



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO



1920 | 2020



MITOFREE: AN AUTOMATED PIPELINE FOR MITOGENOME ASSEMBLY, ANNOTATION AND PHYLOGENOMIC ANALYSIS USING PUBLIC DATA

Gabriel Alves Vieira & Francisco Prosdocimi

Laboratório de Genômica e Biodiversidade
Instituto de Bioquímica Médica Leopoldo de Meis
Universidade Federal do Rio de Janeiro

Summary

- NGS and public data
- Sequence Read Archive (SRA)
- Public data and mitogenomics
- MitoFree



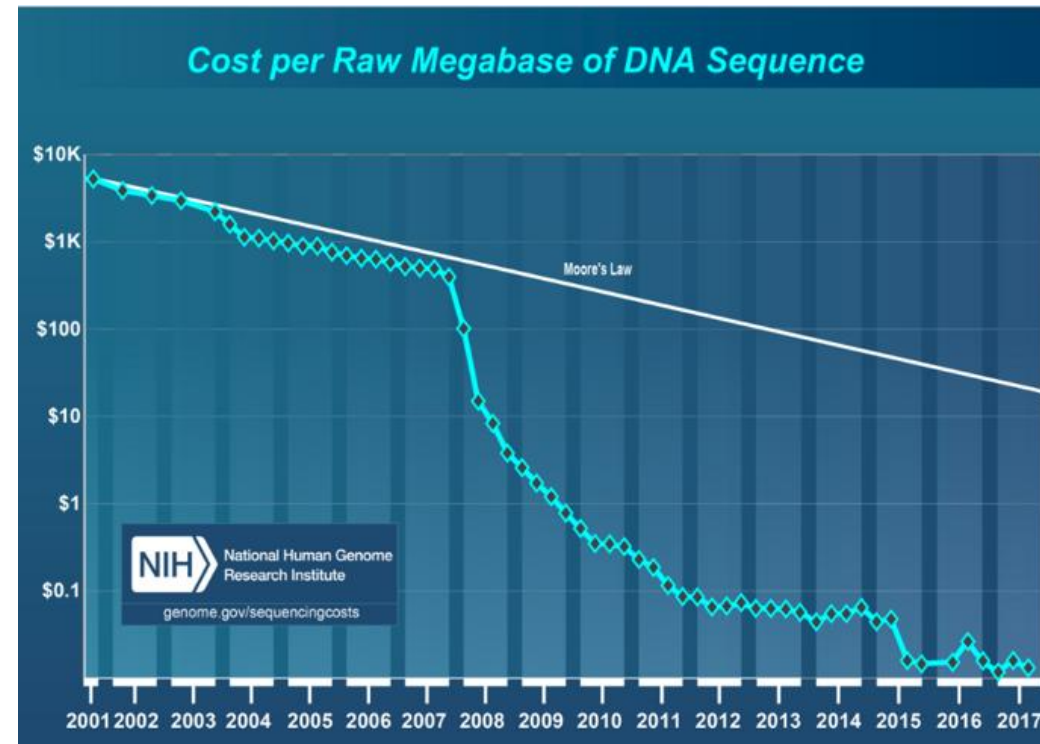
NGS and public data

- Next-Generation Sequencing (NGS)
 - ↑ genomic data produced
 - ↓ cost per sequenced base

Journals: Incentivates raw data publication

Reproducibility

- Growth of public sequencing databases



Sequence Read Archive (SRA)

- Part of an international collaboration (INSDC)

3 databases:

SRA: NCBI Sequence Read Archive

ERA: EBI Sequence Read Archive

DRA: DDBJ Sequence Read Archive

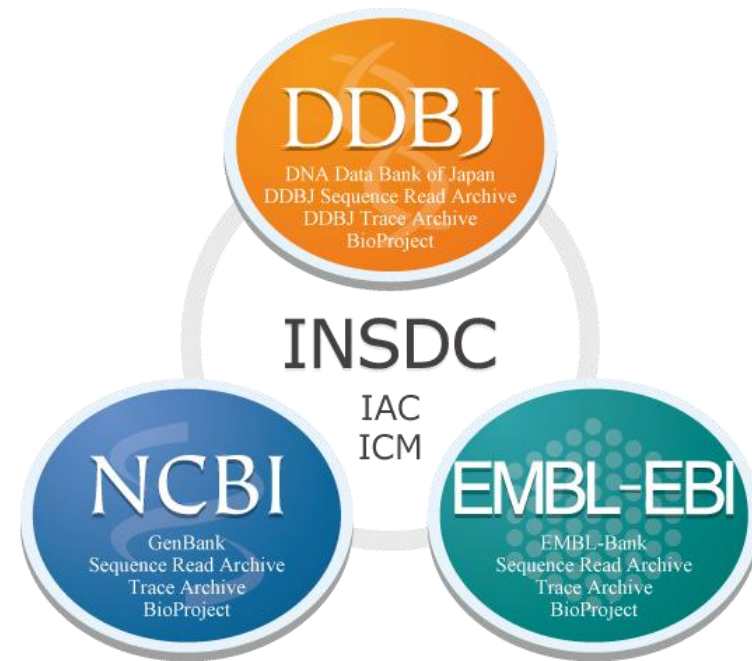
Synchronized

- Several types of sequencing data:

DNA-seq; Bisulfite-seq; RNA-seq...

