

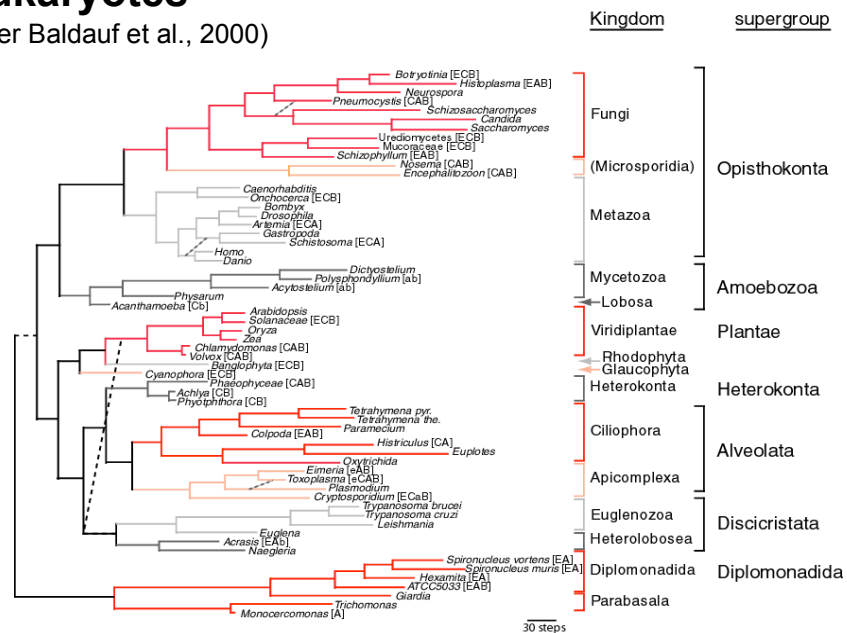
# Eukaryotic Pathogen genomes:

November 15

## Lecture by David Sullivan dsulliva@jhsph.edu

### Eukaryotes

(after Baldauf et al., 2000)



Over 40 species of protozoa and over 100 helminths infect people.

SubKingdom Protista-Protozoa

Phylum Sarcomastigophora

Subphylum Mastigophora-flagellates-*Giardia*, *Trichomonas*,

*Kinetoplastida*-*Leishmania* and *Trypanosoma*

Subphylum Sarcodina-amebae move by pseudopodia- *Entamoeba*,

Phylum Apicomplexan-all obligate parasites with apical organelles

*Plasmodium*, *Toxoplasma*, *Cyptosporidium*, *Cyclospora*, *Babesia*

Phylum Microspora-*Microsporidia*

Phylum Ciliophora-ciliates *Balantidium coli* is the only parasite of humans

SubKingdom Metazoa

Phylum Nematoda- round worms- hookworm, *Ascaris*, *Strongyloides*,

pinworm & whip-worm, filaria

Phylum Platyhelminthes-flatworms

Class Trematode-flukes and Schistosoma

Class Cestode-tapeworms



Max Brodel's 1896  
Cartoon of Osler  
Chasing Away Microbes

Malaria infected RBC's

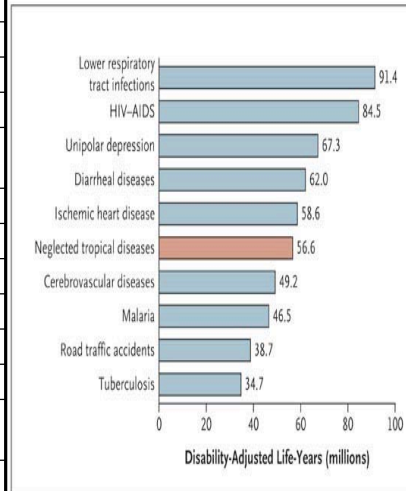


## Neglected Tropical diseases

NEJM 357:1018-27 Sept 2007

Disease	Global Prevalence (millions)	Population at risk
Ascariasis	807	4.2 billion
Trichuriasis	604	3.2 billion
Hookworm	576	3.2 billion
Schistosomiasis	207	779 million
Lymphatic filariasis	120	1.3 billion
Onchocerciasis	37	90 million
Dracunculiasis	0.01	ND
Trachoma	84	590 million
Leprosy	0.4	ND
Leishmaniasis	12	350 million
Chagas disease	9	25 million
Human African Trypanosomiasis	0.3	60 million
Amoebiasis	48	

### The 10 Leading Causes of Life-Years Lost to Disability and Premature Death



### Global impact of parasites

More parasitic organisms than non parasitic organisms in the world. Most parasitic infections cause chronic disease shortening lifespan and increasing sick days and robbing nutrition from young children.

A **hookworm** consumes about 100 ul of blood a day multiplied by all infections equals **130,000 liters of blood a day** or the blood of 26,000 people that is completely sucked out each day by hookworms alone.

More than 80,000 tons of Ascaris eggs are shed in a year. "A huge nutritional beneficence bestowed upon ascaris by the human host" Stoll, Norman "This wormy world" J. Parasit. 1947 33:1-18. In 1950 over 2.2 billion helminth infections for 2.1 billion people. 85% of all helminth infection are due to **"ineffective insulation from our excretory products"**.

Protozoan diseases-African sleeping sickness and Kala Azar are universally fatal unless treated with present ineffective therapy. Many are severely disfiguring such as espundia-mucocutaneous leishmaniasis and elephantiasis-filariaria. Blinding by onchocerciasis affected most adults in African villages. Wild animal reservoirs for *Trypanosoma* prevent the use of 4.5 million acres of otherwise usable african land.

A **single day of fever from malaria consumes 5,000 calories or about two days of hard labor**. The caloric intake in endemic areas is only about 2,200. About 30% of total energy yield of grain production is wasted on parasitic diseases. The average bout of malaria prevents about 4 working days

## Protozoans

Apicomplexans-

<http://www.apidb.org/apidb/>

Toxoplasma, Cryptosporidium,  
Plasmodium, Babesia

Trypanosoma brucei and cruzi

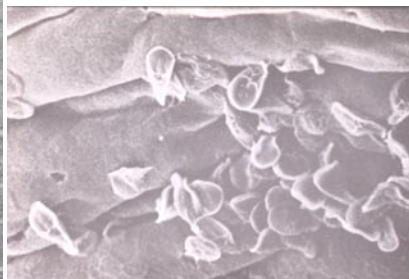
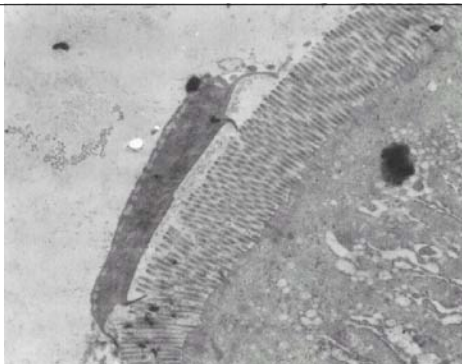
Trichomonas

Entamoeba histolytica

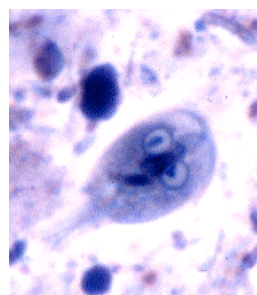
Giardia lamblia

EupathDB

<http://eupathdb.org/eupathdb/>



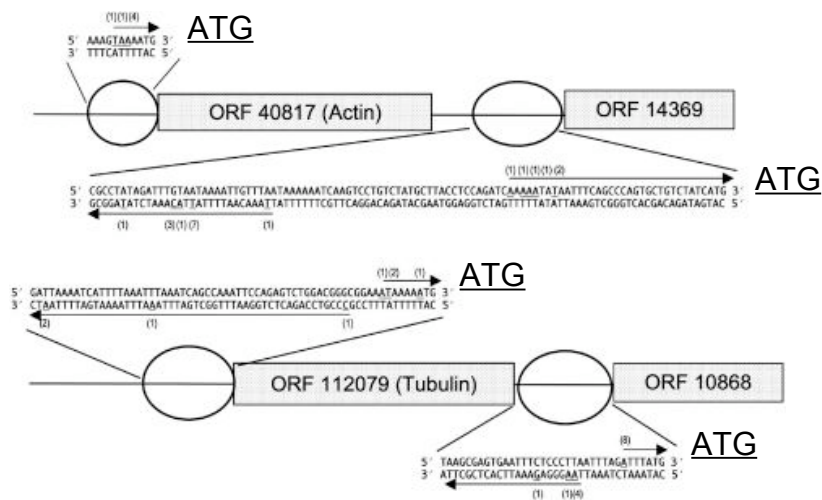
*Giardia lamblia*



## Individual eukaryotic genomes: Protozoans at the base of the tree

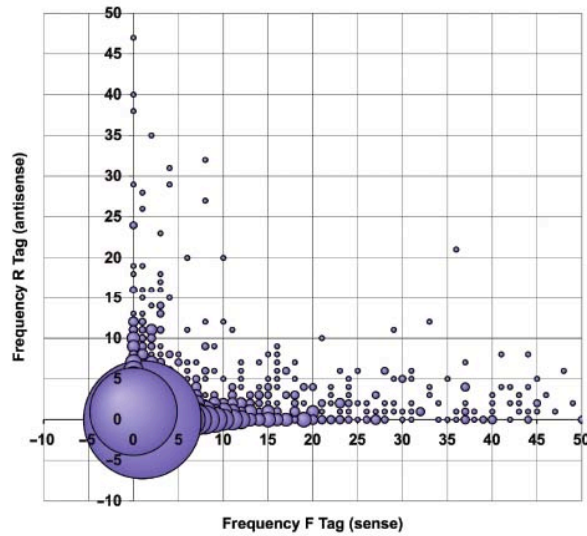
*Giardia lamblia* is a water-borne parasite  
Disease relevance: giardiasis (causes diarrhea)  
Distinguishing features: lack of mitochondria, peroxisomes;  
polyploid (has two nuclei in each cell)  
Genome size: 12 Mb  
Chromosomes: 5 (range 0.7 to >3 Mb)  
Website: <http://www.giardiadb.org/giardiadb/>  
(sequencing in progress)  
No extranuclear mitochondrial DNA  
The genome has just three retrotransposons.  
Also, it appears to have a single intron (ferredoxin gene).  
Short intergenic regions  
Abundant antisense transcripts-20% of cDNAs in libraries  
represent polyadenylated sterile (antisense) transcripts

### Bidirectional transcription in *Giardia*



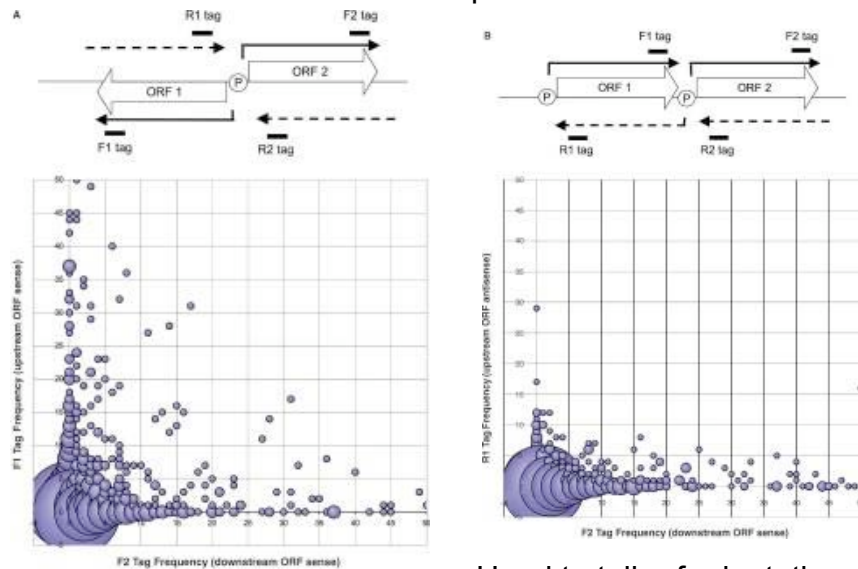
Three promoter sites have bidirectional transcription

Nucleic Acids Res. 2007 April; 35(8): 2544–2553.



