UNIVERSITY of York

This is a repository copy of *Chromosomal assembly of the nuclear genome of the endosymbiontbearing trypanosomatid Angomonas deanei*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/178723/

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Title: Chromosomal assembly of the nuclear genome of the endosymbiont-bearing trypanosomatid *Angomonas deanei*

Authors: John W. Davey[*], Carolina M. C. Catta-Preta[*,†,‡], Sally James[*], Sarah Forrester[*], Maria Cristina M. Motta[§,**], Peter D. Ashton[*], Jeremy C. Mottram[*,†]

Institutions

[*] Department of Biology, University of York, York, YO10 5DD, United Kingdom

[†] York Biomedical Research Institute, University of York, York, YO10 5DD, United Kingdom

[‡] Medicinal Chemistry Center (CQMED)/ Structural Genomics Consortium, Universidade Estadual de Campinas, Campinas, São Paulo 13083-886, Brazil[1]

[§] Laboratório de Ultraestrutura Celular Hertha Meyer, Instituto de Biofísica Carlos Chagas Filho, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

[**] Instituto Nacional de Ciência e Tecnologia em Biologia Estrutural e Bioimagens – RJ, Brazil

[1] Present address

Running title: *A. deanei* chromosomal genome assembly

Corresponding author: John Davey, Department of Biology, University of York, Wentworth Way, York, YO10 5DD, United Kingdom. Email: john.davey@york.ac.uk

Abstract

Angomonas deanei is an endosymbiont-bearing trypanosomatid with several highly fragmented genome assemblies and unknown chromosome number. We present an assembly of the *Angomonas deanei* nuclear genome based on Oxford Nanopore sequence that resolves into 29 complete or close-to-complete chromosomes. The assembly has several previously unknown special features; it has a supernumerary chromosome, a chromosome with a 340 kilobase inversion, and there is a translocation between two chromosomes. We also present an updated annotation of the chromosomal genome with 10 365 protein-coding genes, 59 transfer RNAs, 26 ribosomal RNAs and 62 non-coding RNAs.

Introduction

Angomonas deanei is a trypanosomatid that mutually coevolves with an endosymbiont, a β-Proteobacterium of the Alcaligenaceae family that contains a reduced genome when compared to its ancestral prokaryote. The symbiont divides during the host cell cycle such that each new protozoan contains a single bacterium. Trypanosomatid endosymbiosis involves an intense metabolic exchange: the bacterium supplies the protozoan with amino acids, heme and vitamins, while benefiting from the host's energy and phospholipid production (de Azevedo-Martins *et al.* 2007; Motta *et al.* 2010; Alves *et al.* 2011, 2013; Klein *et al.* 2013; Loyola-Machado *et al.* 2017). Thus, endosymbiosis in trypanosomatids has been used to study cell evolution and the origin of organelles.

Symbiont-harboring trypanosomatids are distributed in four genera: *Angomonas*, *Strigomonas* (Teixeira *et al.* 2011) and *Kentomonas* (Votýpka *et al.* 2014), constituting the Strigomonadinae subfamily, and the phylogenetically distant genus *Novymonas* (Kostygov *et al.* 2016). They have ultrastructural and biochemical features that distinguish them from other monoxenics and human pathogenic trypanosomatids, such as *Trypanosoma cruzi* and *Leishmania* sp., the latter a phylogenetically close genus to symbiont-harboring trypanosomatids. While draft genome assemblies are available for *Angomonas* and *Strigomonas*, there are no complete chromosomal assemblies for any of the four genera of symbiont-harboring trypanosomatids.

The first genome sequencing of *A. deanei* identified the predicted proteins of the protozoan and its symbiont (Motta *et al.* 2013), and two further sequencing efforts have produced fragmented whole genome assemblies (Alves *et al.* 2013; Morales *et al.* 2016). These assemblies have been used to study the loss, transference and interference of genes during symbiosis (Alves *et al.* 2013), as well as to investigate heterologous or endogenous gene and protein expression (Catta-Preta *et al.* 2016; Penha *et al.* 2016; Morales *et al.* 2016). However the structure and full non-coding regions of the genome have not been resolved yet. Here we present a new assembly of the *A. deanei* genome, sequenced using Oxford Nanopore single-molecule technology, which is resolved into 29 chromosomes and reveals several previously unknown special features of the genome. We expect that the new assembly will assist future studies of symbiont-harboring trypanosomatids and other trypanosomatids and monoxenics.

Materials and Methods

Supplemental methods and supplemental file descriptions can be found in File S1.

Organism/Strain origin and derivation

Crithidia deanei Carvalho (ATCC® PRA-265™) (Carvalho 1973), now *Angomonas deanei* (Teixeira *et al.* 2011) was cultivated axenically in Warren's medium (Warren 1960), supplemented with 10% fetal calf serum for 24h at 27 °C. Cells were concentrated to 10^8 by centrifugation (x1200 *g* for 10 minutes) before DNA extraction.

