Comparative Genomics of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major

Separation of chromosomes by pulsed field gel electrophoresis

Saccharomyces cerevisiae chromosomes (245-2190 kb). Run conditions: 180 V, 5.1 V/cm, 34 hrs., 120 angle, 60-120 sec. pulse ramp, 0.5X TBE, 1.2% GTG agarose

Lysing cells embedded in agarose was key technique to avoid breakage of large DNA molecules
*T. brucei* cells have 11 megabase chromosomes and around 100 minichromosomes.

*T. cruzi* has ~25 chromosomes

*L. major* has 36 chromosomes
The Tritryp genomes

Mat Berriman
Sanger Center

Najib El-Sayed
TIGR
<table>
<thead>
<tr>
<th>Feature</th>
<th>Leishmania major</th>
<th>Trypanosoma brucei</th>
<th>Trypanosoma cruzi*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>Leishmaniasis</td>
<td>Human African trypanosomiasis (sleeping sickness), Ngana in cattle</td>
<td>Chagas disease</td>
</tr>
<tr>
<td>Ploidy</td>
<td>Diploid (polyploid for some chromosomes)</td>
<td>Diploid</td>
<td>Diploid</td>
</tr>
<tr>
<td>Haploid genome size (Mb)</td>
<td>33</td>
<td>26</td>
<td>55</td>
</tr>
<tr>
<td>Chromosomes</td>
<td>36</td>
<td>11†</td>
<td>~28§</td>
</tr>
<tr>
<td>Genes</td>
<td>8,272</td>
<td>9,068</td>
<td>~12,000</td>
</tr>
<tr>
<td>Pseudogenes</td>
<td>39</td>
<td>904</td>
<td>3,590Ⅲ</td>
</tr>
<tr>
<td>Non-coding RNA genes</td>
<td>911</td>
<td>&gt;556</td>
<td>1,994Ⅲ</td>
</tr>
</tbody>
</table>

*Draft genome only. †Excludes an unspecified number of intermediate and mini chromosomes. §Exact number is not known. ‡The T. cruzi genome sequenced is a hybrid with significant allelic variation. Data from both haplotypes are included.
Distribution of genes and domains among the kinetoplastid parasites

COGs = clusters of orthologous genes
Three-way COGs = shared by all three species
Two-way COGs = shared by two species
Domains = Pfam motifs

Figure 1. Estimated haploid genome sizes in megabases (Mb) and the estimated number of genes per haploid genome. The Venn diagram depicts the distribution of clusters of orthologous genes among and between the three species: Trypanosoma.
The Pfam database contains information about protein domains and families. Pfam-A is the manually curated portion of the database that contains over 10,000 entries. For each entry a protein sequence alignment and a hidden Markov model is stored.

A protein domain is a part of protein sequence and structure that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact three-dimensional structure and often can be independently stable and folded. Many proteins consist of several structural domains. Domains vary in length from between about 25 amino acids up to 500 amino acids in length. Domains often form functional units, such as the calcium-binding EF hand domain of calmodulin.
Synteny blocks are defined as groups of five or more *T. brucei* genes that possess an ortholog on the same *L. major* chromosome.
The 36 different colors in the *T. brucei* (left) panel represent the locations of the indicated synteny blocks in the 36 chromosomes of *L. major*, and the 11 colors in the *L. major* (right) panel depict the locations of the indicated synteny blocks in the 11 chromosomes of *T. brucei*. 
Tritryp subtelomeric regions

The first new insight of the genome analysis is that most sequenced silent VSGs are defective. Of 806 analyzed, only 57 (7%) are fully functional (that is, encode all recognizable features of known functional VSGs), whereas 9% are atypical (complete genes possibly encoding proteins with inconsistent VSG folding or posttranslational modification), 66% are full-length pseudogenes (with frameshifts and/or in-frame stop codons), and 18% are gene fragments, most of which encode C-terminal domains.

Another new insight is that almost all VSGs form arrays, numbering 3 to 250 (pseudo)genes, and that most of these are subtelomeric and are oriented away from the telomere.

