



Universidade Federal de Santa Catarina  
Centro de Ciéncia Biológicas  
Departamento de Microbiologia, Imunologia e Parasitologia



# *In silico* analysis and biological significance from genomic data

Edmundo Grisard

MIP/CCB/UFSC

## Background:

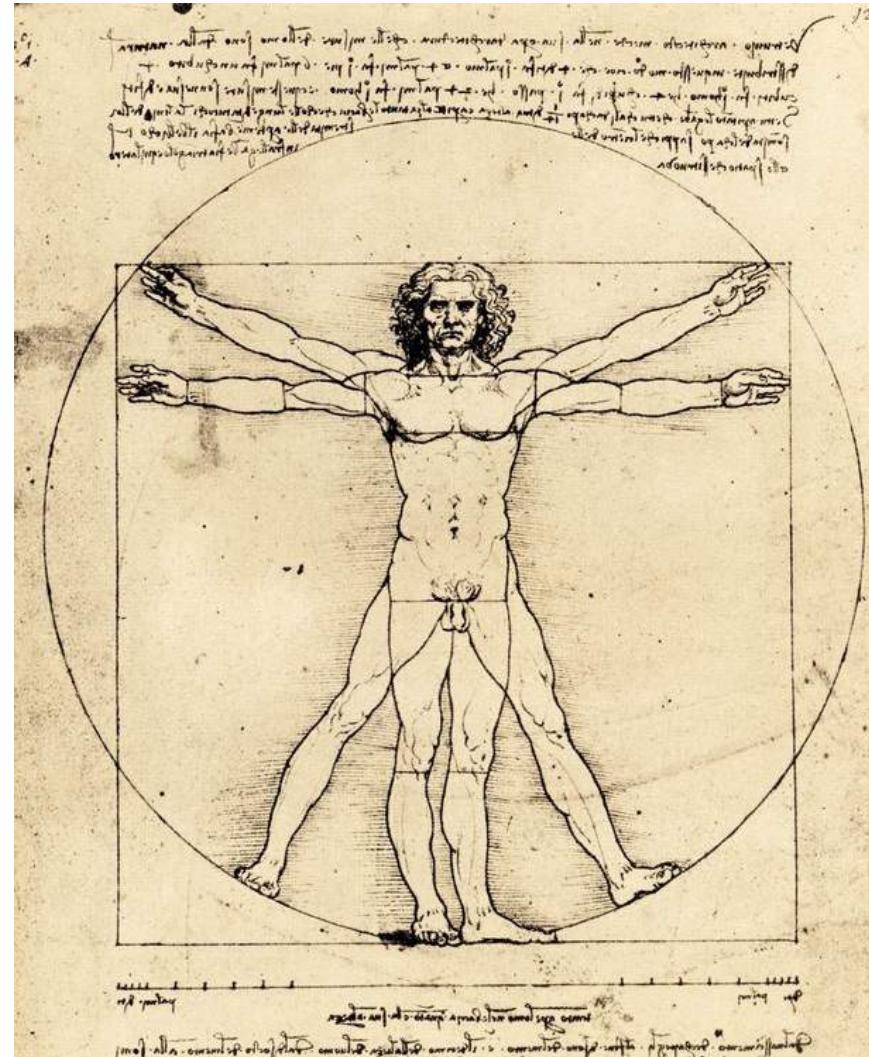
- Biologist
- PhD in Parasitology
- Professor of Cellular and Molecular Parasitology @ UFSC
- CNPq research fellow

## Research interests:

- Hemoflagellate protozoan parasites:
  - *Trypanosoma* spp. – Chagas Disease
  - *Leishmania* spp. – Cutaneous/Visceral leishmaniasis
    - Host-Parasite interaction
    - Comparative Genomics and Proteomics
    - Phylogeny / Taxonomy
    - Diagnostics
    - Strain characterization

Lab webpages: [www.proto.ufsc.br](http://www.proto.ufsc.br) / [www.bioinformatica.ufsc.br](http://www.bioinformatica.ufsc.br)  
E-mail: [edmundo.grisard@ufsc.br](mailto:edmundo.grisard@ufsc.br)

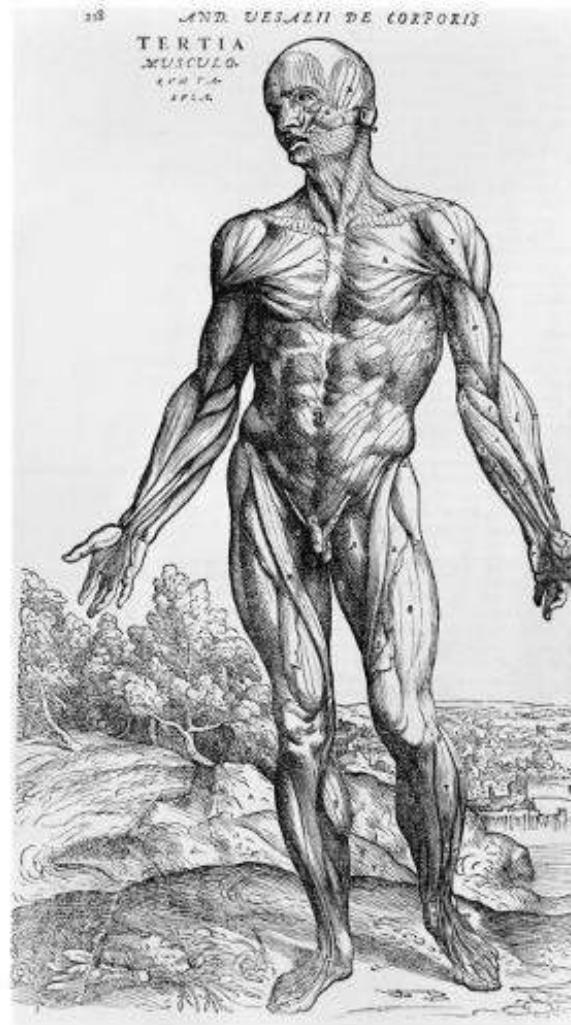
# The complexity of life



*Vitruvian Man (1490)*  
Leonardo Da Vinci (1452-1519)

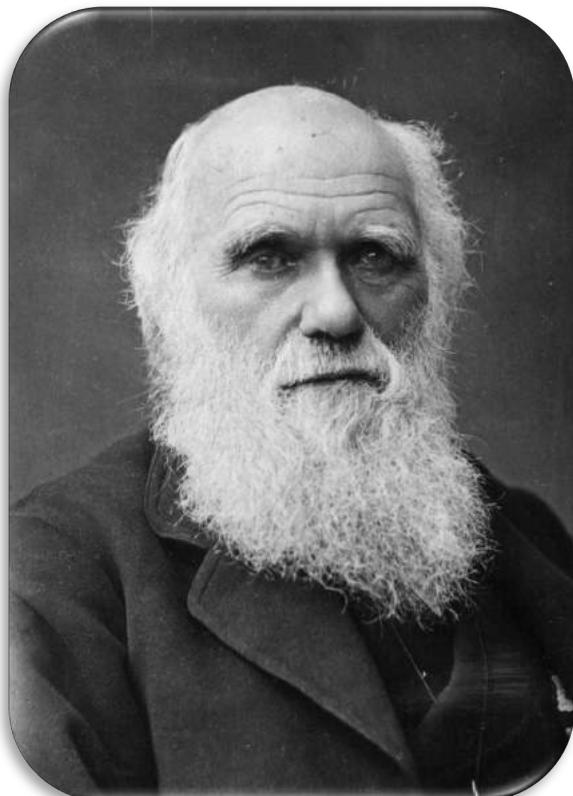
---

# Getting deeper, in detail...

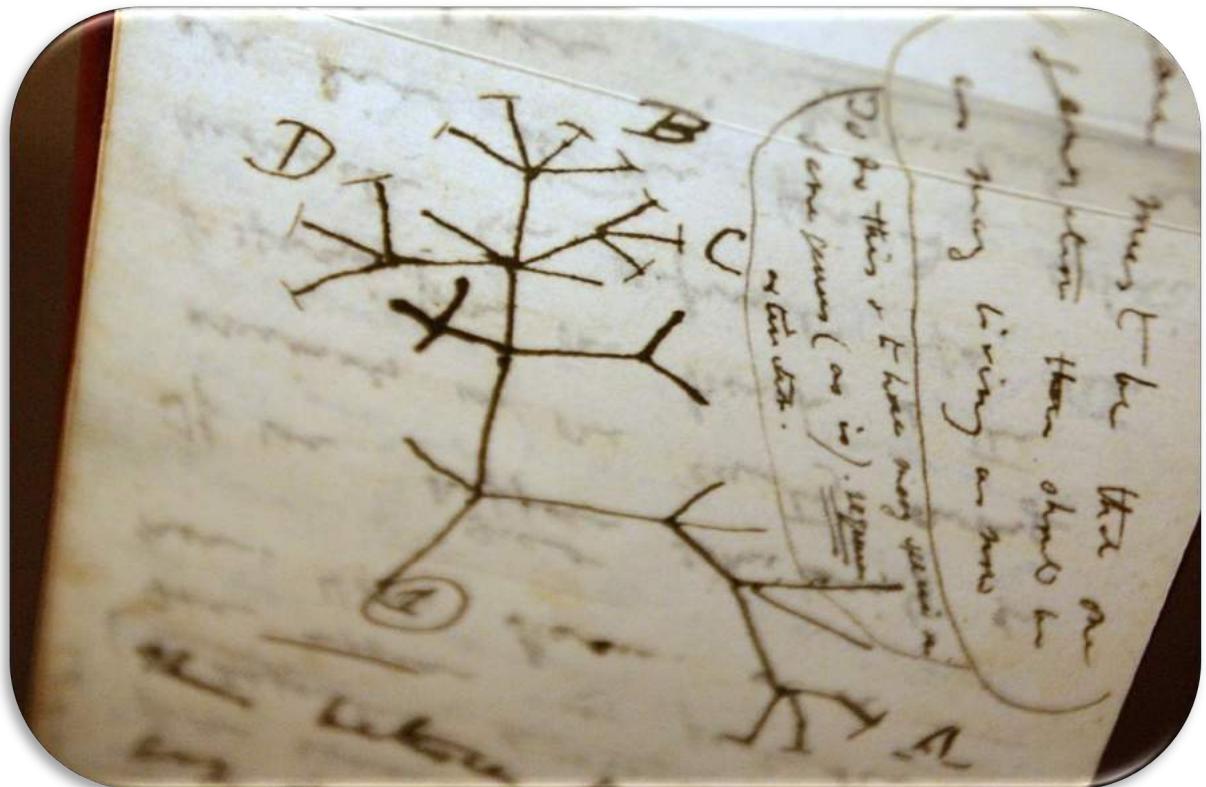


*De Humanis Corporis Fabrica Libri Septem* (1543)  
Andreas Vesalius (1514-1564)

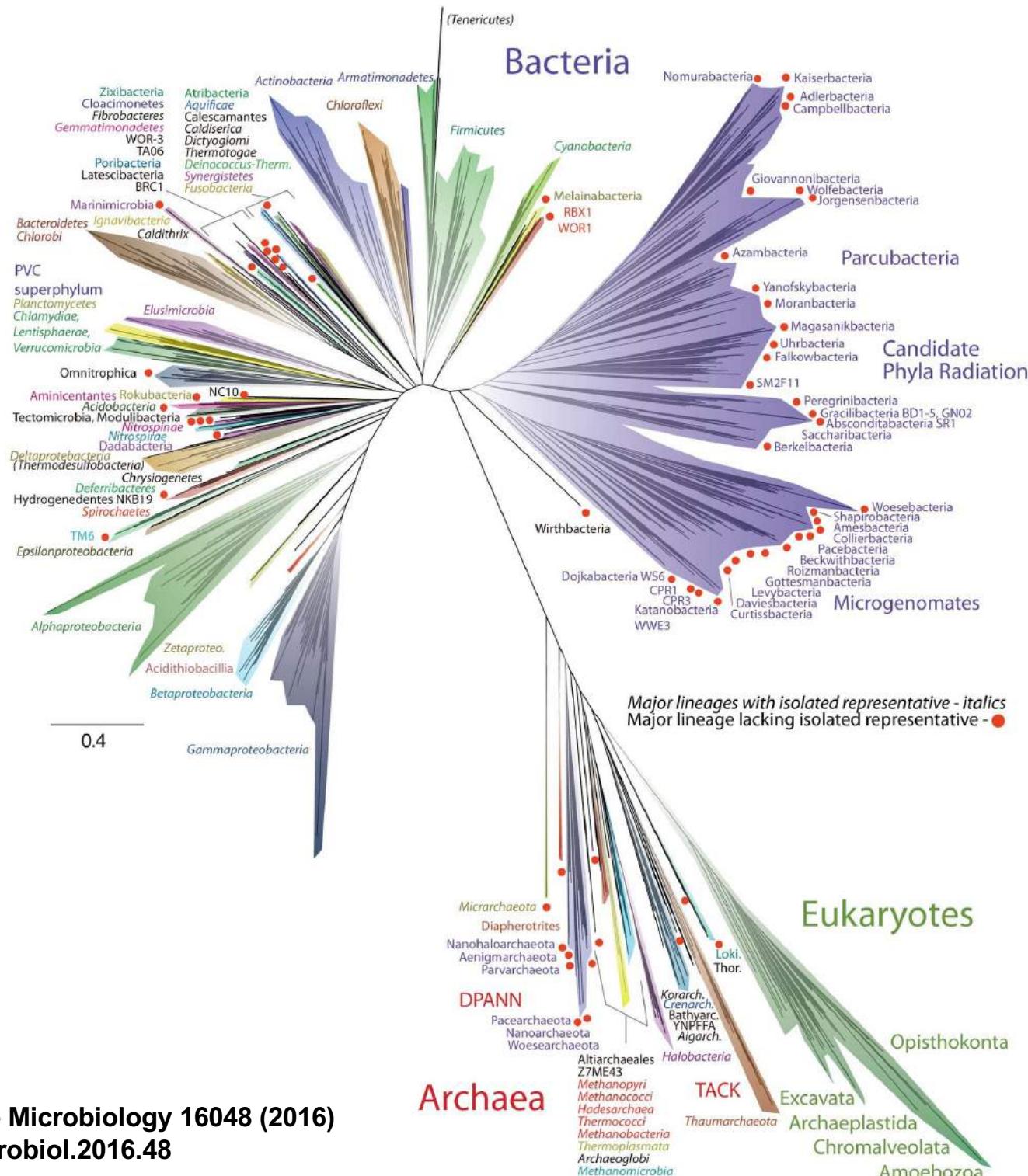
“Organisms evolve through adaptative processes  
in response to the environment “



Charles R. Darwin  
(1809-1882)



Darwin's notebook “B”



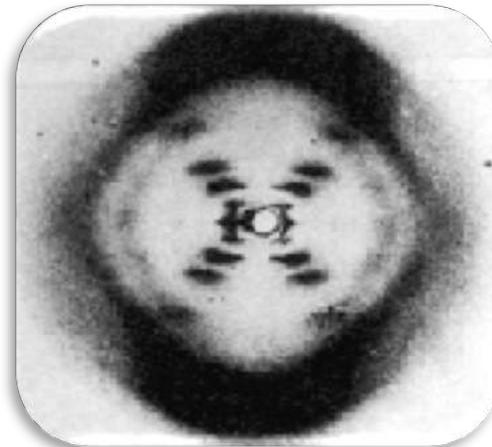
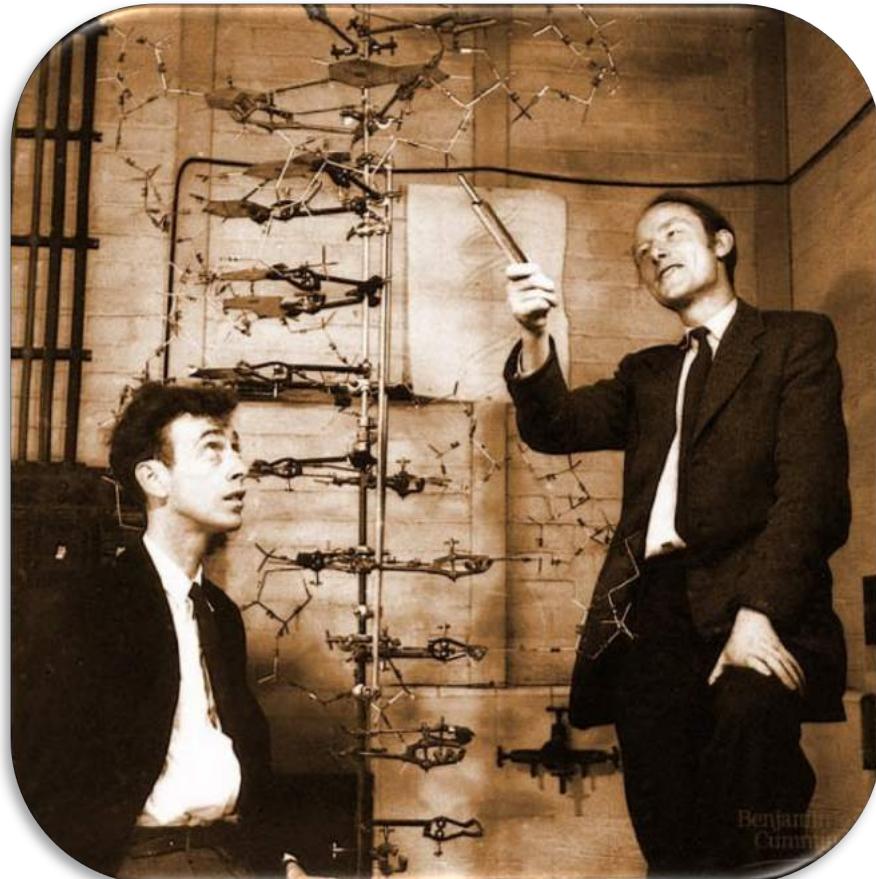
No. 4356 April 25, 1953

NATURE

737

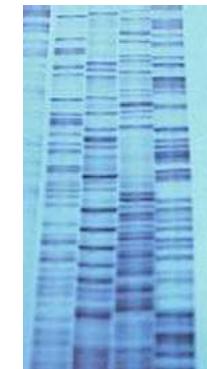
## MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