Sample preparation

DNA was extracted from snap frozen pellets containing approximately 10^8 cells using the Nanobind CBB Big DNA Kit - Beta (Circulomics Inc.), according to the manufacturer's guidelines, using the HMW protocol for gram negative bacteria. Briefly, cell pellets were resuspended in 20 µl PBS before addition of equal volumes of proteinase K and cell lysis buffer CLE, and incubation at 55 °C for 20 minutes. Samples were then treated with RNase A for 5 minutes at room temperature, before the addition of buffer BL3 and a further 15 minute incubation at 55 °C. DNA was precipitated with isopropanol, in the presence of the nanobind disk, washed as per the protocol and eluted from the disk into Tris elution buffer. DNA was left overnight at 4°C to fully resuspend before further processing.

Sequencing

For high accuracy, short read sequencing, a paired-end library was prepared using the NEBNext Ultra II FS DNA library prep kit for Illumina (New England Biolabs), according to the manufacturer's instructions, using 100 ng starting DNA, and using 4 cycles of PCR amplification using NEBNext multiplex oligos for Illumina (unique dual index primers; NEB). The library was then subject to 2 x 150 bp sequencing on an Illumina HiSeq 3000 sequencer, at the University of Leeds Next Generation Sequencing Facility.

Long fragment DNA sequencing was performed using an Oxford Nanopore Technologies MinION sequencer. Approximately 500 ng genomic DNA was subject to shearing using the Covaris g-TUBE™ to a mean fragment size of 20 kb, and mixed with an additional 1µg of unsheared genomic DNA. The sequencing libraries were generated using the SQK-LSK109 ligation sequencing kit (ONT). Library preparation started with DNA repair/A-tailing using the NEBNext® Ultra™ II End Repair/dA-Tailing Module, with additional NEBNext FFPE repair enzyme (NEB), using sequential incubations for 30 minutes at 20 °C and then 65 °C. Following clean up with 0.9 X volumes AMPure XP beads (Beckman Coulter), adapters were ligated to prepared DNA ends using NEBNext quick T4 DNA ligase,and the ligation buffer provided in the SQK-LSK109 kit. An additional clean up with AMPure XP beads, including two washes using the ONT Long Fragment Buffer, was performed prior to elution into the buffer provided. The total eluted library was then loaded onto an ONT FLO-MIN109 R9.4.1 flow cell, following the manufacturer's guidelines, and run for 48 hours using MinKNOW software.

Sequence processing and genome assembly

Oxford Nanopore MinION sequences were basecalled with Guppy 3.1.5. Adapters were trimmed using Porechop 0.2.3 (https://github.com/rrwick/Porechop). The raw MinION reads were assembled with Canu 1.8 (Koren *et al.* 2017) (https://canu.readthedocs.io) with option "genomeSize=23m". The raw assembly (File S2) was manually assessed and edited (see File S1 Section 1, File S3 and File S4 for full details). Contig and read alignments for assessments were produced with minimap2 2.17-r941 (Li 2018) (https://github.com/lh3/minimap2) and were inspected using IGV 2.5.3 (Thorvaldsdóttir *et al.* 2013) (http://software.broadinstitute.org/software/igv/). The genome was edited with seqkit 0.10.0 (Shen *et al.* 2016) (https://bioinf.shenwei.me/seqkit/).

The filtered assembly was polished with medaka 0.7.1 (https://github.com/nanoporetech/medaka), using a filtered set of reads longer than 20 kilobases generated with filtlong 0.2.0 (https://github.com/rrwick/Filtlong) using options --min_length 20000 and --keep_percent 90. The medaka-polished assembly was polished with Illumina data three times using Pilon 1.22 (Walker *et al.* 2014) (https://github.com/broadinstitute/pilon) (File S5 contains the polished assembly). Before polishing, Illumina sequences were adapter trimmed with cutadapt 1.16 (Martin 2011) (https://cutadapt.readthedocs.io) for the Illumina Universal Adapter sequence AGATCGGAAGAG. bwa 0.7.17 (Li 2013) (https://github.com/lh3/bwa) and samtools 1.9 (Li *et al.* 2009) (http://www.htslib.org) were used to align Illumina reads to the assembly for polishing.

Validation with PCR

PCRs to validate assembly features (Figure 2; see File S1 Section 2.2 for further

details) were prepared with 10 ng of *A. deanei* DNA in each reaction (or water for

negative controls), mixed with a low ROX SYBR Green master mix and run on a

QuantStudio 3, using a 2 step fast PCR with a 2 second denaturing step at 95 °C

and 30 second anneal and extend step at 60 °C, for 32 cycles. 15 µl of each product

was run on a 2% agarose gel with an Invitrogen 50bp DNA ladder.

Annotation

The previous genome annotation (Motta *et al.* 2013), NCBI accession

GCA_000442575.2, was transferred with flo (Pracana *et al.* 2017). Duplicate

annotations and erroneous proteins were fixed with a custom Python script (File S6;

output in File S7), and the genome was also annotated using Companion version

1.0.2 (Steinbiss *et al.* 2016) (Files S8-S12). Full details of the annotation process are

in File S1 Section 3. Median TriTrypDB statistics were calculated by downloading a

table of genome information from https://tritrypdb.org (Data Summary -> Genomes

and Data Types, downloaded 11 December 2019) and restricting to reference

genomes only.

