Massey. S[1]; Paxton. E[1]; Steketee. P[1]; Prendergast. J[2]; MacLeod. A[3]; Morrison. L[1]

Affiliations:
1. Division of Infection and Immunity, The Roslin Institute, University of Edinburgh
2. Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh
3. Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow

Analysing the Impact of Sexual Recombination on the Segregation of Virulence Genes in African Trypanosomes

1. Introduction & Aims

Trypanosomes are responsible for both Human and Animal African trypanosomiasis, causing severe health and economic burdens across sub-Saharan Africa. The lifecycle of trypanosomes is shown in Figure 1.

Trypanosomes are extracellular parasites – they frequently change expression of their **variant surface glycoproteins (VSGs),** encoded at their telomeres, which cover the surface of bloodstream parasites to evade host immune responses **[1-2]** (Figure 2).

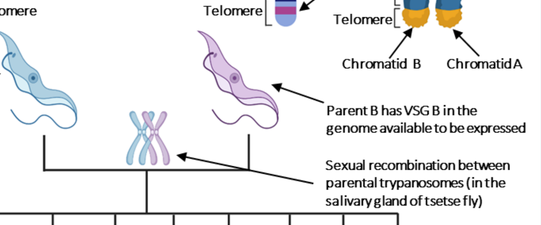

Figure 1. Schematic of the trypanosome lifecycle, adapted from [3-5]

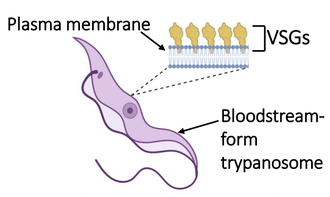

Figure 2. VSGs cover the surface bloodstream-form trypanosome

Aim > Understand how sexual recombination between parental strains of *Trypanosoma brucei* influences VSG content in progeny genomes using long read sequencing

2. Sexual Recombination in Trypanosomes

Trypanosomes undergo sexual recombination (Figure 3), but it is unknown **how this recombination influences the inheritance of the telomeric and subtelomeric genome regions**, which contain important virulence genes and gene families, such as VSGs.

The *T. brucei* genome is **compartmentalised** with VSGs segregated in the **telomeric** regions (Figure 4) **[6].** The aim of the project is to sequence and assemble parental genomes and 6 progeny genomes to map recombination events in VSG regions, to ascertain how parental haplotypes are **inherited in the progeny following sexual recombination.**

Figure 3. Schematic showing sexual recombination between two parental strains of *T. brucei* containing different VSG repertoires and proposing how progeny would inherit VSG content based on Mendelian genetics.

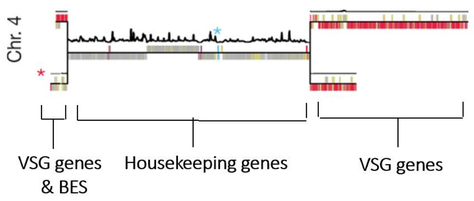

Figure 4. Adapted from [6] showing compartmentalisation of the trypanosome genome with VSG genes (red) segregated into hemizygous telomeric regions

VSGs
- 2,000 VSGs in the *T. brucei* genome, expressed monoallelically **[7]**
- Expressed from 1 of ~20 telomeric bloodstream expression sites (BESs) **[8-9]**
- VSGs are present in repetitive, hemizygous, subtelomeric regions **[6,10]**

3. DNA Extraction and Nanopore Sequencing

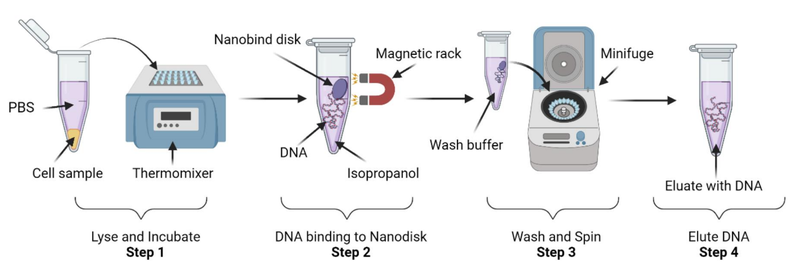

Figure 5. Schematic showing DNA extraction using Circulomics Nanobind CBB Big DNA Kit

VSGs and other virulence genes of interest are found in the notoriously difficult-to-sequence **telomeric regions** of the trypanosome genome. Assembling **complete, contiguous, repetit** arrays can only be achieved through **long read sequencing**.

To generate long-read data**, ultra-high molecular weight DNA extraction methods are required**, such as the Circulomics Nanobi CBB Big DNA kit. The extraction process of this method is outline in Figure 5.

Nanopore sequencing (shown in Figure 6) has generated the **longest read lengths** of any sequencing technology **[11].** As long read data is integral to assembling the repetitive VSG regions, nanopore sequencing was chosen as the primary sequencing technology.

