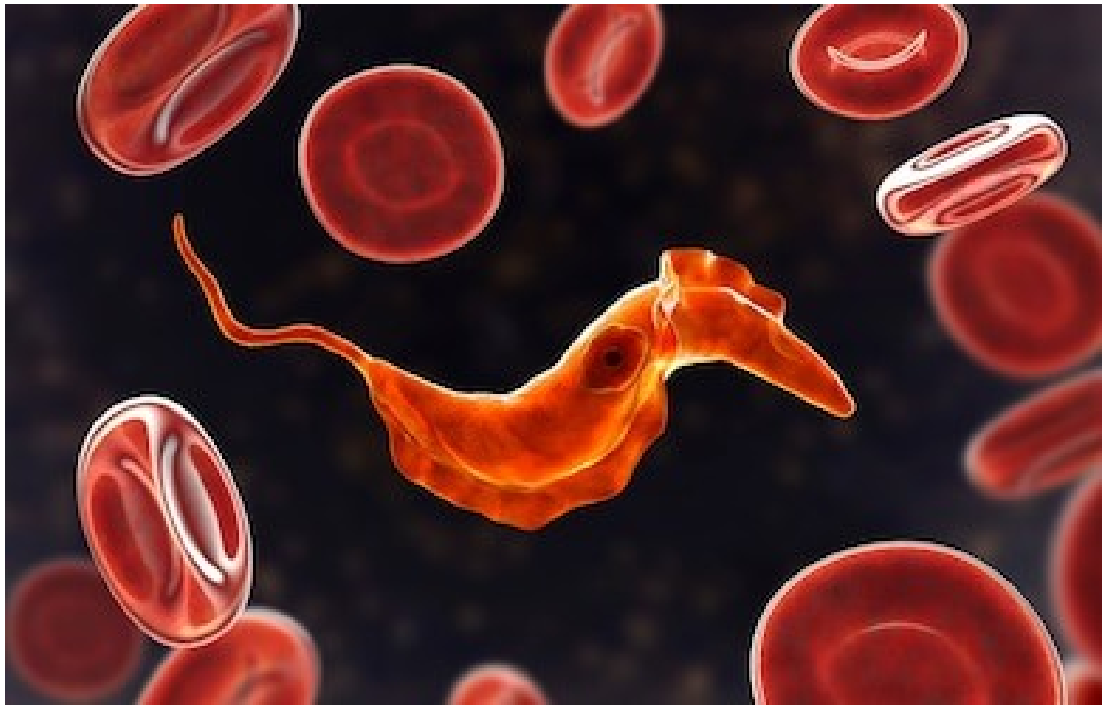
Trypanosomes:
their genomes and
two example of their extreme biology



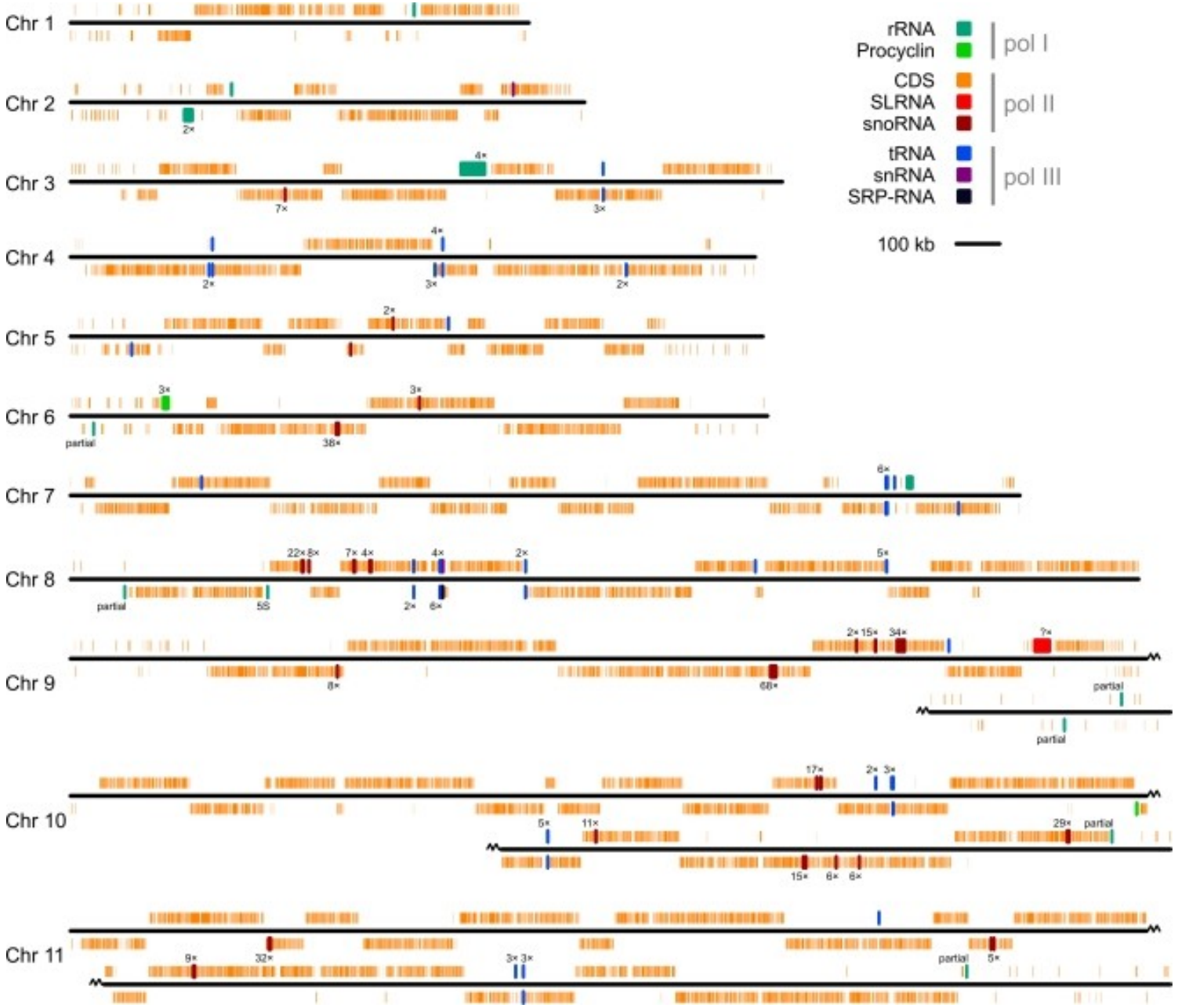
Excavata

Excavata

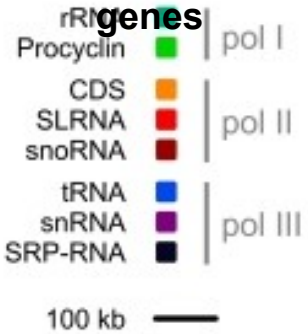
Organism	Relevance	Genome size	Number of genes predicted
Giardia duodenalis (assemblage B)	Human pathogen (Giardiasis)	11.7 Mb	4,470
Giardia duodenalis (assemblage A)	Human pathogen (Giardiasis)	11.7 Mb	6,470
Leishmania braziliensis	**Human pathogen (Leishmaniasis)**	**33 Mb**	**8,314**
Leishmania infantum	**Human pathogen (Visceral leishmaniasis)**	**33 Mb**	**8,195**
Leishmania major	**Human pathogen (Cutaneous leishmaniasis)**	**32.8 Mb**	**8,272**
Naegleria gruberi	Human pathogen (Primary amoebic meningoencephalitis)	41 Mb	15,727
Trichomonas	Human pathogen (Trichomoniasis)	160 Mb	59,681
Entamoeba histolytica	Human pathogen (amoebic dysentery)	23.8 Mb	9,938
brucei			
Trypanosoma cruzi	**Human pathogen (Chagas disease)**	**34 Mb**	**22,570**

Amoebozoa

genome of *Trypanosoma*

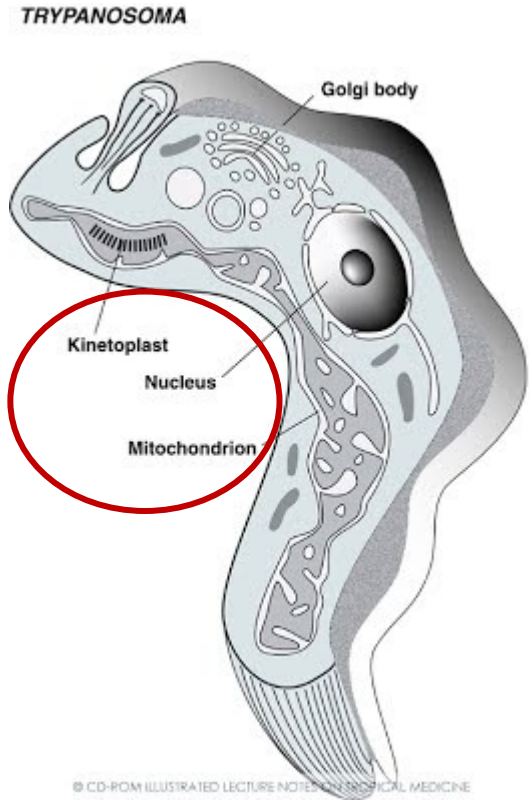


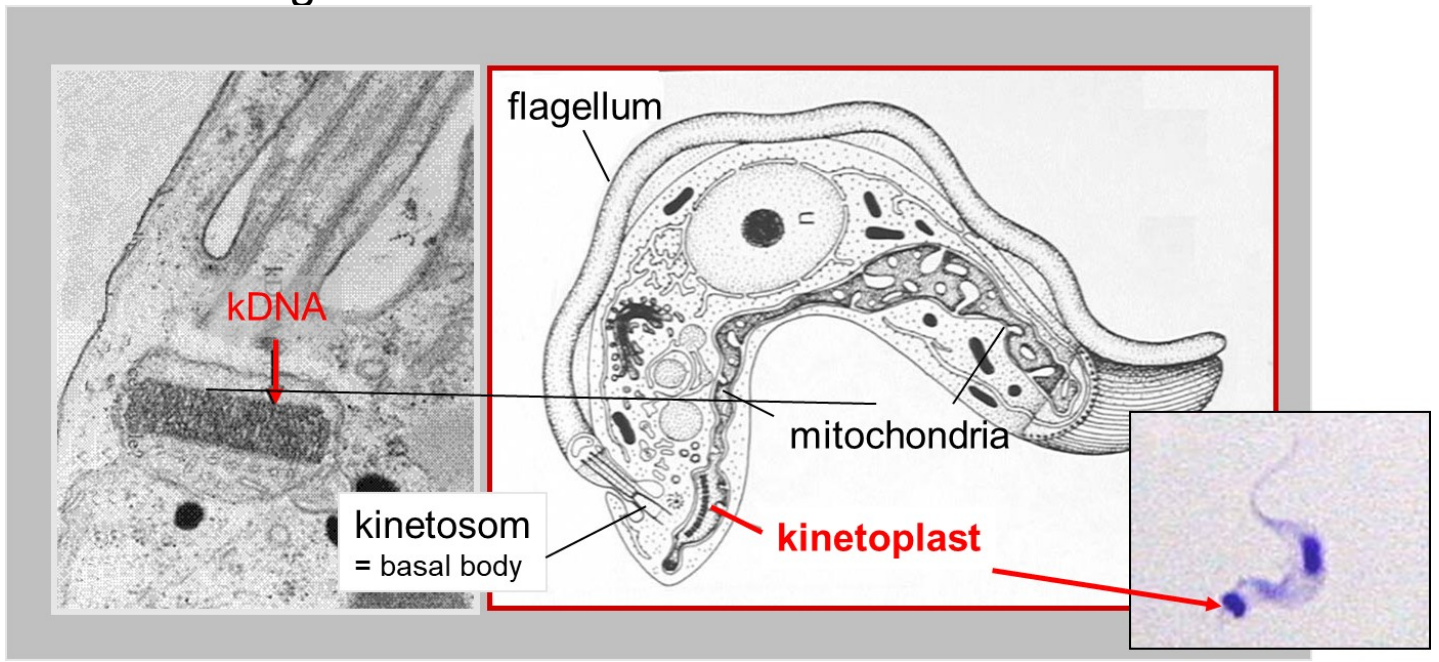
- complete genome **~105 Mb *Trypanosma cruzi***
- complete genome **~35 Mb *Trypanosma cruzi***
- **11 pairs of large chromosomes** of 1 to 6 megabase pairs
- protein-coding genes are organized in large directional clusters - **polycistronic transcription**
- 3–5 **intermediate chromosomes** of 200 to 500 kilobase pairs
- around 100 **minichromosomes** of around 50 to 100 kilobase pairs (these may be present in multiple copies per haploid genome)
- most genes are held on the large chromosomes, with the **minichromosomes carrying only VSG genes**



„EXCAVATA“: kinetoplastida

- mitochondria, nucleus
- unique organelle called the **kinetoplast**
  - **accumulation of mitochondrial DNA**
  - **kDNA**
  - an appendix of their single mitochondrion located **near the basal body of the flagellum** (kinetosome)
  - contains a giant network of thousands of small interlocking **circular DNAs**

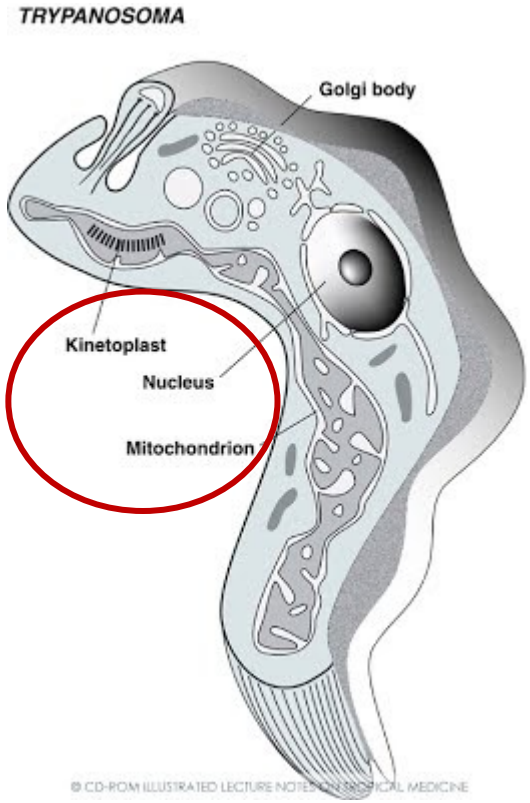


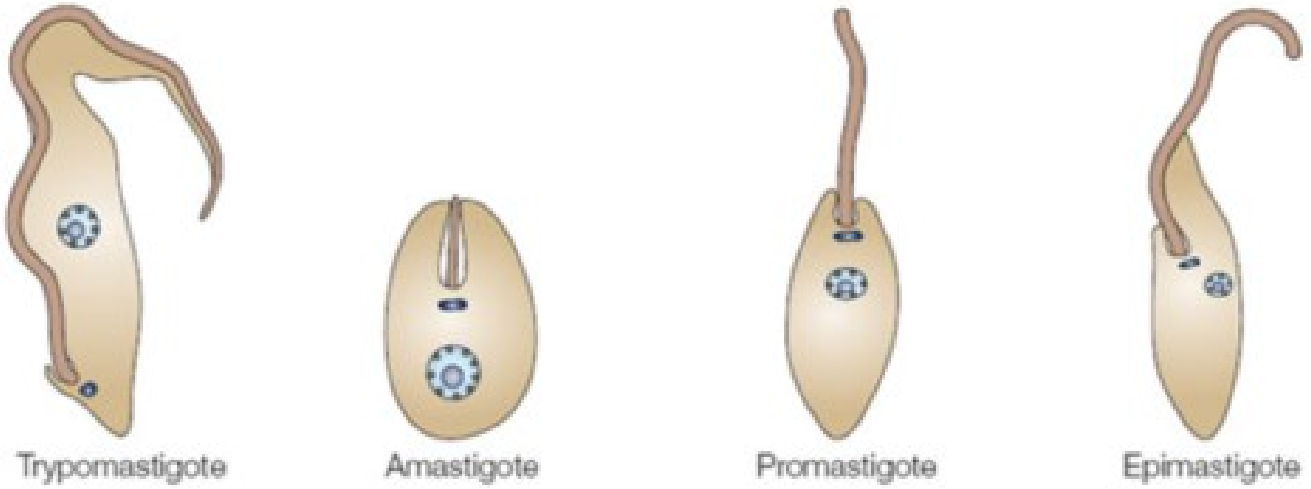


1910-20 Feulgen reaction + „ kinetonucleus“
(J. Kulda a E. Nohýnková „Buňka prvoků“, 2006)

„EXCAVATA“: kinetoplastida

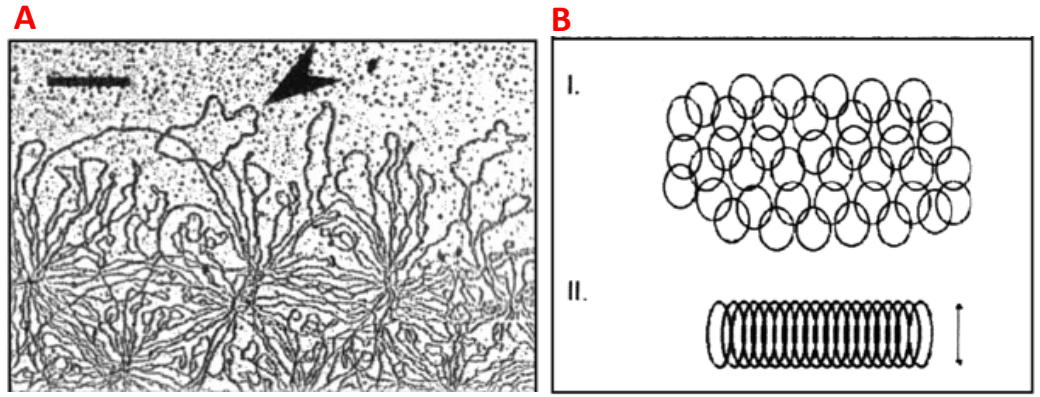
- mitochondria, nucleus
- unique organelle called the **kinetoplast**
  - **accumulation of mitochondrial DNA**
  - **kDNA**
  - an appendix of their single mitochondrion located **near the basal body of the flagellum** (kinetosome)
  - contains a giant network of thousands of small interlocking **circular DNAs**