Genome analysis

Redundancy of genome assemblies (Figure 1) was assessed by aligning genomes

to themselves with minimap2 2.17-r941 (Li 2018) using options -x ava-ont and -a to

output SAM format; alignments were then sorted and indexed with samtools 1.9 (Li

et al. 2009). Copy numbers were calculated with mosdepth 0.2.5 (Pedersen and

Quinlan 2018) (https://github.com/brentp/mosdepth) using option -F 0 to include all

alignments. A script (File S13) was run to calculate the number of bases assigned to each copy number from the mosdepth output.

chr02 was identified as supernumerary (Figure 3) by aligning *A. deanei* Illumina reads used for polishing to the polished assembly with bwa 0.7.17 (Li 2013) and calling variants with freebayes v1.1.0-3-g961e5f3 (Garrison and Marth 2012) (https://github.com/ekg/freebayes) with option -F 0 to accept variants with any alternate fraction. The freebayes VCF was filtered to heterozygote SNPs only using perl and awk, and filtered to only unique regions of the genome using bcftools 1.9 (Li 2011) (https://www.htslib.org) and the mosdepth BED file from the genome redundancy analysis (see above).

Assemblies were assessed with BUSCO v4.0.6 (Seppey *et al.* 2019) (https://busco.ezlab.org) using lineage eukaryota_odb10 and options -m genome and --long.

Data availability

Raw reads are available in the European Nucleotide Archive (project, PRJEB36170; study, ERP119328; sample, ERS4235756; Oxford Nanopore reads, ERR3813852; Illumina reads, ERR3813853). The assembly and annotation are available at accession GCA_903995115; chromosome sequences, LR877145-LR877173. 13 supplemental files (described in File S1) have been uploaded to G3 FigShare.

Results and Discussion

The genome size and chromosome number for *A. deanei* are unknown. Three previous genome assemblies are available (Table 1) (Motta *et al.* 2013; Alves *et al.* 2013; Morales *et al.* 2016). The first (Motta *et al.* 2013) is annotated and available as the reference genome in Ensembl (https://protists.ensembl.org/Angomonas_deanei); it is a reference-guided assembly aimed at identifying protein-coding gene sequences, using a set of 73.8 thousand protein sequences from TriTrypDB 3.3 as a reference, but also including contigs assembled from reads that did not align to the reference. All three assemblies are fragmented and two contain many gaps. They are also of varying sizes (34.1, 23.1 and 19.3 Mb). However, the first assembly contains only 16.6 Mb of unique material, with a further 13.2 Mb of sequence occurring two or three times in the genome (Figure 1). Non-unique material may be accurate expansions of highly repetitive sequence, but could also be extra haplotypic material that should be removed. Of 129 complete eukaryotic BUSCOs found in this assembly, 88 (68% of complete BUSCOs) are duplicated (Table 1). This suggests the first assembly contains many haplotypic sequences, not found to such an extent in the other assemblies, and so the true genome size is likely to be closer to 20 Mb than 35 Mb.

We sequenced 2 051 753 Oxford Nanopore MinION reads containing 13 302 088 880 bp of sequence after adapter trimming (665 times coverage of a 20 Mb genome) with a read N50 of 14 610 bp, and 9 775 722 Illumina HiSeq 3000 read pairs totalling 2 952 268 044 bp (read length 150bp, 148 times coverage of a 20 Mb genome). We assembled the MinION sequence with Canu (Koren *et al.* 2017) to produce an initial

raw genome assembly containing 212 contigs, 27 914 394 bp long (File S2), with a contig N50 of 646 966 bp and no gaps, already an improvement on any existing assembly.

To improve the raw Canu assembly (File S2), we ran the assembly through Tapestry (Davey *et al.* 2020) (https://github.com/johnomics/tapestry) to calculate quality information for each contig (Table S1, File S3), and then filtered and edited the genome based on this information (Table S1, File S1 Section 1, File S4, Figures S1-S13). The assembly contained a symbiont genome in one contig (File S1 Section 1.1), 127 contigs from the kinetoplast minicircle (which were removed from the assembly; File S1 Section 1.2) (Lukes *et al.* 2002) and three contigs from the kinetoplast maxicircle (which were reduced to one unique copy of the maxicircle) (File S1 Section 1.3, Figure S1). As full-length accessory genomes are already publicly available (symbiont: NCBI GenBank GCA_000319225.1 (Motta *et al.* 2013) and GCF_000340825.1 (Alves *et al.* 2013), maxicircle: NCBI GenBank KJ778684.1), these have been removed from our public assembly (NCBI GenBank GCA_903995115), but they are available in our polished assembly included with this paper (File S5).

This left 81 contigs from the nuclear genome. Of these, 49 contigs were extra repeat copies, subtelomeric, or haplotypic and were removed from the assembly, leaving 32 contigs (see Table S1 for details). Manual inspections resolved these contigs to 29 complete or close-to-complete chromosomal sequences, with incomplete contigs explainable due to a translocation (Figures S2-S3), an inversion (Figure S4), and several misassemblies (Figures S5-S10) (all discussed in detail in File S1 Sections

1.4-1.8; genome edits and translocation and inversion haplotypes summarised in Table S2). 56 of 58 contig ends have multiple copies of the trypanosome telomere sequence CCCTAA/TTAGGG (Dreesen *et al.* 2007); although the remaining 2 contig ends do not contain telomeres, the majority of reads that align to these ends do contain telomeres, so these ends are likely to be almost complete (File S1 Section 1.9, Figures S11-S13). The translocation and inversion were validated with read alignments (File S1 Section 2.1, Table S2, Figures S14-S21) and with PCR (Figure 2, Methods, File S1 Section 2.2, Tables S3-S4).

