ProtozoaDB: comparative genomics approaches for the study of protozoan species

Alberto M. R. Dávila
Email: davila@fiocruz.br

Laboratório de Biologia Computacional e Sistemas

Instituto Oswaldo Cruz – FIOCRUZ, Rio de Janeiro, Brasil.

Background

- Pathogenic protozoan species are commonly studied because they are a major cause of mortality and morbidity in humans and domestic animals throughout the world.

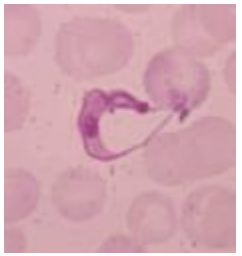
- We have developed an automatic methodology to reconstruct the phylogenetic species tree in Protozoa integrating different phylogenetic algorithms and programs.

- We demonstrate the utility of a supermatrix approach to construct phylogenomic trees using 31 universal representative orthologs.

- These universal orthologs are involved in metabolic pathways, thus might potentially provide some extra information for the understanding of the phylogenetic relationships and evolutionary processes of Protozoa.

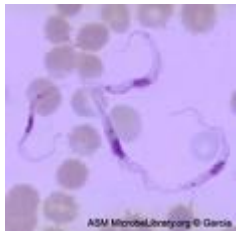
Introduction

- Illness caused by parasitic Protozoa are a major cause of disease worldwide, but because they are concentrated in low socio-economic parts of the world, they receive relatively little attention from the pharmaceutical industry and major scientific funding agencies.



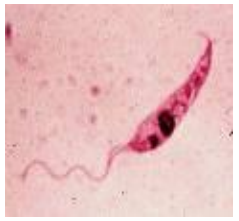
Trypanosoma cruzi

- Of the ten diseases targeted as research priorities by the World Health Organization's Special Program for Research and Training in Tropical Diseases (http://www.who.int/tdr) four are caused by protozoan parasites (malaria, leishmaniasis, Chagas disease, African tripanosomiasis).



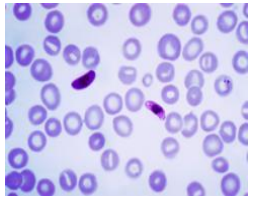
Trypanosoma brucei

- These diseases and other less dangerous ones (e.g. amoebiasis and trichomoniasis) are having an alarming increase of refractoriness cases to the main treatment. Treatment failure has potentially a multifactorial origin with one of the major concerns being drug resistance.



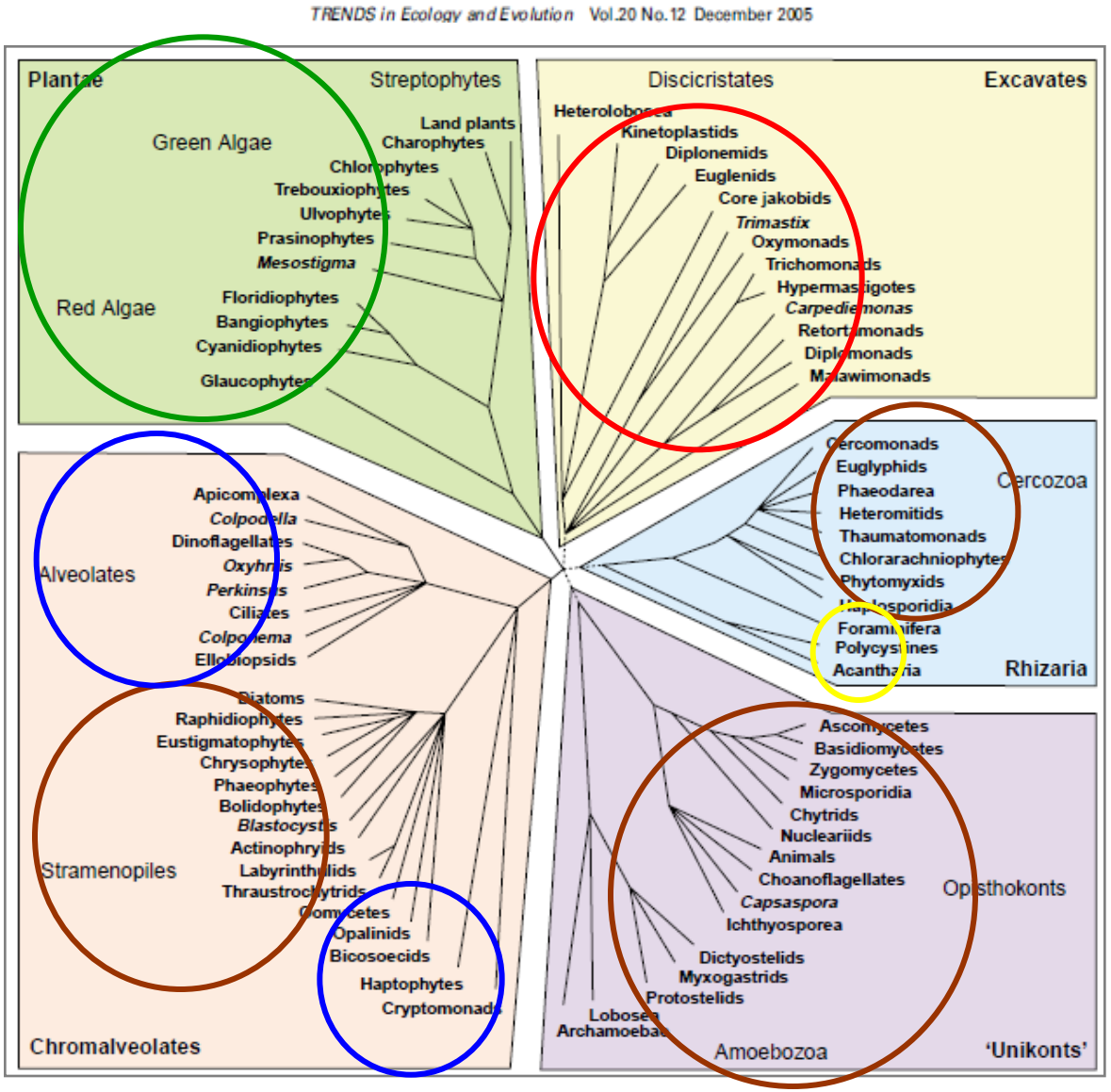
Leishmania major

- Phylogenomics is being intensively used in the "Post Genomic Era". Many different phylogenetic trees have been published on the basis of different models of sequence evolution and utilization of different parameter settings and algorithms.



Plasmodium falciparum

Introduction



- Although, the rRNA and other single genes have been extremely valuable for phylogenetic studies but they have their limitations.

- Chaudhary (2005) and Keeling (2005) showed results on available sequencing data.

- Phylogenetic analysis defined five supergroups of eukaryotes: **(i)** the plant and red/green algal lineage;

(ii) a clade comprised of animals, fungi, slime molds and amebozoans;

and three groups that are entirely protozoan:

(iii) chromalveolates,

(iv) excavates and

(v) rhizaria.

Universal Orthologs:

Ciccareli et al (2006): UO for tree of life

Serra et al (2008): Orthosearch

Toward Automatic Reconstruction of a Highly Resolved Tree of Life

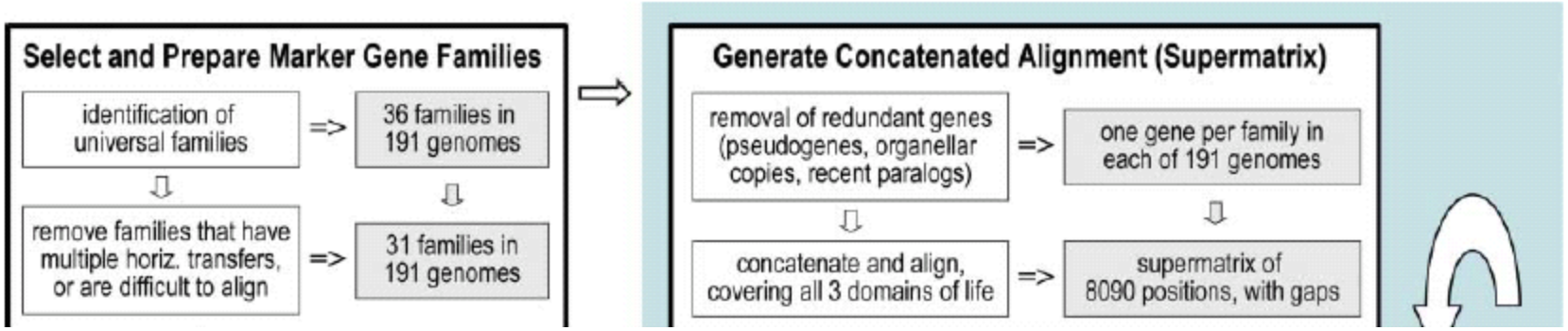
Francesca D. Ciccarelli,[1,2,3*] Tobias Doerks,[1*] Christian von Mering,[1] Christopher J. Creevey,[1] Berend Snel,[4] Peer Bork[1,5]†

We have developed an automatable procedure for reconstructing the tree of life with branch lengths comparable across all three domains. The tree has its basis in a concatenation of 31 orthologs occurring in 191 species with sequenced genomes. It revealed interdomain discrepancies in taxonomic classification. Systematic detection and subsequent exclusion of products of horizontal gene transfer increased phylogenetic resolution, allowing us to confirm accepted relationships and resolve disputed and preliminary classifications. For example, we place the phylum Acidobacteria as a sister group of δ-Proteobacteria, support a Gram-positive origin of Bacteria, and suggest a thermophilic last universal common ancestor.

Reconstructing the phylogenetic relationships among all living organisms is one of the fundamental challenges in biology. Numerous attempts to derive a tree of life using various methods have been published [for a review, see (*1*)], and its principal existence has been questioned recently (*2*, *3*). Moreover, even under the assumption of a tree of life, numerous groupings and taxonomic entities still remain heavily debated, and the advent of molecular and genomic data has increased the variety of classifications rather than reducing the problem (*1*). Theoretical and practical limits to reconstructing a tree of life have been put forward, such as the insufficient amount of discriminating characters available, even in information-rich genomic data sets (*4*), and the computing resources required to cope with large numbers of species (*1*). Furthermore, there are factors that hamper accurate reconstruction of phylogenetic trees regardless of the methods used, such as sampling biases of species included (*5*) and dilution of phylogenetic signal by horizontal gene transfer (HGT) (*6*), the

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany. [2]Department of Experimental Oncology, European Institute of Oncology, Via Ripamonti 435, 20141 Milan, Italy. [3]Institute of Molecular Oncology Foundation, Fondazione Italiana per la Ricerca sul Cancro, Via Adamello 16, 20139 Milan, Italy. [4]Center for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen Medical Center, Toernooiveld 1, 6525 ED, Nijmegen, Netherlands. [5]Max Delbrück Centrum, 13092 Berlin-Buch, Germany.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: peer.bork@embl.de



Science. 2006 Mar 3;311(5765):1283-7.

Selection and preparation of marker gene families

- We based our methodology on Ciccarelli *et al.* (2006) study for the selection and construction of the species tree.

- Thirty-one universal orthologous (UO) genes were downloaded in fasta format from GenBank and then aligned using MAFTT (version 5.861)

Selection and preparation of marker gene families for protozoan species trees

- **Hidden Markov Models** (HMM) profiles were constructed for the **31 UO** (local database) set and queried against all **Protozoa sequences** available in Refseq (RefSeq-release35.catalog-01/11/2009) and Genbank (NCBI-Flat File Release 172.0-06/15/2009).