Enormous potential for “recycling”

Scientific advances at low costs

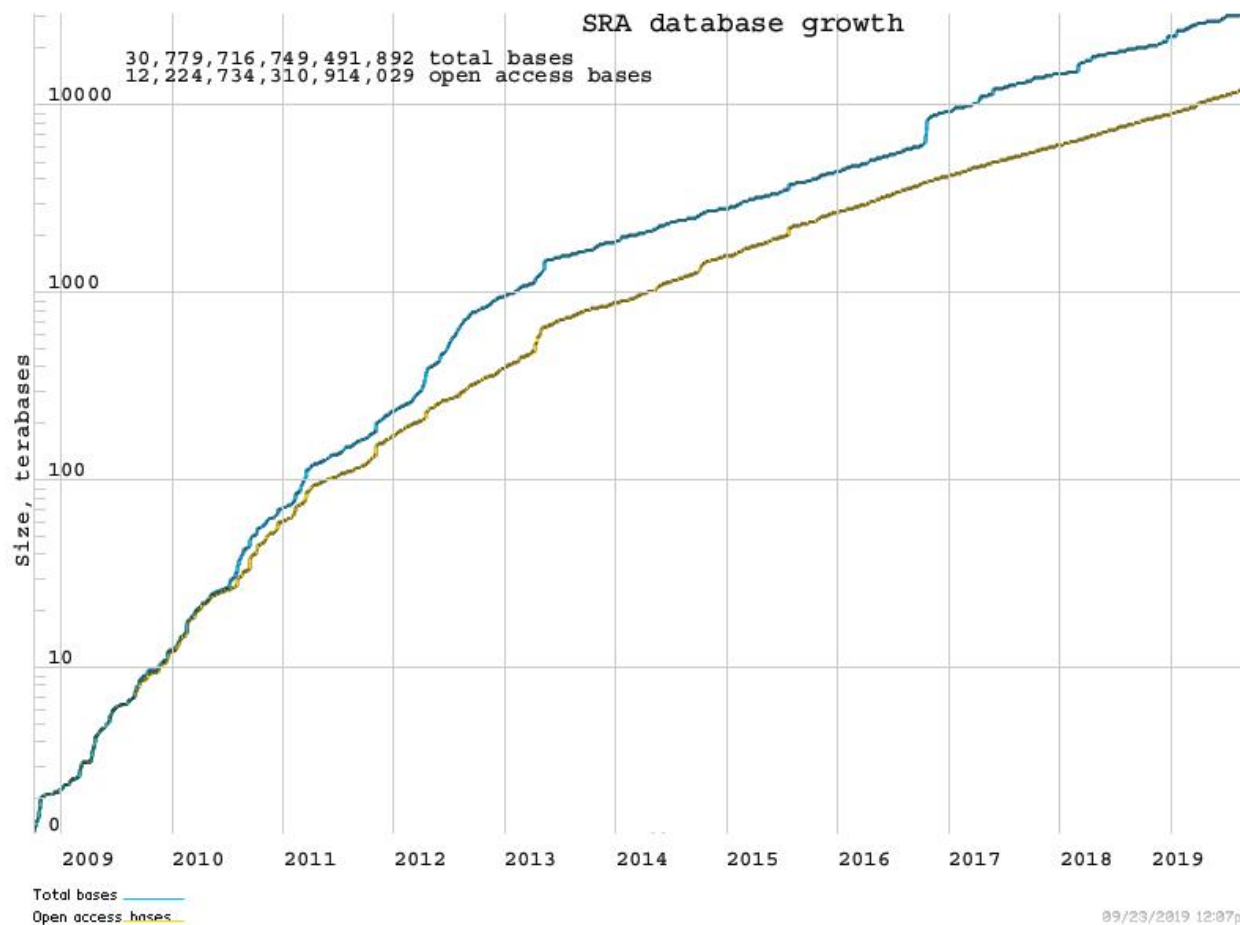


NEWS • 19 AUGUST 2019

Brazil's budget cuts threaten more than 80,000 science scholarships

If the country's main science-funding agency doesn't get more cash soon, young researchers will stop getting paid.