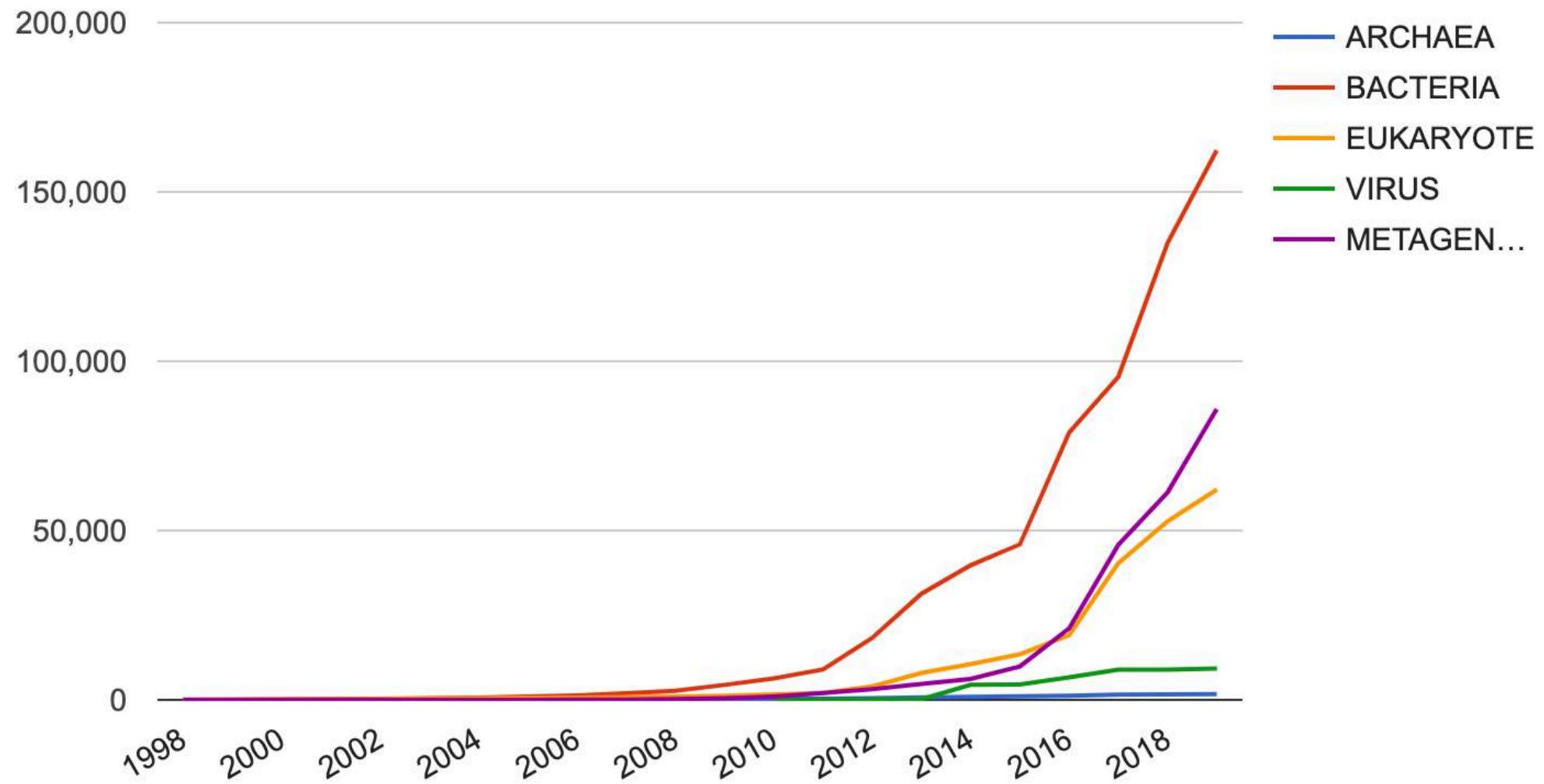


DNA structure (1953)  
James D. Watson and Francis Crick

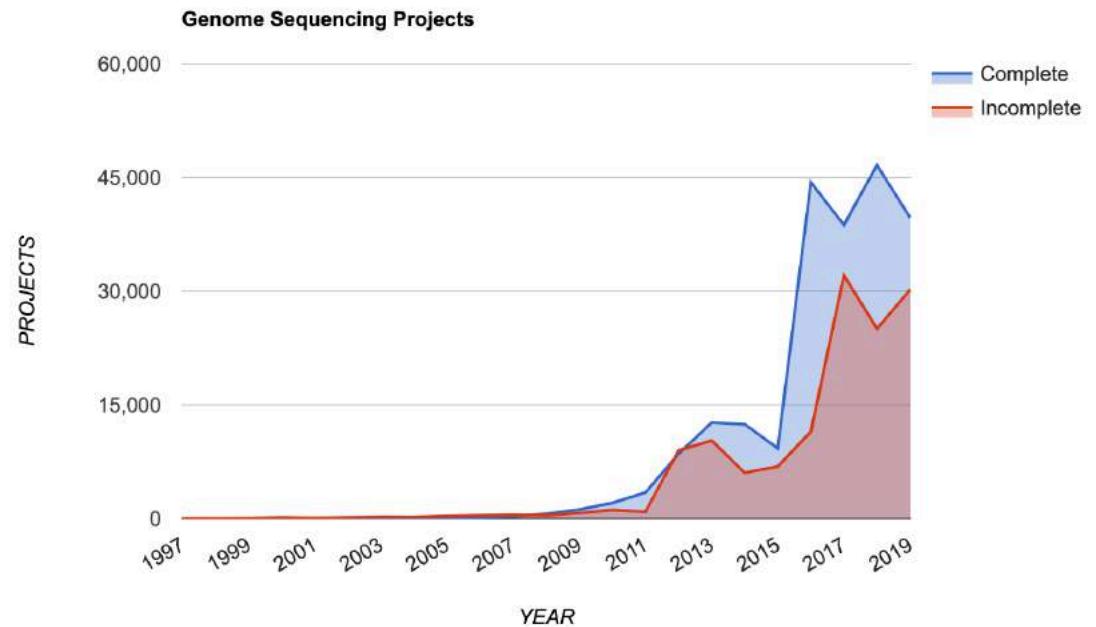
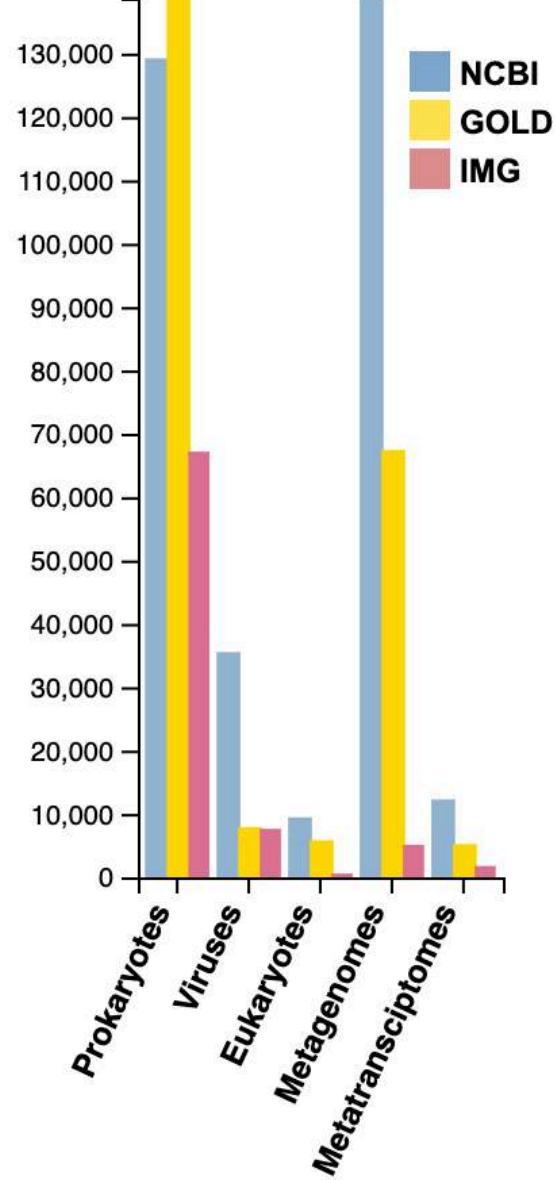
- **Gilbert & Sanger (70's):**
  - ~10 bp / year
- **Manual sequencing - slab gels (80-90's):**
  - Radioisotopes
  - 150-200 bp / week
- **Automated sequencers (90's onwards):**
  - 1<sup>st</sup> generation: 300-400 bp<sup>+</sup>
  - 2<sup>nd</sup> generation: 700 bp – 1.5 Kb
  - 3<sup>rd</sup> generation (output per run):
    - PacBio: >10 Gb (50% > 50Kb)
    - Illumina: >2,000 Gb
    - Mini Ion: >30 Gb



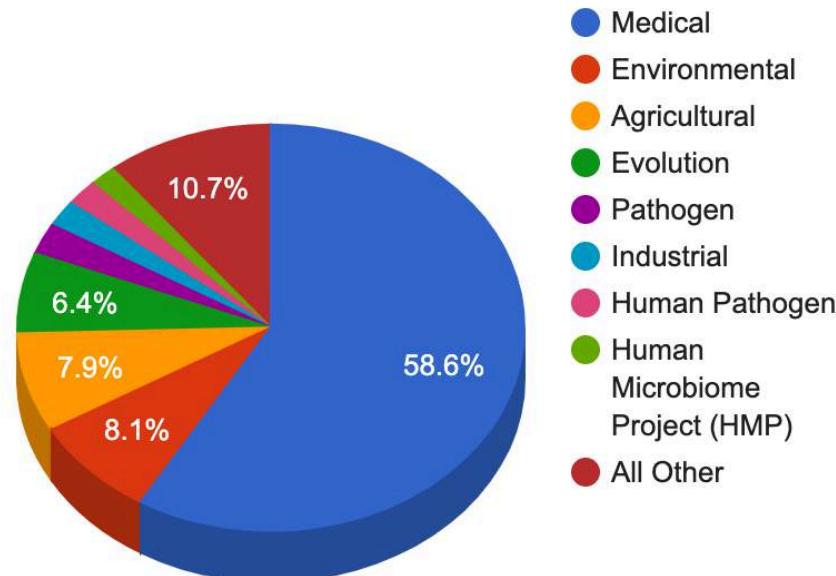
## Projects by Domain



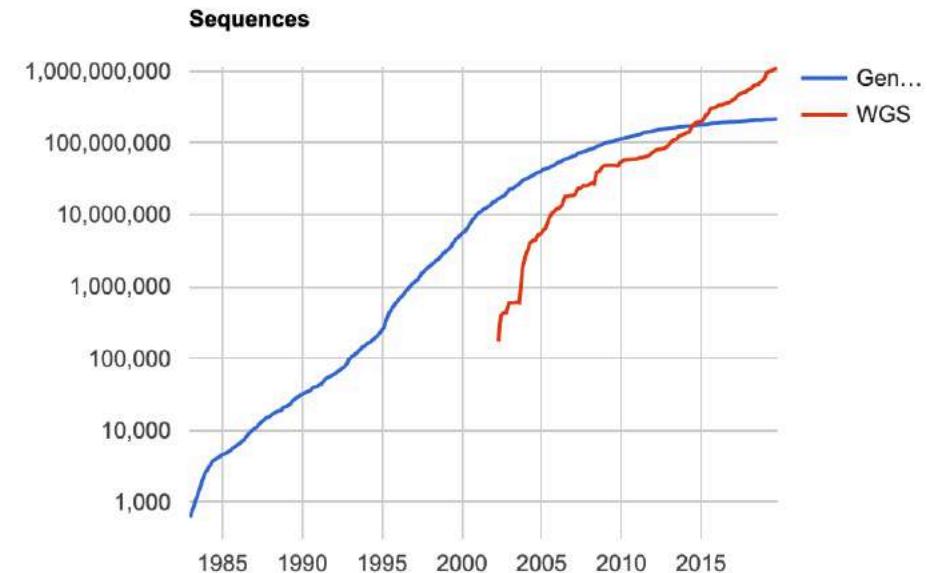
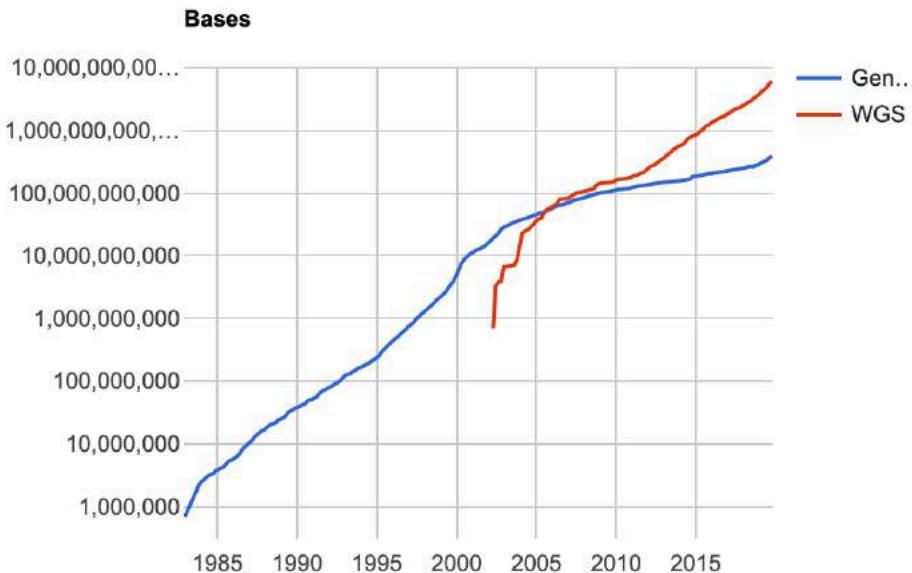
## NCBI Import Tracker



## Project Relevance of Bacterial Projects



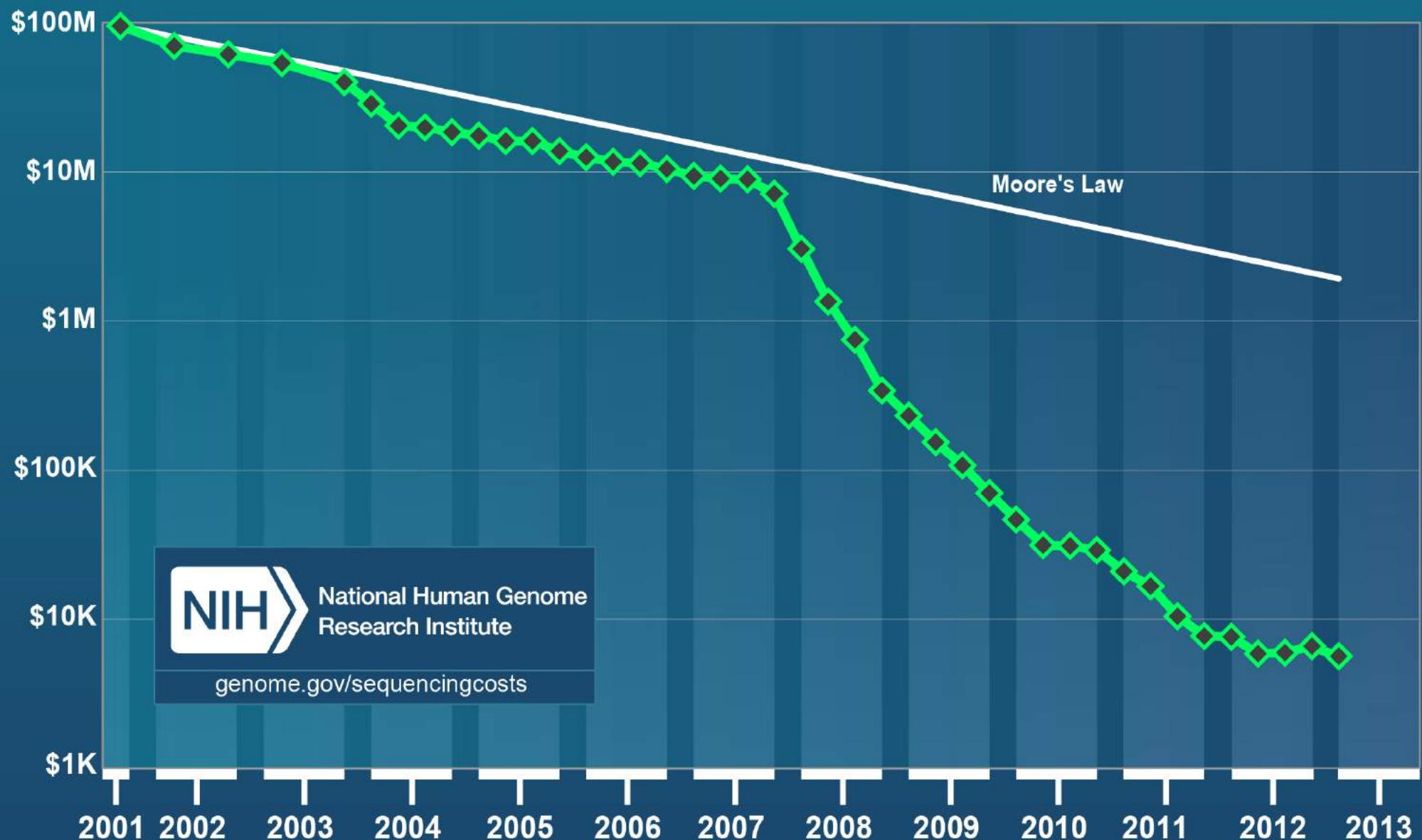
# Growth of the GenBank dataset



## GENBANK AND WGS STATISTICS

		GenBank		WGS	
Release	Date	Bases	Sequences	Bases	Sequences
3	Dec 1982	680338	606		
234	Oct 2019	386197018538	216763706	5985250251028	1097629174

## *Cost per Genome*

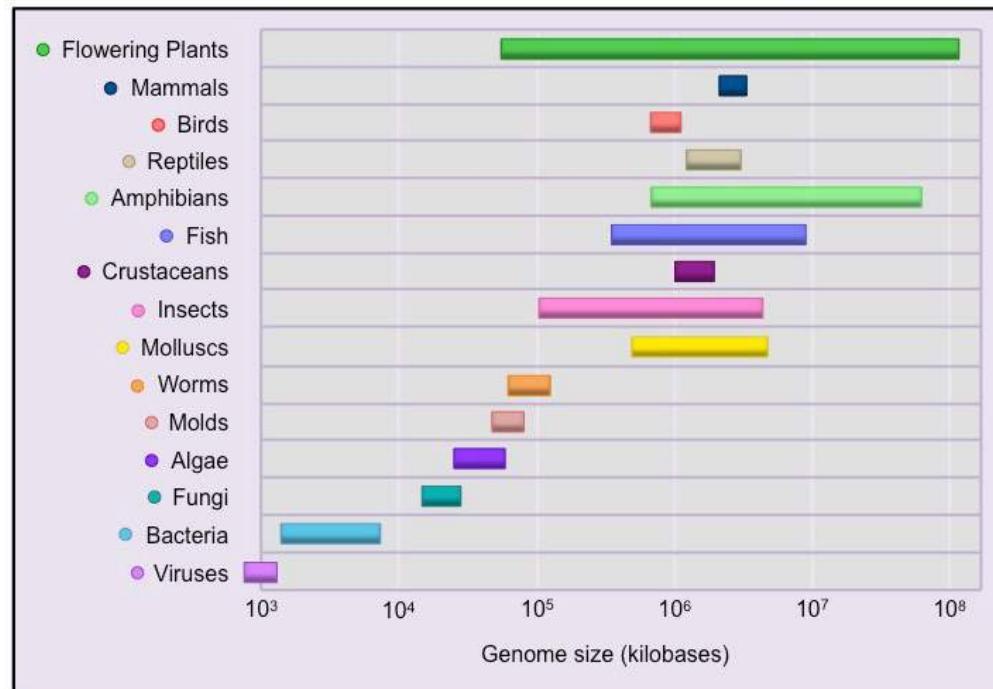


National Human Genome  
Research Institute

[genome.gov/sequencingcosts](http://genome.gov/sequencingcosts)

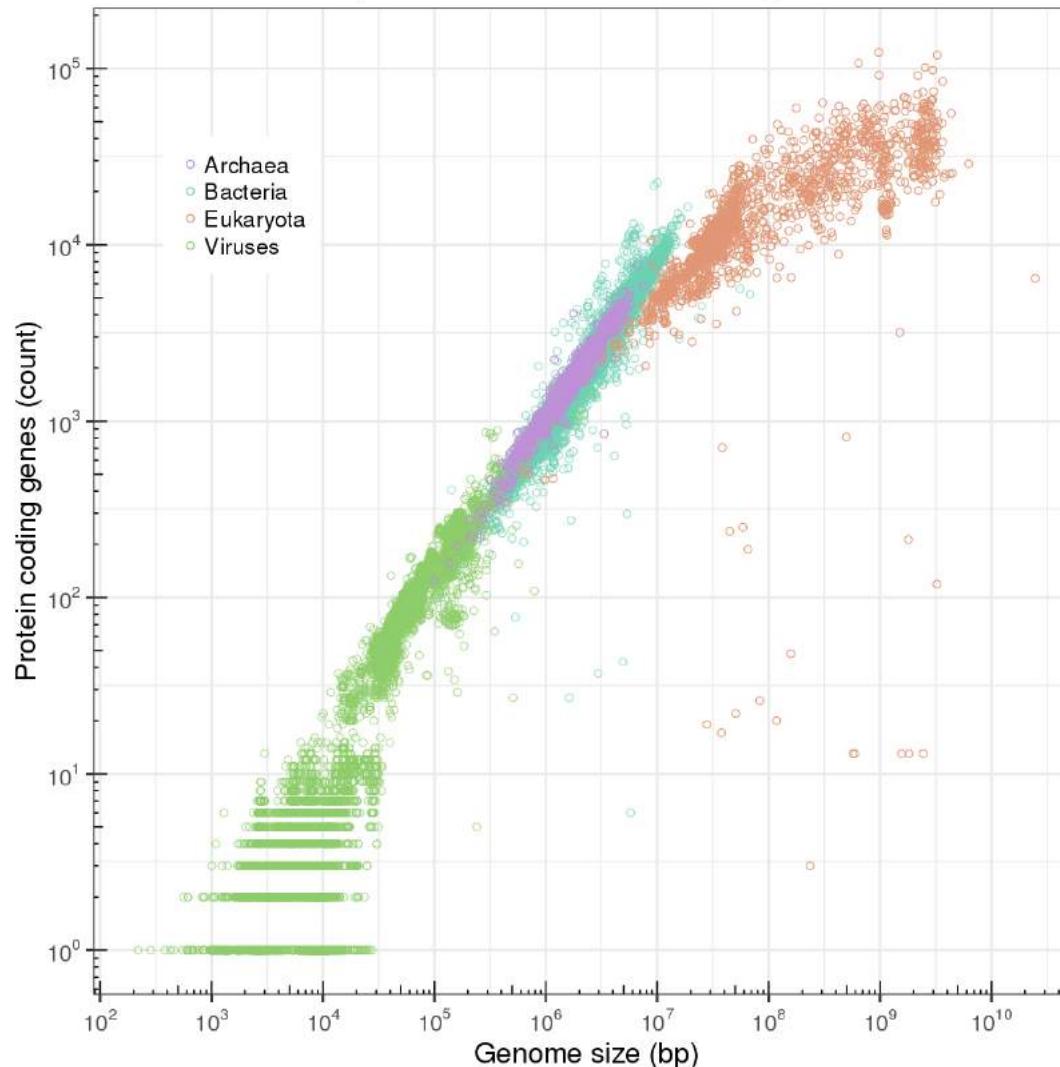
# Genome size X Organism complexity

Species	<i>T2 phage</i>	<i>Escherichia coli</i>	<i>Drosophila melanogaster</i>	<i>Homo sapiens</i>	<i>Paris japonica</i>
Genome Size	170,000 bp	4.6 million bp	130 million bp	3.2 billion bp	150 billion bp
Common Name	 Virus	 Bacteria	 Fruit fly	 Human	 Canopy Plant