- HMM profiles were created (**hmmbuild**) and calibrated (**hmmcalibrate**) using HMMER version 2.3.2 and searches done using **hmmpfam** with e-value "1e-5" as cut-off.

- Those best hmmpfam hits of each of the 31 UO against protozoan genomes were used to construct individual multiple alignments (Mafft v5.861).

- **Methodology 1 (M1):**

  - ✓Those individual alignments were **concatenated**, then resulting in a global supermatrix of 21,260 positions in a total of 74 species.

  - ✓ This supermatrix was used to generate a tree with PHYML using 100 bootstrap replicates and the JTT matrix.

- **Methodology 2 (M2):**

  - ✓ Those individual alignments were **trimmed** using TrimAl v1.2 aiming to eliminate the most variable regions of each of them. Those remaining conserved blocks were then concatenated, resulting in a global supermatrix of 12,807 positions in a total of 74 species.

  - ✓ This supermatrix was used to generate a tree with PHYML using 100 bootstrap replicates and the JTT matrix.

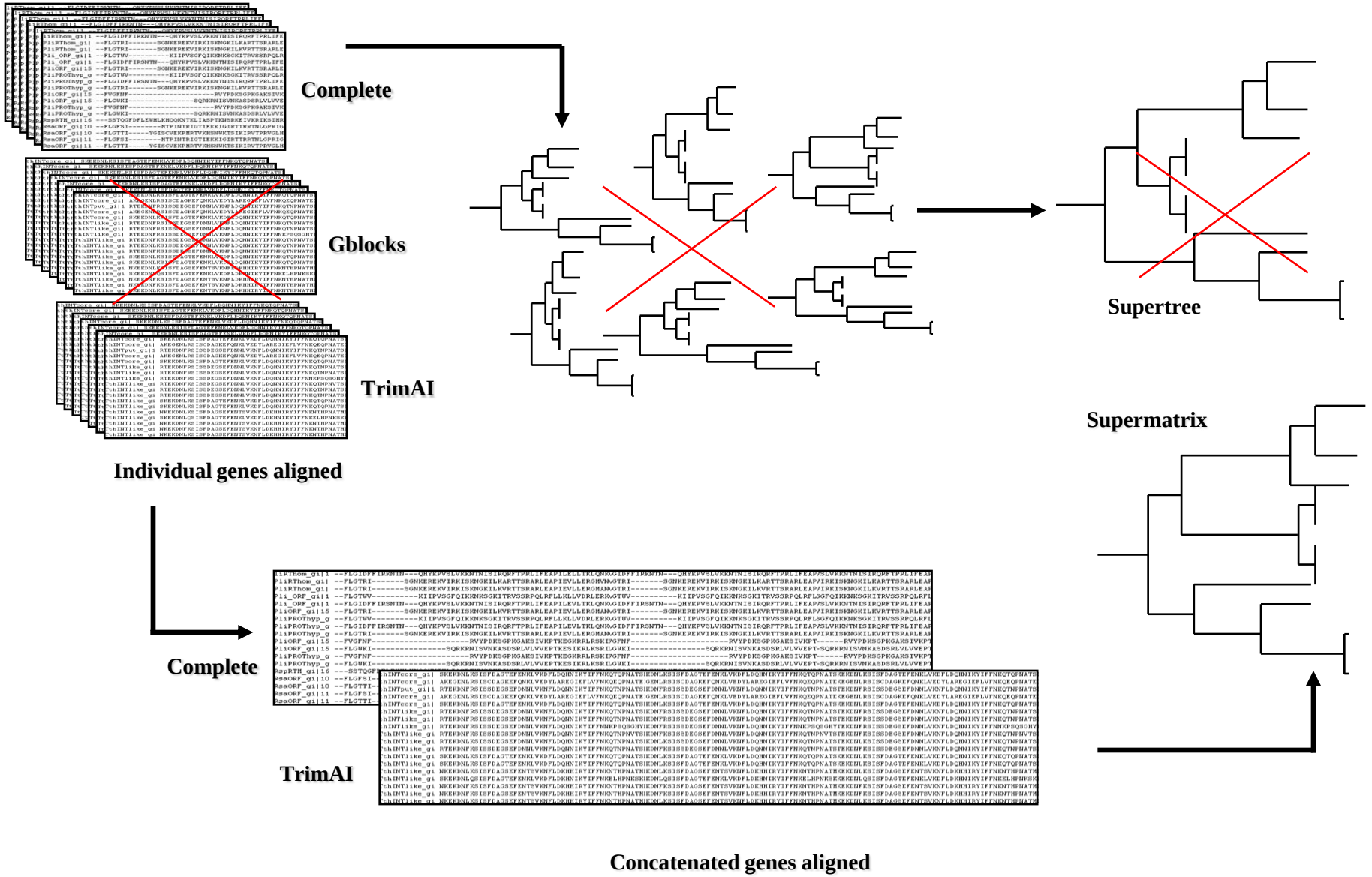


Figure 2. Algorithm for Supermatrix versus Supertree construction.

- We reconstructed the protozoan species tree using several automated bioinformatics tools: downloaded 31 universal UO from GenBank, built HMM profiles with them, searched those profiles in protozoan genomes (43 species) and related species (31 species) showing different level of completeness/coverage.

- The species tree was formed by 3 major clades that were related to 3 groups of data:

(i) 26 species composed of at least 80% of UO (25/31) called **G1**,
(ii) 12 species composed of 50-79% (15-24/31) of UO called **G2** and
(iii) 36 species composed of less than 50% (1-14/31) of UO called **G3**.

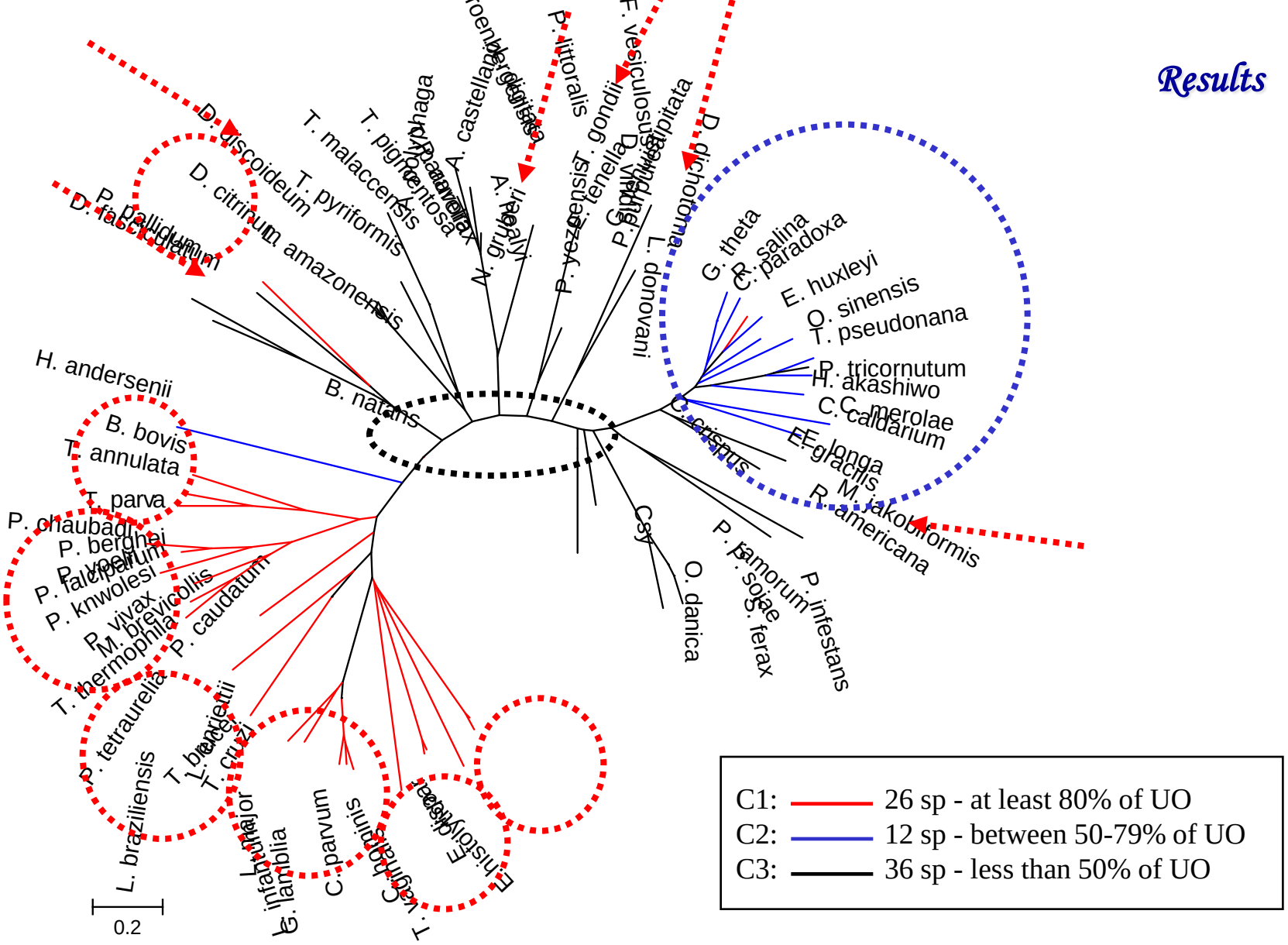


Figure 1. Phylogenomic supermatrix radiation tree of protozoan using masked sequences with TrimAl. Was used a supermatrix of 12,807 positions in the total 74 protozoan species. Maximum likelihood tree was constructed with Phyml, JTT as evolutionary model and bootstrap 100. (C1 is red, C2 blue and C3 black).

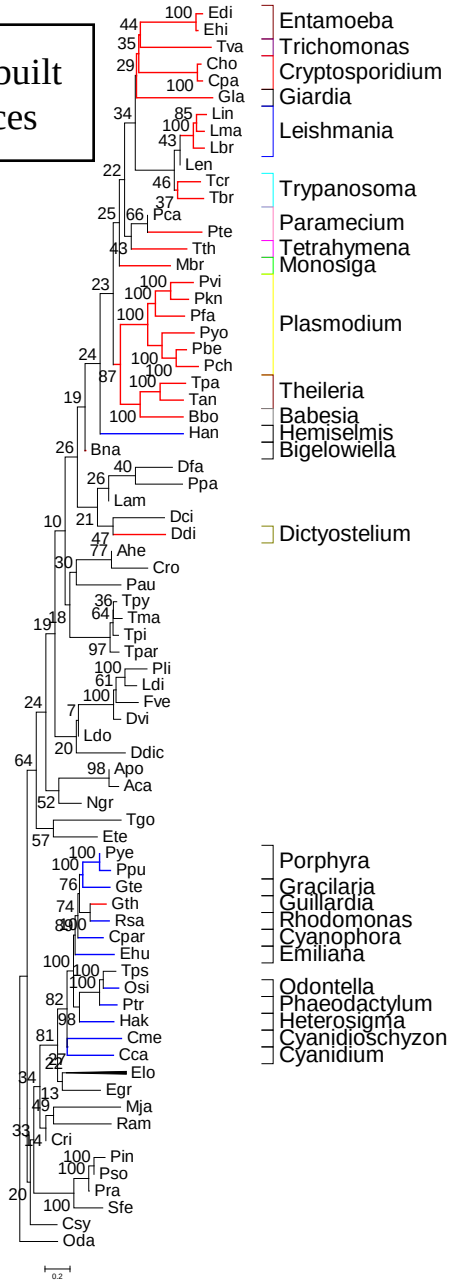
Results

- C1 was composed by only protozoan species, C2 was composed by rodophyta, cryptophyta, glaucocystophyceae, stramenopiles species, and C3 was composed by some species of C1 and C2.

