

Zihao Chen¹; S Cooper¹, E Wadsworth¹; L Tavernier²; L Vermeiren²; J Van Den Abbeele²; N Savill^{1*}; A Schnauffer^{1*}; F Van den Broeck^{2,3*}
¹Institute of Immunology and Infection Research, University of Edinburgh, UK; ²Institute of Tropical Medicine, Antwerp, Belgium; ³Rega Institute, KU Leuven, Leuven, Belgium
 *: shared senior authorship in alphabetical order

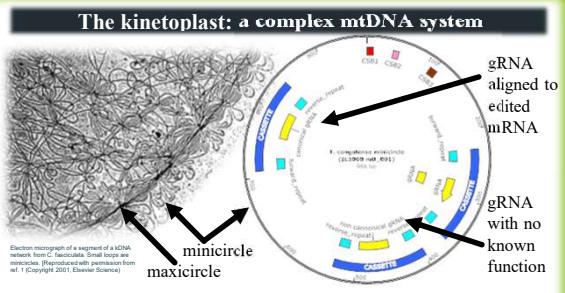


Fig. 1: Maxi and mini-circles linked like chain mail armour in disk. Fig. 2: *Tc* minicircle structure and encoded gRNAs

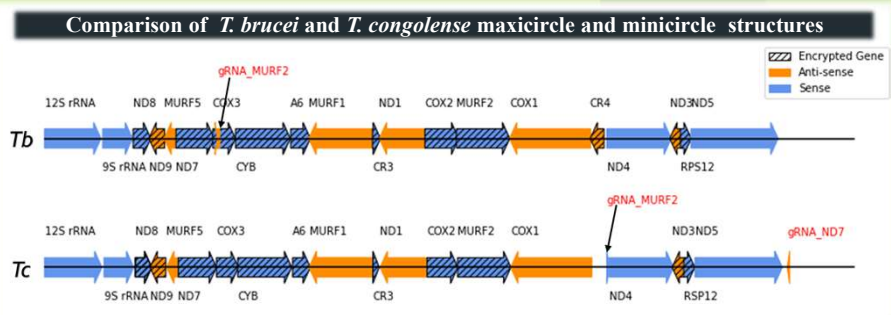
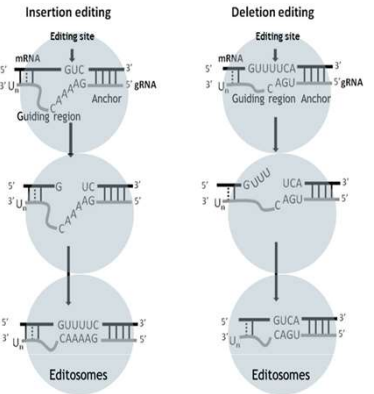
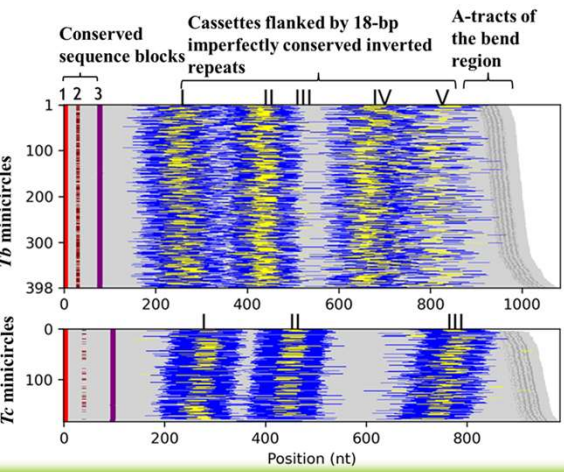


Fig. 4: *Tb* and *Tc* maxicircles are syntenic for mRNA and rRNA genes, but not for gRNA genes



The aim of this study is to compare structure and gene content of the kinetoplast DNA (kDNA) of *T. congolense* (*Tc*, strain IL3000) and *T. brucei* (*Tb*, strain EATRO 1125¹). *Tc* maxicircles and minicircles are being assembled from Illumina reads using the KOMICS pipeline². Short and long transcriptomes are analysed using Illumina and PacBio sequencing and bespoke bioinformatics tools.

Fig. 3 Guide RNAs (gRNAs) direct U insertion and deletion editing of mitochondrial mRNAs



CSB motifs (No. of minicircles)		
	<i>Tb</i>	<i>Tc</i>
CSB1	GGGCGTGCA: 397 GGGCGTGTA: 1	GGGCGTGCA: 115 GGGCGTGCA: 65 GGGCGTCT: 4
CSB2	TCCCGTGC: 288	TCCCGTAC: 31 TCCCGTGT: 3 TCCCGTGC: 1
CSB3	GGGGTTGGTGA: 395 GGGGTTGGTGA: 3	GGGGTTGGTGA: 179 GGGGTTGGTGA: 5

Bold: canonical CSB motif

Fig. 5: Some, but not all CSB motifs are conserved between *Tc* and *Tb*

Fig. 6: *Tc* has fewer minicircles classes, and its minicircles are structurally more homogeneous. Color coding: blue are cassettes and yellow are expressed gRNAs

1. Cooper et al., 2019, Nucleic Acids Res 47(21):11304-11325. doi: 10.1093/nar/gkz28
 2. Van den Broeck et al., 2020, PNAS 117(40):25159-25168. doi: 10.1073/pnas.1920136117