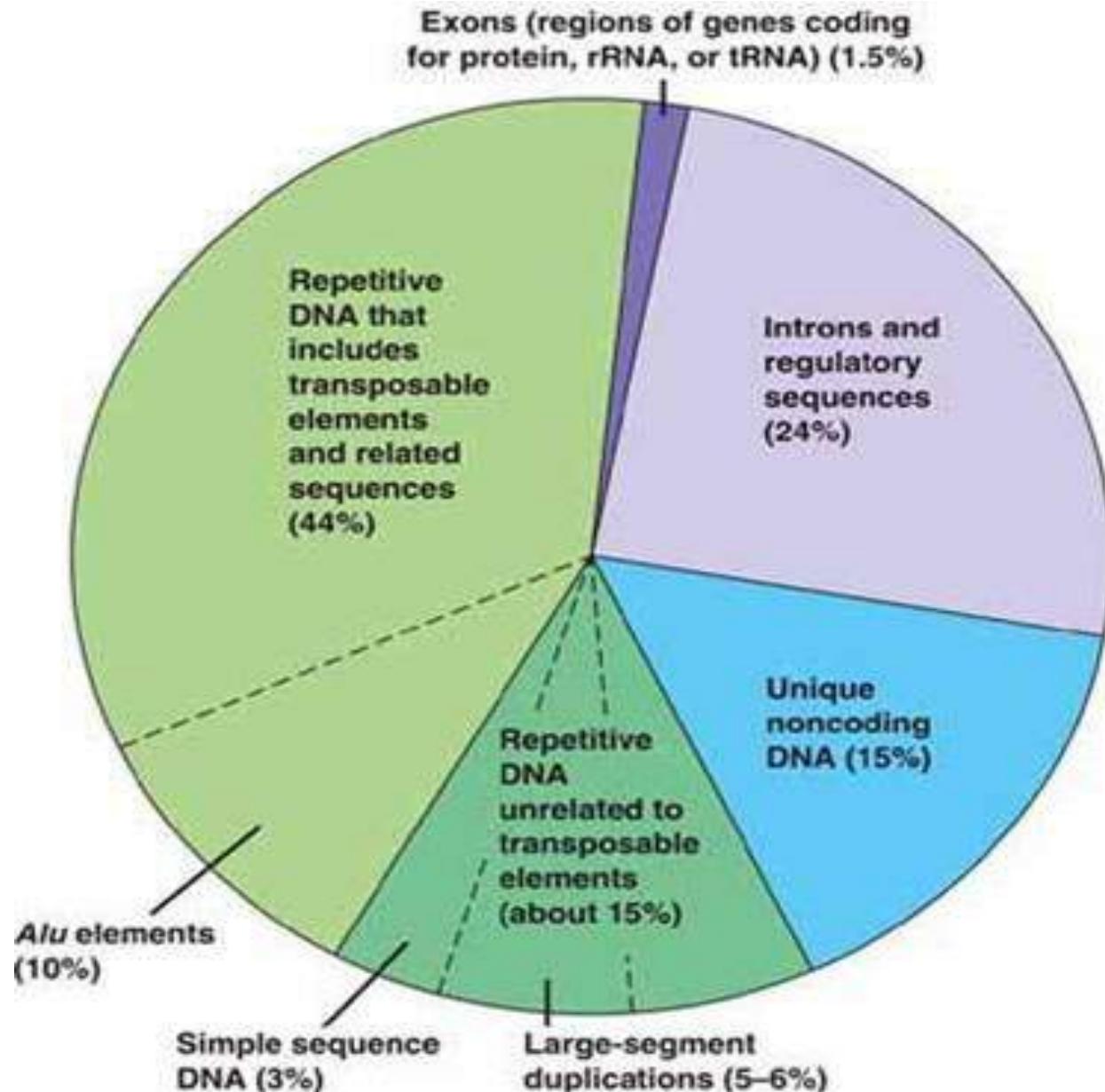


# Genome size X Coding genes

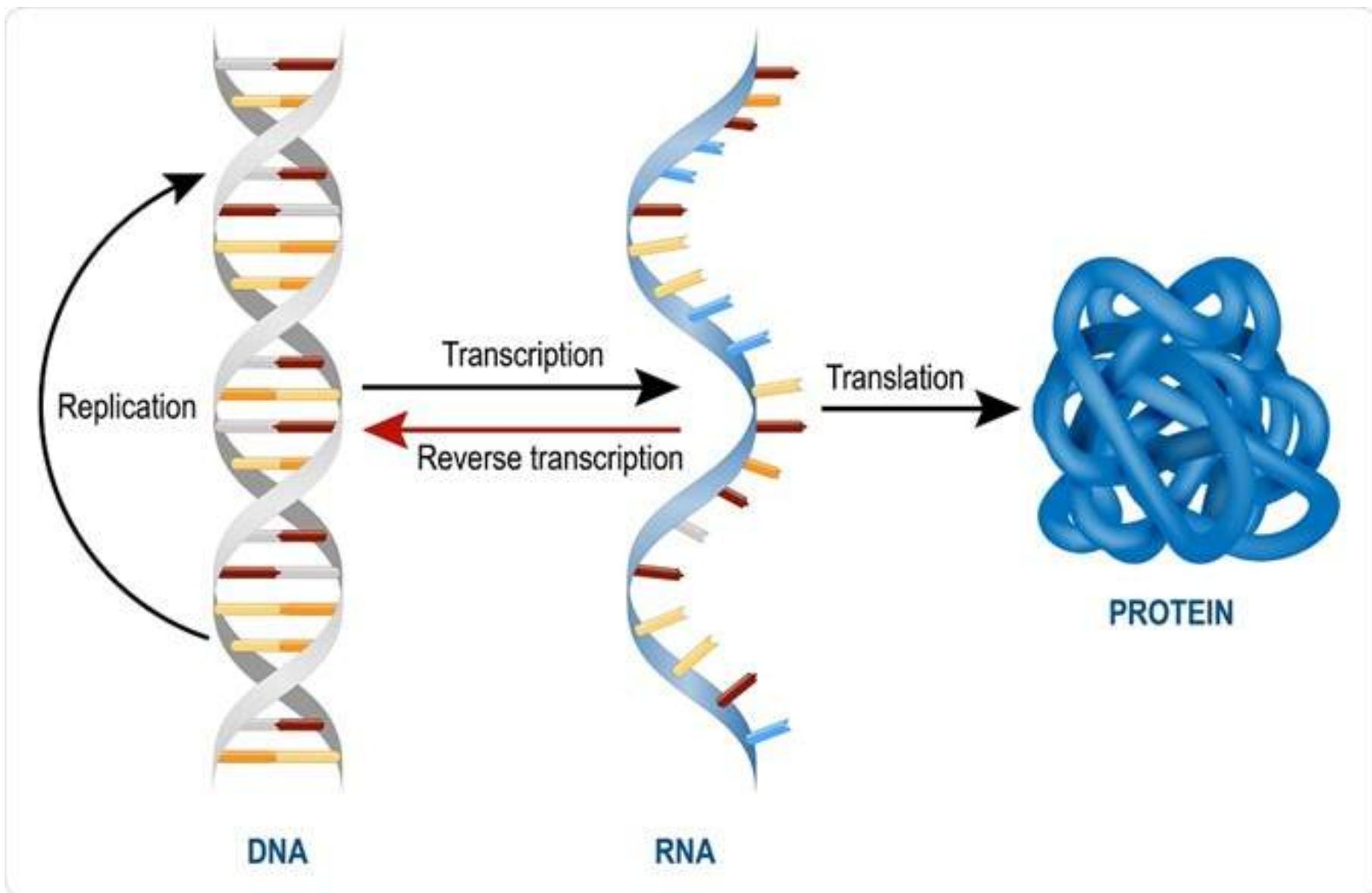
Genome size vs. protein count across NCBI genomes



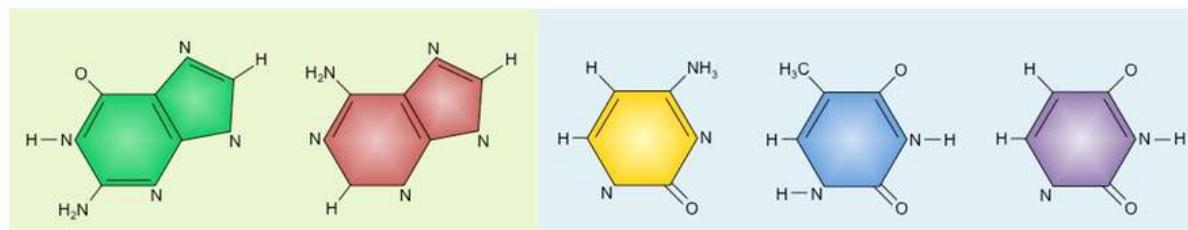
# Genome contents



# The biological dogma



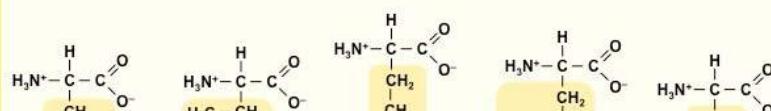
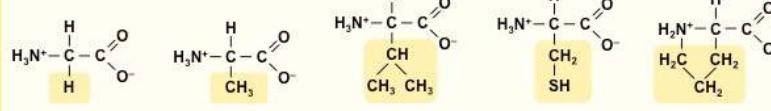
## DNA / RNA



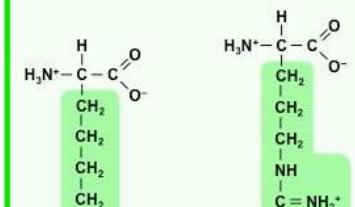
### Purines

### Pyrimidines

### NON-POLAR



### + CHARGE



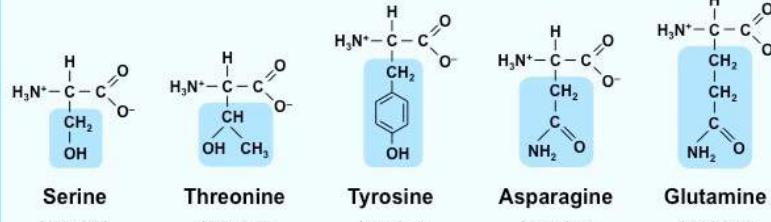
### Histidine

(His / H)

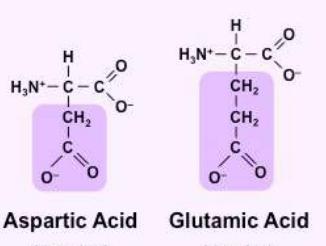
## PROTEIN

DEGENERATE

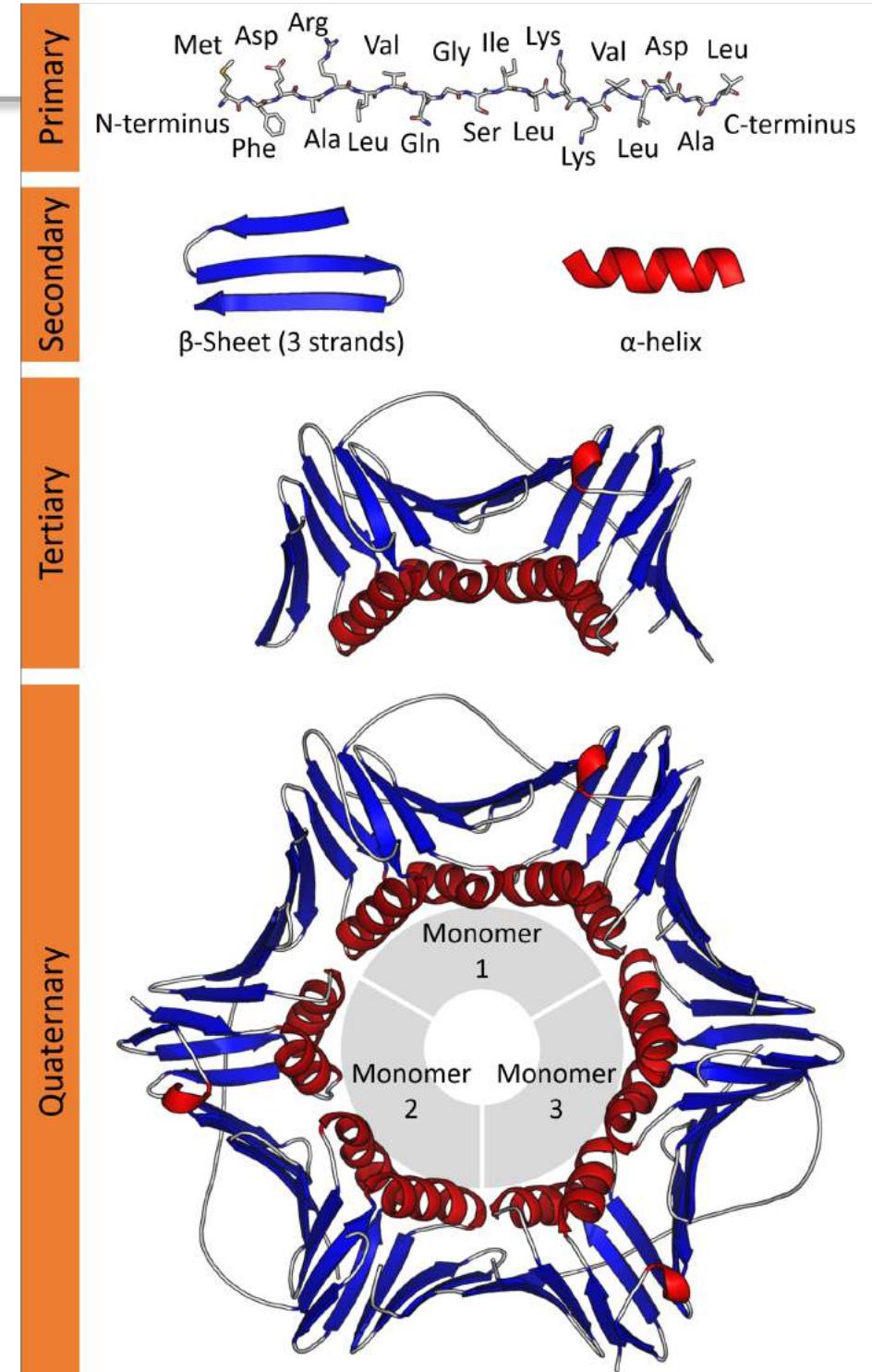
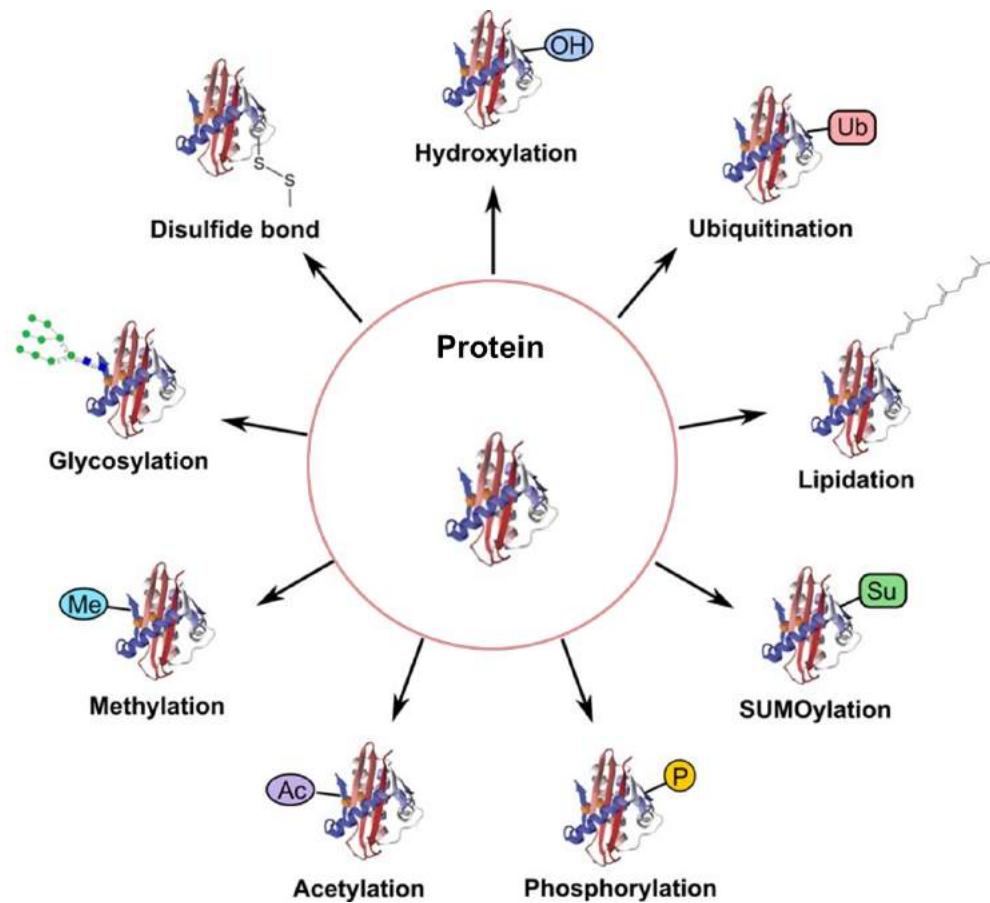
### POLAR

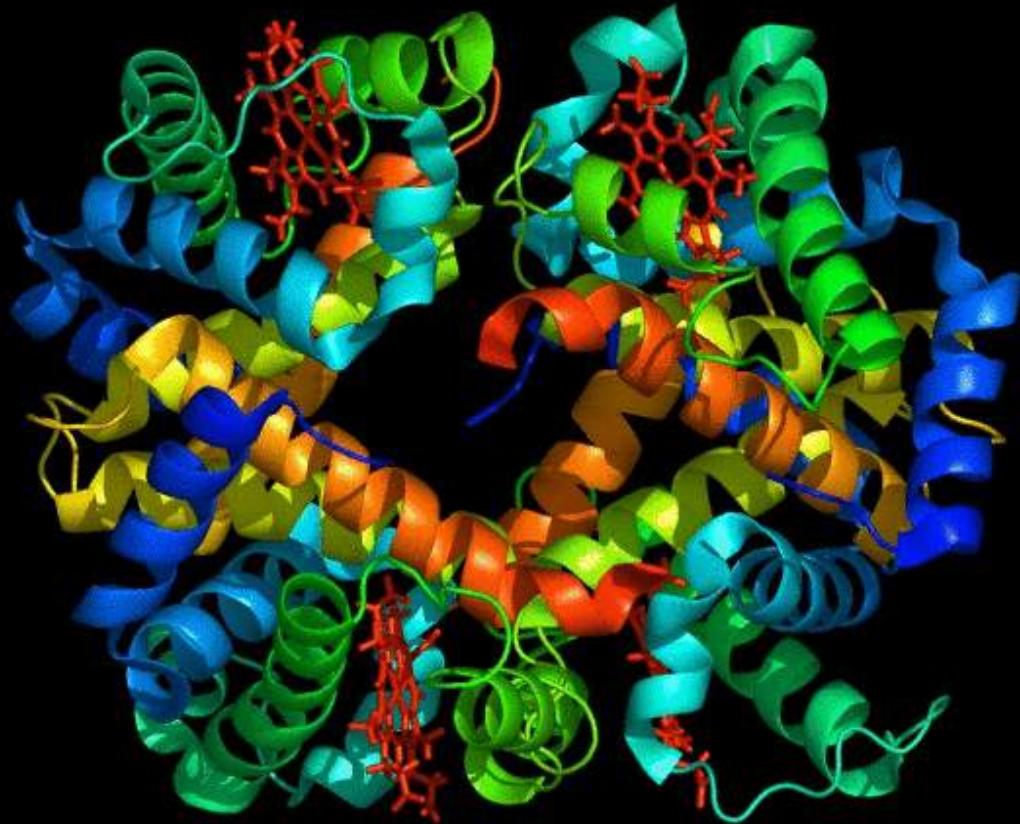


### - CHARGE



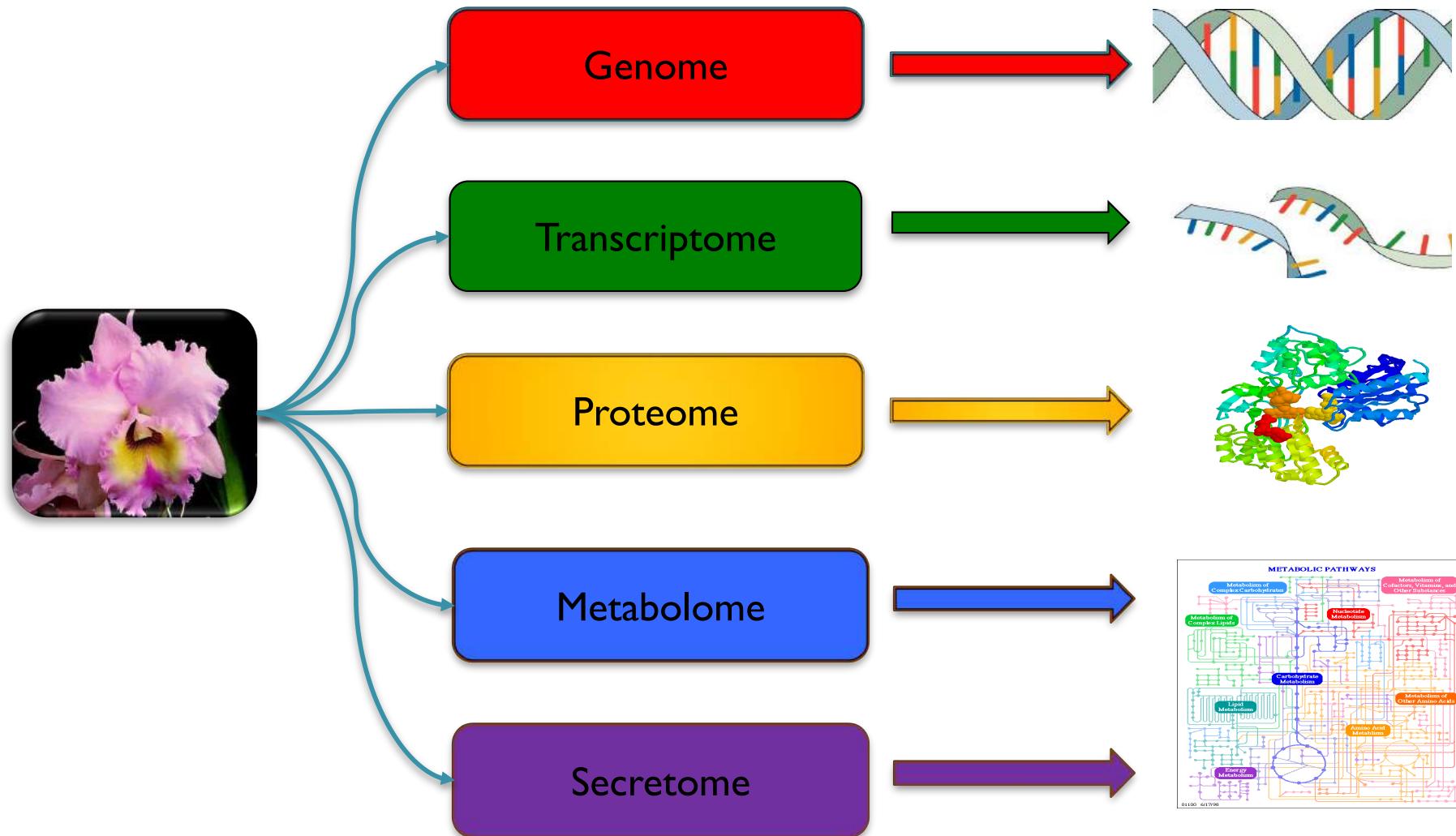
# Post-translational modifications & Structure