Serial analysis of Gene Expression (SAGE) tags were tabulated for sense or antisense for each Open reading Frame (ORF). The loaction was expressed as number of sense to antisense per ORF

### Bidirectional transcription in *Giardia*

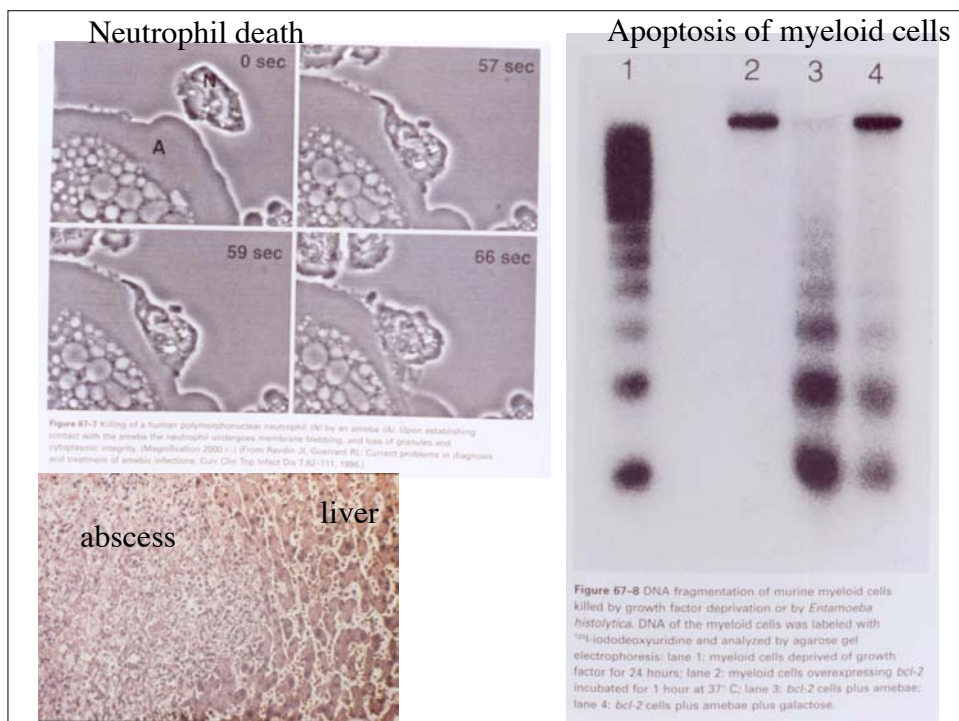
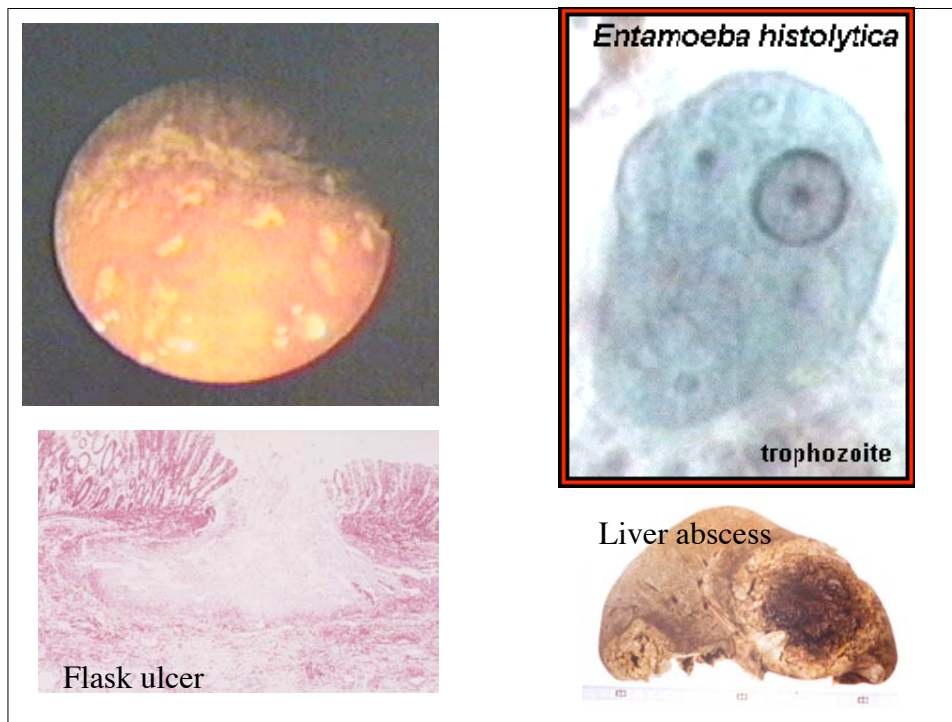


Head to head orf orientation

Head to tail orf orientation

Nucleic Acids Res. 2007 April; 35(8): 2544–2553.





Entamoeba histolytica

[http://www.sanger.ac.uk/Projects/E\\_histolytica/](http://www.sanger.ac.uk/Projects/E_histolytica/)

The genome of the protist parasite Entamoeba histolytica. Nature 2005;433;7028;865-8

*Entamoeba histolytica* is the causal agent of amoebic dysentery and amoebic liver abscess, with an estimated 40 million cases worldwide and 100,000 deaths.

Amitochondriate protozoa

Genome size: 24 Mb

Chromosomes: 14 (range 0.6 to 3.3 Mb)

functionally tetraploid

Lateral gene transfer from bacteria

Epigenetic silencing

no extranuclear mitochondrial DNA

tRNA's encoding amino acids

tRNA genes in clustered arrays making up about 10% of genome.

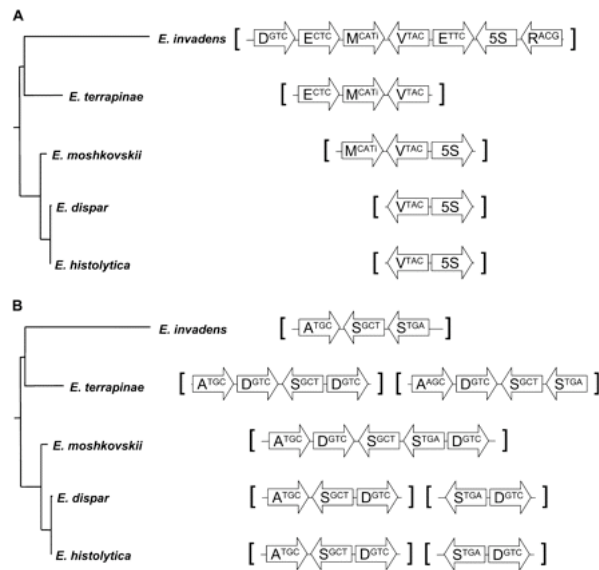
4,500 copies (10 x that of human genome)

Intergenic regions have simple sequence repeats 7-12 bp but up to 44 bp.

*E. histolytica* lacks mini/microsatellites elsewhere.

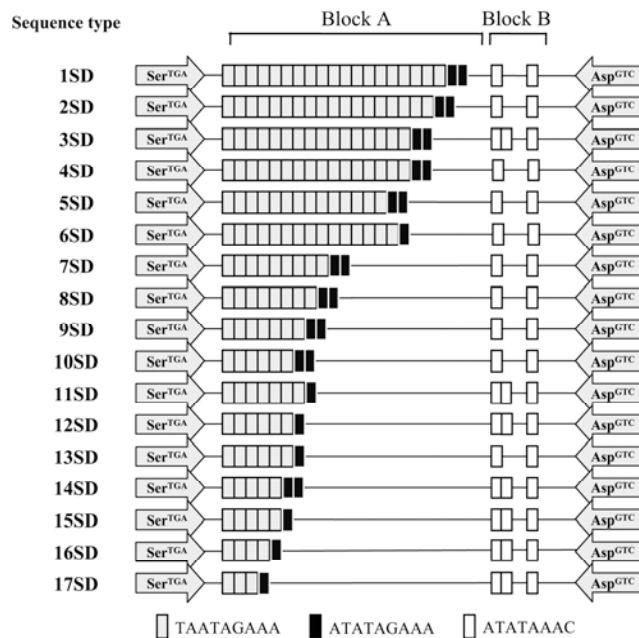
tRNA-linked *E. histolytica* short tandem repeats (STRs) form part of a larger unit that is itself tandemly arrayed





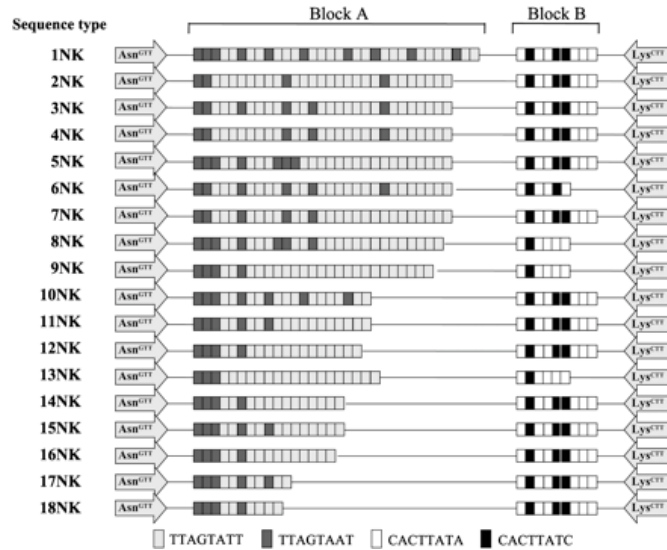
(A) The array unit organization involving the gene encoding tRNA Val<sup>TAC</sup>.  
 (B) The array unit organization involving the genes encoding tRNAs Ser<sup>GCT</sup> and Ser<sup>TGA</sup>.