We transferred the first *A. deanei* gene annotation (NCBI genome GCA_000442575.2) to our new nuclear genome assembly, and also predicted new genes and RNAs where possible (see Methods, File S1 Section 3, Files S6-12). The new annotation has 10 365 protein-coding genes (7 199 transferred, 3 166 newly predicted), 59 tRNAs covering all 20 standard amino acids and 1 tRNA for selenocysteine, 26 ribosomal RNAs and 62 non-coding RNAs (45 ncRNA, 14 snoRNA, 3 snRNA). This compares well to other reference genomes in the Kinetoplastid Genomics database TriTrypDB, which have median 8 652 protein-coding genes (median absolute deviation 387) and 110 non-protein coding genes (median absolute deviation 27).

We therefore propose that *A. deanei* has 29 chromosomes, and have named the remaining 29 contigs chr01 to chr29 in order of size, largest first (File S9). These 29 chromosomes make a nuclear genome of 20 976 081 bp, N50 774 942 bp, with no gaps (Table S5). The assembly has a supernumerary chromosome, in common with other trypanosomatids (Downing *et al.* 2011; Rogers *et al.* 2011; Reis-Cunha *et al.*

2018), with chromosome 2 (chr02) having considerably higher read depth than other chromosomes (Figure 3A); the contig has a mixture of 1:1 and 3:1 ratios for SNP calls (Figure 3B), which suggests there are four copies of this chromosome, not two, as for the remaining diploid chromosomes. There is an inversion on chromosome 5 (chr05) between 157.6 kb and 498.1 kb, 340.5 kb long (1.61% of the nuclear genome), containing 173 genes (1.67% of the protein-coding genes in the nuclear genome). Chromosomes 13 and 18 (chr13, chr18) reciprocally translocate at chromosome 13 196.6 kb and chromosome 18 141.1 kb. Figure 4 shows the genome with these features; the lengths of the contigs are summarised in Table S5.

All four public *A. deanei* genome assemblies have very similar BUSCO scores (Table 1), indicating that all four assemblies have similar gene coverage, despite the excess of duplicated genes in the GCA_000442572.2 assembly. The low yet consistent percentages of eukaryotic BUSCO genes across all *A. deanei* assemblies suggest this eukaryotic gene set is not representative of the *A. deanei* genome, rather than suggesting a large number of *A. deanei* genes are missing from all of these assemblies; nevertheless the BUSCO gene set is useful for comparing the four assemblies. Our new assembly matches the gene coverage of the other assemblies, with slightly higher complete single-copy gene sequences, while greatly improving genome contiguity.

We expect our new high-quality, close-to-complete genome assembly, including full chromosome sequences and many non-coding RNAs and non-genic regions, will be useful for future research. It is the first chromosomal assembly for any endosymbiont-bearing trypanosomatid. MicroRNAs has been reported as important

regulators of symbiosis in plants (Hussain *et al.* 2018; Hossain *et al.* 2019), an interesting mechanism that can now be investigated in *A. deanei*, the model of symbiosis in trypanosomatids. Recently a Brazilian patient presenting symptoms of leishmaniasis was non-responsive to available treatments and was found to be infected with a new trypanosomatid phylogenetically related to *Crithidia fasciculata*, a monoxenic trypanosomatid for which only an unpublished draft genome is available (Maruyama *et al.* 2019). There are few monoxenic genomes that can be used as a reference in such cases, as well as in coinfections of pathogens and the so-called non-pathogens (Pacheco *et al.* 1998; Srivastava *et al.* 2010; Ghosh *et al.* 2012; Kraeva *et al.* 2015). This new *A. deanei* assembly can now be used to assist in the identification of new, possibly pathogenic, species. Moreover, a toolkit for reverse genetic studies is being developed for *A. deanei,* which will illuminate more of the biology of the protozoan and its symbiotic relationship with a prokaryote, and the evolutionary leap from symbiont to organelle. Finally, the assembly provides another example of small genomes being almost completely resolved with single runs of long read sequencing, with close examination of the sequences revealing special features of the genome never known before.