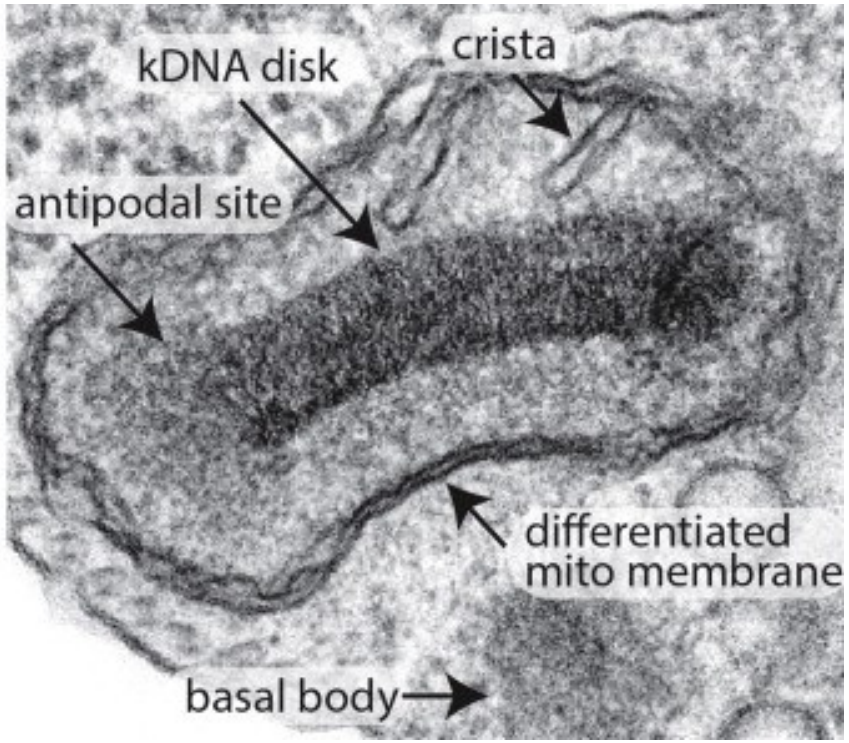


Kinetoplast

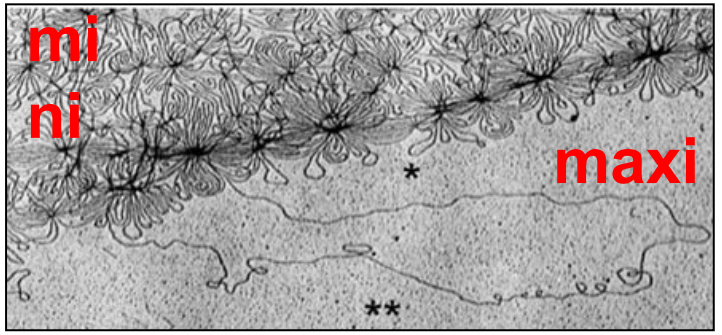
- contains circular DNA in two forms
  - **maxicircles and minicircles**
- **10 – 20% cell DNA** (trypanoplasms 40%)
- system of circular molecules
- **maxicircles: between 20 and 40kb in size, a few dozen identical copies per kinetoplast**
- **minicircles: between 0.5 and 10kb in size, several thousand copies usually nearly identical in size but heterogeneous in sequence**



kDNA network structure. (A) Electron micrograph of the periphery of an isolated kDNA network from *T. avium*. Loops represent interlocked **minicircles (the arrowhead indicates a clear example)**. Bar, 500 nm. (B) Diagrams showing the **organization of minicircles**. (I) Segment of an isolated network showing interlocked minicircles in a planar array. (II) Section through a condensed network disk *in vivo* showing stretched-out minicircles. The double-headed arrow indicates the thickness of the disk, which is about half the circumference of a minicircle.

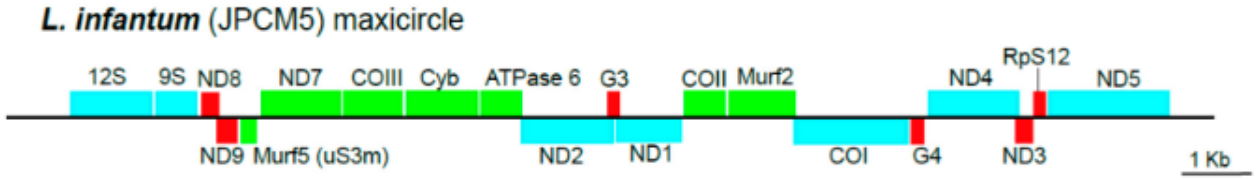


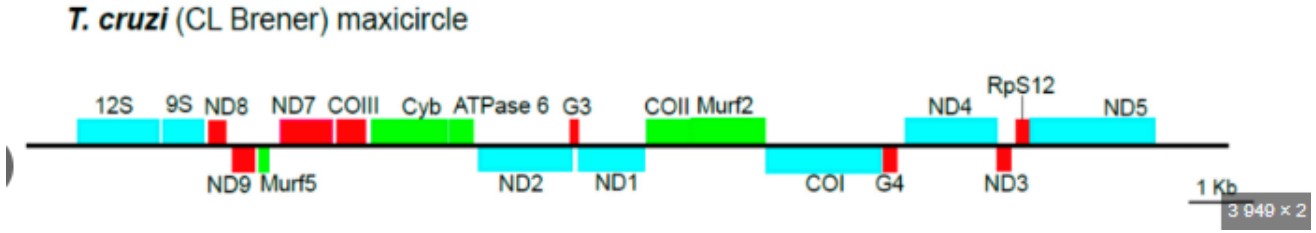
kinetoplast



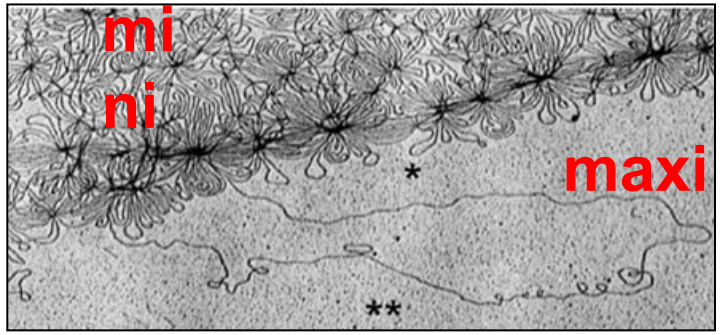
maxicircles

- encode typical **mitochondrial gene products**
  - e.g. rRNAs and subunits of respiratory chain complexes
- some of the protein-coding genes are encrypted
- to generate functional mRNAs, the **cryptic maxicircle transcripts** undergo posttranscriptional modification via an intricate **RNA editing process**
  - **involves insertion and deletion of uridine residues at specific sites in the transcripts**
  - the genetic information for editing is provided by **guide RNAs (gRNAs)**
  - mostly encoded by minicircles, although a few are encoded by maxicircles



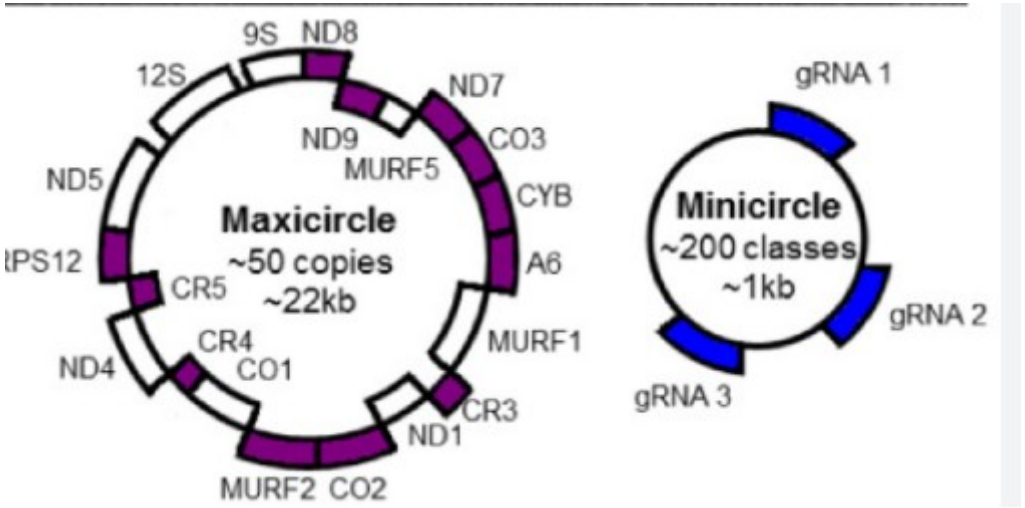


kinetoplast



minicircles

- **encoding gRNAs** is the only known function
- some organisms that edit extensively (such as *Trypanosoma brucei*) possess about 200 different minicircle sequence classes in their network to provide sufficient gRNAs



mitochondriální DNA (kDNA)

↓

transkripce

↓

RNA editing

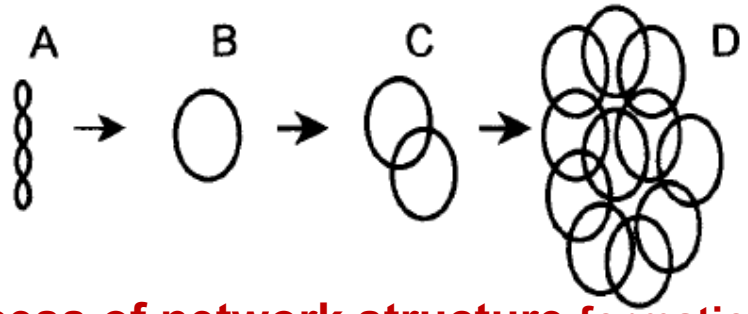
AAAGU**AGAGAA**CCUGGUAGGU
AAAGU**AGAUUGUAUA**CCUGGU

↓

translace

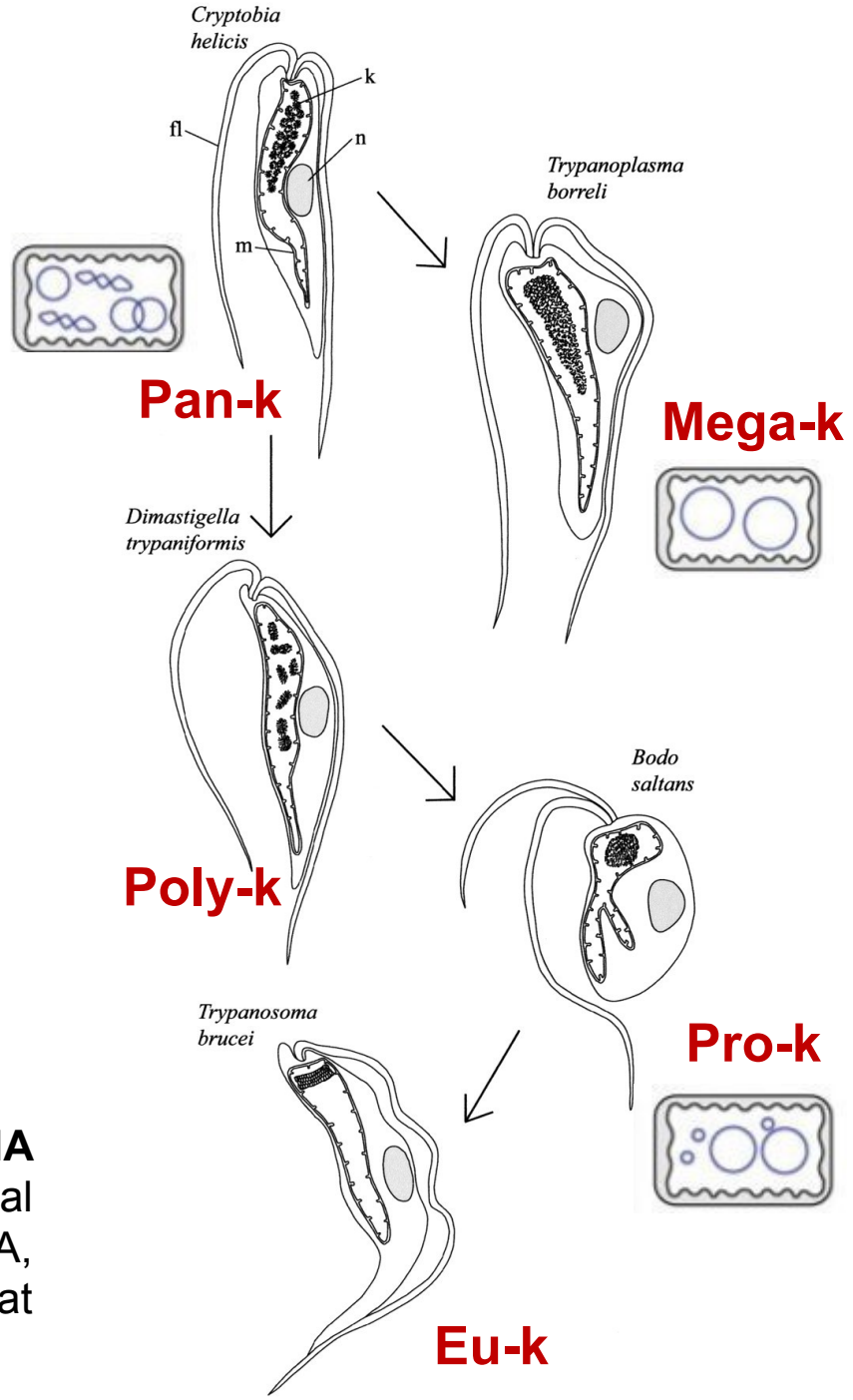
kinetoplast

› kinetoplasts not forming networks and forming networks



process of network structure formation (molecule „relaxation“ is required!)

› different members of kinetoplastida reveals different structure

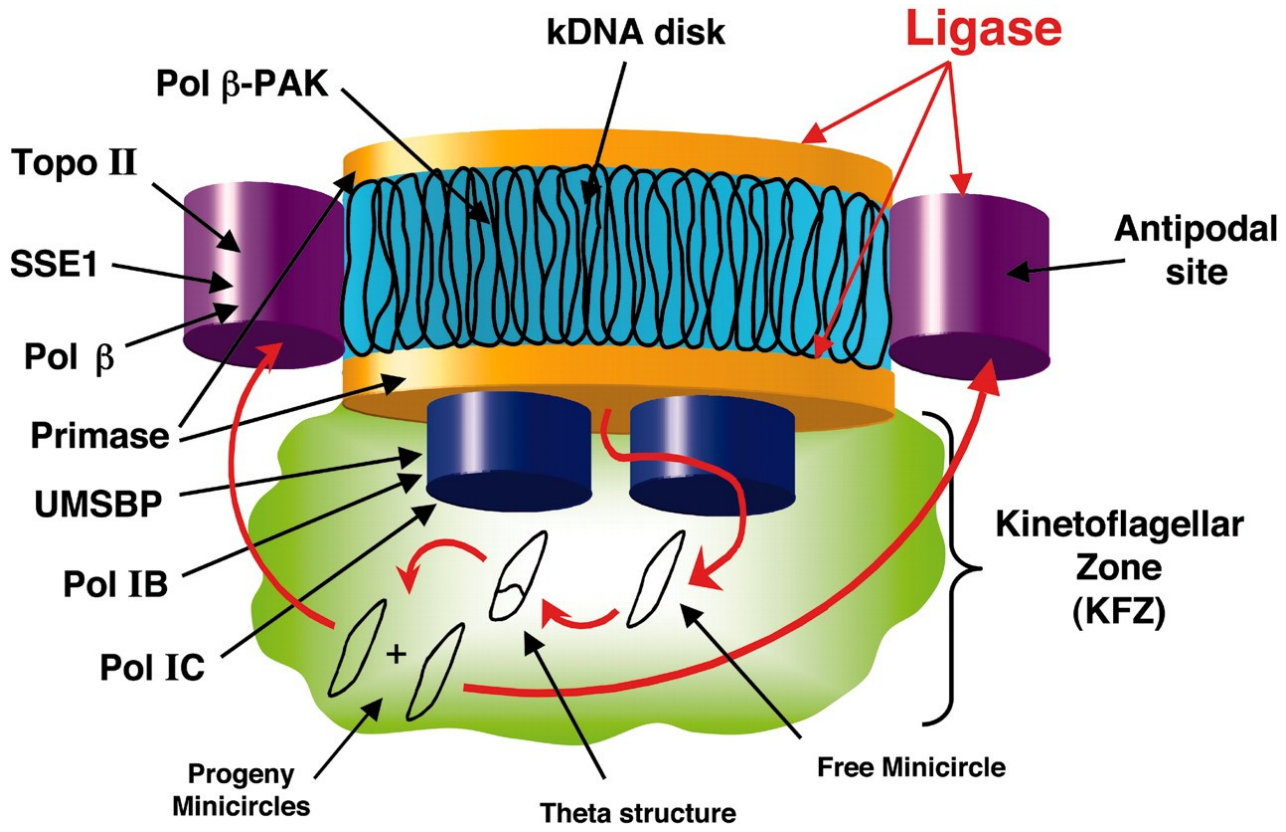


Proposed evolution of kinetoplastids, emphasizing differences in kDNA organization and compaction. kDNA (k) is the structure within the mitochondrial matrix. fl, flagellum; m, mitochondrion; n, nucleus. kDNA in *C. helicis* is pan-kDNA, that in *T. borreli* is mega-kDNA, that in *D. trypaniformis* is poly-kDNA, that in *B. saltans* is pro-kDNA, and that in *T. brucei* is a kDNA network.