Molecular Biology and Evolution 2008 25(1):187-198



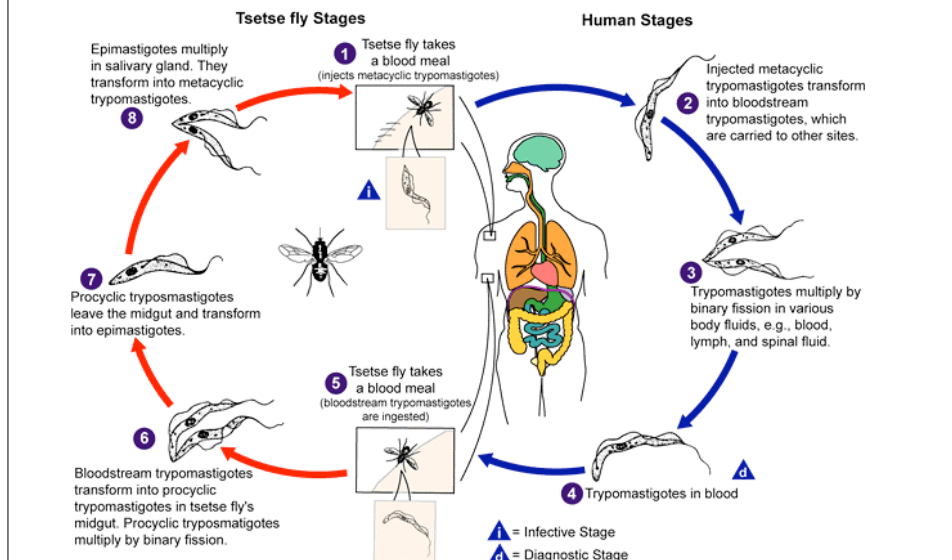
Molecular Biology and Evolution 2008 25(1):187-198

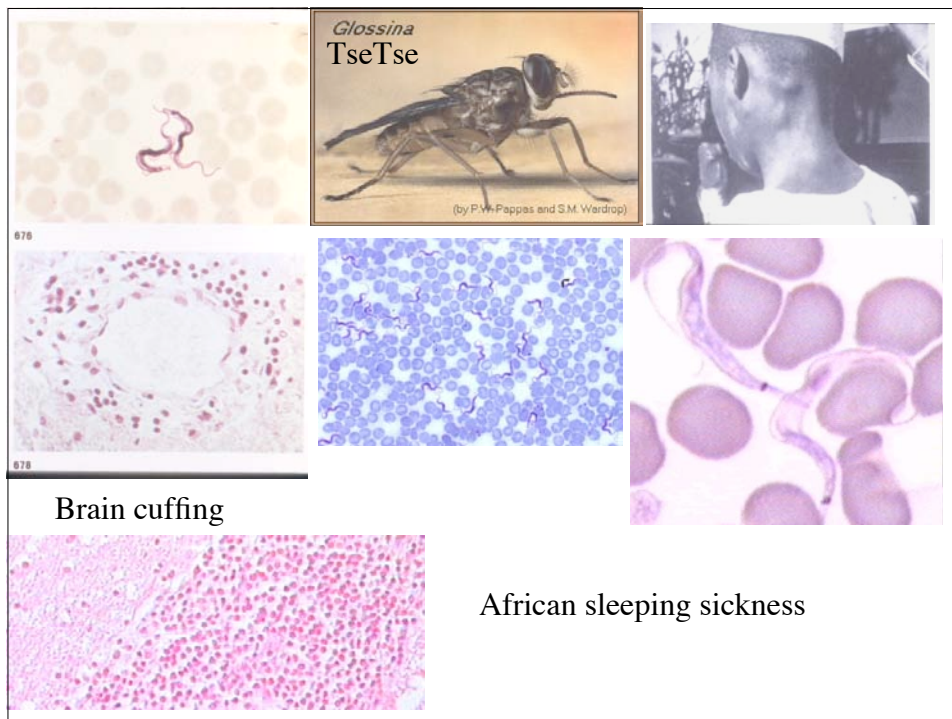
Intraspecific differences in STR organization in the *Entamoeba histolytica* N-K2 intergenic region.



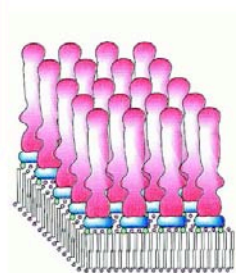
Molecular Biology and Evolution 2008 25(1):187-198

## African Sleeping Sickness- *Trypanosoma brucei rhodesiense* *Trypanosoma brucei gambiense*



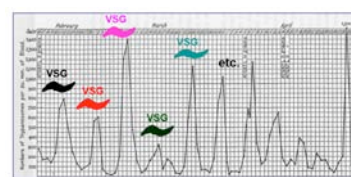


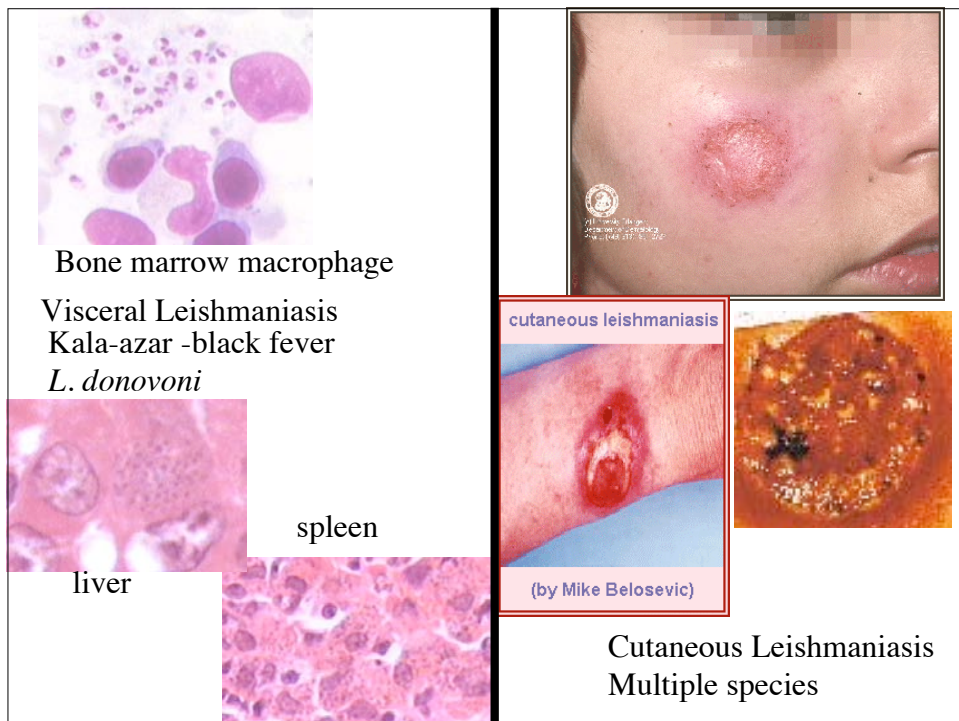
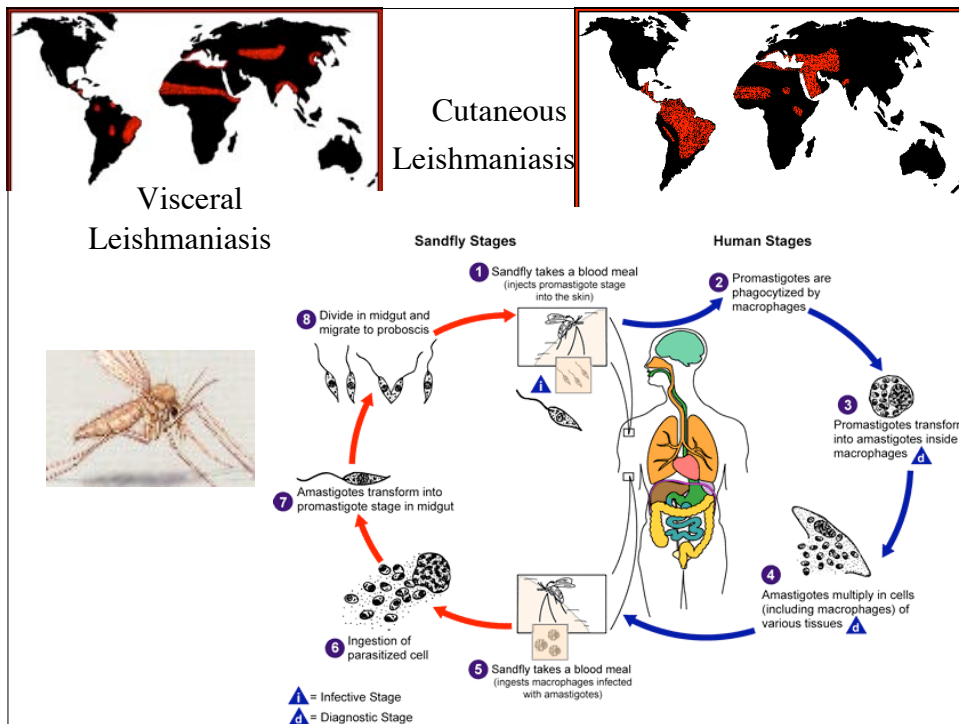
### Antigenic variation and The Variant Surface Glycoprotein (VSG)



Mehrlert et al, 1998

Only one kind expressed at a time  
 $10^7$  molecules per cell (10% of total protein)  
 Forms a uniform coat covering trypanosome  
 ~500 amino acids and 10% carbohydrate  
 Linked to plasma membrane by GPI anchor  
*Target of host's antibody response*





## Individual eukaryotic genomes: *Leishmania*

*Leishmania major* causes leishmaniasis

Intracellular pathogen of the immune system targeting macrophages and dendritic cells. Affects the populations of 88 countries worldwide with symptoms ranging from disfiguring cutaneous and muco-cutaneous lesions that can cause widespread destruction of mucous membranes to visceral disease affecting the haemopoietic organs.

Genome size: 34 Mb

Chromosomes: 36 (range 0.3 to 2.5 Mb)

Genes: about 9800 (all are being manually annotated)

Website: [http://www.sanger.ac.uk/Projects/L\\_major/](http://www.sanger.ac.uk/Projects/L_major/)

*Leishmania* chromosome 1 has 79 protein-coding genes.

The first 29 (from the left telomere) are all transcribed from one strand, and the next 50 from the opposite strand.

**Table 1.** General features of the Trityp genomes. We found 5812 syntenic three-way COGs and 346 nonsyntenic three-way COGs. Mbp, mega-base pairs; NC, not computed.

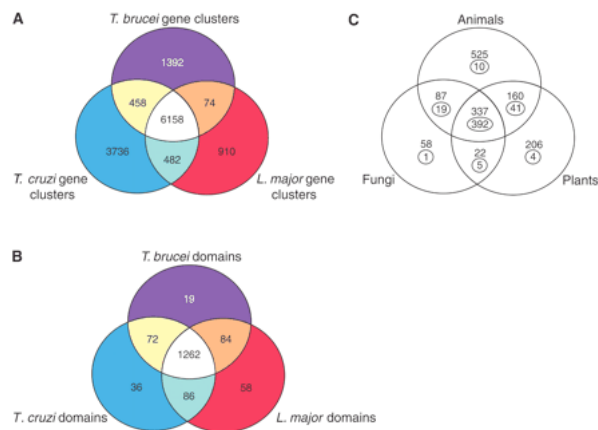
	<i>T. brucei</i>	<i>T. cruzi</i>	<i>L. major</i>
Haploid genome size (Mbp)	25*	55	33
No. of chromosomes (per haploid genome)	11*	~28†	36
No. of genes (per haploid genome)	9068‡	~12,000§	8311
Total regions with synteny blocks (Mbp)	19.9	NC	30.7
Mean CDS size (bp) in syntenic three-way COGs	1511	1457	1731
Mean inter-CDS size (bp) between syntenic three-way COGs	721	561	1431

\*Excluding ~100 mini- and intermediate-sized chromosomes (totaling ~10 Mb).  
and homologs can differ substantially in size.

‡Includes 904 pseudogenes.

||Includes 34 pseudogenes.

†The exact number is not known  
§The exact number of haploid genes



Distribution of genes and domains among the kinetoplastid parasites and other organisms. **(A)** Gene distribution, calculated with the use of Jaccard-filtered COGs (2). **(B)** Domain distribution calculated with the use of TIGRFAMs and Pfam domains. The numbers indicate all domains that score above the trusted cutoff after manual curation with the three-way genome comparisons (2). **(C)** Pfam domain distribution among the three kingdoms of life and the Trityps. Numbers in small circles indicate the number of domains that occur more than once in Trityp parasite genomes. The numbers above the small circles indicate Pfam domains that are not present in the Trityps.