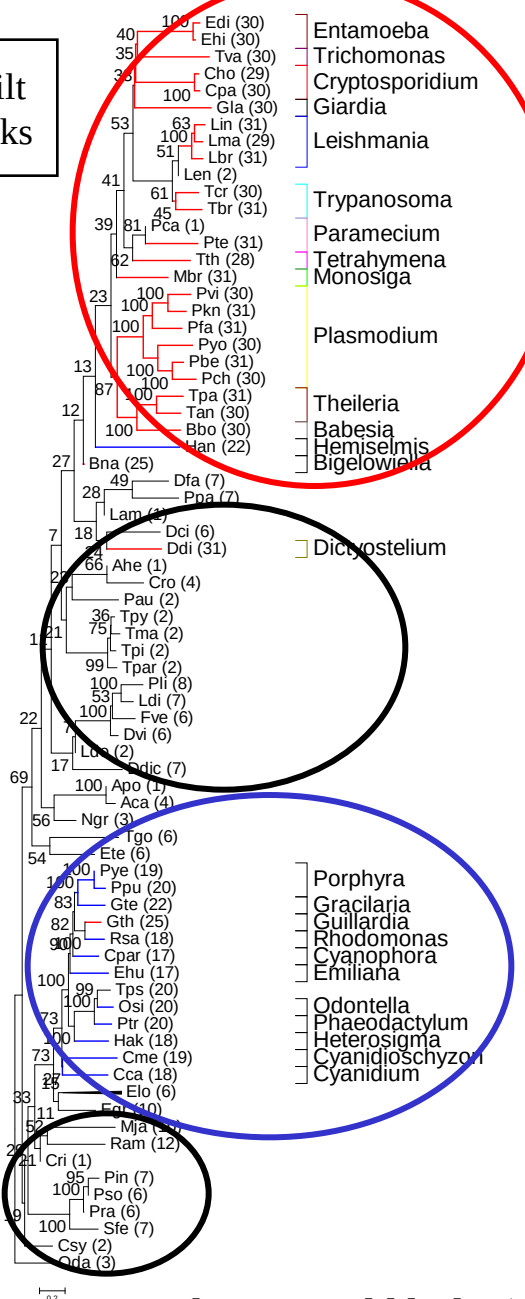
- C1 presented monophyly of Protozoa, C2 presented monophyly of species related to Protozoa, and C3 polytomy of some Protozoa and related species. The more UO included in the supermatrix the better was the tree robustness inside the 3 clades.

- Figure 2 shows the comparison of trees using no masked (M1) and TrimAl masked (M2) sequences. Both methodologies showed very similar topologies, however the tree utilizing TrimAl masked sequences showed higher bootstrap values. Figure 1 is the same tree using TrimAl masked sequences but presented as radiation tree.

Supermatrix tree built with total sequences



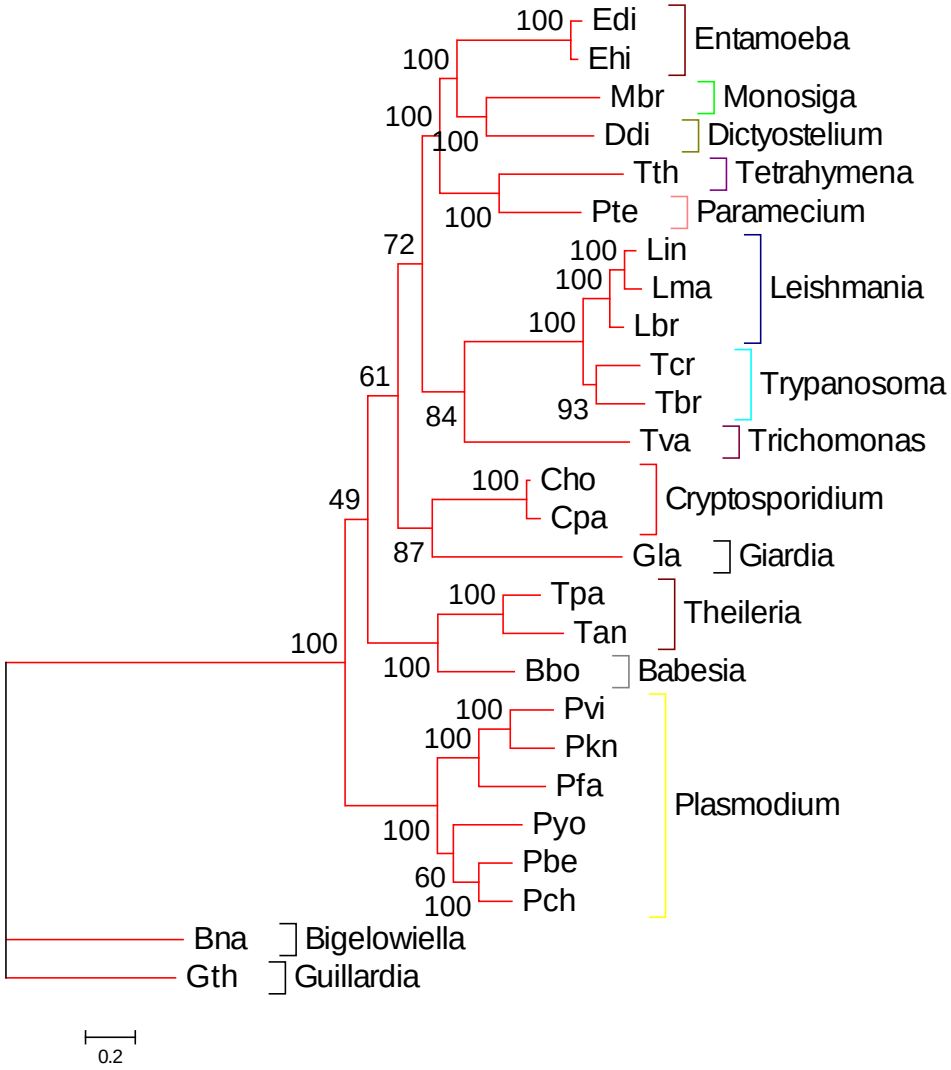
Supermatrix tree built with conserved blocks



Results

Figure 2. Phylogenomic supermatrix trees of protozoan using total sequences and conserved blocks (TrimAl) sequences. Were used supermatrixes of 12,807 and 21,260 positions respectively in 74 species. Maximum likelihood tree was constructed with Phyml, JTT as evolutionary model and bootstrap 100. (C1 is red, C2 blue and C3 black).

Most variable regions trimmed (>80 UO): Clade 1 (zoom)



Conclusion

- We have presented a phylogenomic-based overview for Protozoa. Relationships between protozoan groups are in agreement with previous studies, showing monophyly for protozoan.

- On the other hand, phylogenetic information inferred from C3 is questionable due to incomplete genomes, suggesting that using less than 15 universal UO for phylogenomic reconstruction is not reliable.

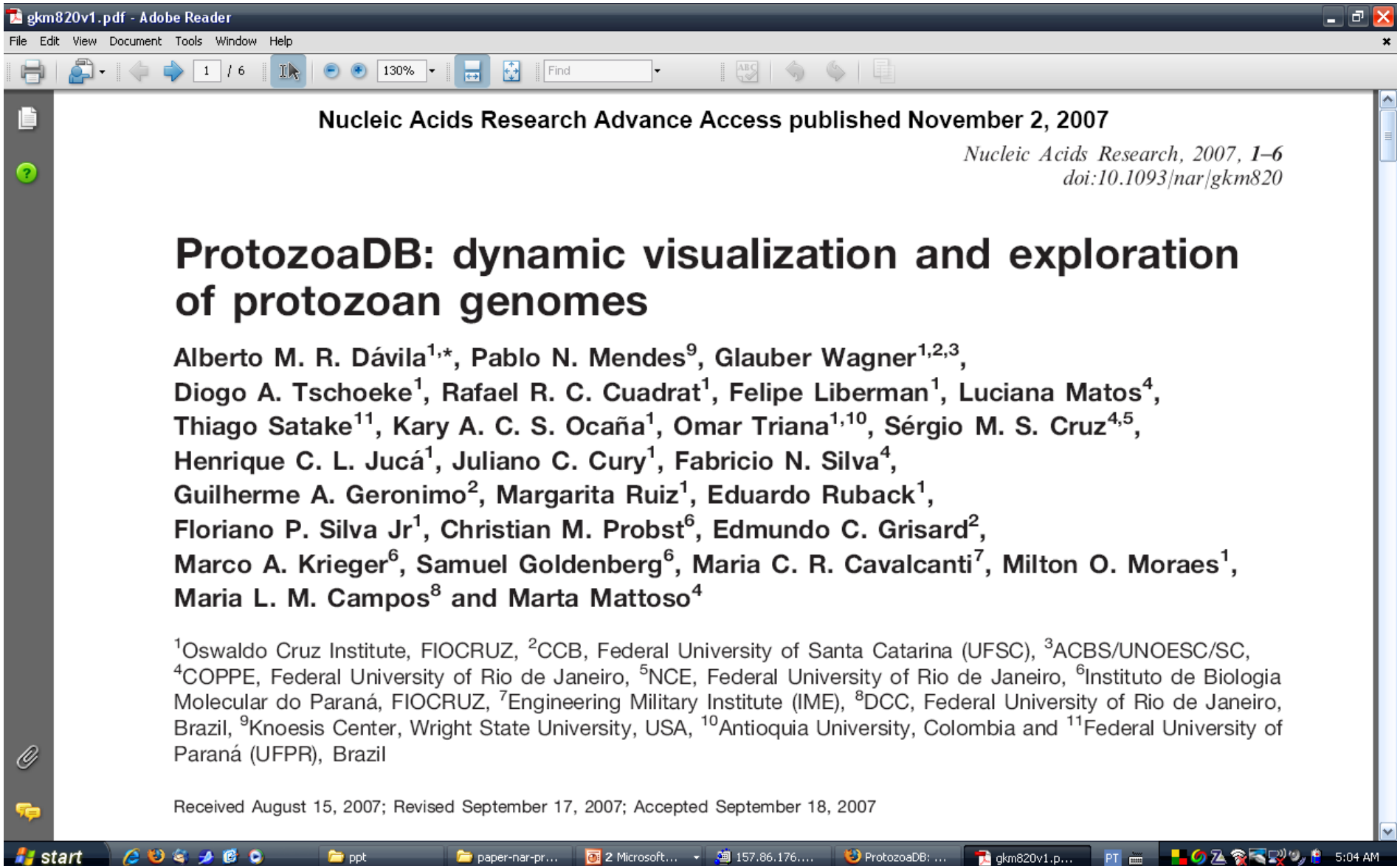
- The inclusion of more data/genes is necessary to obtain a robust tree.