Rodrigo de Oliveira Andrade

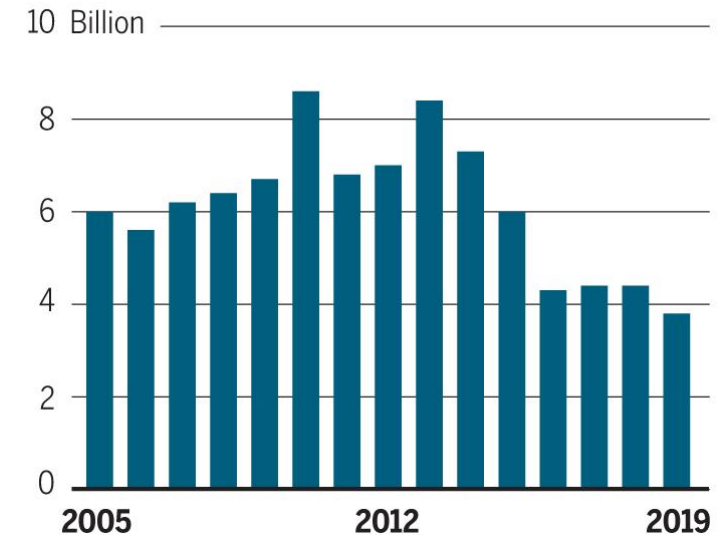


09/23/2019 12:07pm

Science in decline

Even without the freeze, the Brazilian science ministry's budget for R&D—adjusted for inflation—has declined sharply the past few years.

Budget in Brazilian reais



Brazil: ↓ Scientific funding

↑ Public sequencing data

How can these datasets foster new research?

Public data and mitogenomics

☞ Data used to answer a specific question

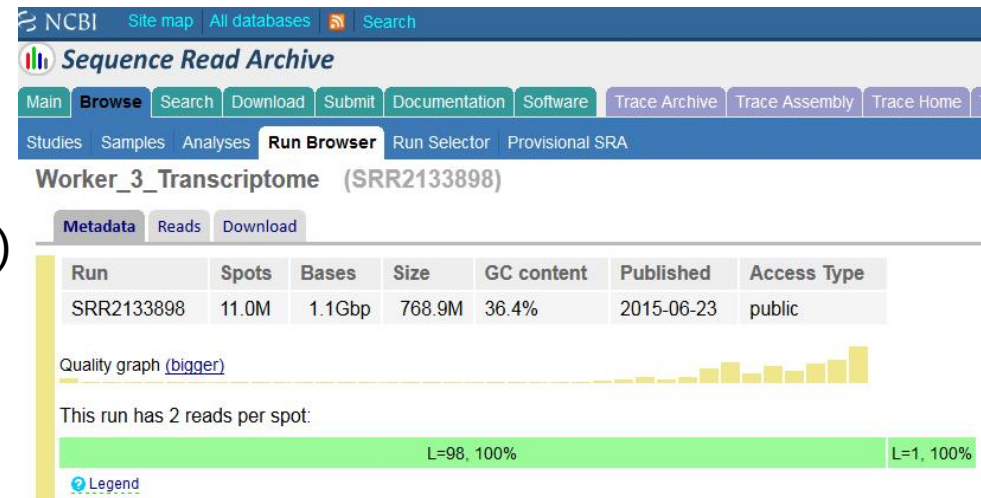
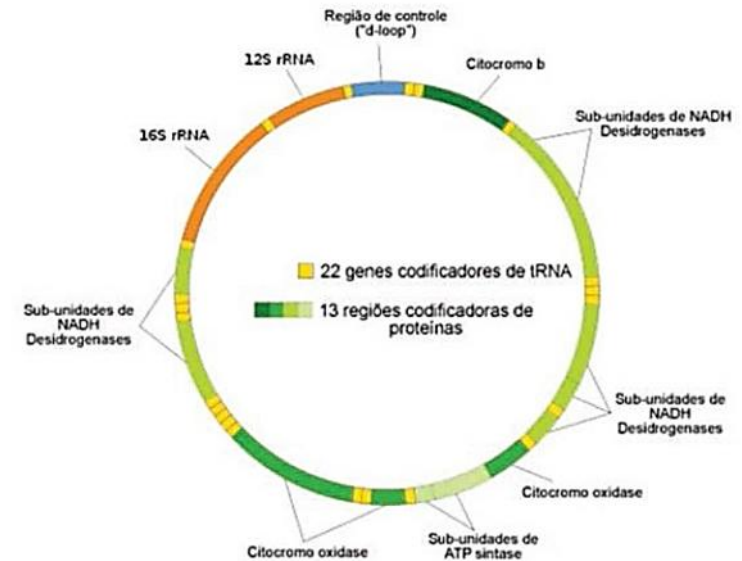
- Not related to the mitogenome

☞ Sequencing

- Mitochondrial sequences often present
- 1 in each 200-1000 sequences is mitochondrial
- Many species with public data available and no mitogenome published