*Science* 15 July 2005: Vol. 309.

## Individual eukaryotic genomes: trypanosomes

*Trypanosoma brucei* causes sleeping sickness (Africa)

Distinguishing features: transmitted by tsetse flies

Genome size: 29 Mb (+/- 25% in various isolates) 11,000 genes

Chromosomes: 11 (range 1 to >6 Mb); also has intermediate chromosomes (200-700 kB) and 100 linear minichromosomes (30-150 kb)

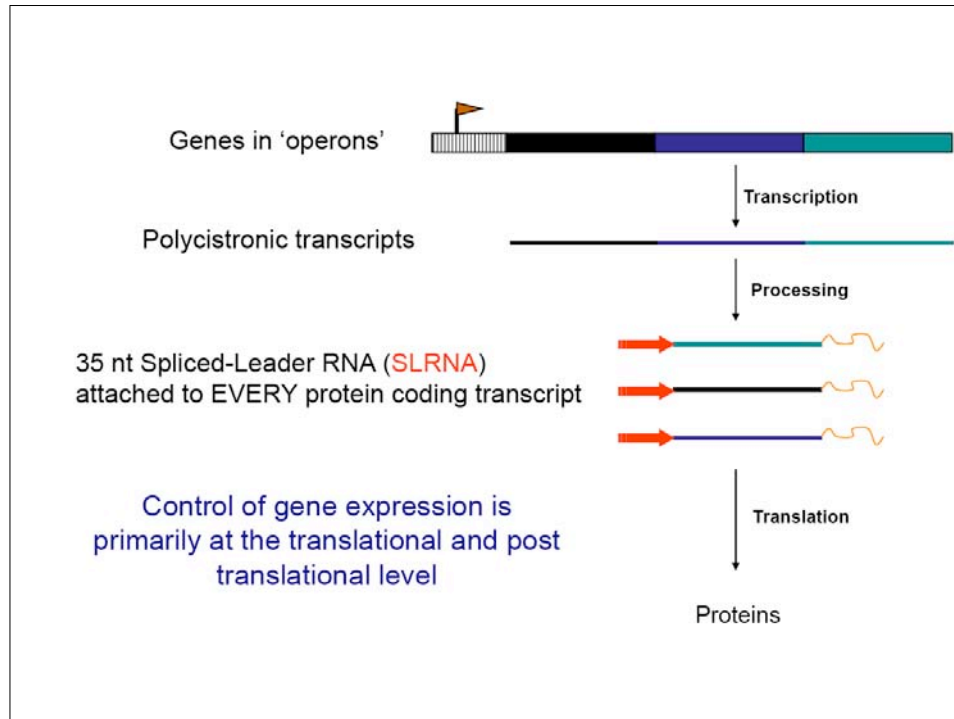
Chromosomes do not condense at metaphase

Only a single intron in the poly (A) polymerase gene

Intermediate and minichromosomes serve as repositories for VSGs

Trypanosomes have kinetoplast DNA (circular rings of mitochondrial DNA)(studied by Paul Englund's lab here).





Morris *et al.*, (2001) Int. J. Parasitol. **31**, 453-458.

Guilbride & Englund (1998) J. Cell Science **111**, 675-679.

Drew & Englund (2001) J. Cell Biol. **153**, 735-743.

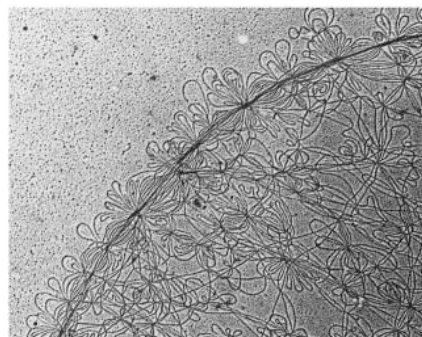
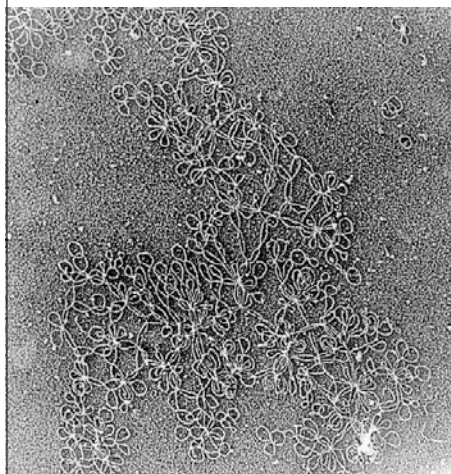
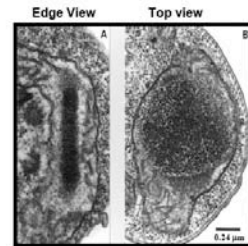
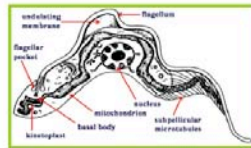


Fig. 1. EM showing a segment of a purified *C. fasciculata* kinetoplast DNA network. Small loops are the 2.5 kb minicircles, and long strands threading through the network interior are parts of the 38 kb maxicircles. EM by David Pérez-Morga

## ORGANIZATION

Mitochondrial genome

*Kinetoplast DNA*



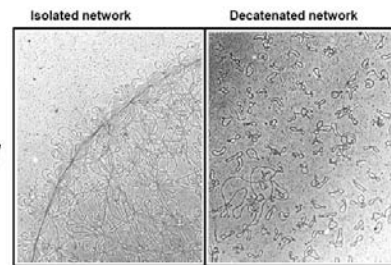
Chain-mail network of about 5000 minicircles and 25 maxicircles

One network per mitochondrion

Maxicircles - mitochondrial gene products

Maxicircle genes are cryptic. Need 'editing'

Minicircles - guide RNAs that control 'editing'



## MANAGEMENT: DNA topoisomerases

Catalyze topological changes in DNA

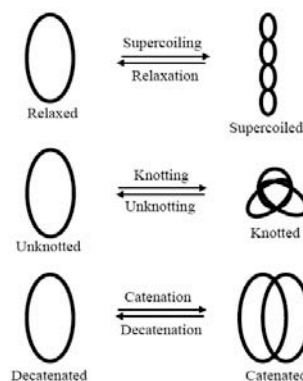
Needed for many nuclear processes - replication, transcription, recombination, maintenance of nuclear structure

Targets of drugs like Ciprofloxacin (anti-bacterial) and etoposide (anti-cancer)

Two types: Type I and Type II

Humans have six total

Trypanosomes encode SEVEN – two (?) dedicated to the mitochondrion!



Topoisomerase targeting drugs can kill trypanosomes *in vitro*

## RNA editing

Cytochrome oxidase subunit III



Feagin et al., 1988

Addition of residues not encoded in original DNA sequence

### Deletion of DNA-encoded residues from transcribed RNA

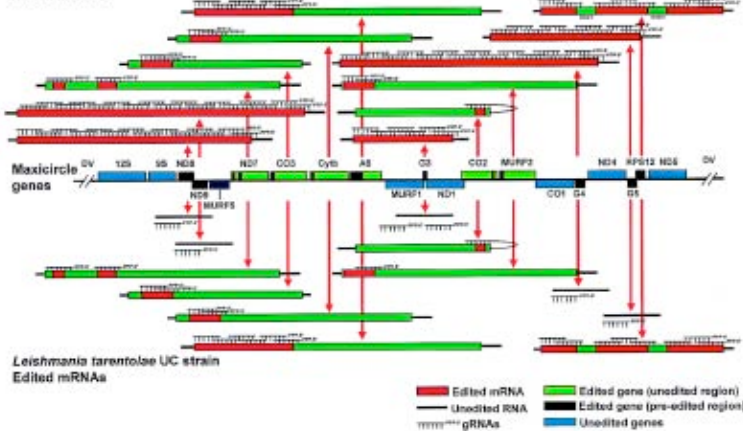
Most mitochondrial genes (on kDNA maxicircles) subject to editing

Governed by guide RNA molecules (gRNA) encoded on kDNA minicircles

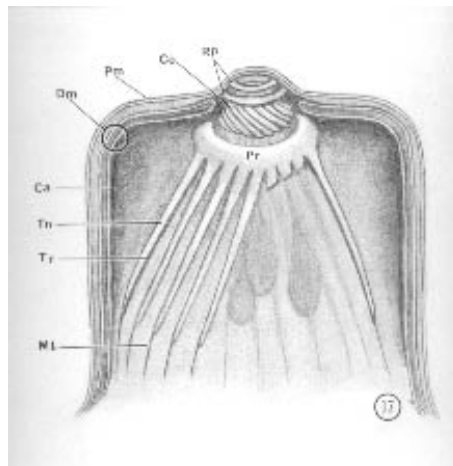
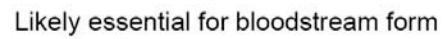
## RNA editing in trypanosome mitochondria

*Leishmania tarentolae* LEM125 strain

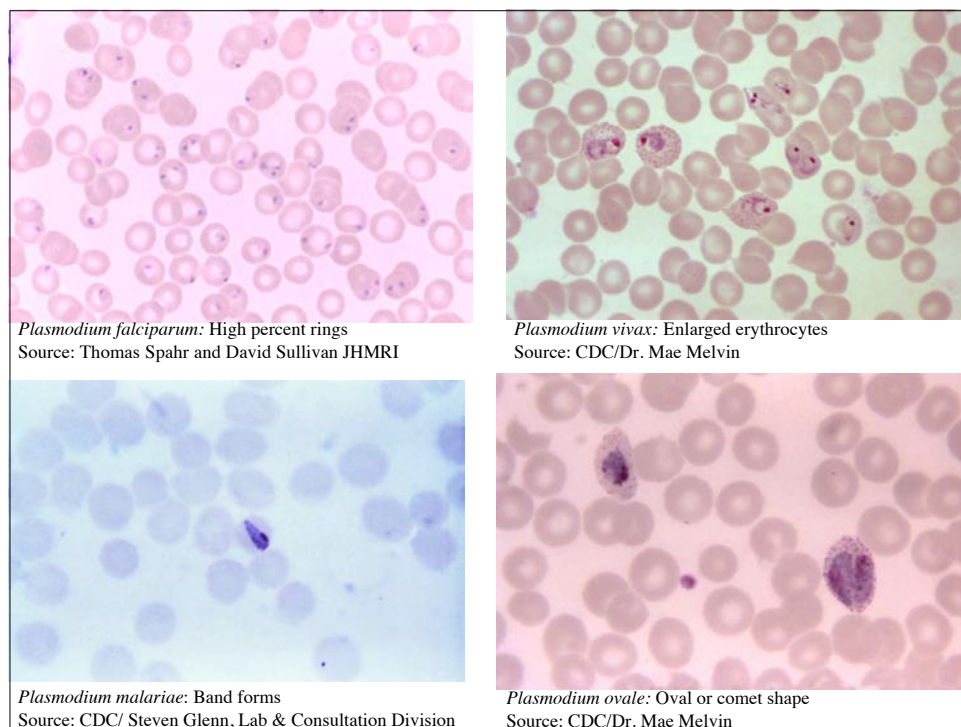
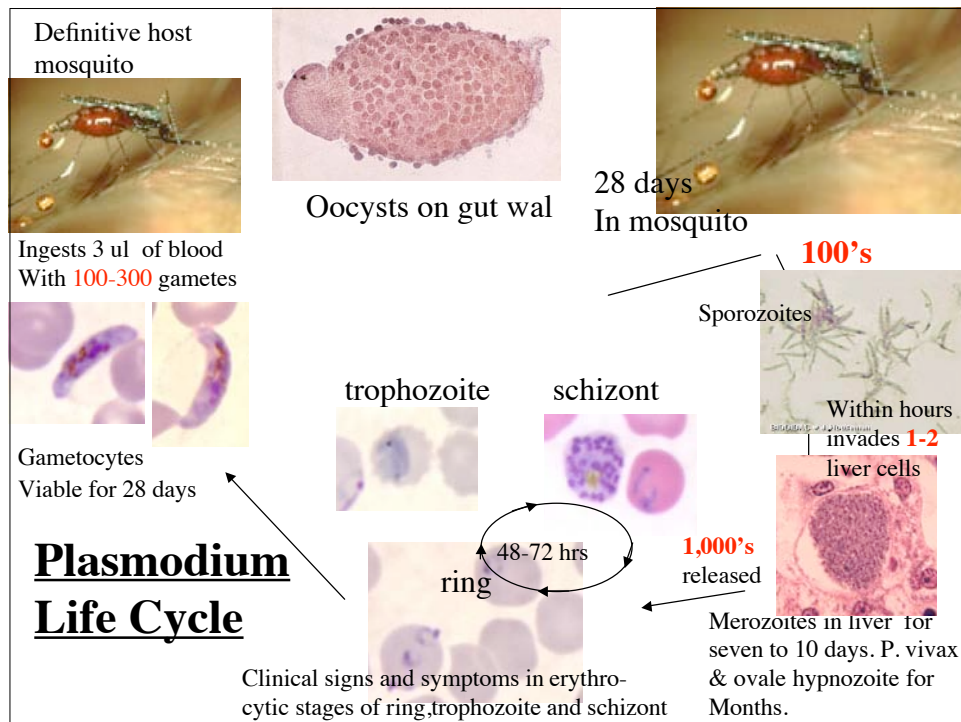
### Edited mRNAs



