

Table 5.5: RBP10 (Tb927.8.27800) clade-specific mutations in naturally occurring monomorphic *T. brucei.*

Nucleotide location	Amino acid location alteration	Clade	Impact
253	R85W	*T. b. evansi* type IVM-t1	Heterozygous non-synonymous mutation
452	N151S	*T. b. equiperdum* type BoTat and *T. b. equiperdum* type OVI	Homozygous non-synonymous mutation. The SNP is heterozygous in TREU927/4 but homozygous in the monomorphic clade. Creates a putative phosphorylation site.
885		*T. b. equiperdum* type BoTat	Homozygous synonymous mutation

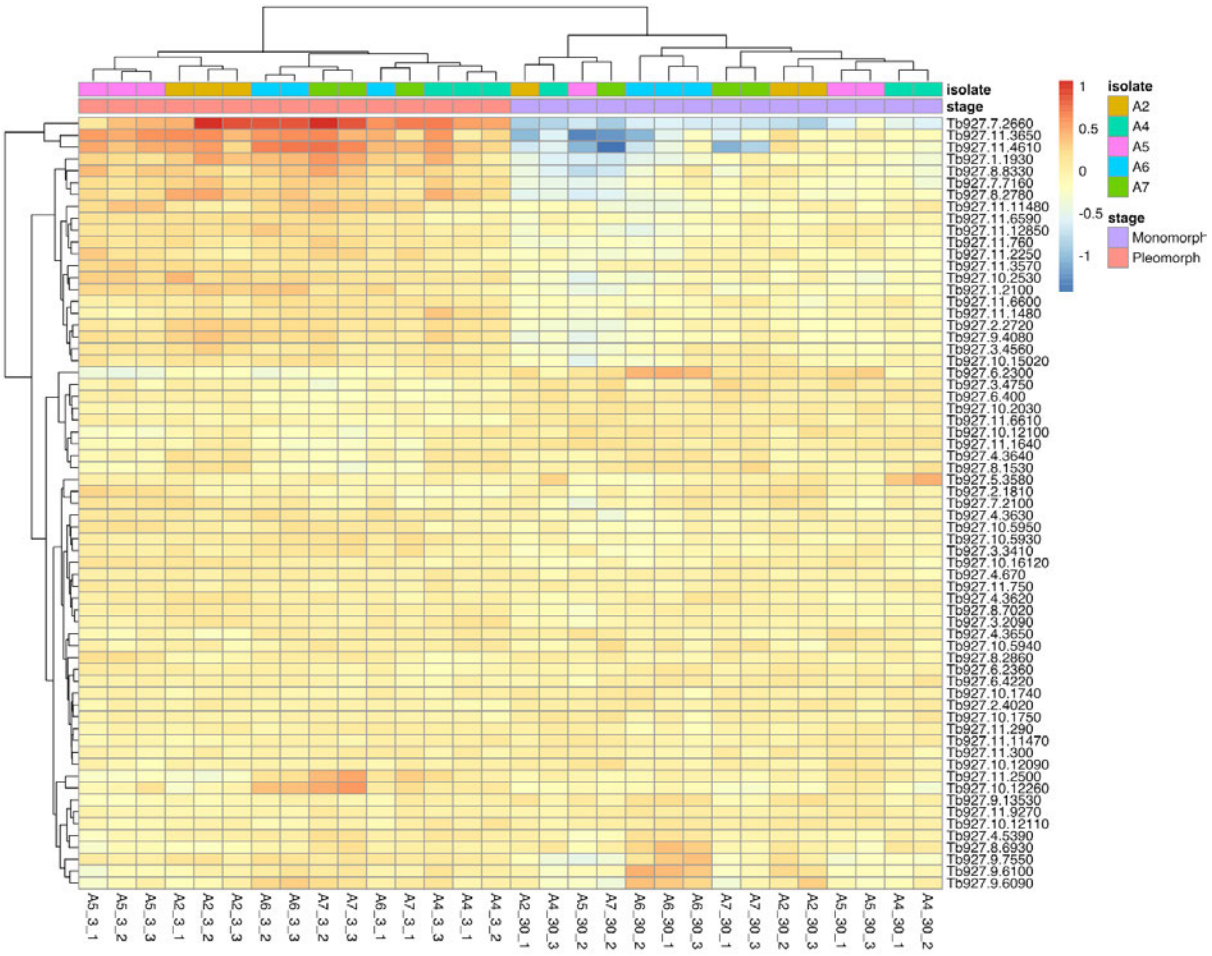


Figure 5.13. Differentially expressed genes in the QS pathway are shared between clones selected for monomorphism. Normalised rlog of a manually curated list of genes in the QS pathway at the start (pleomorphic: A2_3, A4_3, A5_3, A6_3 and A7_3) and end (monomorphic: A2_30, A4_30, A5_30, A6_30 and A7_30) of the selection for monomorphism in five clonal cell lines. Three biological replicates were generated for each stage of each clonal selection (AX_X_1, AX_X_2 and AX_X_3). Normalisation was achieved by calculating the deviation of expression from each gene in a specific sample from the gene's mean expression across all samples. Blocks of genes and samples which covary are clustered, indicated by branches. The split between each stage of the selection was not specified.

5.4. Discussion

In this Chapter, we have begun to highlight early events in the generation of monomorphism, such as the downregulation of ZC3H20 mRNA, which we were able to counteract via inducible overexpression of the same gene. We highlight 20 genes whose expression was consistently altered during the development of monomorphism. These genes were significantly enriched for those bound in an mRNA tethering assay by RBP10, a well-characterised maintainer of the slender form differentiation state (Mugo and Clayton, 2017; Mugo et al., 2017). Although ZC3H20 just fell outside of our filtering criteria which identified the 20 common differentially expressed genes in the selected monomorphic lines, ZC3H20 was downregulated in all clonal selections and is directly bound by RBP10 (Liu et al., 2020).

RBP10 is downregulated in stumpy-form parasites (Dejung et al., 2016) and initiates a regulatory cascade by binding proteins and targeting them for translational repression and destruction (Mugo and Clayton, 2017). RNAi against RBP10 in slender forms initiated the transition to procyclic forms when grown in procyclic media and overexpression of RBP10 in the procyclic form primes the parasites for growth as the bloodstream form (Mugo and Clayton, 2017; Mugo et al., 2017), suggesting that the downregulation of RBP10 should induce differentiation. These results seem to oppose our findings as we see downregulation of RBP10 in rapidly dividing slender-form parasites which do not readily differentiate *in vivo* or *in vitro*. However, Mugo and Clayton, (2017) did not directly investigate the transition from the slender form to the stumpy form. Instead, they focussed on the transition from slender directly to the procyclic form, which can be achieved (Schuster et al., 2021) with low efficiency (Matthews and Larcombe, 2022). RBP10-bound genes were often differentially expressed in the opposite direction in our study compared to RNAi downregulation of RBP10 (Mugo and Clayton, 2017). Therefore, RBP10 could perform opposing roles in the same

gene pathways at different stages of developmental progression from the slender to stumpy form. In future experiments, we will perform overexpression of RBP10 in the clonally-selected monomorphic cell lines in an attempt to rescue pleomorphism, similar to the analysis of ZCH20. Should the inducible overexpression of RBP10 restore developmental competence, we can then dissect the hierarchy of the RBP10 cascade through the manipulation of genes bound by RBP10 which are differentially expressed in selected monomorphic parasites.

Although there were only 20 differentially expressed genes above our threshold cutoffs, variation in gene expression between the clonally selected monomorphic lines was similar, with biological replicates of one clone often grouping with a different clone. This could suggest that the starting population was already clonal. Alternatively or additionally, as the clones were under the same selection pressure as one another, they could have undergone the same alterations in gene expression during the serial passage, suggesting a common path toward monomorphism in each line. In future, we will knock out these 20 differently expressed genes to validate their role in developmental competence. Our results do not corroborate the findings of previously selected monomorphism studies performed *in vivo* (Cai et al., 2021) since the two genes which were consistently downregulated in their study (Tb927.11.370 and Tb927.5.810) were not consistently significantly downregulated ours. Whilst this could suggest that there are multiple routes to generate a monomorphic phenotype, we have concerns regarding the clustering of the *in vivo* selected RNAseq replicates (Cai et al., 2021), which could have precluded the identification of common differentially expressed transcripts with their data.

As for Chapter 4, we have largely used *in vitro* techniques to validate the differentiation capacity of the selected monomorphic and rescue cell lines. Ultimately, our findings will be validated *in vivo*. At P6, the BHI-selected clone displayed a rapid growth rate *in vivo* and reduced developmental competence. In contrast, developmental competence had been lost

by P31 but the growth rate was slower than in P6. We suggest this was caused by culture adaptation of parasites at P31 rendering them less able to grow effectively *in vivo* compared to earlier passages. Future experiments will begin the infection from the bloodstock of a donor mouse, ensuring an equivalent infection dose with viable parasites is delivered, and the mice will be cyclophosphamide treated before the infection to counteract any differential sensitivity to the host immune system between the lines. Although we saw a phenotype upon induction of ZC3H20 overexpression, we were unable to detect the overexpression of ZC3H20 via a western blot against the Ty1 tag. In future, we will confirm the expression of the construct with RT-qPCR. This will also inform on the potential leakiness of the overexpression plasmid.

Our preliminary results suggest that monomorphism is initially a reversible phenotype, associated with the misregulation of genes such as RBP10 and ZC3H20, which can be counteracted by their re-expression. We suggest that such parasites could be described as proto-monomorphic. In the field, parasites which have escaped the geographical range of the tsetse fly could display this proto-monomorphic phenotype. Continual evolution as a proto-monomorph would enforce obligate asexual reproduction, reduced population genetic diversity, erosion of kDNA and the accumulation of deleterious mutations in genes required for a developmentally competent complex life cycle.

5.5. Supplementary material

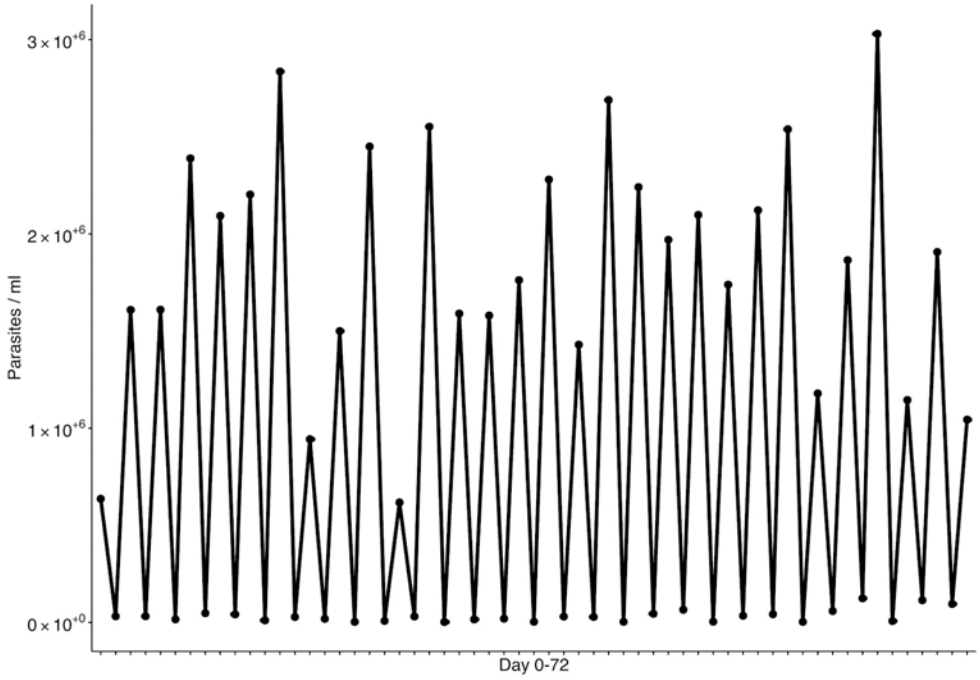


Figure S5.1: Parasite density throughout the 72-day selection of the UPA cell line in HMI-9.

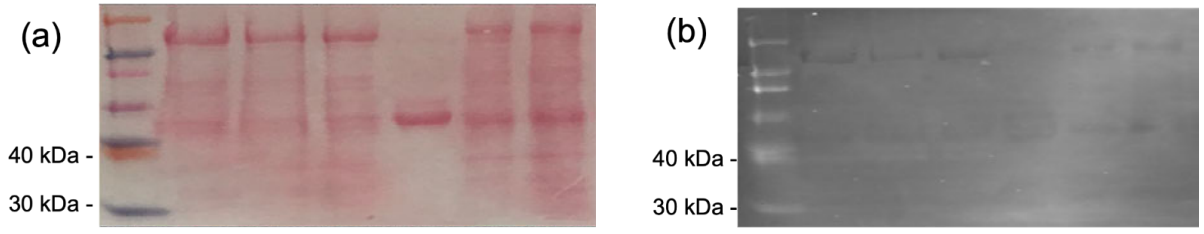


Figure S5.2: Western blot was unable to confirm the expression of Tb927.7.2660-ZC3H20-pDEX-577y. Samples of the UPA P30-ZC3H20-pDEX-577y cell line with doxycycline-induced expression (lane 5) and uninduced (lane 6) compared to four positive control samples (lanes 1-4) (two bands at 40kDa). **(a)** Ponceau S stain loading control and **(b)** BB2 anti-Ty1 western blot. Tb927.7.2660-Ty1 expected size = 44 kDa.

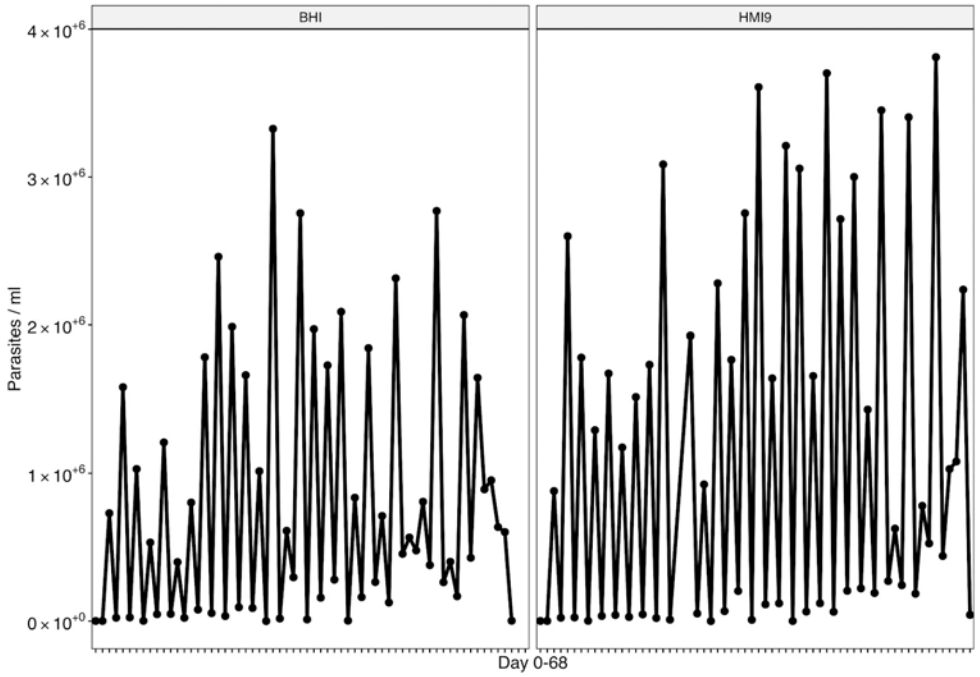


Figure S5.3: Parasite density throughout the 68-day selection of clone A1 in either HMI-9 or HMI-9 supplemented with BHI oligopeptides.

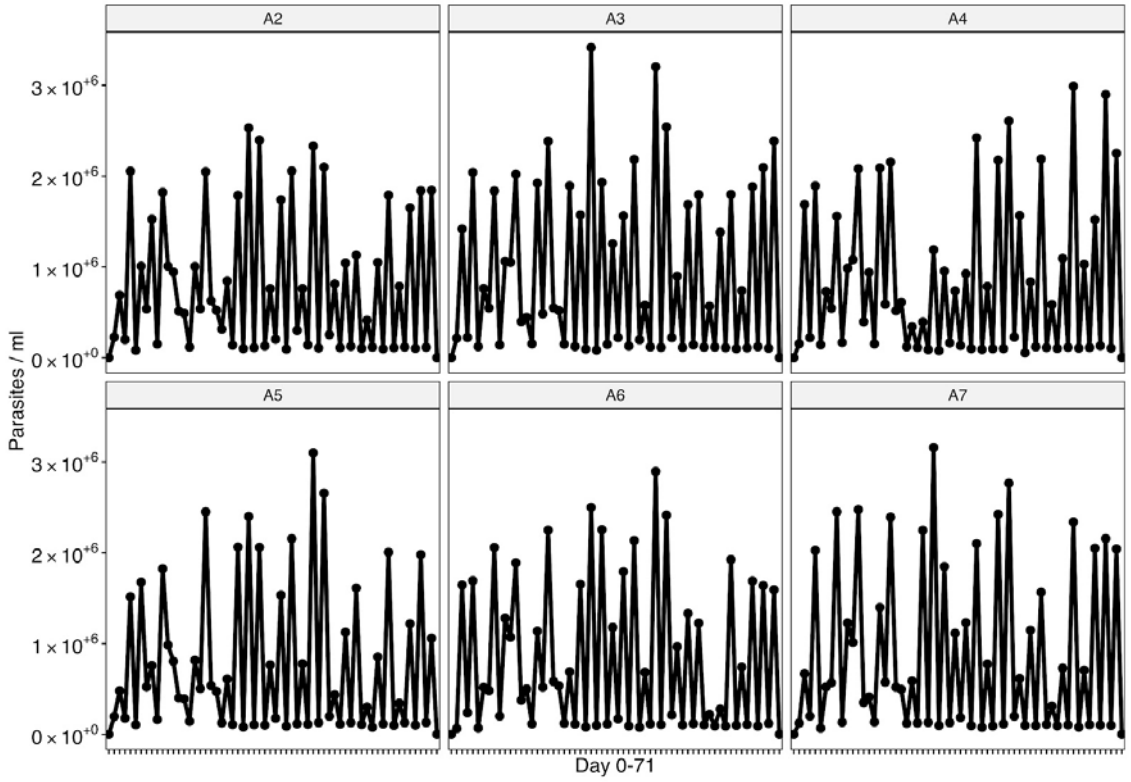


Figure S5.4: Parasite density throughout the 71-day selection of clones A2-A7 in HMI-9 supplemented with BHI oligopeptides.

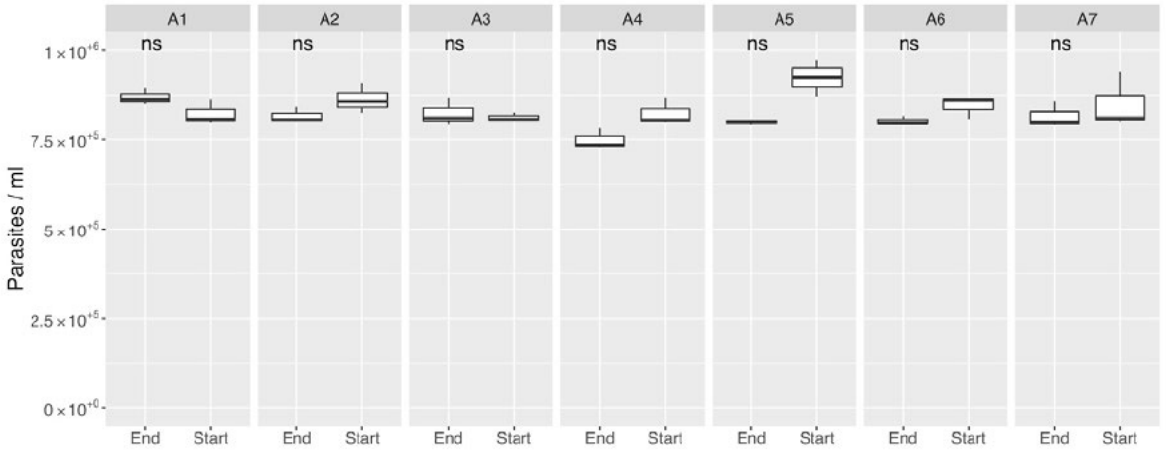


Figure S5.5: Clonal monomorph (A1-A7) density at RNA harvest. RNA was extracted from three replicates at the start and end of the selection for monomorphism.