Table 1. Summary of *A. deanei* genome assemblies

BUSCO statistics are from a set of 255 eukaryotic benchmarking universal single copy orthologs (BUSCOs) (Seppey et al. 2019). Percentages are calculated from all 255 BUSCOs.

| NCBI ID	GCA_000442575.2	GCA_000482225.1	GCA_001659865.1	GCA_903995115.1
NAME	Angomonas_deanei_Genome	Adea_1.0	Angomonas_deanei_v1.0	Adeanei_nanopore_chromosomes
REFERENCE	Motta *et al.* 2013	Alves *et al.* 2013	Morales *et al.* 2016	This paper
SCAFFOLDS	17 339	5 616	408	29
LENGTH (bp)	34 103 807	23 079 371	19 282 250	20 976 081
SCAFFOLD N50	2 497	11 595	300 798	774 942
GAPS (bp)	30 204	197	1 728 731	0
COMPLETE BUSCOs	129 (50.6%)	125 (49.1%)	127 (49.8%)	128 (50.2%)
COMPLETE, SINGLE-COPY BUSCOs	41 (16.1%)	120 (47.1%)	124 (48.6%)	125 (49.0%)
COMPLETE, DUPLICATED BUSCOs	88 (34.5%)	5 (2.0%)	3 (1.2%)	3 (1.2%)
FRAGMENTED BUSCOs	22 (8.6%)	21 (8.2%)	20 (7.8%)	21 (8.2%)
MISSING BUSCOs	104 (40.8%)	109 (42.7%)	108 (42.4%)	106 (41.6%)

Figures

**Figure 1** Redundancy of genome assemblies
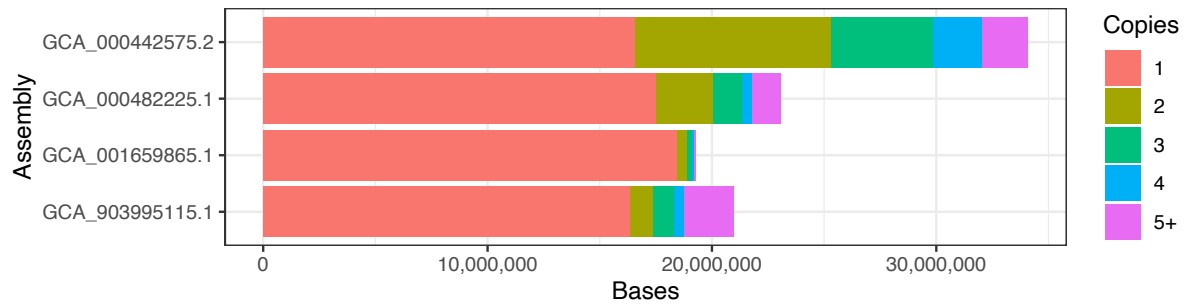
Bars show number of bases in assemblies coloured by copy number. Unique

material has only one copy in the assembly (red). Highly repetitive material has

many copies. Large amounts of material with two or three copies suggests

haplotypic variation has been retained, although some non-unique material is

expected due to common repeats.

Figure 2 PCR validation of special features
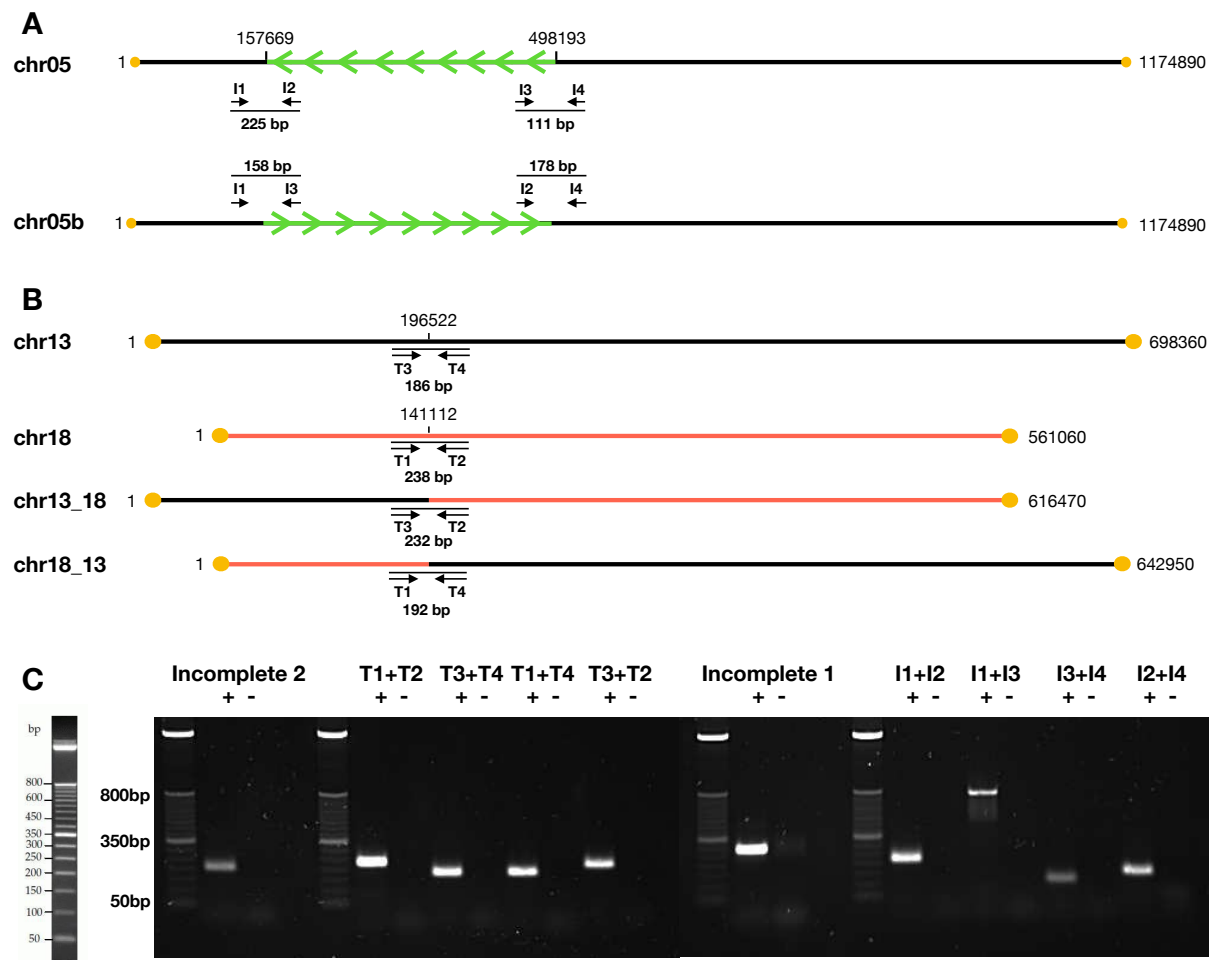
A. Chromosome 5 inversion. Inversion shown with arrows in green. Primers I1, I2, I3 and I4 were designed to span the breakpoints of the two inversion haplotypes chr05 and chr05b. Primer products shown as thin black lines; expected primer product size given below each primer pair. Breakpoint positions in polished genome (File S5) given above chr05. Yellow dots are telomeres. **B.** Chromosome 13/18 translocation. Primers T1, T2, T3 and T4 were designed to span the breakpoints of the four translocation haplotypes chr13, chr18, chr13_18 and chr18_13. Key as in 4A. **C.** PCR products shown via gel electrophoresis against a 50bp Invitrogen DNA ladder (left) for the inversion (I1-I4), the translocation (T1-T4) and two incomplete chromosomes (File S1 Section 2.2, File S5, Table S3, Table S4). '+' and '-' lanes show product and negative control (water) respectively.