PNAS June 20, 2000 vol. 97 no. 13 6986-6993







## Toxoplasmosis

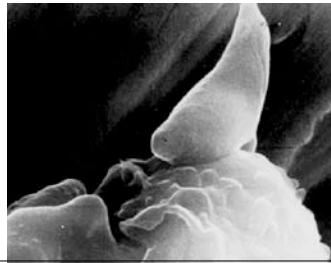
Cat is definitive host

The cat releases oocyst into feces.

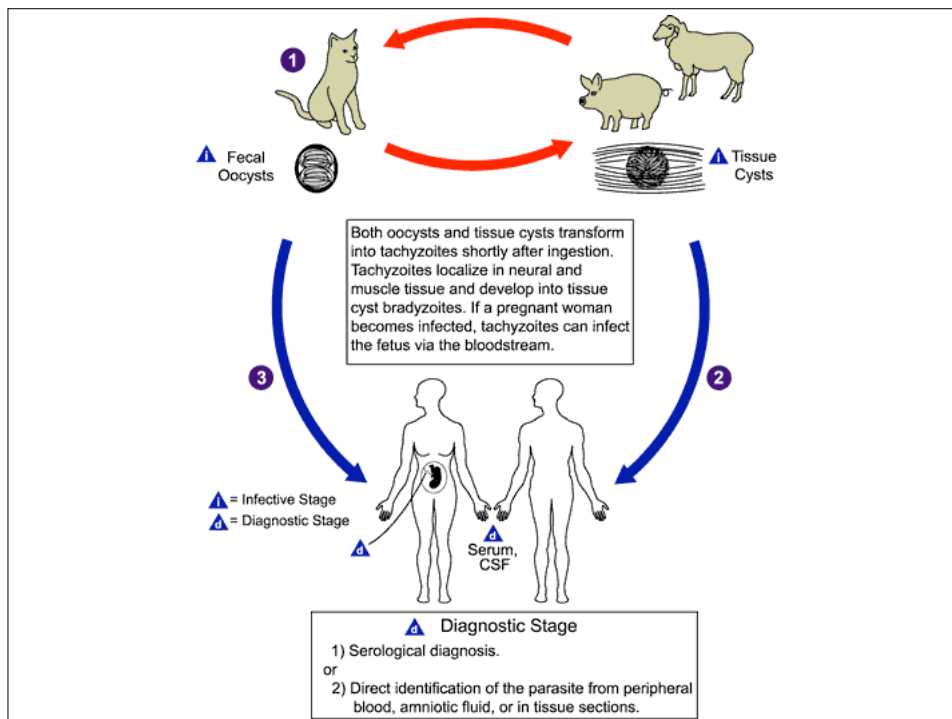
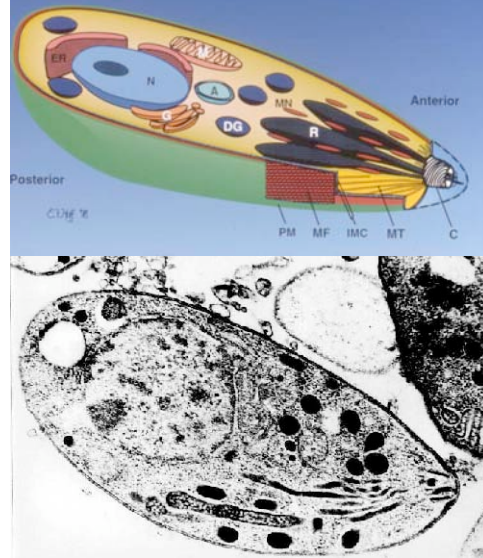
Cat litter infective for pregnant woman. Infection by undercooked meat with bradyzoite cyst.

Tachyzoite in tissue invasive form.

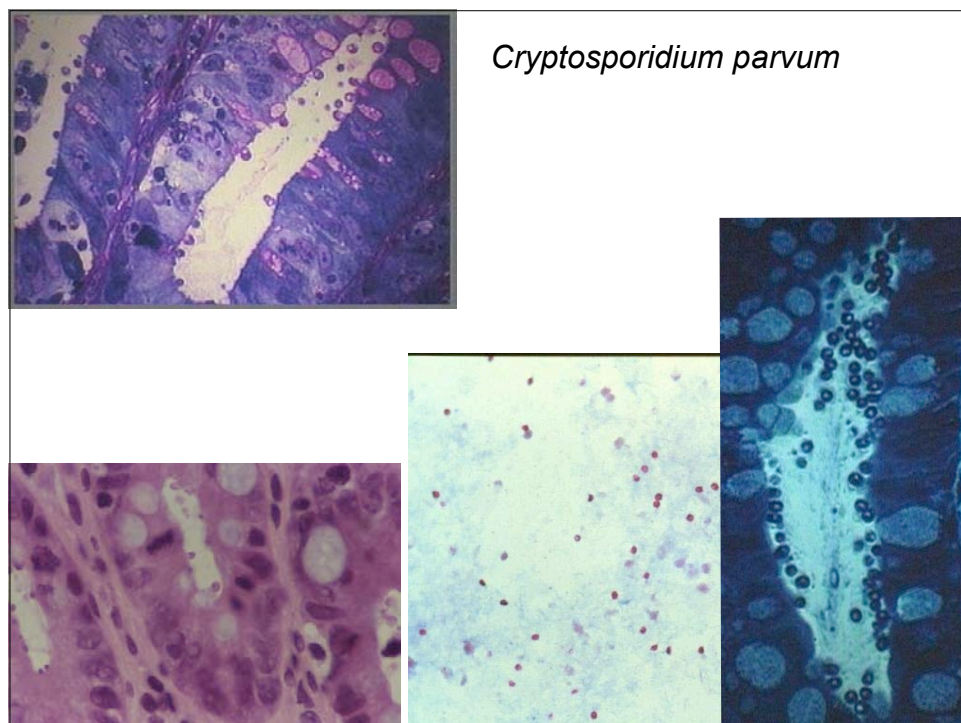
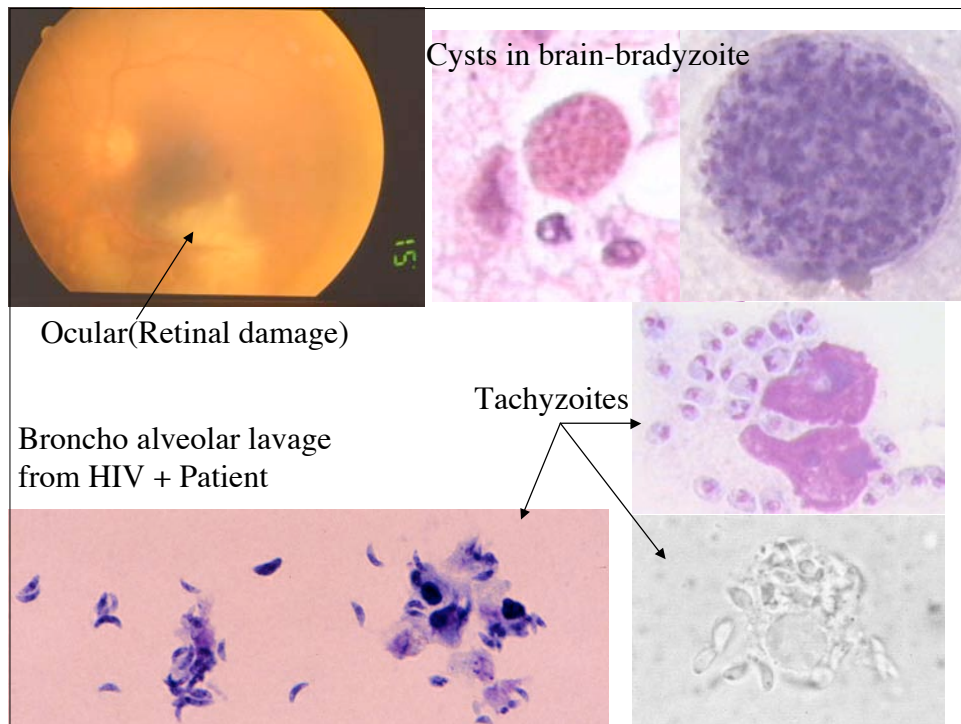
*Toxoplasma gondii* is single species that invades all cell types in all species from cats, humans, rodents, birds, sheep and pigs.



### Sub-Cellular Anatomy of *Toxoplasma gondii*







*Cryptosporidium parvum*

Genome size: 10.4 Mb

Chromosomes: 8 (range 0.6 to 3.3 Mb)

Genes: 5268 (comparable to *S. pombe*)(1 gene/4300 bp)

Website: <http://www.cryptodb.org>

Mean density of one predicted gene per 2.46 kb

Few introns

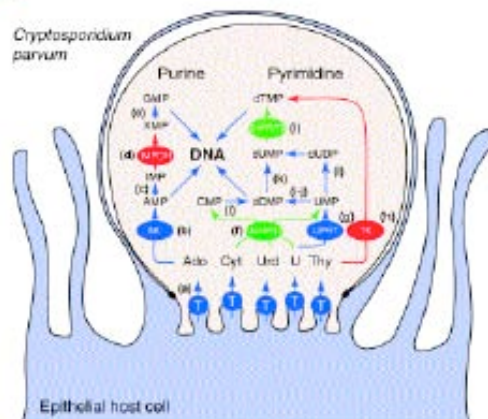
*Cryptosporidium* lacks an apicoplast

The presence of a mitochondrion, focused on the assembly of iron sulfur clusters, has now been confirmed for *C. parvum*

Lacks all six genes encoding enzymes for pyrimidine de novo synthesis, a pathway present in all other apicomplexans.