- Our phylogenomic-based methodology using a supermatrix approach proved to be reliable with protozoan genome data, suggesting that (a) more universal UO used the better, and (b) using the entire UO sequence or just a conserved block of it produce similar reliable results.

- Finally, we need to investigate further if this methodology could be extrapolated or reproduced to other taxonomic groups.

ProtozoaDB

http://protozoadb.biowebdb.org



Nucleic Acids Research Advance Access published November 2, 2007

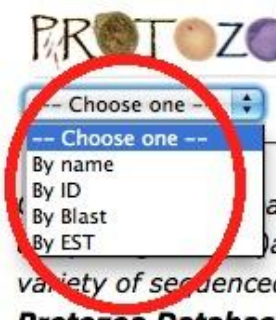
Nucleic Acids Research, 2007, 1–6
doi:10.1093/nar/gkm820

ProtozoaDB: dynamic visualization and exploration of protozoan genomes

Alberto M. R. Dávila[1,*], Pablo N. Mendes[9], Glauber Wagner[1,2,3], Diogo A. Tschoeke[1], Rafael R. C. Cuadrat[1], Felipe Liberman[1], Luciana Matos[4], Thiago Satake[11], Kary A. C. S. Ocaña[1], Omar Triana[1,10], Sérgio M. S. Cruz[4,5], Henrique C. L. Jucá[1], Juliano C. Cury[1], Fabricio N. Silva[4], Guilherme A. Geronimo[2], Margarita Ruiz[1], Eduardo Ruback[1], Floriano P. Silva Jr[1], Christian M. Probst[6], Edmundo C. Grisard[2], Marco A. Krieger[6], Samuel Goldenberg[6], Maria C. R. Cavalcanti[7], Milton O. Moraes[1], Maria L. M. Campos[8] and Marta Mattoso[4]

[1]Oswaldo Cruz Institute, FIOCRUZ, [2]CCB, Federal University of Santa Catarina (UFSC), [3]ACBS/UNOESC/SC, [4]COPPE, Federal University of Rio de Janeiro, [5]NCE, Federal University of Rio de Janeiro, [6]Instituto de Biologia Molecular do Paraná, FIOCRUZ, [7]Engineering Military Institute (IME), [8]DCC, Federal University of Rio de Janeiro, Brazil, [9]Knoesis Center, Wright State University, USA, [10]Antioquia University, Colombia and [11]Federal University of Paraná (UFPR), Brazil

Received August 15, 2007; Revised September 17, 2007; Accepted September 18, 2007



aches involving the integrative use of heterogeneous databases, analyses tools, distributed annotation systems as well as sensitive similarity detection algorithms have been catalyzed by a variety of sequenced genomes. In this purpose, the BiowebDB Consortium, partially funded by CNPq, present the ***Protozoa Database****, a integrated, user-friendly and flexible platform that contains the several protozoa genomes:* ***Trypanosoma cruzi****,* ***Trypanosoma brucei****,* ***Trypanosoma rangeli****,* ***Leishmania major****,* ***Plasmodium falciparum*** *and* ***Entamoeba histolytica****.*

Latest Version: 1.0 beta (August 2010)