Members of up to 11 families of ESAGs are normally associated with VSGs in the polycistronic BESs. All BESs appear to harbour an ESAG6 and ESAG7, and most have a total of five to 10 ESAGs and pseudo-ESAGs.
Motility is mediated by a single flagellum

Direction of motility

Insect Form

Bloodstream Form
Tissue tropisms underlie parasite development and disease pathogenesis

In the tsetse

Stage 1 (Bloodstream)
- Waves of fever, chills, achy joints
- Antigenic Variation
- Immunosuppression and anemia
- Can be asymptomatic

Stage 2 (CNS)
- Neurologic dysfunction, confusion, personality changes
- Somnolence and insomnia
- Secondary Infections
- Coma
- Death
- 100% Fatal if untreated
- Drug resistance

In the human host

Parasites penetrate the blood-brain barrier

Foregut Midgut PV Ectoperitrophic Space
Salivary Gland PM

Salivary Gland

Foregut PV Midgut
Roles for the *Trypanosoma* flagellum

- Motility
  - development, pathogenesis, drug resistance
- Host cell attachment
- Cell division and morphogenesis
  - cell division
  - cell size, shape and form
  - organelle positioning and inheritance
- Immune Evasion
- Endocytosis and secretion
- Sensory organelle
Procyclic or insect form of *T. brucei*
The Flagellar Apparatus of *T. brucei*

**Conserved Features:**
- 9 + 2 Axoneme, Outer and Inner Dyneins, Radial Spokes, Nexin Links

**Novel Features:**
- Attached to cell body (FAZ), PFR, Stationary Central Pair, Expanded Gene Families, e.g. DRC, Nexins, Tip-to-base beat
Simultaneous dynein inactivation $\implies$ paralysis

Dynein activity must be coordinately regulated

Dynein regulatory complex or DRC

- DRC is part of a signal transduction pathway that regulates Dynein
- Trypanin = only known DRC subunit in tryps
What’s in a flagellum?

- 9+2 Axoneme (~250 proteins)
- Membrane/Matrix (~300 proteins)
- PFR (trypanosome specific)

What are the genes/proteins?
- Bioinformatics
- Proteomics/mass spec
- Mutant screens

What do these proteins do and how do they do it?
- Inducible RNAi
- Gene knockouts
- Site-directed mutants

How are these proteins assembled into supramolecular complexes to make a flagellum?
- EM: TEM, SEM, CryoET
- X-ray crystallography
Phylogenetic profiling identifies novel motility genes

88 Proteins

Confirm Homology
- Reciprocal BLAST
- Individual protein alignments

50 CMF Proteins
Tet-inducible RNAi (42)

Phenotypic analysis
- Motility assays
  - Sedimentation, Video Microscopy
- Ultrastructure
- Localization
  - Inducible GFP tags and Fluorescence Microscopy, Biochemical Fractionation

WT, Forward
“Dizzy” Reverse
Paralyzed Lethal

Motile Flagella: intersect
T. brucei
Non-motile /Lack flagella: subtract

Phylogenetic profiling identifies novel motility genes
A

Class 1
"Unaffected"
10/41 (24.4%)

Uninduced
Induced

Class 2
Mild; tiny clumps
6/41 (14.6%)

Class 3
Moderate; large clumps; slow growth
14/41 (34.1%)

Class 4
Severe; large clumps, lethal
11/41 (26.8%)

B

Class 1 (Unaffected)
RNAi Target Tet Probe
TbCMF 15 + +
TbCMF 33 + +

Class 2 (Mild)
TbCMF 2 + +
TbCMF 3 + +

Class 3 (Moderate)
TbCMF 46 + +
TbCMF 49 + +

Class 4 (Severe)
TbCMF 5 + +
TbCMF 9 + +
Flagellum mutants have defective organelle positioning and cytokinesis.
Experimental Proteomics Approach identifies 331 *T. brucei* flagellar proteins

Isolation of axoneme with associated PFR and basal body

552 non-redundant proteins
Mass spec validation confirmed 380 proteins
Removed ribosomal proteins - 331 proteins
*In silico* screen - 208 proteins are trypanosome specific

For each comparison, green indicates present in both; orange/red indicates present in opposing genome but not opposing proteome; blue/yellow indicates no homologue in opposing genome.
Determined human chromosomal loci for homologues

34 genes mapped to 25 loci where diseases suggestive of ciliary dysfunction mapped