kinetoplast: replication of the kDNA network

› replication of the kinetoplast occurs simultaneously to the duplication of the adjacent flagellum and just prior to the nuclear DNA replication

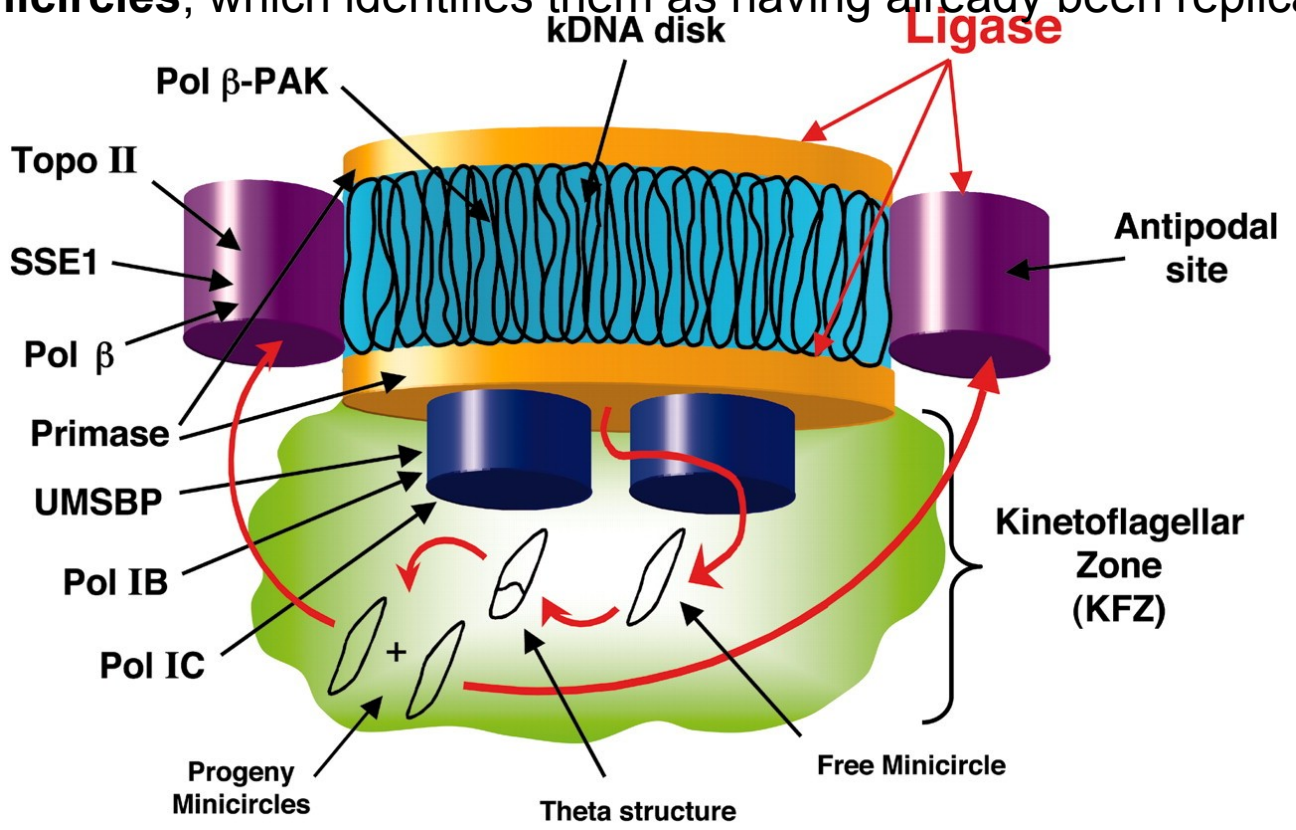
› **minicircles are released from the network** into a **kinetoflagellar zone** (region between the kinetoplast and the mitochondrial membrane) in which they initiate replication



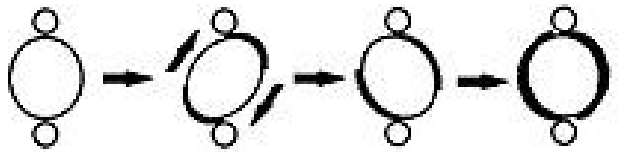
kinetoplast: replication of the kDNA network

› this process occurs one minicircle at a time, and only a small number of minicircles are unlinked at any given moment

› to keep track of **which minicircles have been replicated,** upon rejoining to the kDNA network **a small gap remains in the nascent minicircles**, which identifies them as having already been replicate

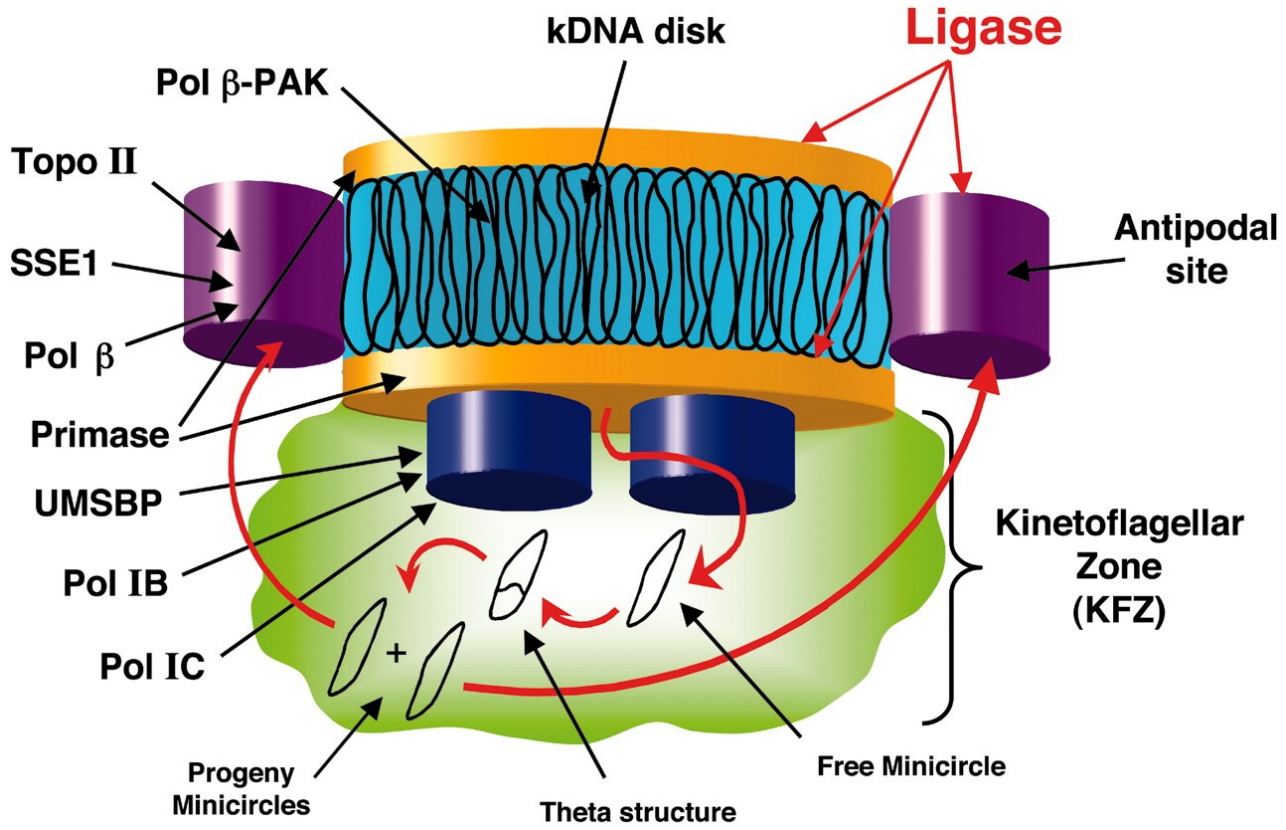


kinetoplast: replication of the kDNA network



› to prevent the build-up of new minicircles, **the entire kDNA network will rotate** around the central axis of the disk

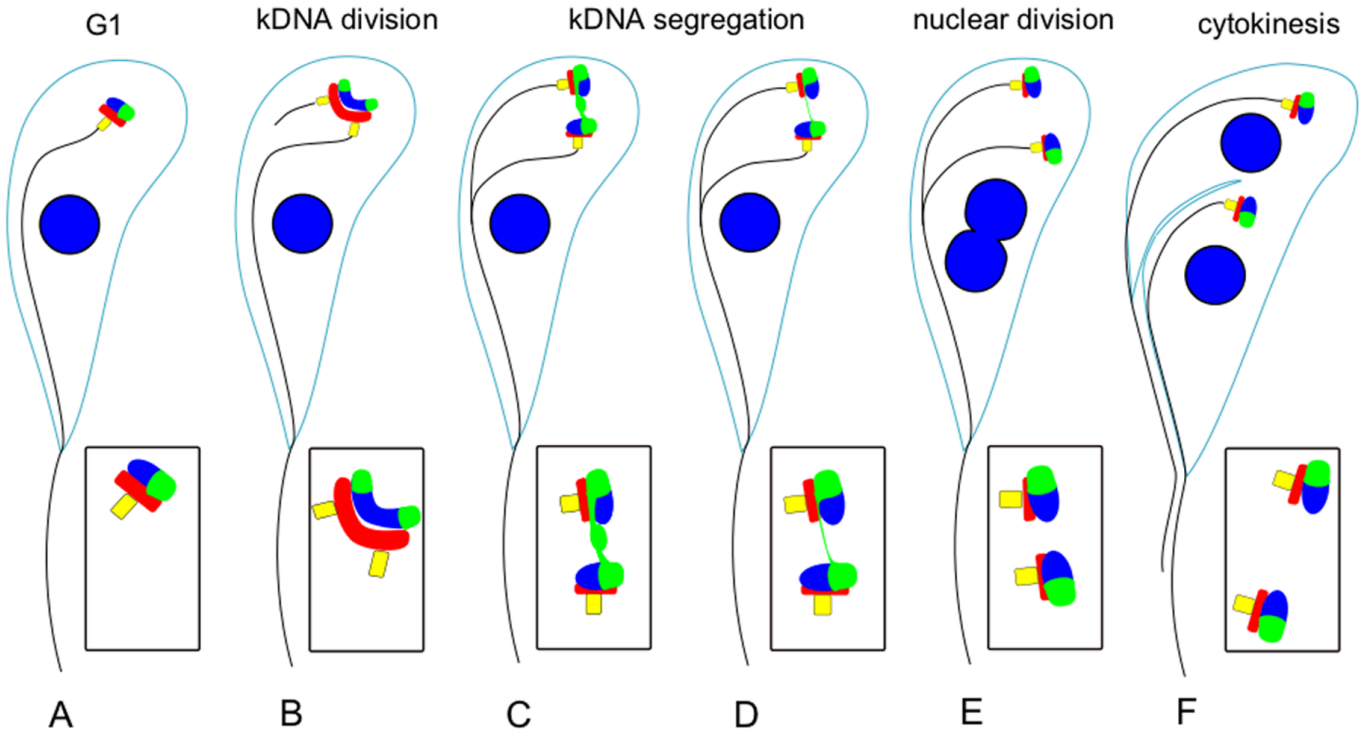
› the rotation is believed to be **directly connected to** the replication of **the adjacent flagellum**

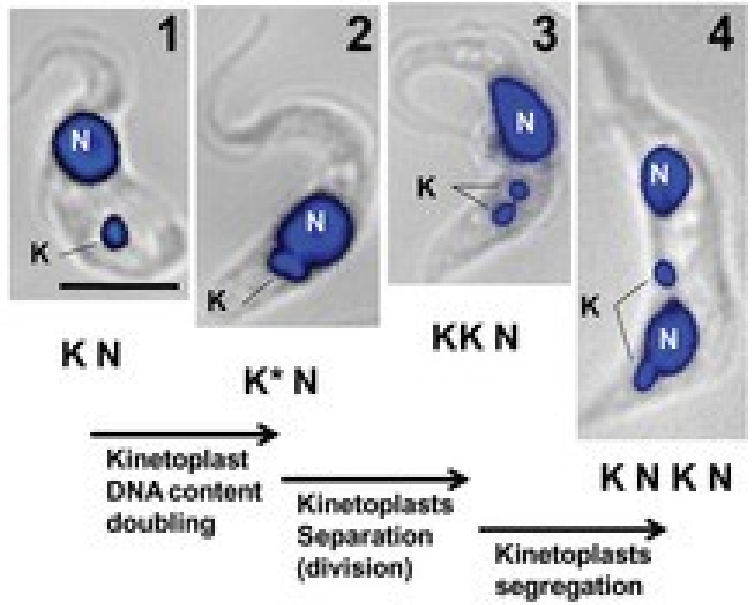


kinetoplast: replication of the kDNA network

- the exact mechanisms for **maxicircle kDNA** have yet to be determined in the same detail
- a structure called a **nabelschnur** (German for "umbilical cord")

Kinetoplast replication is linked to nuclear DNA replication and cell division





Basal body (yellow), flagellum (black line), tripartite attachment complex (TAC) (red), kDNA and nucleus (blue),

RNA editing

- **site-specific posttranscriptional changes in an RNA sequence**
  - (other than pre-mRNA splicing and 3'-polyadenylation)
- was first described in trypanosomatids
- **a widespread phenomenon throughout eukaryotes**

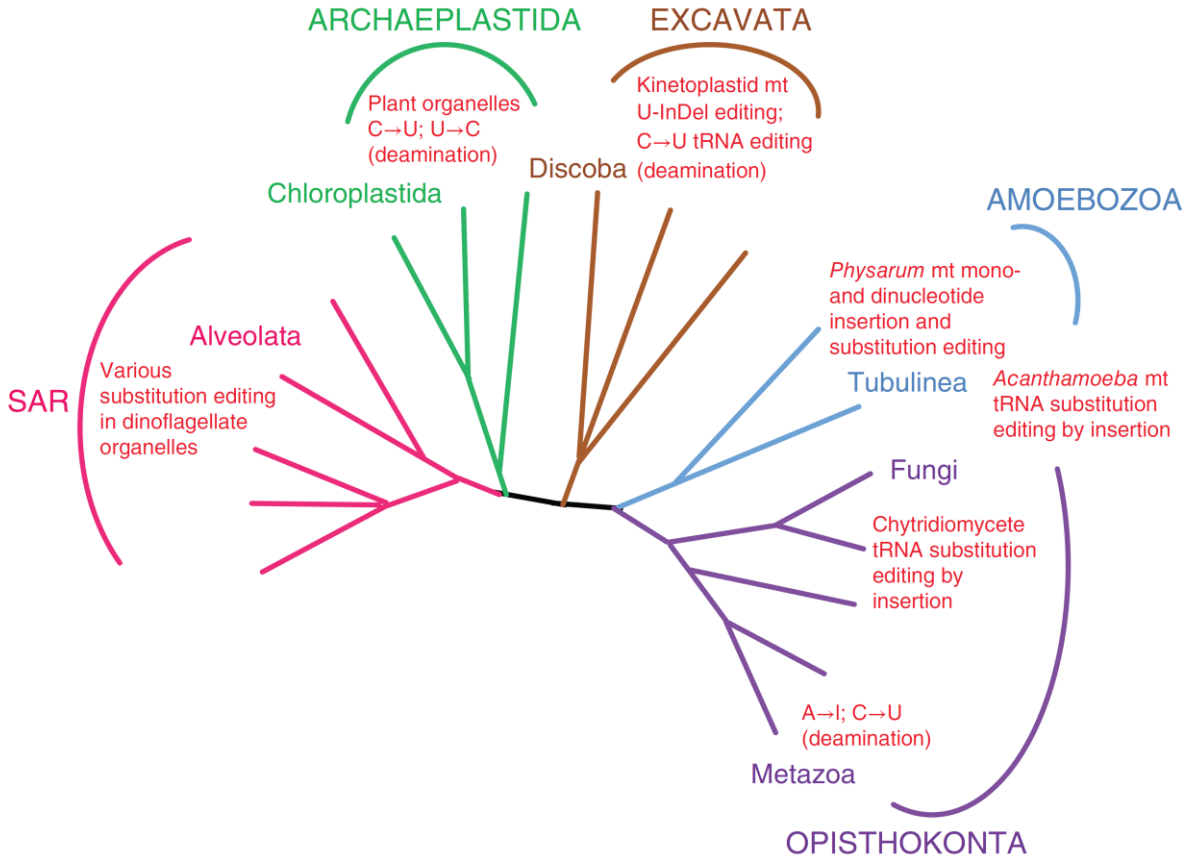
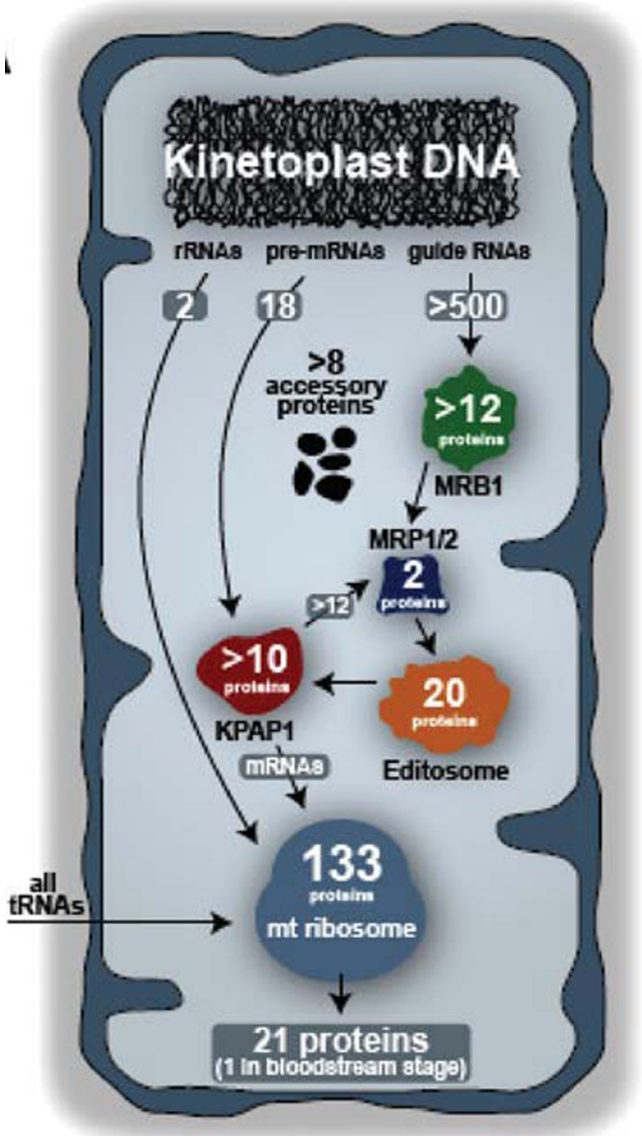


FIGURE 1 | Distribution of multiple types of RNA editing across eukaryotes. Phylogenetic tree based on Adl et al.[4] Only branches with clades that have a demonstrated type of RNA editing are labeled. Adjacent red text summarizes the type of editing.