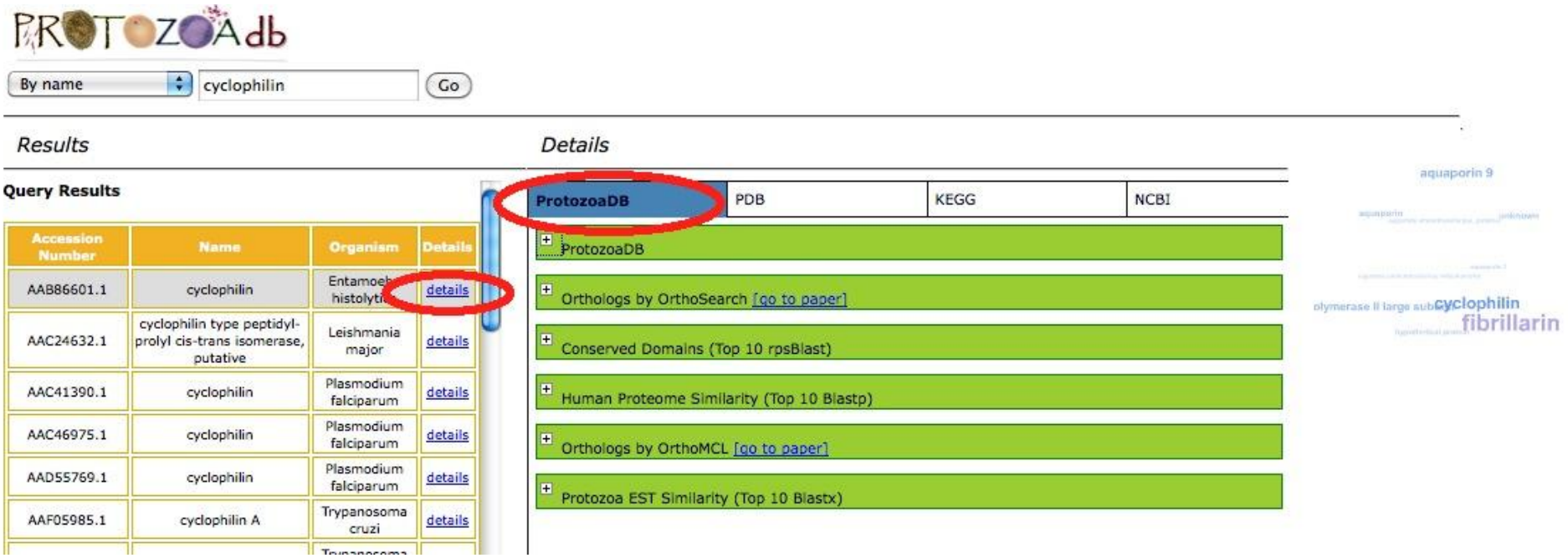
fibrillarin

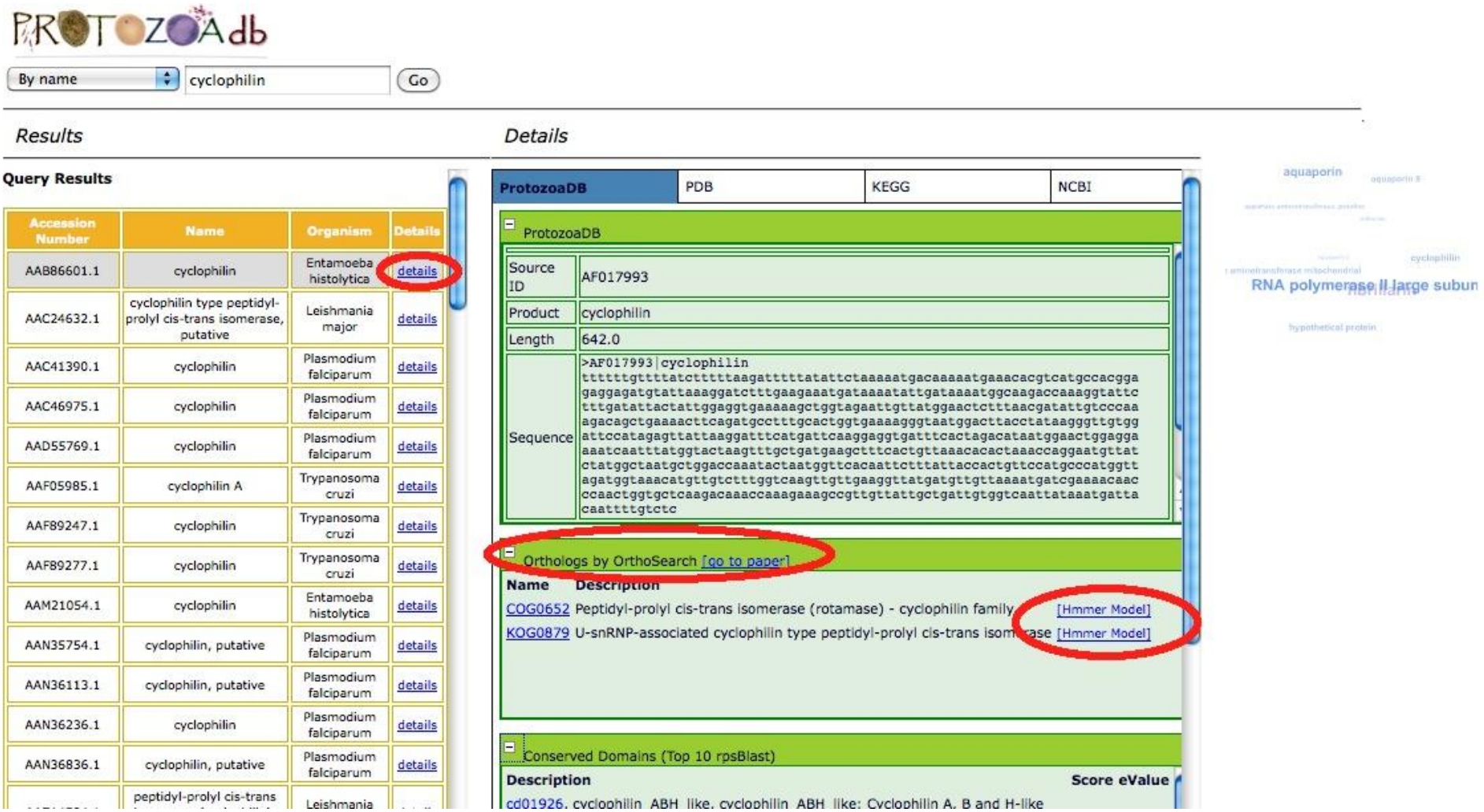
notransferase, putative

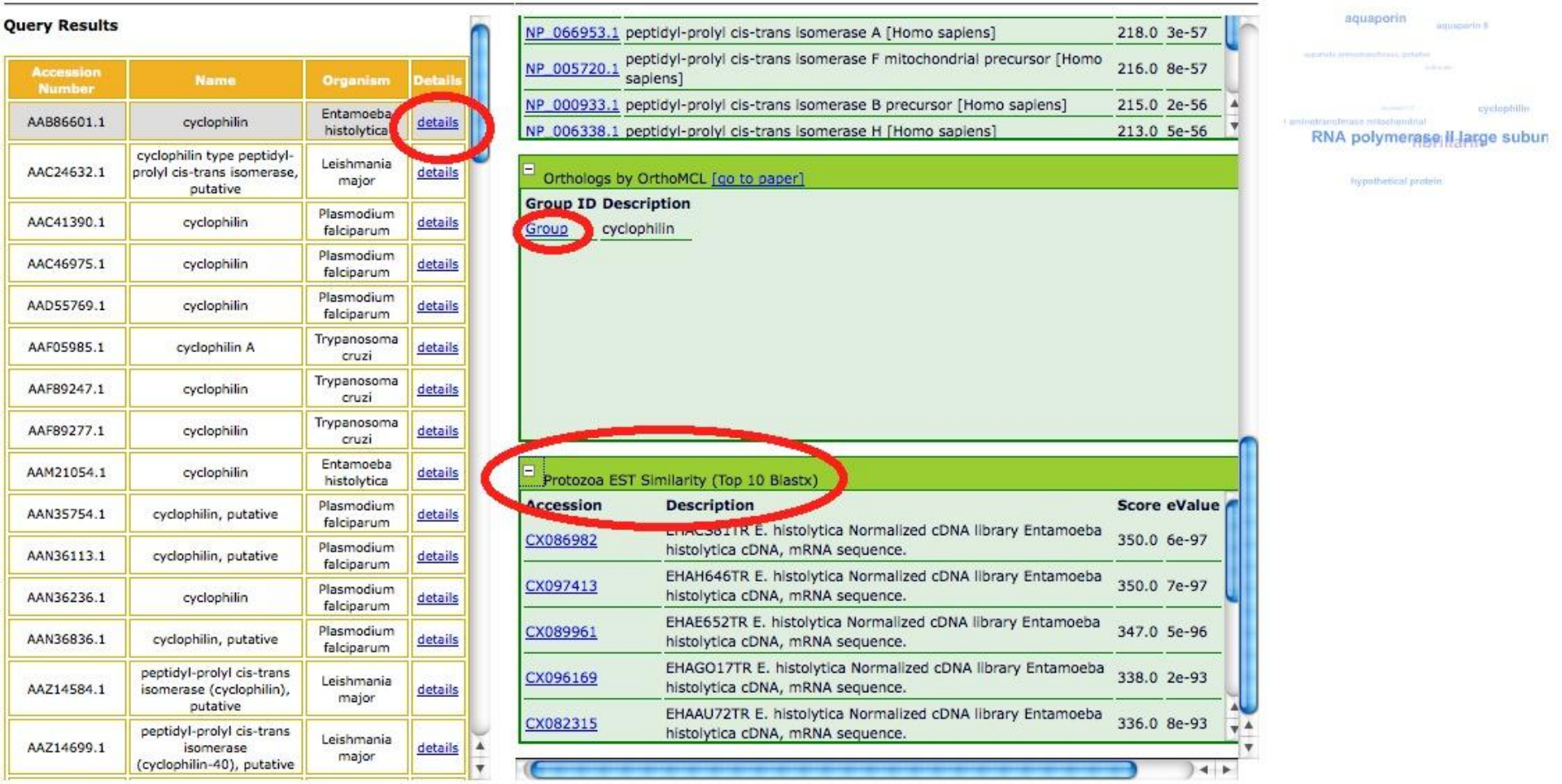
cyclophi

RNA polymerase II large su

aquaporin


PROTOZOAdb
By name
cyclophilin
Go
Results
Details
Query Results
Accession Number
Name
Organism
Details
AAB86601.1
cyclophilin
Entamoeb histolyti
details
AAC24632.1
cyclophilin type peptidyl-prolyl cis-trans isomerase, putative
Leishmania major
details
AAC41390.1
cyclophilin
Plasmodium falciparum
details
AAC46975.1
cyclophilin
Plasmodium falciparum
details
AAD55769.1
cyclophilin
Plasmodium falciparum
details
AAF05985.1
cyclophilin A
Trypanosoma cruzi
details
ProtozoaDB
PDB
KEGG
NCBI
ProtozoaDB
Orthologs by OrthoSearch [go to paper]
Conserved Domains (Top 10 rpsBlast)
Human Proteome Similarity (Top 10 Blastp)
Orthologs by OrthoMCL [go to paper]
Protozoa EST Similarity (Top 10 Blastx)
aquaporin 9
olymerase II large sub
cyclophilin
fibrillarin


PROTOZOAdb
By name
cyclophilin
Go
Results
Details
Query Results
Accession Number
Name
Organism
Details
AAB86601.1
cyclophilin
Entamoeba histolytica
details
AAC24632.1
cyclophilin type peptidyl-prolyl cis-trans isomerase, putative
Leishmania major
details
AAC41390.1
cyclophilin
Plasmodium falciparum
details
AAC46975.1
cyclophilin
Plasmodium falciparum
details
AAD55769.1
cyclophilin
Plasmodium falciparum
details
AAF05985.1
cyclophilin A
Trypanosoma cruzi
details
AAF89247.1
cyclophilin
Trypanosoma cruzi
details
AAF89277.1
cyclophilin
Trypanosoma cruzi
details
AAM21054.1
cyclophilin
Entamoeba histolytica
details
AAN35754.1
cyclophilin, putative
Plasmodium falciparum
details
AAN36113.1
cyclophilin, putative
Plasmodium falciparum
details
AAN36236.1
cyclophilin
Plasmodium falciparum
details
AAN36836.1
cyclophilin, putative
Plasmodium falciparum
details
peptidyl-prolyl cis-trans
Leishmania
ProtozoaDB
PDB
KEGG
NCBI
ProtozoaDB
Source ID
AF017993
Product
cyclophilin
Length
642.0
Sequence
>AF017993|cyclophilin
ttttttgttttatctttttaagattttatattctaaaaatgacaaaaatgaaacacgtcatgccacgga
gaggagatgtattaaaggatctttgaagaaatgataaaatattgataaaatggcaagaccaaaggtattc
tttgatattactattggaggtgaaaaagctggtagaattgttatggaactctttaacgatattgtcccaa
agacagctgaaaacttcagatgcctttgcactggtgaaaagggtaatggacttacctataagggttgtgg
attccatagagttattaaggatttcatgattcaaggaggtgatttcactagacataatggaactggagga
aaatcaatttatggtactaagtttgctgatgaagctttcactgttaaacacactaaaccaggaatgttat
ctatggctaatgctggaccaaatactaatggttcacaattctttattaccactgttccatgcccatggtt
agatggtaaacatgttgtctttggtcaagttgttgaaggttatgatgttgttaaaatgatcgaaaacaac
ccaactggtgctcaagacaaaccaaagaaagccgttgttattgctgattgtggtcaattataaatgatta
caattttgtctc
Orthologs by OrthoSearch [go to paper]
Name
Description
COG0652 Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family [Hmmer Model]
KOG0879 U-snRNP-associated cyclophilin type peptidyl-prolyl cis-trans isom ase [Hmmer Model]
Conserved Domains (Top 10 rpsBlast)
Description
Score eValue
cd01926, cyclophilin_ABH_like, cyclophilin_ABH_like: Cyclophilin A, B and H-like
aquaporin
cyclophilin
RNA polymerase II large subun
hypothetical protein



Query Results

Accession Number	Name	Organism	Details
AAB86601.1	cyclophilin	Entamoeba histolytica	details
AAC24632.1	cyclophilin type peptidyl-prolyl cis-trans isomerase, putative	Leishmania major	details
AAC41390.1	cyclophilin	Plasmodium falciparum	details
AAC46975.1	cyclophilin	Plasmodium falciparum	details
AAD55769.1	cyclophilin	Plasmodium falciparum	details
AAF05985.1	cyclophilin A	Trypanosoma cruzi	details
AAF89247.1	cyclophilin	Trypanosoma cruzi	details
AAF89277.1	cyclophilin	Trypanosoma cruzi	details
AAM21054.1	cyclophilin	Entamoeba histolytica	details
AAN35754.1	cyclophilin, putative	Plasmodium falciparum	details
AAN36113.1	cyclophilin, putative	Plasmodium falciparum	details
AAN36236.1	cyclophilin	Plasmodium falciparum	details
AAN36836.1	cyclophilin, putative	Plasmodium falciparum	details
AAZ14584.1	peptidyl-prolyl cis-trans isomerase (cyclophilin), putative	Leishmania major	details
AAZ14699.1	peptidyl-prolyl cis-trans isomerase (cyclophilin-40), putative	Leishmania major	details

NP_066953.1	peptidyl-prolyl cis-trans isomerase A [Homo sapiens]	218.0	3e-57
NP_005720.1	peptidyl-prolyl cis-trans isomerase F mitochondrial precursor [Homo sapiens]	216.0	8e-57
NP_000933.1	peptidyl-prolyl cis-trans isomerase B precursor [Homo sapiens]	215.0	2e-56
NP_006338.1	peptidyl-prolyl cis-trans isomerase H [Homo sapiens]	213.0	5e-56

Orthologs by OrthoMCL [go to paper]

Group ID	Description
Group	cyclophilin

Protozoa EST Similarity (Top 10 Blastx)

Accession	Description	Score	eValue
CX086982	EHAC381TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	350.0	6e-97
CX097413	EHAH646TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	350.0	7e-97
CX089961	EHAE652TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	347.0	5e-96
CX096169	EHAGO17TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	338.0	2e-93
CX082315	EHAAU72TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	336.0	8e-93

Dávila A.M. et al. ProtozoaDB: dynamic visualization and exploration of protozoan genomes. Nucleic Acids Research, 10:1-6, November 2007.
BiowebDB Consortium

Biowebdb Consortium - Mozilla Firefox

Arquivo Editar Exibir Histórico Favoritos Ferramentas Ajuda

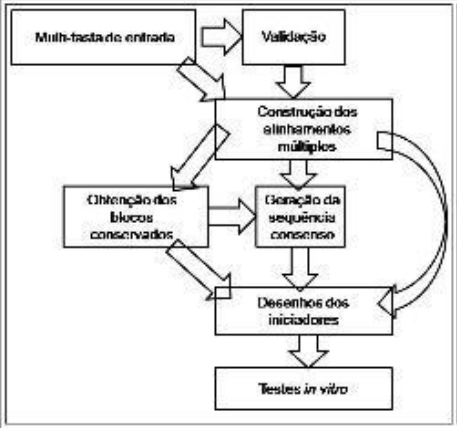
http://kineto2.biowebdb.org/genotype_system/

Mais visitados Guia rápido Últimas notícias orkut - home BBC Brasil - Primeira P... Le Monde diplomatiq... Folha Online - Mundo Álbuns da

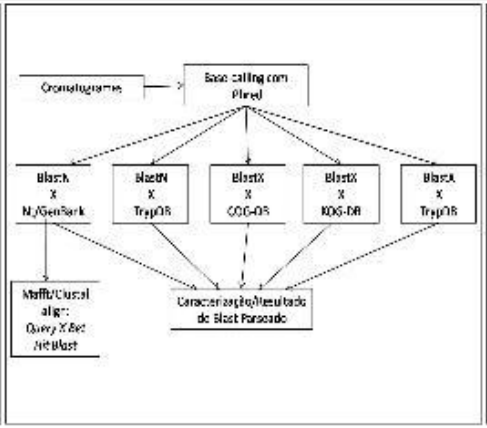
Biowebdb Consortium

Integrated System for genotipyng Parasitic Protozoa based on Universal ortologous genes

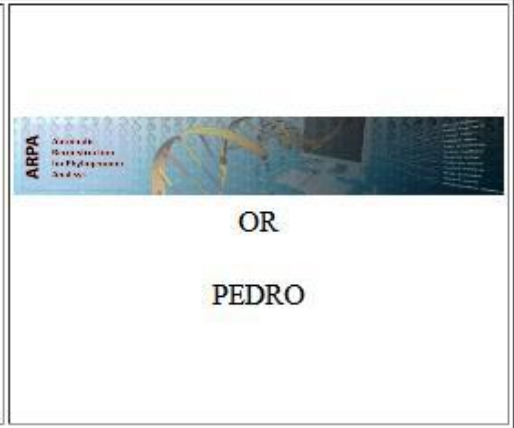
Desing primer for genotypyng Parasitic Protozoa Based on Universal ortologous genes → Base Calling chromato files and caracterize sequences using Blast → Do Phylogenetic Analyses using: Arpa or Evolutionary Analyses using: Pedro