Lacks enzymes for amino acid synthesis

The *Cryptosporidium* nucleotide biosynthetic pathway is a phylogenetic mosaic. Enzymes labeled in red show strong phylogenetic association with eubacteria, those in green with plants and algae, and those in blue with other eukaryotes



Trends Parasitol.  
2004 Aug;20(8):355-8

Enzymes involved in the nucleotide biosynthetic pathway are listed as follows: (a) adenosine transporter; (b) adenosine kinase; (c) adenosine monophosphate deaminase; (d) IMPDH; (e) guanosine monophosphate synthase; (f) UTP/UTP; (g) UTP; (h) TK; (i) ribonucleotide diphosphate reductase; (j) cytosine triphosphate synthetase; (k) deoxycytosine monophosphate deaminase; and (l) DHFR-TS.

## **Individual eukaryotic genomes: malaria parasite *Plasmodium falciparum***

---

*Plasmodium falciparum* causes malaria, killing 2.7 million people each year.

Distinguishing features: Four *Plasmodium* species infect humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*.  
The life cycle is extremely complex.

Genome size: 22.8 Mb

Chromosomes: 14 (range 0.6 to 3.3 Mb)

Genes: 5268 (comparable to *S. pombe*)(1 gene/4300 bp)

Website: <http://www.plasmodb.org>

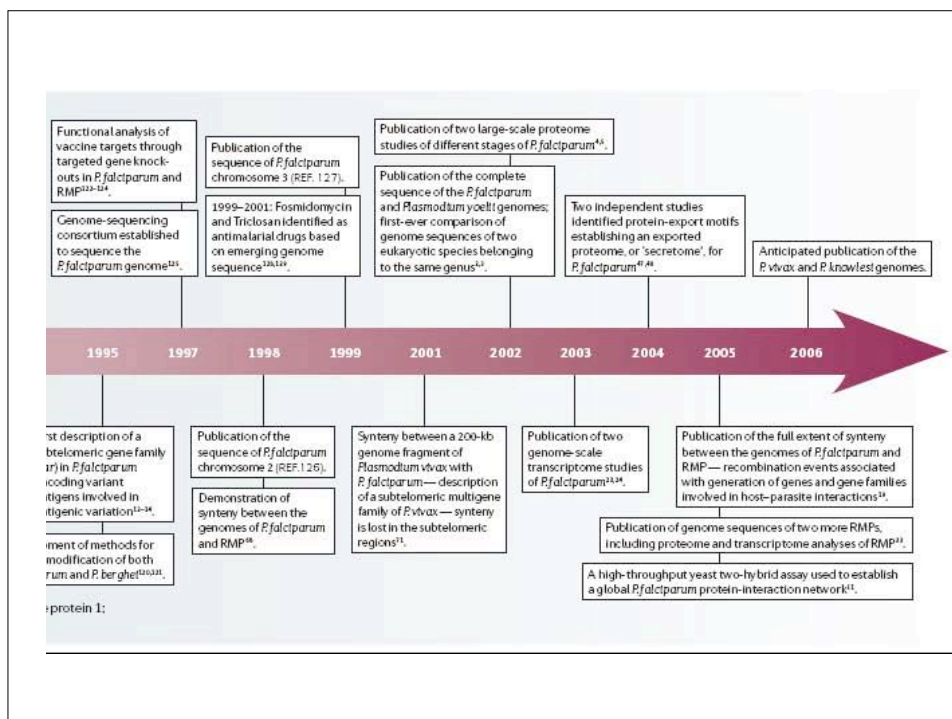
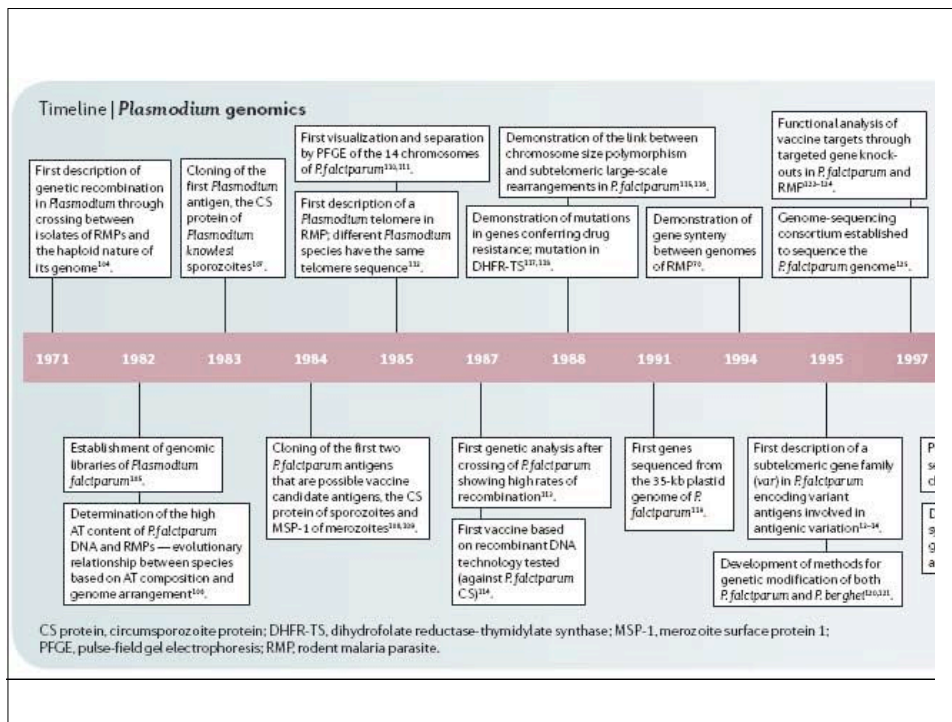
*P. falciparum* has an adenine+thymine (AT) content of 80.6%.  
The *P. yoelli yoelli* genome was also sequenced (infects rats).

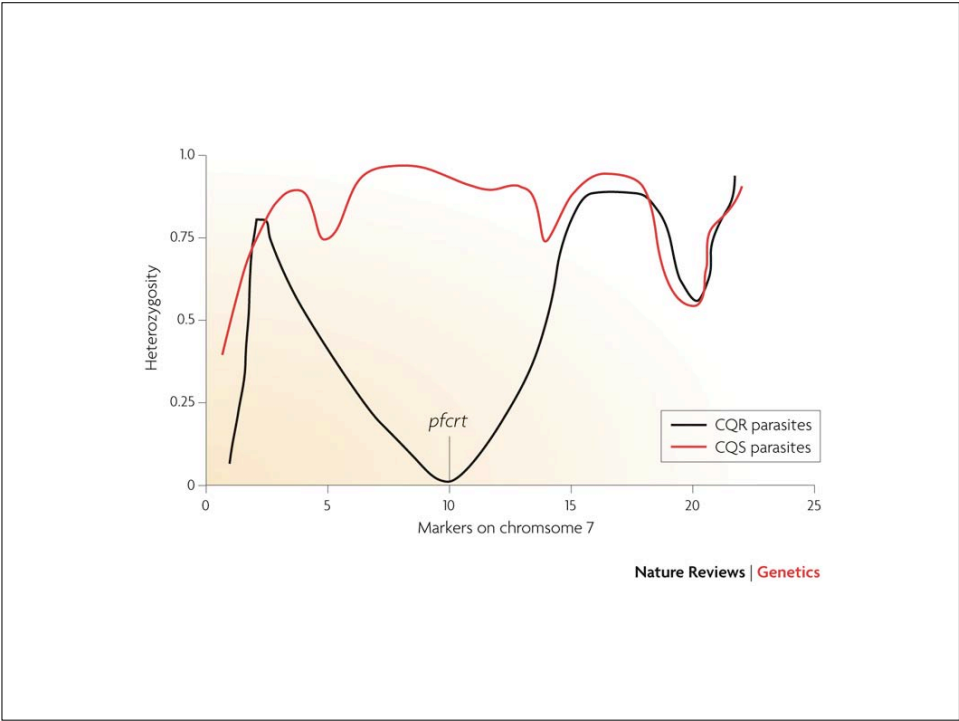
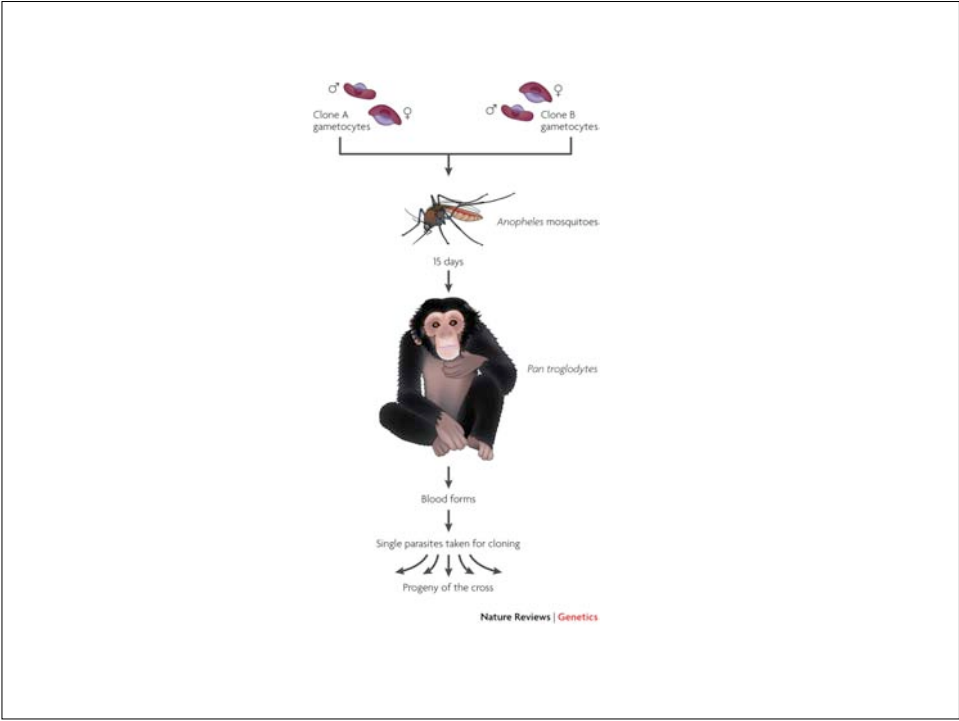
## **Individual eukaryotic genomes: malaria parasite *Plasmodium falciparum***