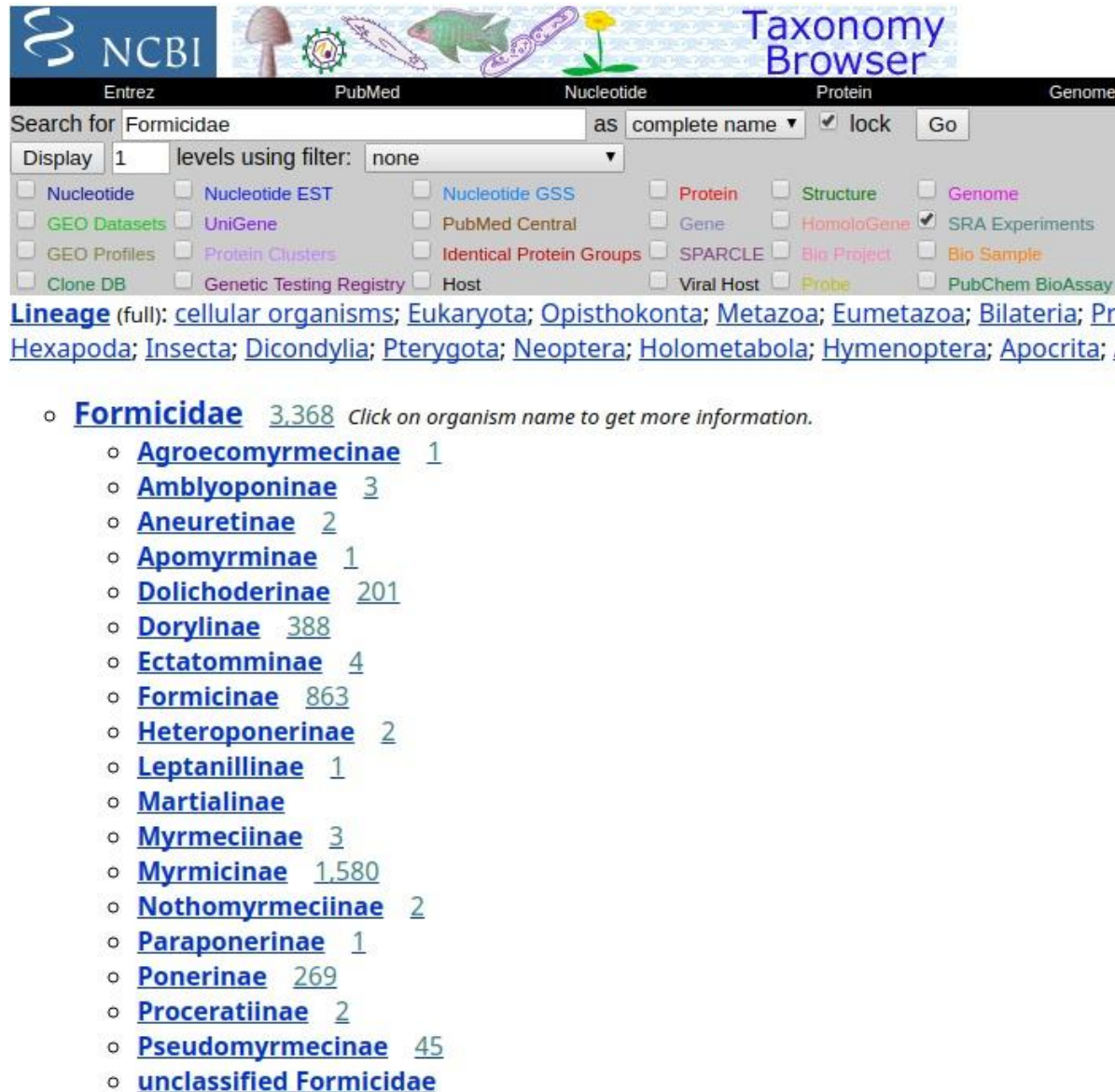
- Most dataset types can be used to assemble the mitogenome

- Phylogenetics (Krzemińska et al., 2017)
- Phylogeography (Fields et al., 2018)
- Populational genetics (Kilińç et al., 2018)
- Conservation (Rosel et al., 2017)



Ants (Hymenoptera: Formicidae)

- More than 13000 *spp.*
- Huge amount of genomic sequencing data for the clade
 - \approx 3.91 Terabytes
- 35 complete mitogenomes
- Why not use this data?



The screenshot shows the NCBI Taxonomy Browser interface. The search bar contains "Formicidae" and the results are displayed as a list of taxa with their respective counts. The interface includes navigation tabs for Entrez, PubMed, Nucleotide, Protein, and Genome. The search results are organized into a hierarchical list, starting with the lineage and then listing the subfamilies of Formicidae.