Figure 3 Chromosome 2 is a supernumerary chromosome
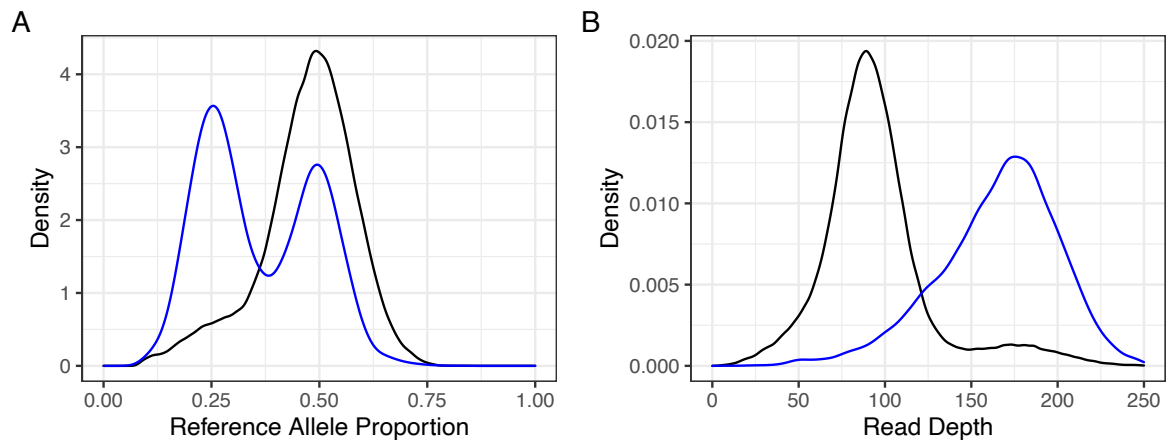
A. Read depths at SNPs in unique regions across the whole nuclear genome (black) or on chr02 only (blue). Chr02 median depth (170) is roughly double the whole genome median depth (88), indicating chr02 may have double the copy number of the rest of the genome. **B.** Proportion of reads with reference allele for all SNPs in unique regions across the genome (black) or on chr02 only (blue). 0.5 (1:1 ratio) indicates two copies; a mixture of 0.25 (1:3 ratio) and 0.5 indicates four copies.