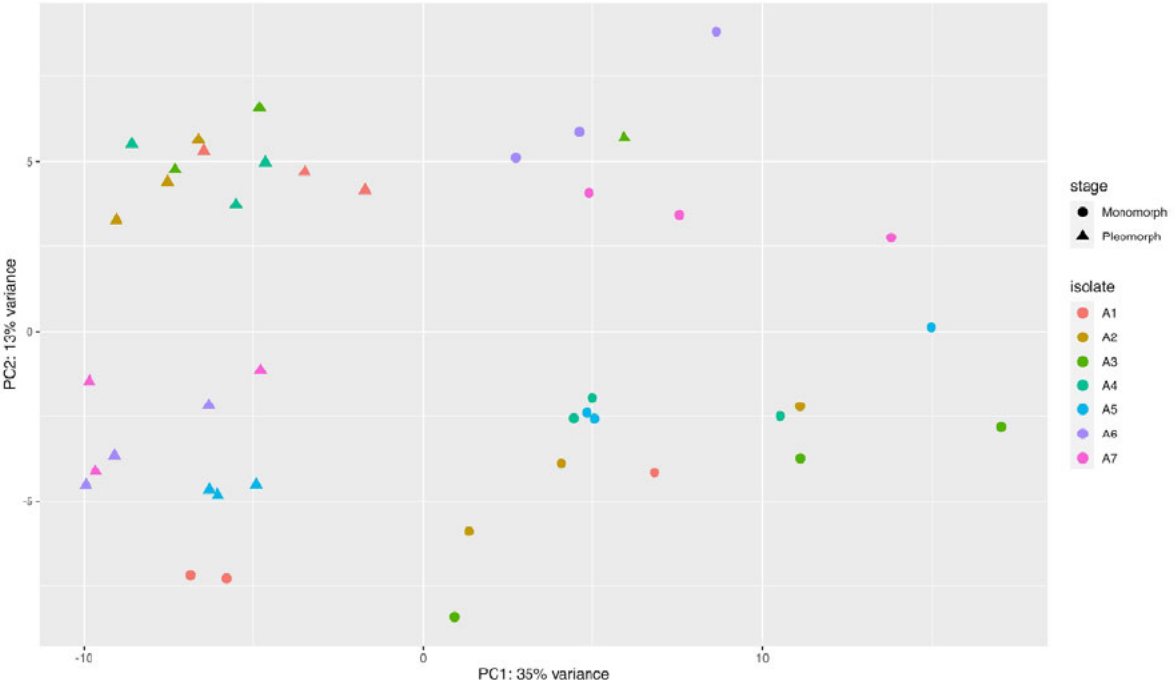


Figure S5.6: Principal component analysis of RNAseq libraries from clonal cell lines (A1-A7) selected for monomorphism in BHI-supplemented HMI-9. RNAseq was performed on three replicates of each clone at the start and end of the selection for monomorphism.

6 General discussion

6.1. Summary and future directions

6.1.1. Summary of aims

To identify the cause of developmental incompetence (monomorphism) in naturally occurring and laboratory-selected *T. brucei*, we proposed three aims;

1. Confirm the evolutionary relationships between naturally occurring monomorphic isolates and highlight features of monomorphism at the genome level.
2. Identify mutations in naturally occurring monomorphic *T. brucei* isolate, which could give rise to developmental incompetence, and validate the role of these mutations in the laboratory.
3. Identify the early transcriptional events in the generation of monomorphism, by selecting for monomorphism with a pleomorphic cell line. Confirm the role of these transcriptional modifications via manipulation of the target transcripts.

Our results highlight potential early and late events in the evolution of monomorphism, and how the complex QS pathway is constructed and perturbed, providing molecular predictors for increased virulence associated with monomorphic *T. brucei* in the field.

6.1.2. Aim 1: Monomorphic *T. brucei* form at least four independent clades

Developmentally incompetent monomorphic clades have arisen from developmentally competent *T. brucei* at least four times independently of one another (Cuypers et al., 2017;

Oldrieve et al., 2022). To fully acknowledge their polyphyletic origin, and distinct evolutionary histories, we suggest the taxonomy of monomorphic Trypanozoon should be based on whole genome analysis alone, which is quantitative, non-subjective and can be assessed from samples lacking detailed records. The disease manifestation and tissue specificity of the IVM-t1 isolate also suggest that *T. brucei evansi* type B or a fifth monomorphic clade can manifest dourine symptoms with sexual transmission, although direct transmission evidence is needed to confirm this.

The incongruity between the parasites' evolutionary position and the induced pathology, mode of transmission and tissue tropism highlights the potential for the host physiology and immune response to contribute to clinical disease manifestation, rather than being solely parasite-driven (Büscher et al., 2019). This provides an important example of the potential complexity regarding the described taxonomic position of monomorphic trypanosomes and the diseases they cause. Notably, the separate clades display lineage-specific genomic features whilst sharing a remarkably similar phenotype. The convergent phenotype can confuse attempts to study the basis of their developmental incompetence. Future studies will analyse the emergence of each monomorphic clade, using tools such as the Bayesian Evolutionary Analysis by Sampling Trees (BEAST) software package which can date phylogenetic events via analysis of sequence data (Suchard et al., 2018). Using Markov chain Monte Carlo integration, BEAST adds divergence time dating to phylogenetic reconstructions. Dating the emergence of each clade would allow us to confirm our hypothesis that *T. b. equiperdum* type OVI is the most ancient monomorphic clade and finalise the independent evolutionary origins of these monomorphic clades.

6.1.3. Aim 2: Monomorphic *T. brucei* have lineage-specific mutations which affect developmental competence, growth rate and motility

After confirming the relationships between each monomorphic clade, we identified distinct mutations in APPBP1 (Tb927.2.4020) and a hypothetical mitochondrial protein (Tb927.5.2580) which were present in independent naturally occurring monomorphic clades and hinder the generation of the stumpy form. We have also identified monomorphic clade-specific mutations in GPR89 (Tb927.8.1530) which reduces the growth rate and FAZ41 (Tb927.11.3400) which alters motility. A recent study highlighted a gene deletion and/or frameshift in PAG3 which corroborates our evidence for an accumulation of deleterious mutations in monomorphic Trypanozoon which affect developmental competence (Wen et al., 2022).

Despite the identification of these distinct mutations in different clades we are unable to confirm that these mutations were the initial cause of monomorphism. The mutation could have arisen after, before or during the transition to monomorphism. The disparate phenotypes we identified could be indicative of an accumulation of deleterious mutations which is often found in obligate asexual organisms (Jaron et al., 2021; Schwander et al., 2013). For the APPBP1 and hypothetical mitochondrial protein (Tb927.5.2580) experiments, we were unable to recreate the incompetent phenotype with a single mutation, instead, the mutations took effect when in combination with a background of mutations specific to the gene of interest in the monomorphic clade. In future, we will explore why the full complement of monomorphic mutations is required.

6.1.4. Aim 3: Monomorphism is initially generated via downregulation of key transcripts, such as ZC3H20, which can be reversed

We propose that in the field, the generation of monomorphism may be associated with the migration of infected mammals away from tsetse endemic regions, or an alteration in the range of the tsetse fly, which removes them from a geographic region containing infected mammals. In the absence of, or with reduced exposure to, the tsetse fly vector, reducing the control of parasitaemia would improve the opportunity, although not the efficiency, of mechanical transmission (Desquesnes et al., 2009). In this scenario, there would be strong selection pressure to suppress the expression of QS-associated genes, such as ZC3H20 and RBP10, which restrict parasitaemia at high population densities. Continual mechanical transmission, facilitated by the downregulation of QS genes in the absence of the tsetse vector, would begin to create a monomorphic (proto-monomorphic) phenotype which is potentially reversible in the first instance by restoration of the expression of the repressed genes. We have recreated this scenario through the selected monomorphic cell lines we generated in Chapter 5, where the re-expression of ZC3H20 restored stumpy formation. In future, re-adapting parasites to a cyclical life cycle, via continual passage through the tsetse fly vector, followed by monitoring of ZC3H20 expression levels would prove the true reversibility of the proto-monomorphic phenotype.

Proto-monomorphs could be obligately or facultatively asexual, depending upon whether they were transmitted through the insect vector. They would likely exist on a spectrum of developmental competence, displaying the ability to transmit through the tsetse, but a trend towards mechanical transmission. Indeed, mechanical transmission occurs even when tsetse flies are present, which increases the circulation of parasites (Desquesnes et al., 2009; Moloo et al., 2000). Furthermore, slender forms can generate a tsetse infection in laboratory studies (Schuster et al., 2021) at low efficiency (Matthews and Larcombe, 2022).

However, continual mechanical transmission, contributed to by the downregulation of QS genes and asexual reproduction, could thereafter lead to the accumulation of mutations in QS genes and the loss or removal of kDNA (Dewar et al., 2018), locking the proto-monomorphs into a fully monomorphic phenotype. Mechanical transmission would be advantageous in the short run, providing an abundance of new geographical and host niches. Muller's ratchet suggests that the accumulation of deleterious mutations, with an inability to reverse these mutations through recombination in the obligate asexual population, could eventually doom the monomorphic clades to extinction. This prediction is supported by modelling the population dynamics of asexual parasites. Howard and Lively, (2002) showed that a parasite population with high levels of virulence and transmission will endure population bottlenecks during host-parasite coevolution. The rapid accumulation of mutations, in the frequently bottlenecked asexual population, led to rapid elimination of the asexual lineage (Howard and Lively, 2002). These findings could indicate why sexual reproduction is common amongst parasitic organisms.

When we selected monomorphic *T. brucei* via serial passage and then tested their developmental competencies *in vitro*, we found that insensitivity to the oligopeptide signal was observed after just six passages (P6), even when initiated with clonal parasite lines. Thus, when P6 cells were used to infect three mice, the cells retained developmental competence but the population grew rapidly, potentially associated with a loosening of the QS density control mechanisms. This raises the question of whether at P6 the whole population gradually loosened the density control mechanisms or whether a small percentage of the population had rapidly become monomorphic. In the second scenario, this small percentage would then be selected over time to establish a monomorphic population. Future studies could address the heterogeneity of the population via scRNA-seq of P6. Sequencing RNA from 1000s of cells will enable us to identify any sub-populations of monomorphic cells within the population as a whole, which is impossible using traditional bulk RNAseq (Luzak et al., 2021). We will also extend our analysis of the selected

monomorphic cell lines to understand underlying molecular changes via analysis of transposase-accessible chromatin with ATAC-Seq which could give rise to the downregulation of transcripts during the selection for monomorphism (Buenrostro et al., 2013; Müller et al., 2018).

A comparison of molecular changes found in the field and laboratory monomorphic *T. brucei* might help identify common changes which were missed when the data were analysed separately. This would be performed by building a database from the results of this study, and publicly available data available on TriTrypDB (Aslett et al., 2010), for each gene in the TREU927/4 reference genome. Using machine learning, we could then analyse the database to predict genes which are associated with developmental competence. The analysis could be performed using the Python-based machine learning tool Scikit-learn (Pedregosa et al., 2012) and validated with known developmental regulators as a 'truth set' (Mony et al., 2014).

6.2. Implications

6.2.1. Costs and benefits of a complex life cycle

The transition from a complex to a simple life cycle is relevant to wildlife and livestock but also provides a fascinating opportunity to study the evolutionary consequences of life history alterations. Many species, from bacteria (Pal and Fikrig, 2003) to arthropods (Riley, 1983), have a complex life cycle that can require up to four obligate host species (Cribb et al., 2003). Complex life cycles might appear cumbersome and risky, with parasites relying upon a precise sequence of transmission events, however, they are abundant and present fascinating examples of evolutionary trade-offs.

Complex life cycles allow the parasite to separate growth and reproduction depending on its host environment (Auld and Tinsley, 2015). For example, parasites such as *Plasmodium spp.* and *Theileria spp.* reproduce sexually in their insect vector and asexually in their intermediate host (Roos, 2005). As the insect stage is predicted to be the ancestral stage of the parasite's life cycle (Flegontov et al., 2013), the expansion into the mammalian host likely increased the growth rate, prevalence (asexual reproduction) and potential for transmission that the parasite can achieve, whilst retaining sexual reproduction in the insect stage. This can be seen in *Plasmodium chabaudi chabaudi* which displays increased virulence when serially passaged without transferring through its insect vector. The virulence can be attenuated via transmission of the parasite back through its mosquito host (Spence et al., 2013). Although the parasite population grows within the insect host, sexual reproduction is absent from the mammalian host stage. Therefore, in the mammalian host, a parasite can focus on adaptations specific to growth.

Toxoplasma gondii normally establishes an infection in an intermediate vertebrate, where it asexually reproduces. Upon infection of a felid, *T. gondii* sexually reproduces. However, clades of *T. gondii* that are capable of sexual reproduction have forgone the felid host and therefore, sexual reproduction. These clonal clades can be transmitted orally and display increased virulence in humans (Su et al., 2003). It has been proposed that the drastically reduced genetic diversity of the clonal *T. gondii* clades arose due to a genetic sweep which fixed the loci of many genes via the hitchhiking effect (Smith and Haigh, 1974). The selective advantage of oral transmission ultimately led to alterations in the parasites' complex life cycles, drastically influencing the biology of the asexual *T. gondii* clades. However, the increased virulence seen in these clades, and serially passaged *P. c. chabaudi*, could be maladaptive in the long term as their increased virulence leads to a reduction in the host's lifespan and a reduction in the genetic diversity of the parasite population.

Genetic diversity directly affects the lag time in co-evolution between host and parasite and therefore the evolutionary potential of the parasite to genetically adapt to environmental pressure. The fungal parasite *Podosphaera plantaginis* employs both sexual and asexual reproduction to increase its chances of infecting its plant host, *Plantago lanceolata* (ribwort plantain). Areas where there are higher rates of coinfection, and therefore increased likelihood of genetic diversity, have an increased chance of parasite survival through winter (Laine et al., 2019). Pockets of diverse parasite populations emerged depending on each area's evolutionary rate, leading to local clusters of new genotypes (Thompson, 2005). Such new genotypes could prove costly if the genotype has a lower fitness. However, the parasites hedge their bets to create rare genotypes with the potential to thrive in stressful situations (Starrfelt and Kokko, 2012).

Asexual monomorphic *T. brucei* have forgone sexual reproduction, potentially reducing their ability to respond to future environmental stresses. In developmentally competent *T. brucei*, the parasite's complex life cycle facilitates attenuated virulence and the maintenance of

genetic diversity, monomorphic *T. brucei* have forgone these benefits to expand into new geographic and host niches.

6.2.2. Climate change and the expansion of monomorphism

Given the initial advantage of monomorphism, monomorphic *T. brucei* clades will likely continue to emerge. Indeed, increasing sampling efforts might identify new monomorphic clades already in circulation. We further suggest that the rapidly changing climate may facilitate an explosion of new monomorphic clades, as the range of the tsetse fly vector adapts to changing climatic norms (Longbottom et al., 2020; Lord et al., 2018). Thus, parasites in areas where tsetse was previously located, but have now been displaced due to a changing climate, can only persist through mechanical transmission. These monomorphic clades, which have lost or reduced their growth control mechanism, may both exhibit high virulence and pose the risk of further spread outside traditional disease boundaries. Escaping transmission by the tsetse fly facilitated the expansion of monomorphic trypanosomes to other regions of Africa, Asia, Europe and the Americas, although they have subsequently been eradicated from North America and limited to local outbreaks in Europe (Radwanska et al., 2018). A single disease outbreak of *T. vivax* in South America in 1995 was estimated to cause more than $160 million in damages to cattle and equids, highlighting the risk of the emergence of new strains of monomorphic *T. brucei* (Seidl et al., 1999). The risk is further exacerbated by the potential for human infective *T. b. rhodesiense*, which is zoonotic, spreading beyond the tsetse belt and at higher virulence.

6.2.3. Molecular tools to track monomorphism in the field

We have discussed that new monomorphic clades likely continually emerge from pleomorphic *T. brucei* in the field. To explore this, future studies could sample field isolates from areas where the tsetse fly range has recently altered. Should the area still contain mammals with *T. brucei* infections, these isolates may display proto-monomorphic traits. By sequencing the genomes of these organisms, or analysing their transcriptomes, we could build a database of early genomic events during the transition from proto-monomorph to monomorph which could be tracked. This would build on the detection of mutations, and transcriptional changes, we have identified in this study. Using miniaturised sequencing technologies, such as the minION nanopore, will allow us to do this in real-time in a low-resource field-based setting. The rapid improvement of Nanopore data quality means that it is not necessary to sequence the genome using short-read technology and that adaptive genome sequencing might allow for the enrichment of target parasite DNA by depleting the host background (Marquet et al., 2022). This approach could allow for sampling directly from host tissue/blood rather than requiring resource-intense culture facilities, avoiding the pitfalls of methods currently suggested by the World Organisation for Animal Health (discussed in detail in Chapter 2).

6.3. Concluding remarks

We have highlighted immediate and long-term alterations that occur during the generation of developmental incompetence in *T. brucei,* providing new insight into *T. brucei* developmental biology*,* and how it can be disrupted. Developmentally incompetent *T. brucei* present a fascinating system to study the evolutionary consequences of transitions between complex and simple lifecycles, and the associated loss of sexual reproduction. Furthermore, we have highlighted the threat of new developmentally incompetent lineages, whilst providing molecular predictors to help identify and understand these isolates in the field.

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

7.1. Preface

Trypanosoma theileri, a non-pathogenic trypanosome which infects cattle and is transmitted by tabanid flies is closely related to *Trypanosoma melophagium*, which infects sheep and is transmitted by the sheep ked. The *T. theileri* genome was published in 2017 and we sequenced and assembled the *T. melophagium* genome to identify the genomic basis of the host and vector specificity exhibited by these closely related species (Oldrieve et al., 2022). Whilst our ultimate motive was to answer this fascinating biological question, the project could be completed almost entirely from home (at a time when laboratories were closed) and involved learning new techniques, such as using the portable minION genome sequencer. The project also provided an opportunity to gain supervision experience, as the *T. melophagium* genome was assembled with the help of Beatrice Malacart, an undergraduate student in the Matthews lab and a co-author on the following publication (Oldrieve, 2022). Our orthologous protein clustering results have been used to compare the conservation of cell cycle-regulated genes across trypanosomatids in a recent scRNA-Seq analysis (Briggs et al., 2023).

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

© 2022. Published by The Company of Biologists Ltd | Biology Open (2022) 11, bio059237. doi:10.1242/bio.059237

RESEARCH ARTICLE

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 Trypanosomiases (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 Molyneux, 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; Pays 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 leishmanolysin (Yao, 2010).

Studies have largely focused on pathogenic dixenous trypanosomatids, which can hold broad host niches, capable of

Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, UK.

*Author for correspondence (guy.oldrieve@ed.ac.uk)

G.R.O., 0000-0003-1428-0608; B.M., 0000-0002-7705-1506; J.L., 0000-0002-0144-2316; K.R.M., 0000-0003-0309-9184

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Received 19 January 2022; Accepted 25 March 2022

Biology Open (2022) 11, bio059237. doi:10.1242/bio.059237

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 parasitaemia 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; Seifi, 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 grayi* (Kelly et al., 2014, 2017), distinct from African trypanosomes, such as *T. brucei*. *T. grayi* 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 trypanosome 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; Martinković 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 (Martinković et al., 2012). Other surveys via blood culture found 7.8% of sheep to be infected with *T. melophagium* (Serpil and Zafer, 2008). It has historically been argued that *T. melophagium* is a monoxenous parasite of the sheep ked and that the mammalian host is obsolete for transmission (Flu, 1908; Porter; Swingle, 1911). However, extensive studies demonstrated a mammalian host is required (Hoare, 1923). Experimental infections of sheep with *T. melophagium* suggested the longest infection lasts 3 months and there is no lasting immunity as sheep can be reinfected with *T. melophagium* after several months of isolation (Gibson et al., 2010; Hoare, 1972)

Molecular markers place *T. melophagium* as a close relative to *T. theileri* (Martinković et al., 2012). SSU rRNA shares ∼98% identity between *T. theileri* and *T. melophagium* isolates (Gibson et al., 2010). Presumably the divergence of *T. theileri* and *T. melophagium* is associated with their discrete host niches (Gibson et al., 2010; Martinković 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;

Table 1. Genome assessment

	T. melophagium	*T. theileri* (32)
k-mer based genome survey		
Heterozygosity (%)	0.30	0.41
Genome length (Mbp) / Repeat / Unique	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
Minimium / maximium / mean (bp)	5,186 / 1,230,212 / 364,120	791 / 1,635,300 / 117,005
N50	505,851	517,122
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.

