



Obtenção de genomas mitocondriais e filogenômica de insetos usando dados públicos

Aluno: Gabriel Alves Vieira

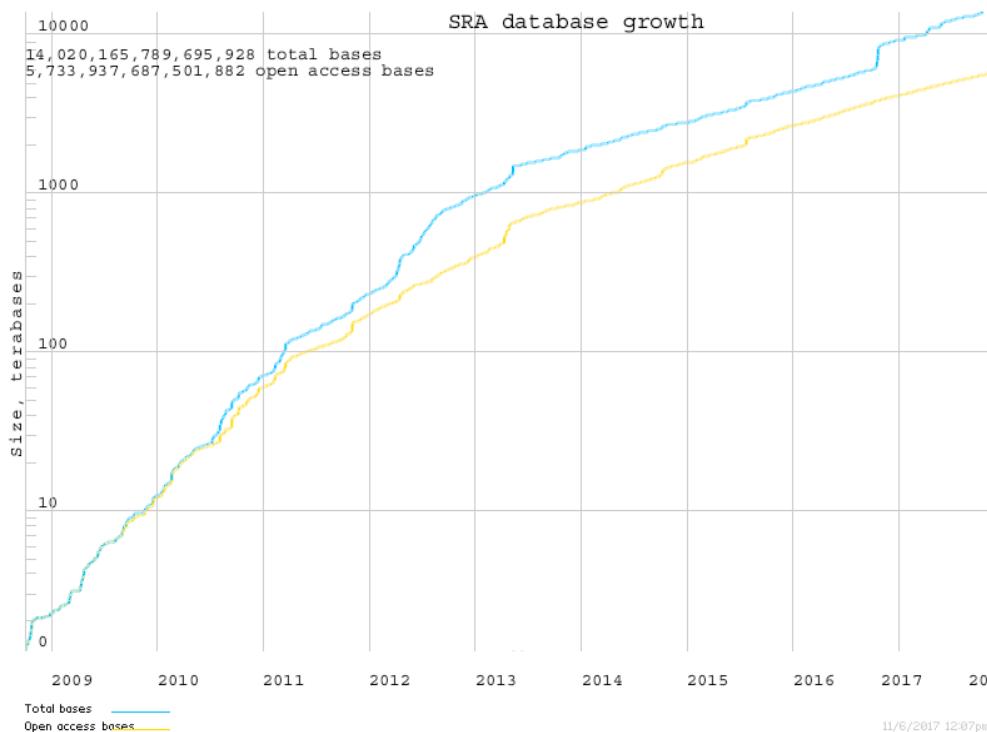
Orientador: Francisco Prosdocimi



```
AAGTCAGCTGCTGTGGCTGTGATCTGCCAAACCCCACAGCTGGTAGCAGG  
AGGACCTTGATGCTCTGGCACAGATGAGGAGAACCTCTCTTCTCTGTTGAAG  
GACAGACATGACTTGGATTTCAGGAGGACTTGGCAACCAGITCCAAAAGGCT  
GAAACCATCCCTGCTCCATGAGATGATCCACCAGATCTCAATCTTCACTGCACA  
AAGGACTCAGTCAGCTGCTGGATGAGACCCCTCTAGACAAATTCTACACTGAAC  
TACCGAGCTGAATGACCTGGAAAGCTGTGTGATAACAGGGGTGGGGTACAGAG  
ACTCCCCCTGATGAAGGAGGACTCATTCTGGCTGTGAGGAAATACTTCAAAGAAC  
ACTCTCTATCTGAAAGAGAAGAAATACAGCCCTTGCTGGAGGTTGTCAGAGCA  
GAAATCATGAGATCTTCTCTGCAACAAACTGCAAGAAAGTTAAGAAGTAAG  
GAATGA, TCTGATCTGCTCAAACCCACAGCTGGTAGCAGGAGGACCTGTG  
TCCTGGCACAGATGAGGAGAACCTCTCTCTGCTGTTGAAGGACAGACATGACT  
TTGGATTTCCCAGGAGGAGTTGGCAACCAGTCCAAAGGCTGAACCATCCCTG  
TCCTCCATGAGATGATCCAGCAGATCTCAATCTTCAGCACAAGGACTCATCTG  
CTGCTTGGGATGAGACCCCTCTAGACAAATTCTACACTGAACCTACAGCAGCTGA
```

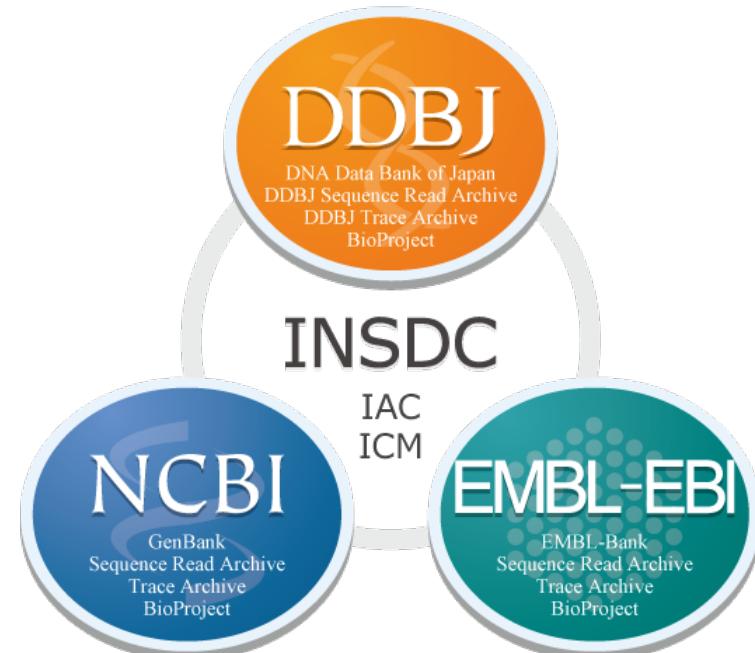
International Nucleotide Sequence Database Collaboration

- Crescimento dos bancos de dados públicos de sequenciamento
- Sequenciamento de Nova Geração (NGS)
 - ↑ produção de dados
 - ↓ custo por base
- Revistas: Incentivam disponibilização dos dados
 - Reprodutibilidade



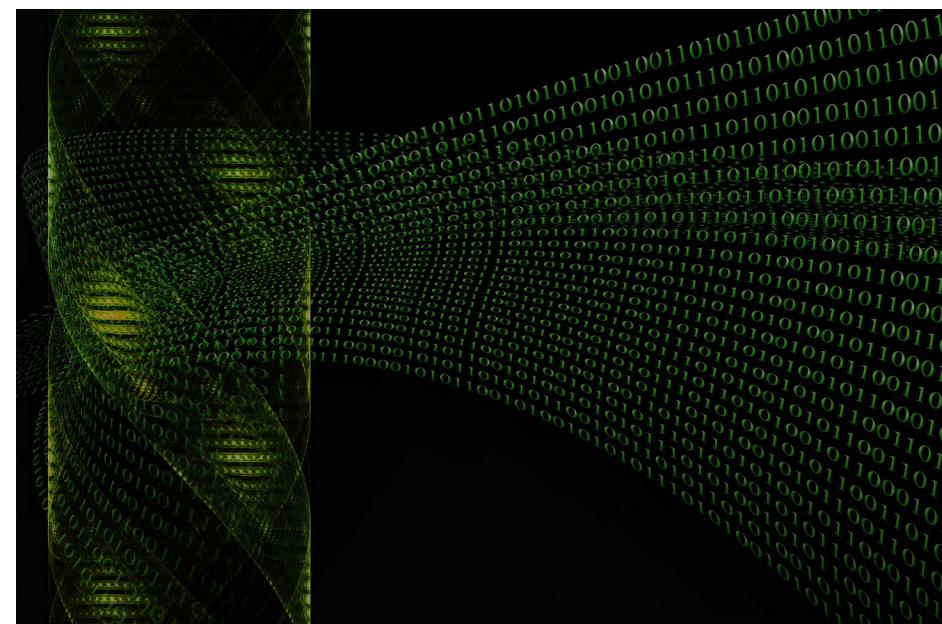
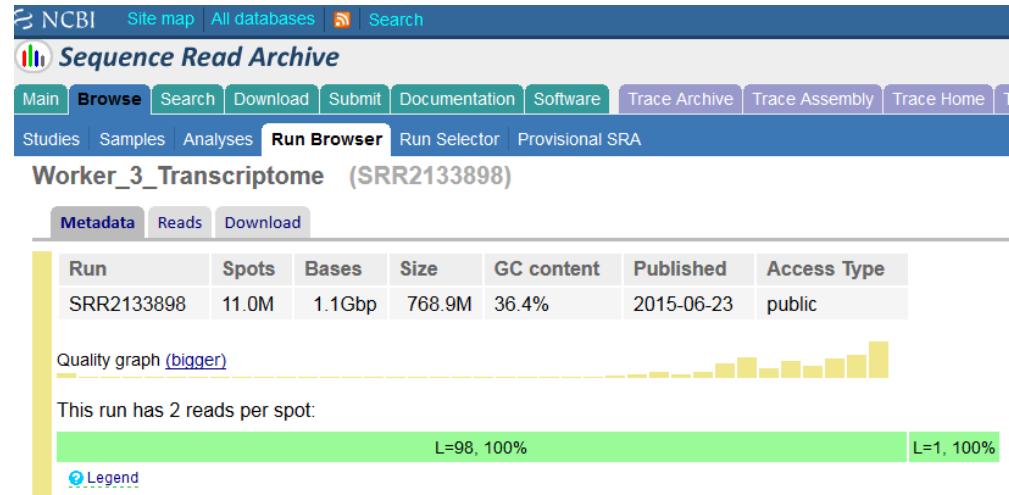
Sequence Read Archive (SRA)