***Table: Ciliary disease genes in the TbFP***

<table>
<thead>
<tr>
<th>Trypanosoma brucei accession number</th>
<th>Homo sapiens accession number</th>
<th>Name</th>
<th>Known flagellar component?</th>
<th>Disease</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tb11.02.0760</td>
<td>Q0NYC9</td>
<td>DNAHC11</td>
<td>dynein heavy chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb11.02.2040</td>
<td>AAH30583</td>
<td>DNAI1</td>
<td>dynein intermediate chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb11.02.10060</td>
<td>NP_038575</td>
<td>DNAHC11</td>
<td>dynein heavy chain</td>
<td>situs inversus</td>
<td>Bartoloni et al 2002</td>
</tr>
<tr>
<td>Tb927.3.930</td>
<td>NP_001360</td>
<td>DNAHC5</td>
<td>dynein heavy chain</td>
<td>Hydrocephalus</td>
<td>Ibanez-Tallon et al 2002</td>
</tr>
<tr>
<td>Tb927.3.2310</td>
<td>NP_089623</td>
<td>TbPACRGA</td>
<td>dynein heavy chain</td>
<td>Male infertility</td>
<td>Lorenzetti et al 2004</td>
</tr>
<tr>
<td>Tb09.211.1470</td>
<td>NP_089623</td>
<td>TbPACRB</td>
<td>dynein heavy chain</td>
<td></td>
<td>Sapiro et al 2002</td>
</tr>
<tr>
<td>Tb10.70.1720</td>
<td>BAA02648</td>
<td>KIAA1410</td>
<td>dynein heavy chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb11.01.0360</td>
<td>BAA02648</td>
<td>KIAA1410</td>
<td>dynein heavy chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb927.2.12670</td>
<td>NP_038575</td>
<td>SPAG6/PF16</td>
<td>central apparatus protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb927.3.1060/40</td>
<td>NP_060570</td>
<td>Rib72/EFHC1</td>
<td>protofilament ribbon component</td>
<td>Juvenile myoclonic epilepsy</td>
<td>Suzuki et al 2004</td>
</tr>
<tr>
<td>Tb927.6.2560</td>
<td>NP_060570</td>
<td>Rib72/EFHC1</td>
<td>protofilament ribbon component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb10.6.15.2920</td>
<td>NP_060570</td>
<td>Rib72/EFHC1</td>
<td>protofilament ribbon component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb11.47.0006</td>
<td>NP_060570</td>
<td>Rib72/EFHC1</td>
<td>protofilament ribbon component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tb10.61.2870</td>
<td>AAH43348</td>
<td>RP2</td>
<td>membrane protein</td>
<td>Retinitis pigmentosa</td>
<td>Schwahn et al 1998</td>
</tr>
<tr>
<td>Tb10.70.1880</td>
<td>NP_878899</td>
<td>Scorpion</td>
<td>NO</td>
<td>Polycystic kidney disease</td>
<td>Sun et al 2004</td>
</tr>
</tbody>
</table>
Physical connection between basal body of flagellum and kDNA

Hypotonic lysis of Leishmania tarentolae cells

Addition of 0.25 M sucrose

(The early years!)
Kinetoplasts are attached to basal bodies after hypotonic or detergent lysis (1991)
Flagella and kinetoplast organization in cell cycle of T. brucei

Antibodies against paraflagellar rod (ROD1) or basal body BBA4) DNA stained with DAPI
Effect of inhibitors of kDNA segregation or basal body segregation on kinetoplast segregation

a-d ethidium bromide and teniposide (inhibit topoisomerase)

e-f ansamitocin (anti-microtubule)
"Exclusion zone" between basal body and outer mitochondrial membrane

Unilateral filaments only on flagellar face of the kinetoplast - link it to membrane

TAC

Tripartite Attachment Zone
Schematic diagram of the TAC complex in trypanosomes
Kinetoplast DNA replication and segregation occurs simultaneously in *T. brucei* procyclic cells.
Acriflavine blocks the postreplication mitochondrial genome linkage to the TAC.
Cell cycle of the procyclic form of T. brucei
The flagellum plays key roles in the morphogenesis, motility and pathogenicity of trypanosomes. Its basal body complex defines both the site of the flagellar pocket, the unique region for surface membrane traffic, and the position of the kinetoplast, the single mass of mitochondrial DNA. The nuclear-external left-handed helical rows of the distal flagellum along the cell body also reflects the internal axis and polarity of cytoskeletal elements involved in organelle positioning and cytokinesis. Duplication or remodeling of these cellular features occurs in a strict temporal order within the cell cycle. Clarifying these events has revealed genetic aspects of morphogenesis of eukaryotic flagella and cells as well as highlighting the requirement for a three-dimensional understanding of complex cell processes.

The immature basal body "bud" (shown in red), which was termed during the latest cell cycle, is included into a perinuclear region by the mitotic signal. When the nuclear envelope re-forms, it is guided into a new cell division cycle.

The immature basal body "bud" (shown in red), which is capable of initiating a new flagellum. Two new immature basal bodies (yellow) are formed close to the next two mature basal bodies.

The new flagellum is extended out of the flagellum pocket and positioned alongside the old flagellum. Ultimately, a new flagellum pocket forms, and basal bodies move apart, thereby segregating the kinetoplast DNA connected to the posterior end of the basal body. The chitin of the new flagellum remains attached by the perinuclear structure to the side of the old flagellum, hence influencing the helical internal cytoplastin. Flagellum duplication, after initiation from the anterior end of the cell, ultimately segregates all displaced fibrils and molecular components into daughter trypanosomes.