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

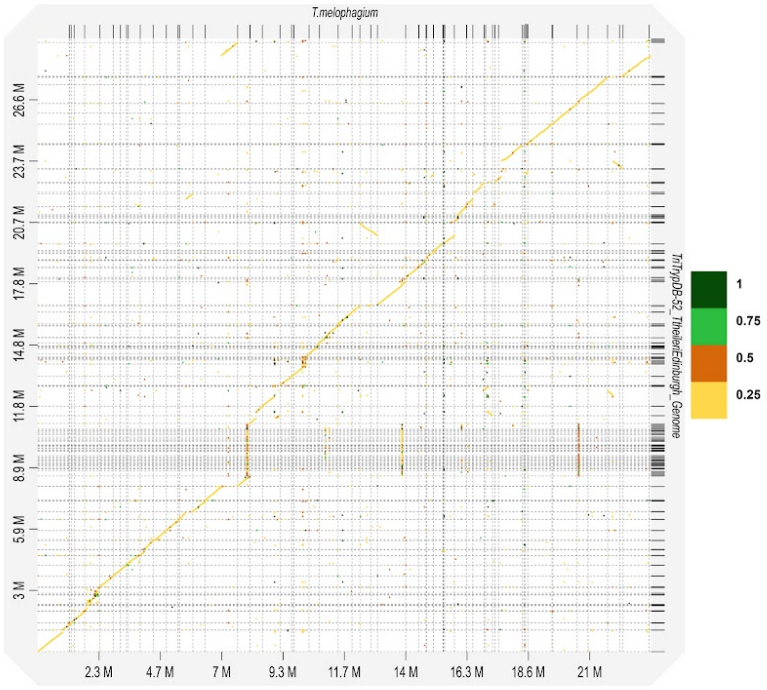


Fig. 1. Synteny of the *T. melophagium* and *T. theileri* genome sequences highlights conservation and identity between the two species. The legend refers to the percentage identity between the sequences.

Biology Open (2022) 11, bio059237. doi:10.1242/bio.059237

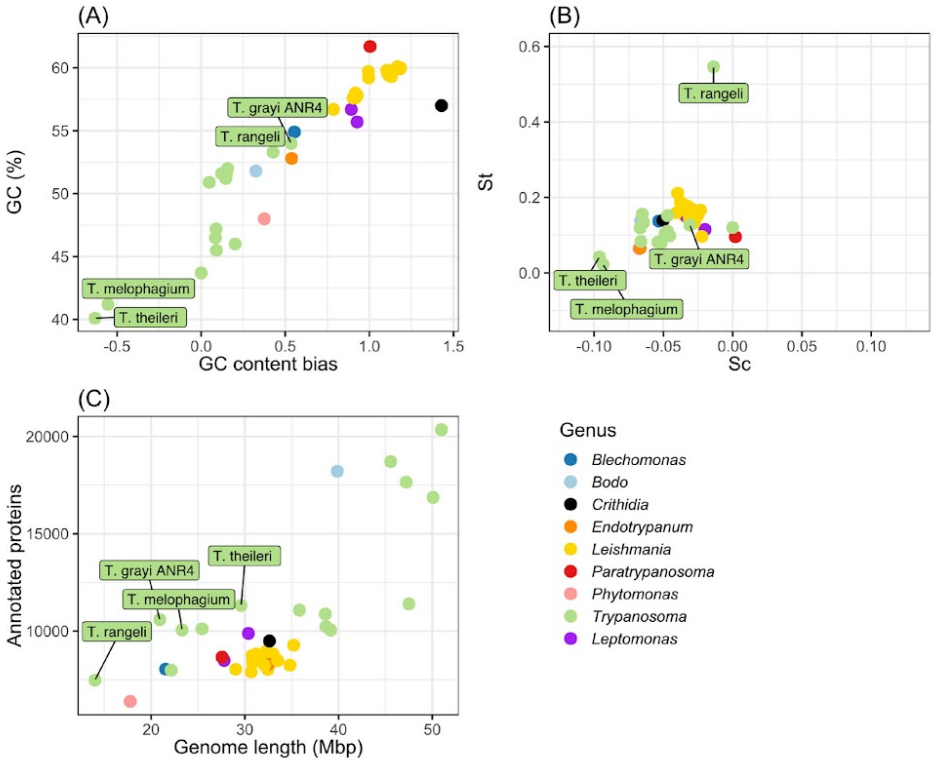


Fig. 2. Kinetoplastid genome comparison. (A) GC content across the whole genome and GC content bias in the CDS of kinetoplastid universal single copy orthologues (*n*=992). GC content (GC)>0=GC content bias. GC <0=AT content bias. (B) Selection acting on translational efficiency (St) and selection acting 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.

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; Martinković 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

Biology Open (2022) 11, bio059237. doi:10.1242/bio.059237

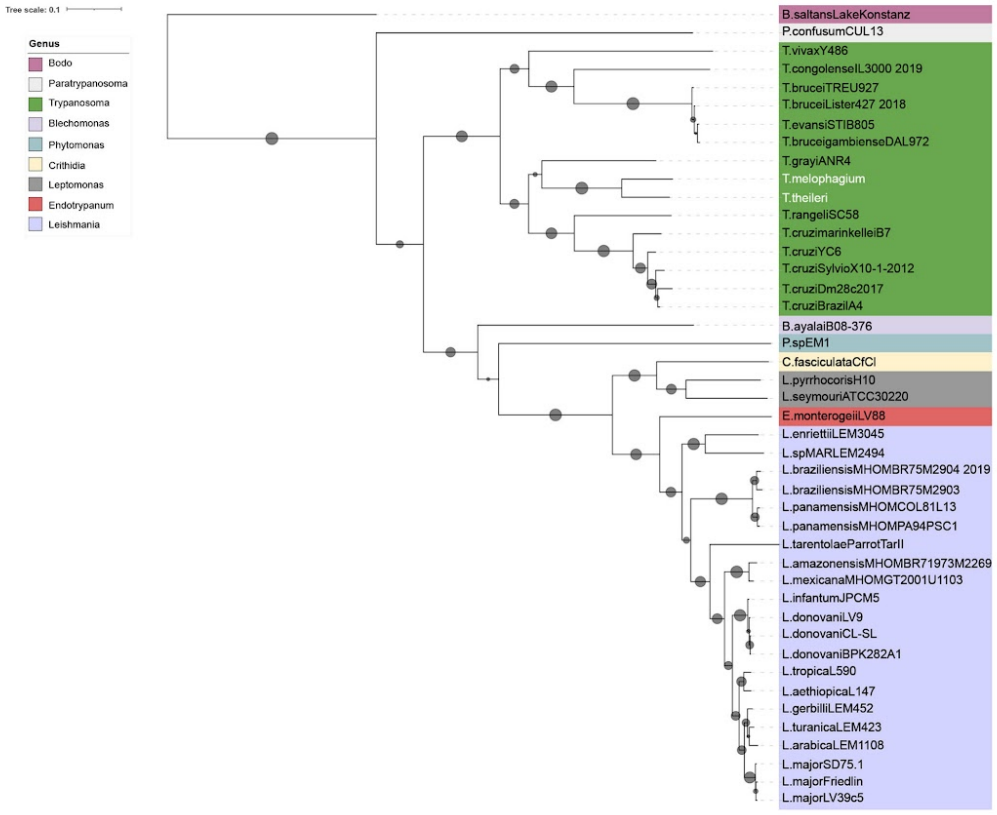


Fig. 3. Species consensus tree based on 2,312 gene trees created by STAG and STRIDE, OrthoFinder. The support values are represented by circles. Support values correlate to the proportion of times that the bipartition is seen in each of the individual trees used to create the consensus tree. The scale represents substitutions per site.

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

Table 2. Orthologous protein clustering statistics of *T. melophagium* and *T. theileri*

	T. melophagium	*T. theileri*
Number of genes	10,057	11,312
Genes in orthogroups (%)	97.3	97.4
Orthogroups containing species (%)	43.9	44.2
Number of species-specific orthogroups	76	121
Genes in species-specific orthogroups (%)	2.7	12.9
Comparatively expanded orthogroups	787	563

The full species-specific clustering summary can be found in File S3.

Biology Open (2022) 11, bio059237. doi:10.1242/bio.059237

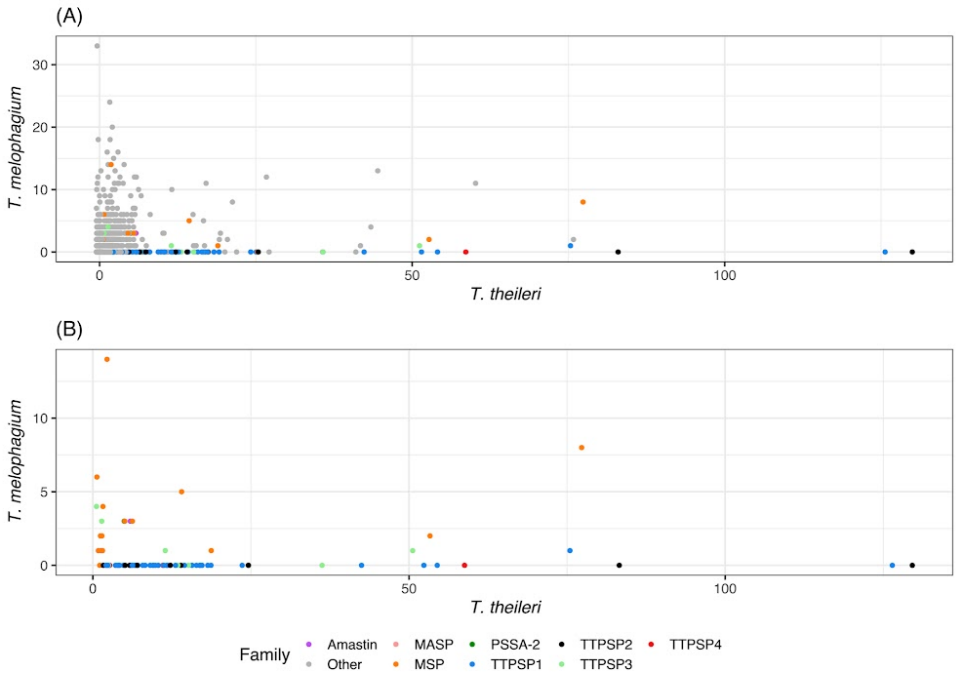


Fig. 4. (A) All orthogroups and (B) orthogroups associated with host interaction size comparison between *T. melophagium* and *T. theileri*. Each dot represents the numbers of genes found in each orthogroup for both species. The orthogroups have been annotated with their designation as either a putative cell surface protein family or 'other' (Kelly et al., 2017).

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

		Orthogroup count		Gene count	
Cell surface conservation	Annotation	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 orthogroups. Of the nine remaining annotated orthogroups, OG0008096 is the largest and contains trans-sialidsases. The orthogroup is expanded in *T. theileri* (n=17) compared to *T. melophagium* (n=5).

T. theileri has 121 species specific orthogroups. Sixty-four of these were putative host interaction genes (Kelly et al., 2017). Two annotated orthogroups remained, which included a leucine-rich repeat family associated with protein binding (OG0011868) and a calpain cysteine protease family (OG0018204). *T. melophagium* has 76 species-specific orthogroups, 68 of these were unannotated. MSP families were annotated in three of the remaining orthogroups (OG0009903, OG0013746 and OG0013747). Other annotated species-specific orthogroups consisted of an actin family (OG0018163) along with several families associated with protein binding, WD domain G-beta repeat (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 so are likely to have been lost by *T. melophagium* and *T. grayi*. In *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 *Phytomonas* (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 glycosomal 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; Ling et al., 2011; Liu 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 et al., 1994; Gale and Parsons, 1993; Jones et al., 2014 a; Müller et al., 2002; Szöőr 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 KRIPP14, 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 SPO11, MND1, 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; Martinković et al., 2012). Here, the genome of *T. melophagium* was sequenced

Biology Open (2022) 11, bio059237. doi:10.1242/bio.059237

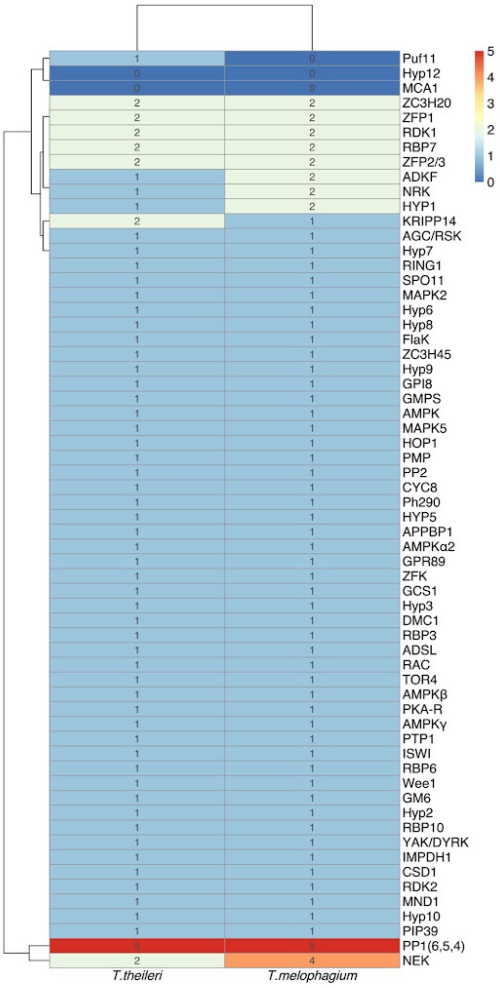


Fig. 5. Genes associated with development, and related proteins, at various stages throughout the *T. brucei* life cycle. The number of genes in orthogroups associated with developmental regulation have been quantified in *T. theileri* and *T. melophagium*.

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.

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 confusum* 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

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 (Serpil 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 glycolysis 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 also a stercorarian trypanosome, but which lacks the requisite molecular machinery (Ullu 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 DNBseq (4.201Gb, 150 base pair reads). RNA was extracted with the RNeasy mini kit (Qiagen) including a DNAse step, following the manufacturer's instructions and sequenced with BGI's DNBseq (5.019Gb, 100 bp reads). Raw DNA and RNA DNBseq 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 Wtdbg2 draft assembly and three iterations of Racon (Vaser et al., 2017) followed by one round of Medaka (available at https://github.com/nanoporetech/medaka) were performed. DNBseq reads were mapped to the Medaka polished assembly, and two iterations of Racon were performed. Short and long reads were aligned to the Racon polished assembly to complete two final iterations of polishing with Pilon (Walker et al., 2014). At each stage of polishing, the quality of the draft genome was assessed using scaffold_stats.pl (available at https://github.com/sujaikumar/assemblage) and BUSCO (Seppey et al., 2019). BUSCO provides a metric for genome assembly and annotation completeness, based on the presence of near-universal single-copy orthologues in a genome assembly or the corresponding annotated proteins.

Both sets of reads were mapped to the draft assembly. Each contig was subjected to a DIAMOND blastx search against the InterProScan database (Buchfink et al., 2015; Jones et al., 2014b). The resulting alignments and DIAMOND hits were visualised with BlobTools (Laetsch and Blaxter, 2017b). Every contig had a DIAMOND hit to sequences from the *Trypanosoma* genus. Therefore, the assembly was confirmed to be free from contamination. However, two contigs were outliers in comparison to the rest of the assembly at under 100x coverage and consisting of only 3,975 and 1,892 base pairs (Fig. S2A). These contigs were removed from the assembly. BUSCO and BlobTools were re-run on this trimmed assembly to assess the final assembly's completeness and coverage (Fig. S2B).

Repeat sequences in the genome were identified and soft-masked using RepeatModeler2 and RepeatMasker (Flynn et al., 2020). BRAKER2 was used to annotate the soft-masked genome in ETP mode. The OrthoDB v10 protozoa database was utilised for protein hints (Kriventseva et al., 2019) with the addition of the *T. theileri* proteome and *T. melophagium* RNAseq evidence (Barnett et al., 2011; Brůna et al., 2021, 2020; Buchfink et al., 2015; Gotoh, 2008; Hoff et al., 2016, 2019; Iwata and Gotoh, 2012; Li et al., 2009; Lomsadze et al., 2014; Stanke et al., 2008, 2006). BRAKER2 produced the protein and transcript files used in the following analysis. The *T. melophagium* proteome was functionally annotated using InterProScan (Jones et al., 2014b) using the Pfam and SignalP databases (Mistry et al., 2021; Nielsen, 2017). Genome conservation of collinearity was compared using D-Genies (Cabanettes and Klopp, 2018).

Publicly available genomes, transcriptomes and proteomes were accessed from TriTrypDB (Aslett et al., 2010) along with the *Phytomonas* EM1 assembly which was accessed from NCBI (Leinonen et al., 2011). The quality of the proteomes were assessed using BUSCO; only isolates which had >85% complete proteomes were included, which left 43 isolates along with *T. melophagium*. A list of all the isolates used in this study can be found in File S1. The assembly statistics of the 44 genomes were assessed using scaffold_stats.pl (available at https://github.com/sujaikumar/assemblage).

The genomes from each of these kinetoplastid isolates were screened for transfer RNA genes using tRNAscan-SE (Chan et al., 2021). The outputs of these results were used to infer the strength of selection acting on translational efficiency and nucleotide cost for each isolate, along with the background mutation bias, using CodonMuSe (Seward and Kelly, 2016, 2018). Each isolate was assessed using (1) every CDS, (2) single-copy universal orthologs (identified in the orthologous protein clustering steps below) (n=992) and 3) single-copy universal orthologs, which are essential for every stage of the *T. brucei* life cycle (n=158). The last list of genes were identified by screening the universal single-copy orthologs for genes that had a significant reduction in transcript levels in every library of an RNAi phenotype screen (>1.5 log fold decrease) (Alsford et al., 2011).

Each kinetoplastid genome was also screened for retrotransposons and long terminal repeat (LTR) retrotransposons using TransposonPSI (available at http://transposonpsi.sourceforge.net) and LTR-harvest (Ellinghaus et al., 2008), respectively.

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 absence 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-25).