RNA editing in Trypanosomes



mitochondriální DNA (kDNA)

↓

transkripce

↓

RNA editing

AAAGU**AGAGAA**CCUGGUAGGU

AAAGU**AGAUUGUAUA**CCUGGU

↓

translace

RNA editing in Trypanosomes

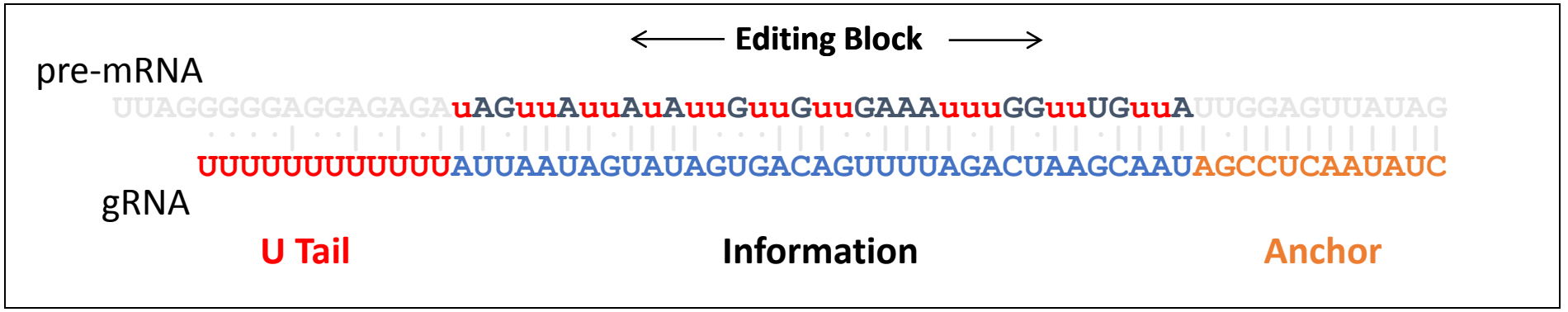
T. brucei A6 RNA editing

edited

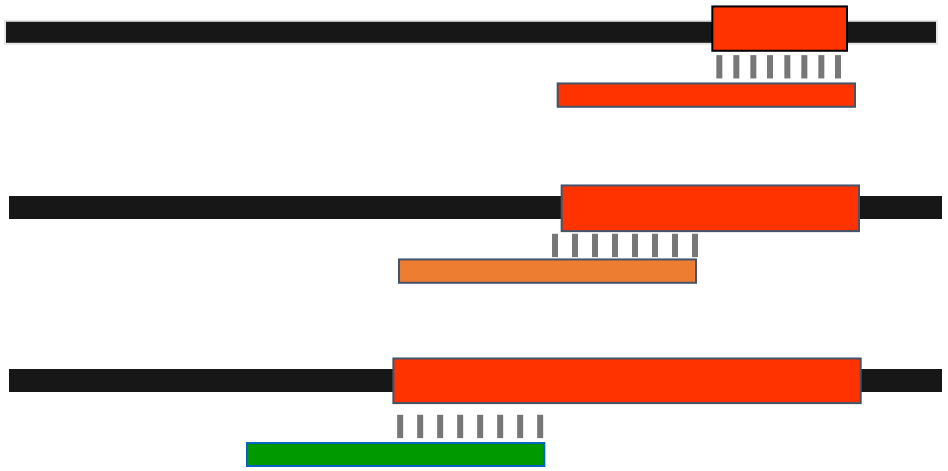
```
                                    M  F  L  F  F  F  C  D
L  F  W  L  R  L  L  L  C  M  Y  Y  C  V  W  S  R  L  C  F
   I  V  Y  F  N  C  L  M  L  I  F  D  F  L  L  F  C  L  F
   D  L  Y  L  F  V  G  L  C     L  F  L  L  L  W  F  M  L
   F  N  L  Y  S  L  I  L  Y  Y  C  I  T  Y  L     N  L  Y
   L  L  F  C  I  V  F  L  L  Y  I  A  F  L  F  L  F  C  F
   L  C  D  F  F  L  F  N  N  L  L  V  G  D     S  F  M  D
   V  F  F  I     R  F  L  L  C  F  L  E  C  F  S  L  L  C  R
   C  L  S  T  F  L  R  L  F  C  N  L  L  S  S  H  F  L  L
   L  M  F  F  D  F  F  Y  F  I  F  V  F  F  F  W  C  F  L
L  L  I  Y  F  I  Y  F  C  V  L  F  L  F  I  I  L  C  V  F
   I  F  V  G  F  I  C     R  H  I  T     V  I  Y  F  L ter
```

Benne et al. (1986), Bhat et al. (1990)

RNA editing in Trypanosomes

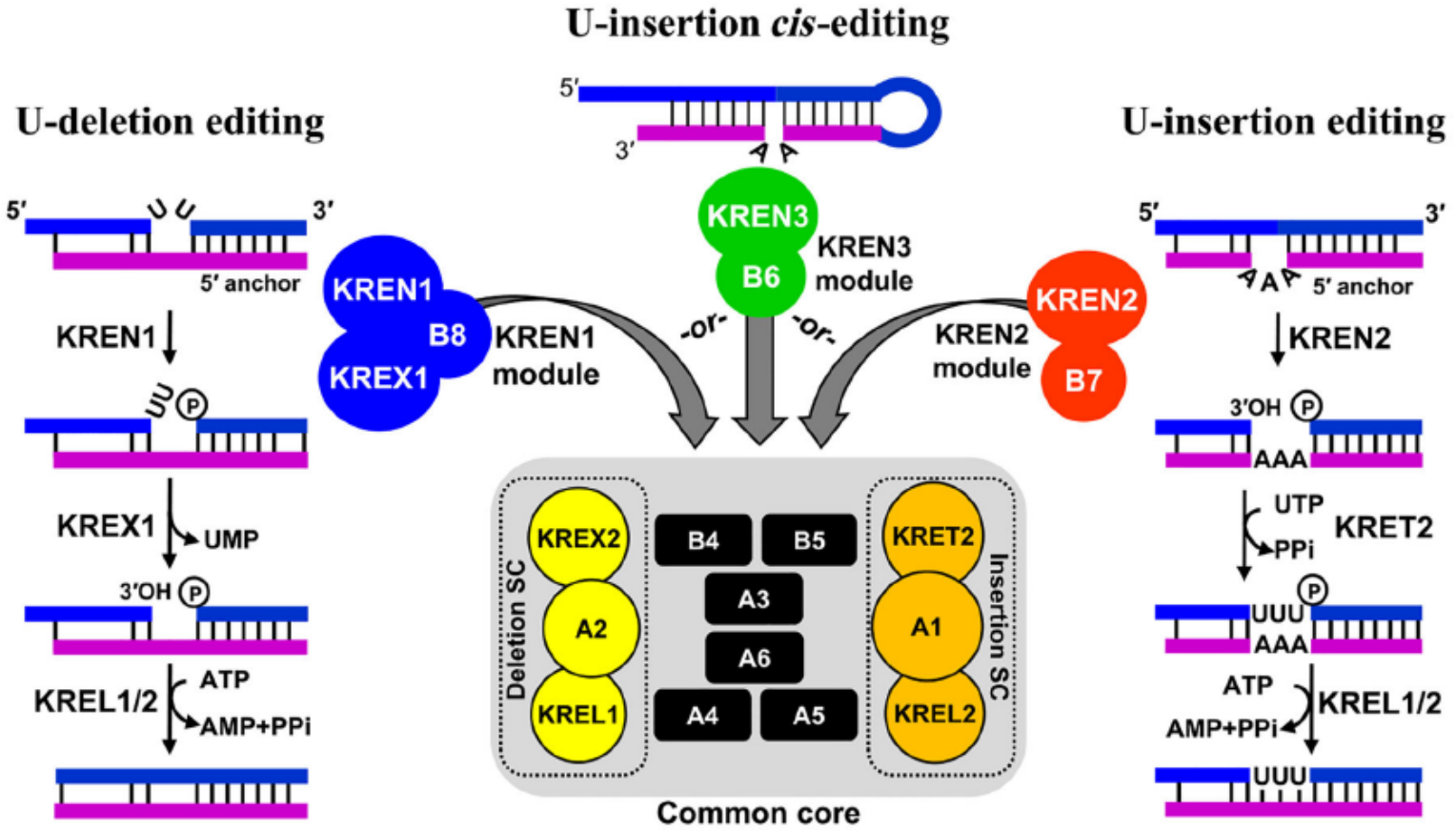


Více gRNA je nutných pro editaci jednoho transkriptu



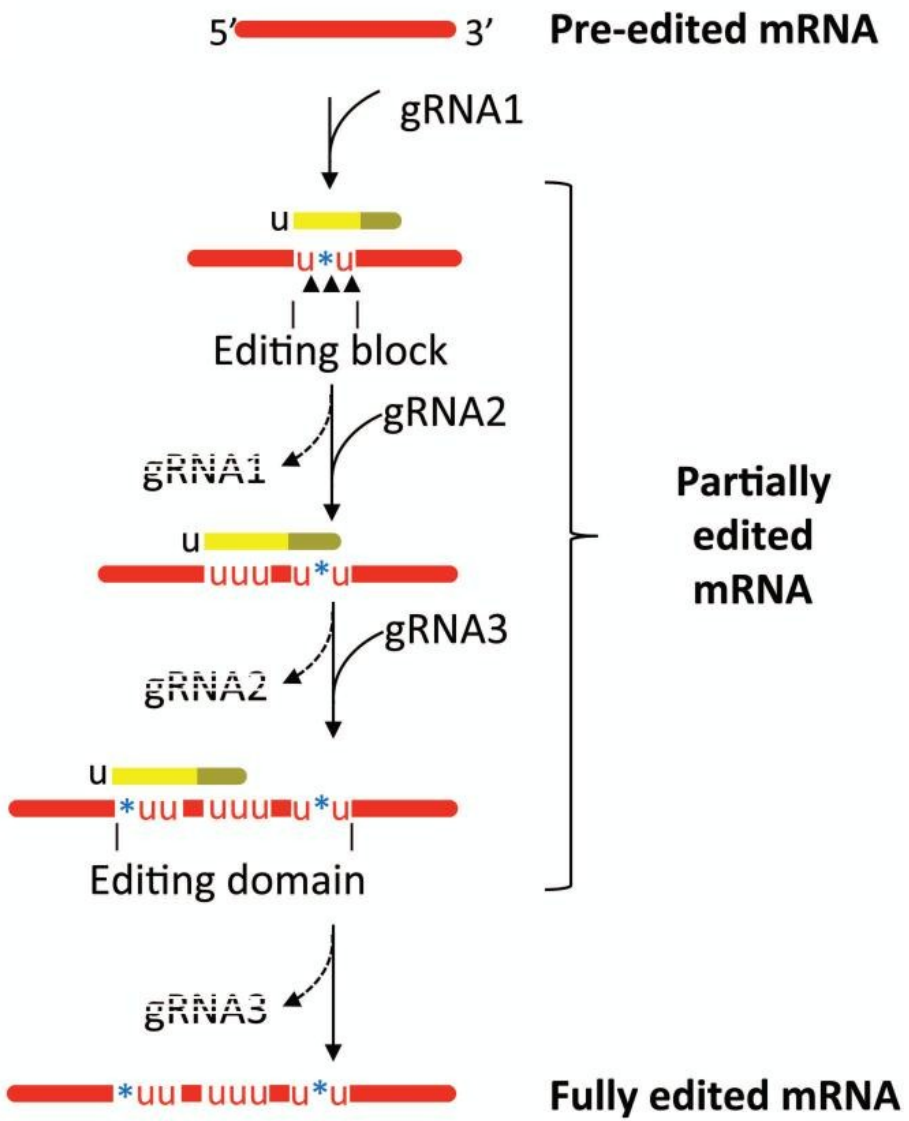
RNA editing in Trypanosomes

- the uridine (U) insertion/deletion editing
- occurs in the **kinetoplast**
- *trans*-acting **guide RNAs** and entails the **insertion** of hundreds **and deletion** of dozens of **U residues** from mitochondrial RNAs to produce mature, translatable mRNAs
- catalytic machinery termed the **editosome or RNA editing core complex (RE**



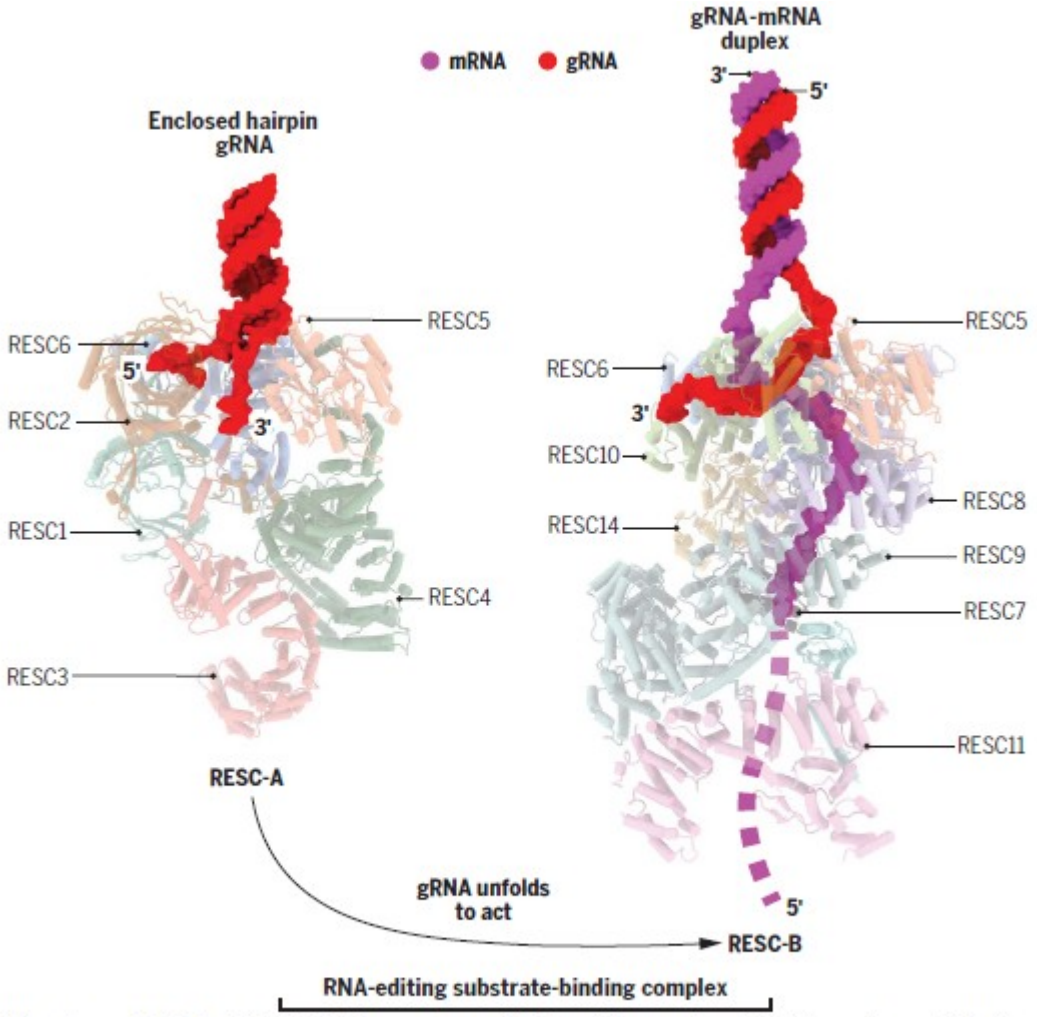
Pre-edited mRNAs are transcribed from mitochondrial **maxicircles**, while the majority of **gRNAs** are transcribed from the **minicircle** component of mitochondrial DNA. The gRNA 5' anchor region (olive) basepairs with the mRNA and the **gRNA** information region (yellow) **directs the number of U's inserted or deleted.** The gRNA 3' oligo(U) tail stabilizes the gRNA/mRNA interaction. Enzymes contained within the RNA editing core complex (**RECC**) catalyze **mRNA endonucleolytic cleavage** at an editing sites, **U insertion** by a 3' TUTase, and **U deletion** by a U-specific exoribonuclease as directed by the sequences of

RNA editing in Trypanosomes



Multi-round editing entails sequential utilization of **multiple gRNAs**. Because the anchor region of a given gRNA basepairs with edited mRNA sequence specified by the prior gRNA, editing progresses in a 3' to 5' direction along an mRNA. Multiple black arrowheads symbolize multiple editing sites within an editing block, as defined by the hybridized gRNA. Dashed gRNA labels indicate that they are turned over during/after an editing block has been processed. An editing domain is a stretch of mRNA sequence that requires the gRNA cascade for its processing.