- Parte de colaboração internacional
 - 3 bancos:
 - SRA: NCBI Sequence Read Archive
 - ERA: EBI Sequence Read Archive
 - DRA: DDBJ Sequence Read Archive
 - Sincronizados
- Diferentes tipos de dados:
 - DNA-seq; Bisulfite-seq; RNA-seq...
 - Um pouco de reflexão e imaginação
 - Gerar novos conhecimentos



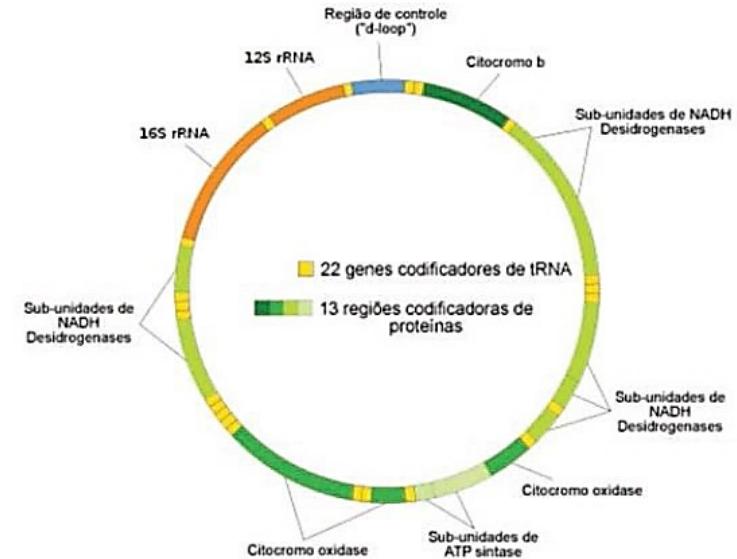
Dados públicos

- Dados usados para responder uma pergunta
 - Nada a ver com a mitocôndria
- Sequenciamento
 - Não separa a mitocôndria
 - 0,05% dos dados é mitocondrial
 - Espécies com dados públicos e sem mitogenoma descrito
- Esses dados podem ser usados para obter o mitogenoma



Mas por que o mitogenoma?

- Menor genoma da célula ($\approx 16\text{kb}$)
 - Relativamente simples de montar
 - Excelente treino para iniciantes
 - Obtido de diversos tipos de datasets
 - WGS, RNA-Seq, Targeted Sequencing

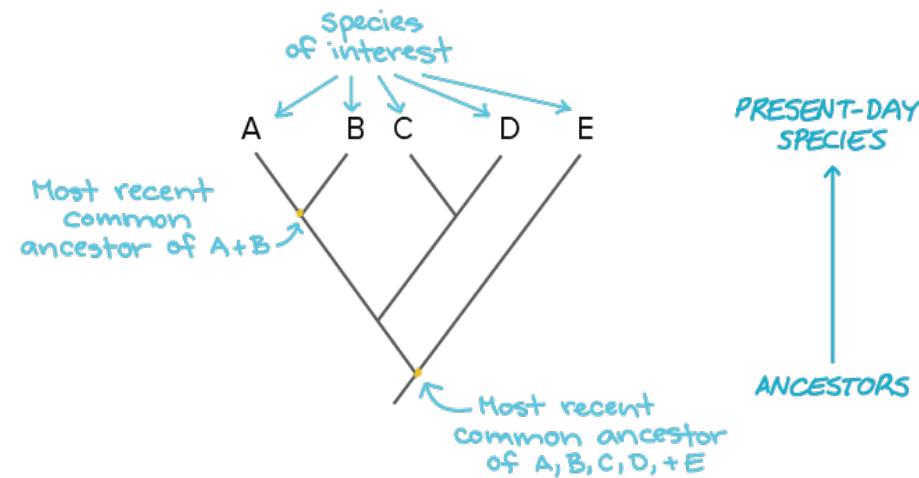


- Utilizado como marcador molecular

- Sem recombinação
- Herança materna

- Uso de genes mitocondriais para estudar relações evolutivas
 - Corroborar ou questionar filogenias anteriores

Filogenia e
Filogeografia



The complete mitochondrial genome of the golden mussel *Limnoperna fortunei* and comparative mitogenomics of Mytilidae

Marcela Uliano-Silva ^{a,*}, Juliana Alves Americo ^a, Igor Costa ^b, Alex Schomaker-Bastos ^{b,1},
Mauro de Freitas Rebelo ^a, Francisco Prosdocimi ^b

^a BioMA Laboratory, Biophysics Institute Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Brazil

^b Genomics and Biodiversity Laboratory, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