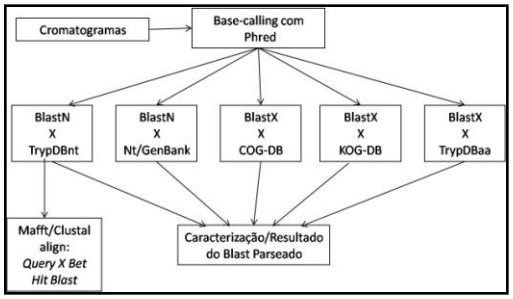


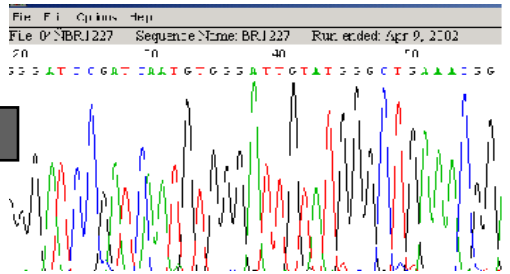
→

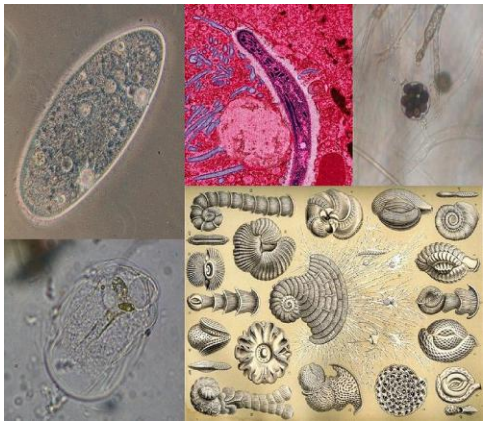


→









Methodology

DNA extraction

PCR using 33 pair of designed primers

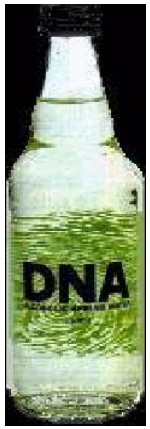
Agarose gel visualization

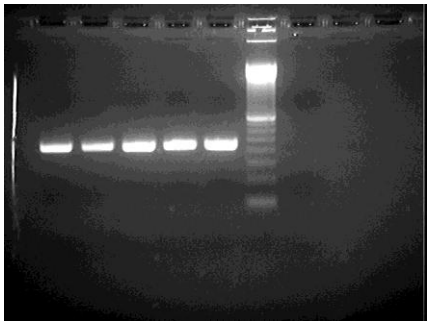
PCR product purification

Sequencing

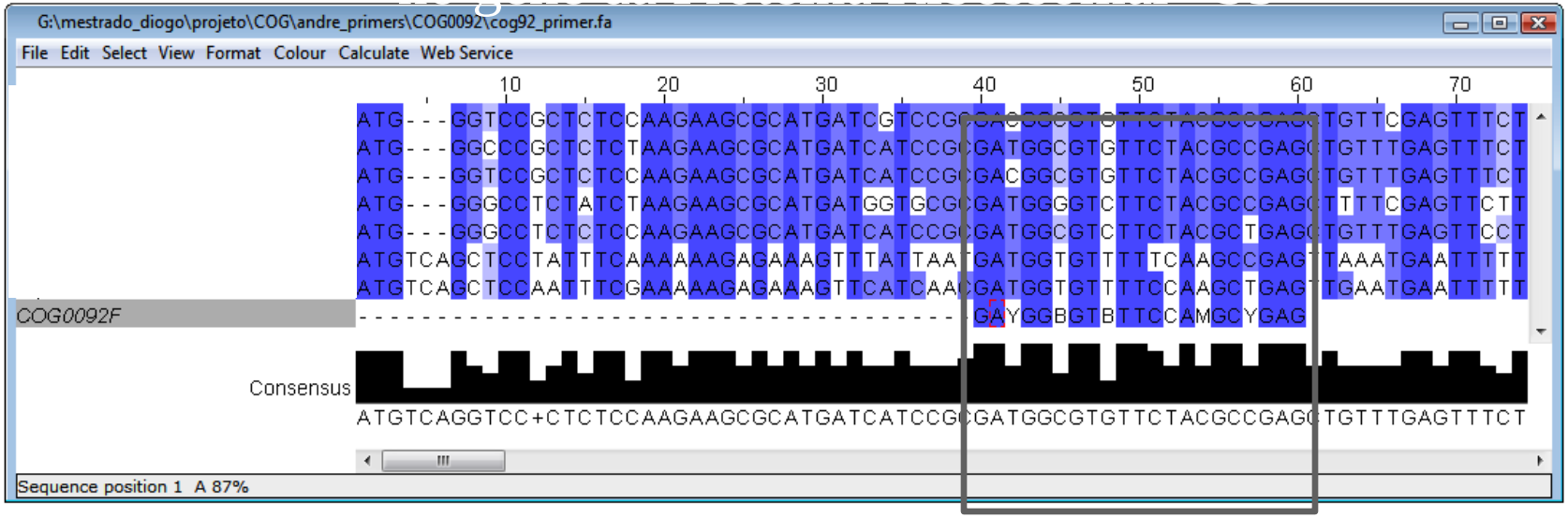
Annotation

Phylogenetic analyses

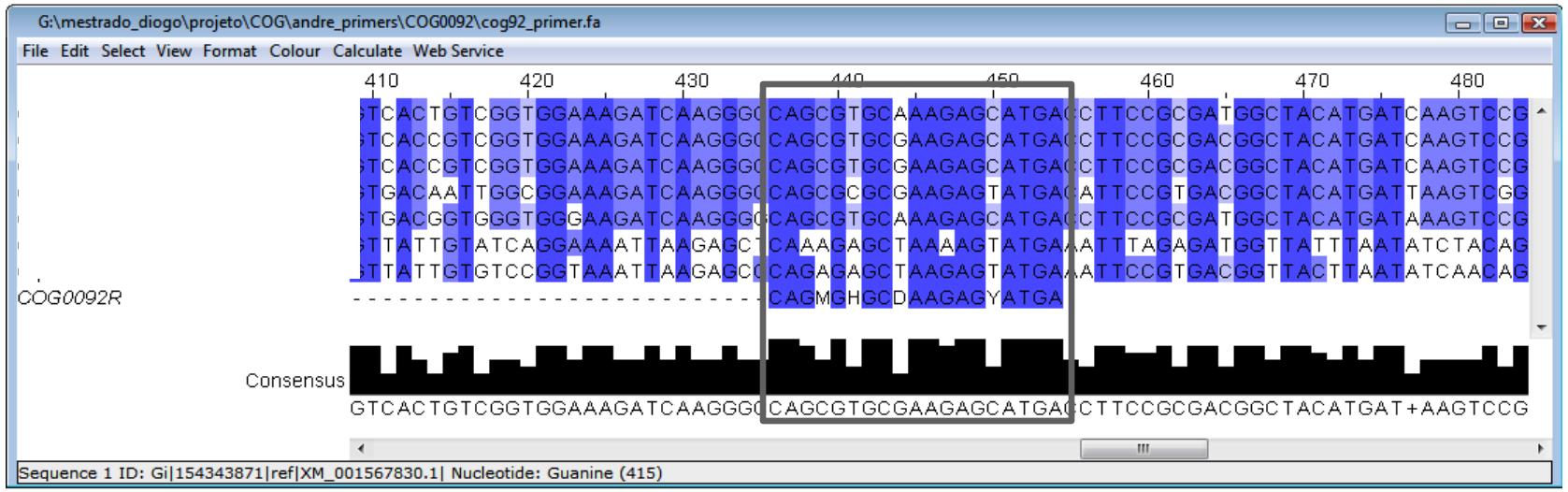


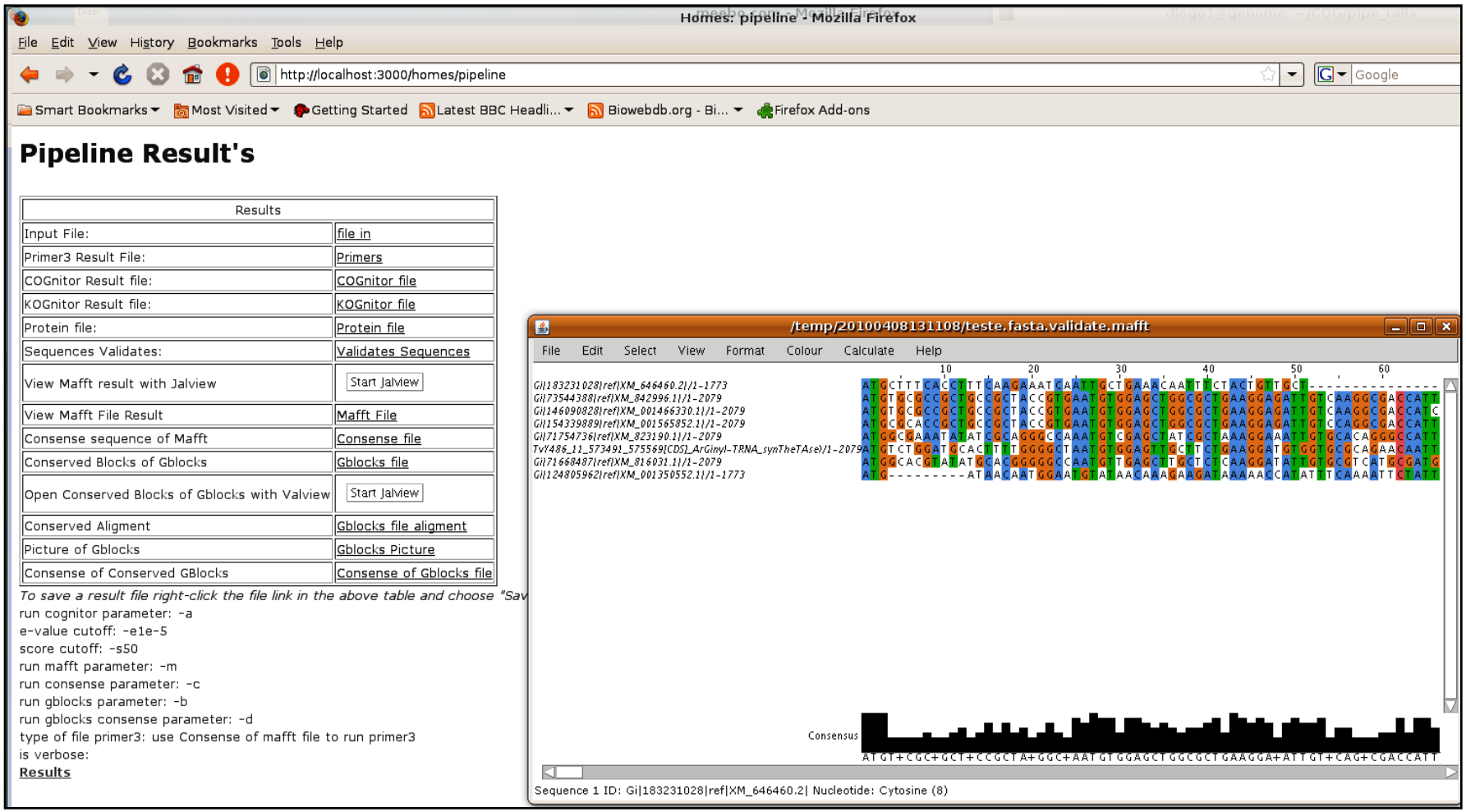