RNA editing in Trypanosomes



Structures of RESCs. (Left) RESC-A sequesters gRNA termini, promoting hairpin formation and blocking mRNA access. (Right) RESC-A conversion into RESC-B unfolds gRNA and allows mRNA recognition, likely exposing editing sites to RECC-embedded enzymes.

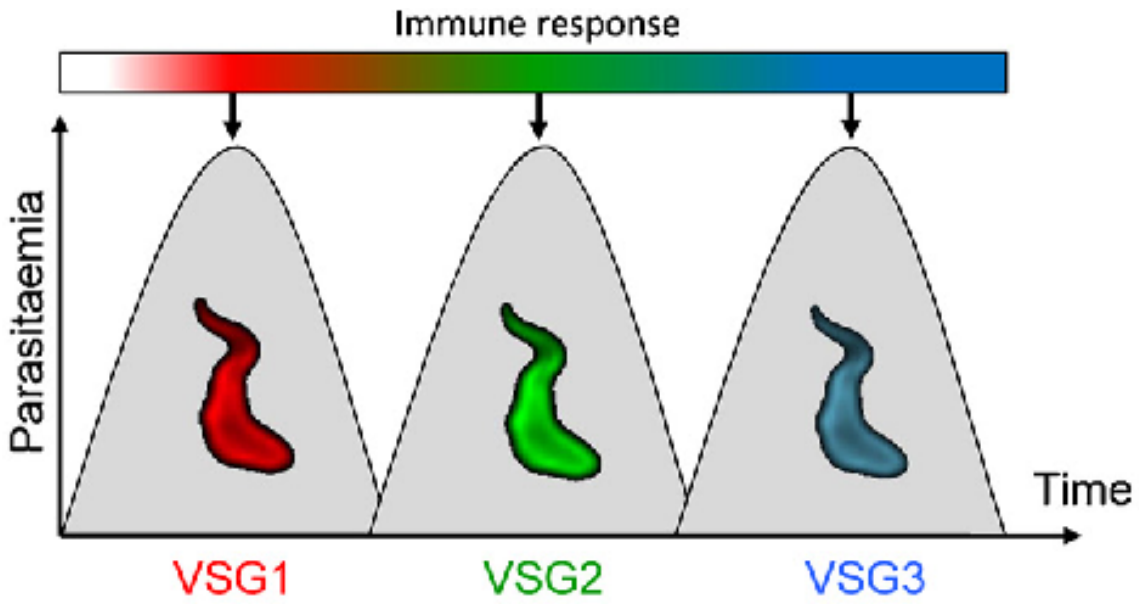
RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Structural basis of gRNA stabilization and mRNA recognition in trypanosomal RNA editing

Shiheng Liu†, Hong Wang†, Xiaorun Li, Fan Zhang, Jane K.J. Lee, Zihang Li, Clinton Yu, Jason J. Hu, Xiaojing Zhao, Takuma Suematsu, Ana L. Alvarez-Cabrera, Qiushi Liu, Liye Zhang, Lan Huang, Inna Aphasizheva, Ruslan Aphasizhev*, Z. Hong Zhou*

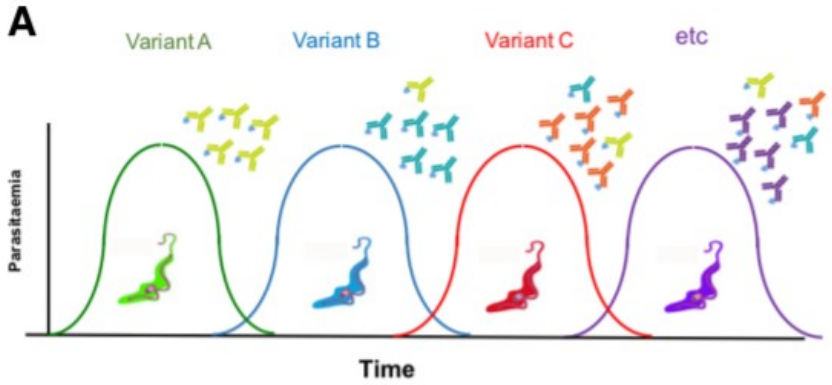
Antigenic variation: *Trypanosoma*



Consequences of antigenic variation
→ prolonging infection

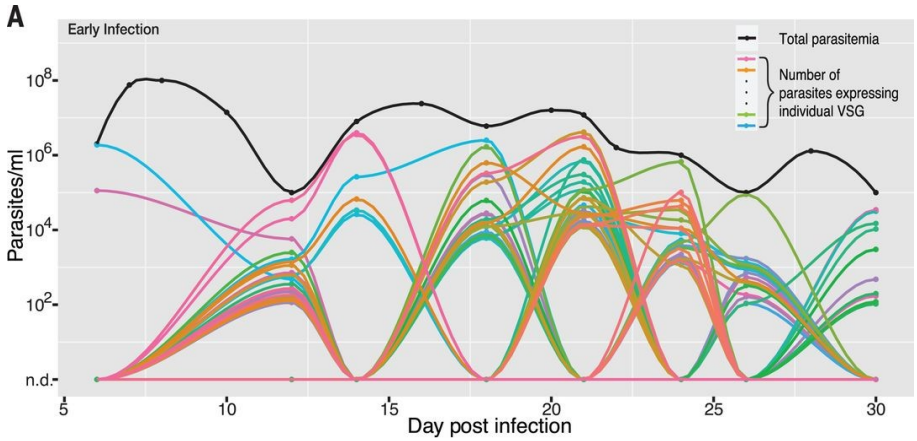
Traditional view

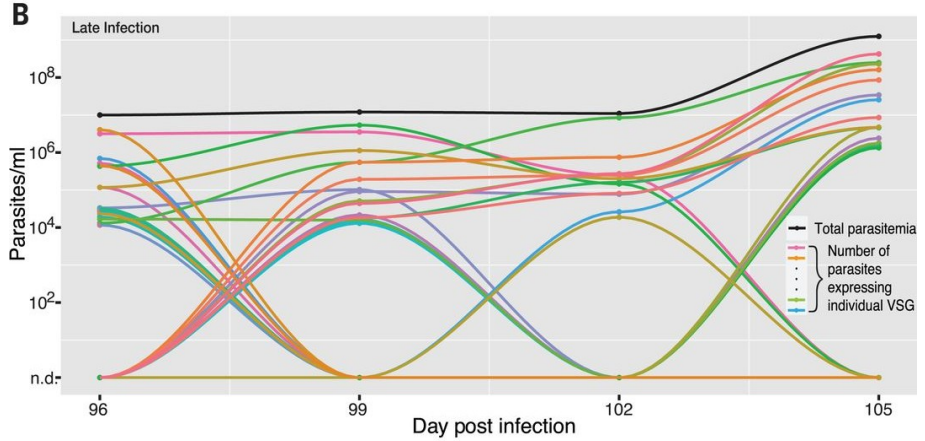
› each growth peak contains one variant
› reduction of each growth peak is due to antibodies against each VSG



Reality

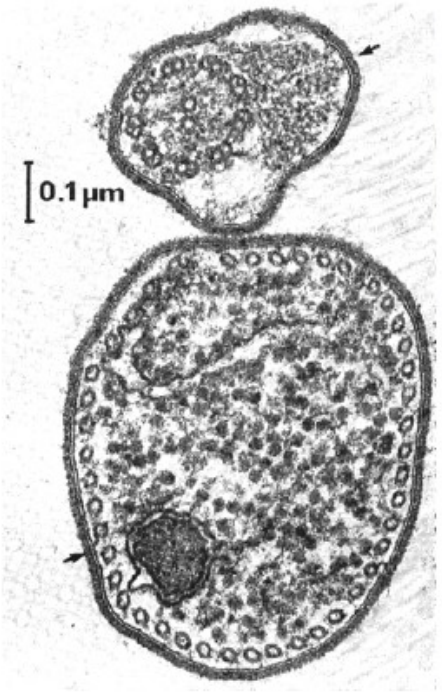
› each growth peak generally contains **many variants**

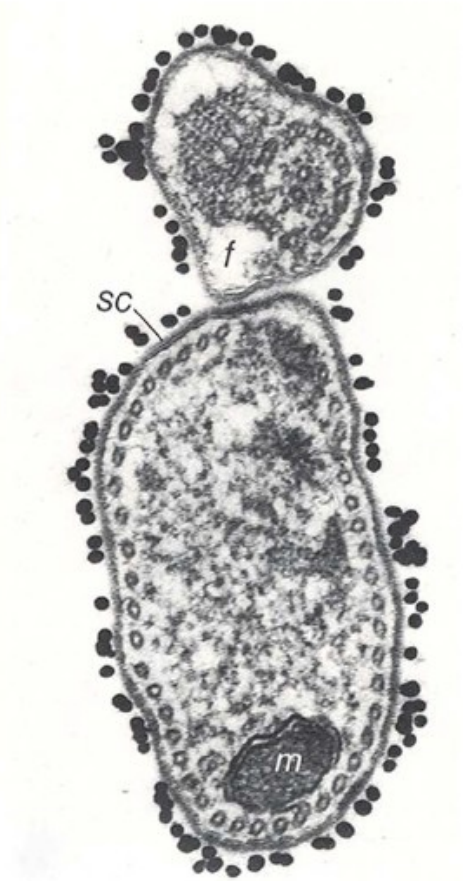


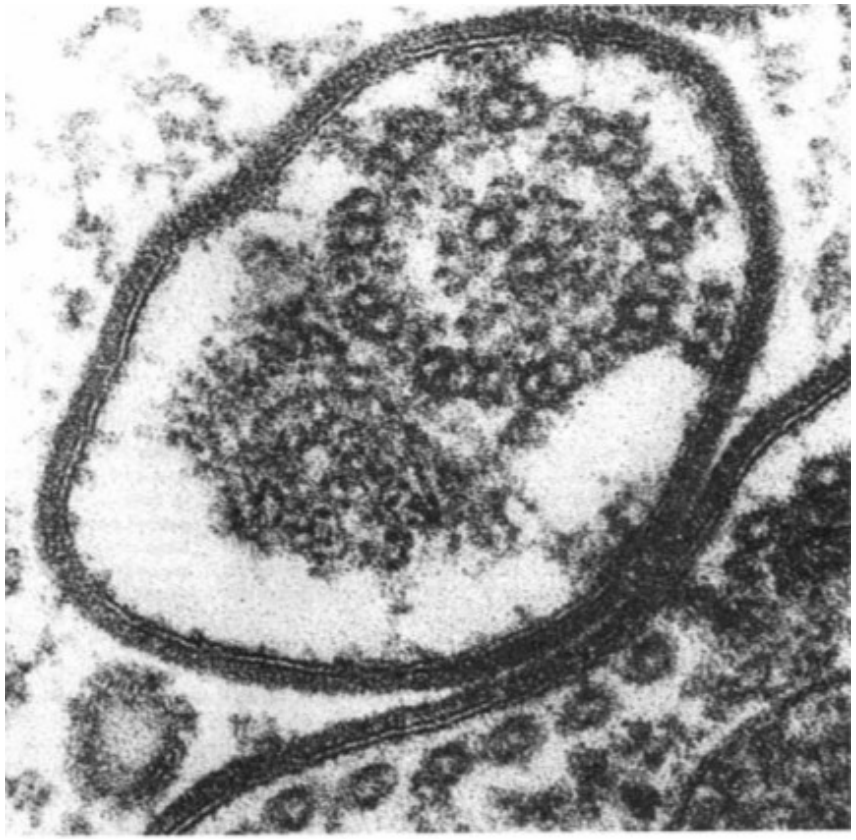


› **reduction of each growth peak depends on two factors:**
  › **differentiation to the non-dividing stumpy stage**
  › **anti-VSG antibodies**

Trypansoma antigenic variation
VSG (variant sufrace glycoprotein)

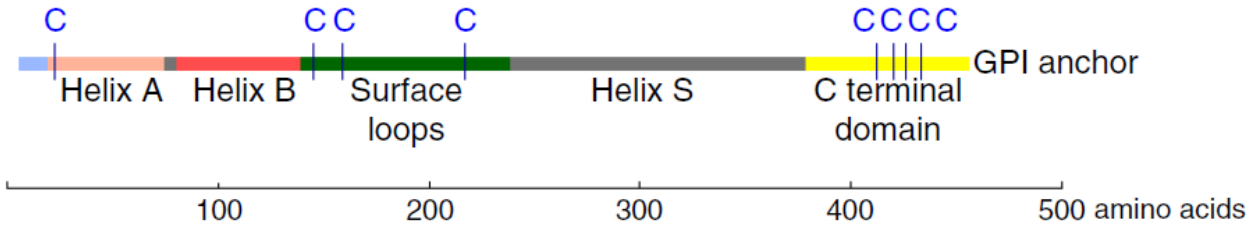


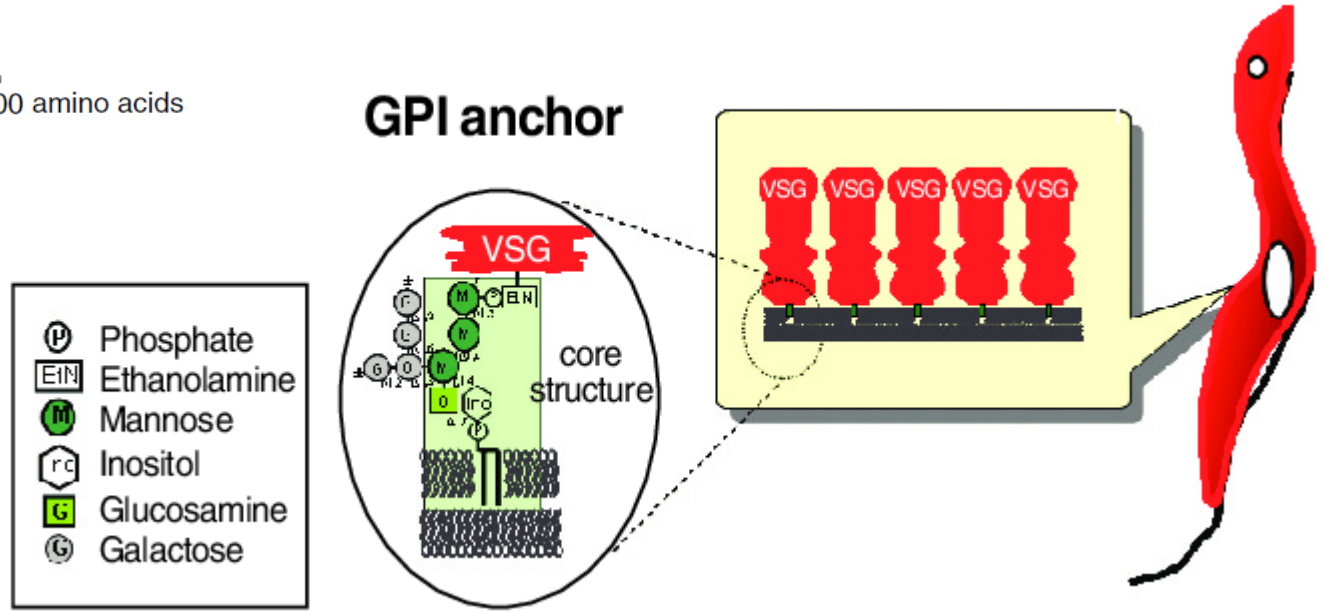




TRYPANOSOME'S **SURFACE COAT** of **VSG's** is visible as a diffuse, dark layer in an electron micrograph. A cross section of the parasite's body and flagellum. The double membrane just inside the surface coat is the cell membrane.

Variable surface glycoprotein (VSG)



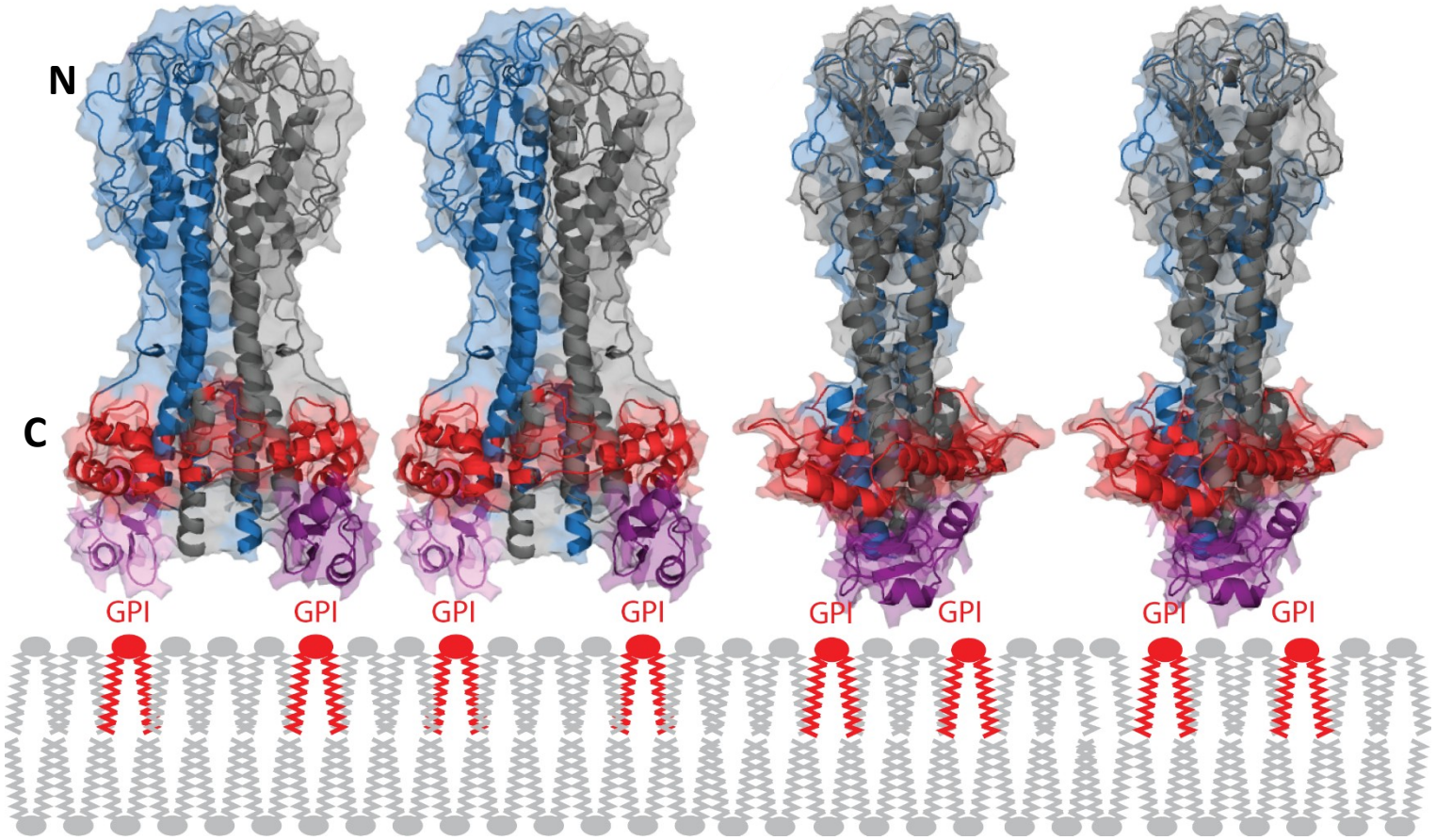


Glycosylphosphatidylinositol is a **phosphoglyceride** attached to the C-terminus of a protein during posttranslational modification. The hydrophobic C-terminal sequence is then cleaved off and replaced by the GPI-anchor.