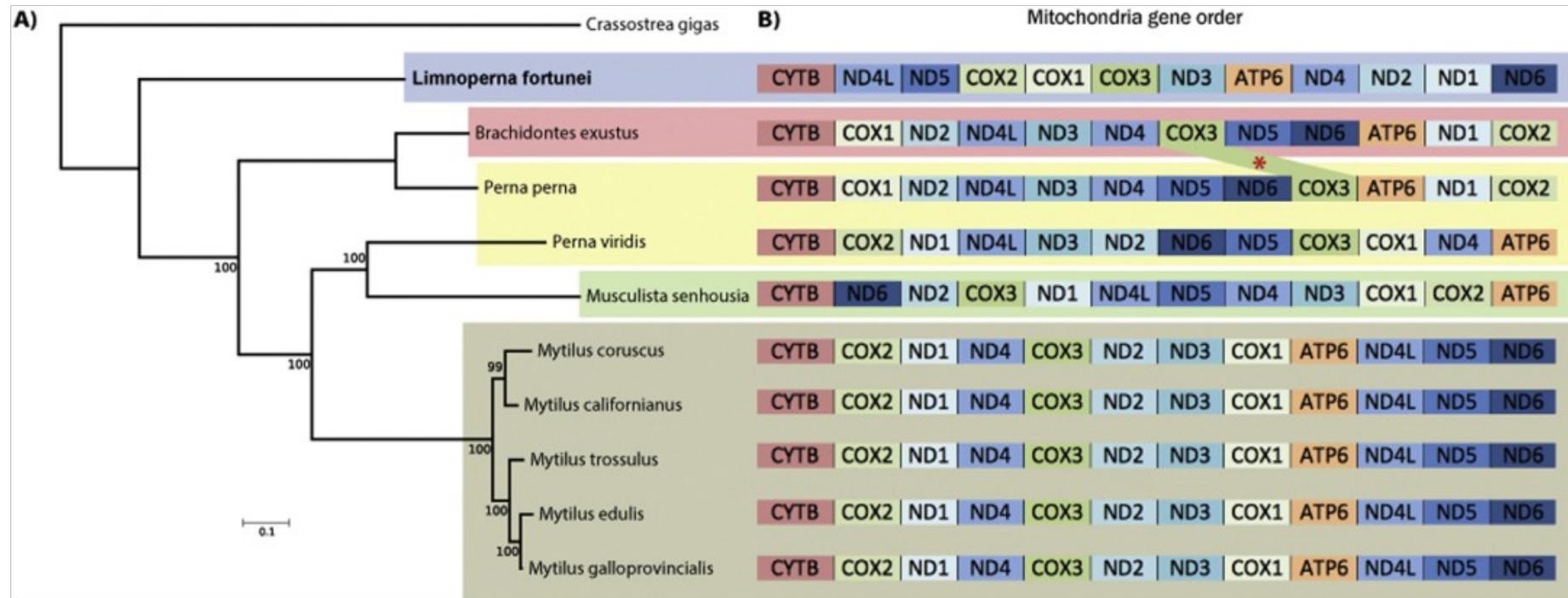


Fig. 4. A) Maximum likelihood phylogeny using a super-matrix from the concatenation of all genes for the 10 Mytilidae mitochondrial genomes present in public databases up to date, including the new released mitochondrial genome of *Limnoperna fortunei*. *Crassostrea gigas* mitogenome was used as an outgroup. B) The gene order of all Mytilidae mitogenomes were organized starting with Cytochrome B. Colored boxes indicate different genes. Polyphyly detected for *Perna* genus was observed using both (i) sequence data as indicated by the tree and synteny, once one single rearrangement can explain the different gene order between *B. exustus* and *P. perna* (green box with a red asterisk). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The complete mitochondrial genome of the golden mussel *Limnoperna fortunei* and comparative mitogenomics of Mytilidae

Marcela Uliano-Silva ^{a,*}, Juliana Alves Americo ^a, Igor Costa ^b, Alex Schomaker-Bastos ^{b,1},
Mauro de Freitas Rebelo ^a, Francisco Prosdocimi ^b

^a BioMA Laboratory, Biophysics Institute Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Brazil

^b Genomics and Biodiversity Laboratory, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

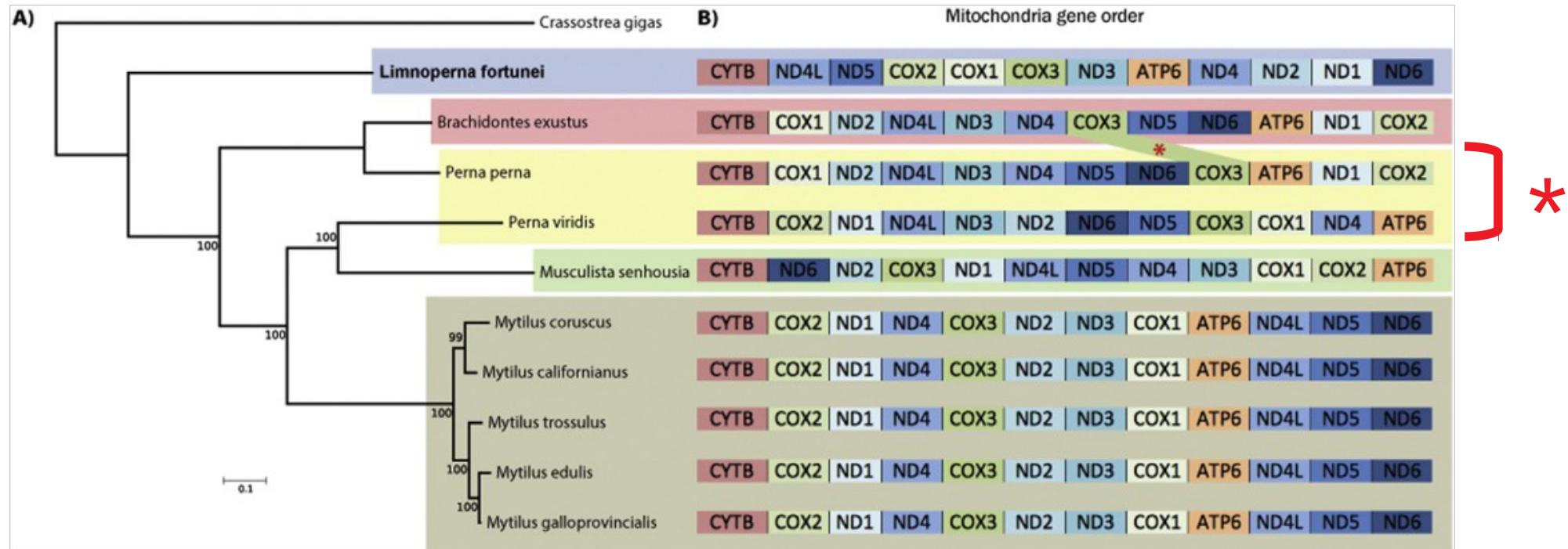
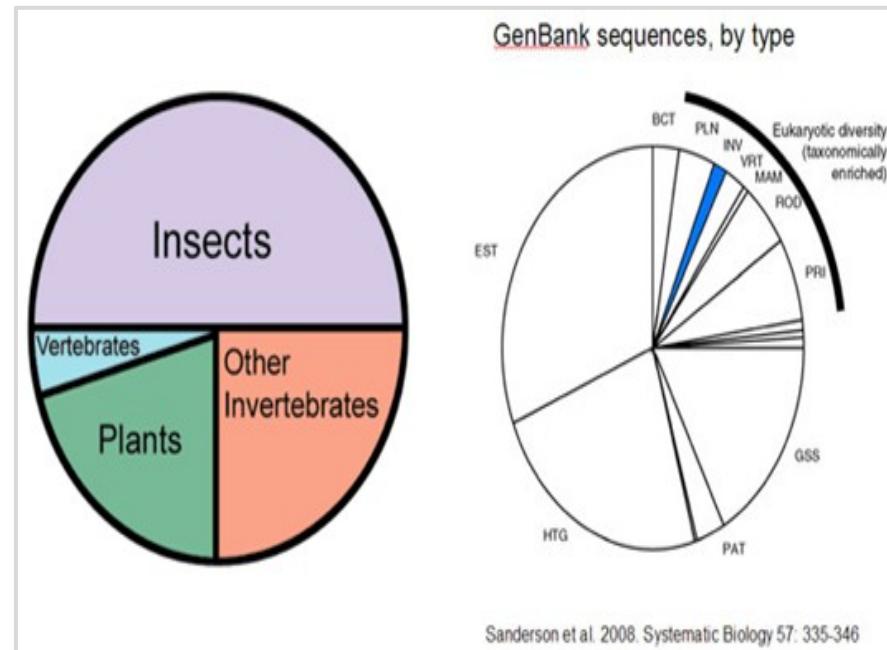


Fig. 4. A) Maximum likelihood phylogeny using a super-matrix from the concatenation of all genes for the 10 Mytilidae mitochondrial genomes present in public databases up to date, including the new released mitochondrial genome of *Limnoperna fortunei*. *Crassostrea gigas* mitogenome was used as an outgroup. B) The gene order of all Mytilidae mitogenomes were organized starting with Cytochrome B. Colored boxes indicate different genes. Polyphyly detected for *Perna* genus was observed using both (i) sequence data as indicated by the tree and synteny, once one single rearrangement can explain the different gene order between *B. exustus* and *P. perna* (green box with a red asterisk). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Formigas (Hymenoptera: Formicidae)

- Engenheiras ecossistêmicas
 - Alteram características do solo
 - Serviços ecossistêmicos
- Importância econômica
 - Pragas
 - Agentes de controle biológico
- Grande biodiversidade
 - ≈ 13.000 spp.
 - GenBank
 - 15 mitocondrias descritas



Formigas (Hymenoptera: Formicidae)

- Grande quantidade de informação para o clado:
 - ≈ 3.91 Terabytes
- Todas as subfamílias
 - Exceto Martialinae
- Pseudomyrmecinae:
 - Subfamília pequena
 - Sem mitogenoma descrito
 - Muitos dados

The screenshot shows the NCBI Taxonomy Browser interface. The search bar at the top contains 'Formicidae'. Below the search bar, there are several filter options and a lineage tree. The lineage tree starts with 'Lineage (full)' followed by a list of taxonomic groups. The 'Pseudomyrmecinae' link is highlighted with a red arrow.