Comparison of *T. brucei* and *T. congolense* gRNA gene cassettes

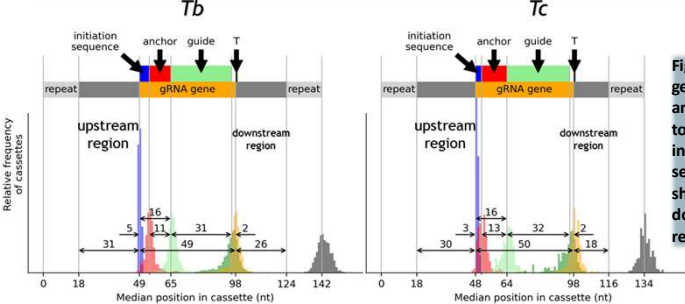


Fig. 7: *Tc* gRNA gene cassettes are shorter due to shorter gRNA initiation sequences and shorter downstream regions

Each gRNA has a unique 5' "anchor" region of 6-10 nt length that binds via Watson-Crick base-pairing to pre- and partially edited mRNA transcripts. The central "guiding" region of each gRNA directs U-insertion/deletion editing of the mRNA using both Watson-Crick and GU wobble base-pairing.

Number (percentage) of cassettes				
gRNA gene expression	expressed		non-expressed	
	<i>Tb</i>	<i>Tc</i>	<i>Tb</i>	<i>Tc</i>
canonical	917 (98%)	292(71%)	14 (2%)	120 (29%)
non-canonical	295 (76%)	71(61%)	92(24%)	46(39%)
total	1212 (92%)	363(68%)	106 (8%)	166(31%)

Fig. 8: *Tb* EATRO1125 has 2.5 x more gRNA genes. Both strains have a significant number of gRNA genes of unknown function (non-canonical genes). A higher percentage of canonical genes is expressed compared to non-canonical genes.

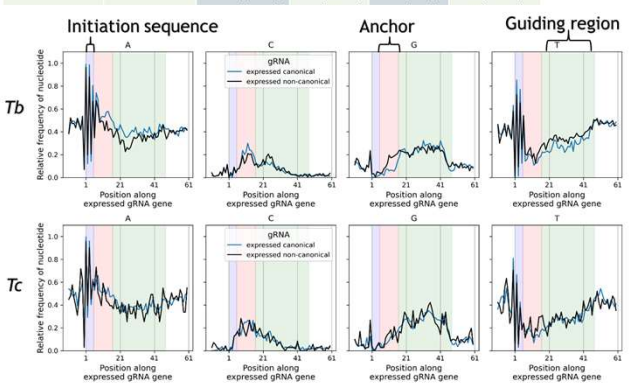


Fig. 9. gRNA genes show a strong nucleotide bias, reflecting frequent ATATA initiation sequences, low G frequency in anchor regions, and low C frequency in guiding regions. Non-canonical gRNA of *Tb* show an increased G frequency in anchor sequences, perhaps reflecting relaxed selective pressure to maintain functionality.

Comparison of *T. brucei* and *T. congolense* gRNAs

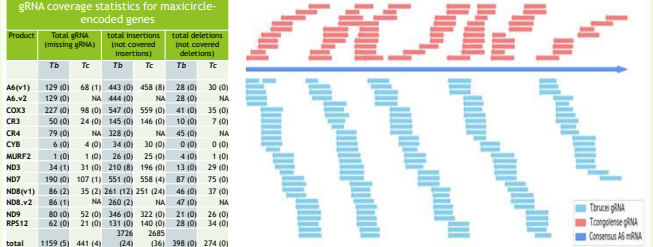


Fig. 10: The identified gRNAs cover >98% of known editing events in both strains

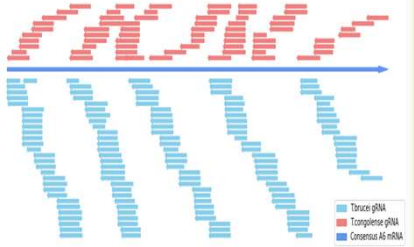


Fig. 11: *Tc* has lower gRNA coverage compared to *Tb*, consistent with a more 'streamlined' minicircle genome of lower complexity. This is exemplified here by the mRNA and gRNAs for subunit A6 of the F1Fo-ATP synthase.

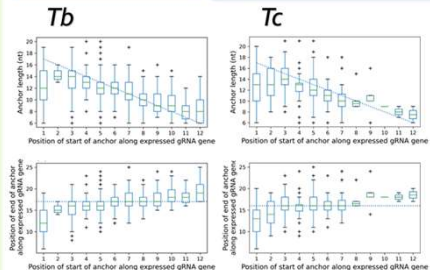


Fig. 12: gRNAs for both species show longer initiation sequences are paired with shorter anchors to conserve the combined length, indicating a 'molecular ruler' that restricts the acceptable distance between the triphosphate 5' end of the gRNA and the 3' end of the anchor sequence.

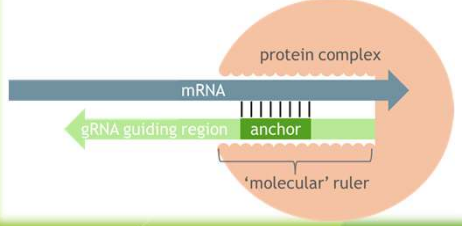


Fig. 13: graphic presentation of 'molecular ruler'

Acknowledgements and references

This work was supported by the UK Medical Research Council [G0600129 and MR/L019701/1 to A.S.]. Z.C is on a joint PhD programme funded by EGRS School of Biological Sciences Scholarship within the EASTBio Programme.
 1. Cooper et al., (2019) Nucleic Acids Res 47(21):11304-11325. PMID: 31665448
 2. Van den Broeck et al. (2020) PNAS 117(40):25159-25168. PMID: 32958676