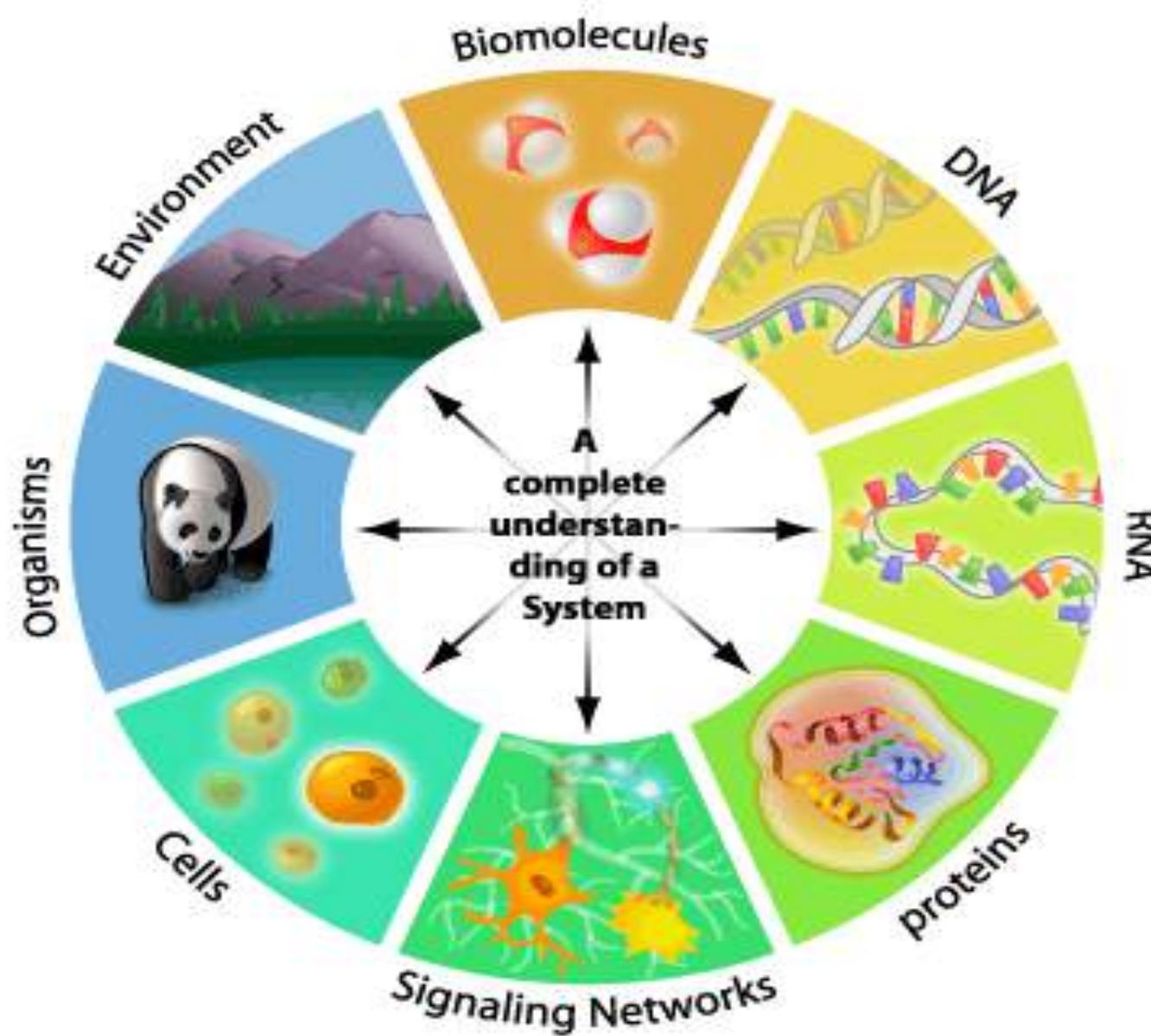


## Hemoglobin Flexible quaternary structure

# Each cell within an organism have a unique:



## Lamarck's & Darwin's & Wallace's concerns...



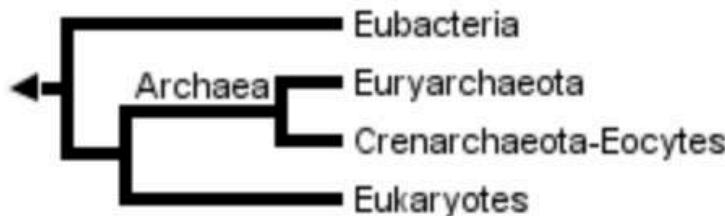
- Type of data:
  - Eukaryotic or Prokaryotic
  - DNA, RNA, Protein, Structure, ...
  - Genes, Proteins, Families
  - Functional domains, Antigens, Signatures
  - Intra- and Inter-specific variability
    - Concept of species ~ Phylogeny
      - Similarity:
        - Measurable !
        - Phylogeny
        - Distance, molecular clock, barcoding,...
      - Homology:
        - Non-Measurable ! Imply on ancestrality
        - Orthology and Paralogy
      - Search for patterns:
        - *In silico* prediction
        - Benchwork validation

## TREE OF LIFE web project

## Discussion of Phylogenetic Relationships

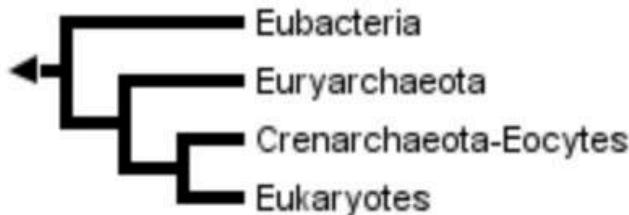
Two alternative views on the relationship of the major lineages (omitting viruses) are shown below

The "archaea tree":



**“DNA-based”**

The "eocyte tree":



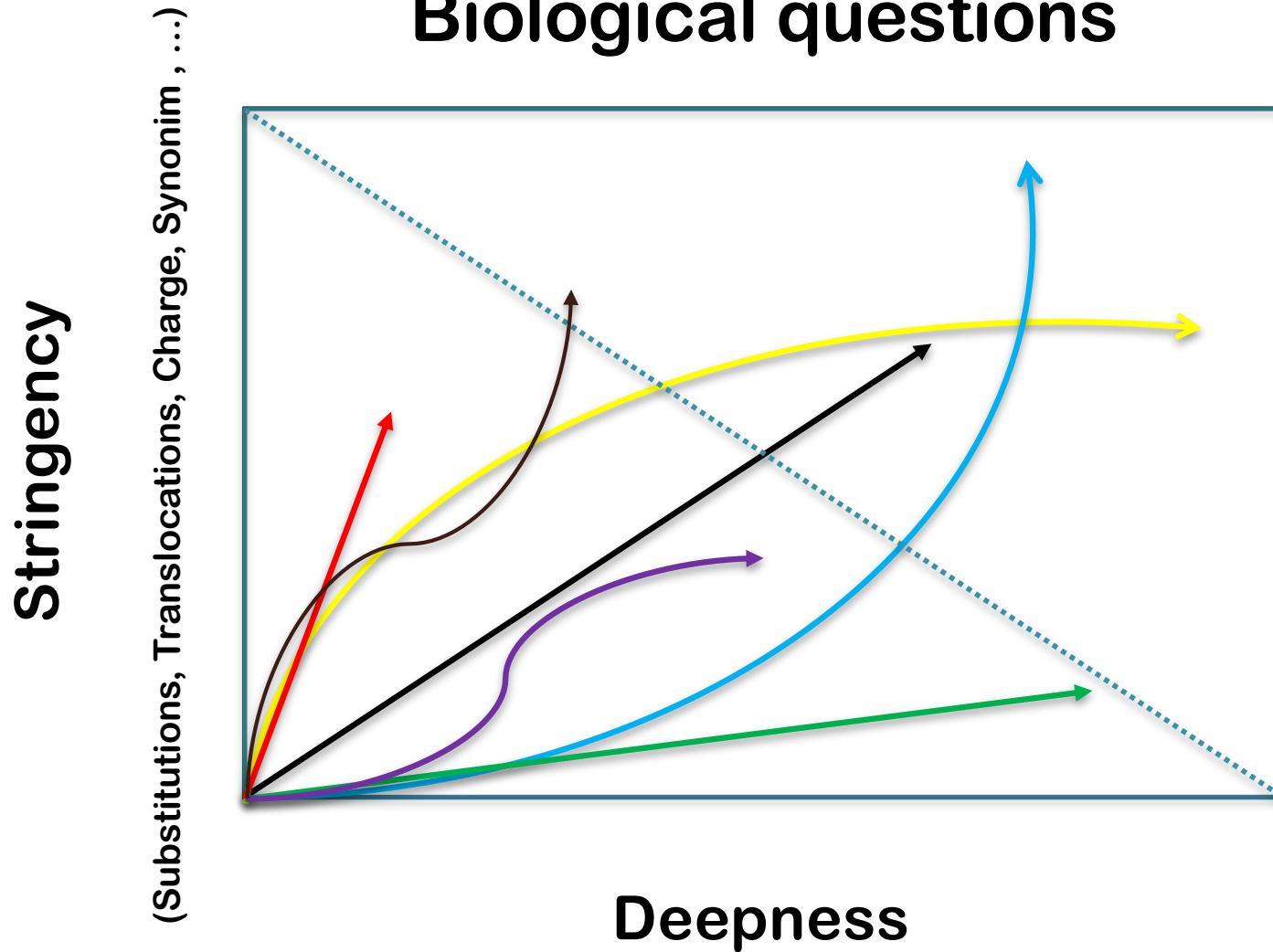
**“RNA-based”**

# (Few) Biological questions

What to compare? / What to look for?

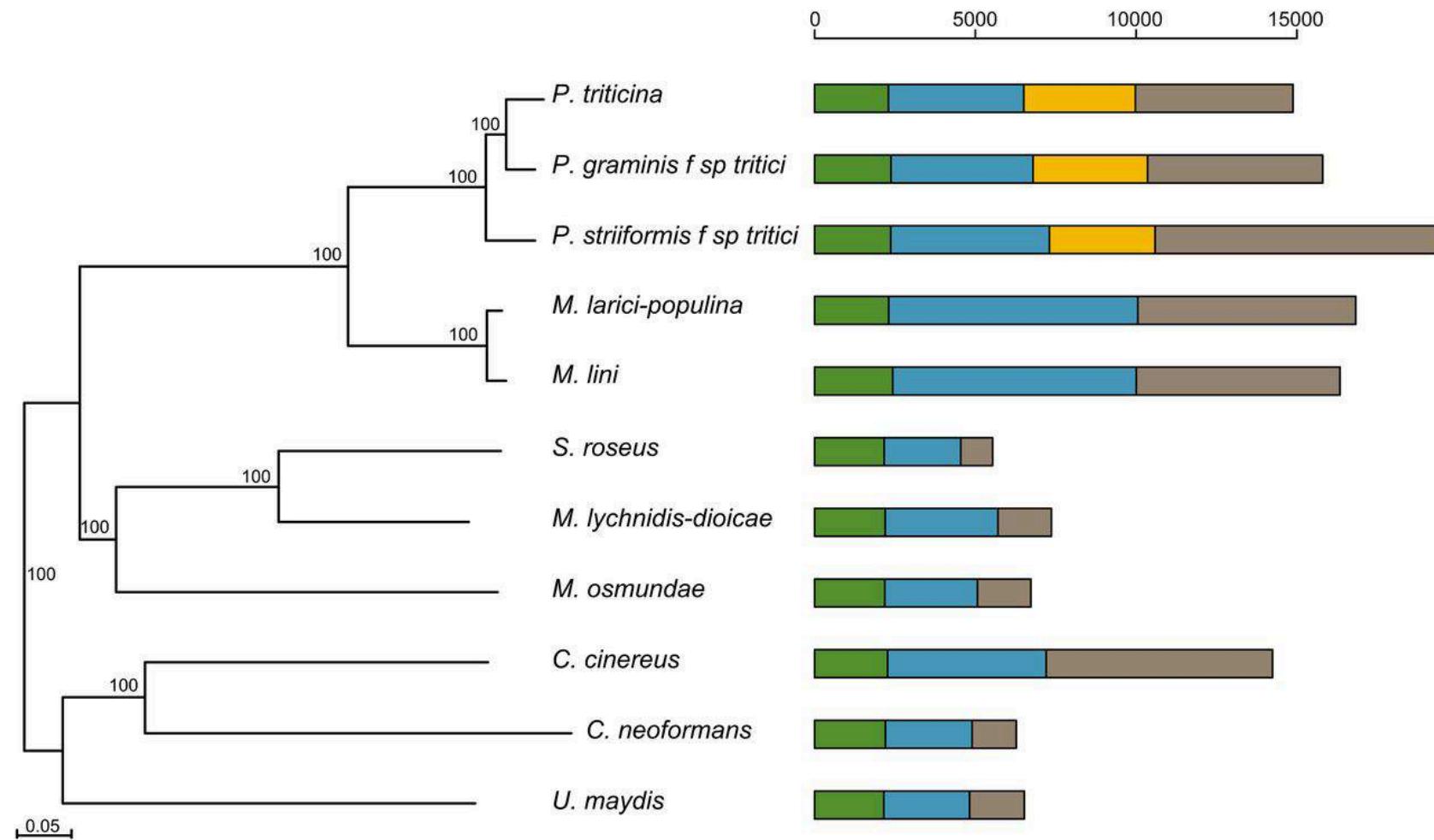
- Sequence (Size/Composition)
- Presence / Absence of a gene:
  - Introns, Pseudogenes, Duplications, Transposons, ITS, ...
- Syntheny
- Transcription / Expression levels
- Post-Transcriptional / Post-Translational modifications
- Protein structure
- Function
- Activity
- Life cycle stages:
  - Stage-specific expression
- Homeostatic *Versus* Pathological stages:
  - Mutations
  - Drug resistance

# Biological questions

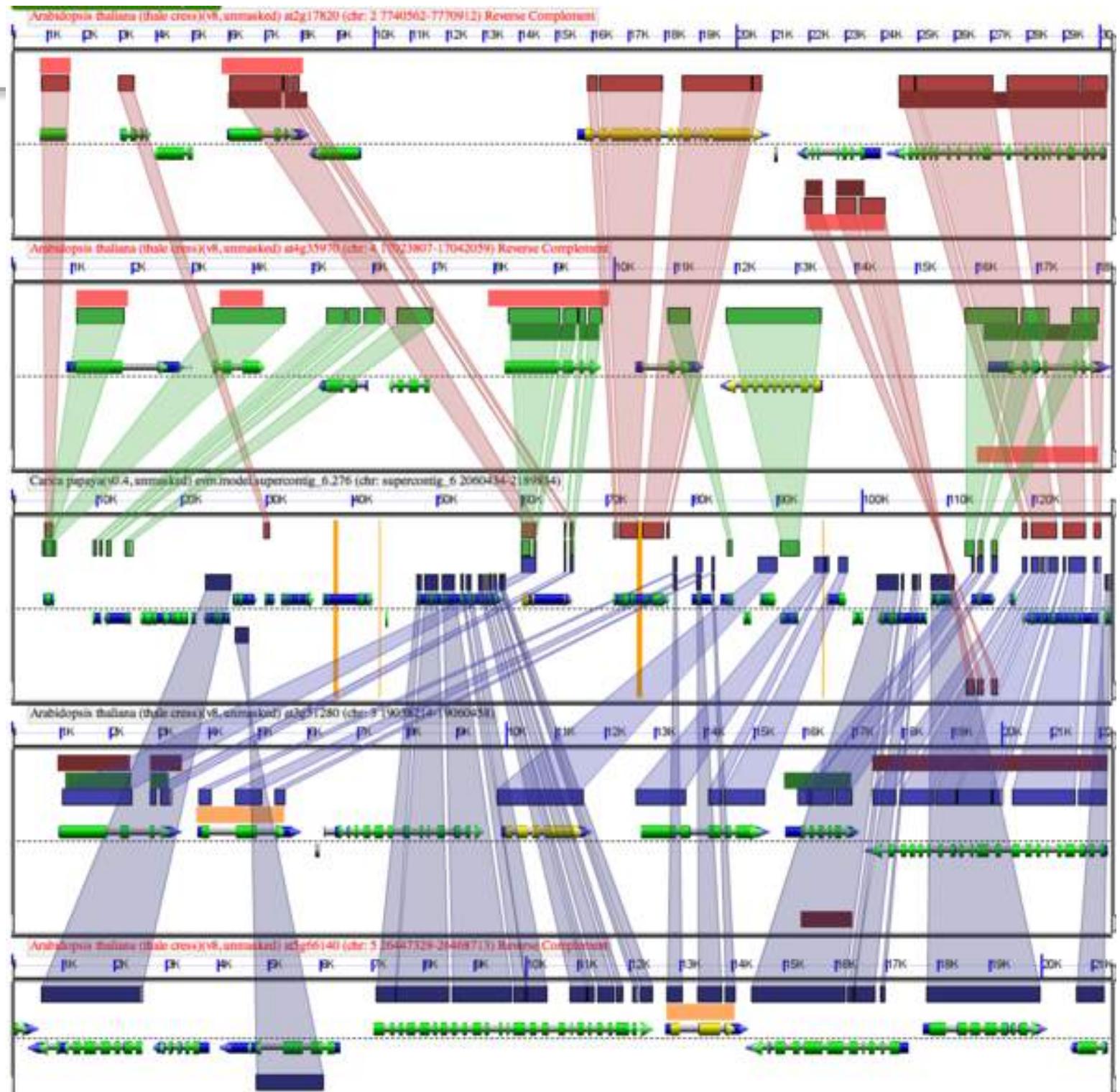


(Data type, number of markers and specimens, broadness of taxa, rooting, algorithm, ...)

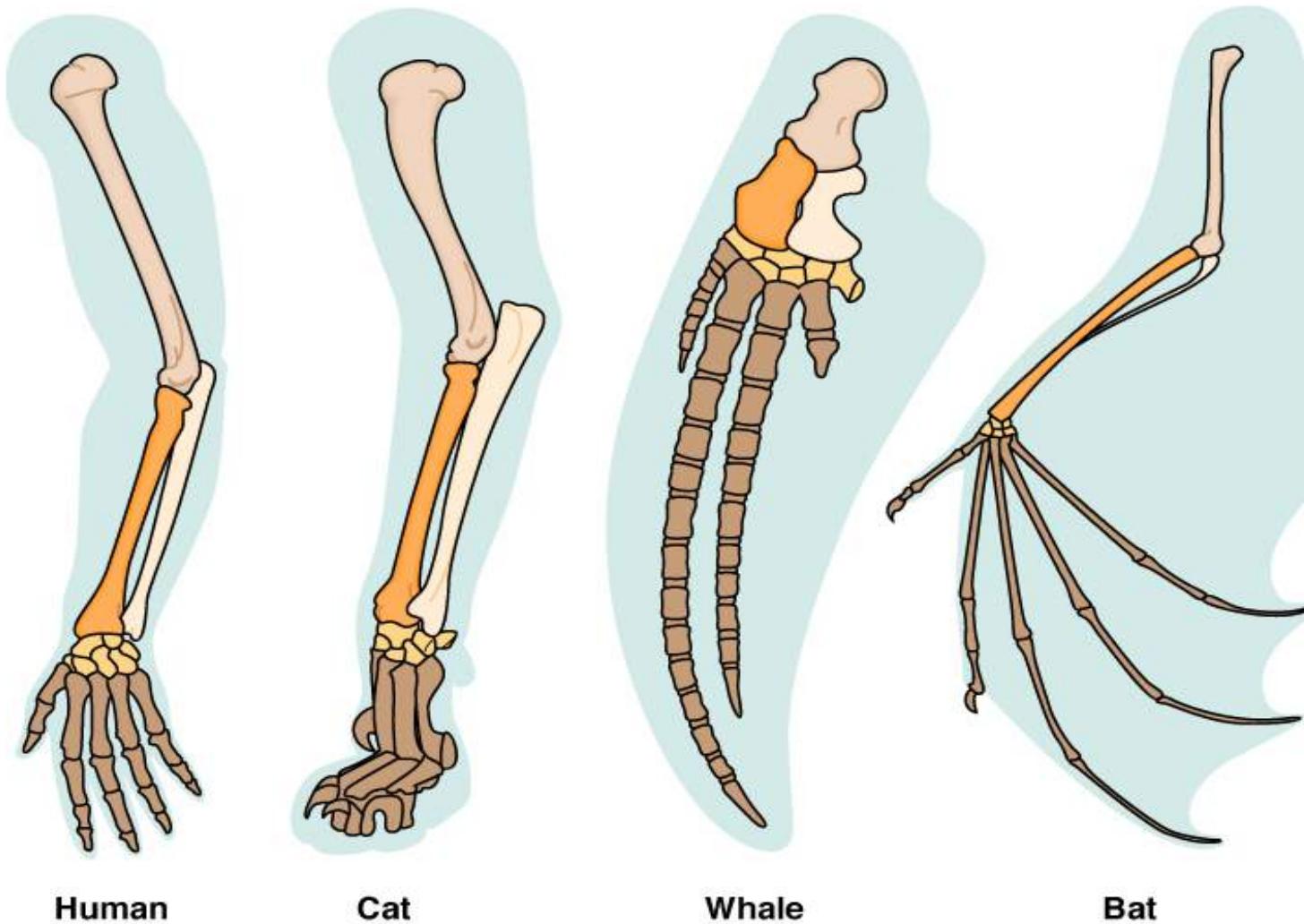
# Comparative pattern recognition



*A.thaliana*



# Similarity & Homology



©1999 Addison Wesley Longman, Inc.

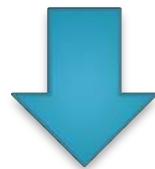
Common evolutive origin? Similar/shared characteristics?

Same / Distinct function ?

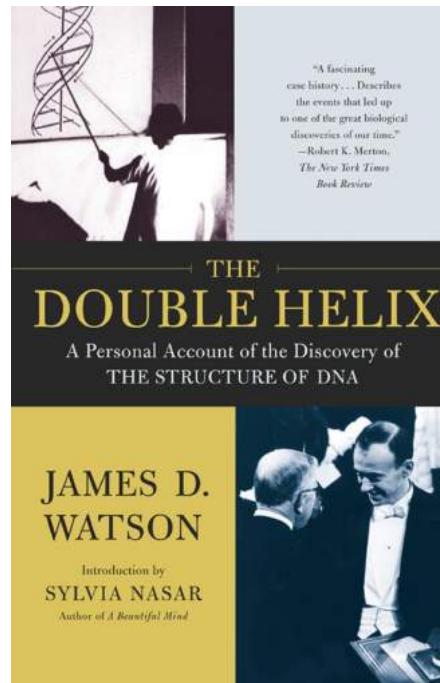
.....ASVJSIRWQEKVNLSDMAFKOAPDSFLASDKFLASVASMRDSVKS'DALSKDF  
LASDFKDSMKLSMCKLAMSDKGFH AISVNJD SVANSNINZFDNSGND SIOVNSAN  
DIFNOINVDSNVSMDAFEASDVAMVDSNGFIOANEINWIOENVIA NDIOSNVNSODV  
NSODNVUNPNRUNVUIDNVIUSNDUINAUI NSDUNVRUNGUNERUVNDFPSUNVA  
SDUNRUINVUNWINRU VANUIBVUIBRIUBVUIEBNRTUIBVUIBPUBAUIAWPIUEB  
RUABVIYBAIPBWQIPWIDVBWBNVODVINWIRHUVWNVUNUNRUNOUNVDUNU  
OQNEDNVEURNVUNE OV NQNUVDNQUBUOE BRUOVBERBVDOUBWEUOVBOU  
NDVODQVOSIVNDIONERO GNERUOVNROUVBWVUQBODUBVDOJB NVUGBER  
UBVN WJOEWNFINWV IFHYNEWNERO QHTHE DOUBLEPBWQIPWHELIXFDJAHD  
SHFHASDFHLASDFISABOOKASKDFKASJDFDSFKSDFFROMJSDHFJKADJFKA  
SDHFLAHSDFJKHAJSDJAMESDWATSONIPBIUWBEIUBQPIUBWUEPIBCVY WV  
BDYIYBSBCVABSDPIUF BQWERUVBRQBWEJBWEJT BQWHTVYSGSVNASRNTI  
OWHFAJSBFSDFJAWBDFGUWRBQUDPIUVSJ KDBVUIPWE GUQFADSUGHOSD  
HGJBQEPWIGUDUIABVSUAIGFDABWEFUIVBWUDFBGAUGE FUDNFGANDSVE  
UDJAJSBDVIUAUSBURBGUBWQPQJDCJPWICIRRYBQVIBWIUBVUIEBNRTUIB  
VUIBPUBAUIAWPIUEBRUABVIYBAIPBWQIPWIDVBWBNVODVINWIRHUVWNVU  
NUNRUNOUNVDUNUOQNEDNVEURNVUNE OV NQNUVDNQUBUOE BRUOVBER  
BVDOUBWEUOVBOUNDVODQVOSIVNDIONERO GNERUOVNROUVBWVUQBO  
DUBVDOJB NVUGBERUBVN WJOEWNFINWV IFHYNEWNERO QHODVSBDPVAS  
DVARNOFVSUODBVAUSBUVIPBIUWBEIUBQPIUBWUEPIBCVY WVBDYIVWDYB  
SBCVABSDPIUFIUBVUIEBNRTUIBVUIBPUBAUIAWPIUEBRUABVIYBAIPBWQIP  
WIDVBWBNVODVINWIRHUVWNVUNUNRUNOUNVDUNUOQNEDNVEURNVUN  
EOVNQNUVDNQUBUOE BRUOVBERBVDOUBWEUOVBOUNDVODQVOSIVNDIO  
NEROGNERUOVNROUVBWVUQBODUBVDOJB NVUGBERUBVN WJOEWNFINW  
VIFHYNEWNERO QHODVSBDPVASDVARNOFVSUODBVAUSBUVIPBIUWBEIUB  
QPIUBWUEPIBCVY WV.....