Figure 4 New *A. deanei* nuclear genome assembly
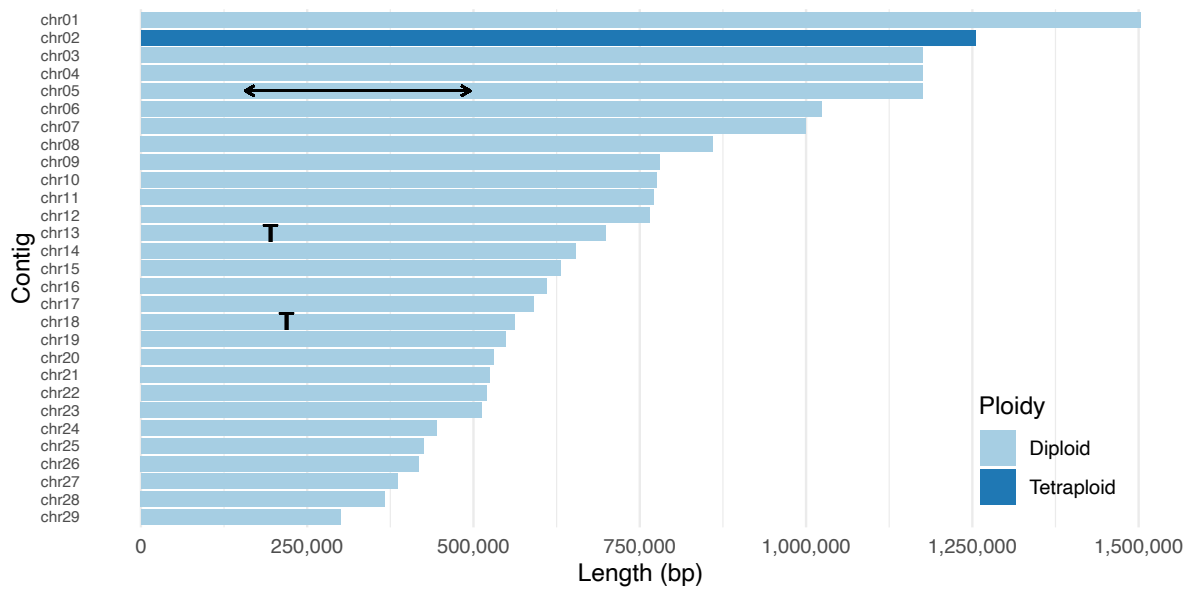
Lengths of chromosomes in new assembly. chr02 is supernumerary (dark blue), chr05 has a 340 kb inversion (line with arrows), and chromosomes 13 and 18 translocate at the points marked 'T'.

Literature Cited

Alves, J. M. P., C. C. Klein, F. M. da Silva, A. G. Costa-Martins, M. G. Serrano *et al.*, 2013 Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. BMC Evol. Biol. 13: 190.

Alves, F. M., N. Olifiers, R. de C. Bianchi, A. C. Duarte, P. M. T. Cotias *et al.*, 2011 Modulating variables of Trypanosoma cruzi and Trypanosoma evansi transmission in free-ranging Coati (Nasua nasua) from the Brazilian Pantanal region. Vector Borne Zoonotic Dis. 11: 835–841.

de Azevedo-Martins, A. C., M. L. Frossard, W. de Souza, M. Einicker-Lamas, and M. C. M. Motta, 2007 Phosphatidylcholine synthesis in Crithidia deanei: the influence of the endosymbiont. FEMS Microbiol. Lett. 275: 229–236.

Carvalho, A. L. M., 1973 Estudos sobre a posição sistemática, a biologia e a transmissão de tripanosomatídeos encontrados em Zelus leucogrammus (Perty, 1834) (Hemiptera, Reduviidae). Revista de Patologia Tropical 2: 223–274.

Catta-Preta, C. M. C., B. Dos Santos Pascoalino, W. de Souza, J. C. Mottram, M. C. M. Motta *et al.*, 2016 Reduction of Tubulin Expression in Angomonas deanei by RNAi Modifies the Ultrastructure of the Trypanosomatid Protozoan and Impairs Division of Its Endosymbiotic Bacterium. J. Eukaryot. Microbiol. 63: 794–803.

Davey, J. W., S. J. Davis, J. C. Mottram, and P. D. Ashton, 2020 Tapestry: validate and edit small eukaryotic genome assemblies with long reads. bioRxiv https://doi.org/10.1101/2020.04.24.059402.

Downing, T., H. Imamura, S. Decuypere, T. G. Clark, G. H. Coombs *et al.*, 2011 Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance.

Genome Res. 21: 2143–2156.

Dreesen, O., B. Li, and G. A. M. Cross, 2007 Telomere structure and function in trypanosomes: a proposal. Nat. Rev. Microbiol. 5: 70–75.

Garrison, E., and G. Marth, 2012 Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN].

Ghosh, S., P. Banerjee, A. Sarkar, S. Datta, and M. Chatterjee, 2012 Coinfection of Leptomonas seymouri and Leishmania donovani in Indian leishmaniasis. J. Clin. Microbiol. 50: 2774–2778.

Hossain, M. S., N. T. Hoang, Z. Yan, K. Tóth, B. C. Meyers *et al.*, 2019 Characterization of the Spatial and Temporal Expression of Two Soybean miRNAs Identifies SCL6 as a Novel Regulator of Soybean Nodulation. Front. Plant Sci. 10: 475.

Hussain, S. S., M. Hussain, M. Irfan, and K. H. M. Siddique, 2018 Legume, Microbiome, and Regulatory Functions of miRNAs in Systematic Regulation of Symbiosis, pp. 255–282 in *Plant Microbiome: Stress Response*, edited by D. Egamberdieva and P. Ahmad. Springer Singapore, Singapore.

Klein, C. C., J. M. P. Alves, M. G. Serrano, G. A. Buck, A. T. R. Vasconcelos *et al.*, 2013 Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. PLoS One 8: e79786.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman *et al.*, 2017 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27: 722–736.