Lineage (full): [cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Pr](#)
[Hexapoda](#); [Insecta](#); [Dicondylia](#); [Pterygota](#); [Neoptera](#); [Holometabola](#); [Hymenoptera](#); [Apocrita](#); ,

- **Formicidae** [3,368](#) *Click on organism name to get more information.*
 - [Agroecomyrmecinae](#) [1](#)
 - [Amblyoponinae](#) [3](#)
 - [Aneuretinae](#) [2](#)
 - [Apomyrminae](#) [1](#)
 - [Dolichoderinae](#) [201](#)
 - [Dorylinae](#) [388](#)
 - [Ectatomminae](#) [4](#)
 - [Formicinae](#) [863](#)
 - [Heteroponerinae](#) [2](#)
 - [Leptanillinae](#) [1](#)
 - [Martialinae](#)
 - [Myrmeciinae](#) [3](#)
 - [Myrmicinae](#) [1,580](#)
 - [Nothomyrmeciinae](#) [2](#)
 - [Paraponerinae](#) [1](#)
 - [Ponerinae](#) [269](#)
 - [Proceratiinae](#) [2](#)
 - [Pseudomyrmecinae](#) [45](#)
 - [unclassified Formicidae](#)

PeerJ

Accessible molecular phylogenomics at no cost: obtaining 14 new mitogenomes for the ant subfamily Pseudomyrmecinae from public data

Gabriel A. Vieira and Francisco Prosdocimi

Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

Submitted 7 November 2018

Accepted 10 December 2018

Published 24 January 2019

Corresponding authors

Gabriel A. Vieira



Mitochondrial DNA Part B Resources

ISSN: (Print) 2380-2359 (Online) journal homepage: <https://www.tandfonline.com/loi/tmdn20>

Complete mitochondrial genomes for three lizards (*Anolis punctatus*, *Sceloporus woodi*, and *S. grammicus*): a contribution to mitochondrial phylogenomics of Iguanoidea

Ana Teresa Nogueira Dumans, Gabriela Warwar Teixeira, Gabriel Alves Vieira, Deise Schroder Sarzi, Carolina Furtado, W. Bryan Jennings & Francisco Prosdocimi

Mitogenomics: is it accessible?

- Two major problems:
 - Lack of bioinformatics expertise
 - Lack of robust servers for mitogenome assembly
- MitoFree:
 - Simple command-line interface and input file

##General usage:

```
$> python3 /path/to/mitofree.py input_file.txt
```

##Example of input file for MitoFree:

##Each line corresponds to a different assembly

##Three tab-separated columns:

##(1)SRA ACCESSION

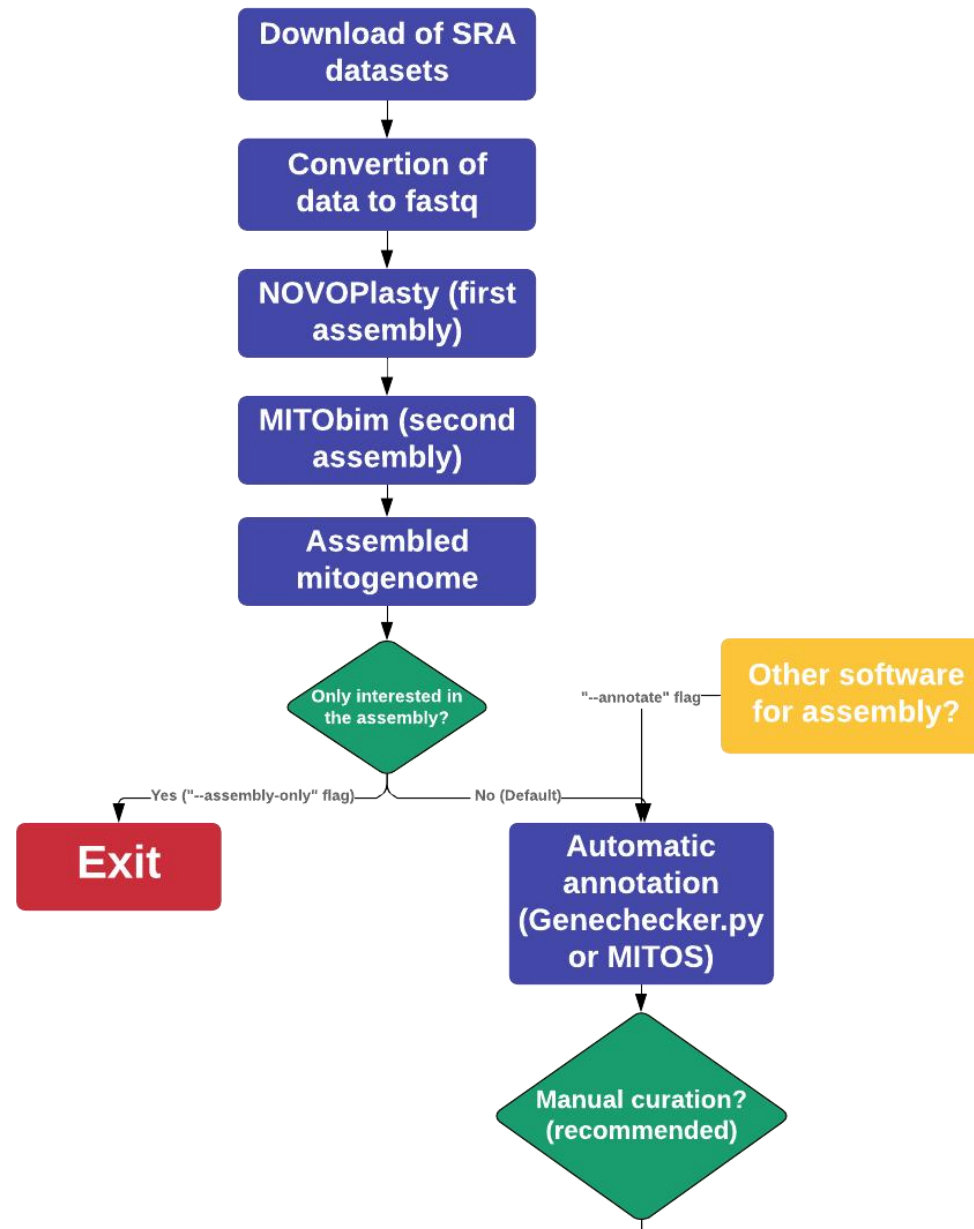
##(2)SPECIES NAME

##(3)SEED ACCESSION

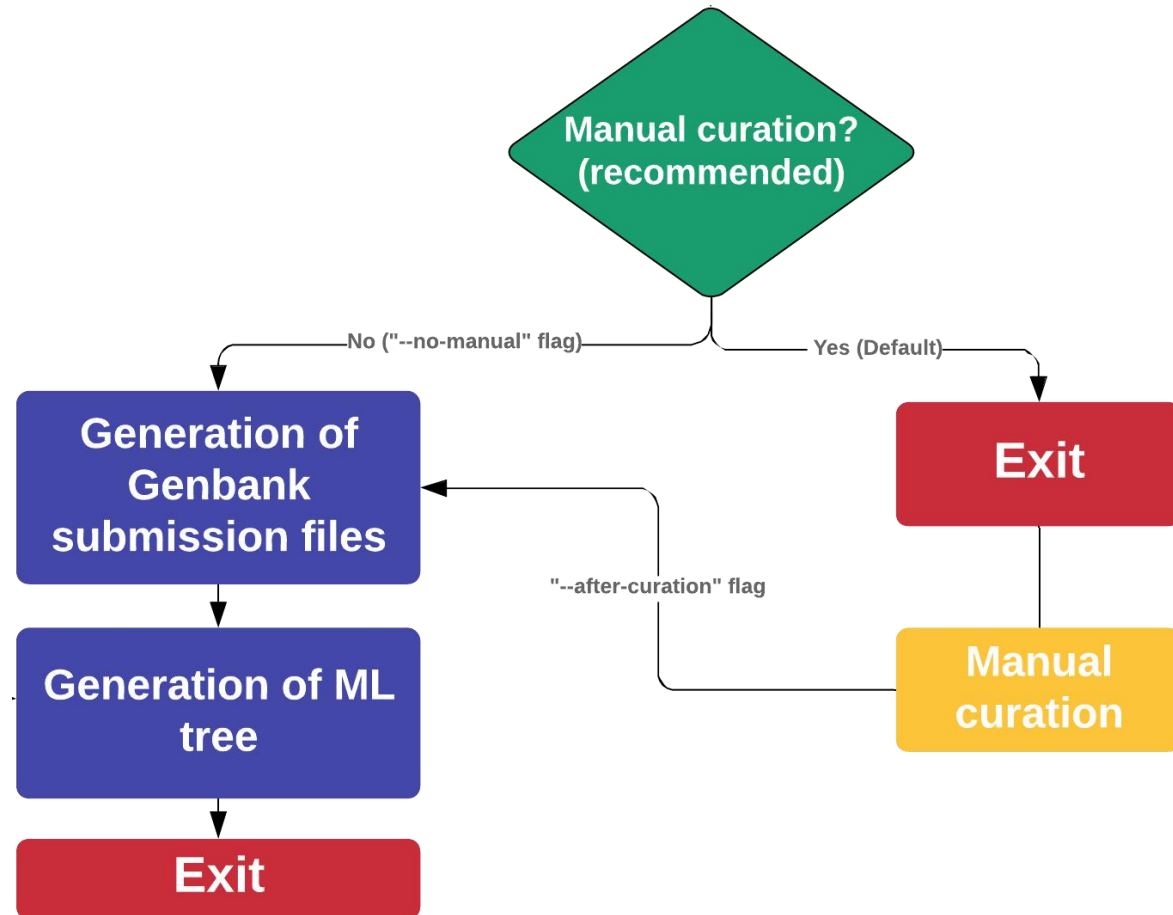
ERR1306022	Species1	MK297287
ERR7295163	Species2	MK297241
ERR1306034	Species3	MK291745
SRR4409513	Species4	MK291678