.....ASVJSIRWQEKVNLSDMAFKOAPDSFLASDKFLASVASMRDSVKS'DALSKDF  
LASDFKDSMKLSMCKLAMSDKGFH AISVNJD SVANSNINZFN DSGNDSIOVNSAN  
DIFNOINVDSNVSMDAFEASDVAMVDSNGFIOANEINWIOENVIA NDIOSNVNSODV  
NSODNVUNPNRUNVUIDNVIUSNDUINAUI NSDUNVRUNGUNERUVNDFPSUNVA  
SDUNRUINVUNWINRU VANUIBVUIBRIUBVUIEBNRTUIBVUIBPUBAUIAWPIUEB  
RUABVIYBAIPBWQIPWIDVBWBNVODVINWIRHUVWNVUNUNRUNOUNVDUNU  
OQNEDNVEURNVUNE OV NQNUVDNQUBUOE BRUOVBERBVDOUBWEUOVBOU  
NDVODQVOSIVNDIONEROGNERUOVNROUVBWVUQBODUBVDOJB NVUGBER  
UBVN WJOEWNFINWV IFHYNEWNEROQH **THE DOUBLE** PBWQIPW **HELIX** FDJAHD  
SHFHASDFHLASDF **IS A BOOK** ASKDFKASJDFDSFKSDF **FROM** JSDHFJKADJFKA  
SDHFLAHSDFJKHAJSD **JAMES DWATSON** IPBIUWBEIUBQPIUBWUEPIBCVY WV  
BDYIYBSBCVABSDPIUF BQWERUVBRQBWEJBWEJT BQWHTVYSGSVNASRNTI  
OWHFAJSBFSDFJAWBDFGUWRBQUDPIUVSJ KDBVUIPWE GUQFADSUGHOSD  
HGJBQEPWIGUDUIABVSUAIGFDABWEFUIVBWUDFBGAUGE FUDNFGANDSVE  
UDJAJSBDVIUAUSBURBGUBWQPQJDCJPWICIRRYBQVIBWIUBVUIEBNRTUIB  
VUIBPUBAUIAWPIUEBRUABVIYBAIPBWQIPWIDVBWBNVODVINWIRHUVWNVU  
NUNRUNOUNVDUNUOQNEDNVEURNVUNE OV NQNUVDNQUBUOE BRUOVBER  
BVDOUBWEUOVBOUNDVODQVOSIVNDIONEROGNERUOVNROUVBWVUQBO  
DUBVDOJB NVUGBERUBVN WJOEWNFINWV IFHYNEWNEROQHODVSBDPVAS  
DVARNOFVSUODBVAUSBUVIPBIUWBEIUBQPIUBWUEPIBCVY WVBDYIVWDYB  
SBCVABSDPIUF IUBVUIEBNRTUIBVUIBPUBAUIAWPIUEBRUABVIYBAIPBWQIP  
WIDVBWBNVODVINWIRHUVWNVUNUNRUNOUNVDUNUOQNEDNVEURNVUN  
EOVNQNUVDNQUBUOE BRUOVBERBVDOUBWEUOVBOUNDVODQVOSIVNDIO  
NEROGNERUOVNROUVBWVUQBODUBVDOJB NVUGBERUBVN WJOEWNFINW  
VIFHYNEWNEROQHODVSBDPVASDVARNOFVSUODBVAUSBUVIPBIUWBEIUB  
QPIUBWUEPIBCVY WV.....

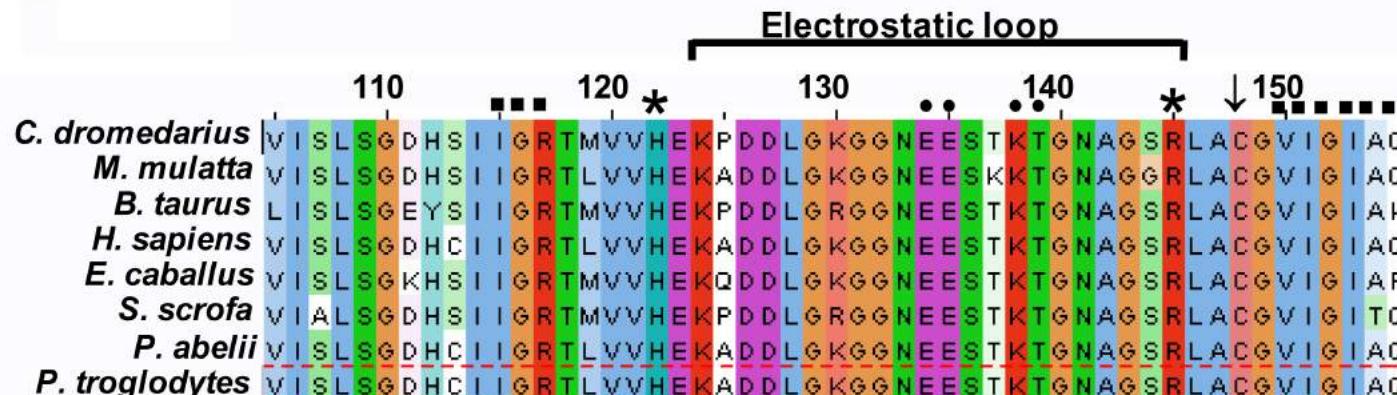
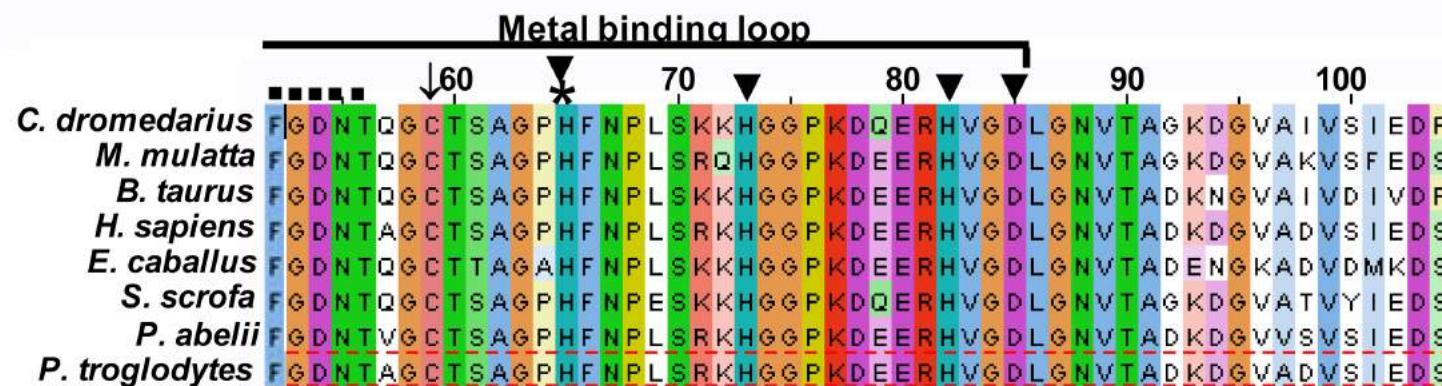
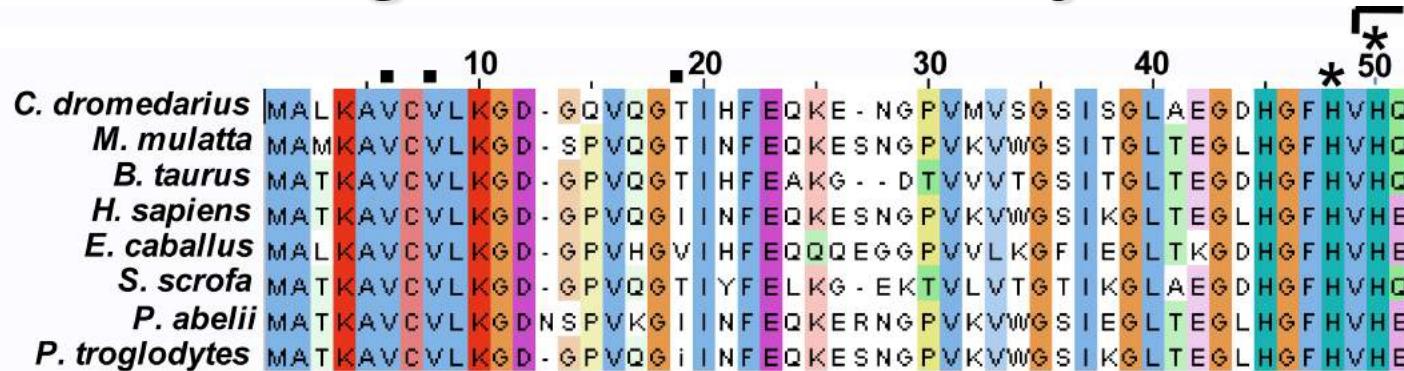
...HYNEWNEROQHTHE DOUBLE HELIXPBWQIPWHELIXFDJAHDHFHASDF  
HLASDFIS A BOOKASKDFKASJDFDSFKSDFFFROMJSDHFJKADJFKA  
SDHFLAHSDFJKHAJSDJAMES DWATSONIPBIUWBEIUBQPI...



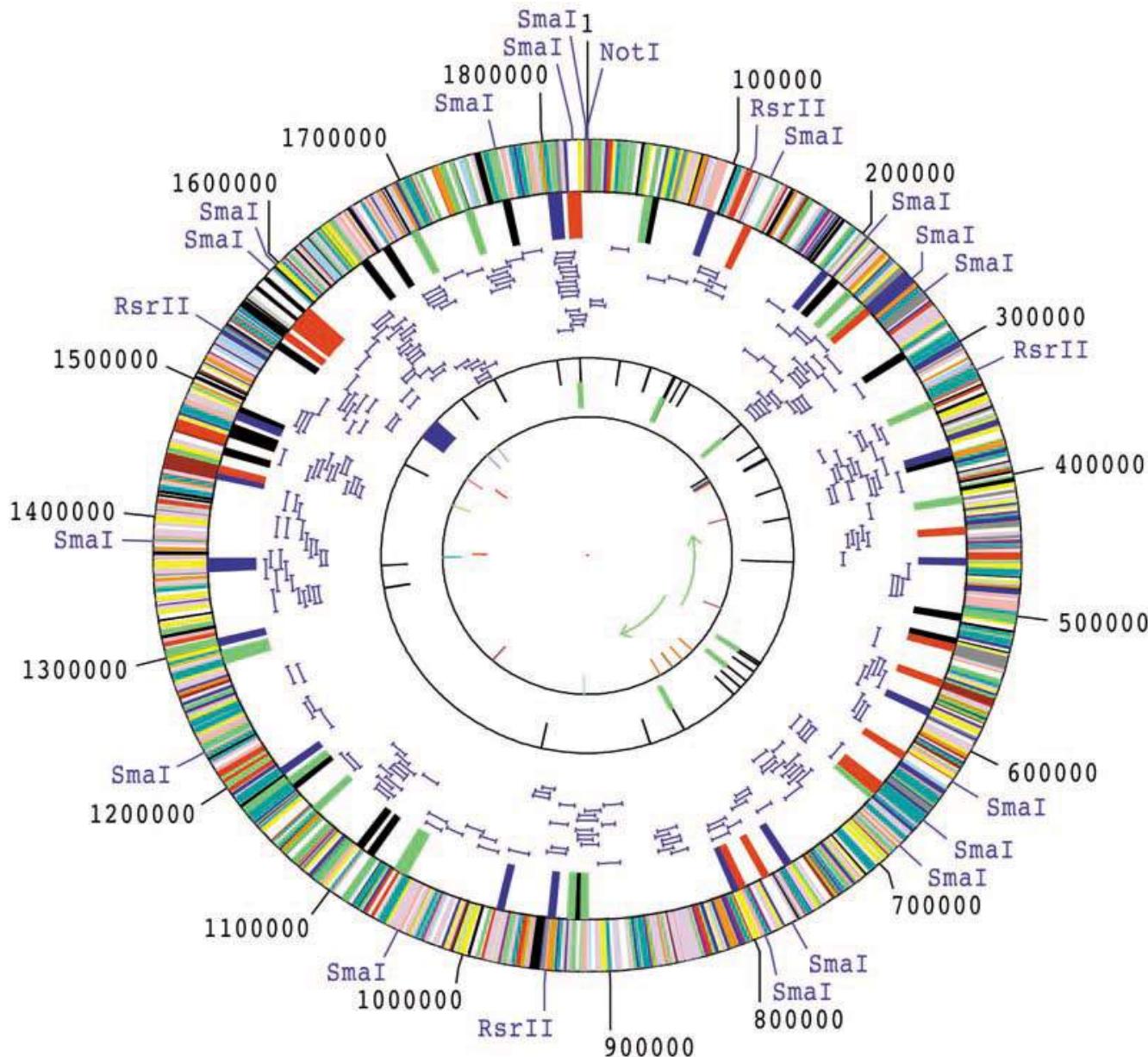
THE DOUBLE HELIX IS A BOOK FROM JAMES WATSON



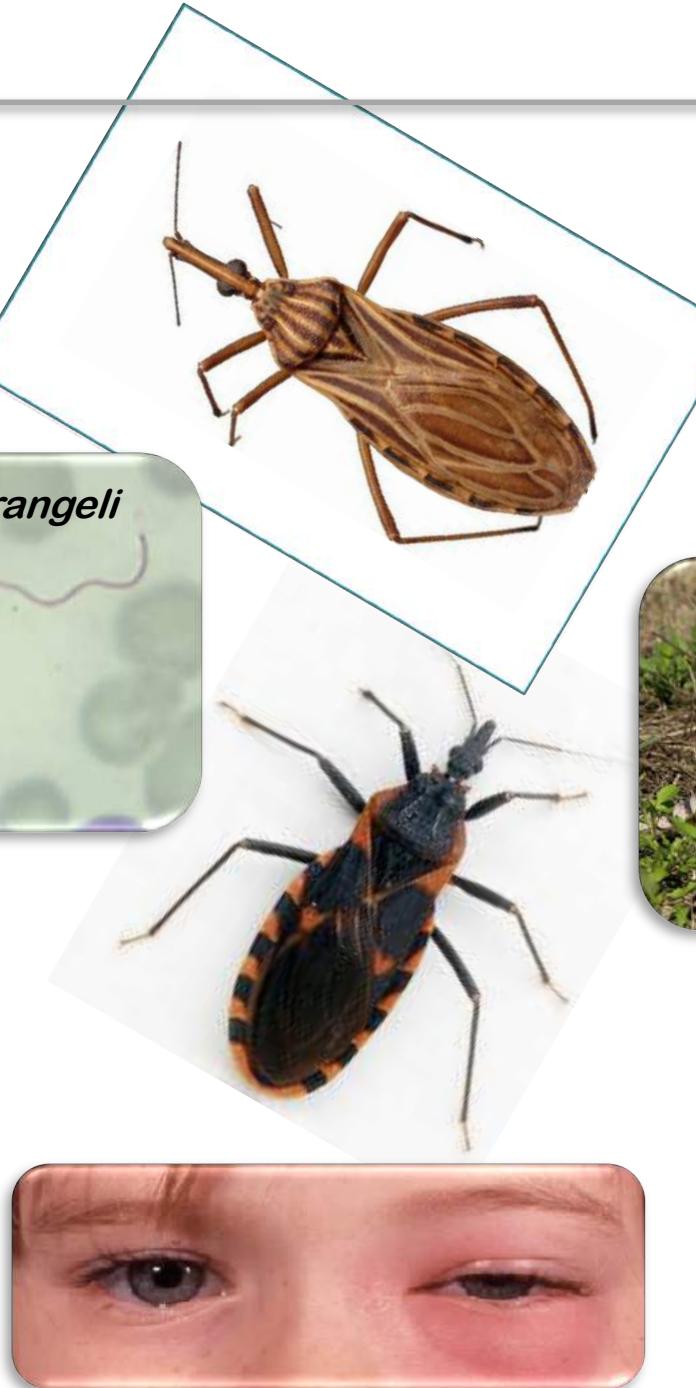
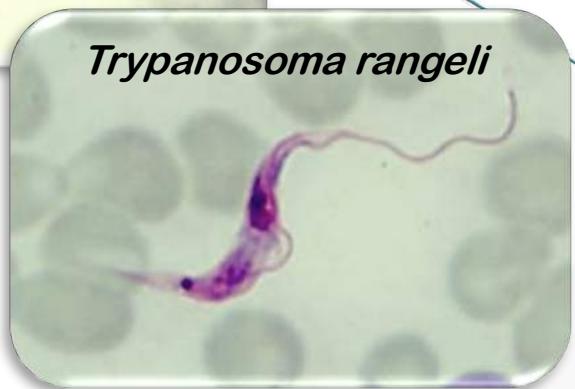
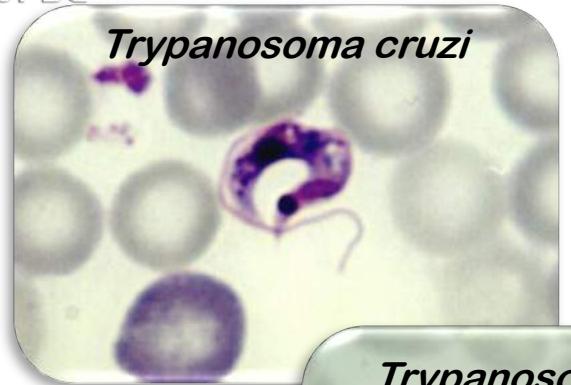
# Patterns might be evolutionary conserved



# Patterns might be evolutionary conserved



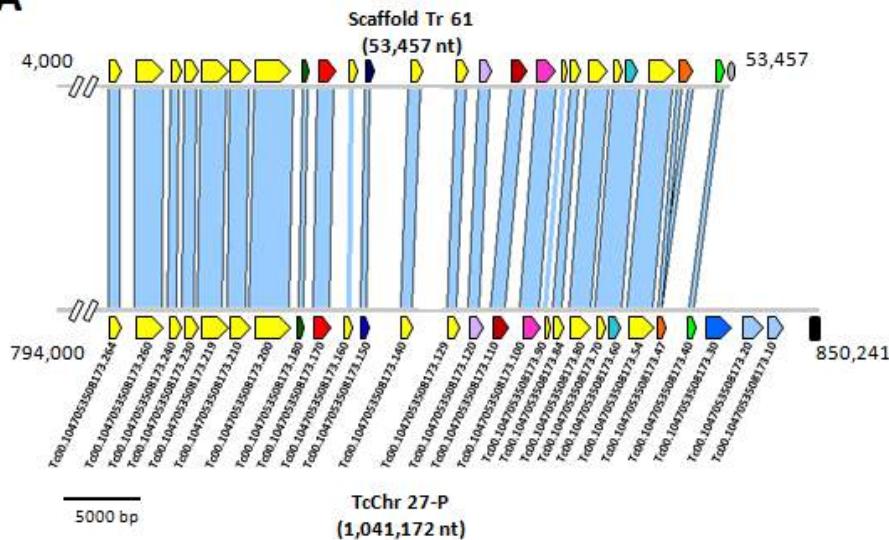
# Study models



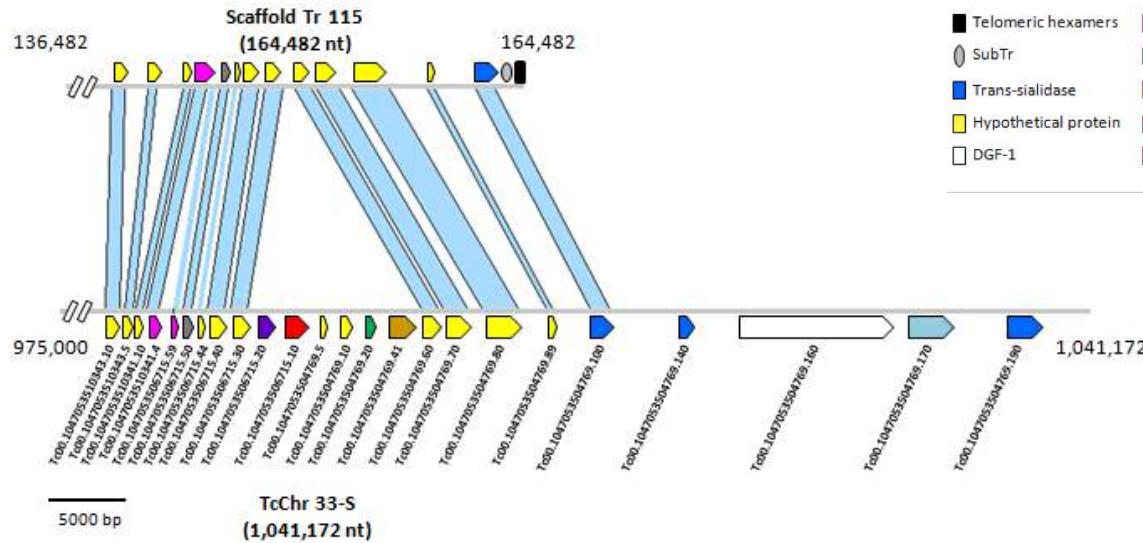
Sharing of geographical distribution, hosts, vectors, genes, proteins, epitopes, ...