Forward primer (COG0092F) in the multiple alignment

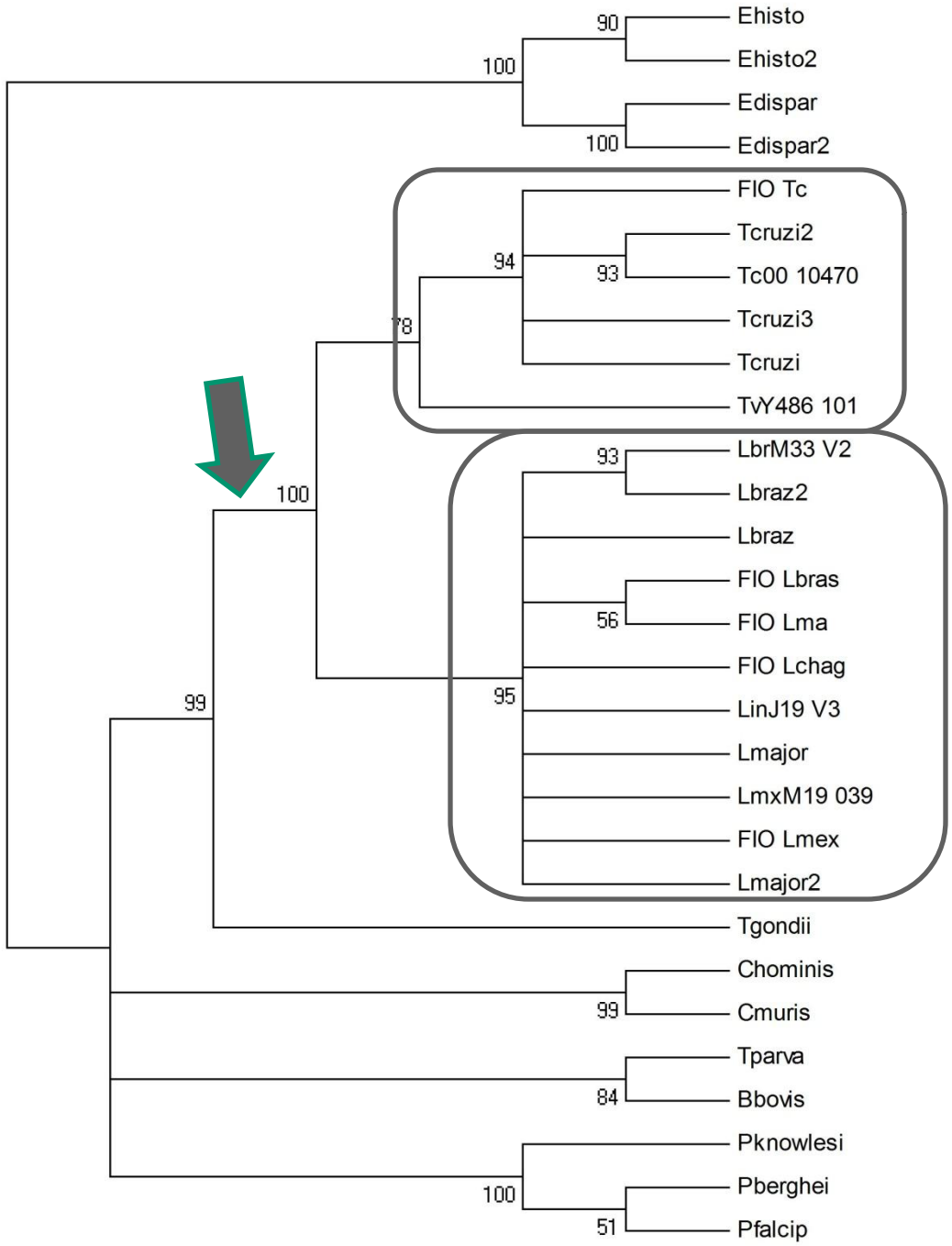


Reverse primer (COG0092F) in the multiple alignment



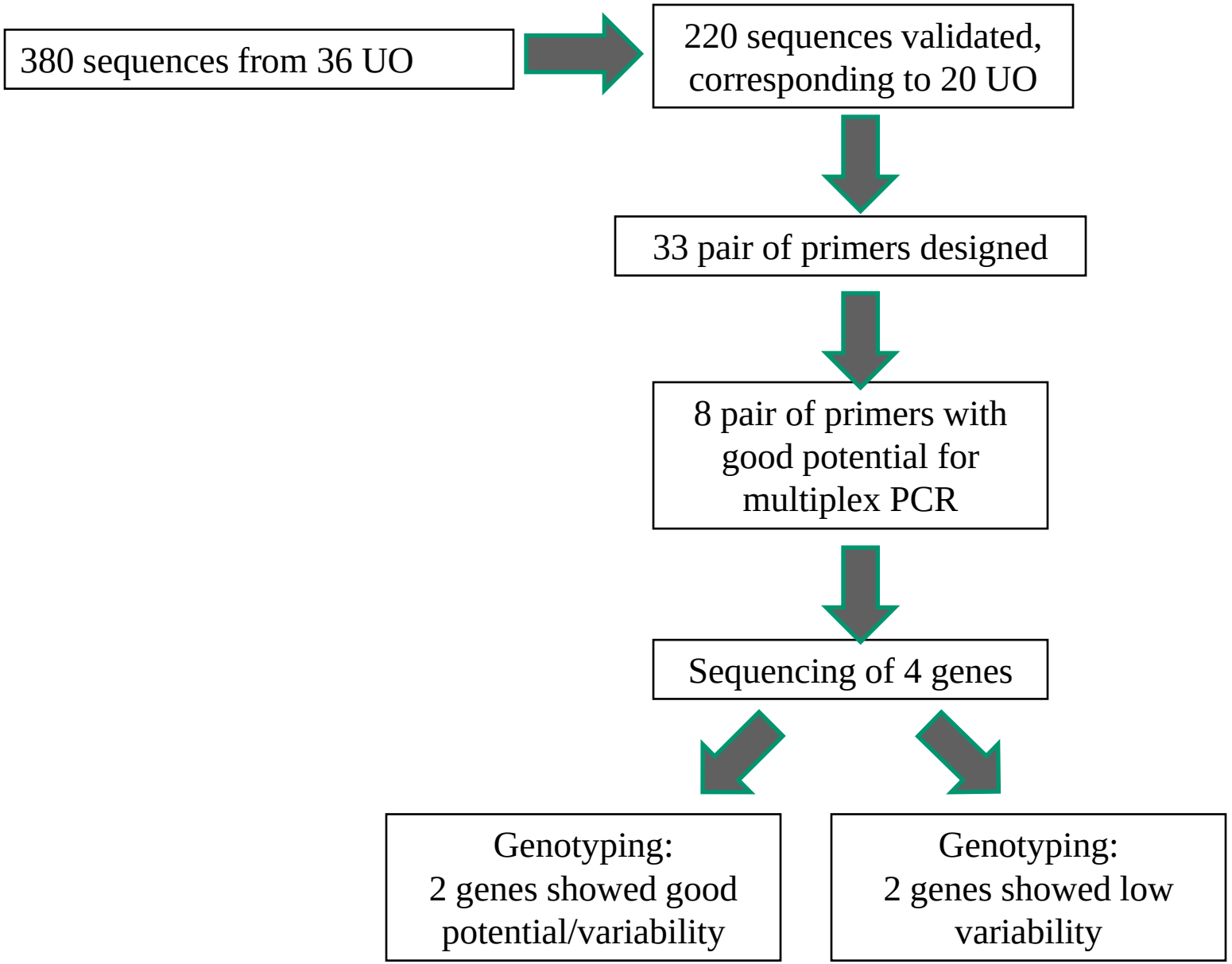


Genotyping system visualization (multiple alignment) using Jalview

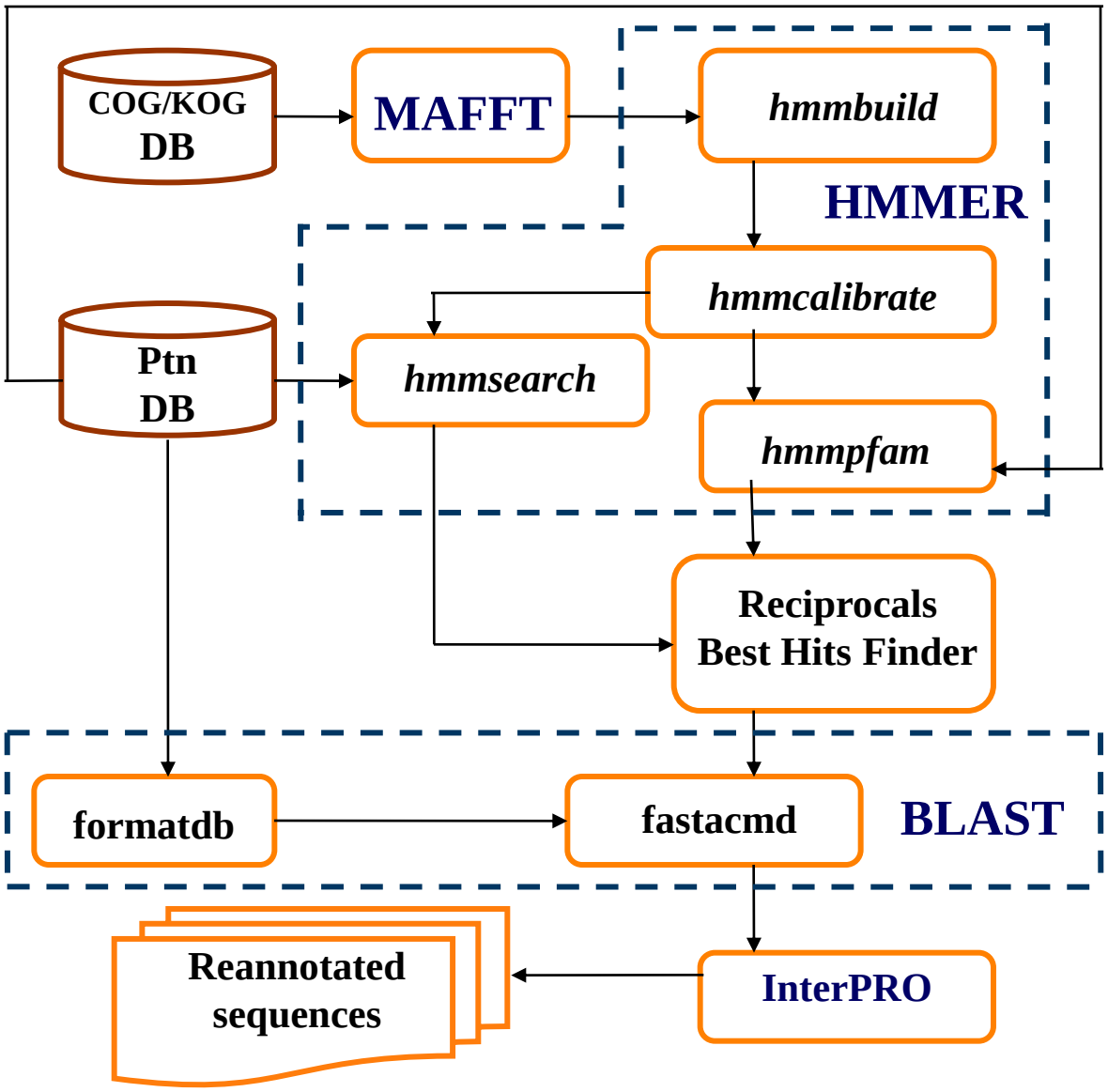


Phylogenetic tree inferred with Maximum Likelihood for KOG0400 (Ribosomal protein S15P/S13E). 1000 replications (*bootstrap)*. Sequences from genbank and generated with our primers are shown.