- Low RAM requirements
 - Optimized assemblers - organelle assembly

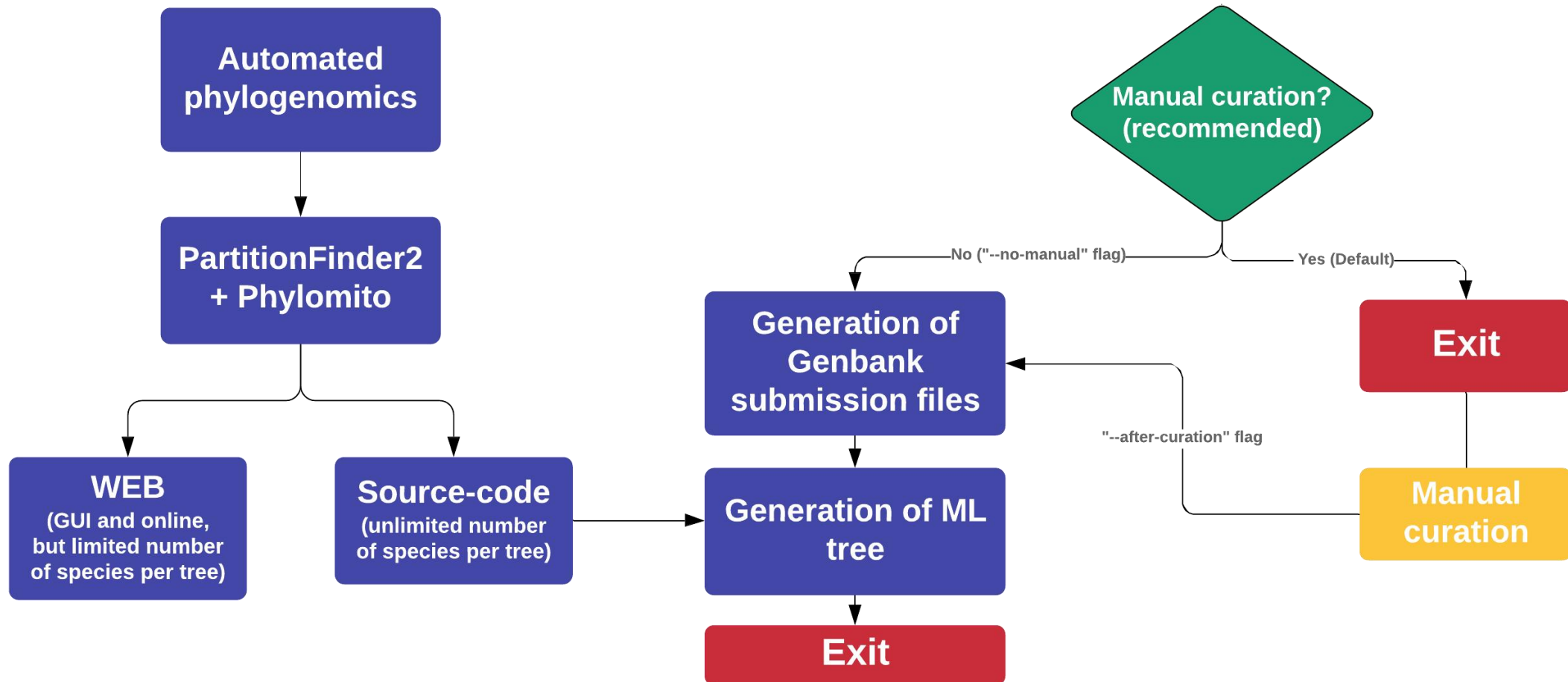
ALGORITHM (First half)



ALGORITHM (Second half)

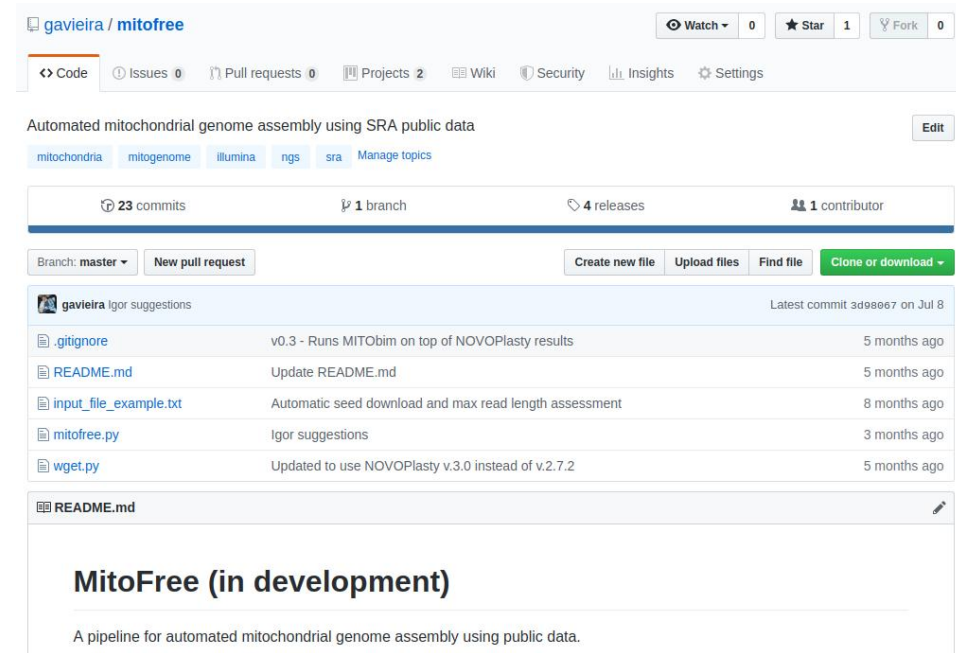


ALGORITHM (Second half)