Search for **Formicidae** as complete name lock

Display 1 levels using filter: none

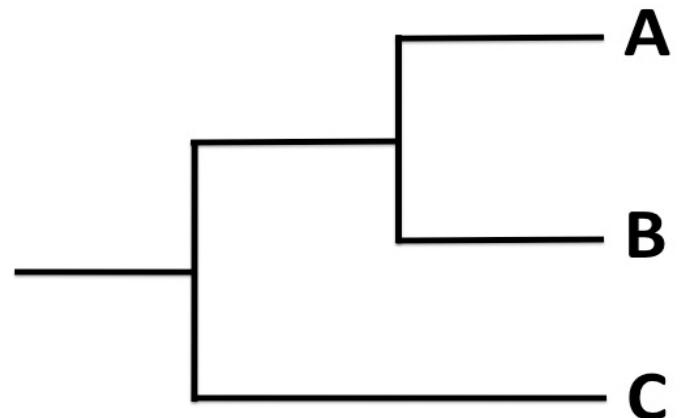
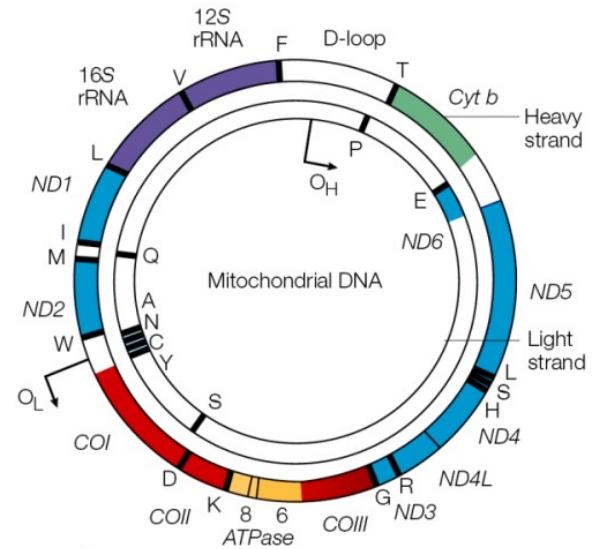
Nucleotide Nucleotide EST Nucleotide GSS Protein Structure Genome
 GEO Datasets UniGene PubMed Central Gene HomoloGene SRA Experiments
 GEO Profiles Protein Clusters Identical Protein Groups SPARCLE Bio Project Bio Sample
 Clone DB Genetic Testing Registry Host Viral Host Probe PubChem BioAssay

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

- **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](#)

Objetivos

- Montar, anotar e disponibilizar mitogenomas da família Formicidae usando dados públicos
- Qual a história evolutiva que a mitocôndria nos conta?
- Auxiliar a resolver as relações filogenéticas, especialmente as de organismos não modelo



Obtenção dos dados

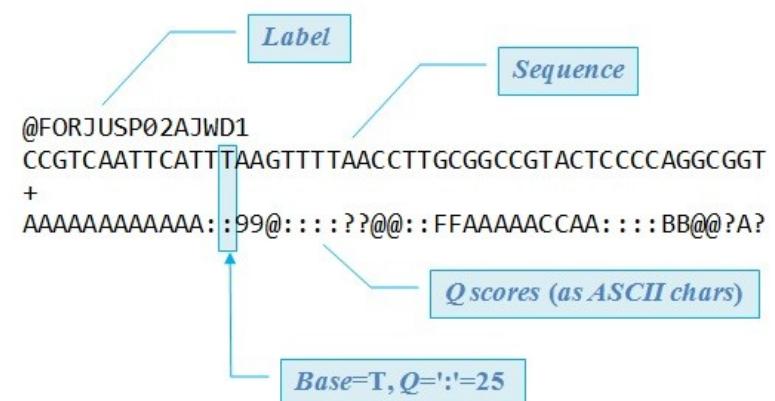
- Arquivos *.sra*

- Utilizado pelos 3 bancos
- Disponíveis para download direto
- Precisa ser convertido (SRA Toolkit) para *.fastq*

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	CRAM Index files (Galaxy)
PRJNA360290	SAMN06208930	SRS1901018	SRX2468701	SRR5150611	55425	Paraponera clavata	Illumina HiSeq 2500	PAIRED	File 1 File 2	File 1 File 2			File 1	File 1		

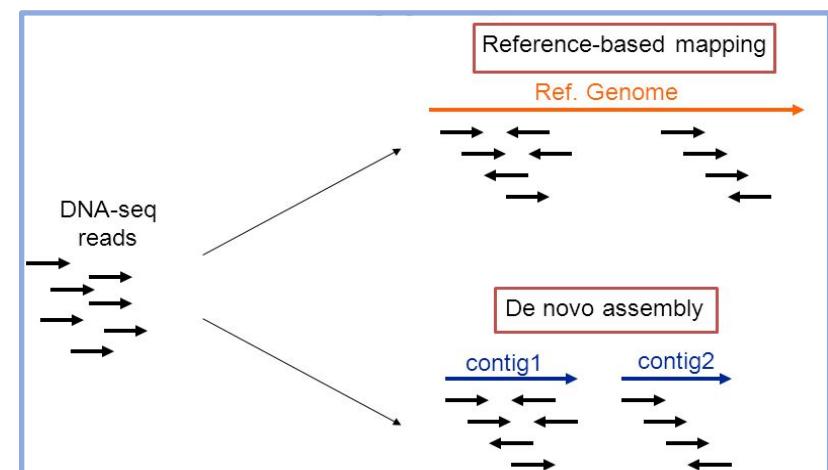
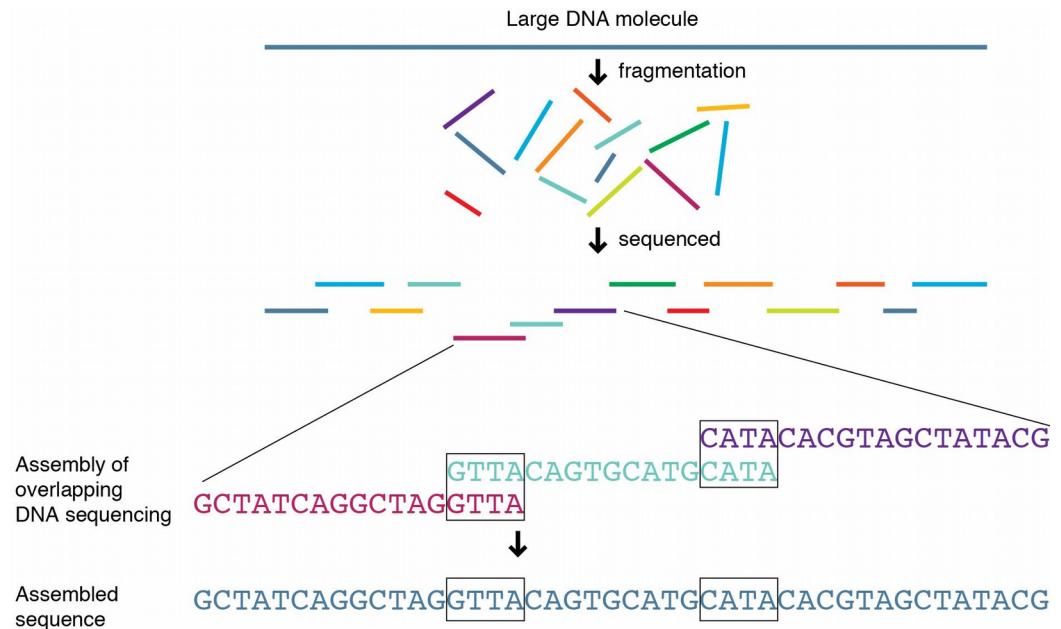
- Arquivos *.fastq*