Methodology: short version



OrthoSearch Workflow (Orthologs inference)



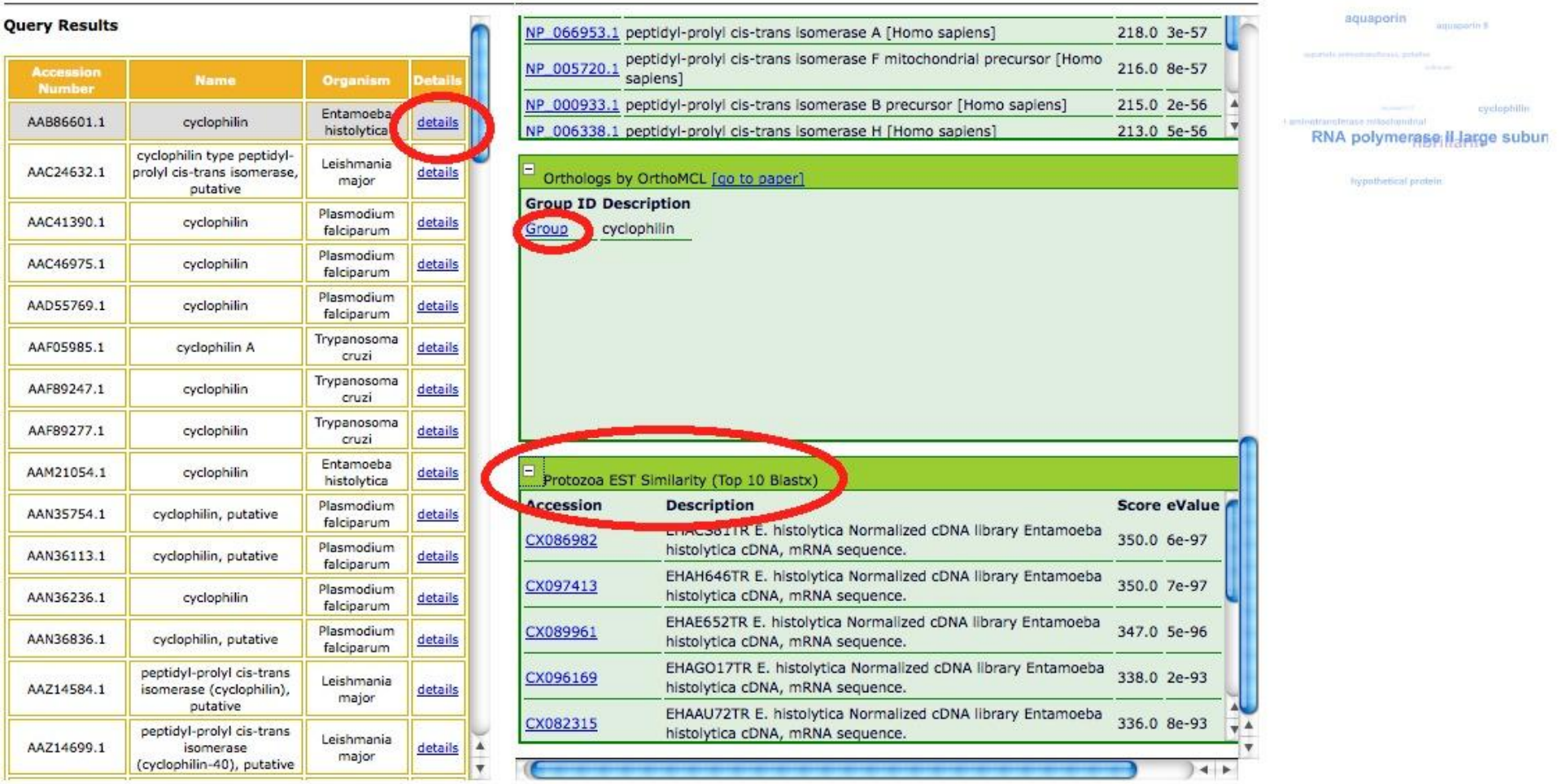
Conceptual Pipeline

OrthoSearch results

Parasite	Total Hits	Confirmed annotation (Interpro)	Partial annotations (re)annotated	Hypothetical protein (re)annotated
T. cruzi	147	141	6	12
T. brucei	135	129	6	11
L. major	165	157	9	14
P. falciparum	111	107	6	12
E. histolytica	98	90	5	6

Protozoan Orthologs (OrthoMCL-based)

* Some can be protozoa-specific



Query Results

Accession Number	Name	Organism	Details
AAB86601.1	cyclophilin	Entamoeba histolytica	details
AAC24632.1	cyclophilin type peptidyl-prolyl cis-trans isomerase, putative	Leishmania major	details
AAC41390.1	cyclophilin	Plasmodium falciparum	details
AAC46975.1	cyclophilin	Plasmodium falciparum	details
AAD55769.1	cyclophilin	Plasmodium falciparum	details
AAF05985.1	cyclophilin A	Trypanosoma cruzi	details
AAF89247.1	cyclophilin	Trypanosoma cruzi	details
AAF89277.1	cyclophilin	Trypanosoma cruzi	details
AAM21054.1	cyclophilin	Entamoeba histolytica	details
AAN35754.1	cyclophilin, putative	Plasmodium falciparum	details
AAN36113.1	cyclophilin, putative	Plasmodium falciparum	details
AAN36236.1	cyclophilin	Plasmodium falciparum	details
AAN36836.1	cyclophilin, putative	Plasmodium falciparum	details
AAZ14584.1	peptidyl-prolyl cis-trans isomerase (cyclophilin), putative	Leishmania major	details
AAZ14699.1	peptidyl-prolyl cis-trans isomerase (cyclophilin-40), putative	Leishmania major	details

NP_066953.1	peptidyl-prolyl cis-trans isomerase A [Homo sapiens]	218.0	3e-57
NP_005720.1	peptidyl-prolyl cis-trans isomerase F mitochondrial precursor [Homo sapiens]	216.0	8e-57
NP_000933.1	peptidyl-prolyl cis-trans isomerase B precursor [Homo sapiens]	215.0	2e-56
NP_006338.1	peptidyl-prolyl cis-trans isomerase H [Homo sapiens]	213.0	5e-56

Orthologs by OrthoMCL [go to paper]

Group ID	Description
Group	cyclophilin

Protozoa EST Similarity (Top 10 Blastx)

Accession	Description	Score	eValue
CX086982	EHAC381TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	350.0	6e-97
CX097413	EHAH646TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	350.0	7e-97
CX089961	EHAE652TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	347.0	5e-96
CX096169	EHAGO17TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	338.0	2e-93
CX082315	EHAAU72TR E. histolytica Normalized cDNA library Entamoeba histolytica cDNA, mRNA sequence.	336.0	8e-93

Dávila A.M. et al. ProtozoaDB: dynamic visualization and exploration of protozoan genomes. Nucleic Acids Research, 10:1-6, November 2007.
BiowebDB Consortium

PRIMER19:

KOG1487 (GTP-binding protein DRG1 - ODN superfamily)

COG1163 (Predicted GTPase)

PRIMER20:

KOG0217 (Mismatch repair ATPase MSH6 - MutS family)

COG0249 (MutS-like ATPases involved in mismatch repair, family 2)

PRIMER23:

KOG0291 (WD40-repeat-containing subunit of the 18S rRNA processing complex)

COG2319 (WD40 repeat protein)

PRIMER143:

COG0143 (Methionyl tRNA Synthetase)

PRIMER172:

COG0172 (Methionyl tRNA Synthetase)

	PRIMER 19	PRIMER 20	PRIMER 23	PRIMER 143	PRIMER 172
L. mexicana	X	X	X	X	-
L. brasiliensis	X	X	X	X	X
L. major	X	X	X	X	X
L. chagasi	X	X	X	X	X
L. amazonensis	-	X	X	X	X
L. guyanensis	-	X	X	X	X
T. cruzi	X	X	X	X	X
T. brucei	X	X	X	-	-
T. evansi	-	X	X	X	-
T. vivax	-	X	X	X	X
T. congolense	-	X	-	X	-

Home Recent Results

Choose program to use and database to search:

Program blastn Database refseq_genomic

Enter sequence below in FASTA format

Or load it from disk Choose File no file selected

Optional parameters

Threshold	Expect	Matrix	Filter	Descriptions	Alignments	Hits	Gapped
0	10	BLOSUM-62	default	500	250	0	yes

Alignment View Option

pairwise

Search type : DNA sequences

Enter with the sequence in the format FASTA :

Or select multifasta file : Choose File no file selected

nucleotideo

PHYML : Yes Not

Evolutionary model : HKY

Bootstrap [10 - 1000]: 100

Substitution Rate Categories : 4

Welcome to STINGRAY

Version 2.0 Beta - Feb 17, 2009

Login / Email :
guest
Password:
•••••
Enter

About | **Trial** | **Projects Avaliable** | **Pipeline** | **Team** | **FAQ** | **Contact** | **Publications**

The **System for Integrate Genomic Resources and Analyses (STINGRAY)** is a user-friendly web-based system designed to analyse genomic data in the context of a pipeline. It was developed for Linux systems using Perl, Bioperl, CGI, Apache and MySQL. EST and GSS data can especially benefit using the system since it can accept a) chromatograms, b) download of sequences from GenBank, c) FASTA files stored locally or d) a combinations of all 3 STINGRAY uses the phred/phrap package to process chromatograms, evaluate quality of traces and remove vector contamination. Clusterization is done using the CAP3 program. Sequences are submitted to Blast analyses with database choosen by the user. Conserved domain searches are also performed using the CDD tool of NCBI. Interpro have been implemented in the pipeline as well, and tRNA-Scan, Psort, SignalP and OrthoMCL are avaliable too. Beside that, it is possible to find genes or ORFs using GLIMMER, GlimmerHMM and by Orthologous Groups aproach (ORF by COGs), and then search the analysis mentionated above. Once all the above similarities searches have been performed, the user can select some contigs and results for multiple alignment using ClustalW and phylogeny analyses using Phylip. STINGRAY offers flexibility to the users by allowing them to configure the parameters of the used programs. All the results can be filtered according to algorithm and database used, facilitating user visualization. Furthermore, statistics and graphical results of the all analysis and processed chromatograms are presented.

Better performance in Firefox

Collaborators

Marta Mattoso
M. Luiza Machado Campos
Edmundo C. Grisard
“Yoko” Cavalcanti

Students

Sérgio Serra
Kary Soriano Ocaña
Fabio Coutinho
Edgar Sarmiento
Diogo Tschoeke
Rafael Cuadrat
Daniel Loureiro
Rodrigo Jardim
Luisa Pitaluga

Funding

CNPq
FINEP
CAPES
FAPERJ
European Union (EELA-2)

Recent collaborators:

Elisa Cupolillo
Claudia D. Levy