Gene Family	<i>T. rangeli</i>	<i>T. cruzi</i>
	SC-58	CL Brener
MASP	50	1465
GP63	444	449
Trans-sialidases	120	1481
Amastins	72	27
DGF	422	569
KMP-11	148	40
Tuzin	34	83
RHS	689	777
Mucin	15	992
MSH6	2	2
MSH2	2	2
GPI8	2	2

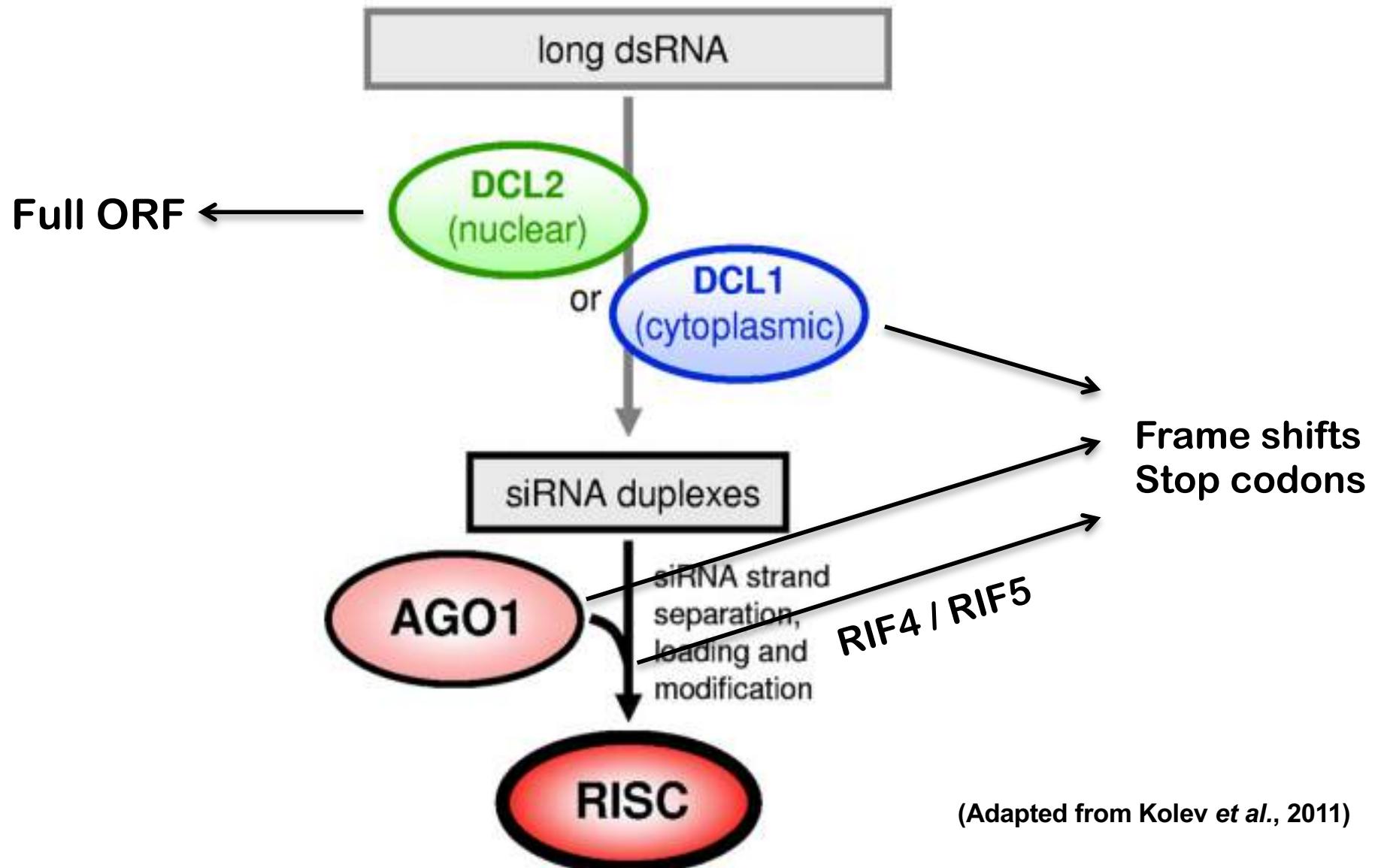
## Multicopy gene families

**A**


**Sinteny  
No DGF  
No TS  
< Transposable elements (*Tc*, *Tb*, *Lm*)**

**B**


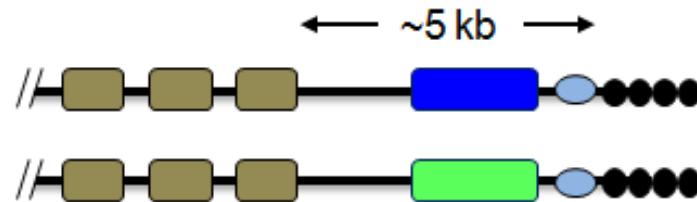
Telomeric hexamers	Monoxygenase	GP63	Nuclear transport factor 2 protein (NFT2)
SubTr	RHS	MASP	SNARE domain-containing protein
Trans-sialidase	Protein kinase	Disulfide isomerase	Mercaptopyruvate sulfurtransferase
Hypothetical protein	Metallopeptidase	Cis-prenyltransferase	Microtubule-associated protein 1A/1B, light chain 3
DGF-1	Cysteine protease	Procyclic formsurface glycoprotein	



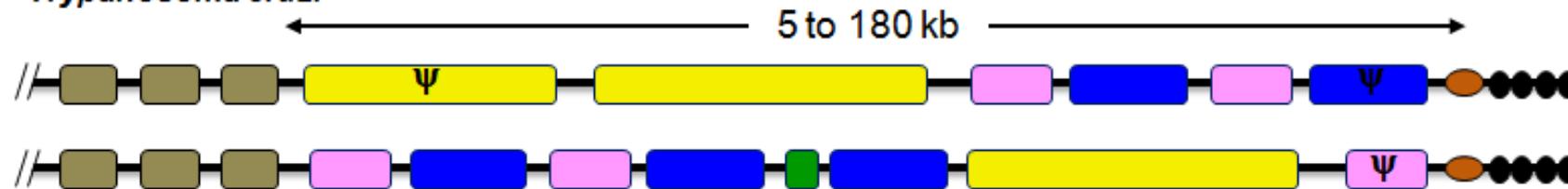
(Adapted from Kolev et al., 2011)

SC-58 and Choachí strains

## Trypanosoma rangeli



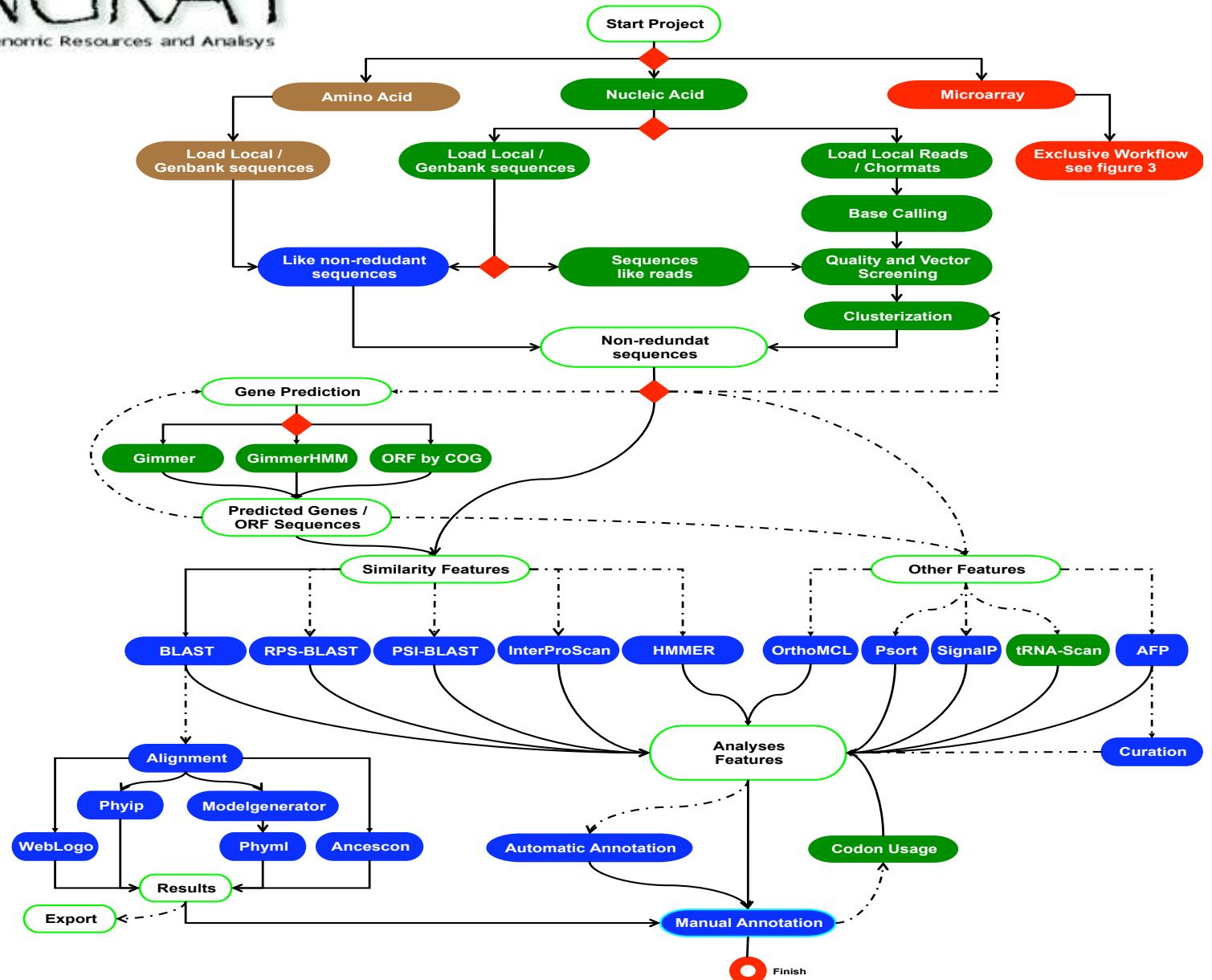
## Trypanosoma cruzi



## Trypanosoma brucei



● Hexamer repeat	● <i>T. cruzi</i> /telomeric junction	■ VSG	■ Interstitial genes	■ Retrotransposons: SIRE, VIPER
▶ 70-bp repeat repeat	● <i>T. rangeli</i> /telomeric junction (SubTr)	■ TS family	□ ESAG	■ Retrotransposons: <i>ingi</i> , RIME
◀ 29-bp repeat repeat	Ψ pseudogene	■ DGF-1	■ RHS	■ Mercaptopyruvate sulfurtransferase

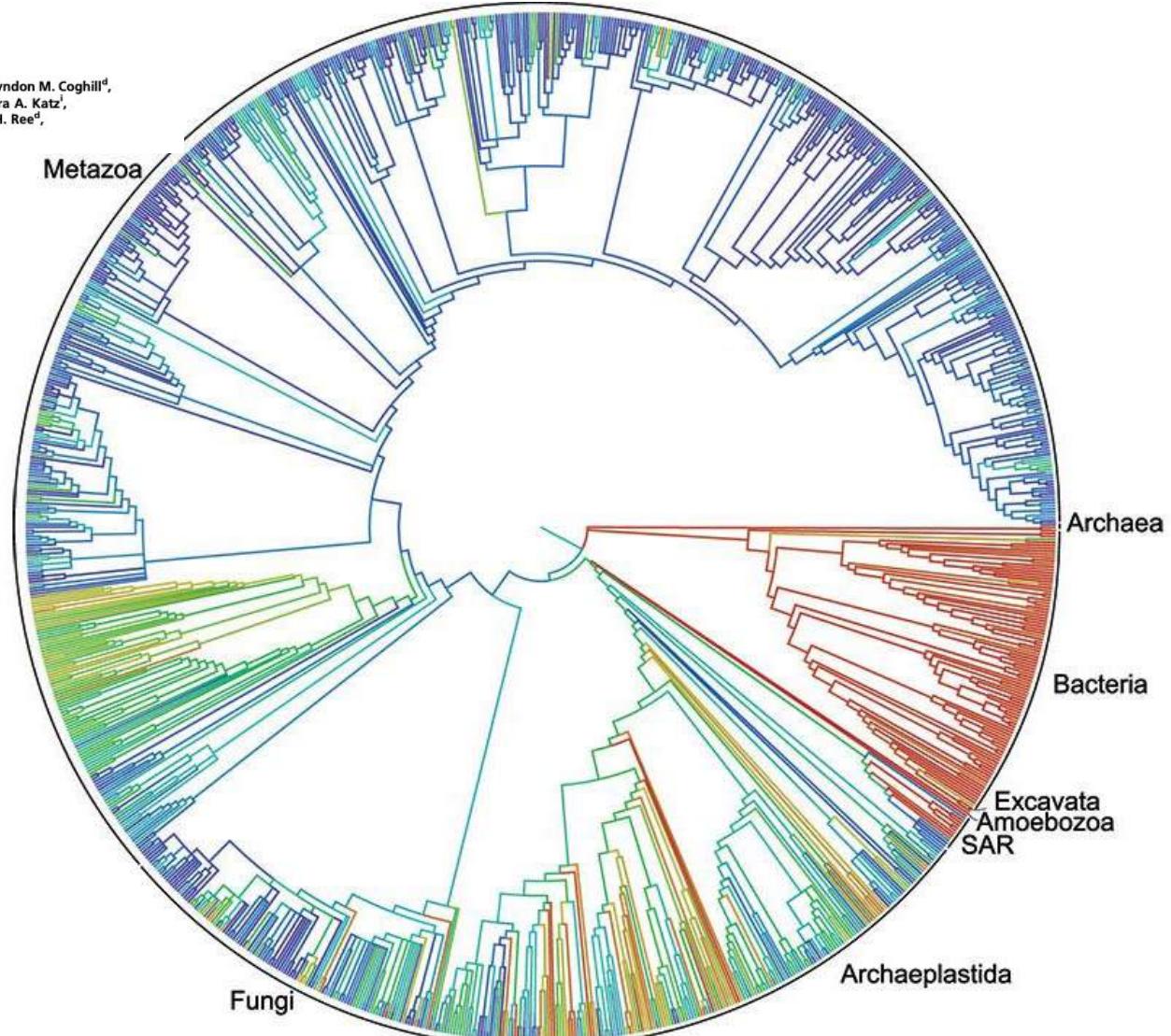




## Synthesis of phylogeny and taxonomy into a comprehensive tree of life

Cody E. Hinchliff<sup>a,1</sup>, Stephen A. Smith<sup>a,1,2</sup>, James F. Allman<sup>b</sup>, J. Gordon Burleigh<sup>c</sup>, Ruchi Chaudhary<sup>c</sup>, Lyndon M. Coghill<sup>d</sup>, Keith A. Crandall<sup>e</sup>, Jiaxin Deng<sup>f</sup>, Bryan T. Drew<sup>f</sup>, Romina Gazis<sup>g</sup>, Karl Gude<sup>h</sup>, David S. Hibbett<sup>g</sup>, Laura A. Katz<sup>i</sup>, H. Dail Laughinghouse IV<sup>j</sup>, Emily Jane McTavish<sup>j</sup>, Peter E. Midford<sup>d</sup>, Christopher L. Owen<sup>c</sup>, Richard H. Ree<sup>d</sup>, Jonathan A. Rees<sup>k</sup>, Douglas E. Soltis<sup>c,l</sup>, Tiffani Williams<sup>m</sup>, and Karen A. Cranston<sup>b,2</sup>

NAS



~2.7 million species  
~1.4 million synonyms  
~0.5 million unknowns

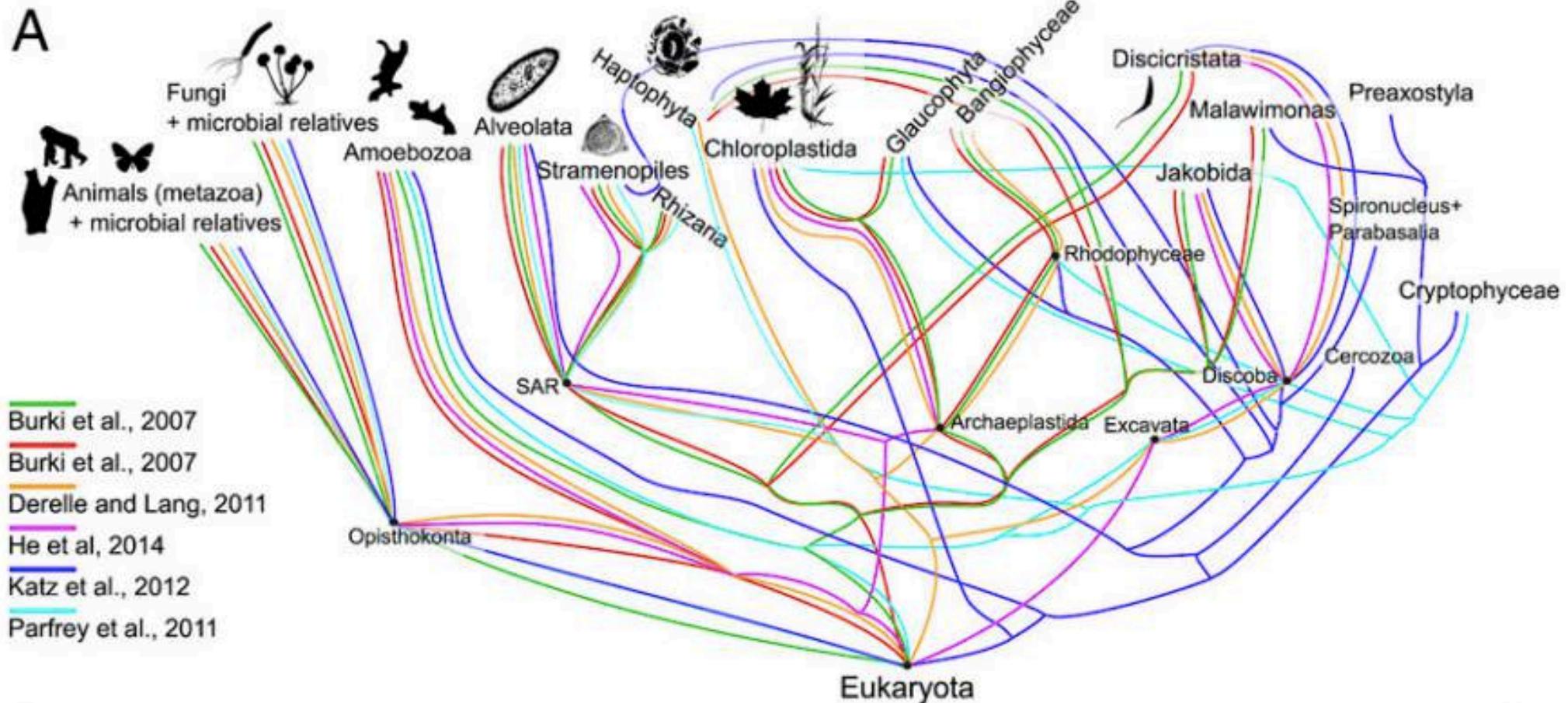


## Synthesis of phylogeny and taxonomy into a comprehensive tree of life

Cody E. Hinchliff<sup>a,1</sup>, Stephen A. Smith<sup>a,1,2</sup>, James F. Allman<sup>b</sup>, J. Gordon Burleigh<sup>c</sup>, Ruchi Chaudhary<sup>c</sup>, Lyndon M. Coghill<sup>d</sup>, Keith A. Crandall<sup>e</sup>, Jiaxin Deng<sup>f</sup>, Bryan T. Drew<sup>f</sup>, Romina Gazis<sup>g</sup>, Karl Gude<sup>h</sup>, David S. Hibbett<sup>g</sup>, Laura A. Katz<sup>i</sup>, H. Dail Laughinghouse IV<sup>j</sup>, Emily Jane McTavish<sup>j</sup>, Peter E. Midford<sup>d</sup>, Christopher L. Owen<sup>c</sup>, Richard H. Ree<sup>d</sup>, Jonathan A. Rees<sup>k</sup>, Douglas E. Soltis<sup>c,l</sup>, Tiffani Williams<sup>m</sup>, and Karen A. Cranston<sup>b,2</sup>

NAS

## Conflicts / Lack of resolution

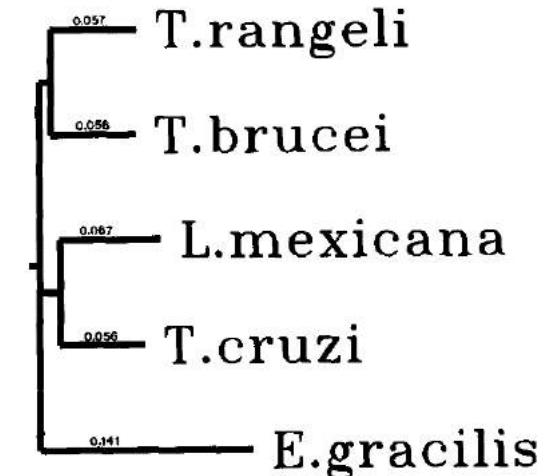


*Trypanosoma rangeli*: sequence analysis of  $\beta$ -tubulin gene suggests closer relationship to *Trypanosoma brucei* than to *Trypanosoma cruzi*

M. Isabel Amorim, Hooman Momen and Yara M. Traub-Cseko

*Instituto Oswaldo Cruz, FIOCRUZ, Departamento de Bioquímica e Biologia Molecular, Rio de Janeiro, Brazil*