- Cabecalho, sequência e qualidade
- Aceitos pela maioria dos programas



MONTAGEM

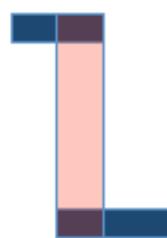
- Quebra de múltiplas cópias de DNA
- Obtenção de pequenas sequências (reads)
- Quebra-cabeça – Ferramentas computacionais
 - NOVOPlasty, MIRA, MITObim
- Dois tipos de montagem
 - *De novo*
 - Sobreposição
 - Referência
 - Alinhamento com sequência
 - + fácil; menos gaps
 - Qualidade da referência



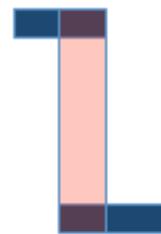
Montagem



Montagem



Montagem



Sequência contínua ou contig



Montagem



Fragmentos
alinhados

ACCGCGATTCA~~G~~GT~~T~~ACCACG
GCGATTCA~~G~~GT~~T~~ACCACGCG
GATTCA~~G~~GT~~T~~ACCACGCGTA
TTCAGGTTACCACGCGTAGC
CAGGTTACCACGCGTAGCGC
GGTTACCACGCGTAGCGCAT
TTACCACGCGTAGCGCATT
ACCACGCGTAGCGCATTACA
CACCGCGTAGCGCATTACACA
CGCGTAGCGCATTACACAGA
CGTAGCGCATTACACAGATT
TAGCGCATTACACAGATTAG

Contig consenso

ACCGCGATTCA~~G~~GT~~T~~ACCACGCGTAGCGCATTACACAGATTAG

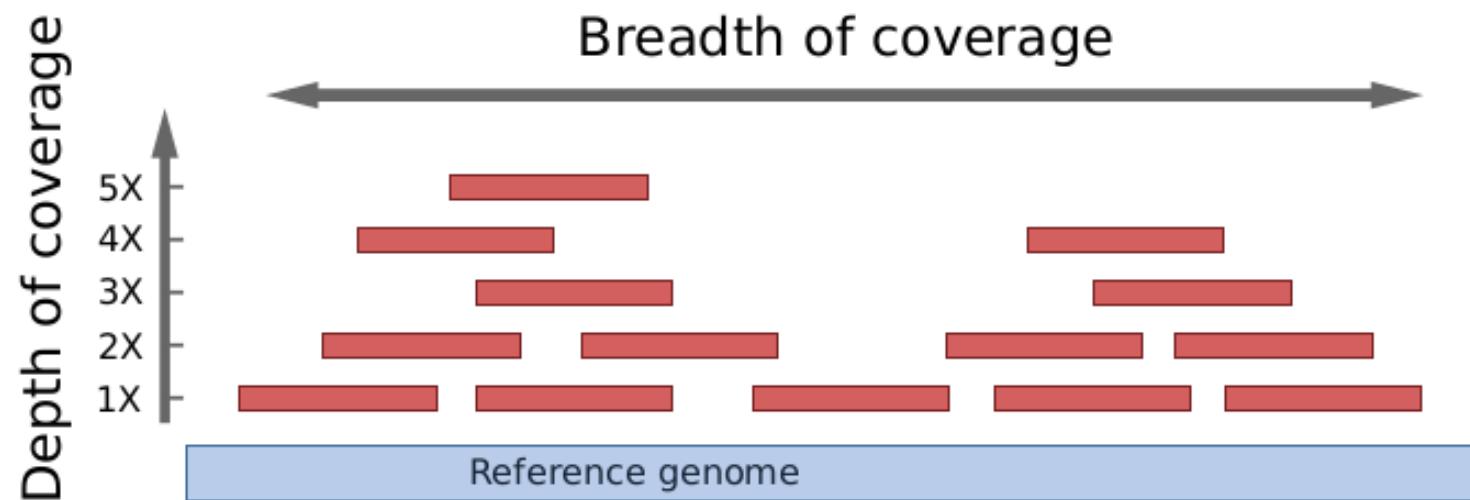
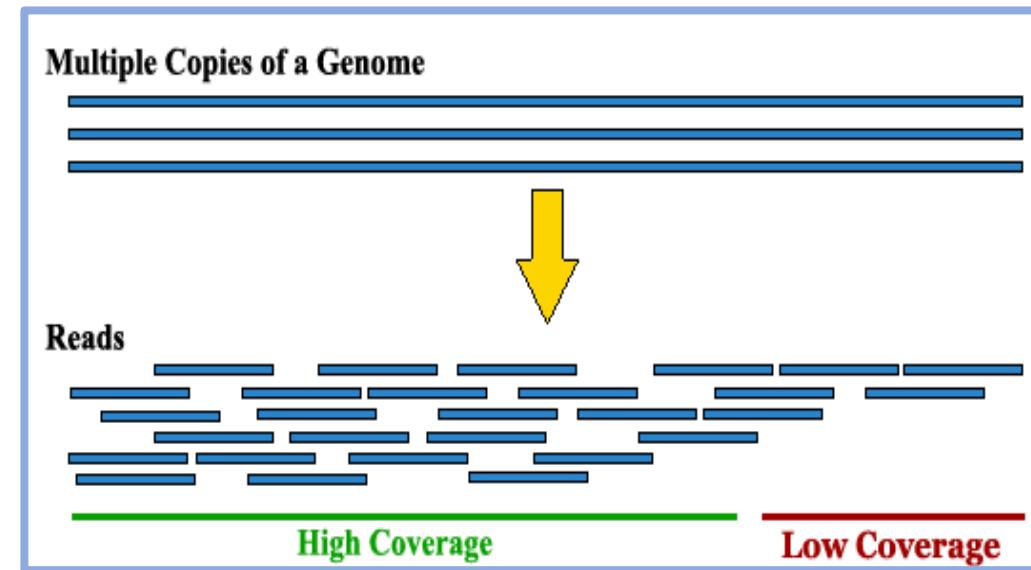
Montagem