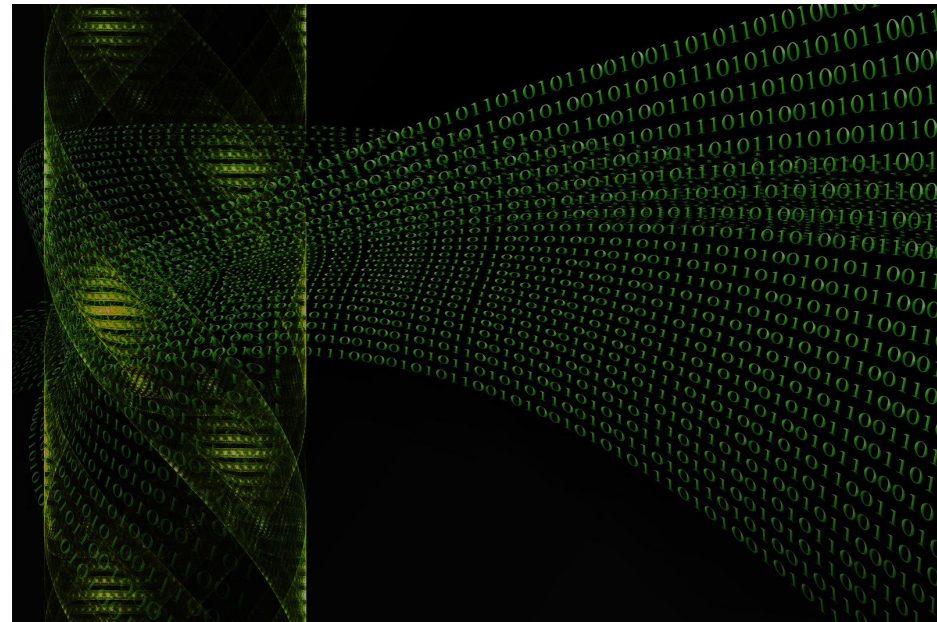


MitoFree

- Python3
- Several dependencies
 - Biopython
 - Third-party software
 - Future - Single package
- Beta version
 - Downloads and assembles mitogenomes using public data
 - Available for download (google it!)

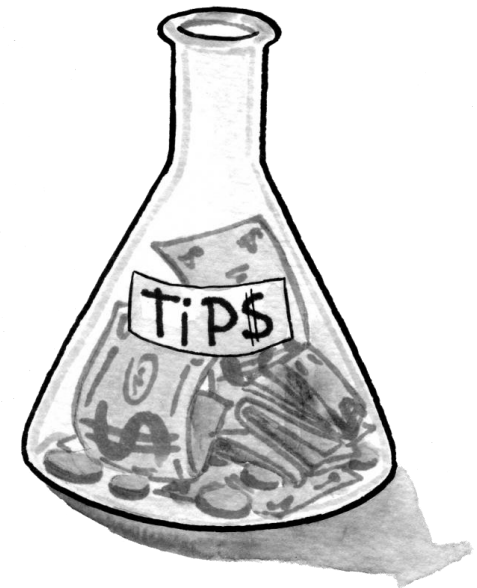


<https://github.com/gaviiera/mitofree>



Conclusion

- Free datasets available
 - Species with no mitogenome described
- Mitogenomes
 - Ubiquitous in sequencing datasets
 - Versatile in analyses
- NO/LOW BUDGET MITOGENOMICS
 - No sequencing costs
 - Free bioinformatics software
 - No need for supercomputers
- MitoFree hopefully will:
 - Highlight the untapped potential of public sequencing databases
 - Make mitogenomics using public data more accessible
 - ↑ mitogenomes : Evolution and Conservation (Non-model)



REFERENCES

- Fields, Peter D., et al. "Mitogenome phylogeographic analysis of a planktonic crustacean." *Molecular Phylogenetics and Evolution* (2018).
- Kılınç, Gülşah Merve, et al. "Investigating Holocene human population history in North Asia using ancient mitogenomes." *Scientific reports* 8.1 (2018): 8969.
- Krzemińska, Urszula, et al. "Population mitogenomics provides insights into evolutionary history, source of invasions and diversifying selection in the House Crow (*Corvus splendens*)." *Heredity* 120.4 (2018): 296.
- Rosel, Patricia E., et al. "Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences." *Marine Mammal Science* 33.S1 (2017): 76-100.

OBRIGADO!

Email: gabriel.vieira@bioqmed.ufrj.br