The cell surface orthogroups 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

Conceptualization: G.O., 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., Hertz-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

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, 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/btr174

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, 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

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. (2002). Antigenic variation and allelic exclusion. *Cell* **109**, 5-8. doi:10.1016/S0092-8674(02)00711-0
Borst, P. and Fase-Fowler, F. (1979). The maxi-circle of Trypanosoma brucei kinetoplast DNA. *Biochim. Biophys. Acta* **565**, 1-12. doi:10.1016/0005-2787(79)90078-9
Böse, R. and Heister, N. C. (1993). Development of Trypanosoma (M.) theileri in tabanids. *J. Eukaryot. Microbiol.* **40**, 788-792. doi:10.1111/j.1550-7408.1993.tb04475.x
Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108. doi:10.1093/nargab/lqaa108
Brůna, T., Lomsadze, A. and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform* **2**, lqaa026. doi:10.1093/nargab/lqaa026
Buchfink, B., Xie, C. and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59-60. doi:10.1038/nmeth.3176
Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016147. doi:10.1101/cshperspect.a016147
Buscaglia, C. A., Campo, V. A., Frasch, A. C. and Di Noia, J. M. (2006). Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat. Rev. Microbiol.* **4**, 229-236. doi:10.1038/nrmicro1351
Cabanettes, F. and Klopp, C. (2018). D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958. doi:10.7717/peerj.4958
Cavalier-Smith, T. (2010). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol. Lett.* **6**, 342-345. doi:10.1098/rsbl.2009.0948
Cayla, M., Rojas, F., Silvester, E., Venter, F. and Matthews, K. R. (2019). African trypanosomes. *Parasit. Vectors* **12**, 190. doi:10.1186/s13071-019-3355-5
Cayla, M., McDonald, L., MacGregor, P. and Matthews, K. (2020). An atypical DYRK kinase connects quorum-sensing with posttranscriptional gene regulation in Trypanosoma brucei. *Elife* **9**, e51620. doi:10.7554/eLife.51620
Chainey, J. E. (1993). Horse-flies, deer-flies and clegs (Tabanidae). In *Medical Insects and Arachnids* (ed. R. P. Lane and R. W. Crosskey), pp. 310-332. Dordrecht: Springer Netherlands.
Chan, P. P., Lin, B. Y., Mak, A. J. and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077-9096. doi:10.1093/nar/gkab688
Clayton, C. (2019). Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.* **9**, 190072. doi:10.1098/rsob.190072
Cooper, D. and Eleftherianos, I. (2017). Memory and specificity in the insect immune system: current perspectives and future challenges. *Front Immunol* **8**, 539. doi:10.3389/fimmu.2017.00539
Daniels, J. P., Gull, K. and Wickstead, B. (2010). Cell biology of the trypanosome genome. *Microbiol. Mol. Biol. Rev.* **74**, 552-569. doi:10.1128/MMBR.00024-10
Doherty, M. L., Windle, H., Voorheis, H. P., Larkin, H., Casey, M., Clery, D. and Murray, M. (1993). Clinical disease associated with Trypanosoma theileri infection in a calf in Ireland. *Vet. Rec.* **132**, 653-656. doi:10.1136/vr.132.26.653
Domingo-Sananes, M. R., Szöőr, B., Ferguson, M. A., Urbaniak, M. D. and Matthews, K. R. (2015). Molecular control of irreversible bistability during trypanosome developmental commitment. *J. Cell Biol.* **211**, 455-468. doi:10.1083/jcb.201506114
Ellinghaus, D., Kurtz, S. and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18. doi:10.1186/1471-2105-9-18
Emms, D. M. and Kelly, S. (2017). STRIDE: species tree root inference from gene duplication events. *Mol. Biol. Evol.* **34**, 3267-3278. doi:10.1093/molbev/msx259
Emms, D. M. and Kelly, S. (2018). STAG: Species tree inference from all genes. *BioRxiv*.
Emms, D. M. and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. doi:10.1186/s13059-019-1832-y
Emms, D. M. and Kelly, S. (2020). Benchmarking orthogroup inference accuracy: revisiting orthobench. *Genome Biol. Evol.* **12**, 2258-2266. doi:10.1093/gbe/evaa211
Erben, E. D., Fadda, A., Lueong, S., Hoheisel, J. D. and Clayton, C. (2014). A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. *PLoS Pathog.* **10**, e1004178. doi:10.1371/journal.ppat.1004178
Eyrewalker, A. C. (1991). An analysis of codon usage in mammals - selection or mutation bias. *J. Mol. Evol.* **33**, 442-449. doi:10.1007/BF02103136
Farrar, R. G. and Klei, T. R. (1990). Prevalence of Trypanosoma theileri in Louisiana cattle. *J. Parasitol.* **76**, 734-736. doi:10.2307/3282992
Flu, P. C. (1908). Über die Flagellaten im Darm von Melophagus ovinus. *Arch. f. Protist.* **12**, 147-153.
Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C. and Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**, 9451-9457. doi:10.1073/pnas.1921046117
Funk, S., Nishiura, H., Heesterbeek, H., Edmunds, W. J. and Checchi, F. (2013). Identifying transmission cycles at the human-animal interface: the role of animal reservoirs in maintaining gambiense human african trypanosomiasis. *PLoS Comput. Biol.* **9**, e1002855. doi:10.1371/journal.pcbi.1002855
Gale, M., Jr, Carter, V. and Parsons, M. (1994). Translational control mediates the developmental regulation of the Trypanosoma brucei Nrk protein kinase. *J. Biol. Chem.* **269**, 31659-31665. doi:10.1016/S0021-9258(18)31746-0
Gale, M., Jr and Parsons, M. (1993). A Trypanosoma brucei gene family encoding protein kinases with catalytic domains structurally related to Nek1 and NIMA. *Mol. Biochem. Parasitol.* **59**, 111-121. doi:10.1016/0166-6851(93)90012-M
Gibson, W., Pilkington, J. G. and Pemberton, J. M. (2010). Trypanosoma melophagium from the sheep ked Melophagus ovinus on the island of St Kilda. *Parasitology* **137**, 1799-1804. doi:10.1017/S0031182010000752
Gotoh, O. (2008). A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res.* **36**, 2630-2638. doi:10.1093/nar/gkn105
Hoare, C. A. (1923). An experimental study of the sheep-trypanosome (T. melophagium Flu, 1908), and its transmission by the sheep-ked (Melophagus ovinus L.). *Parasitology* **15**, 365-424. doi:10.1017/S0031182000014888
Hoare, C. A. (1929). Studies on Trypanosoma grayi. II. Experimental transmission to the crocodile. *Trans. R. Soc. Trop. Med. Hyg.* **23**, 39-56. doi:10.1016/S0035-9203(29)90831-2
Hoare, C. A. (1931). Studies on Trypanosoma grayi. III. Life-cycle in the tsetse-fly and in the crocodile. *Parasitology* **23**, 449-484. doi:10.1017/S0031182000013858
Hoare, C. A. (1972). *The Trypanosomes of Mammals. A Zoological Monograph*. Oxford, England: Blackwell.
Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769. doi:10.1093/bioinformatics/btv661
Hoff, K. J., Lomsadze, A., Borodovsky, M. and Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65-95. doi:10.1007/978-1-4939-9173-0_5
Hu, H., Gao, J., He, J., Yu, B., Zheng, P., Huang, Z. Q., Mao, X. B., Yu, J., Han, G. Q. and Chen, D. W. (2013). Codon optimization significantly improves the expression level of a keratinase gene in pichia pastoris. *PLoS One* **8**, e58393. doi:10.1371/journal.pone.0058393
Iwata, H. and Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* **40**, e161. doi:10.1093/nar/gks708
Jaskowska, E., Butler, C., Preston, G. and Kelly, S. (2015). Phytomonas: trypanosomatids adapted to plant environments. *PLoS Pathog.* **11**, e1004484. doi:10.1371/journal.ppat.1004484
Jones, N. G., Thomas, E. B., Brown, E., Dickens, N. J., Hammarton, T. C. and Mottram, J. C. (2014a). Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. *PLoS Pathog.* **10**, e1003886. doi:10.1371/journal.ppat.1003886
Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. et al. (2014b). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240. doi:10.1093/bioinformatics/btu031
Kaufer, A., Ellis, J., Stark, D. and Barratt, J. (2017). The evolution of trypanosomatid taxonomy. *Parasit. Vectors* **10**, 287. doi:10.1186/s13071-017-2204-7
Kelly, S., Ivens, A., Manna, P. T., Gibson, W. and Field, M. C. (2014). A draft genome for the African crocodilian trypanosome Trypanosoma grayi. *Sci. Data* **1**, 140024. doi:10.1038/sdata.2014.24
Kelly, S., Ivens, A., Mott, G. A., O'Neill, E., Emms, D., Macleod, O., Voorheis, P., Tyler, K., Clark, M., Matthews, J. et al. (2017). An alternative strategy for trypanosome survival in the mammalian bloodstream revealed through genome and transcriptome analysis of the ubiquitous bovine parasite trypanosoma (Megatrypanum) theileri. *Genome Biol. Evol.* **9**, 2093-2109. doi:10.1093/gbe/evx152
Kleisen, C. M. and Borst, P. (1975). Sequence heterogeneity of the mini-circles of kinetoplast DNA of Crithidia luciliae and evidence for the presence of a component more complex than mini-circle DNA in the kinetoplast network. *Biochim. Biophys. Acta.* **407**, 473-478. doi:10.1016/0005-2787(75)90301-9
Kniepert, F. W. (1980). Blood-feeding and nectar-feeding in adult tabanidae (Diptera). *Oecologia* **46**, 125-129. doi:10.1007/BF00346976
Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A. and Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807-D811. doi:10.1093/nar/gky1053
Laetsch, D. R. and Blaxter, M. L. (2017a). KinFin: software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda)* **7**, 3349-3357. doi:10.1534/g3.117.300233

Laetsch, D. R. and Blaxter, M. L. (2017b). BlobTools: interrogation of genome assemblies. *F1000Research* **6**, 1287. doi:10.12688/f1000research.12232.1
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.10.2.228
Leger, A. and Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J. Open Source Softw.* **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 phylogenetic 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, TbZC3H20, stabilizes two developmentally regulated mRNAs in trypanosomes. *J. Biol. Chem.* **286**, 20152-20162. doi:10.1074/jbc.M110.139261
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. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119. doi:10.1093/nar/gku557
Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D. and Erben, E. D. (2016). Gene expression regulatory networks in Trypanosoma brucei: insights into the role of the mRNA-binding proteome. *Mol. Microbiol.* **100**, 457-471. doi:10.1111/mmi.13328
Marçais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770. doi:10.1093/bioinformatics/btr011
Martinković, F., Matanović, K., Rodrigues, A. C., Garcia, H. A. and Teixeira, M. M. (2012). Trypanosoma (Megatrypanum) melophagium in the sheep ked Melophagus ovinus from organic farms in Croatia: phylogenetic inferences support restriction to sheep and sheep keds and close relationship with trypanosomes from other ruminant species. *J. Eukaryot. Microbiol.* **59**, 134-144. doi:10.1111/j.1550-7408.2011.00599.x
Maslov, D. A., Opperdoes, F. R., Kostygov, A. Y., Hashimi, H., Lukeš, J. and Yurchenko, V. (2019). Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology* **146**, 1-27. doi:10.1017/S0031182018000951
Matthews, D. M., Kingston, N., Maki, L. and Nelms, G. (1979). Trypanosoma theileri Laveran, 1902, in Wyoming cattle. *Am. J. Vet. Res.* **40**, 623-629.
Mehlitz, D. and Molyneux, D. H. (2019). The elimination of Trypanosoma brucei gambiense? Challenges of reservoir hosts and transmission cycles: expect the unexpected. *Parasite Epidemiol Control* **6**, e00113. doi:10.1016/j.parepi.2019.e00113
Melville, S. E., Leech, V., Gerrard, C. S., Tait, A. and Blackwell, J. M. (1998). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei and the assignment of chromosome markers. *Mol. Biochem. Parasitol.* **94**, 155-173. doi:10.1016/S0166-6851(98)00054-1
Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J. et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412-D419. doi:10.1093/nar/gkaa913
Mony, B. M. and Matthews, K. R. (2015). Assembling the components of the quorum sensing pathway in African trypanosomes. *Mol. Microbiol.* **96**, 220-232. doi:10.1111/mmi.12949
Mony, B. M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D. and Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. *Nature* **505**, 681-685. doi:10.1038/nature12864
Mott, G. A., Wilson, R., Fernando, A., Robinson, A., MacGregor, P., Kennedy, D., Schaap, D., Matthews, J. B. and Matthews, K. R. (2011). Targeting cattle-borne zoonoses and cattle pathogens using a novel trypanosomatid-based delivery system. *PLoS Pathog.* **7**, e1002340. doi:10.1371/journal.ppat.1002340
Müller, I. B., Domenicali-Pfister, D., Roditi, I. and Vassella, E. (2002). Stage-specific requirement of a mitogen-activated protein kinase by Trypanosoma brucei. *Mol. Biol. Cell* **13**, 3787-3799. doi:10.1091/mbc.e02-02-0093
Nardy, A. F., Freire-de-Lima, C. G., Pérez, A. R. and Morrot, A. (2016). Role of trypanosoma cruzi trans-sialidase on the escape from host immune surveillance. *Front. Microbiol.* **7**, 348. doi:10.3389/fmicb.2016.00348
Nielsen, H. (2017). Predicting secretory proteins with SignalP. *Methods Mol. Biol.* **1611**, 59-73. doi:10.1007/978-1-4939-7015-5_6
Novoa, E. M. and de Pouplana, L. R. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* **28**, 574-581. doi:10.1016/j.tig.2012.07.006
Oldrieve, G., Verney, M., Jaron, K. S., Hebert, L. and Matthews, K. R. (2021). Monomorphic Trypanozoon: towards reconciling phylogeny and pathologies. *Microb. Genom.* **7**, 000632. doi:10.1099/mgen.0.000632
Parsons, M., Nelson, R. G., Watkins, K. P. and Agabian, N. (1984). Trypanosome mRNAs share a common 5′ spliced leader sequence. *Cell* **38**, 309-316. doi:10.1016/0092-8674(84)90552-X
Pays, E., Vanhamme, L. and Pérez-Morga, D. (2004). Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries. *Curr. Opin. Microbiol.* **7**, 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 polypurine tracts in mRNA sequences of thermophilic prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2951-2956. doi:10.1073/pnas.0308594100
Peacock, L., Kay, C., Farren, 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
Plotkin, 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**, 12588-12591. 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 genomic 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 Sci.* **218**, 189-224. doi:10.1242/jcs.s2-55.218.189
Rao, Y. S., Wu, G. Z., Wang, Z. F., Chai, X. W., Nie, Q. H. and Zhang, X. Q. (2011). Mutation bias is the driving force of codon usage in the Gallus gallus genome. *DNA Res.* **18**, 499-512. doi:10.1093/dnares/dsr035
Ruan, J. and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155-158. doi:10.1038/s41592-019-0669-3
Sanchez-Moreno, M., Lasztity, D., Coppens, I. and Opperdoes, F. R. (1992). Characterization of carbohydrate metabolism and demonstration of glycosomes in a Phytomonas sp. isolated from Euphorbia characias. *Mol. Biochem. Parasitol.* **54**, 185-199. doi:10.1016/0166-6851(92)90111-V
Schlafer, D. H. (1979). Trypanosoma theileri: a literature review and report of incidence in New York cattle. *Cornell Vet* **69**, 411-425.
Schnaufer, A., Domingo, G. J. and Stuart, K. (2002). Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *Int. J. Parasitol.* **32**, 1071-1084. doi:10.1016/S0020-7519(02)00020-6
Seifi, H. A. (1995). Clinical trypanosomosis due to trypanosoma-theileri in a cow in iran. *Trop. Anim. Health Prod.* **27**, 93-94. doi:10.1007/BF02236319
Seppey, M., Manni, M. and Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227-245. doi:10.1007/978-1-4939-9173-0_14
Serpil, N. and Zafer, K. (2008). Trypanosoma melophagium in blood cell culture. *Ank. Üniv. Vet. Fak. Derg.* **50**, 1-1. doi:10.1501/Vetfak_0000000325
Seward, E. A. and Kelly, S. (2016). Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome Biol.* **17**, 226. doi:10.1186/s13059-016-1087-9
Seward, E. A. and Kelly, S. (2018). Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria. *Genome Biol.* **19**, 102. doi:10.1186/s13059-018-1480-7
Shah, P. and Gilchrist, M. A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10231-10236. doi:10.1073/pnas.1016719108
Simpson, L. (1979). Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. *Proc. Natl. Acad. Sci. USA* **76**, 1585-1588. doi:10.1073/pnas.76.4.1585
Sorensen, M. A., Kurland, C. G. and Pedersen, S. (1989). Codon usage determines translation rate in escherichia-coli. *J. Mol. Biol.* **207**, 365-377. doi:10.1016/0022-2836(89)90260-X
Stanke, M., Schoffmann, O., Morgenstern, B. and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62. doi:10.1186/1471-2105-7-62
Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637-644. doi:10.1093/bioinformatics/btn013
Steinert, M. (1960). Mitochondria associated with the kinetonucleus of trypanosoma mega. *J. Biophys. Biochem. Cytol.* **8**, 542-546. doi:10.1083/jcb.8.2.542

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

Szöőr, 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

Szöőr, 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., Blauwiekel, 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

Vurture, 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., Sakthikumar, 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

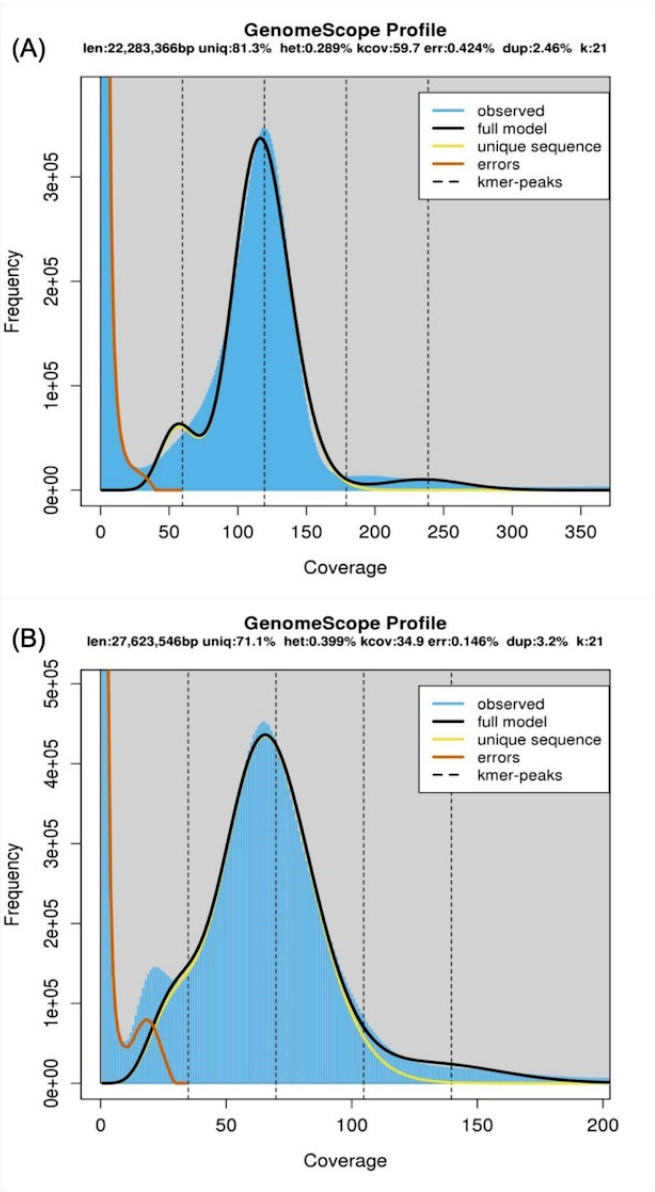


Fig. S1. k-mer spectra of (A) *T. melophagium* and (B) *T. theileri* using a k-mer size of 21. The statistics presented in these images represent the lowest estimate size, the highest estimate was used in Table 1 as these had a greater model fit.

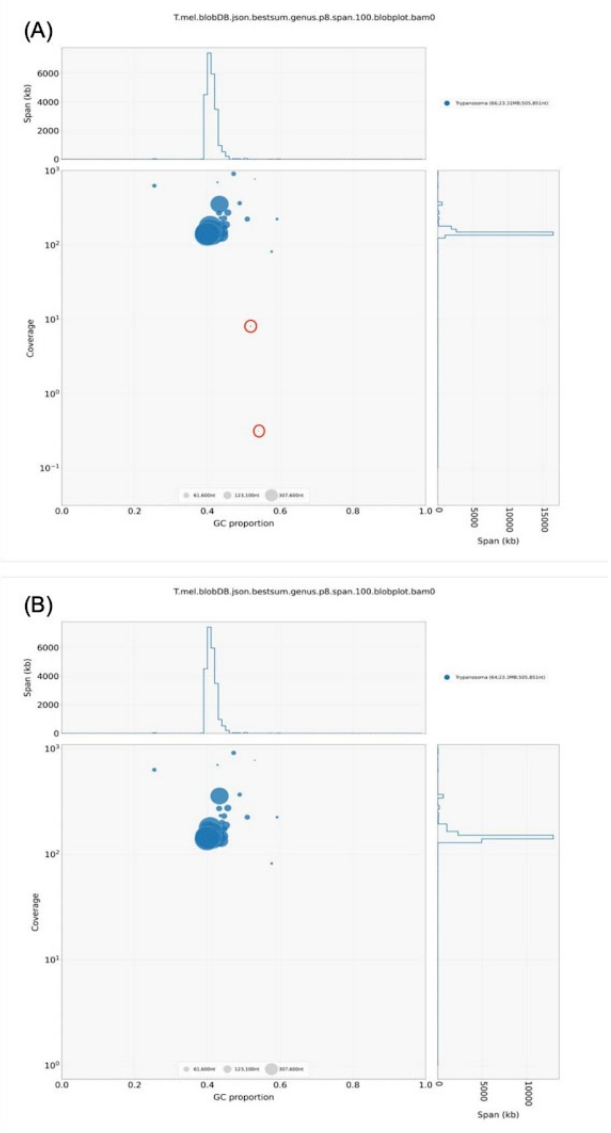


Fig. S2. Coverage and blast annotation for (A) the polished assembly and (B) the manually trimmed assembly.