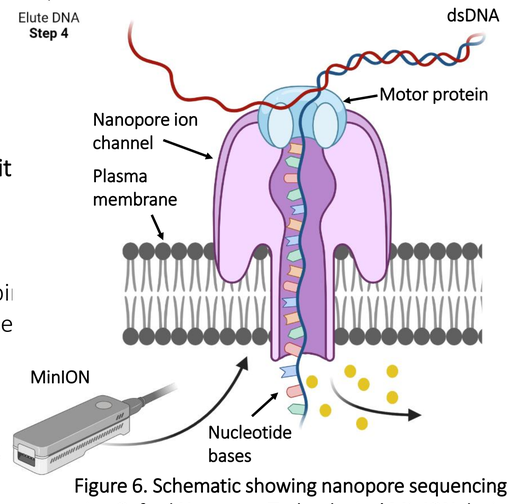

Figure 6. Schematic showing nanopore sequencing using Oxford Nanopore Technologies' MinION device

4. Results: MinION Run Statistics

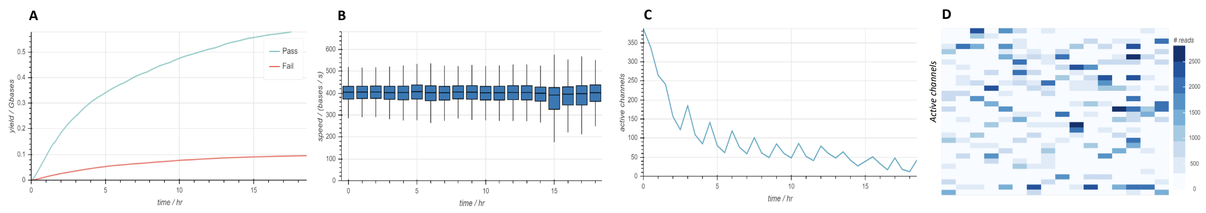

Figure 7. Showing the performance of MinION Run 1 through time. (A) Total yield from flow cell over time, (B) Speed of bases passing through nanopore every second over time, (C) Number of active nanopore channels over time, (D) Number of reads passing through each channel over time

Run performance through time (Figure 7) is monitored, which indicates whether any issues, such as air bubbles, appeared during the run. Following each run, **QC metrics and assembly statistics are calculated** (Table 1). As my project focuses on assembling VSG regions and requires long read data, it is important to have **a high N50 statistic** (which shows that 50% of the genome is contained in fragments of that size or larger), and **low numbers of fragments and scaffolds** (Runs 3 and 4, shown below).

MinION Run #	Clone	Extraction Kit Protocol	Library Sequencing Kit	MinION Pores	Total Reads (K)	Total Yield (Gb)	Mean Read Length (Kb)	N50 Read Length (Kb)	Mean Qscore	Total Length (Mb)	Fragments	Fragments N50 (Mb)	Largest Fragment (Mb)	Scaffolds	Mean Coverage
1	STIB 247 Clone 1	HMW DNA Extraction of Cultured Cells (Circulomics)	Rapid Sequencing Kit (SQK-RAD004)	690	186.1	0.6799	3.1	5.6	10.2	31.80	610	0.32	1.55	34	13
2	STIB 247 Clone 1	HMW DNA Extraction of Cultured Cells (Circulomics)	Rapid Sequencing Kit (SQK-RAD004)	1065	97.2	0.5349	3.6	6.2	9.8	31.10	661	0.095	0.53	26	9
3	STIB 247 Clone 1	UHMW DNA Extraction of Cultured Cells (Circulomics)	Rapid Sequencing Kit (SQK-RAD004)	905	71.1	1.1	**8.5**	**18.3**	9.5	**36.03**	**156**	**1.50**	**5.23**	**9**	20
4	TREU 927 Clone 1	UHMW DNA Extraction of Cultured Cells (Circulomics)	Nanobind UL Library Prep Kit (Circulomics)/ Ultra-Long DNA Sequencing Kit (ONT; SQK-ULK001)	1415	214.7	**2.0**	1.4	**17.9**	9.7	33.84	**185**	**2.14**	**5.60**	**6**	**41**

Table 1. Showing MinION Run quality control and assembly statistics for runs 1-4, for both parental strains of *Trypanosoma brucei*

5. Results: Parental Genome Assemblies

***De novo* assemblies of parental genomes.** Both genomes were assembled using Flye and error corrected and polished using Medaka. Dotplots were generated to compare each assembled parental genome against the TREU 927 reference (Figure 8).