---

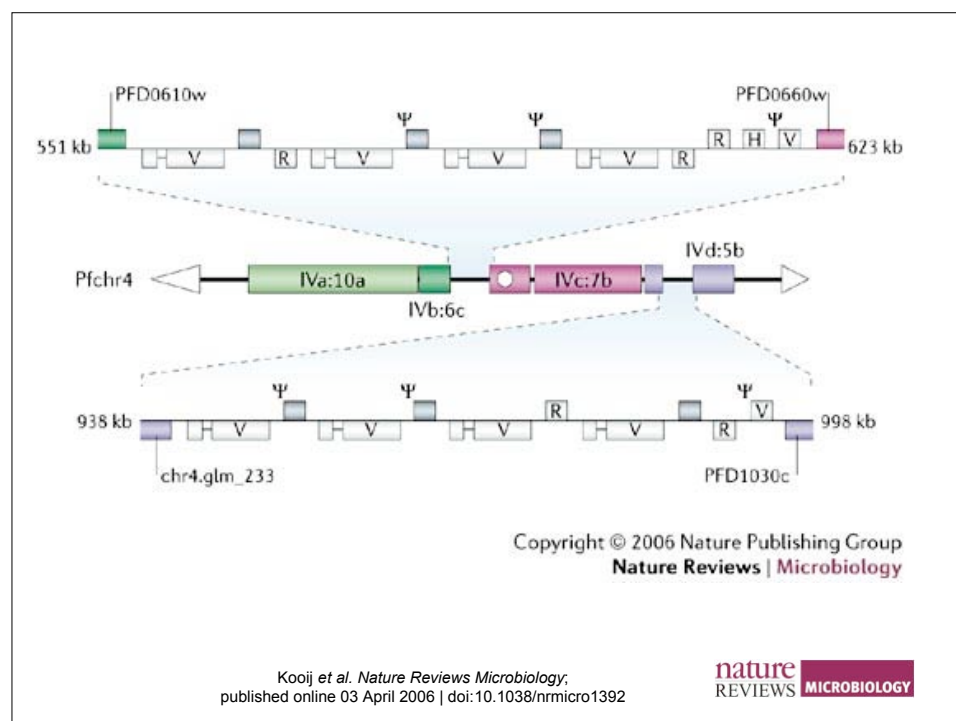
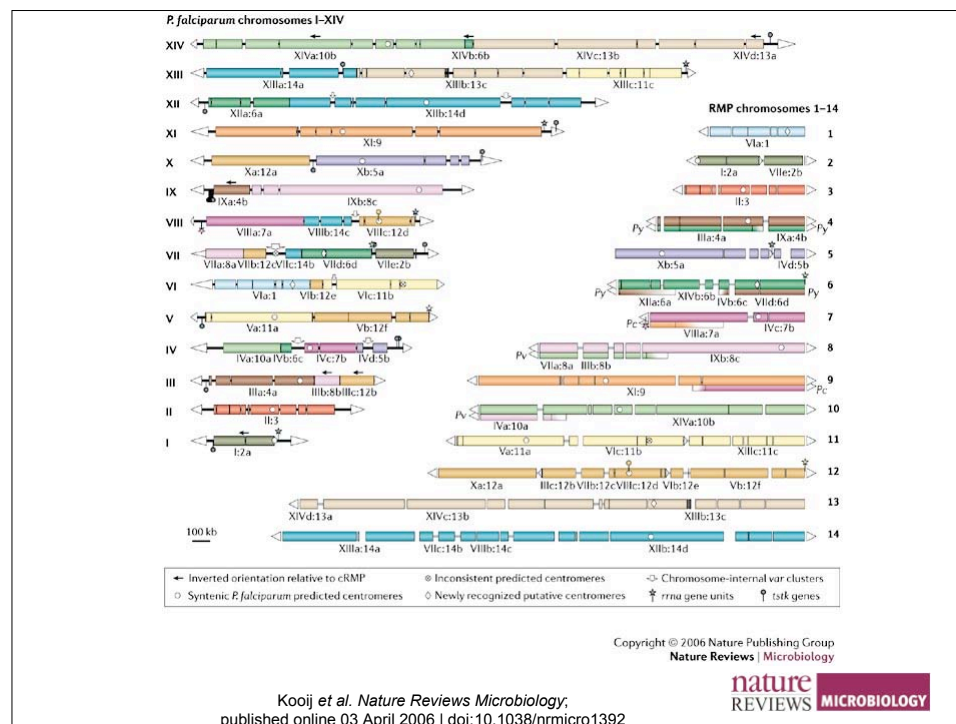
Bioinformatics approaches to *Plasmodium falciparum*:

- The apicoplast (relic plastid; fatty acid, isoprene metabolism) is a potential drug target. Apicoplast signal sequences found.
- Comparative genomics defines some gene functions, identifies genes lacking in closely related species
- Genes implicated in antigenic variation and immune system evasion can be identified (e.g. 1000 copies of *vir*)
- Proteomics applied to four stages of the life cycle (sporozoites, merozoites, trophozoites, gametocytes)
- Atypical metabolic pathways may be exploited, e.g. use of 1-deoxy-D-xylulose 5-phosphate (DOXP) in isoprene biosynthesis.

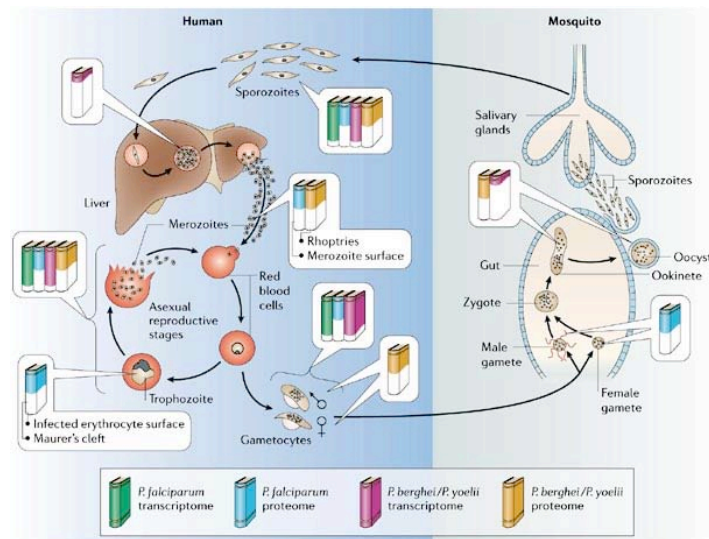












Copyright © 2006 Nature Publishing Group  
Nature Reviews | Microbiology

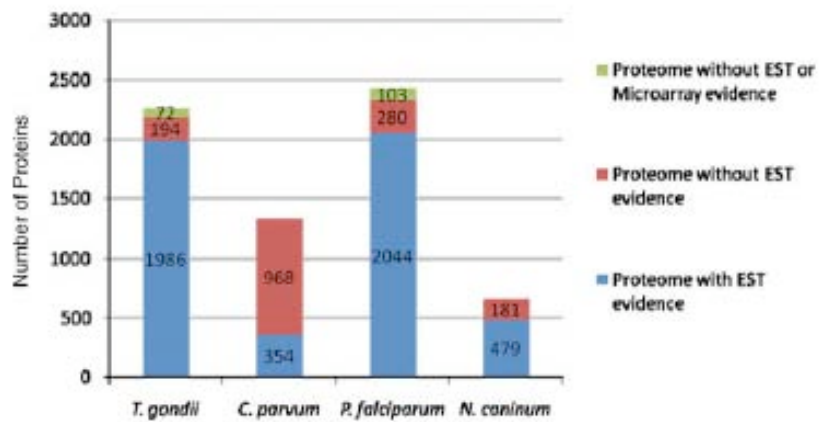
Kooij *et al.* Nature Reviews Microbiology,  
published online 03 April 2006 | doi:10.1038/nrmicro1392

nature  
REVIEWS MICROBIOLOGY

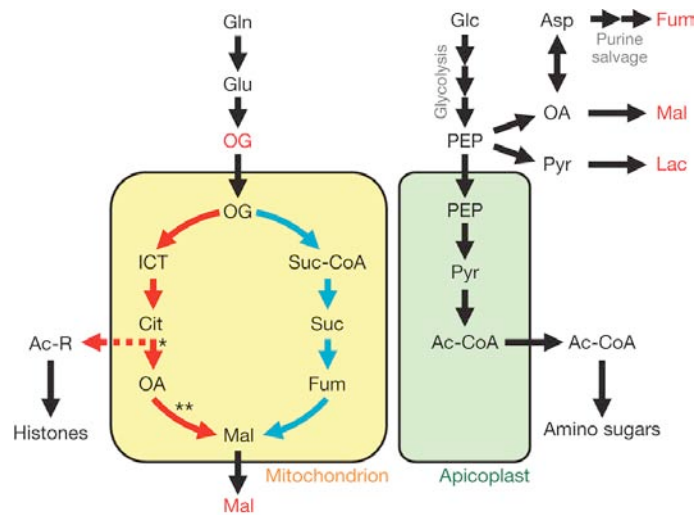
Arrays	Platform	Probe length	Number of probes per gene/SNP	Number of probes	Number of genes targeted	References or contacts
cDNA array	Custom	PCR products	-	3,648 inserts	-	79
DeRisi array	Custom	70mer	1+/gene	7,462	5,803	21
Scripps array	Affymetrix	25mer	1–20/gene	328,989	5,159	22
JHU array	Affymetrix	25mer	11/gene	>500,000	4,700	Johns Hopkins Malaria Research Institute
PfSanger array	Affymetrix	25mer	Tiling	2,500,000	5,486	Sanger Institute (A. Ivens, personal communication)
NIH SNP array	Affymetrix	25mer	1/SNP	3,354	3,354 SNP	Affymetrix/NIH (X.S., unpublished work)
Toxo_Pf array	Affymetrix	25mer	4/SNP	8,120	2,030	UPenn (X.S. & D. Roos, unpublished work)
<i>Plasmodium yoelii</i> array	Custom	65mer	1–4/ORF	7,271	6,500	118
<i>Plasmodium berghei</i> array	Custom	PCR products	-	6,354	-	25
<i>Plasmodium vivax</i> array	Custom	70mer	1/exon/2 kb	16,120	16,824 ORF	The J. Craig Venter Institute *

The arrays that are listed here are either published or available to the public freely or on a collaborative basis. This is probably a sample of the arrays that have been produced — we apologize for not including those that we are unaware of in this table. \*Formerly The Institute for Genomic Research, JHU, John Hopkins University (USA); NIAID, National Institute of Allergy and Infectious Diseases (USA); NIH, National Institutes of Health (USA); UPenn, University of Pennsylvania (USA).

## Proteomic validation of RNA detection



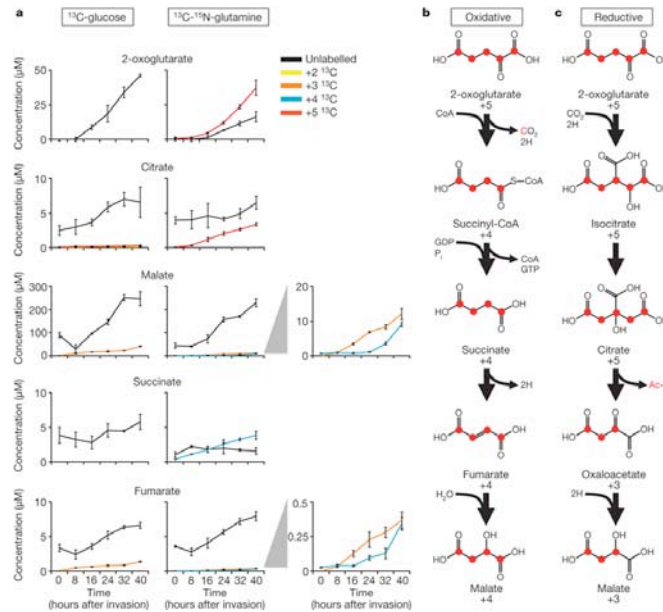
## An integrated model for central carbon metabolism in *P. falciparum*.