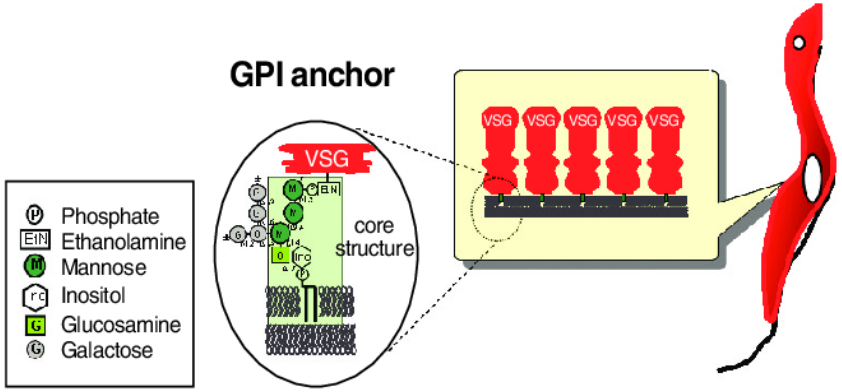
GPI is composed of a phosphatidylinositol group linked through a carbohydrate-containing linker (glucosamine and mannose glycosidically bound to the inositol residue) and via an ethanolamine phosphate (EtNP) bridge. **The two fatty acids within the hydrophobic phosphatidyl-inositol group anchor the protein to the cell membrane.**

Variable surface glycoprotein (VSG) bloodstream form

VSG homodimer

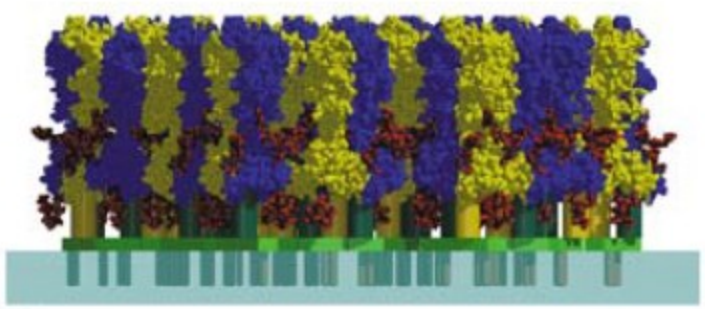


GPI anchor

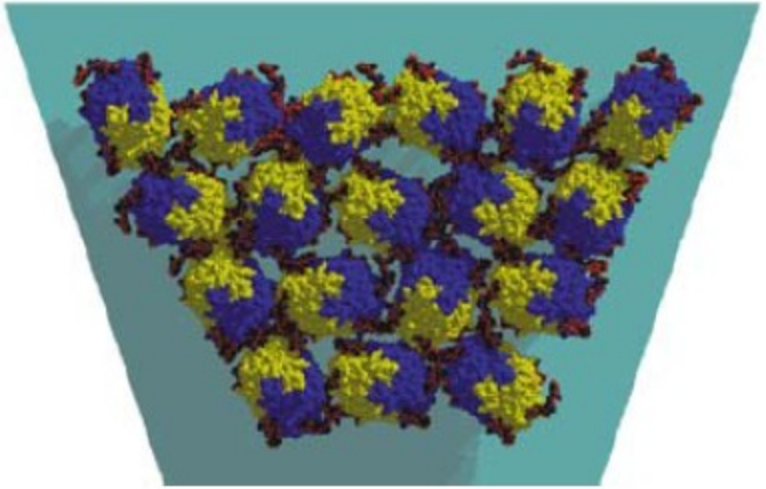


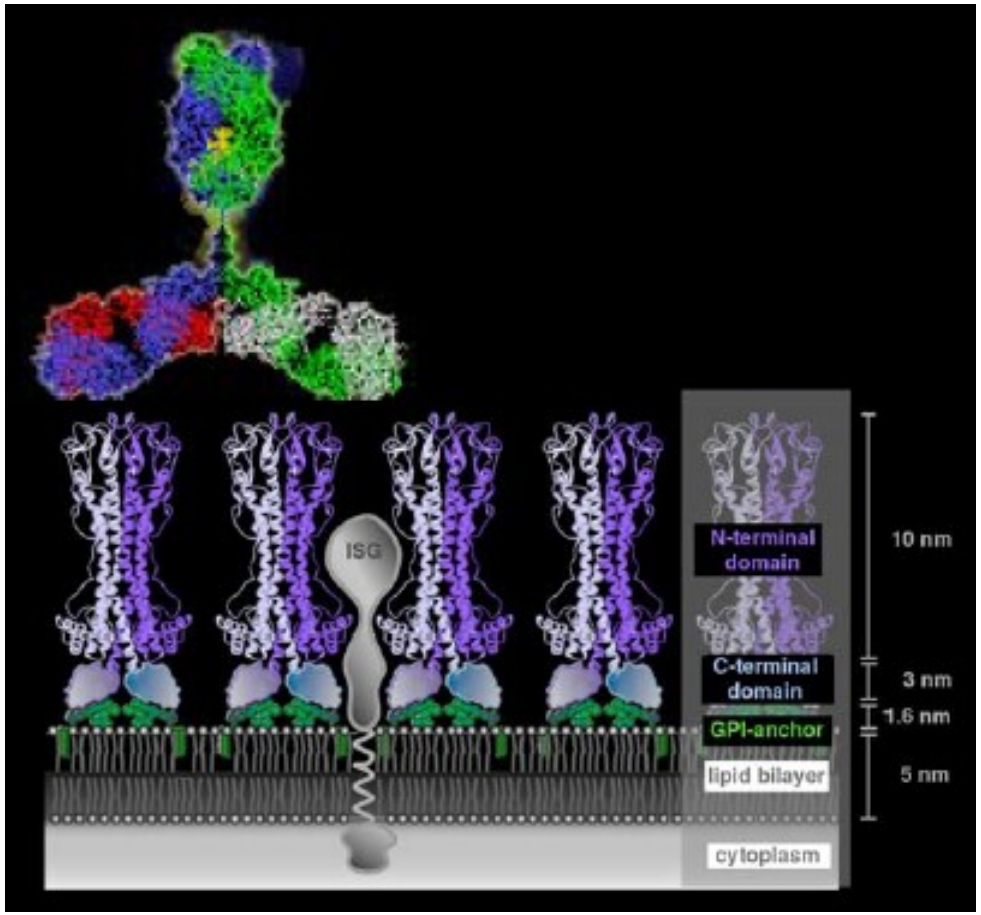
VSG in the surface coat: PROTECTION

… like a dense homogenous forest



a dense monolayer of ~5×10^6 identical VSG dimers

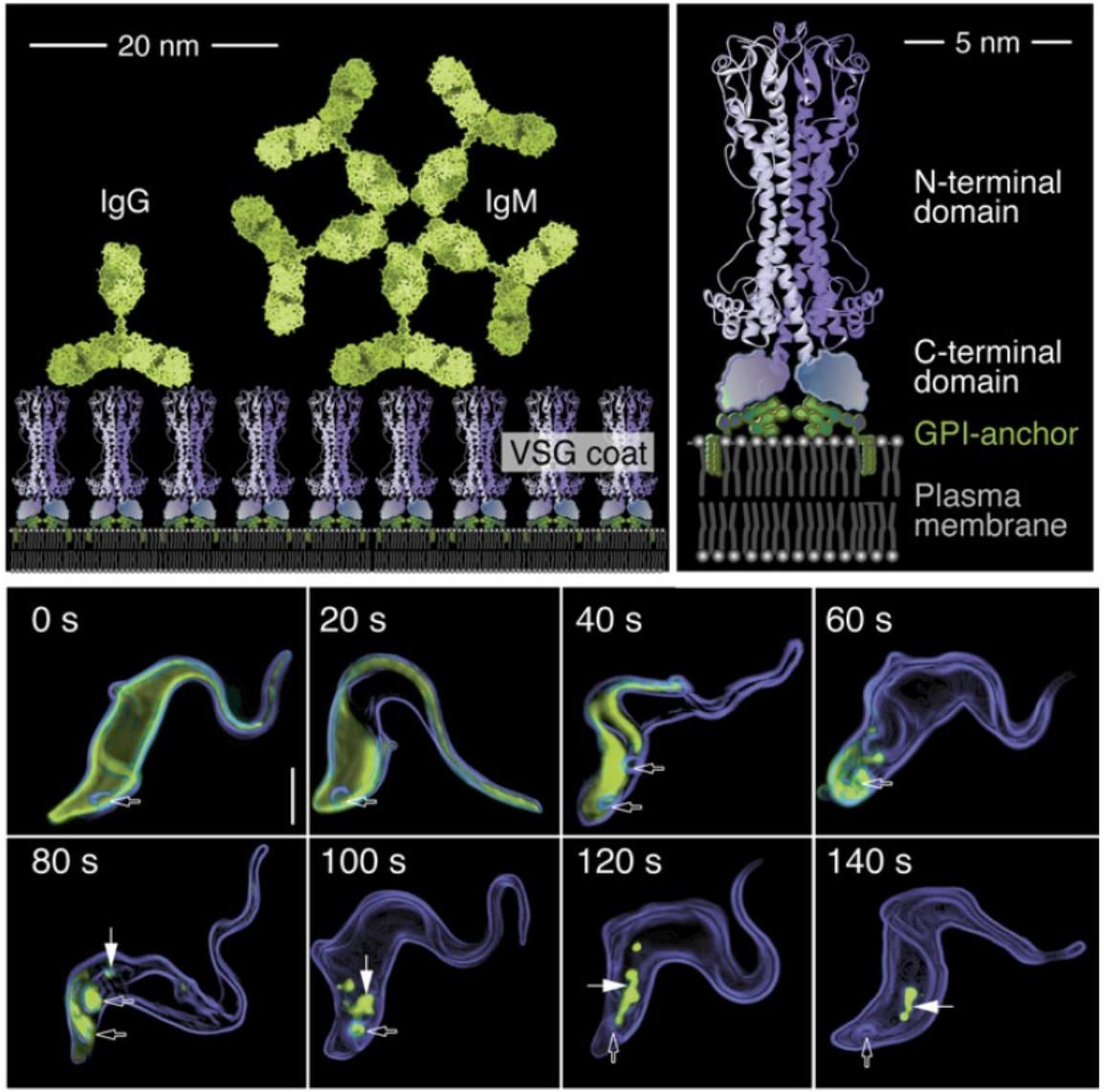




Antibody cannot access invariant antigen, only the VSG (this is why coat change is required).

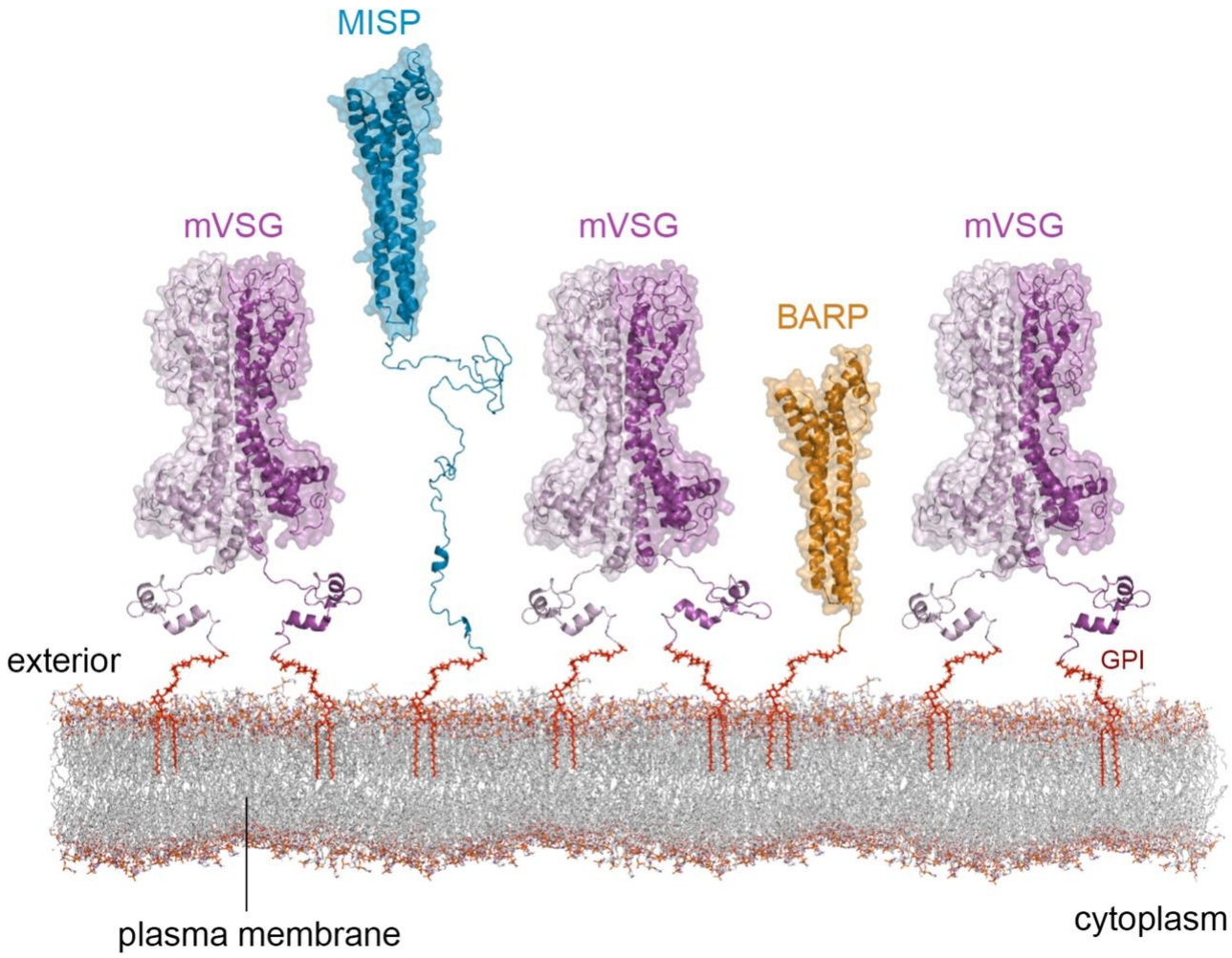
VSG in the surface coat: PROTECTION

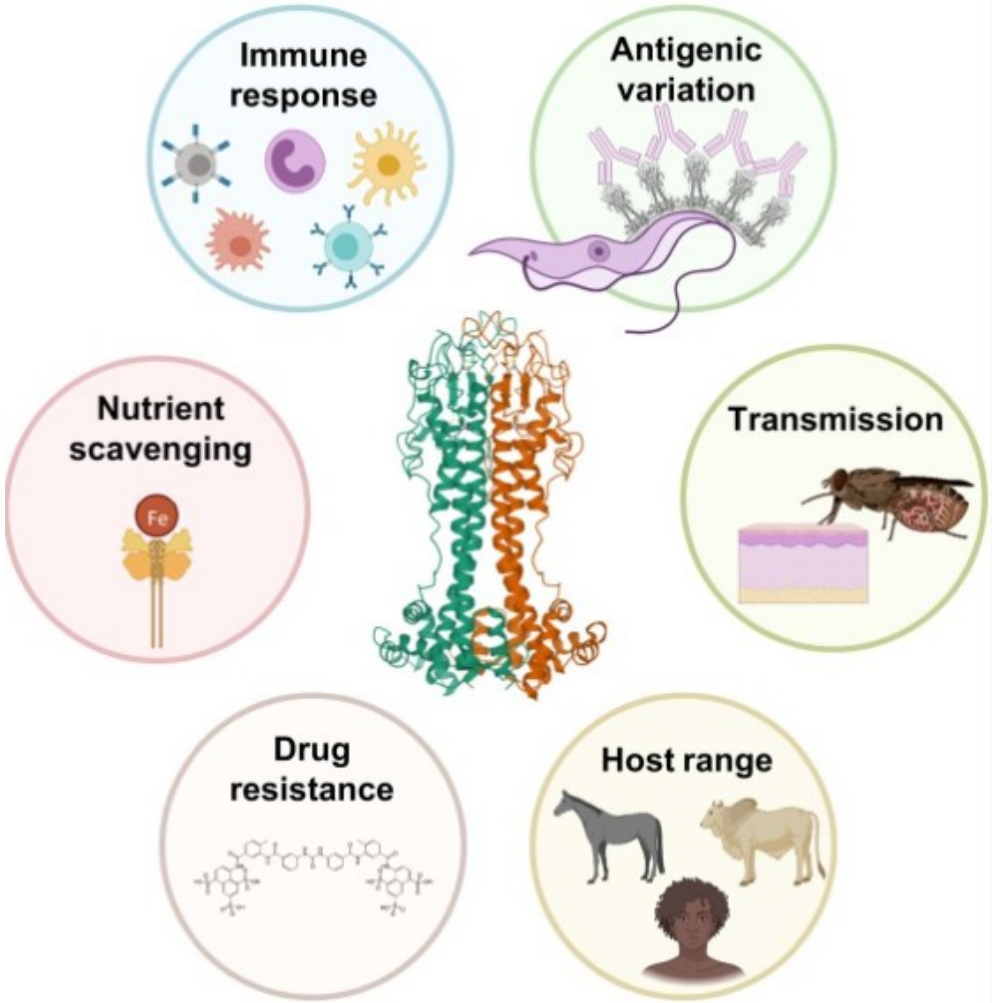
Host antibodies are removed from the surface