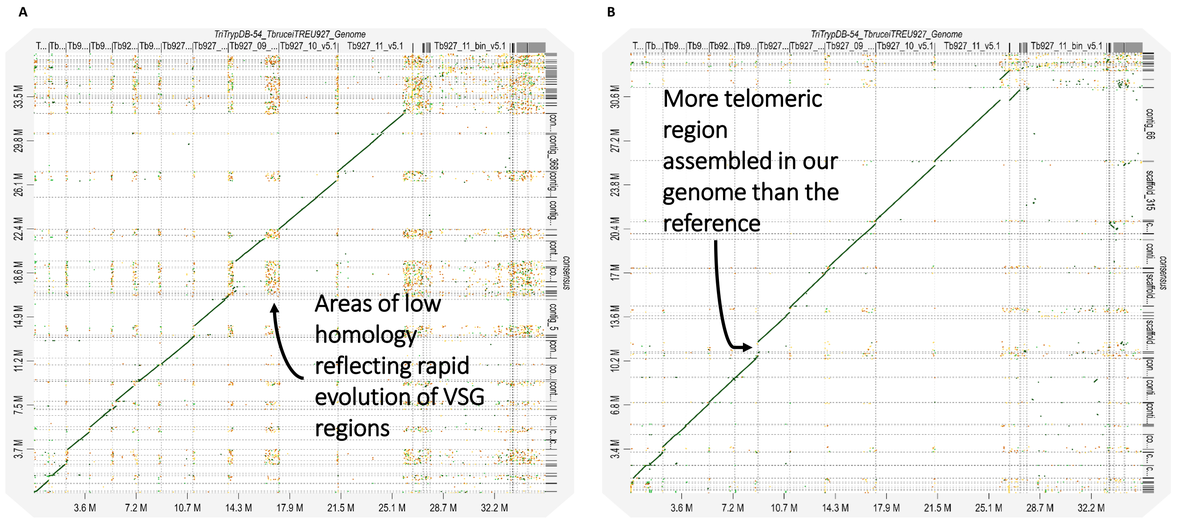

Figure 8. Dotplots showing the results of the de novo genomes of both parental T. brucei strains (A) STIB 247 (Runs 1-3 combined; 37.14Mb) and (B) TREU 927 (Run 4; 33.83Mb) compared to the TREU 927 reference genome (TriTrypDB)

The *de novo* TREU 927 genome aligns well to the TREU 927 reference, but we have managed to assemble more telomeric sequence for several chromosomes. As expected STIB 247 aligns less well to the TREU 927 reference, and especially at the rapidly evolving VSG regions. Additional **Hi-C sequencing and PacBio HiFi data will be used to further refine these assemblies and VSG regions.**

6. Future Work

1. Finalise PacBio assembly of two parental genomes
2. Perform Hi-C Sequencing to improve the assembly of parental telomeric regions
3. Identify VSG content in parental genomes
4. Analyse the inheritance of parental haplotypes, including the segregation of the 2,000 VSGs in F1 progeny assemblies, and how this might influence the progression of antigenic variation.

7. References

1. Daniels, J.-P., Gull, K. & Wickstead, B. Cell Biology of the Trypanosome Genome. Microbiol. *Mol. Biol. Rev.* **74**, 552–569 (2010)
2. Matthews, K. R, McCulloch, R. & Morrison, L, J. The within-host dynamics of African trypanosome infections. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, (2015)
3. K, V. Developmental cycles and biology of pathogenic trypanosomes. *Br. Med. Bull.* **41**, 105–114 (1985).
4. David Barry, J. & McCulloch, R. Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite. *Adv. Parasitol.* **49**, 1–70 (2001).
5. Gibson, W. & Peacock, L. Fluorescent proteins reveal what trypanosomes get up to inside the tsetse fly. *Parasites and Vectors* vol. **12** (2019).
6. Müller, L. S. M. *et al.* Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**, 121–125 (2018)
7. Horn, D. Antigenic variation in African trypanosomes. *Molecular and Biochemical Parasitology* vol. **195** 123–129 (2014)
8. Morrison, L. J., Majiwa, P., Read, A. F. & Barry, J. D. Probabilistic order in antigenic variation of Trypanosoma brucei. *Int. J. Parasitol.* **35**, 961–972 (2005)
9. Sheader, K., Te Vruchte, D. & Rudenko, G. Bloodstream Form-specific Up-regulation of Silent VSG Expression Sites and Procyclin in Trypanosoma brucei after Inhibition of DNA Synthesis or DNA Damage. *J. Biol. Chem.* **279**, 13363–13374 (2004).
10. Berriman, M. *et al.* The genome of the African trypanosome Trypanosoma brucei. *Science (80-.).* **309**, 416–422 (2005)
11. Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* vol. **21** 1–16 (2020)
12. Figures 1, 2, 3, 5 and 6 generated using *biorender.com*

Biotechnology and Biological Sciences Research Council
THE UNIVERSITY of EDINBURGH The Royal (Dick) School of Veterinary Studies
Precision Medicine
University of Glasgow
MRC Medical Research Council