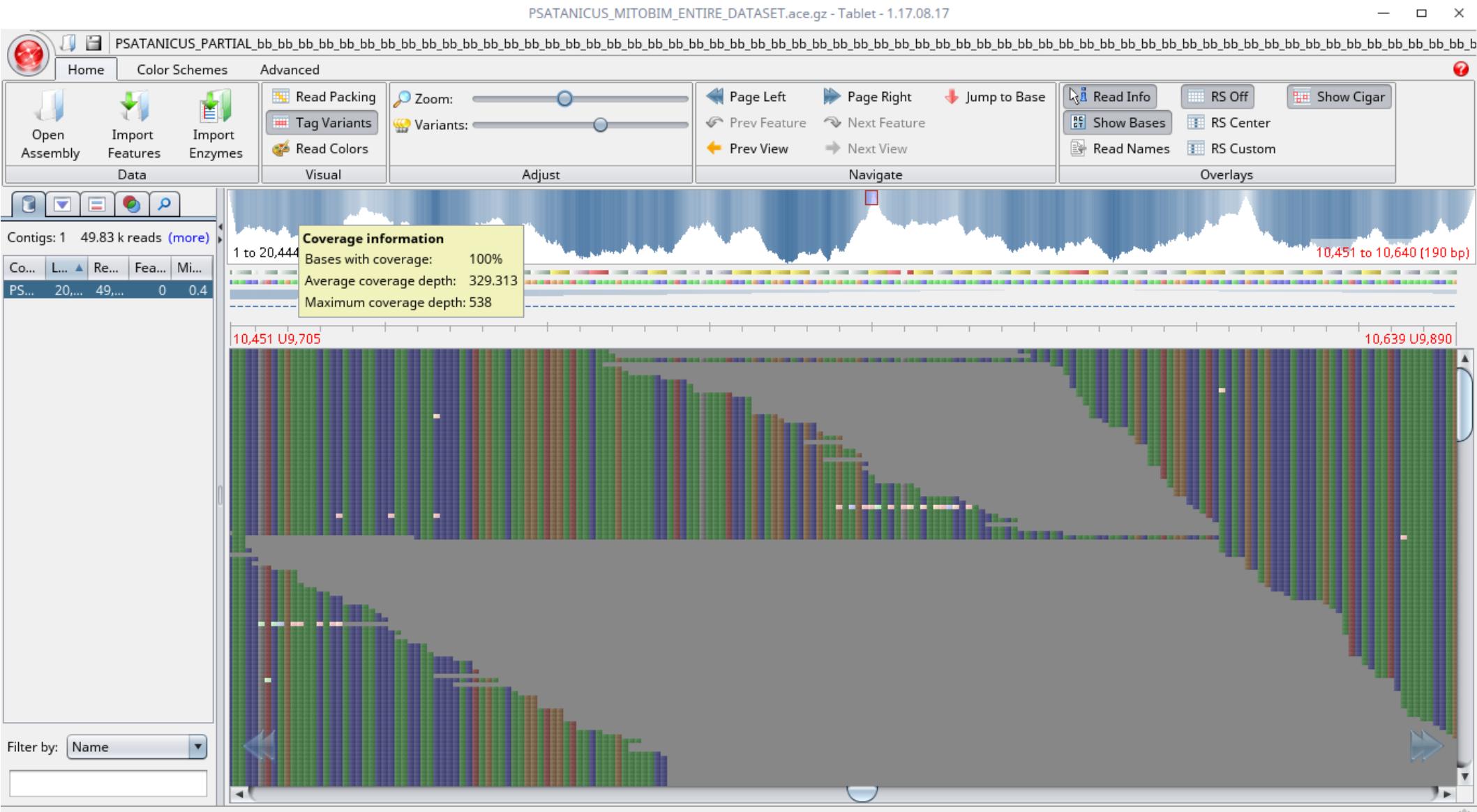
MONTAGEM

- Checagem (TABLET)

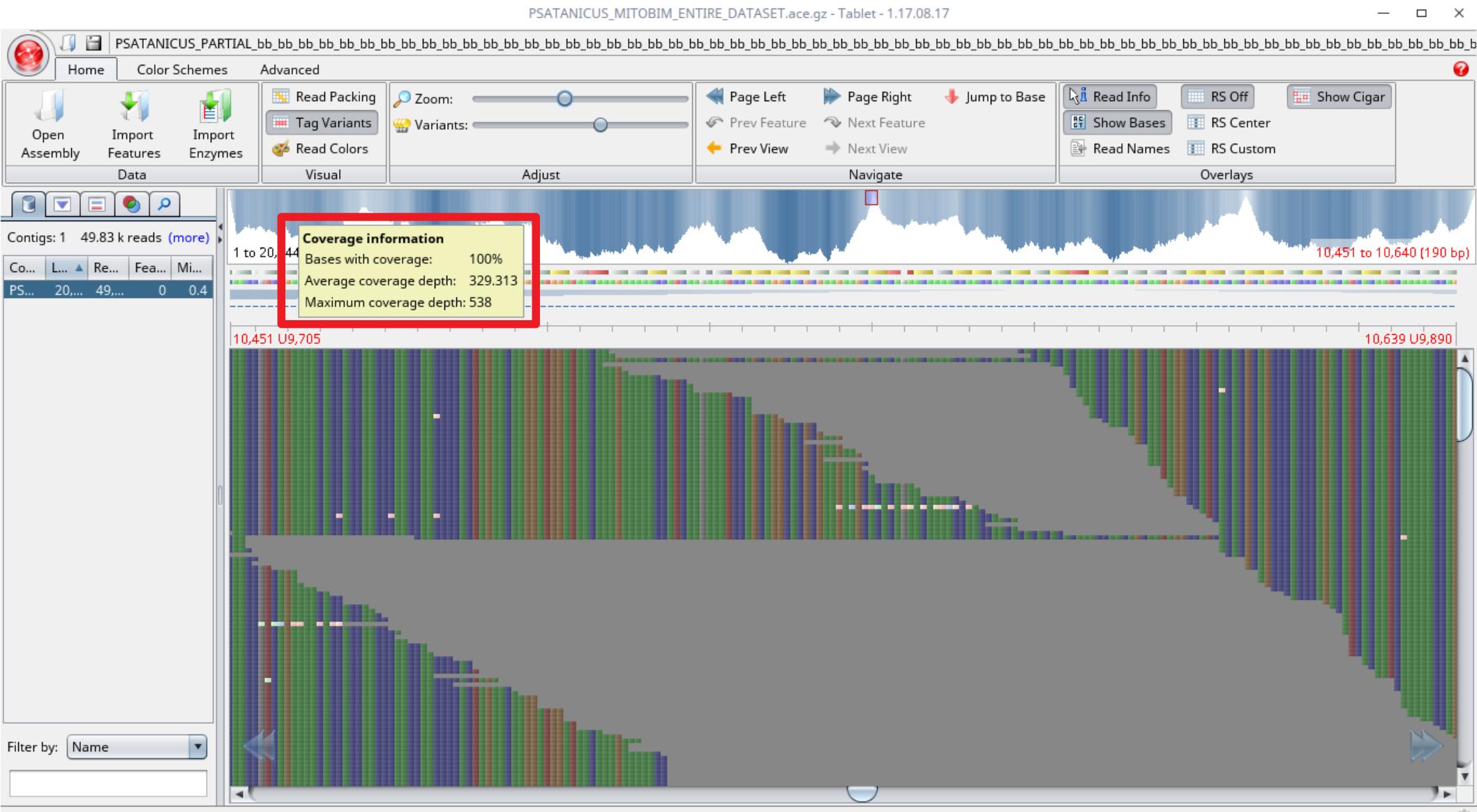
- Cobertura
 - Quantas sequências corroboram a montagem
 - Circularização