Tcruzi	G G T G G T A C G G G G C T C T G G C A T G G G A A C T C T G C T C A T C T C G A A G C T G C G C	48
Trangeli	G G T G G C A G G G G C T C G G G A A T G G G A A C G C T G C T C A T T T C G A A G C T T T C G G	48
Tbrucei	G G T G G T A C T G G C T C C G G C A T G G G T A C G C T G C T C A T C T C G A A G C T T T C G C	48
Lmexicana	G G C G G G C A C G G G G C T C C G G C A T G G G C A C G C T G C T C A T C T C C A A G C T G C G C	48
Euglena	G G T G G C A C G G G G T C C G G A A T G G G C A C C C T C C T C A T C T C C A A G A T C C G T	48
Tcruzi	G A G G G A G T A C C C A G A C C G G C A T C A T G A T G A C C T T C - - A A C A T C A A G T C C	93
Trangeli	G A G G C A G T A C C C G G A C C G G C A T C A T G A T G A C C T T C - - A A C A T C A A G T C G	93
Tbrucei	G A G G C A G T A C C C T G A C C G G T A T C A T G A T G A C T T T C - - A A C A T C A A G T C C	93
Lmexicana	G A G G G A G T A C C C G G A C C G G C A T G A T G A C C T T C - - A A C A T C A A G T C C	93
Euglena	G A G G G A G T A C C C T G A C C G G C A T G A T G A C C T T C - - A A C A T C A A G T C C	93
Tcruzi	T C C A T C T G C G A C A T C C C G C C C A A G G G G C T C A A G A T G G C C G T C A C C T T T	141
Trangeli	T C C G T C T G C G A C A T C C C C C C C A A G G G G C T C A A G A T G G C T G T C A C C T T C	141
Tbrucei	T C T G T T T G C G A T A T C C C A C C C A A G G G G A C T C A A G A T G G G T G T C A C C T T C	141
Lmexicana	T C C A T C T G C G A T A T C C C G C C C A A G G G G T C T C A A G A T G T C C G T C A C C T T C	141
Euglena	T C T G T C T G C G A C A T C C C T C C C A A G G G G C T G A A G A T G T C T G C C A C C T T C	141
Tcruzi	G T C G G C A A C A A C A C C T G C A T C C C A G G G A G A T G T T T C G C C C G C G T G G G C G A G	189
Trangeli	G T T G G C A A C A A C A C C T G C A T C C C A G G G A G A T G T T T C C G C C G G T C G G A G A G	189
Tbrucei	A T T G G C A A C A A C A C C T G C A T C C C A G G G A G A T G T T T C C G C C G G T G T G G G A G A G	189
Lmexicana	A T C G G C A A C A A C A C C T G C A T C C C A G G G A G A T G T T T C C G C C G G T C G G G T G A G	189
Euglena	A T T G G G A A C A A C A C C G G C C A T C C C A G G G A G A T G T T T C A A G C G A G T G T C T G A G	189
Tcruzi	C A G T T C A C G G C A T G T T C C G C C G C A A G G G C G T T C T T G C A C T G G T A C	234
Trangeli	C A G T T C A C G G C T G A T G T T C C G C C G C A A G G G C G T T C T T G C A C T G G G T A C	234
Tbrucei	C A G T T C A C C C T C A T G T T C C G C T G C A A G G G C G T T C T T G C A C T G T A C	234
Lmexicana	C A G T T C A C G G G T A T G T T C C G C C G C A A G G G C G T T C T T G C A C T G G G T A C	234
Euglena	C A G T T C A C T G C C A T G T T C C G T G C A A G G G C T T C T T G C A T T G G G T A C	234



## Mini-exon gene sequence polymorphism among *Trypanosoma rangeli* strains isolated from distinct geographical regions

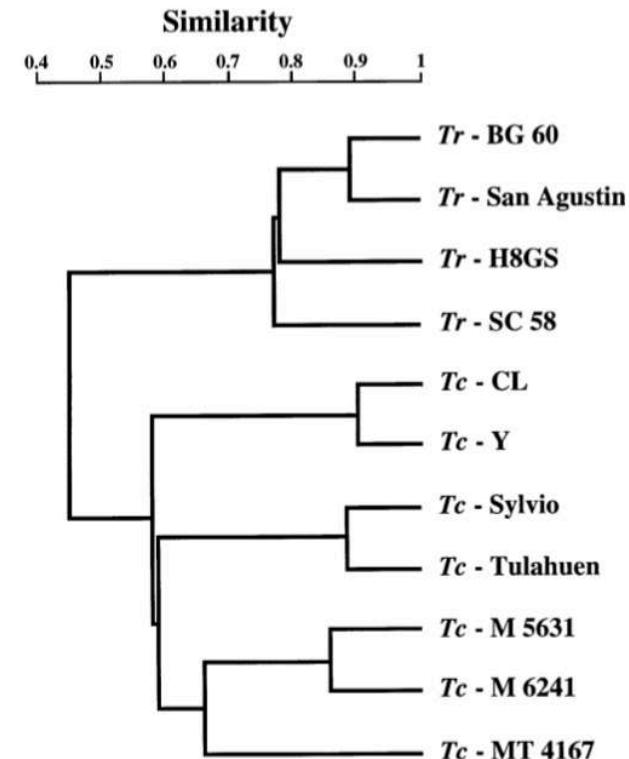
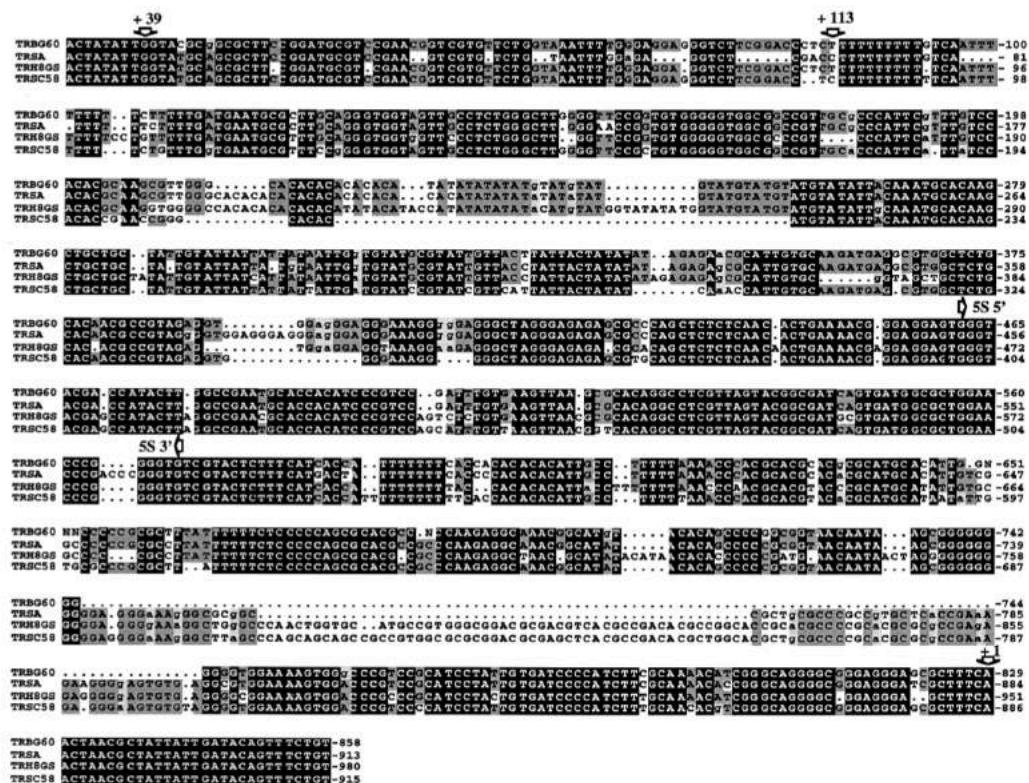
E. C. GRISARD<sup>1, 2, 4†</sup>, D. A. CAMPBELL<sup>2\*</sup> and A. J. ROMANHA<sup>3</sup>

<sup>1</sup> Departamento de Microbiologia e Parasitologia, Universidade Federal de Santa Catarina, Florianópolis, SC, Caixa postal 476, 88040-900, Brazil

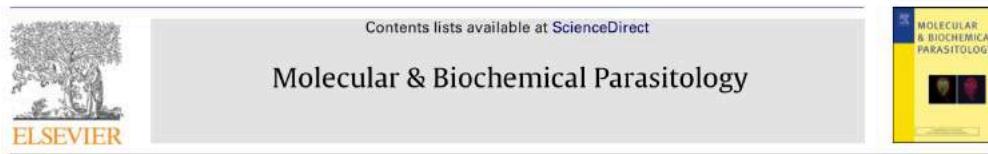
<sup>2</sup> Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, CA, USA

<sup>3</sup> Centro de Pesquisas René Rachou, FIOCRUZ, Belo Horizonte, MG, Brazil

<sup>4</sup> Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil



Molecular & Biochemical Parasitology 174 (2010) 18–25

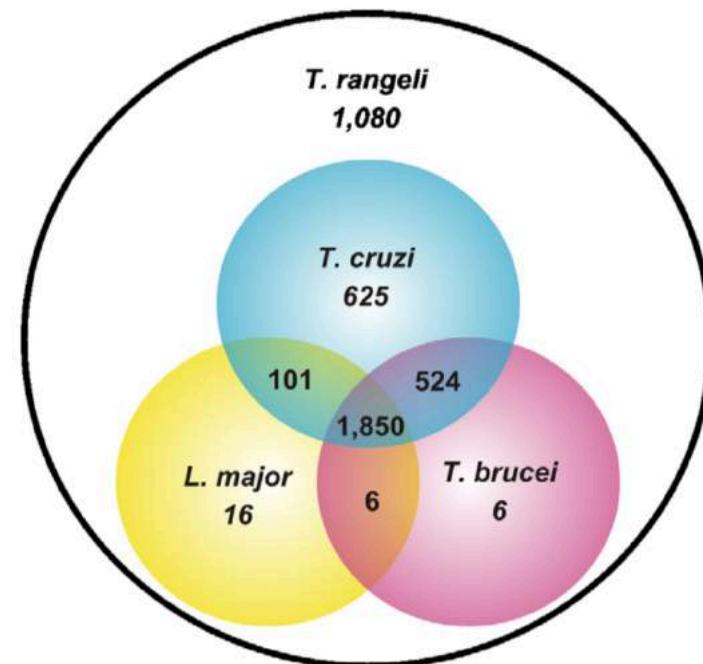


## Transcriptomic analyses of the avirulent protozoan parasite *Trypanosoma rangeli*<sup>☆</sup>

Edmundo C. Grisard<sup>a,f,\*1</sup>, Patrícia H. Stoco<sup>a,f,1</sup>, Glauber Wagner<sup>a,b,1</sup>, Thaís C.M. Sincero<sup>a,1</sup>, Gianinna Rotava<sup>a,2</sup>, Juliana B. Rodrigues<sup>a,1</sup>, Cristiane Q. Snoeijer<sup>a,1</sup>, Leonardo B. Koerich<sup>a,1</sup>, Maísa M. Sperandio<sup>a,2</sup>, Ethel Bayer-Santos<sup>a,2</sup>, Stenio P. Fragoso<sup>c,3</sup>, Samuel Goldenberg<sup>c,3</sup>, Omar Triana<sup>d,2</sup>, Gustavo A. Vallejo<sup>e,2</sup>, Kevin M. Tyler<sup>f,1</sup>, Alberto M.R. Dávila<sup>g,2</sup>, Mário Steindek<sup>a,1</sup>

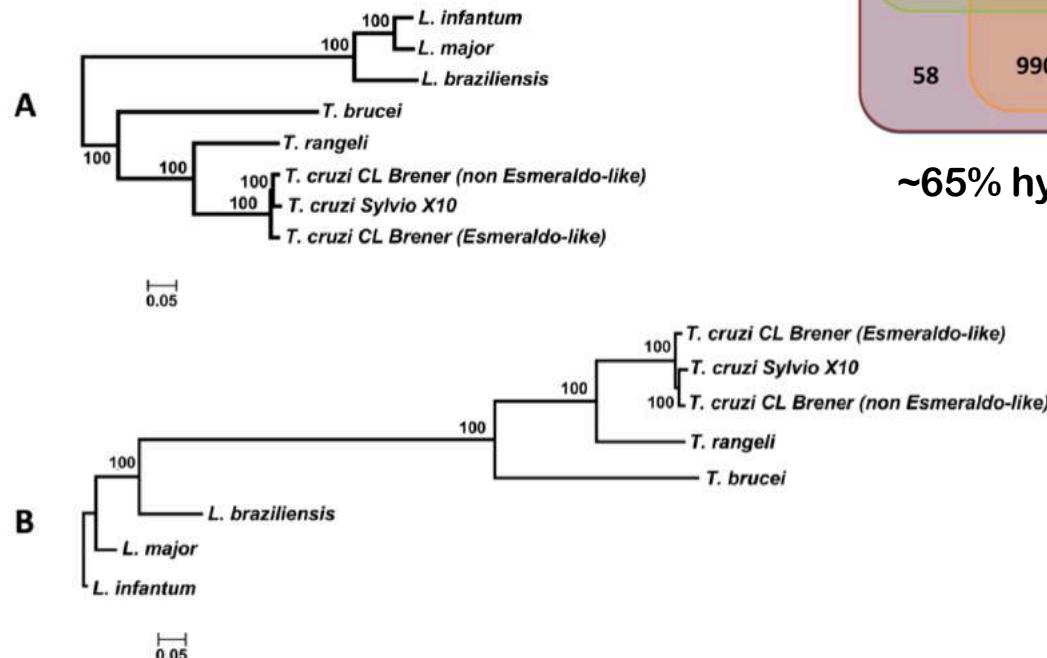
**Table 2**  
General characteristics of the *Trypanosoma rangeli* transcriptome analysis.

Parameter	Value
Total number of sequenced clones	10,617
Total number of non-redundant sequences	4208
Average length of non-redundant sequences (bp)	370
Number of non-redundant sequences with hits (blast/RPS-blast/Interpro/HMMER)	3236
Sequences without hits (blast/RPS-blast/Interpro/HMMER)	972
Number of validated sequences/parasite form	2942
Epimastigote	1198
Trypomastigote	1530
Epimastigote and trypomastigote	214
Hypothetical/conserved hypothetical	1076
Number of distinct annotations	459

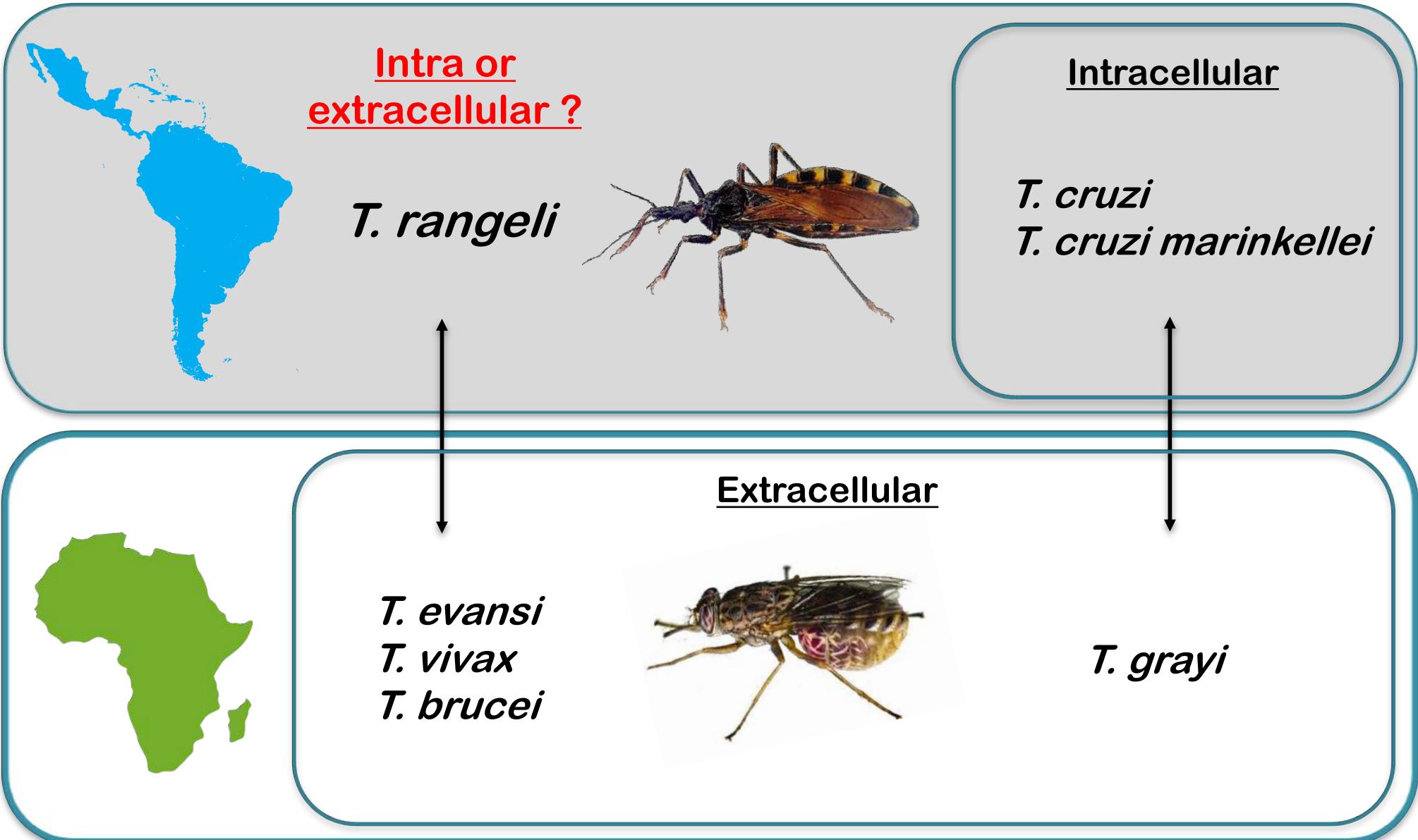


## Genome of the Avirulent Human-Infective Trypanosome—*Trypanosoma rangeli*

Patrícia Hermes Stoco<sup>1\*</sup>, Glauber Wagner<sup>1,2</sup>, Carlos Talavera-Lopez<sup>3</sup>, Alexandra Gerber<sup>4</sup>, Arnaldo Zaha<sup>5</sup>, Claudia Elizabeth Thompson<sup>4</sup>, Daniella Castanheira Bartholomeu<sup>6</sup>, Débora Denardin Lückemeyer<sup>1</sup>, Diana Bahia<sup>6,7</sup>, Elgion Loreto<sup>8</sup>, Elisa Beatriz Prestes<sup>1</sup>, Fábio Mitsuo Lima<sup>7</sup>, Gabriela Rodrigues-Luiz<sup>6</sup>, Gustavo Adolfo Vallejo<sup>9</sup>, José Franco da Silveira Filho<sup>7</sup>, Sérgio Schenkman<sup>7</sup>, Karina Mariante Monteiro<sup>5</sup>, Kevin Morris Tyler<sup>10</sup>, Luiz Gonzaga Paula de Almeida<sup>4</sup>, Mauro Freitas Ortiz<sup>5</sup>, Miguel Angel Chiurillo<sup>7,11</sup>, Milene Höehr de Moraes<sup>1</sup>, Oberdan de Lima Cunha<sup>4</sup>, Rondon Mendonça-Neto<sup>6</sup>, Rosane Silva<sup>12</sup>, Santuza Maria Ribeiro Teixeira<sup>6</sup>, Silvane Maria Fonseca Murta<sup>13</sup>, Thais Cristine Marques Sincero<sup>1</sup>, Tiago Antonio de Oliveira Mendes<sup>6</sup>, Turán Peter Urményi<sup>12</sup>, Viviane Grazielle Silva<sup>6</sup>, Wanderson Duarte DaRocha<sup>14</sup>, Björn Andersson<sup>3</sup>, Álvaro José Romanha<sup>1</sup>, Mário Steindel<sup>1</sup>, Ana Tereza Ribeiro de Vasconcelos<sup>3</sup>, Edmundo Carlos Grisard<sup>1\*</sup>

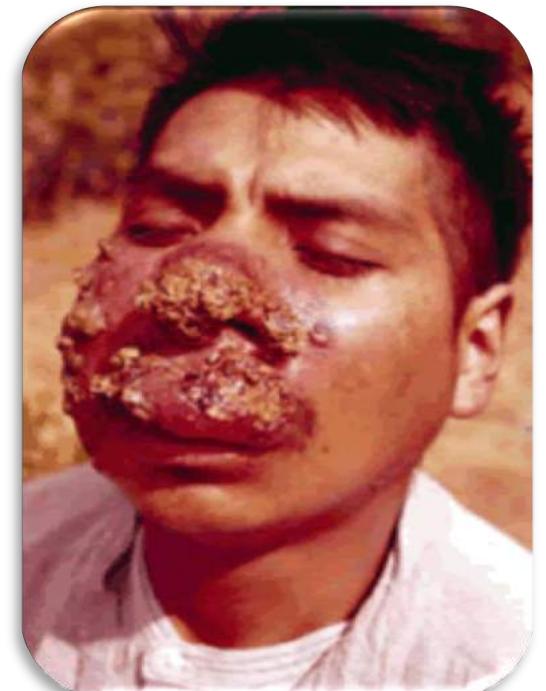


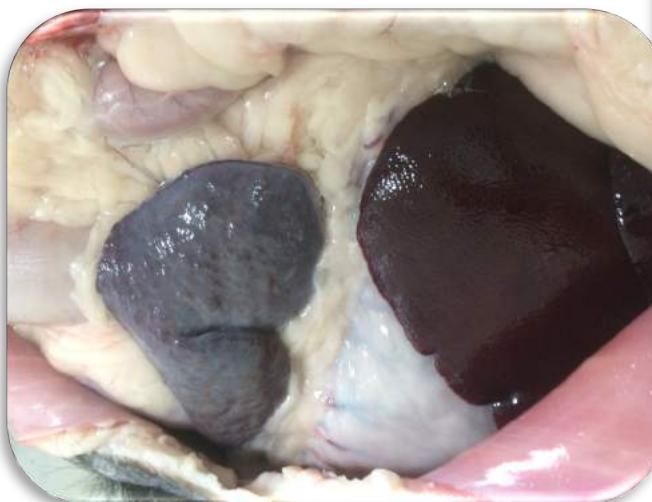
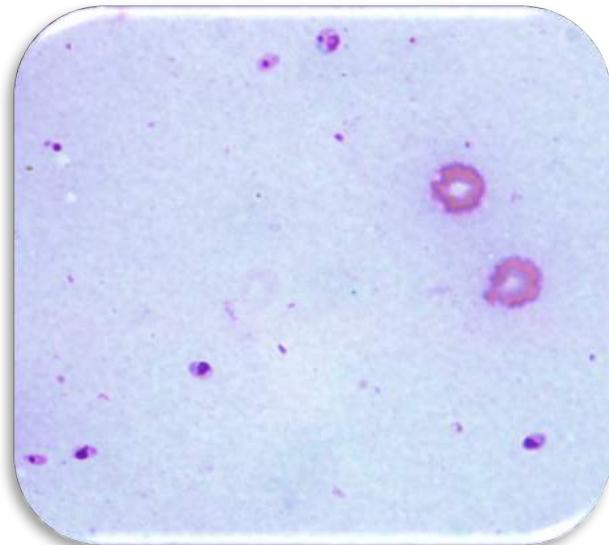
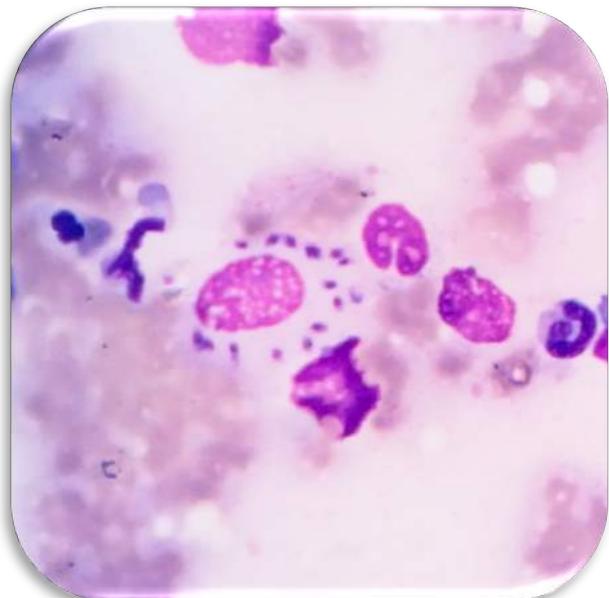
**Figure 3. Evolutionary history of the Trypanosomatidae family obtained through a phylogenomic approach using (A) the neighbor joining (NJ) or (B) the maximum likelihood (ML) methods.** In the NJ results, the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. In the ML results, each internal branch indicates, as a percentage, how often the corresponding cluster was found among the 1,000 intermediate trees. The scale bar represents the number of amino acid substitutions per site.  
doi:10.1371/journal.pntd.0003176.g003