Engstler et al., Cell

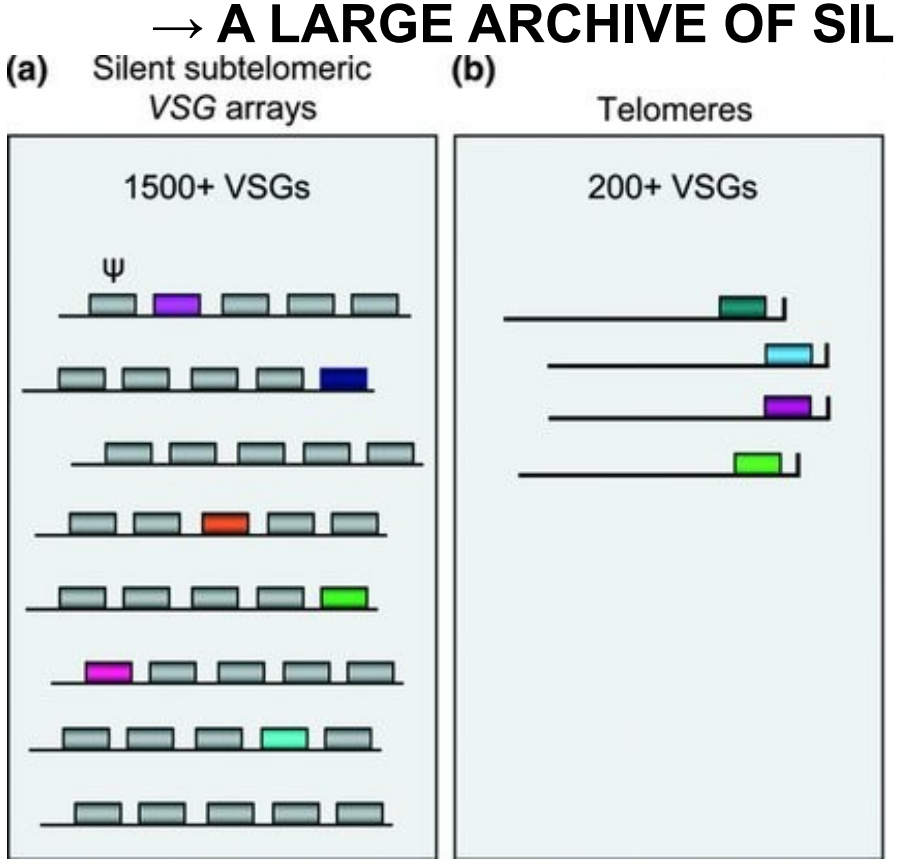
Variable surface glycoprotein (VSG) metacyclic form

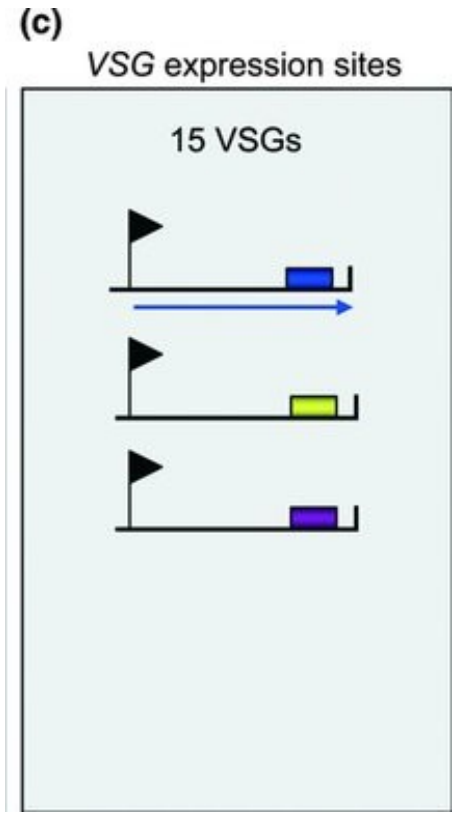


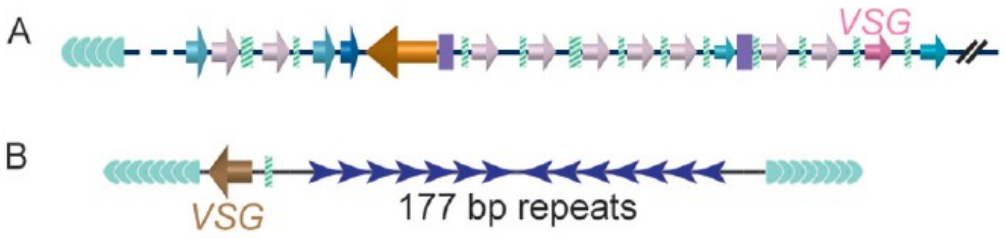


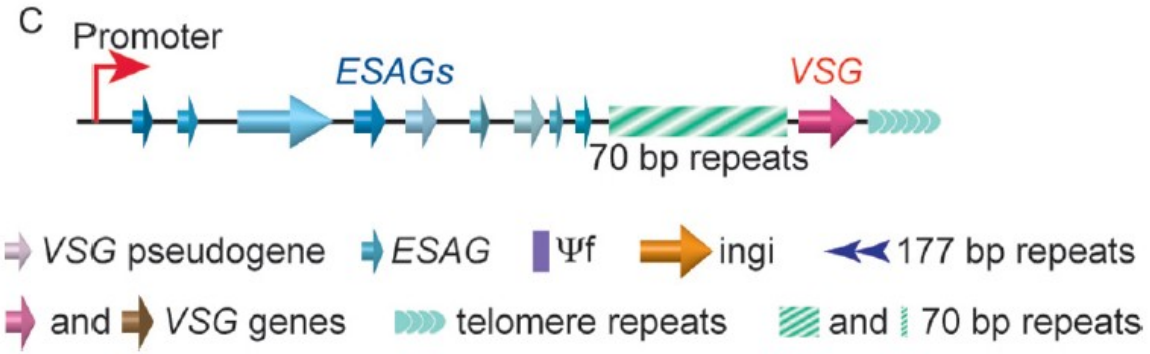
ANTIGENIC VARIATION IS BASED ON SILENT INFORMATION

→ A LARGE ARCHIVE OF SILENT VSG GENES







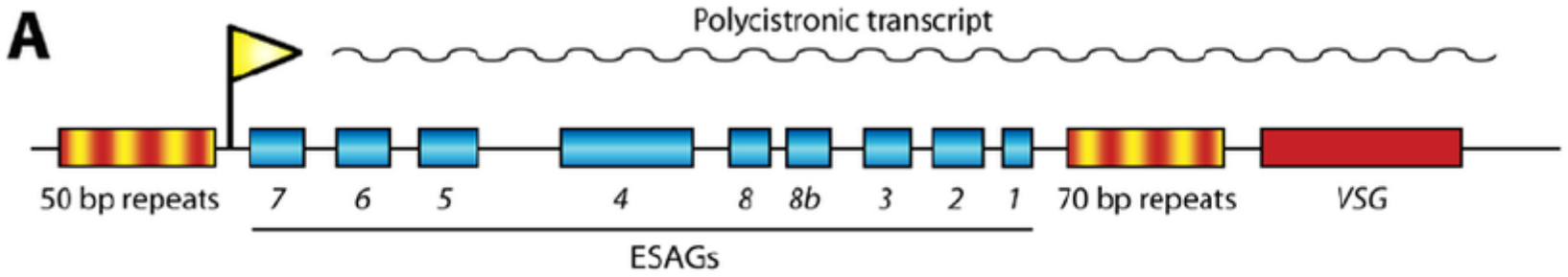


Monoallelic gene expression

„one gene at a time“

- **more than 1,000 VSG genes** and pseudogenes are packed as **gene arrays or located at subtelomeres**
- transcription occurs only from specialised subtelomeric transcription units known as **Bloodstream Expression Sites, BESs** (sometimes called telomeric expression sites, ESs)
  - ~15 BESs in a cell
  - polymorphic in size and structure
  - reveal a surprisingly conserved architecture in the context of extensive recombination

- **in a given cell only one BES is active at any time, and therefore only one VSG protein expressed**
- a diverse range of polymorphic genes called **Expression Site Associated Genes** (ESAGs)
  - membrane-associated or membrane-targeted proteins, transmembrane receptors, etc.

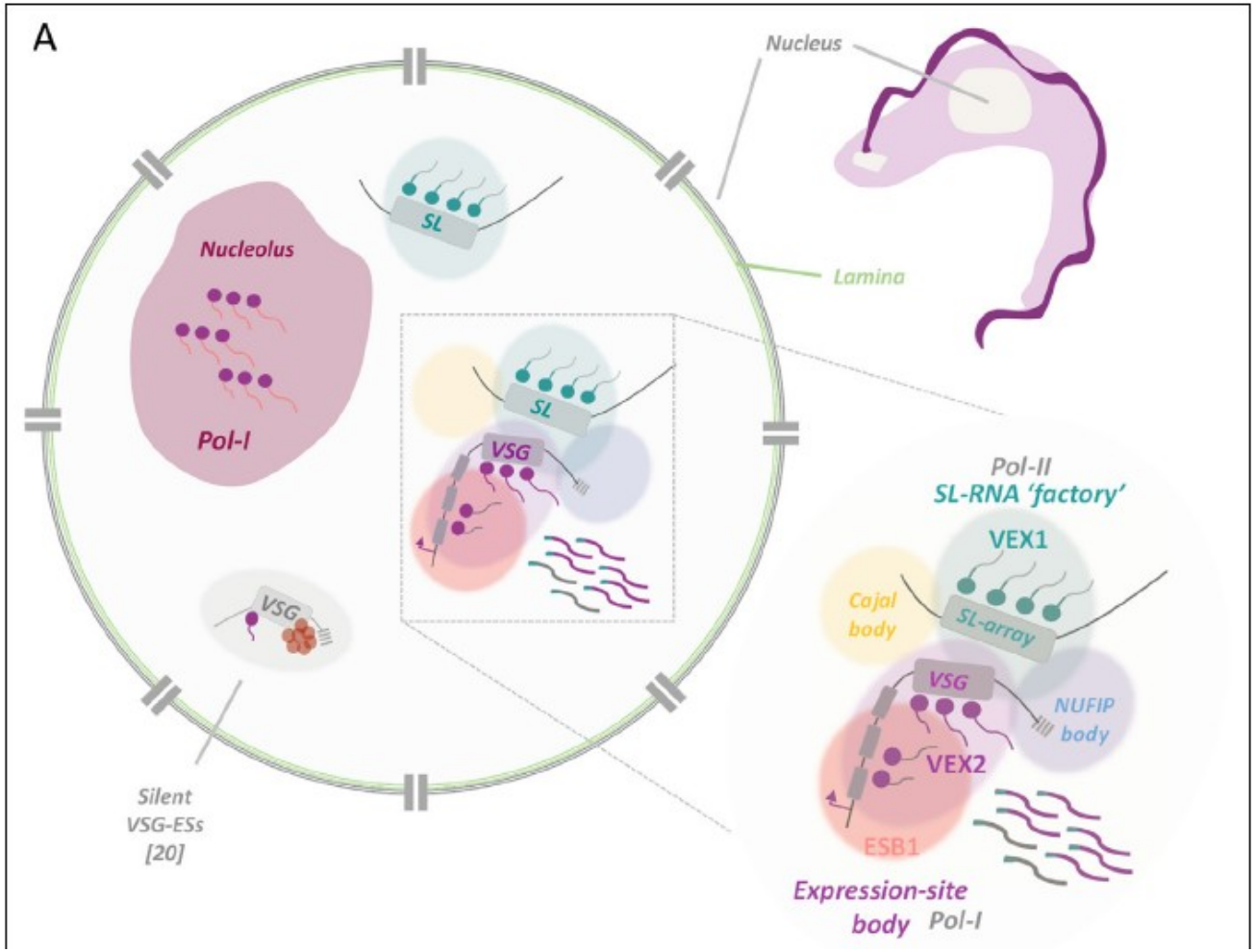


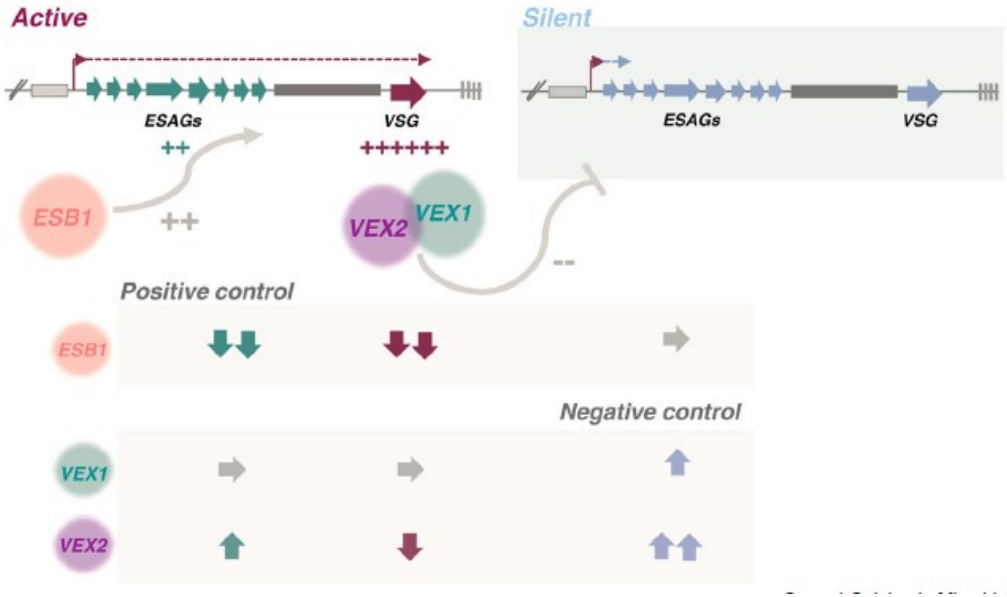
- polycistronic unit contains a number of **ESAGs** all expressed along with the active VSG
- **BES transcription results from the recruitment of RNA polymerase I** (pol I) on a promoter of the ribosomal type, in a non-nucleolar nuclear structure termed the “BES body“

Monoallelic gene expression

Bloodstream Expression Sites, BESs

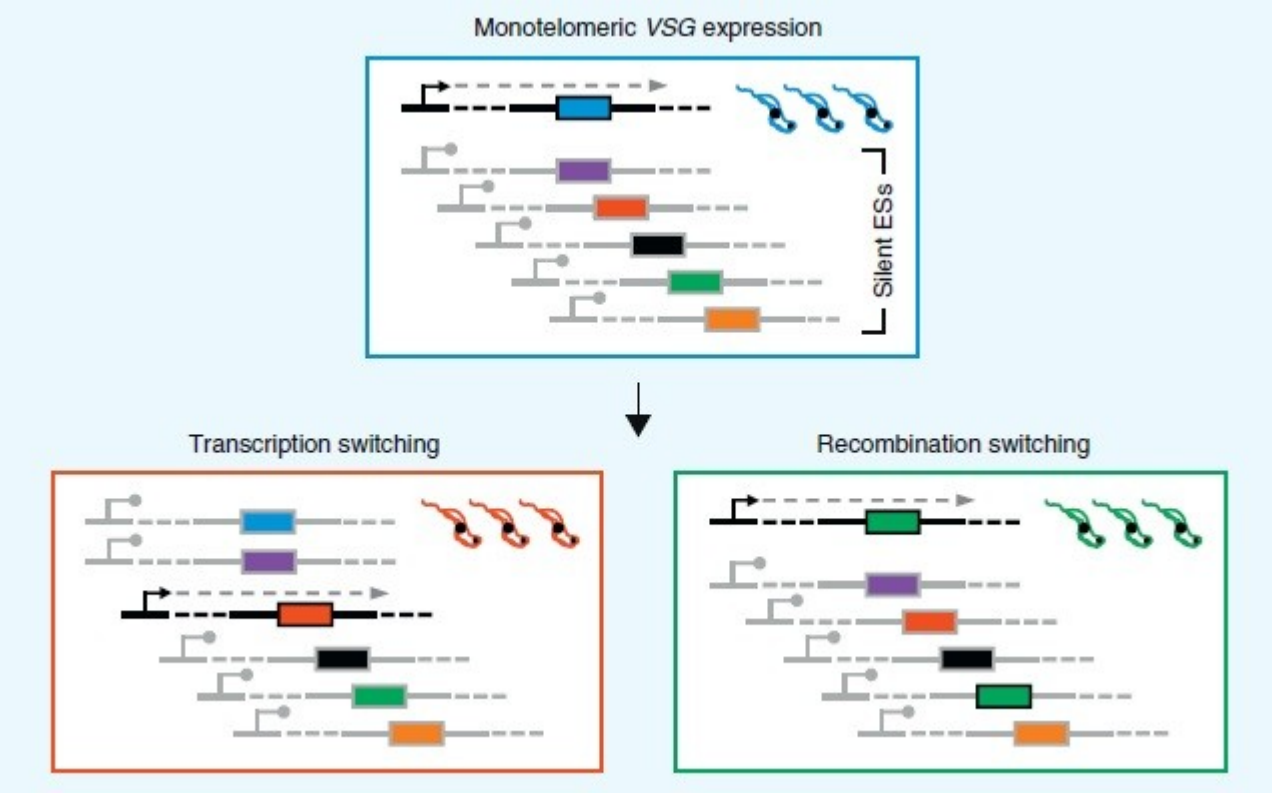
- spatial integration of transcription and splicing
- VEX2 – sustains exclusive interaction between a single VSG ES and SL-array