TABLET

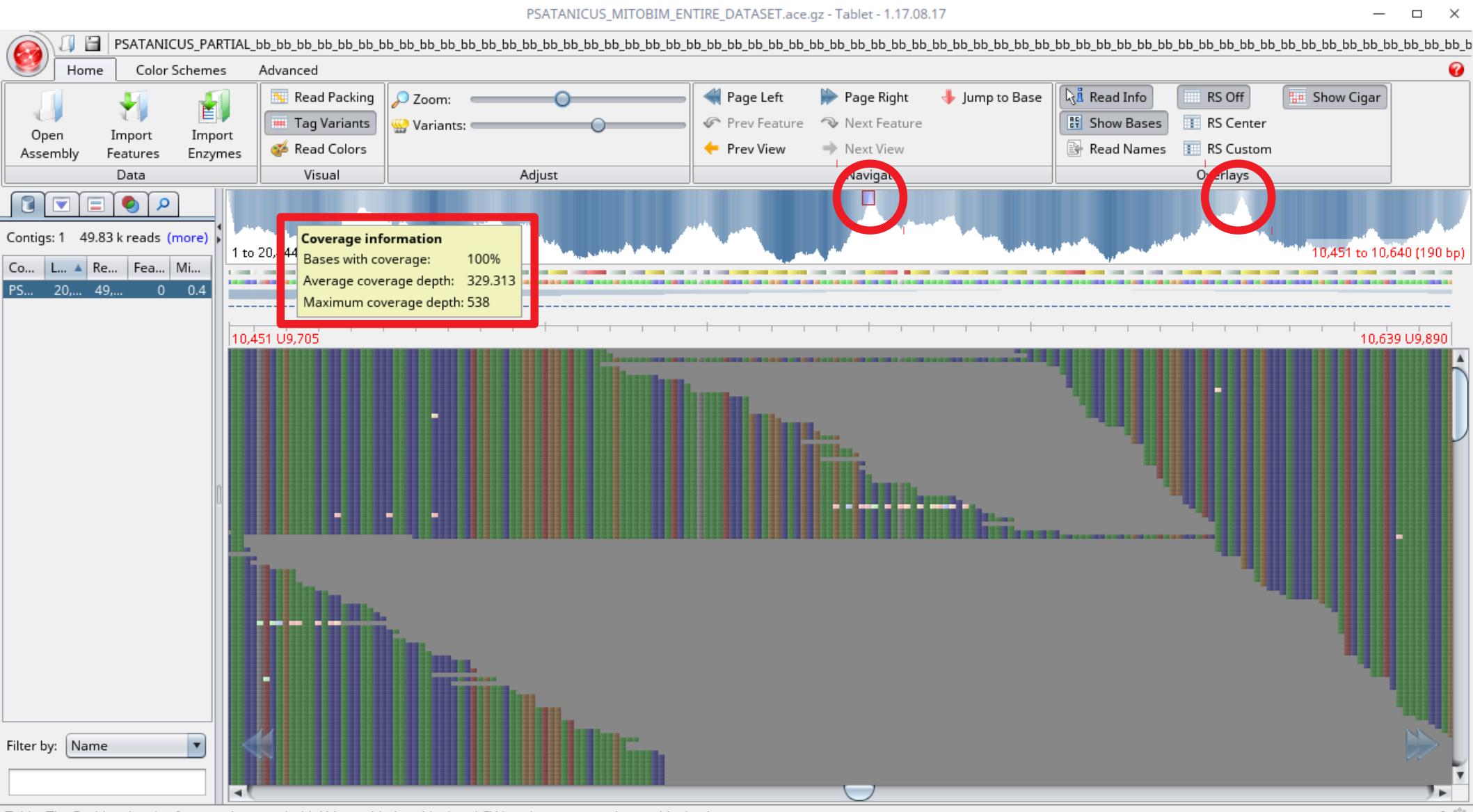


TABLET



Tablet Tip: Position data is often supplemented with U (unpadded position) and CV (read coverage at that position) values so

TABLET



Montagem

- Resultado da montagem: arquivo fasta
 - Arquivo de texto
 - Não apresenta valor de qualidade

- Apresenta apenas:
 - 1^a linha: Cabeçalho
 - Sequência

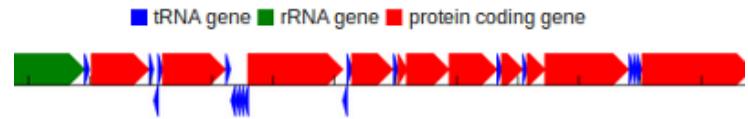
- Vantagem: Muito leve
 - Desvantagem: Pouca informação

>Paraponera clavata mitochondrion. complete genome
taactaatatattttttaatttatctcccccctaattttataatatttagtt
attaataaaaatttaataatcttatttatattatccttaaaataatttttaatt
cttctatgtataatatttaatagctatatttttatttattttactttta
aatttaataaaatacaatattgattttaatataagttttagataaaatttatattt
ataataataattatattttatctattaaactaattttataatataatttaa
ttaaaaatttaattaaatttaatttatattttattttattttattttatataat
tgattattattattatgttatatctaagcactttatataaattttatataat
aaataatttaatttataatatttaatgattaataataaataattttatattttat
ttatattatataatataatataatattttatattttatattttatattttat
aaataatttatataatataatataatgtatattttatattttatattttat
ttcaattttaaattctaaataatataattttcatatttatattttatcttttatt
ttaaaaattttaaaaatttatcttaatactcgaggcgcctccccccttccccaa
cattaaatataatcattttaaaatttaatctattaaatcataaaatcacttactattt
tagactaacataataatttaatttcatttatttaattttatacttattttattttat
aataaaaataagtctggctaaataataaagcttttaggttcataccctaaacatagaatc
taatctttatatttcatgaaatgcctgataaaaggatttttgcataatc
atacaatatttaatttttttttattttatattttatattttatattttat
taaaatttcaaaaattttcatgcttttacattaaagtatataatataatattttat
attttatataatataatattttatattttatattttatattttatattttat
ttttttacaatatgatttttagaaattttaaattttatattttatattttat

TCTACCGCAAACAACACCGTCGAGGTGACTTACGAGGAACTCTCCCCAAGGGTTGACCGACTAACGCAGAGTGAAAGTGGTCGTGAGGCCGGAGAGAGAGAGGG
GGGGGTGAGGAGTTGGTATGCTTCCCTACCTGTGCGTCTACCATTGGCAGTGTAGTCCTCTGAGATGCCTCATTGGCGGTACTCTGACCTCGCTTTCTCTCTC
CCGTCGCTGAGCCGTGACTTCTGCTCTGGTTGGCGGTAGGCTAGGGGAGGGAGGGGGACGGCGGTGTTCGTGCCTCCCCCTTGCTGTGG
GATGCAGGCATGTTTCATGTGCTGAGGACAACGATGTACACTGGCTGCAACCTGGAAGTGTGGGCAGAGAGAGAGGCTGAGGAATGGAGCATGGAGG
CTCATGGAGAGGTCTTGTGGTGCCTTGATGTGGTGGCATCACACACCTCACTCGAAAACTACTCGCCAGCGAGCCTCTCCTTCCCAGTCGTTGGCTCTCT
TCCTCTTCTCTTCCGGTCTCGGCTCTTCAGGGCAGCGCCAACCAGCGCAAAAAACAAAGCGAGGCACAGAGGAGGAGTACTCACGACACGAGTAATGCC
GAAGCAGGTCAATTTCATGCAGCAAACATGCCGGCAGGAGAAACGCCATGGCGCTGTCCGCCGGACTGCAGGTGTGCCTATGGCGAGCGCCCCCTCCCCCTCA
CTGAGCGCGTGCCTGCCGGCACGCCCTGCGCACCATGCCACGCGCTCTTTGGTTGCAGGCATTGGCTCTCAACCTCACGACATTGCGC
CTCCGTCTCCTGCCCTGCACACTCCCTCACCTCCCTCCCTCTTCTCTCTCCCTCCCTCGACGCCGGCGCCACGCGCACACAGGCACAGGGACAGAC
ACACCTATGGACAGCGTGCCTGTGCAAACATGTGGCAGACTCTGCCAGCGATATCGTCTTCAGCCCACATCCCTCGCAGCCTCTGGAGTCCA
ACTGCACGGTCTCCGACGGCGCTACACCGCAACTCGAGCCTCGTCGCCGAAACCAACAGCCTCTGCATCTCGAAAACCTCGTCGAGAGGAATCACT
CGCTCGTGGCTTCCCTCGACAAGACCGCAGCGGCTGCCCTCTTCTACGTGAACCGCCCTCGTCGTGGTGGGCCGCAACCAGAACCTCTCCAGGAGGTTGC
GCTGCCGGCGGCGCTGCCGACGGCGTTCCCGTCGCTGCCGTGCCCTGCCGTTGCCTCTTCATGACGAGGGGAACCTCTGCCTCTGCTTCATCACGC
ACCGCACGCCATGCACCGGAGAAGACAATCCAGCTGATCCGCCCTGGCTCTGTGTGAATTATCGATGACCCCTGCACGGCTGACCA
CGACGACAGGCCACCGGAAAGATCACGGATCTGCAATGCCGGTGAGCGCGAAATTGCGTACCAACTGCACGCTGCTCGTCGATACCCACTGGCGTC
GCTCGGTGCTACCTCACCCGAAGGGGAGTACGTGGCGTTCAAGACGTCGTCGGTGGCTCGTGAAGCCCTGTCACCACACTCGCGGAGTCGGTCCACA
TTGCGAGTGGGAGGGCGCCATGGCCTCACTCAAGAGGAACATGGCGGAGTTTTCTAAGTGGGTGATCGAGTGCTGGAAAGCAGCGGACCGTGGAG
CGACGTGAGGGAGCTGCCAGTCCTGCCACCGCACGTCACTGCTGCCGGACACGCCCTCTTCCCTGATGTTGCGAGGCCGTCGCGGAGCATGTC
TTTCATCGAGGGCGAGGGCAGGCCGGCGGAGCTGGTGAACCTGCCACCCCTGGCGAGGCCGTTACAAGGCTGCGTCGAAGGACTGGGCTACCGATGCC
AGCCTCACATCCACGGTGCTCCTCAGCAGCGGTGAGCTCCAGCGGCCCTACAAGCGCTTCCGTTGGCCGATGTAGTGCAGGCTATCCTCTCGCGGAGGA
GCAGTTGCTGGCGGACTGCAGCAATGTGTCTTGCAGCTGGCTGCCGGAGTCGTGGCAGAGACCGAAGTGGGTTGCACCTGCTTACACAGTC