Biology Open • Supplementary information

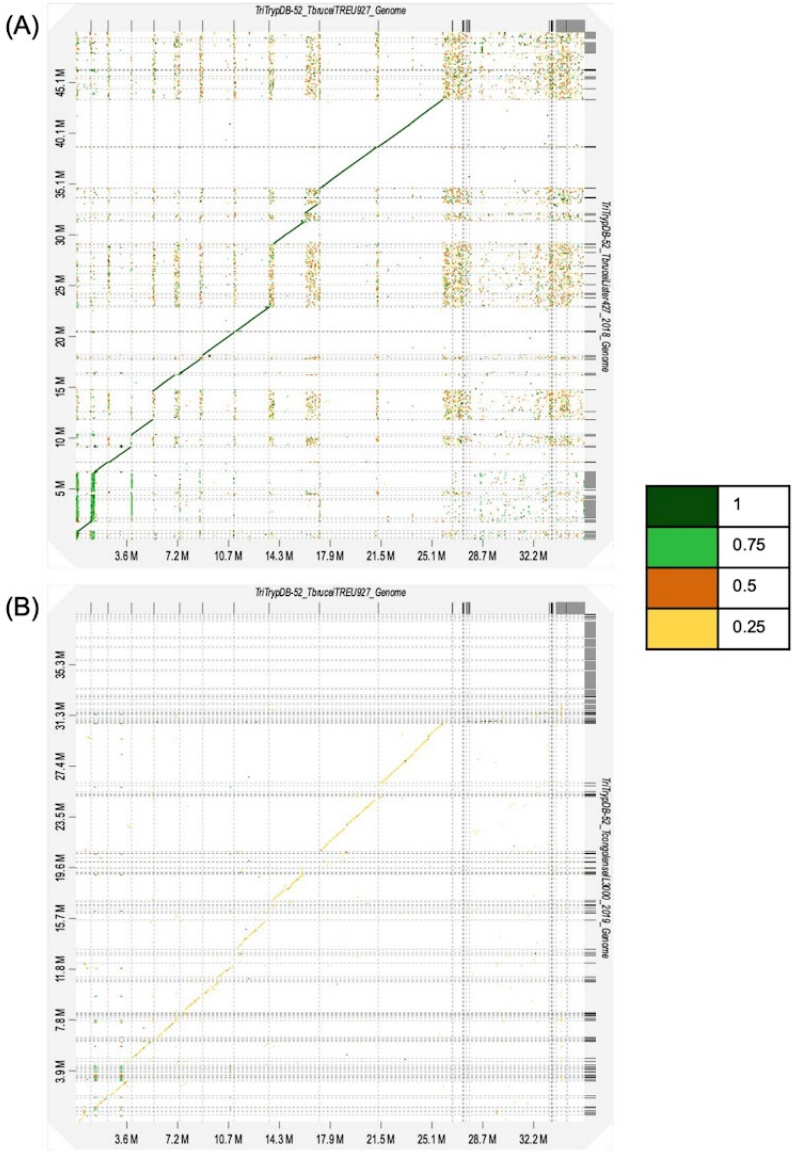


Fig. S3. Synteny plot of (A) *T. brucei* TREU927/4 and *T. brucei* Lister 427 (2018) and (B) *T. brucei* TREU927/4 *and T. congolense* IL3000 2019 genome sequences. The legend refers to the identity between the sequences.

Biology Open • Supplementary information

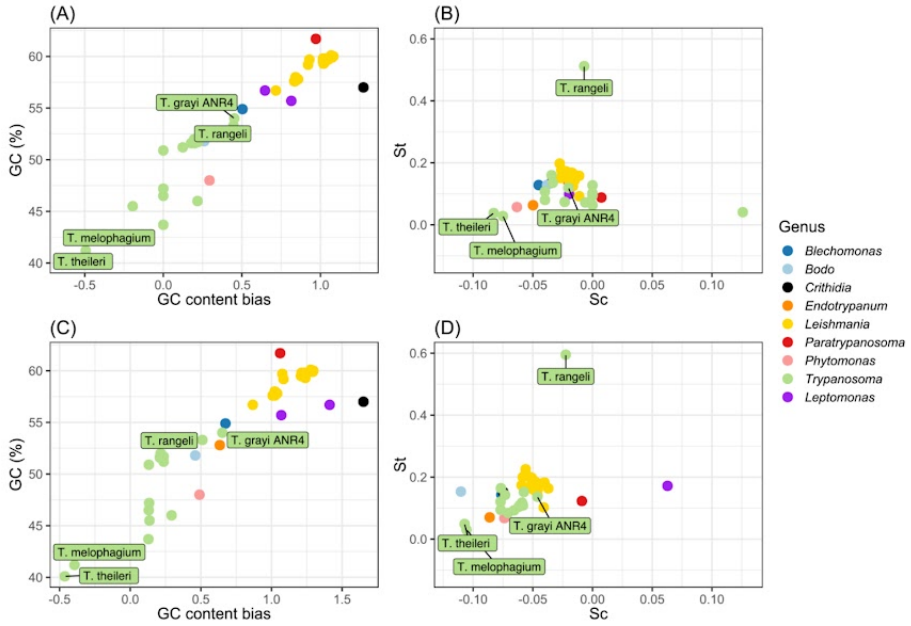


Fig. S4. (A) GC content across the whole genome and GC content bias in a every CDS. (B) Selection acting on translational efficiency (St) and selection acting on nucleotide cost (Sc). (C) GC content across the whole genome and GC content bias in CDS of a trypanosomatid universal single copy orthologue which is essential in every life cycle stage in *T. brucei* (n=158). (D) Selection acting on translational efficiency (St) and selection acting on nucleotide cost (Sc) in trypanosomatid universal single copy orthologues which is essential in every life cycle stage in *T. brucei* (n=158).

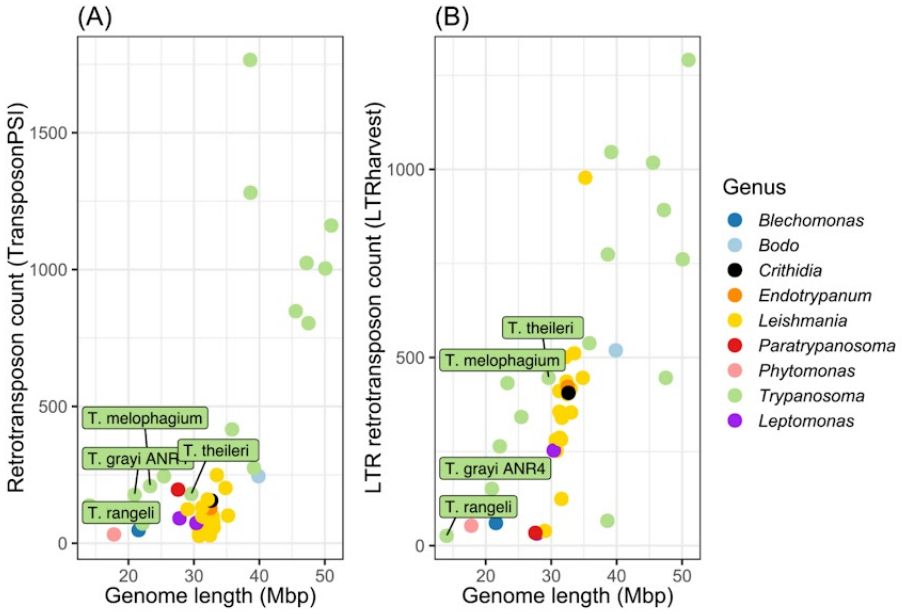


Fig. S5. (A) Retrotransposon (B) long terminal repeat (LTR) retrotransposon counts compared to genome length.

Biology Open • Supplementary information

Table S1. Software and any options used.

Tool	Version	Flag	Available at
Guppy	4.0.15	--flowcell FLO-MIN106 --kit SQK-RAD004	https://community.nanoporetech.com/
PycoQC	2.5.0.3		https://github.com/a-slide/pycoQC
Trimmomatic	0.39	-SLIDINGWINDOW 4:20 -MINLEN 50	https://github.com/timflutre/trimmomatic
GenomeScope	1	-Kmer 21 -Max kmer cov. 1000	http://qb.cshl.edu/genomescope/
Jellyfish	2.2.10	-C -m 21 -s 1000000000	https://github.com/gmarcais/Jellyfish
Wtdbg2	3	-x ont -g 30m -A -t 10	https://github.com/ruanjue/wtdbg2
Minimap2	2.17-r941	-ax map-ont	https://github.com/lh3/minimap2
Racon	v1.4.13	-m 8 -x -6 -g -8 -w 500	https://github.com/isovic/racon
Medaka	V5	-m r941_min_high_g360	https://github.com/nanoporetech/medaka
BWA-MEM	0.7.17-r1188		https://github.com/lh3/bwa
Pilon	1.24	--genome --diploid –bam --bam	https://github.com/broadinstitute/pilon
Blast	2.6.0+	blastn -task megablast -outfmt '6 qseqid	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/

Biology Open • Supplementary information

		staxids bitscore std' -max_target_seqs 1 -max_hsps 1 -evalue 1e-25 or 1e-5	
DIAMOND	2.0.5.143	Blastx --sensitive --max-target-seqs 1 --evalue 1e-25	http://www.diamondsearch.org
BlobTools	1.1.1	-r genus	https://github.com/DRL/blobtools
Scaffold_stats.pl	N/A		https://github.com/sujaikumar/assemblage
BUSCO	V5	-l euglenozoa_odb10	https://busco.ezlab.org/
RepeatModeler	N/A	-database - LTRStruct -pa 8	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker	N/A	-lib consensi.fa.classified -pa 8 -xsmall -nolow –gff	http://www.repeatmasker.org/RMDownload.html
BRAKER2	2.1.6	--prot_seq --bam --softmasking --etpmode --gff3	https://github.com/Gaius-Augustus/BRAKER
D-Genies	1.2.0	-Minimap2	http://dgenies.toulouse.inra.fr/run
tRNAscan-SE	2.0.9	-E	http://lowelab.ucsc.edu/tRNAscan-SE/
CodonMuse	0.1.0	-f -tscan	https://github.com/easeward/CodonMuSe

TransposonPSI	1.0.0	-nuc	http://transposonpsi.sourceforge.net
gt suffixerator	1.6.2	-tis -suf -lcp -des -ssp -sds -dna	http://genometools.org/tools/gt_ltrharvest.html
gt LTRharvest	1.6.2		http://genometools.org/tools/gt_ltrharvest.html
OrthoFinder	2.5.2	-S diamond_ultra_sens	https://github.com/davidemms/OrthoFinder
InterProScan	5.52-86.0	--dp --goterms -appl SignalP-EUK-4.1, Pfam -f TSV	https://www.ebi.ac.uk/interpro/
KinFin	1.0.3		https://github.com/DRL/kinfin
iTOL	6		https://itol.embl.de
R	3.6.1		https://www.r-project.org/
ggplot2	3.3.3		https://cran.r-project.org/web/packages/ggplot2/index.html
ggrepel	0.9.1		https://cran.r-project.org/web/packages/ggrepel/index.html

Biology Open • Supplementary information

Table S2.

Click here to download Table S2

Table S3.

Click here to download Table S3

Table S4.

Click here to download Table S4

Table S5.

Click here to download Table S5

Biology Open • Supplementary information

Bibliography

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I. and Englund, P. T. (2001). The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1513–1518.

Ahmed, Y., Hagos, A., Merga, B., Van Soom, A., Duchateau, L., Goddeeris, B. and Govaere, J. (2018). Trypanosoma equiperdum in the horse : a neglected threat? *Vlaams Diergeneeskd. Tijdschr.* **87**, 66–75.

Akoda, K., Harouna, S., Marcotty, T., De Deken, R. and Van den Bossche, P. (2008). Investigations on the transmissibility of Trypanosoma congolense by the tsetse fly Glossina morsitans morsitans during its development in a mammalian host. *Acta Trop.* **107**, 17–19.

Alsford, S., Turner, D. J., Obado, S. O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-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.

Altenhoff, A. M., Train, C.-M., Gilbert, K. J., Mediratta, I., Mendes de Farias, T., Moi, D., Nevers, Y., Radoykova, H.-S., Rossier, V., Warwick Vesztrocy, A., et al. (2021). OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res.* **49**, D373–D379.

Amir, R. E., Iwai, K. and Ciechanover, A. (2002). The NEDD8 Pathway Is Essential for SCFβ-TrCP-mediated Ubiquitination and Processing of the NF-κB Precursor p105 *. *J. Biol. Chem.* **277**, 23253–23259.

Andrews, S. and Others (2010). FastQC: a quality control tool for high throughput sequence data.

Aregawi, W. G., Agga, G. E., Abdi, R. D. and Büscher, P. (2019). Systematic review and meta-analysis on the global distribution, host range, and prevalence of Trypanosoma evansi. *Parasit. Vectors* **12**, 67.

Ashcroft, M. T. (1960). A comparison between a syringe-passaged and a tsetse-fly-transmitted line of a strain of Trypanosoma rhodesiense. *Ann. Trop. Med. Parasitol.* **54**, 44–53.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38**, D457–62.

Auld, S. K. J. R. and Tinsley, M. C. (2015). The evolutionary ecology of complex lifecycle parasites: linking phenomena with mechanisms. *Heredity* **114**, 125–132.

Auty, H., Mundy, A., Fyumagwa, R. D., Picozzi, K., Welburn, S. and Hoare, R. (2008). Health management of horses under high challenge from trypanosomes: a case study from Serengeti, Tanzania. *Vet. Parasitol.* **154**, 233–241.

Auty, H., Torr, S. J., Michoel, T., Jayaraman, S. and Morrison, L. J. (2015). Cattle trypanosomosis: the diversity of trypanosomes and implications for disease epidemiology and control. *Rev. Sci. Tech.* **34**, 587–598.

Balber, A. E. (1972). Trypanosoma brucei: fluxes of the morphological variants in intact and X-irradiated mice. *Exp. Parasitol.* **31**, 307–319.

Baltz, T., Baltz, D., Giroud, C. and Crockett, J. (1985). Cultivation in a semi-defined medium of animal infective forms of Trypanosoma brucei, T. equiperdum, T. evansi, T. rhodesiense and T. gambiense. *EMBO J.* **4**, 1273–1277.

Barry, J. D. (1986). Antigenic variation during Trypanosoma vivax infections of different host species. *Parasitology* **92**, 51–65.

Bassett, A. and Liu, J.-L. (2014). CRISPR/Cas9 mediated genome engineering in Drosophila. *Methods* **69**, 128–136.

Bast, J., Jaron, K. S., Schuseil, D., Roze, D. and Schwander, T. (2019). Asexual reproduction reduces transposable element load in experimental yeast populations. *Elife* **8**, e48548.

Bastin, P., Bagherzadeh, Z., Matthews, K. R. and Gull, K. (1996). A novel epitope tag system to study protein targeting and organelle biogenesis in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **77**, 235–239.

Beaver, A., Crilly, N. P., Hakim, J., Zhang, B., Bobb, B., Rijo-Ferreira, F., Figueiredo, L. and Mugnier, M. R. (2022). Extravascular spaces are reservoirs of antigenic diversity in Trypanosoma brucei infection. *bioRxiv* 2022.06.27.497797.

Beneke, T. and Gluenz, E. (2020). Bar-seq strategies for the LeishGEdit toolbox. *Mol. Biochem. Parasitol.* **239**, 111295.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J. and Gluenz, E. (2017). A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *R Soc Open Sci* **4**, 170095.

Beneke, T., Demay, F., Hookway, E., Ashman, N., Jeffery, H., Smith, J., Valli, J., Becvar, T., Myskova, J., Lestinova, T., et al. (2019). Genetic dissection of a Leishmania flagellar proteome demonstrates requirement for directional motility in sand fly infections. *PLoS Pathog.* **15**, e1007828.

Benz, C., Dondelinger, F., McKean, P. G. and Urbaniak, M. D. (2017). Cell cycle synchronisation of Trypanosoma brucei by centrifugal counter-flow elutriation reveals the timing of nuclear and kinetoplast DNA replication. *Sci. Rep.* **7**, 17599.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, 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.

Betu Kumeso, V. K., Kalonji, W. M., Rembry, S., Valverde Mordt, O., Ngolo Tete, D., Prêtre, A., Delhomme, S., Ilunga Wa Kyhi, M., Camara, M., Catusse, J., et al. (2022). Efficacy and safety of acoziborole in patients with human African trypanosomiasis caused by Trypanosoma brucei gambiense: a multicentre, open-label, single-arm, phase 2/3 trial. *Lancet Infect. Dis.* **22**, 00660–00660.

Bhat, G. J., Koslowsky, D. J., Feagin, J. E., Smiley, B. L. and Stuart, K. (1990). An

extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. *Cell* **61**, 885–894.

Billington, K., Halliday, C., Madden, R., Dyer, P., Carrington, M., Vaughan, S., Hertz-Fowler, C., Dean, S., Sunter, J. D., Wheeler, R. J., et al. (2022). Genome-wide subcellular protein localisation in the flagellate parasite Trypanosoma brucei. *bioRxiv* 2022.06.09.495287.

Black, S. J., Jack, R. M. and Morrison, W. I. (1983). Host-parasite interactions which influence the virulence of Trypanosoma (Trypanozoon) brucei brucei organisms. *Acta Trop.* **40**, 11–18.

Blanco-Ameijeiras, J., Lozano-Fernández, P. and Martí, E. (2022). Centrosome maturation--in tune with the cell cycle. *J. Cell Sci.* **135**, jcs259395.

Blom, N., Gammeltoft, S. and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362.

Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

Borst, P. and Fase-Fowler, F. (1979). The maxi-circle of Trypanosoma brucei kinetoplast DNA. *Biochim. Biophys. Acta* **565**, 1–12.

Breidbach, T., Ngazoa, E. and Steverding, D. (2002). Trypanosoma brucei: in vitro slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms. *Exp. Parasitol.* **101**, 223–230.

Briggs, E. M., Rojas, F., McCulloch, R., Matthews, K. R. and Otto, T. D. (2021). Single-cell transcriptomic analysis of bloodstream Trypanosoma brucei reconstructs cell cycle progression and developmental quorum sensing. *Nat. Commun.* **12**, 5268.

Briggs, E. M., Marques, C. A., Oldrieve, G. R., Hu, J., Otto, T. D. and Matthews, K. D. (2023). Profiling the bloodstream form and procyclic form Trypanosoma brucei cell cycle using single cell transcriptomics. *bioRxiv* 2023.01.09.523263.

Bringaud, F. and Baltz, T. (1994). African trypanosome glucose transporter genes: organization and evolution of a multigene family. *Mol. Biol. Evol.* **11**, 220–230.

Bringaud, F., Rivière, L. and Coustou, V. (2006). Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol. Biochem. Parasitol.* **149**, 1–9.

Bringaud, F., Ghedin, E., El-Sayed, N. M. A. and Papadopoulou, B. (2008). Role of transposable elements in trypanosomatids. *Microbes Infect.* **10**, 575–581.

Bruce, D., Hamerton, A. E., Bateman, H. R. and Mackie, F. P. (1910). Trypanosome diseases of domestic animals in Uganda II.—Trypanosoma brucei (Plimmer and Bradford). *Proc. R. Soc. Lond. B Biol. Sci.* **83**, 1–14.

Brun, R., Hecker, H. and Lun, Z. R. (1998). Trypanosoma evansi and T. equiperdum: distribution, biology, treatment and phylogenetic relationship (a review). *Vet. Parasitol.* **79**, 95–107.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218.

Burki, F., Roger, A. J., Brown, M. W. and Simpson, A. G. B. (2020). The New Tree of Eukaryotes. *Trends Ecol. Evol.* **35**, 43–55.

Büscher, P., Bart, J.-M., Boelaert, M., Bucheton, B., Cecchi, G., Chitnis, N., Courtin, D., Figueiredo, L. M., Franco, J.-R., Grébaut, P., et al. (2018). Do Cryptic Reservoirs Threaten Gambiense-Sleeping Sickness Elimination? *Trends Parasitol.* **34**, 197–207.

Büscher, P., Gonzatti, M. I., Hébert, L., Inoue, N., Pascucci, I., Schnaufer, A., Suganuma, K., Touratier, L. and Van Reet, N. (2019). Equine trypanosomosis: enigmas and diagnostic challenges. *Parasit. Vectors* **12**, 234.

Cai, X.-L., Li, S.-J., Zhang, P., Li, Z., Hide, G., Lai, D.-H. and Lun, Z.-R. (2021). The Occurrence of Malignancy in Trypanosoma brucei brucei by Rapid Passage in Mice. *Front. Microbiol.* **12**, 806626.

Calistri, P., Narcisi, V., Atzeni, M., De Massis, F., Tittarelli, M., Mercante, M. T., Ruggieri, E. and Scacchia, M. (2013). Dourine Reemergence in Italy. *J. Equine Vet. Sci.* **33**, 83–89.

Capewell, P., Clucas, C., DeJesus, E., Kieft, R., Hajduk, S., Veitch, N., Steketee, P. C., Cooper, A., Weir, W. and MacLeod, A. (2013). The TgsGP gene is essential for resistance to human serum in Trypanosoma brucei gambiense. *PLoS Pathog.* **9**, e1003686.

Carnes, J., Anupama, A., Balmer, O., Jackson, A., Lewis, M., Brown, R., Cestari, I., Desquesnes, M., Gendrin, C., Hertz-Fowler, C., et al. (2015). Genome and phylogenetic analyses of Trypanosoma evansi reveal extensive similarity to T. brucei and multiple independent origins for dyskinetoplasty. *PLoS Negl. Trop. Dis.* **9**, e3404.