VSG SWITCHING

- **two main mechanisms** are used to change the expressed VSG gene and therefore perform antigenic variation (but there more mechanisms!)
  - **transcriptional switching between BESs** (a process called "*in situ* activation"), which turns off the active BES and turns on a new one
  - **homologous recombination** (gene conversion or telomere exchange), which replaces the VSG gene in the active BES



Early switches:

- <10% are transcriptional switching among the BES pool
- >90% are duplicative switching from silent archive (mostly from minichromosomes)

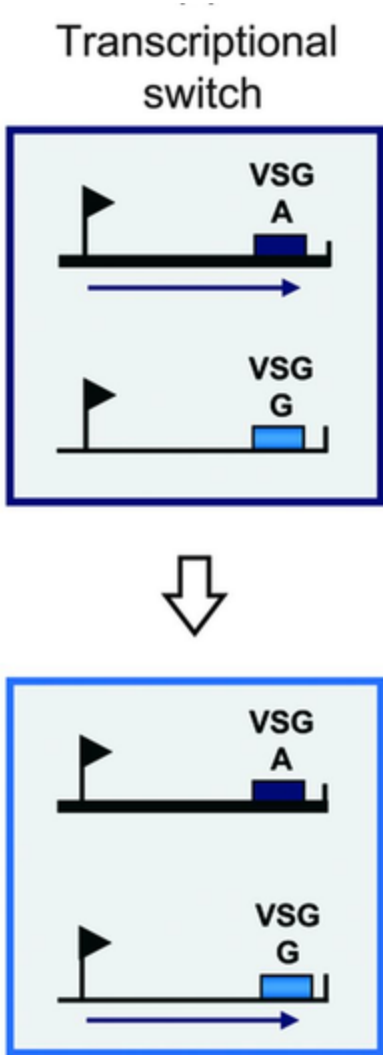
Late switches:

- duplicative switching from silent archive (from array intact genes)
- mosaic gene formation (by partial duplication from array pseudogenes)

VSG SWITCHING

transcriptional switching: "*in situ* activation"

- **to silence the active expression site and activate a new one** (an *in situ* switch)
- this method of switching accesses a **relatively small pool of ~15 VSG genes**
- on and off states differ at the level of transcript elongation
- recent evidence indicates that the different VSG expression sites each contain genes encoding receptor proteins that are optimized for **different hosts**
- this could mean that *in situ* switches are **important during the establishment of the *Trypanosoma* in a new host species**



VSG SWITCHING

recombination(al) switching

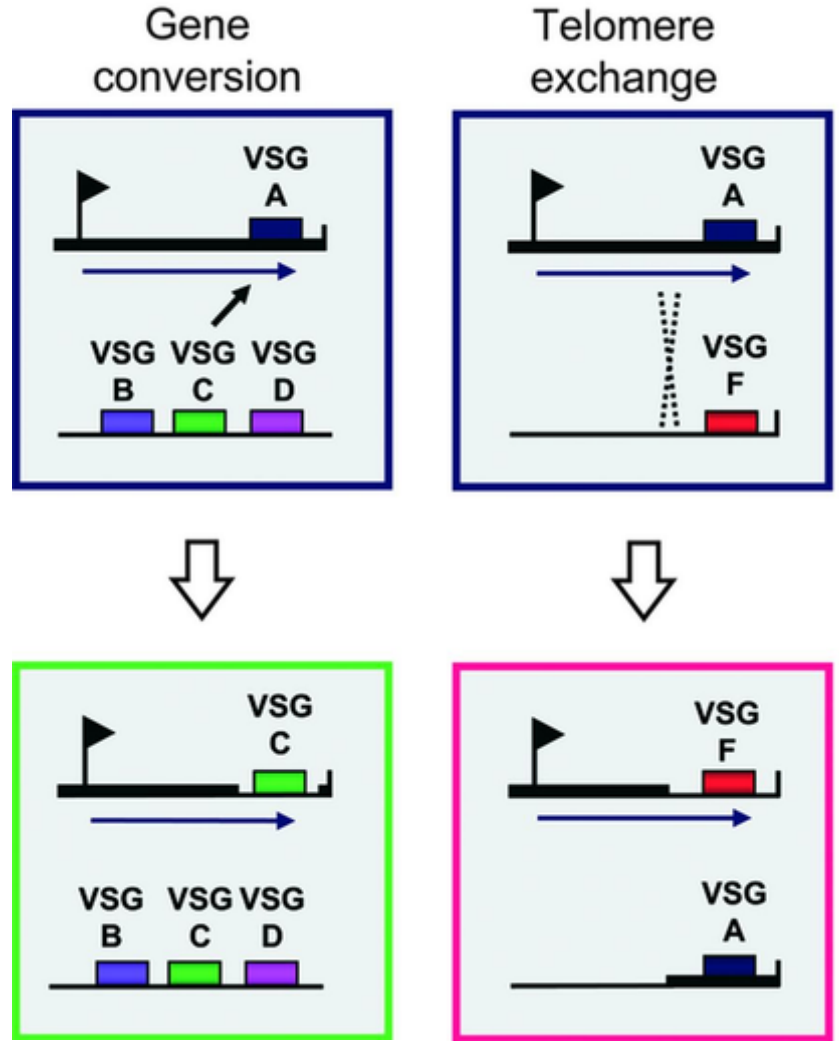
- recombination is central to antigenic variation, allowing the parasite **to utilise complete VSG archive**, typically by copying (**duplication** of) a different gene into the active BESs
- **gene conversions** or **telomere exchange**

gene conversions (array conversions)

- access the largest pool of VSG genes (virtually all of them)
- **a silent VSG gene is copied and inserted into the active expression site, replacing the old VSG gene**
- **1. array gene conversion**
- **2. telomere conversion**

telomere exchange

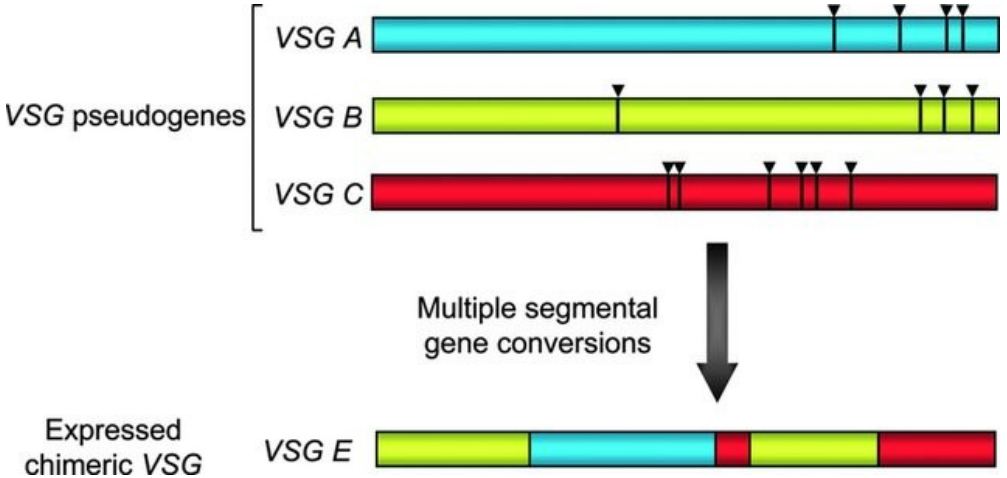
- a silent VSG gene at a **chromosome end** is **flipped into** the active VSG expression site

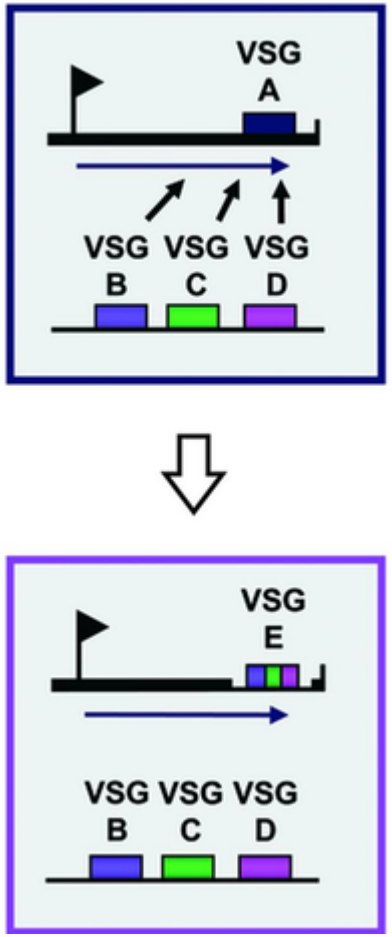


VSG SWITCHING

Segmental VSG conversion: mosaic genes

- during chronic infection, **novel MOSAIC GENES** are expressed
- they are assembled **from segments of damaged VSG genes** (pseudogenes)





Segmental gene conversion of multiple VSG pseudogenes can result in the creation of a **new functional chimeric VSG**. Three different VSG pseudogenes are indicated above, with disruptions of the ORF indicated with arrow heads and vertical lines. Multiple successive gene conversion reactions can take place, resulting in the creation of a new functional VSG which is a mosaic of segments of the different VSG pseudogenes.

VSG SWITCHING

Most VSG switching is recombinational…

› initially, mosaics were found only late in infection, or in distinct infections

Early switches:
› <10% are transcriptional switching among the BES pool
› >90% are duplicative switching from silent archive (mostly from minichromosomes)

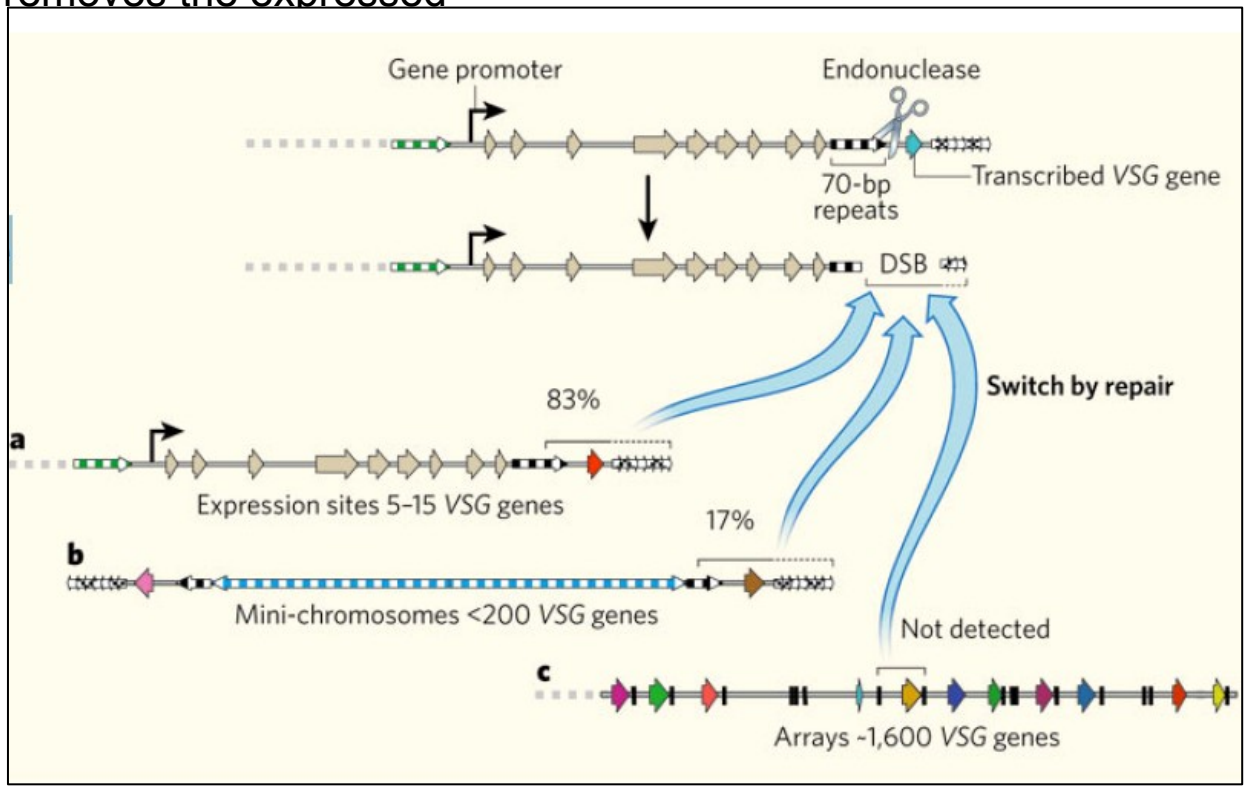
Late switches:
› duplicative switching from silent archive (from array intact genes)
› mosaic gene formation (by partial duplication from array pseudogenes)

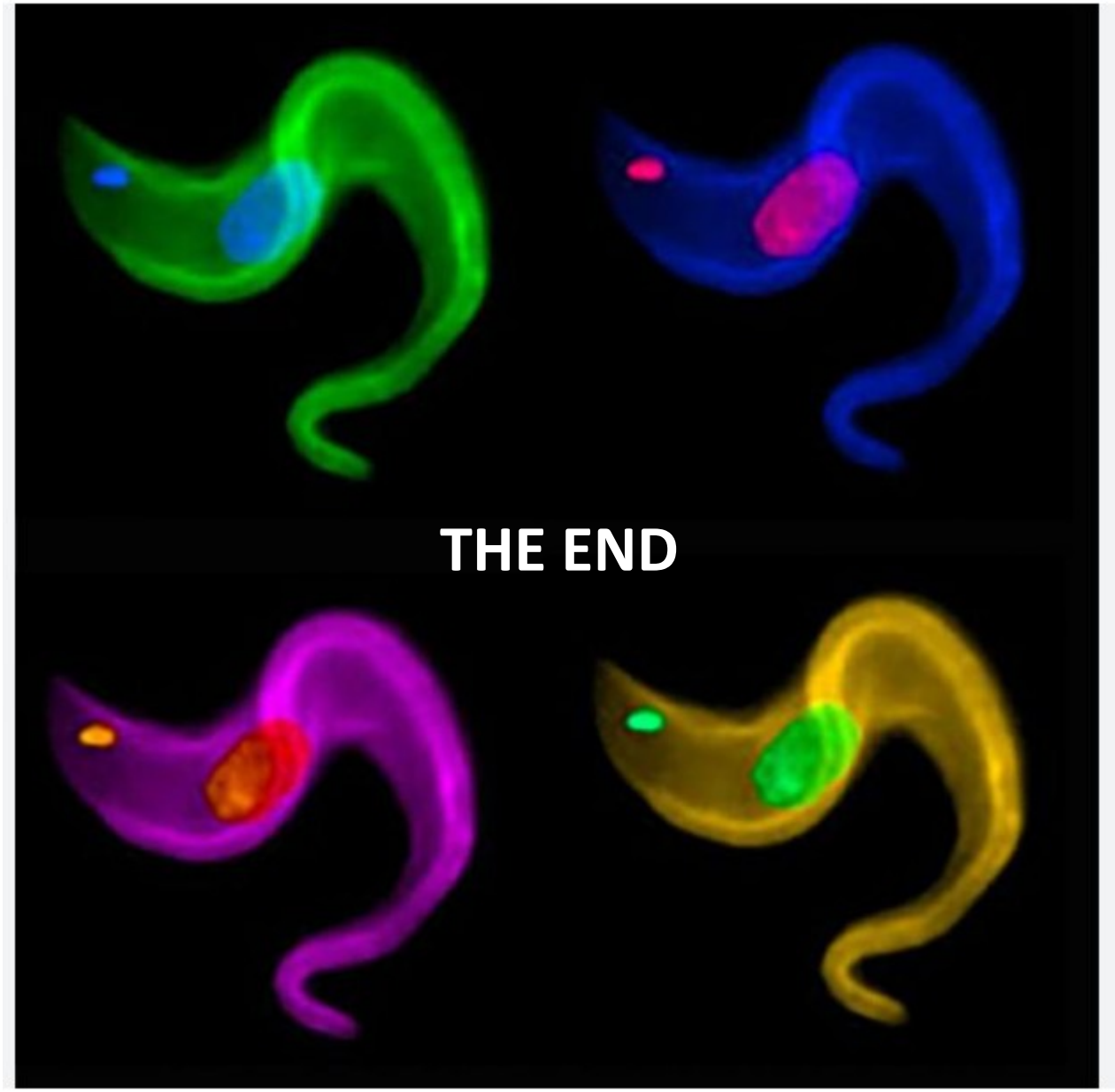
VSG SWITCHING

Most VSG switching is recombinational…

These DNA rearrangements probably are triggered by a DNA double strand break (DSB) in the 70-bp repeats of the BES.

- artificial induction of a DSB triggers recombinational switching
- **DSB appear naturally in the 70-bp repeats of the BES**
- DSB formation is followed by creation of a **gap** (which removes the expressed VSG gene)
- the **gap requires repair** – from a silent gene
- **mosaic formation** probably occurs differently, by recombination within the VSG coding sequence




THE END