Kostygov, A. Y., E. Dobáková, A. Grybchuk-Ieremenko, D. Váhala, D. A. Maslov *et al.*, 2016 Novel Trypanosomatid-Bacterium Association: Evolution of

Endosymbiosis in Action. MBio 7: e01985.

Kraeva, N., A. Butenko, J. Hlaváčová, A. Kostygov, J. Myškova *et al.*, 2015 Leptomonas seymouri: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with Leishmania donovani. PLoS Pathog. 11: e1005127.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993.

Li, H., 2018 Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Loyola-Machado, A. C., A. C. Azevedo-Martins, C. M. C. Catta-Preta, W. de Souza, A. Galina *et al.*, 2017 The Symbiotic Bacterium Fuels the Energy Metabolism of the Host Trypanosomatid Strigomonas culicis. Protist 168: 253–269.

Lukes, J., D. L. Guilbride, J. Votýpka, A. Zíková, R. Benne *et al.*, 2002 Kinetoplast DNA network: evolution of an improbable structure. Eukaryot. Cell 1: 495–502.

Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12.

Maruyama, S. R., A. K. M. de Santana, N. T. Takamiya, T. Y. Takahashi, L. A. Rogerio *et al.*, 2019 Non-Leishmania Parasite in Fatal Visceral Leishmaniasis-Like Disease, Brazil. Emerg. Infect. Dis. 25: 2088–2092.

Morales, J., S. Kokkori, D. Weidauer, J. Chapman, E. Goltsman *et al.*, 2016

Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid Angomonas deanei. BMC Evol. Biol. 16: 247.

Motta, M. C. M., C. M. C. Catta-Preta, S. Schenkman, A. C. de Azevedo Martins, K. Miranda *et al.*, 2010 The bacterium endosymbiont of Crithidia deanei undergoes coordinated division with the host cell nucleus. PLoS One 5: e12415.

Motta, M. C. M., A. C. de A. Martins, S. S. de Souza, C. M. C. Catta-Preta, R. Silva *et al.*, 2013 Predicting the proteins of Angomonas deanei, Strigomonas culicis and their respective endosymbionts reveals new aspects of the trypanosomatidae family. PLoS One 8: e60209.

Pacheco, R. S., M. C. Marzochi, M. Q. Pires, C. M. Brito, M. de F. Madeira *et al.*, 1998 Parasite genotypically related to a monoxenous trypanosomatid of dog's flea causing opportunistic infection in an HIV positive patient. Mem. Inst. Oswaldo Cruz 93: 531–537.

Pedersen, B. S., and A. R. Quinlan, 2018 Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34: 867–868.

Penha, L. L., L. Hoffmann, S. S. de Souza, A. C. de A. Martins, T. Bottaro *et al.*, 2016 Symbiont modulates expression of specific gene categories in Angomonas deanei. Mem. Inst. Oswaldo Cruz 111: 686–691.

Pracana, R., A. Priyam, I. Levantis, R. A. Nichols, and Y. Wurm, 2017 The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. Mol. Ecol. 26: 2864–2879.

Reis-Cunha, J. L., H. O. Valdivia, and D. C. Bartholomeu, 2018 Gene and Chromosomal Copy Number Variations as an Adaptive Mechanism Towards a Parasitic Lifestyle in Trypanosomatids. Curr. Genomics 19: 87–97.

Rogers, M. B., J. D. Hilley, N. J. Dickens, J. Wilkes, P. A. Bates *et al.*, 2011

Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. Genome Res. 21: 2129–2142.

Seppey, M., M. Manni, and E. M. Zdobnov, 2019 BUSCO: Assessing Genome Assembly and Annotation Completeness, pp. 227–245 in *Gene Prediction: Methods and Protocols*, edited by M. Kollmar. Springer New York, New York, NY.

Shen, W., S. Le, Y. Li, and F. Hu, 2016 SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One 11: e0163962.

Srivastava, P., V. K. Prajapati, M. Vanaerschot, G. Van der Auwera, J. C. Dujardin *et al.*, 2010 Detection of Leptomonas sp. parasites in clinical isolates of Kala-azar patients from India. Infect. Genet. Evol. 10: 1145–1150.

Steinbiss, S., F. Silva-Franco, B. Brunk, B. Foth, C. Hertz-Fowler *et al.*, 2016 Companion: a web server for annotation and analysis of parasite genomes. Nucleic Acids Res. 44: W29–34.

Teixeira, M. M. G., T. C. Borghesan, R. C. Ferreira, M. A. Santos, C. S. A. Takata *et al.*, 2011 Phylogenetic validation of the genera Angomonas and Strigomonas of trypanosomatids harboring bacterial endosymbionts with the description of new species of trypanosomatids and of proteobacterial symbionts. Protist 162: 503–524.

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14: 178–192.

Votýpka, J., A. Y. Kostygov, N. Kraeva, A. Grybchuk-Ieremenko, M. Tesařová *et al.*, 2014 Kentomonas gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. Protist 165: 825–838.

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: an

integrated tool for comprehensive microbial variant detection and genome

assembly improvement. PLoS One 9: e112963.

Warren, L. G., 1960 Metabolism of Schizotrypanum cruzi Chagas. I. Effect of culture

age and substrate concentration on respiratory rate. J. Parasitol. 46: 529–539.