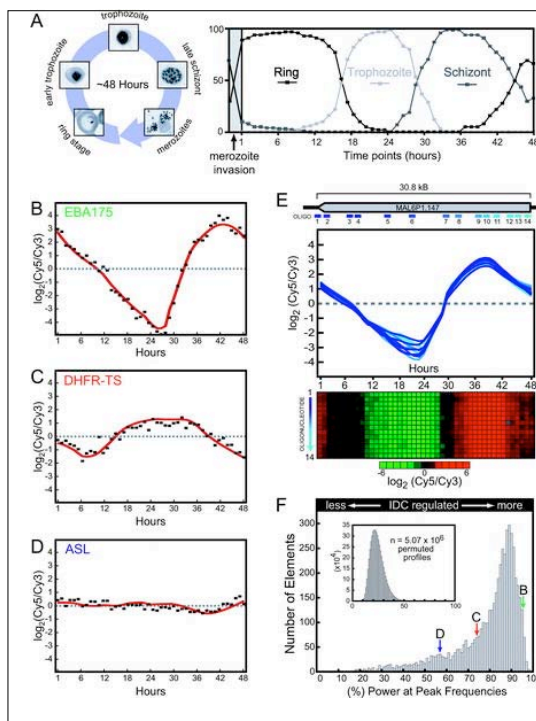
KL Olszewski *et al.* *Nature* **466**, 774-778 (2010) doi:10.1038/nature09301

nature

### Glutamine drives reverse flux through the TCA cycle.



KL Olszewski et al. *Nature* 466, 774-778 (2010) doi:10.1038/nature09301

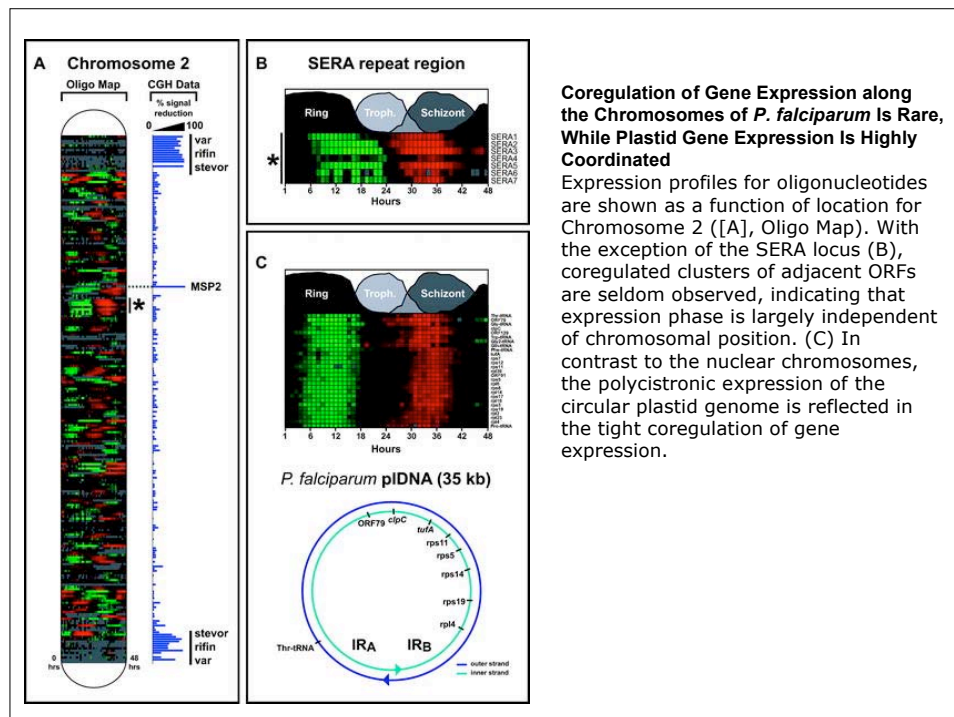
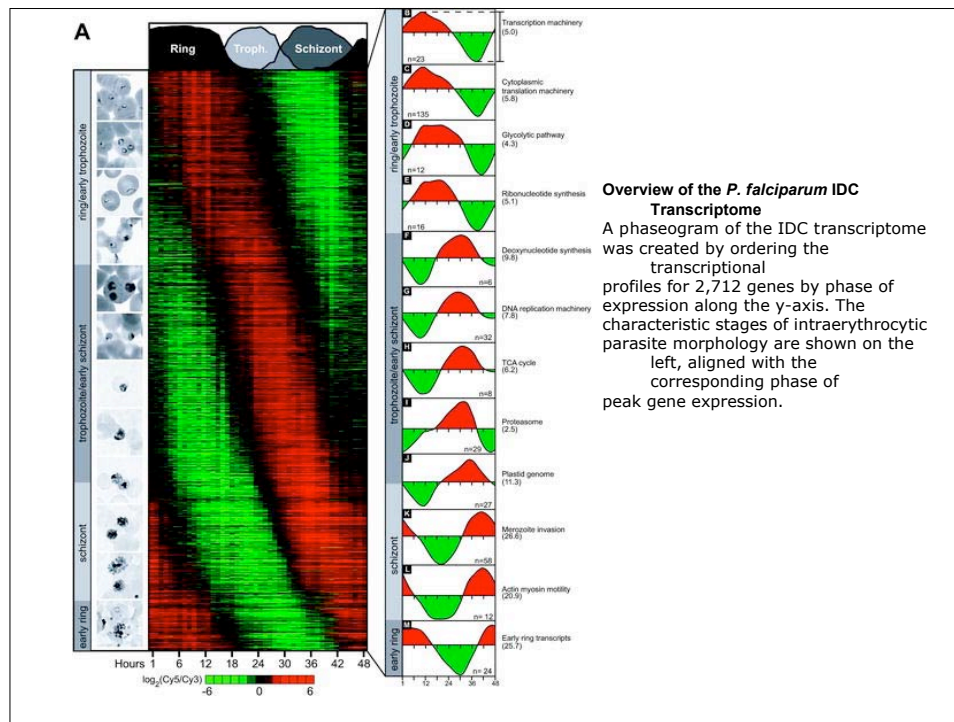


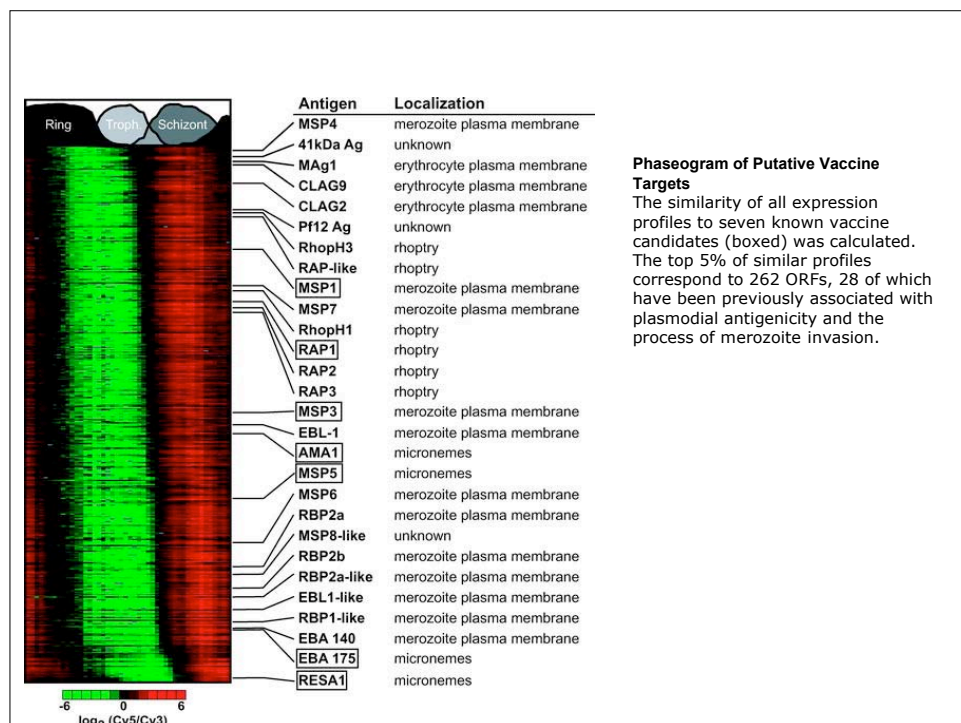
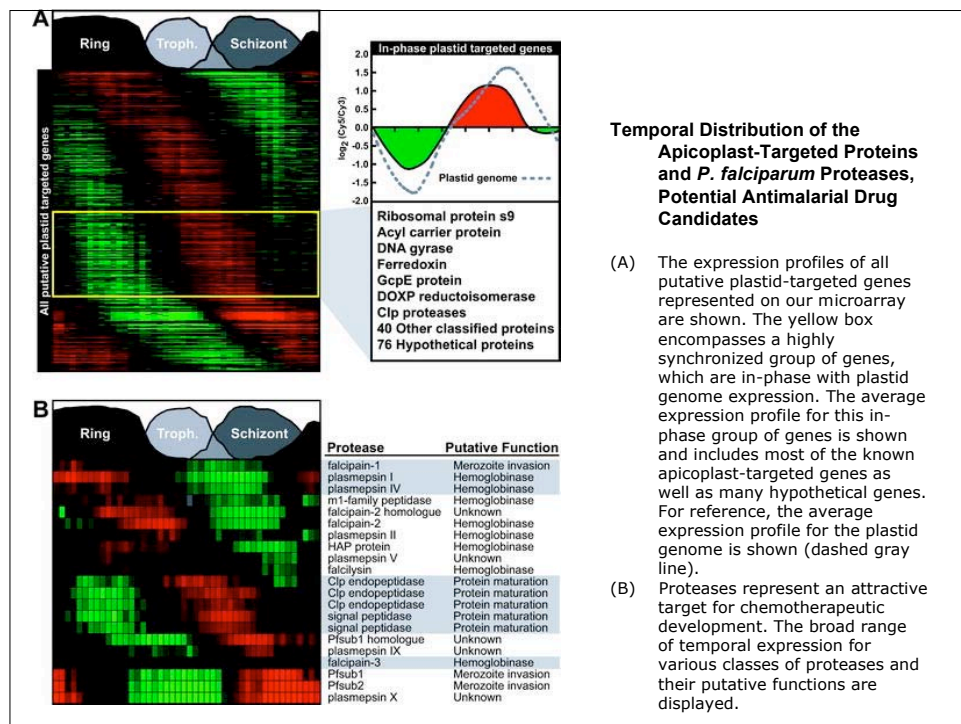
### The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*

Zbynek Bozdech,  
Manuel Llin, Brian Lee  
Pulliam, Edith D. Wong,  
Jingchun Zhu, Joseph  
L. DeRisi

<http://biology.plosjournals.org/perlserv?request=get-document&doi=10.1371/journal.pbio.0000005>

(A) Giemsa stains of the major morphological stages  
(B-D) Example expression profiles for three genes, encoding EBA175, DHFR-TS, and ASL  
(E) MAL6P1.147, the largest predicted ORF in the *Plasmodium* genome, is represented by 14 unique DNA oligonucleotide elements. The location of each of the oligonucleotide elements within the predicted ORF and the corresponding individual expression profiles are indicated (oligo 1±14).







#### Websites

<http://www.plasmodb.org/plasmo/home.jsp>

<http://www.apidb.org/apidb/>

<http://www.ebi.ac.uk/parasites/parasite-genome.html>

Codon usage table

<http://www.kazusa.or.jp/codon/>

#### Pathport vbi vt

<http://pathport.vbi.vt.edu/pathinfo/>

[http://pathport.vbi.vt.edu/pathinfo/pathogens/Plasmodium\\_falciparum\\_Info.shtml](http://pathport.vbi.vt.edu/pathinfo/pathogens/Plasmodium_falciparum_Info.shtml)

[http://pathport.vbi.vt.edu/pathinfo/pathogens/Cryptosporidium\\_parvum\\_Info.shtml](http://pathport.vbi.vt.edu/pathinfo/pathogens/Cryptosporidium_parvum_Info.shtml)