Cauchard, J., Carnicer, D., Madeline, A., Guitton, E., Giraudet, A., Büscher, P., Hébert, L. and Laugier, C. (2016). Evaluation of melarsamine hydrochloride (Cymelarsan®) efficacy for the treatment of dourine nervous form on experimentally infected ponies. *J. Equine Vet. Sci.* **39**, S51.

Cayla, M., McDonald, L., MacGregor, P. and Matthews, K. (2020). An atypical DYRK kinase connects quorum-sensing with posttranscriptional gene regulation in Trypanosoma brucei. *Elife* **9**, e51620.

Ceulemans, H. and Bollen, M. (2004). Functional diversity of protein phosphatase-1, a cellular economizer and reset button. *Physiol. Rev.* **84**, 1–39.

Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. and Pfanner, N. (2009). Importing mitochondrial proteins: machineries and mechanisms. *Cell* **138**, 628–644.

Chanez, A.-L., Hehl, A. B., Engstler, M. and Schneider, A. (2006). Ablation of the single dynamin of T. brucei blocks mitochondrial fission and endocytosis and leads to a precise cytokinesis arrest. *J. Cell Sci.* **119**, 2968–2974.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205.

Chen, P. and Zhang, J. (2021). Asexual Experimental Evolution of Yeast Does Not Curtail Transposable Elements. *Mol. Biol. Evol.* **38**, 2831–2842.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92.

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

Constable, P. D., Hinchcliff, K. W., Done, S. H. and Grünberg, W. eds. (2017). 18 - Diseases Primarily Affecting the Reproductive System. In *Veterinary Medicine (Eleventh Edition)*, pp. 1758–1829. W.B. Saunders.

Cook, D. E. and Andersen, E. C. (2017). VCF-kit: assorted utilities for the variant call format. *Bioinformatics* **33**, 1581–1582.

Cribb, T. H., Bray, R. A., Olson, P. D. and Littlewood, D. T. J. (2003). Life cycle evolution in the digenea: a new perspective from phylogeny. *Adv. Parasitol.* **54**, 197–254.

Cross, G. A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. *Parasitology* **71**, 393–417.

Cully, D. F., Ip, H. S. and Cross, G. A. (1985). Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in Trypanosoma brucei. *Cell* **42**, 173–182.

Cuypers, B., Van den Broeck, F., Van Reet, N., Meehan, C. J., Cauchard, J., Wilkes, J. M., Claes, F., Goddeeris, B., Birhanu, H., Dujardin, J.-C., et al. (2017). Genome-Wide SNP Analysis Reveals Distinct Origins of Trypanosoma evansi and Trypanosoma equiperdum. *Genome Biol. Evol.* **9**, 1990–1997.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008.

Daniels, J.-P., Gull, K. and Wickstead, B. (2010). Cell biology of the trypanosome genome. *Microbiol. Mol. Biol. Rev.* **74**, 552–569.

Das, A., Gale, M., Jr, Carter, V. and Parsons, M. (1994). The protein phosphatase inhibitor okadaic acid induces defects in cytokinesis and organellar genome segregation in Trypanosoma brucei. *J. Cell Sci.* **107**, 3477–3483.

Dean, S., Marchetti, R., Kirk, K. and Matthews, K. R. (2009). A surface transporter family conveys the trypanosome differentiation signal. *Nature* **459**, 213–217.

Dean, S., Gould, M. K., Dewar, C. E. and Schnaufer, A. C. (2013). Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes.

Proceedings of the National Academy of Sciences **110**, 14741–14746.

Dean, S., Sunter, J. D. and Wheeler, R. J. (2017). TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends Parasitol.* **33**, 80–82.

de Clare Bronsvoort, B. M., Thumbi, S. M., Poole, E. J., Kiara, H., Auguet, O. T., Handel, I. G., Jennings, A., Conradie, I., Mbole-Kariuki, M. N., Toye, P. G., et al. (2013). Design and descriptive epidemiology of the Infectious Diseases of East African Livestock (IDEAL) project, a longitudinal calf cohort study in western Kenya. *BMC Vet. Res.* **9**, 171.

Dejung, M., Subota, I., Bucerius, F., Dindar, G., Freiwald, A., Engstler, M., Boshart, M., Butter, F. and Janzen, C. J. (2016). Quantitative Proteomics Uncovers Novel Factors Involved in Developmental Differentiation of Trypanosoma brucei. *PLoS Pathog.* **12**, e1005439.

De Lange, T. and Borst, P. (1982). Genomic environment of the expression-linked extra copies of genes for surface antigens of Trypanosoma brucei resembles the end of a chromosome. *Nature* **299**, 451–453.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.

Desquesnes, M., Biteau-Coroller, F., Bouyer, J., Dia, M. L. and Foil, L. (2009). Development of a mathematical model for mechanical transmission of trypanosomes and other pathogens of cattle transmitted by tabanids. *Int. J. Parasitol.* **39**, 333–346.

Desquesnes, M., Kamyingkird, K. and Vergne, T. (2012). Evaluation of melarsomine hydrochloride for the treatment of surra in Thailand in cattle and dog. *Proc. 1st Workshop Sch. Web Min*.

Desquesnes, M., Dargantes, A., Lai, D.-H., Lun, Z.-R., Holzmuller, P. and Jittapalapong, S. (2013). Trypanosoma evansi and surra: a review and perspectives on transmission, epidemiology and control, impact, and zoonotic aspects. *Biomed Res. Int.* **2013**, 321237.

Desquesnes, M., Gonzatti, M., Sazmand, A., Thévenon, S., Bossard, G., Boulangé, A., Gimonneau, G., Truc, P., Herder, S., Ravel, S., et al. (2022a). A review on the diagnosis of animal trypanosomoses. *Parasit. Vectors* **15**, 1–24.

Desquesnes, M., Sazmand, A., Gonzatti, M., Boulangé, A., Bossard, G., Thévenon, S., Gimonneau, G., Truc, P., Herder, S., Ravel, S., et al. (2022b). Diagnosis of animal trypanosomoses: proper use of current tools and future prospects. *Parasit. Vectors* **15**, 235.

Dewar, C. E., MacGregor, P., Cooper, S., Gould, M. K., Matthews, K. R., Savill, N. J. and Schnaufer, A. (2018). Mitochondrial DNA is critical for longevity and metabolism of transmission stage Trypanosoma brucei. *PLoS Pathog.* **14**, e1007195.

Dewar, C. E., Casas-Sanchez, A., Dieme, C., Crouzols, A., Haines, L. R., Acosta-Serrano, Á., Rotureau, B. and Schnaufer, A. (2022). Oxidative Phosphorylation Is Required for Powering Motility and Development of the Sleeping Sickness Parasite Trypanosoma brucei in the Tsetse Fly Vector. *MBio* **13**, e0235721.

Doleželová, E., Kunzová, M., Dejung, M., Levin, M., Panicucci, B., Regnault, C., Janzen, C. J., Barrett, M. P., Butter, F. and Zíková, A. (2020). Cell-based and multi-omics profiling reveals dynamic metabolic repurposing of mitochondria to drive developmental progression of Trypanosoma brucei. *PLoS Biol.* **18**, e3000741.

Dreesen, O., Li, B. and Cross, G. A. M. (2005). Telomere structure and shortening in telomerase-deficient Trypanosoma brucei. *Nucleic Acids Res.* **33**, 4536–4543.

Dyer, N. A., Rose, C., Ejeh, N. O. and Acosta-Serrano, A. (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends Parasitol.* **29**, 188–196.

Dylus, D., Altenhoff, A. M., Majidian, S., Sedlazeck, F. J. and Dessimoz, C. (2022). Read2Tree: scalable and accurate phylogenetic trees from raw reads. *bioRxiv* 2022.04.18.488678.

Ebhodaghe, F., Isaac, C. and Ohiolei, J. A. (2018). A meta-analysis of the prevalence of bovine trypanosomiasis in some African countries from 2000 to 2018. *Prev. Vet. Med.* **160**, 35–46.

Echeverry, M. C., Bot, C., Obado, S. O., Taylor, M. C. and Kelly, J. M. (2012). Centromere-associated repeat arrays on Trypanosoma brucei chromosomes are much more extensive than predicted. *BMC Genomics* **13**, 29.

Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D. and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. *J Biol Methods* **1**, e10.

Ellis, D. S. and Evans, D. A. (1977). Passage of Trypanosoma brucei rhodesiense through the peritrophic membrane of Glossina morsitans morsitans. *Nature* **267**, 834–835.

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., et al. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**, 404–409.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N. and Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131**, 505–515.

Erben, E. D., Fadda, A., Lueong, S., Hoheisel, J. D. and Clayton, C. (2014). A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. *PLoS Pathog.* **10**, e1004178.

Ershov, D., Phan, M.-S., Pylvänäinen, J. W., Rigaud, S. U., Le Blanc, L., Charles-Orszag, A., Conway, J. R. W., Laine, R. F., Roy, N. H., Bonazzi, D., et al. (2021). Bringing TrackMate into the era of machine-learning and deep-learning. *bioRxiv* 2021.09.03.458852.

Evans, D. A. and Ellis, D. S. (1978). The penetrative ability of sleeping-sickness trypanosomes. *Trans. R. Soc. Trop. Med. Hyg.* **72**, 653–655.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034.

Fèvre, E. M., Coleman, P. G., Odiit, M., Magona, J. W., Welburn, S. C. and Woolhouse, M. E. (2001). The origins of a new Trypanosoma brucei rhodesiense sleeping sickness

outbreak in eastern Uganda. *Lancet* **358**, 625–628.

Field, M. C. and Carrington, M. (2009). The trypanosome flagellar pocket. *Nat. Rev. Microbiol.* **7**, 775–786.

Flegontov, P., Votýpka, J., Skalický, T., Logacheva, M. D., Penin, A. A., Tanifuji, G., Onodera, N. T., Kondrashov, A. S., Volf, P., Archibald, J. M., et al. (2013). Paratrypanosoma is a novel early-branching trypanosomatid. *Curr. Biol.* **23**, 1787–1793.

Franco, J. R., Cecchi, G., Paone, M., Diarra, A., Grout, L., Kadima Ebeja, A., Simarro, P. P., Zhao, W. and Argaw, D. (2022). The elimination of human African trypanosomiasis: Achievements in relation to WHO road map targets for 2020. *PLoS Negl. Trop. Dis.* **16**, e0010047.

Frolov, A. O., Malysheva, M. N., Yurchenko, V. and Kostygov, A. Y. (2016). Back to monoxeny: Phytomonas nordicus descended from dixenous plant parasites. *Eur. J. Protistol.* **52**, 1–10.

Gale, M., Jr, Carter, V. and Parsons, M. (1994). Translational control mediates the developmental regulation of the Trypanosoma brucei Nrk protein kinase. *J. Biol. Chem.* **269**, 31659–31665.

Gibson, W. (2007). Resolution of the species problem in African trypanosomes. *Int. J. Parasitol.* **37**, 829–838.

Gibson, W. (2012). The origins of the trypanosome genome strains Trypanosoma brucei brucei TREU 927, T. b. gambiense DAL 972, T. vivax Y486 and T. congolense IL3000. *Parasit. Vectors* **5**, 71.

Gibson, W. and Bailey, M. (2003). The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid Biol. Dis.* **2**, 1.

Gibson, W., Garside, L. and Bailey, M. (1992). Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. *Mol. Biochem. Parasitol.* **51**, 189–199.

Gibson, W., Peacock, L., Ferris, V., Williams, K. and Bailey, M. (2008). The use of yellow fluorescent hybrids to indicate mating in Trypanosoma brucei. *Parasit. Vectors* **1**, 4.

Gingrich, J. B., Roberts, L. W. and Macken, L. M. (1983). Trypanosoma brucei rhodesiense: mechanical transmission by tsetse, Glossina morsitans (Diptera: Glossinidae), in the laboratory. *J. Med. Entomol.* **20**, 673–676.

Giordani, F., Morrison, L. J., Rowan, T. G., de Koning, H. P. and Barrett, M. P. (2016). The animal trypanosomiases and their chemotherapy: a review. *Parasitology* **143**, 1862–1889.

Gizaw, Y., Ashenafi, H., Demssie, T., Bekana, M., Govaere, J. and Abei, G. (2021). Pathological Observations in Horses Naturally Infected with Trypanosoma equiperdum in Western Arsi Zone, Ethiopia. *JAVR* **11**, 9–16.

Gnipová, A., Panicucci, B., Paris, Z., Verner, Z., Horváth, A., Lukeš, J. and Zíková, A. (2012). Disparate phenotypic effects from the knockdown of various Trypanosoma brucei cytochrome c oxidase subunits. *Mol. Biochem. Parasitol.* **184**, 90–98.

Griffiths, S., Portman, N., Taylor, P. R., Gordon, S., Ginger, M. L. and Gull, K. (2007).

RNA interference mutant induction in vivo demonstrates the essential nature of trypanosome flagellar function during mammalian infection. *Eukaryot. Cell* **6**, 1248–1250.

Guegan, F., Rajan, K. S., Bento, F., Pinto-Neves, D., Sequeira, M., Gumińska, N., Mroczek, S., Dziembowski, A., Cohen-Chalamish, S., Doniger, T., et al. (2022). A long noncoding RNA promotes parasite differentiation in African trypanosomes. *Sci Adv* **8**, eabn2706.

Gull, K. (1999). The cytoskeleton of trypanosomatid parasites. *Annu. Rev. Microbiol.* **53**, 629–655.

Gupta, S., Vohra, S., Sethi, K., Gupta, S., Bera, B. C., Kumar, S. and Kumar, R. (2022). In vitro anti-trypanosomal effect of ivermectin on Trypanosoma evansi by targeting multiple metabolic pathways. *Trop. Anim. Health Prod.* **54**, 240.

Gutierrez, C., Desquesnes, M., Touratier, L. and Büscher, P. (2010). Trypanosoma evansi: recent outbreaks in Europe. *Vet. Parasitol.* **174**, 26–29.

Hajduk, S. and Ochsenreiter, T. (2010). RNA editing in kinetoplastids. *RNA Biol.* **7**, 229–236.

Hajduk, S. L., Moore, D. R., Vasudevacharya, J., Siqueira, H., Torri, A. F., Tytler, E. M. and Esko, J. D. (1989). Lysis of Trypanosoma brucei by a toxic subspecies of human high density lipoprotein. *J. Biol. Chem.* **264**, 5210–5217.

Hall, J. P. J., Wang, H. and Barry, J. D. (2013). Mosaic VSGs and the scale of Trypanosoma brucei antigenic variation. *PLoS Pathog.* **9**, e1003502.

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169.

Hébert, L., Guitton, E., Madeline, A., Géraud, T., Zientara, S., Laugier, C., Hans, A., Büscher, P., Cauchard, J. and Petry, S. (2018). Melarsomine hydrochloride (Cymelarsan®) fails to cure horses with Trypanosoma equiperdum OVI parasites in their cerebrospinal fluid. *Vet. Parasitol.* **264**, 47–51.

Higgins, M. K., Lane-Serff, H., MacGregor, P. and Carrington, M. (2017). A Receptor's Tale: An Eon in the Life of a Trypanosome Receptor. *PLoS Pathog.* **13**, e1006055.

Hill, K. L. (2003). Biology and mechanism of trypanosome cell motility. *Eukaryot. Cell* **2**, 200–208.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522.

Hoare, C. A. (1954). The loss of the kinetoplast in trypanosomes, with special reference to Trypanosoma evansi. *J. Protozool.* **1**, 28–33.

Hoare, C. A. (1965). Vampire bats as vectors and hosts of equine and bovine trypanosomes. *Acta Trop.* **22**, 204–216.

Hoare (1972). *The trypanosomes of mammals. A zoological monograph*. Blackwell.

Holland, W. G., Do, T. T., Huong, N. T., Dung, N. T., Thanh, N. G., Vercruysse, J. and

Goddeeris, B. M. (2003). The effect of Trypanosoma evansi infection on pig performance and vaccination against classical swine fever. *Vet. Parasitol.* **111**, 115–123.

Höög, J. L., Gluenz, E., Vaughan, S. and Gull, K. (2010). Chapter 8 - Ultrastructural Investigation Methods for Trypanosoma brucei. In *Methods in Cell Biology* (ed. Müller-Reichert, T.), pp. 175–196. Academic Press.

Horáková, E., Lecordier, L., Cunha, P., Sobotka, R., Changmai, P., Langedijk, C. J. M., Van Den Abbeele, J., Vanhollebeke, B. and Lukeš, J. (2022). Heme-deficient metabolism and impaired cellular differentiation as an evolutionary trade-off for human infectivity in Trypanosoma brucei gambiense. *Nat. Commun.* **13**, 7075.

Hovel-Miner, G. A., Boothroyd, C. E., Mugnier, M., Dreesen, O., Cross, G. A. M. and Papavasiliou, F. N. (2012). Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. *PLoS Pathog.* **8**, e1002900.

Hovel-Miner, G., Mugnier, M. R., Goldwater, B., Cross, G. A. M. and Papavasiliou, F. N. (2016). A Conserved DNA Repeat Promotes Selection of a Diverse Repertoire of Trypanosoma brucei Surface Antigens from the Genomic Archive. *PLoS Genet.* **12**, e1005994.

Howard, R. S. and Lively, C. M. (2002). The Ratchet and the Red Queen: the maintenance of sex in parasites. *J. Evol. Biol.* **15**, 648–656.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**, S96–104.

Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K. and Rudenko, G. (2007). A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *EMBO J.* **26**, 2400–2410.

Hulsman, P., Savenije, H. H. G. and Hrachowitz, M. (2021). Satellite-based drought analysis in the Zambezi River Basin: Was the 2019 drought the most extreme in several decades as locally perceived? *Journal of Hydrology: Regional Studies* **34**, 100789.

Imhof, S., Knüsel, S., Gunasekera, K., Vu, X. L. and Roditi, I. (2014). Social motility of African trypanosomes is a property of a distinct life-cycle stage that occurs early in tsetse fly transmission. *PLoS Pathog.* **10**, e1004493.

Imhof, S., Vu, X. L., Bütikofer, P. and Roditi, I. (2015). A Glycosylation Mutant of Trypanosoma brucei Links Social Motility Defects In Vitro to Impaired Colonization of Tsetse Flies In Vivo. *Eukaryot. Cell* **14**, 588–592.

Jackson, A. P. (2007a). Tandem gene arrays in Trypanosoma brucei: comparative phylogenomic analysis of duplicate sequence variation. *BMC Evol. Biol.* **7**, 54.

Jackson, A. P. (2007b). Evolutionary consequences of a large duplication event in Trypanosoma brucei: chromosomes 4 and 8 are partial duplicons. *BMC Genomics* **8**, 432.

Jackson, A. P., Quail, M. A. and Berriman, M. (2008). Insights into the genome sequence of a free-living Kinetoplastid: Bodo saltans (Kinetoplastida: Euglenozoa). *BMC Genomics* **9**, 594.

Jamonneau, V., Ilboudo, H., Kaboré, J., Kaba, D., Koffi, M., Solano, P., Garcia, A.,

Courtin, D., Laveissière, C., Lingue, K., et al. (2012). Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal. *PLoS Negl. Trop. Dis.* **6**, e1691.

Jannin, J. (1999). Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomosis. *Bull. World Health Organ.* **77**, 450.

Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M. and Schwander, T. (2021). Genomic Features of Parthenogenetic Animals. *J. Hered.* **112**, 19–33.

Jensen, R. E. and Englund, P. T. (2012). Network news: the replication of kinetoplast DNA. *Annu. Rev. Microbiol.* **66**, 473–491.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589.

Kaufer, A., Ellis, J., Stark, D. and Barratt, J. (2017). The evolution of trypanosomatid taxonomy. *Parasit. Vectors* **10**, 287.

Kay, C., Peacock, L., Williams, T. A. and Gibson, W. (2022). Signatures of hybridization in Trypanosoma brucei. *PLoS Pathog.* **18**, e1010300.

Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B., et al. (2007). Functional genomics in Trypanosoma brucei: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol. Biochem. Parasitol.* **154**, 103–109.

Kelly, S., Ivens, A., Mott, G. A., O'Neill, E., Emms, D., Macleod, O., Voorheis, P., Tyler, K., Clark, M., Matthews, J., et al. (2017). An Alternative Strategy for Trypanosome Survival in the Mammalian Bloodstream Revealed through Genome and Transcriptome Analysis of the Ubiquitous Bovine Parasite Trypanosoma (Megatrypanum) theileri. *Genome Biol. Evol.* **9**, 2093–2109.

Kennedy, P. G. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet Neurol.* **12**, 186–194.

Kieft, R., Capewell, P., Turner, C. M. R., Veitch, N. J., MacLeod, A. and Hajduk, S. (2010). Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16137–16141.