**“Leishmaniasis is a group of heterogeneous clinical manifestations caused by distinct species of protozoan parasites of the Genus *Leishmania*”**

Adapted from Antinori, Schifanella & Corbellino, 2011





Family

*Trypanosomatidae*

Genus

*Leishmania*

Subgenus

*Leishmania* | *Viannia*

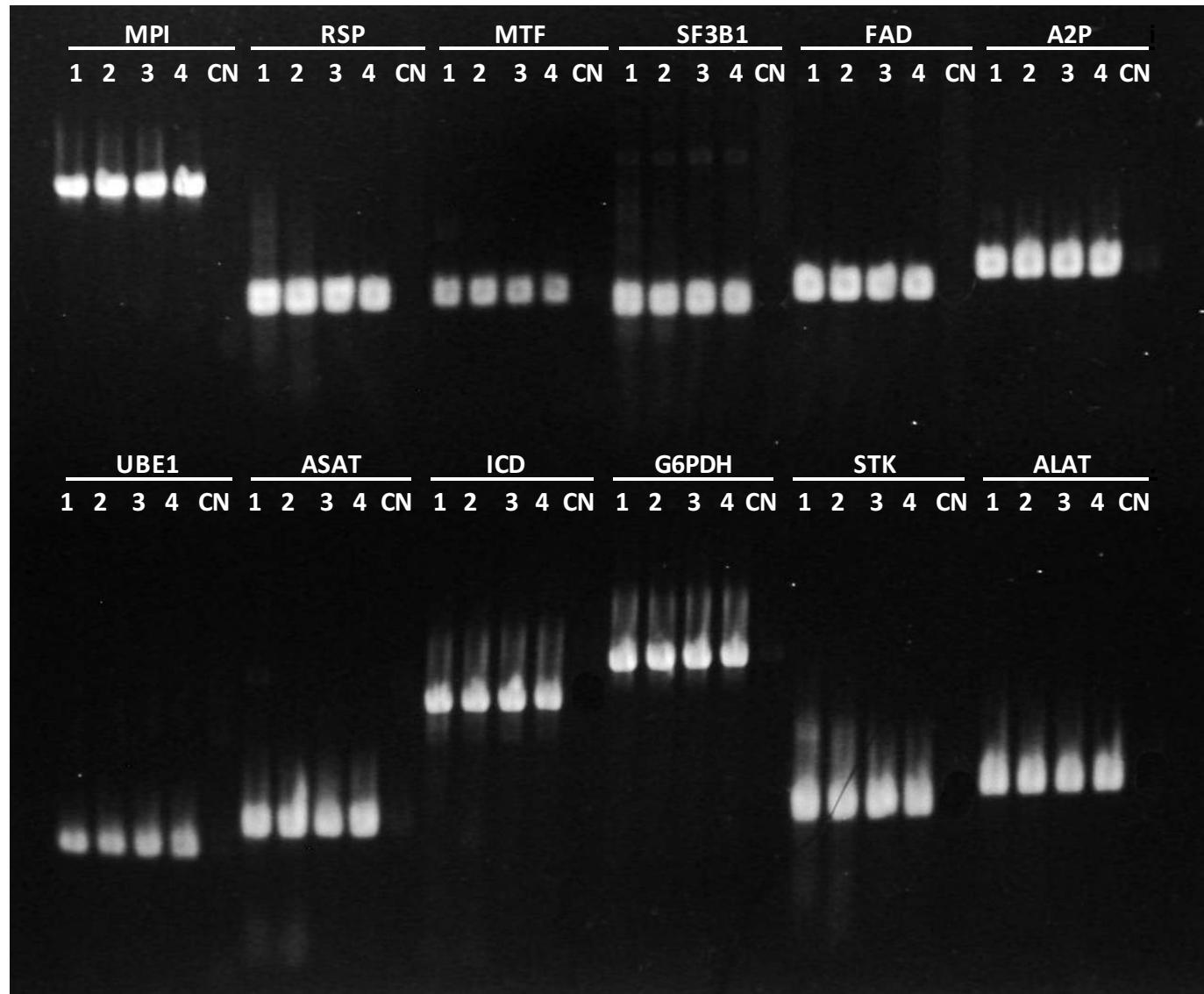
Complex<sup>a</sup> *L. donovani*    *L. tropica*    *L. major*    *L. aethiopica*    *L. mexicana*    *L. braziliensis* — *L. guyanensis*

Species	<i>L. archibaldi</i> <sup>b</sup>	<i>L. killicki</i>	<i>L. major</i>	<i>L. aethiopica</i>	<i>L. amazonensis</i>	<i>L. braziliensis</i>	<i>L. guyanensis</i>
	<i>L. chagasi</i>	<i>L. tropica</i>			<i>L. garnhami</i>	<i>L. peruviana</i>	<i>L. panamensis</i>
	<i>L. donovani</i>				<i>L. mexicana</i>	<i>L. colombiensis</i>	
	<i>L. infantum</i>				<i>L. pifanoi</i>	<i>L. lainsoni</i>	
				<i>L. venezuelensis</i>	<i>L. shawii</i>		
					<i>L. naiffi</i>		

<sup>a</sup> The classification of complexes is based on biochemical and molecular characteristics.

<sup>b</sup> May not be a separate species.

## Caracterização dos isolados de LV de Florianópolis



## Caracterização dos isolados de LV de Florianópolis

### 5' A2rel

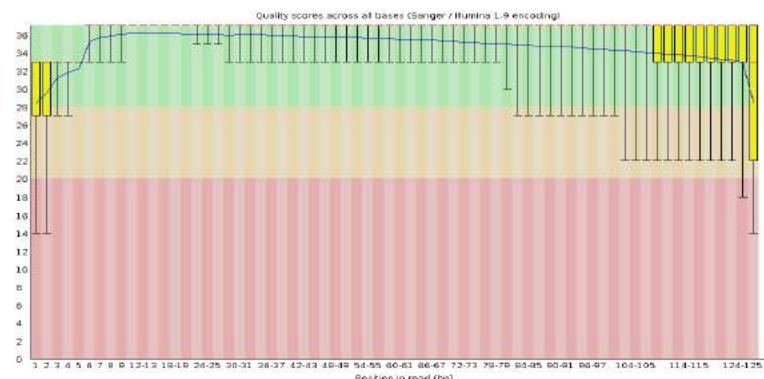
- Amastigote-specific
- Visceralization
- Virulence

	210	220	230	240	250	260	270	280	290	300
REF LLM-877	.... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
2788 H ES	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
2972 H PI	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
PC H SC	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
2666 C MS	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
2919 C DF	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3135 C MT	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3177 C RN	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3186 C RN	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3196 C RS	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3256 C SP	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3257 C SP	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3379 C MG	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
3381 C MG	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
BOB C SC	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
SNOOPY C SC	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
LUMA C SC	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
NICK C SC	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG									
LOBA C SC	GGGGAAAAGGCCGTTCAAGGGCTGGCGCTCAACCTTGCGCGCTGGCGACTGTGGCTGGCGAGGCCGCTTTCGGCGCTGGCTGAATCCGGCG	*****	*****	*****	*****	*****	*****	*****	*****	*****
	310	320	330	340	350	360	370	380	390	400
REF LLM-877	.... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
2788 H ES	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
2972 H PI	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
PC H SC	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
2666 C MS	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
2919 C DF	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3135 C MT	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3177 C RN	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3186 C RN	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3196 C RS	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3256 C SP	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3257 C SP	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3379 C MG	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
3381 C MG	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
BOB C SC	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
SNOOPY C SC	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
LUMA C SC	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
NICK C SC	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG									
LOBA C SC	GGGGCCCCGCAAAGCTGCTGAGGGCGCGCTGGCGACCTCCCACAAACTTGCGGAGCGAGCACAACTAAAGCCTCAAGCGCGCGCAACITTTGTG	*****	*****	*****	*****	*****	*****	*****	*****	*****

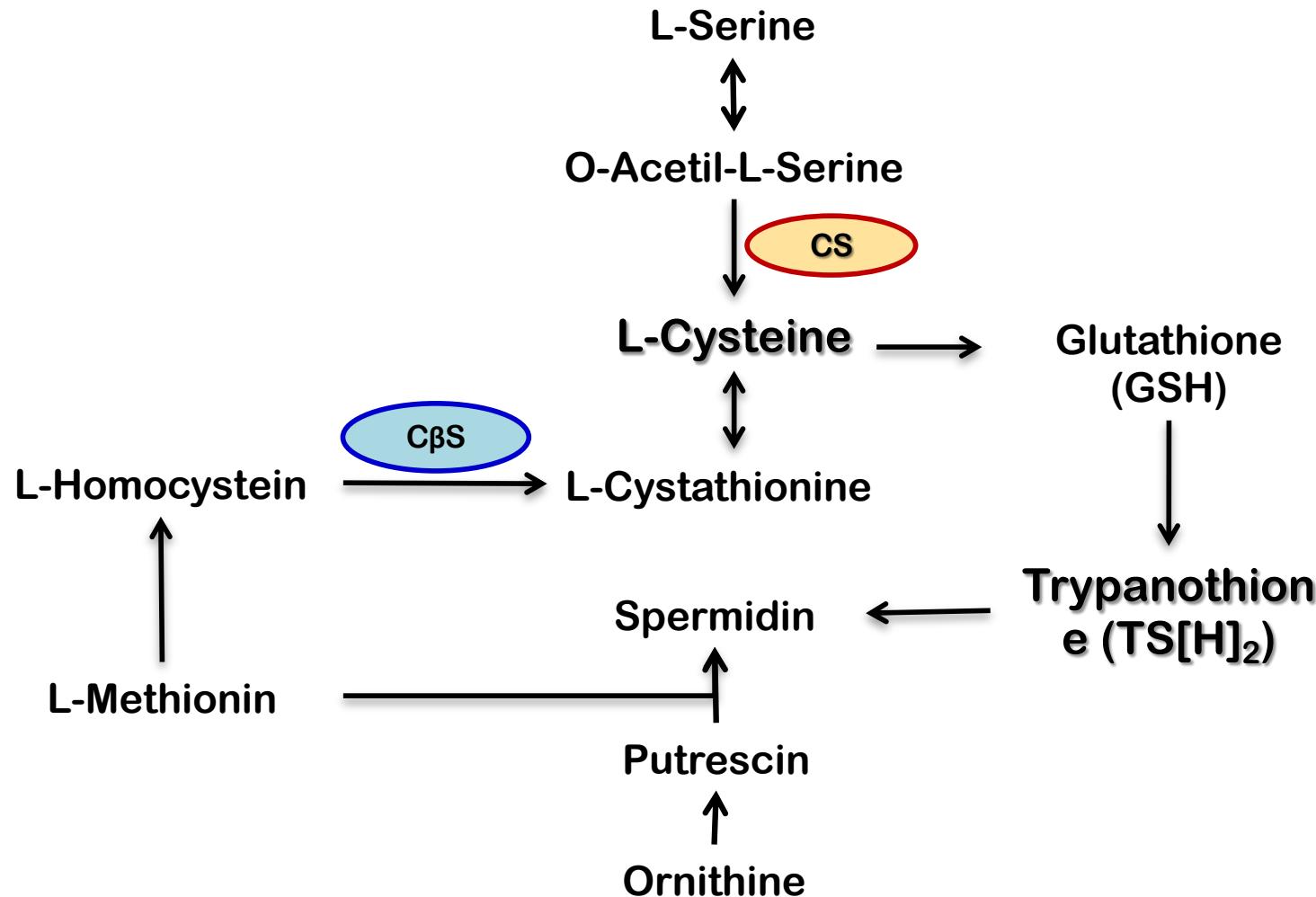
## Genomes of *L. infantum* from Florianópolis

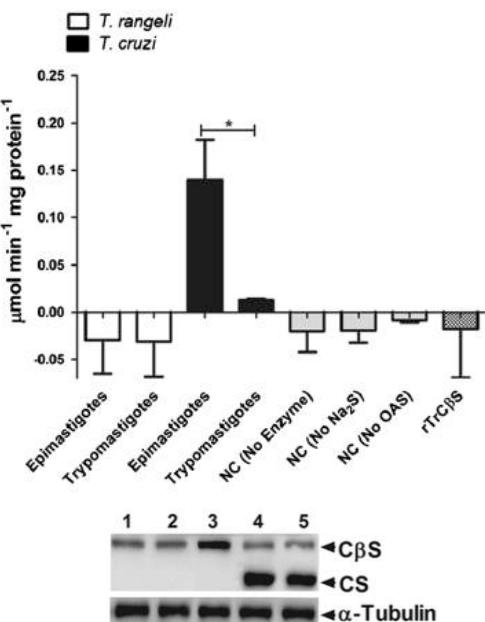
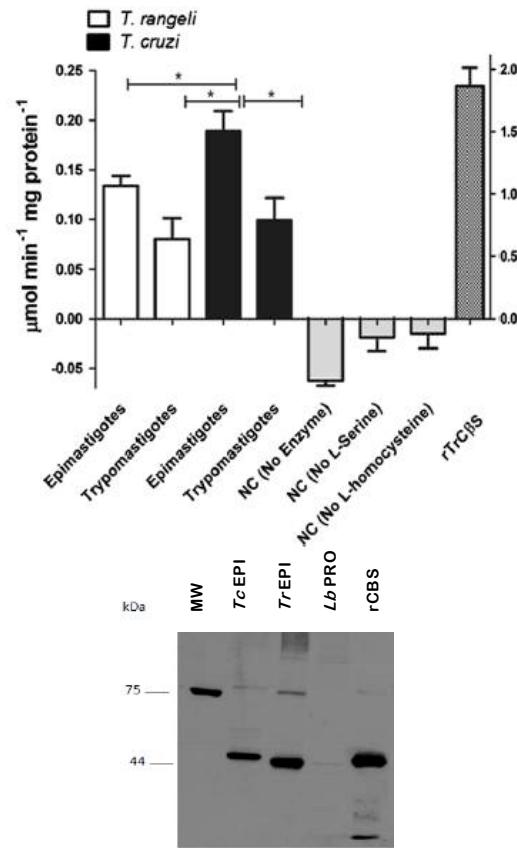
ISOLATES ID	TOTAL NUMBER OF <i>READS</i>	<i>READS MAPPED ON THE REFERENCE GENOME (ABSOLUTE)</i>	<i>READS MAPPED ON THE REFERENCE GENOME (%)</i>
SC1006 (DOG, 2015)	20972393	20550871	97.99%
SC1007 (DOG, 2016)	32328874	31682098	98.00%
SC1008 (DOG, 2010)	21575573	20959774	97.15%
SC1009 (HUMAN, 2017)	24426247	23365451	95.66%

+ 4 isolates



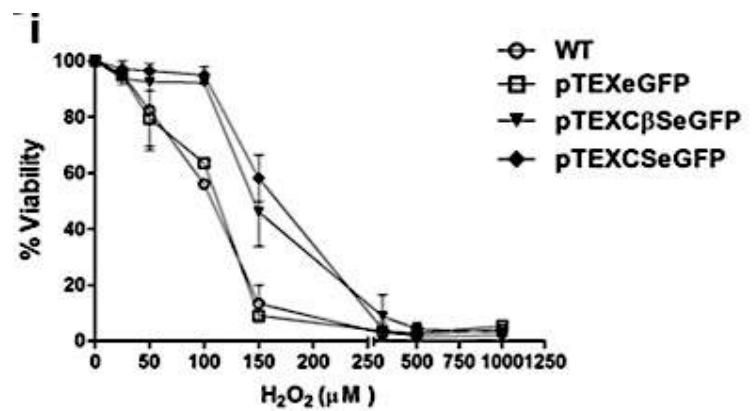
## Cysteine / Cystathione biosynthesis





**TRASC58CS** 336 TGCGGCCTTT-GGCATGAAGTGGGCCGTGACAT-TGGTGAA-CAGCATT--GCCA-GC  
**TRACHOCS** 332 TGCAGGCCTTT-GGCATGAAGTGGGCCGTGACAT-TGGTGAA-CAGCATT--GCCA-GC  
**TCRCS** 378 CGCGGCACCTT-GGTATGAAGGGCGCTTGGAG-CAGCAA-TCGCATT--GTCA-GC  
**LBRACS** 375 CGCCGCTCTC-GGCATGAAGGGTGCCTGGGA-TGGCAA-GAAGATC--GTGCG-C  
**LMJCS** 375 TGCCGCTCTC-GGCATGAAGGGTGCCTGGGA-TGGCAA-GAAAATC--GTGA-GC  
**LINFCS** 375 TGCCGCTCTT-GGCATGAAGGGTGCCTGGGA-TGGCAA-GAAAATC--GTGCG-C  
  
**TRASC58CS** 426 AAAATA-CAAC--ACCCAGATCC--ACCAGGATACCACCGGGCCG[GAGAT]TAGGGC  
**TRACHOCS** 422 AAAATA-CAAT--ACCCAGATCC--ACCAGGATACCACCGGGTCGGAGAT[TAGGGC]  
**TCRCS** 468 CAAATA-TAAT--GCCCAATTC--ACGAAGAAACAAC[GAGAT]TTGGGCG  
**LBRACS** 465 CAAGTA-CAAC-GCCCTATGC-ACGAGGAAACTACA-GGACCT[GAGAT]TTGGGAI  
**LMJCS** 465 CAAGTA-CAAC-GCCCTATAC-ACGAGGAGACCAAGGGCCT[GAGAT]TTGGGAC  
**LINFCS** 465 CAAGTA-CAAC-GCCCTATAC-ACGAGGAAACCAAGGGCCT[GAGAT]TTGGGAC

STOP CODON



Parasite	Plasmid transfected into cells	$\text{IC}_{50}^a$ ( $\mu\text{M}$ )	
		$\text{H}_2\text{O}_2$	SNAP
<i>T. rangeli</i>	WT <sup>b</sup>	$66 \pm 4$	$353 \pm 10$
	pTEXeGFP	$69 \pm 0.3$	$316 \pm 17$
	pTEXCSeGFP	$160 \pm 19^{**}$	$402 \pm 49$
	pTEXC $\beta$ SeGFP	$153 \pm 18^{**}$	$345 \pm 3$
<i>L. braziliensis</i>	WT	$560 \pm 4$	$99 \pm 13$
	pTEXeGFP	$558 \pm 3$	$96 \pm 3$
	pTEXCSeGFP	$1,368 \pm 7^{***}$	$93 \pm 6$
	pTEXC $\beta$ SeGFP	$1,551 \pm 33^{***}$	$80 \pm 17$

Romero *et al.*, AAC, 2015

Romero *et al.*, P&V, 2014

### *T. rangeli*:

- Lacks *de novo* cysteine biosynthesis
- Single route for cysteine biosynthesis – homocysteine
- Uses the transulfuration pathway (as *Tb*)

## Challenges for MASTER/PrInt/CAPES:

Improve the genome assembling strategies:

- Long, repetitive sequences
- Non-coding regions
- Hypothetical / Conserved Hypothetical proteins

Improve genome annotation:

- Algorithm development & training
  - Multiple biological variables
- Time consuming manual curation



**Professor Mário Steindel**  
**Dr<sup>a</sup>. Patricia Stoco**  
**Dr. Glauber Wagner**  
**Dr<sup>a</sup>. Milene de Moraes**  
**Dr<sup>a</sup>. Débora Lückemeyer**  
**Dr<sup>a</sup>. Ibeth Romero**  
**Dr<sup>a</sup>. Elisa Prestes**  
**Dr. Abadio Costa Jr.**  
**Caibe Alves, MSc**  
**Laís Yamanaka, MSc**  
**Carime Pontes, MSc**  
**Adriana Silva, MSc**  
**Fábio Indá, MSc**  
**Tatiana Catecati, MSc**  
**Laryssa de Liz**  
**Viviane Flores**  
**Amábilli Rosar**  
**Natalia Gutierrez**



**Dr Luiz Felipe Oliveira**  
**Dr Aline Sereia**

**Dr Kevin Tyler**



**Dr Gustavo Vallejo**

**Professor John Kelly**  
**Dr Michael Lewis**



**Professor Björn Andersson**  
**Dr Hamid Darban**

**Dr Ana Teresa Vasconcellos**



**Dr Hercules Moura**

**Dr Samuel Goldenberg**  
**Dr Silvane Murta**  
**Dr Elisa Cupolillo**



**Professora Santuza Teixeira**  
**Dr Daniela Bartholomeu**  
**Dr Ricardo Fujiwara**  
**Dr João Cunha**  
**Jennifer Ottino MSc**

**Professor Arnaldo Zaha**  
**Professor Henrique Ferreira**



**Professor José Franco da Silveira**  
**Professora Nobuko Yoshida**  
**Professor Sergio Schenkman**