Kleisen, C. M. and Borst, P. (1975). Sequence heterogeneity of the mini-circles of kinetoplast DNA of Crithidia luciliae and evidence for the presence of a component more complex than mini-circle DNA in the kinetoplast network. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **407**, 473–478.

Kolev, N. G., Ramey-Butler, K., Cross, G. A. M., Ullu, E. and Tschudi, C. (2012). Developmental progression to infectivity in Trypanosoma brucei triggered by an

RNA-binding protein. *Science* **338**, 1352–1353.

Kurz, T., Pintard, L., Willis, J. H., Hamill, D. R., Gönczy, P., Peter, M. and Bowerman, B. (2002). Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **295**, 1294–1298.

Lacomble, S., Vaughan, S., Gadelha, C., Morphew, M. K., Shaw, M. K., McIntosh, J. R. and Gull, K. (2009). Three-dimensional cellular architecture of the flagellar pocket and associated cytoskeleton in trypanosomes revealed by electron microscope tomography. *J. Cell Sci.* **122**, 1081–1090.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Lai, D.-H., Hashimi, H., Lun, Z.-R., Ayala, F. J. and Lukes, J. (2008). Adaptations of Trypanosoma brucei to gradual loss of kinetoplast DNA: Trypanosoma equiperdum and Trypanosoma evansi are petite mutants of T. brucei. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1999–2004.

Laine, A.-L., Barrès, B., Numminen, E. and Siren, J. P. (2019). Variable opportunities for outcrossing result in hotspots of novel genetic variation in a pathogen metapopulation. *Elife* **8**, e47091.

Lamont, G. S., Tucker, R. S. and Cross, G. A. (1986). Analysis of antigen switching rates in Trypanosoma brucei. *Parasitology* **92 (Pt 2)**, 355–367.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. and Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118.

Laxman, S., Riechers, A., Sadilek, M., Schwede, F. and Beavo, J. A. (2006). Hydrolysis products of cAMP analogs cause transformation of Trypanosoma brucei from slender to stumpy-like forms. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19194–19199.

Le Blancq, S. M., Swinkels, B. W., Gibson, W. C. and Borst, P. (1988). Evidence for gene conversion between the phosphoglycerate kinase genes of Trypanosoma brucei. *J. Mol. Biol.* **200**, 439–447.

Lee, M.-C. and Marx, C. J. (2012). Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet.* **8**, e1002651.

Letunic, I. and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, Z., Tu, X. and Wang, C. C. (2006). Okadaic acid overcomes the blocked cell cycle caused by depleting Cdc2-related kinases in Trypanosoma brucei. *Exp. Cell Res.* **312**, 3504–3516.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Liao, S., Hu, H., Wang, T., Tu, X. and Li, Z. (2017). The Protein Neddylation Pathway in Trypanosoma brucei: Functional characterization and substrate identification. *J. Biol.*

Chem. **292**, 1081–1091.

Ling, A. S., Trotter, J. R. and Hendriks, E. F. (2011). A zinc finger protein, TbZC3H20, stabilizes two developmentally regulated mRNAs in trypanosomes. *J. Biol. Chem.* **286**, 20152–20162.

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.

Longbottom, J., Caminade, C., Gibson, H. S., Weiss, D. J., Torr, S. and Lord, J. S. (2020). Modelling the impact of climate change on the distribution and abundance of tsetse in Northern Zimbabwe. *Parasit. Vectors* **13**, 526.

Lopez, M. A., Saada, E. A. and Hill, K. L. (2015). Insect stage-specific adenylate cyclases regulate social motility in African trypanosomes. *Eukaryot. Cell* **14**, 104–112.

Lord, J. S., Hargrove, J. W., Torr, S. J. and Vale, G. A. (2018). Climate change and African trypanosomiasis vector populations in Zimbabwe's Zambezi Valley: A mathematical modelling study. *PLoS Med.* **15**, e1002675.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Love, M. I., Anders, S., Kim, V. and Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Res.* **4**, 1070.

Lugli, E. B., Pouliot, M., Portela, M. D. P. M., Loomis, M. R. and Raper, J. (2004). Characterization of primate trypanosome lytic factors. *Mol. Biochem. Parasitol.* **138**, 9–20.

Lukeš, J., Skalický, T., Týč, J., Votýpka, J. and Yurchenko, V. (2014). Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* **195**, 115–122.

Lukeš, J., Kachale, A., Votýpka, J., Butenko, A. and Field, M. C. (2022). African trypanosome strategies for conquering new hosts and territories: the end of monophyly? *Trends Parasitol.* **38**, 724–736.

Luzak, V., López-Escobar, L., Siegel, T. N. and Figueiredo, L. M. (2021). Cell-to-Cell Heterogeneity in Trypanosomes. *Annu. Rev. Microbiol.* **75**, 107–128.

Mabille, D., Dirkx, L., Thys, S., Vermeersch, M., Montenye, D., Govaerts, M., Hendrickx, S., Takac, P., Van Weyenbergh, J., Pintelon, I., et al. (2022). Impact of pulmonary African trypanosomes on the immunology and function of the lung. *Nat. Commun.* **13**, 7083.

MacGregor, P. and Matthews, K. R. (2012). Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in Trypanosoma brucei. *Nucleic Acids Res.* **40**, 7705–7717.

MacGregor, P., Szöőr, B., Savill, N. J. and Matthews, K. R. (2012). Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nat. Rev. Microbiol.* **10**, 431–438.

MacGregor, P., Rojas, F., Dean, S. and Matthews, K. R. (2013). Stable transformation of pleomorphic bloodstream form Trypanosoma brucei. *Mol. Biochem. Parasitol.* **190**, 60–62.

Macleod, O. J. S., Bart, J.-M., MacGregor, P., Peacock, L., Savill, N. J., Hester, S., Ravel, S., Sunter, J. D., Trevor, C., Rust, S., et al. (2020). A receptor for the complement regulator factor H increases transmission of trypanosomes to tsetse flies. *Nat. Commun.* **11**, 1–12.

Marais, G. A. B., Campos, P. R. A. and Gordo, I. (2010). Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol. Evol.* **2**, 347–357.

Marçais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770.

Marquet, M., Zöllkau, J., Pastuschek, J., Viehweger, A., Schleußner, E., Makarewicz, O., Pletz, M. W., Ehricht, R. and Brandt, C. (2022). Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. *Sci. Rep.* **12**, 4000.

Maslov, D. A., Opperdoes, F. R., Kostygov, A. Y., Hashimi, H., Lukeš, J. and Yurchenko, V. (2019). Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology* **146**, 1–27.

Matthews, K. R. (2021). Trypanosome Signaling—Quorum Sensing. *Annu. Rev. Microbiol.* **75**, 495–514.

Matthews, K. R. and Gull, K. (1994). Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J. Cell Biol.* **125**, 1147–1156.

Matthews, K. R. and Larcombe, S. (2022). Comment on "Unexpected plasticity in the life cycle of Trypanosoma brucei." *Elife* **11**, e74985.

Matthews, K. R., Ellis, J. R. and Paterou, A. (2004). Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol.* **20**, 40–47.

Maudlin, I., Holmes, P. H. and Miles, M. A. (2004). *The Trypanosomiases*. CABI Pub.

Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S. and Matthews, K. (2006). Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in Trypanosoma brucei: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res.* **34**, 5312–5324.

McAllaster, M. R., Ikeda, K. N., Lozano-Núñez, A., Anrather, D., Unterwurzacher, V., Gossenreiter, T., Perry, J. A., Crickley, R., Mercadante, C. J., Vaughan, S., et al. (2015). Proteomic identification of novel cytoskeletal proteins associated with TbPLK, an essential regulator of cell morphogenesis in Trypanosoma brucei. *Mol. Biol. Cell* **26**, 3013–3029.

McBride, H. M., Neuspiel, M. and Wasiak, S. (2006). Mitochondria: more than just a powerhouse. *Curr. Biol.* **16**, R551–60.

McDonald, L., Cayla, M., Ivens, A., Mony, B. M., MacGregor, P., Silvester, E., McWilliam, K. and Matthews, K. R. (2018). Non-linear hierarchy of the quorum sensing signalling pathway in bloodstream form African trypanosomes. *PLoS Pathog.* **14**, e1007145.

McLintock, L. M., Turner, C. M. and Vickerman, K. (1993). Comparison of the effects of immune killing mechanisms on Trypanosoma brucei parasites of slender and stumpy

morphology. *Parasite Immunol.* **15**, 475–480.

McWilliam, K. R., Ivens, A., Morrison, L. J., Mugnier, M. R. and Matthews, K. R. (2019). Developmental competence and antigen switch frequency can be uncoupled in Trypanosoma brucei. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22774–22782.

Mekonnen, G., Mohammed, E. F., Kidane, W., Nesibu, A., Yohannes, H., Van Reet, N., Büscher, P. and Birhanu, H. (2018). Isometamidium chloride and homidium chloride fail to cure mice infected with Ethiopian Trypanosoma evansi type A and B. *PLoS Negl. Trop. Dis.* **12**, e0006790.

Mihok, S., Maramba, O., Munyoki, E. and Kagoiya, J. (1995). Mechanical transmission of Trypanosoma spp. by African Stomoxyinae (Diptera: Muscidae). *Trop. Med. Parasitol.* **46**, 103–105.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682.

Moloo, S. K., Kabata, J. M. and Gitire, N. M. (2000). Study on the mechanical transmission by tsetse fly Glossina morsitans centralis of Trypanosoma vivax, T. congolense or T. brucei brucei to goats. *Acta Trop.* **74**, 105–108.

Mony, B. M. and Matthews, K. R. (2015). Assembling the components of the quorum sensing pathway in African trypanosomes. *Mol. Microbiol.* **96**, 220–232.

Mony, B. M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D. and Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. *Nature* **505**, 681–685.

Moreira, R. S., Calomeno, N. A., das Neves, G. B., do Nascimento, L. F. N., Filho, V. B., Wagner, G. and Miletti, L. C. (2022). Trypanosoma evansi secretome carries potential biomarkers for Surra diagnosis. *J. Proteomics* **10**, 104789.

Morrison, W. I., Wells, P. W., Moloo, S. K., Paris, J. and Murray, M. (1982). Interference in the establishment of superinfections with Trypanosoma congolense in cattle. *J. Parasitol.* **68**, 755–764.

Morrison, L. J., Vezza, L., Rowan, T. and Hope, J. C. (2016). Animal African Trypanosomiasis: Time to Increase Focus on Clinically Relevant Parasite and Host Species. *Trends Parasitol.* **32**, 599–607.

Morrison, L. J., Steketee, P. C., Tettey, M. D. and Matthews, K. R. (2022). Pathogenicity and virulence of African trypanosomes: from laboratory models to clinically relevant hosts. *Virulence* **14**, 2150445.

Mottram, J. C. and Smith, G. (1995). A family of trypanosome cdc2-related protein kinases. *Gene* **162**, 147–152.

Mugnier, M. R., Cross, G. A. M. and Papavasiliou, F. N. (2015). The in vivo dynamics of antigenic variation in Trypanosoma brucei. *Science* **347**, 1470–1473.

Mugo, E. and Clayton, C. (2017). Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. *PLoS Pathog.* **13**, e1006560.

Mugo, E., Egler, F. and Clayton, C. (2017). Conversion of procyclic-form Trypanosoma brucei to the bloodstream form by transient expression of RBP10. *Mol. Biochem.*

Parasitol. **216**, 49–51.

Mulindwa, J., Leiss, K., Ibberson, D., Kamanyi Marucha, K., Helbig, C., Melo do Nascimento, L., Silvester, E., Matthews, K., Matovu, E., Enyaru, J., et al. (2018). Transcriptomes of Trypanosoma brucei rhodesiense from sleeping sickness patients, rodents and culture: Effects of strain, growth conditions and RNA preparation methods. *PLoS Negl. Trop. Dis.* **12**, e0006280.

Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., et al. (2018). Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**, 121–125.

Naguleswaran, A., Fernandes, P., Bevkal, S., Rehmann, R., Nicholson, P. and Roditi, I. (2021). Developmental changes and metabolic reprogramming during establishment of infection and progression of Trypanosoma brucei brucei through its insect host. *PLoS Negl. Trop. Dis.* **15**, e0009504.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C. and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751.

Navarro, M. and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature* **414**, 759–763.

Nelson, C. W., Moncla, L. H. and Hughes, A. L. (2015). SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **31**, 3709–3711.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274.

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I. and Ochsenreiter, T. (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS Pathog.* **6**, e1001037.

Nolan, D. P. and Voorheis, H. P. (1992). The mitochondrion in bloodstream forms of Trypanosoma brucei is energized by the electrogenic pumping of protons catalysed by the F1F0-ATPase. *Eur. J. Biochem.* **209**, 207–216.

Nolan, D. P., Rolin, S., Rodriguez, J. R., Van Den Abbeele, J. and Pays, E. (2000). Slender and stumpy bloodstream forms of Trypanosoma brucei display a differential response to extracellular acidic and proteolytic stress. *Eur. J. Biochem.* **267**, 18–27.

Oberholzer, M., Lopez, M. A., McLelland, B. T. and Hill, K. L. (2010). Social motility in african trypanosomes. *PLoS Pathog.* **6**, e1000739.

Oberle, M., Balmer, O., Brun, R. and Roditi, I. (2010). Bottlenecks and the maintenance of minor genotypes during the life cycle of Trypanosoma brucei. *PLoS Pathog.* **6**, e1001023.

Odeniran, P. O., Onifade, A. A., Omolabi, K. F. and Ademola, I. O. (2021). Financial losses estimation of African animal trypanosomosis in Nigeria: field reality-based model. *Trop. Anim. Health Prod.* **53**, 159.

Oldrieve, G. R. (2022). First person – Guy Oldrieve. *Biol. Open* **11**, bio059369.

Oldrieve, G., Verney, M., Jaron, K. S., Hébert, L. and Matthews, K. R. (2021). Monomorphic Trypanozoon: towards reconciling phylogeny and pathologies. *Microb Genom* **7**, 000632.

Oldrieve, G. R., Malacart, B., López-Vidal, J. and Matthews, K. R. (2022). The genomic basis of host and vector specificity in non-pathogenic trypanosomatids. *Biol. Open* **11**, bio059237.

Olego-Fernandez, S., Vaughan, S., Shaw, M. K., Gull, K. and Ginger, M. L. (2009). Cell morphogenesis of Trypanosoma brucei requires the paralogous, differentially expressed calpain-related proteins CAP5.5 and CAP5.5V. *Protist* **160**, 576–590.

Orr, G. A., Werner, C., Xu, J., Bennett, M., Weiss, L. M., Takvorkan, P., Tanowitz, H. B. and Wittner, M. (2000). Identification of novel serine/threonine protein phosphatases in Trypanosoma cruzi: a potential role in control of cytokinesis and morphology. *Infect. Immun.* **68**, 1350–1358.

Osaka, F., Saeki, M., Katayama, S., Aida, N., Toh-E, A., Kominami, K., Toda, T., Suzuki, T., Chiba, T., Tanaka, K., et al. (2000). Covalent modifier NEDD8 is essential for SCF ubiquitin-ligase in fission yeast. *EMBO J.* **19**, 3475–3484.

Otto, S. P. (2021). Selective Interference and the Evolution of Sex. *J. Hered.* **112**, 9–18.

Pagabeleguem, S., Ravel, S., Dicko, A. H., Vreysen, M. J. B., Parker, A., Takac, P., Huber, K., Sidibé, I., Gimonneau, G. and Bouyer, J. (2016). Influence of temperature and relative humidity on survival and fecundity of three tsetse strains. *Parasit. Vectors* **9**, 520.

Pal, U. and Fikrig, E. (2003). Adaptation of Borrelia burgdorferi in the vector and vertebrate host. *Microbes Infect.* **5**, 659–666.

Parsons, M., Nelson, R. G., Watkins, K. P. and Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. *Cell* **38**, 309–316.

Pascucci, I., Di Provvido, A., Cammà, C., Di Francesco, G., Calistri, P., Tittarelli, M., Ferri, N., Scacchia, M. and Caporale, V. (2013). Diagnosis of dourine in outbreaks in Italy. *Vet. Parasitol.* **193**, 30–38.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419.

Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M. and Gibson, W. (2011). Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3671–3676.

Peacock, L., Bailey, M., Carrington, M. and Gibson, W. (2014). Meiosis and haploid gametes in the pathogen Trypanosoma brucei. *Curr. Biol.* **24**, 181–186.

Peacock, L., Bailey, M. and Gibson, W. (2016). Dynamics of gamete production and mating in the parasitic protist Trypanosoma brucei. *Parasit. Vectors* **9**, 404.

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., et al. (2012). Scikit-learn: Machine Learning in Python. *arXiv [cs.LG]* 2825–2830.

Peregrine, A. S., Kemei, S. and Ndoutamia, G. (1995). Cross-resistance phenotypes associated with induction of resistance to isometamidium chloride and quinapyramine sulphate in Trypanosoma congolense.OAU/STRC.

Pita, S., Díaz-Viraqué, F., Iraola, G. and Robello, C. (2019). The Tritryps Comparative Repeatome: Insights on Repetitive Element Evolution in Trypanosomatid Pathogens. *Genome Biol. Evol.* **11**, 546–551.

Podlipaev, S. (2001). The more insect trypanosomatids under study-the more diverse Trypanosomatidae appears. *Int. J. Parasitol.* **31**, 648–652.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178.

Portman, N. and Gull, K. (2014). Identification of paralogous life-cycle stage specific cytoskeletal proteins in the parasite Trypanosoma brucei. *PLoS One* **9**, e106777.

Povelones, M. L. (2014). Beyond replication: division and segregation of mitochondrial DNA in kinetoplastids. *Mol. Biochem. Parasitol.* **196**, 53–60.

Pozo, J. C., Timpte, C., Tan, S., Callis, J. and Estelle, M. (1998). The ubiquitin-related protein RUB1 and auxin response in Arabidopsis. *Science* **280**, 1760–1763.

Priest, J. W. and Hajduk, S. L. (1994). Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei. *J. Bioenerg. Biomembr.* **26**, 179–191.

Radwanska, M., Vereecke, N., Deleeuw, V., Pinto, J. and Magez, S. (2018). Salivarian Trypanosomosis: A Review of Parasites Involved, Their Global Distribution and Their Interaction With the Innate and Adaptive Mammalian Host Immune System. *Front. Immunol.* **9**, 2253.

Ranallo-Benavidez, T. R., Jaron, K. S. and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432.

Ranjithkumar, M., Saravanan, B. C., Yadav, S. C., Kumar, R., Singh, R. and Dey, S. (2014). Neurological trypanosomiasis in quinapyramine sulfate-treated horses--a breach of the blood-brain barrier? *Trop. Anim. Health Prod.* **46**, 371–377.

Raper, J., Portela, M. P., Lugli, E., Frevert, U. and Tomlinson, S. (2001). Trypanosome lytic factors: novel mediators of human innate immunity. *Curr. Opin. Microbiol.* **4**, 402–408.

R Core Team (2022). R: A Language and Environment for Statistical Computing.

Read, M. A., Brownell, J. E., Gladysheva, T. B., Hottelet, M., Parent, L. A., Coggins, M. B., Pierce, J. W., Podust, V. N., Luo, R. S., Chau, V., et al. (2000). Nedd8 modification of cul-1 activates SCF(beta(TrCP))-dependent ubiquitination of IkappaBalpha. *Mol. Cell. Biol.* **20**, 2326–2333.

Reuner, B., Vassella, E., Yutzy, B. and Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol.*

Biochem. Parasitol. **90**, 269–280.

Riley, J. (1983). Recent advances in our understanding of pentastomid reproductive biology. *Parasitology* **86 (Pt 4)**, 59–83.

Roberts, L. W., Wellde, B. T., Reardon, M. J. and Onyango, F. K. (1989). Mechanical transmission of Trypanosoma brucei rhodesiense by Glossina morsitans morsitans (Diptera:Glossinidae). *Ann. Trop. Med. Parasitol.* **83 Suppl 1**, 127–131.

Robinson, D. R. and Gull, K. (1991). Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. *Nature* **352**, 731–733.

Rodríguez, J. A., Lopez, M. A., Thayer, M. C., Zhao, Y., Oberholzer, M., Chang, D. D., Kisalu, N. K., Penichet, M. L., Helguera, G., Bruinsma, R., et al. (2009). Propulsion of African trypanosomes is driven by bihelical waves with alternating chirality separated by kinks. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19322–19327.

Rojas, F. and Matthews, K. R. (2019). Quorum sensing in African trypanosomes. *Curr. Opin. Microbiol.* **52**, 124–129.

Rojas, F., Silvester, E., Young, J., Milne, R., Tettey, M., Houston, D. R., Walkinshaw, M. D., Pérez-Pi, I., Auer, M., Denton, H., et al. (2019). Oligopeptide Signaling through TbGPR89 Drives Trypanosome Quorum Sensing. *Cell* **176**, 306–317.e16.

Rojas, F., Cayla, M. and Matthews, K. R. (2021). Basement membrane proteins as a substrate for efficient Trypanosoma brucei differentiation in vitro. *PLoS Negl. Trop. Dis.* **15**, e0009284.

Roos, D. S. (2005). Genetics. Themes and variations in apicomplexan parasite biology. *Science* **309**, 72–73.

Rose, C., Casas-Sánchez, A., Dyer, N. A., Solórzano, C., Beckett, A. J., Middlehurst, B., Marcello, M., Haines, L. R., Lisack, J., Engstler, M., et al. (2020). Trypanosoma brucei colonizes the tsetse gut via an immature peritrophic matrix in the proventriculus. *Nat Microbiol* **5**, 909–916.

Rotureau, B. and Van Den Abbeele, J. (2013). Through the dark continent: African trypanosome development in the tsetse fly. *Front. Cell. Infect. Microbiol.* **3**, 53.

Rotureau, B., Subota, I., Buisson, J. and Bastin, P. (2012). A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly. *Development* **139**, 1842–1850.

Rotureau, B., Ooi, C.-P., Huet, D., Perrot, S. and Bastin, P. (2014). Forward motility is essential for trypanosome infection in the tsetse fly. *Cell. Microbiol.* **16**, 425–433.

Saldivia, M., Ceballos-Pérez, G., Bart, J.-M. and Navarro, M. (2016). The AMPKα1 Pathway Positively Regulates the Developmental Transition from Proliferation to Quiescence in Trypanosoma brucei. *Cell Reports* **17**, 660–670.

Santos, A. L. S., Branquinha, M. H. and D'Avila-Levy, C. M. (2006). The ubiquitous gp63-like metalloprotease from lower trypanosomatids: in the search for a function. *An. Acad. Bras. Cienc.* **78**, 687–714.

Savage, V. L., Christley, R., Pinchbeck, G., Morrison, L. J., Hodgkinson, J. and Peachey, L. E. (2021). Co-infection with Trypanosoma congolense and Trypanosoma brucei is a significant risk factor for cerebral trypanosomosis in the equid population of

the Gambia. *Prev. Vet. Med.* **197**, 105507.

Sazmand, A., Bahari, A., Papi, S. and Otranto, D. (2020). Parasitic diseases of equids in Iran (1931-2020): a literature review. *Parasit. Vectors* **13**, 586.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.

Schnaufer, A. (2010). Evolution of dyskinetoplastic trypanosomes: how, and how often? *Trends Parasitol.* **26**, 557–558.

Schnaufer, A., Domingo, G. J. and Stuart, K. (2002). Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *Int. J. Parasitol.* **32**, 1071–1084.

Schnaufer, A., Clark-Walker, G. D., Steinberg, A. G. and Stuart, K. (2005). The F1-ATP synthase complex in bloodstream stage trypanosomes has an unusual and essential function. *EMBO J.* **24**, 4029–4040.

Schneider, A. and Ochsenreiter, T. (2018). Failure is not an option–mitochondrial genome segregation in trypanosomes. *J. Cell Sci.* **131** , 221820.

Schumann Burkard, G., Jutzi, P. and Roditi, I. (2011). Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Mol. Biochem. Parasitol.* **175**, 91–94.

Schuster, S., Lisack, J., Subota, I., Zimmermann, H., Reuter, C., Mueller, T., Morriswood, B. and Engstler, M. (2021). Unexpected plasticity in the life cycle of Trypanosoma brucei. *Elife* **10**, e66028.

Schwander, T., Crespi, B. J., Gries, R. and Gries, G. (2013). Neutral and selection-driven decay of sexual traits in asexual stick insects. *Proc. Biol. Sci.* **280**, 20130823.

Seed, J. R., Sechelski, J. B. and Loomis, M. R. (1990). A survey for a trypanocidal factor in primate sera. *J. Protozool.* **37**, 393–400.

Seidl, A., Dávila, A. M. and Silva, R. A. (1999). Estimated financial impact of Trypanosoma vivax on the Brazilian pantanal and Bolivian lowlands. *Mem. Inst. Oswaldo Cruz* **94**, 269–272.

Shapiro, S. Z., Naessens, J., Liesegang, B., Moloo, S. K. and Magondu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Trop.* **41**, 313–323.

Sharma, R., Gluenz, E., Peacock, L., Gibson, W., Gull, K. and Carrington, M. (2009). The heart of darkness: growth and form of Trypanosoma brucei in the tsetse fly. *Trends Parasitol.* **25**, 517–524.

Shaw, A. P. M., Cecchi, G., Wint, G. R. W., Mattioli, R. C. and Robinson, T. P. (2014). Mapping the economic benefits to livestock keepers from intervening against bovine trypanosomosis in Eastern Africa. *Prev. Vet. Med.* **113**, 197–210.

Shaw, S., DeMarco, S. F., Rehmann, R., Wenzler, T., Florini, F., Roditi, I. and Hill, K. L. (2019). Flagellar cAMP signaling controls trypanosome progression through host tissues. *Nat. Commun.* **10**, 803.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.

Silva Pereira, S., de Almeida Castilho Neto, K. J. G., Duffy, C. W., Richards, P., Noyes, H., Ogugo, M., Rogério André, M., Bengaly, Z., Kemp, S., Teixeira, M. M. G., et al. (2020). Variant antigen diversity in Trypanosoma vivax is not driven by recombination. *Nat. Commun.* **11**, 844.

Silvester, E., McWilliam, K. R. and Matthews, K. R. (2017a). The Cytological Events and Molecular Control of Life Cycle Development of Trypanosoma brucei in the Mammalian Bloodstream. *Pathogens* **6**, 29.

Silvester, E., Young, J., Ivens, A. and Matthews, K. R. (2017b). Interspecies quorum sensing in co-infections can manipulate trypanosome transmission potential. *Nat Microbiol* **2**, 1471–1479.

Simpson, L. (1979). Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1585–1588.

Smith, J. M. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35.

Smith, A. B., Esko, J. D. and Hajduk, S. L. (1995). Killing of trypanosomes by the human haptoglobin-related protein. *Science* **268**, 284–286.

Smith, T. K., Bringaud, F., Nolan, D. P. and Figueiredo, L. M. (2017). Metabolic reprogramming during the Trypanosoma brucei life cycle. *F1000Res.* **6**, 683.

Soneson, C., Love, M. I. and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521.

Spence, P. J., Jarra, W., Lévy, P., Reid, A. J., Chappell, L., Brugat, T., Sanders, M., Berriman, M. and Langhorne, J. (2013). Vector transmission regulates immune control of Plasmodium virulence. *Nature* **498**, 228–231.

Stanne, T. M., Kushwaha, M., Wand, M., Taylor, J. E. and Rudenko, G. (2011). TbISWI regulates multiple polymerase I (Pol I)-transcribed loci and is present at Pol II transcription boundaries in Trypanosoma brucei. *Eukaryot. Cell* **10**, 964–976.

Starrfelt, J. and Kokko, H. (2012). Bet-hedging-a triple trade-off between means, variances and correlations. *Biol. Rev. Camb. Philos. Soc.* **87**, 742–755.

Steinert, M. (1960). Mitochondria associated with the kinetonucleus of Trypanosoma mega. *J. Biophys. Biochem. Cytol.* **8**, 542–546.

Steketee, P. C., Dickie, E. A., Iremonger, J., Crouch, K., Paxton, E., Jayaraman, S., Alfituri, O. A., Awuah-Mensah, G., Ritchie, R., Schnaufer, A., et al. (2021). Divergent metabolism between Trypanosoma congolense and Trypanosoma brucei results in differential sensitivity to metabolic inhibition. *PLoS Pathog.* **17**, e1009734.

Stephens, J. L., Lee, S. H., Paul, K. S. and Englund, P. T. (2007). Mitochondrial fatty acid synthesis in Trypanosoma brucei. *J. Biol. Chem.* **282**, 4427–4436.

Su, C., Evans, D., Cole, R. H., Kissinger, J. C., Ajioka, J. W. and Sibley, L. D. (2003). Recent expansion of Toxoplasma through enhanced oral transmission. *Science* **299**, 414–416.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J. and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016.

Suganuma, K., Narantsatsral, S., Battur, B., Yamasaki, S., Otgonsuren, D., Musinguzi, S. P., Davaasuren, B., Battsetseg, B. and Inoue, N. (2016). Isolation, cultivation and molecular characterization of a new Trypanosoma equiperdum strain in Mongolia. *Parasit. Vectors* **9**, 481.

Sunter, J. D. and Gull, K. (2016). The flagellum attachment zone: “the cellular ruler” of trypanosome morphology. *Trends Parasitol.* **32**, 309–324.

Sunter, J. D., Yanase, R., Wang, Z., Catta-Preta, C. M. C., Moreira-Leite, F., Myskova, J., Pruzinova, K., Volf, P., Mottram, J. C. and Gull, K. (2019). Leishmania flagellum attachment zone is critical for flagellar pocket shape, development in the sand fly, and pathogenicity in the host. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6351–6360.

Szöőr, B. and Cayla, M. (2020). Gel-Based Methods for the Investigation of Signal Transduction Pathways in Trypanosoma brucei. In *Trypanosomatids: Methods and Protocols* (ed. Michels, P. A. M.), Ginger, M. L.), and Zilberstein, D.), pp. 497–522. New York, NY: Springer US.

Szöőr, 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.

Szöor, 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.

Szöőr, B., Simon, D. V., Rojas, F., Young, J., Robinson, D. R., Krüger, T., Engstler, M. and Matthews, K. R. (2019). Positional Dynamics and Glycosomal Recruitment of Developmental Regulators during Trypanosome Differentiation. *MBio* **10**, e00875–19.

Tamarit, A., Gutierrez, C., Arroyo, R., Jimenez, V., Zagalá, G., Bosch, I., Sirvent, J., Alberola, J., Alonso, I. and Caballero, C. (2010). Trypanosoma evansi infection in mainland Spain. *Vet. Parasitol.* **167**, 74–76.

Taylor, J. E. and Rudenko, G. (2006). Switching trypanosome coats: what’s in the wardrobe? *Trends Genet.* **22**, 614–620.

Tettey, M. D., Rojas, F. and Matthews, K. R. (2022). Extracellular release of two peptidases dominates generation of the trypanosome quorum-sensing signal. *Nat. Commun.* **13**, 1–14.

Thompson, J. N. (2005). Coevolution: the geographic mosaic of coevolutionary arms races. *Curr. Biol.* **15**, R992–4.

Tiengwe, C., Koeller, C. M. and Bangs, J. D. (2018). Endoplasmic reticulum-associated degradation and disposal of misfolded GPI-anchored proteins in Trypanosoma brucei. *Mol. Biol. Cell* **29**, 2397–2409.

Timms, M. W., van Deursen, F. J., Hendriks, E. F. and Matthews, K. R. (2002). Mitochondrial development during life cycle differentiation of African trypanosomes: evidence for a kinetoplast-dependent differentiation control point. *Mol. Biol. Cell* **13**, 3747–3759.

Toro, M., León, E., López, R., Pallota, F., Garcia, J. A. and Ruíz, A. (1983). Effect of isometamidium on infections by Trypanosoma vivax and T. evansi in experimentally-infected animals. *Vet. Parasitol.* **13**, 35–43.

Trindade, S., De Niz, M., Costa-Sequeira, M., Bizarra-Rebelo, T., Bento, F., Dejung, M., Narciso, M. V., López-Escobar, L., Ferreira, J., Butter, F., et al. (2022). Slow growing behavior in African trypanosomes during adipose tissue colonization. *Nat. Commun.* **13**, 7548.

Tsirigos, K. D., Peters, C., Shu, N., Käll, L. and Elofsson, A. (2015). The TOPCONS web server for combined membrane protein topology and signal peptide prediction. *Nucleic Acids Res.* **43**, W401–W407.

Tuntasuvan, D., Trongwanichnam, K. and Sukruen, A. (2003). Efficacy of diminazene aceturate on the treatment of trypanosomosis in pigs. *Journal of the Thai* **54**, 49–55.

Turner, C. M. (1990). The use of experimental artefacts in African trypanosome research. *Parasitol. Today* **6**, 14–17.

Turner, C. M. (1997). The rate of antigenic variation in fly-transmitted and syringe-passaged infections of Trypanosoma brucei. *FEMS Microbiol. Lett.* **153**, 227–231.

Turner, C. M. and Barry, J. D. (1989). High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. *Parasitology* **99 Pt 1**, 67–75.

Tyler, K. M. and Engman, D. M. (2001). The life cycle of Trypanosoma cruzi revisited. *Int. J. Parasitol.* **31**, 472–481.

Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homblé, F., Grélard, A., Zhendre, V., Nolan, D. P., Lins, L., et al. (2013). Mechanism of Trypanosoma brucei gambiense resistance to human serum. *Nature* **501**, 430–434.

Van Den Abbeele, J., Caljon, G., De Ridder, K., De Baetselier, P. and Coosemans, M. (2010). Trypanosoma brucei modifies the tsetse salivary composition, altering the fly feeding behavior that favors parasite transmission. *PLoS Pathog.* **6**, e1000926.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D. P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., et al. (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**, 83–87.

van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Gilchrist, C. L. M., Söding, J. and Steinegger, M. (2022). Foldseek: fast and accurate protein structure search. *bioRxiv* 2022.02.07.479398.

Van Xong, H., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P. and Pays, E. (1998). A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in Trypanosoma rhodesiense. *Cell* **95**, 839–846.

Vassella, E., Reuner, B., Yutzy, B. and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle

arrest via the cAMP pathway. *J. Cell Sci.* **110 (Pt 21)**, 2661–2671.

Venter, F., Matthews, K. R. and Silvester, E. (2022). Parasite co-infection: an ecological, molecular and experimental perspective. *Proc. Biol. Sci.* **289**, 20212155.

Ventura, R. M., Takata, C. S., Silva, R. A., Nunes, V. L., Takeda, G. F. and Teixeira, M. M. (2000). Molecular and morphological studies of Brazilian Trypanosoma evansi stocks: the total absence of kDNA in trypanosomes from both laboratory stocks and naturally infected domestic and wild mammals. *J. Parasitol.* **86**, 1289–1298.

Vercesi, A. E., Docampo, R. and Moreno, S. N. (1992). Energization-dependent Ca2+ accumulation in Trypanosoma brucei bloodstream and procyclic trypomastigotes mitochondria. *Mol. Biochem. Parasitol.* **56**, 251–257.

Verner, Z., Basu, S., Benz, C., Dixit, S., Dobáková, E., Faktorová, D., Hashimi, H., Horáková, E., Huang, Z., Paris, Z., et al. (2015). Malleable mitochondrion of Trypanosoma brucei. *Int. Rev. Cell Mol. Biol.* **315**, 73–151.

Verney, M., Gautron, M., Lemans, C., Rincé, A., Hans, A. and Hébert, L. (2022). Development of a microsphere-based immunoassay for the serological diagnosis of equine trypanosomosis. *Sci. Rep.* **12**, 1308.

Vickerman, K. (1965). Polymorphism and Mitochondrial Activity In Sleeping Sickness Trypanosomes. *Nature* **208**, 762–766.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. *J. Cell Sci.* **5**, 163–193.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *Br. Med. Bull.* **41**, 105–114.

Vigneron, A., O'Neill, M. B., Weiss, B. L., Savage, A. F., Campbell, O. C., Kamhawi, S., Valenzuela, J. G. and Aksoy, S. (2020). Single-cell RNA sequencing of *Trypanosoma brucei* from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 2613–2621.

Votýpka, J., d'Avila-Levy, C. M., Grellier, P., Maslov, D. A., Lukeš, J. and Yurchenko, V. (2015). New Approaches to Systematics of Trypanosomatidae: Criteria for Taxonomic (Re)description. *Trends Parasitol.* **31**, 460–469.

Vurture, 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.

Walden, H., Podgorski, M. S., Huang, D. T., Miller, D. W., Howard, R. J., Minor, D. L., Jr, Holton, J. M. and Schulman, B. A. (2003). The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Mol. Cell* **12**, 1427–1437.

Walsh, B. and Hill, K. L. (2021). Right place, right time: Environmental sensing and signal transduction directs cellular differentiation and motility in Trypanosoma brucei. *Mol. Microbiol.* **115**, 930–941.

Walshe, D. P., Lehane, M. J. and Haines, L. R. (2011). Post eclosion age predicts the prevalence of midgut trypanosome infections in Glossina. *PLoS One* **6**, e26984.

Waters, C. M. and Bassler, B. L. (2005). Quorum sensing: cell-to-cell communication in

bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346.

Weir, W., Capewell, P., Foth, B., Clucas, C., Pountain, A., Steketee, P., Veitch, N., Koffi, M., De Meeûs, T., Kaboré, J., et al. (2016). Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *Elife* **5**, e11473.

Weiss, B. L., Wang, J., Maltz, M. A., Wu, Y. and Aksoy, S. (2013). Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. *PLoS Pathog.* **9**, e1003318.

Weiss, B. L., Savage, A. F., Griffith, B. C., Wu, Y. and Aksoy, S. (2014). The peritrophic matrix mediates differential infection outcomes in the tsetse fly gut following challenge with commensal, pathogenic, and parasitic microbes. *J. Immunol.* **193**, 773–782.

Welburn, S. C. and Maudlin, I. (1992). The nature of the teneral state in Glossina and its role in the acquisition of trypanosome infection in tsetse. *Ann. Trop. Med. Parasitol.* **86**, 529–536.

Wen, Y.-Z., Tang, H.-T., Cai, X.-L., Wu, N., Xu, J.-Z., Su, B.-X., Hide, G., Lun, Z.-R. and Lai, D.-H. (2022). PAG3 promotes the differentiation of bloodstream forms in Trypanosoma brucei and reveals the evolutionary relationship among the Trypanozoon trypanosomes. *Front. Cell. Infect. Microbiol.* **12**,.

Wheeler, R. J. (2010). The trypanolytic factor–mechanism, impacts and applications. *Trends Parasitol.* **26**, 457–464.

Wheeler, R. J., Gluenz, E. and Gull, K. (2013a). The limits on trypanosomatid morphological diversity. *PLoS One* **8**, e79581.

Wheeler, R. J., Scheumann, N., Wickstead, B., Gull, K. and Vaughan, S. (2013b). Cytokinesis in Trypanosoma brucei differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis. *Mol. Microbiol.* **90**, 1339–1355.

Wickham, H. (2016). Data Analysis. In *ggplot2: Elegant Graphics for Data Analysis* (ed. Wickham, H.), pp. 189–201. Cham: Springer International Publishing.

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.

Wijers, D. J. (1958). Factors that may influence the infection rate of Glossina palpalis with Trypanosoma gambiense. I. The age of the fly at the time of the infected feed. *Ann. Trop. Med. Parasitol.* **52**, 385–390.

Woodward, R. and Gull, K. (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of Trypanosoma brucei. *J. Cell Sci.* **95 (Pt 1)**, 49–57.

Yang, Z. and dos Reis, M. (2011). Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* **28**, 1217–1228.

Yorke, W. and Blacklock, B. (1912). A note on the morphology of a strain of Trypanosoma equiperdum. *Br. Med. J.* **2**, 473.

Zhou, Q., Gu, J., Lun, Z.-R., Ayala, F. J. and Li, Z. (2016). Two distinct cytokinesis pathways drive trypanosome cell division initiation from opposite cell ends. *Proc. Natl.*

Acad. Sci. U. S. A. **113**, 3287–3292.

Zhou, Q., An, T., Pham, K. T. M., Hu, H. and Li, Z. (2018). The CIF1 protein is a master orchestrator of trypanosome cytokinesis that recruits several cytokinesis regulators to the cytokinesis initiation site. *J. Biol. Chem.* **293**, 16177–16192.

Ziegelbauer, K., Quinten, M., Schwarz, H., Pearson, T. W. and Overath, P. (1990). Synchronous differentiation of Trypanosoma brucei from bloodstream to procyclic forms in vitro. *Eur. J. Biochem.* **192**, 373–378.

Zíková, A., Panigrahi, A. K., Uboldi, A. D., Dalley, R. A., Handman, E. and Stuart, K. (2008). Structural and functional association of Trypanosoma brucei MIX protein with cytochrome c oxidase complex. *Eukaryot. Cell* **7**, 1994–2003.

Zimmermann, H., Subota, I., Batram, C., Kramer, S., Janzen, C. J., Jones, N. G. and Engstler, M. (2017). A quorum sensing-independent path to stumpy development in Trypanosoma brucei. *PLoS Pathog.* **13**, e1006324.