TCTACCGCAAACAACACCGT **CGAGGTGACTTACGAGGA** Regiões reguladoras TAAACGCAGAGTGAAGTGGTCGTGAGGCCGGAGAGAGAGGG
GGGGGTGAGGAGTTGGTATGCTTCCCTACCTGTGCGTCTACCATTGGCAGTGTAGTCCTCTGAGATGCCTCATGGCGGTACTCTGACCTCGCTTCTCTCTC
CCGTTCGCTGAGCCGTGACTTCTGCTCTGGTTGGCGGTAGGCTAGGGGAGGGAGGGGGACGGCGGTGTCGTGCCTCCCC **CTCTTGCTTGTGGA**
GAT CDS SAGGACAACGATGTACACTGGGCTGCAACCTGGAAGTGTGGGGCAGAGAGAGAGGCTCGAGGAATGGAGCATGGAGG
CTCATGGAGAGGGTCTGTGGTGCCTTGATGTGGTGGCATCACACACCTCACTCGAAAACTACTGCCAGCGAGCCTCCCTTCCGCATGCGTGTGGCTCTCT
TCCTCTTCTCTTCCGGTCTCGGCTCTCTCAGGGCAGCGCCAACCAGCGCAAAAAAACAAAGCGAGGCAC **AGAGGAGGGAGTACTCACGACACGAGTAATGCC**
GAAGCAGGTCAATTATGCAGCAAACATGCCCGCAGGAGAAACGCCATGGCGCTGTCCGCCGGACTGCAGGTGTGCCTATGGCGAGCGCCCCCTCCCCCTCA
CTGAGCGCGTGCCTGCCGGCACGCCCTGCGCACCATGCCACGCGTCTTTGGTTGCAGGCGATTGGGCTCTTCGTCCTCTAACCTTCACGACATTGCGC
CTCCGTCTCGCTCGCACACTCCCTCACCTC **CTCCCTCCC** rRNA **TCGACGCCGGGCCACCGCACACAGGCACAGGGACAGAC**
ACACCTATGGACAGCGTGCTCGTGTGCAAACATGTGGCAGACTCTGCCAGCGATATCGTCTCAGCCCACATCCCTCGCAGCCTCTGGAGTCCAATGCA
GGTGTGAAAGTCTCCGACGGCGCGCTACACCGCAACTCGAGCCTCGTCGCCGAAACCAACAGCCTCTGCATCTCGAAAACCTCGTCGAGAGGAATCA
CGCTCGTGGCTTCCCTCGACAAGACGCAGCGGCTGCTCCTCTACGTGAACCGCCCCTGCGTGGTGGGCCGCAACCAGAACCTCTCCAGGGAGGTGGC
GCTGCCGGCGGCGGCTGCCGACGGCGTTCCGTCGCTGCCGTGCCCTGGCGGTGGCGCGTCTTCATGACGAGGGAACCTCTGCCCT **TGCTTCATCACGC**
tRNA GAAGACAATCCAGCTGATCCGCCCTGGGCTCTGTGTGAATTATGCGATCGACCCCTGCACGGCTGACCACGACAAGCGGCCAC
GACCTCTTCTGGACGGAAAAAAAGATCACCGGATCTGCAATGCCGGTGCAGCGCAAATTGCGTACCAACTGCACGCTGCTCGATAACCCACTGGCGTC
GCTCGGTGCTACCTTCACCCCGAAGGGGAGTACGTGGCGTTCAAGACGTCGTCGGTGGCTCGTGAAG **CCC** Padrões específicos **TCCACA**
TTGCGAGTGGGCAGGGCGCCATGCCCTACTCAAGAGGAACATGGCGGAGTTTTCTAACTGAGGGTATCGAGTGTGGAAAGCAGCGGCCACCGTGGAGCT
CGACGTGAGGGAGCTCGGGAGTCCTCGCCACCGCACGTCACTGCTGCCGGACACGCCCTCTTTCCCTGATGTTGCGAGCGTCGCGGGAGCATGTC
TTTCATCGAGGGCGAGGGCAGGCGGGCCAGTGGTACCTCGCCACCCCTGGCGAGGCGGTTACAAGGCTCGTGAAGGACTGGGCTACGCGATGCC
AGCCTCACATCCACGGTGCTCCTCAGCAGCGGTGAGCTCAGCGGCCCTACAAG Elementos repetitivos **GTGCGGCT**ATCCTCTCGCGGGAGGA
GCAGTTGCTGGCGGCAGCAATGTGTCTTGCGGACTTGGTGGCTGCCGGGGAAAGTCGTGGCAGAGACCGAAGTGGGTTGCACCTGCTTACACAGTC

ANOTAÇÃO



• ANOTAÇÃO

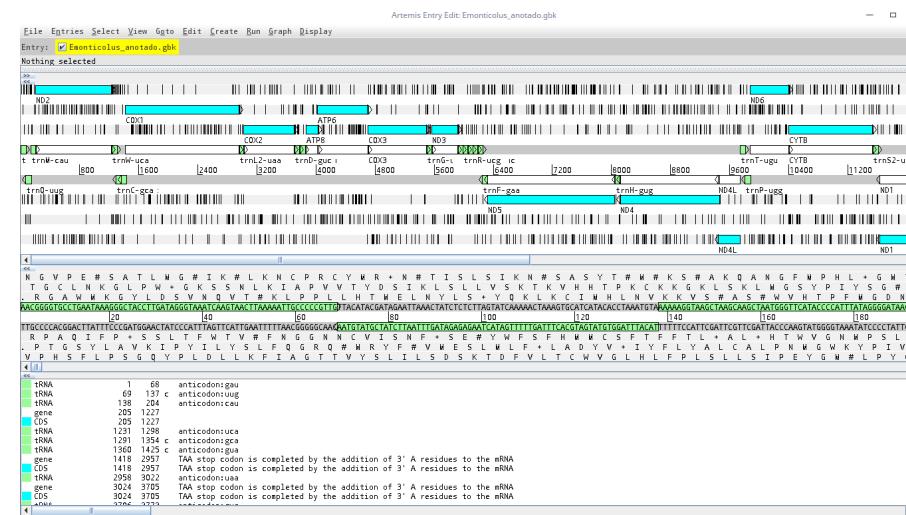
- Identificar regiões no genoma (Ex: genes)
 - Automática: MITOS Web Server
 - Manual: Artemis

MITOS WebServer

Name: Gabriel
Email: gabrieleusdeth@gmail.com
Job identifier: Pclavata_mitogenome
Genetic Code*: 05 - Invertebrate
Fasta File*: Choose File PCLAVATA_CIRCULAR.fasta
* = required Proceed >

A tutorial on how to use MITOS, including an [example](#) and the used sample data, can be found [here](#).

Advanced >



• Submissão ao GenBank

- BankIt (Plataforma Online)



- Sequência disponível online no formato genbank (.gb)

GenBank:

Header

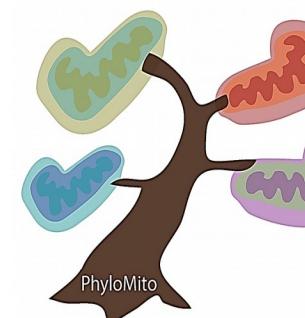
Feature Table

Sequence

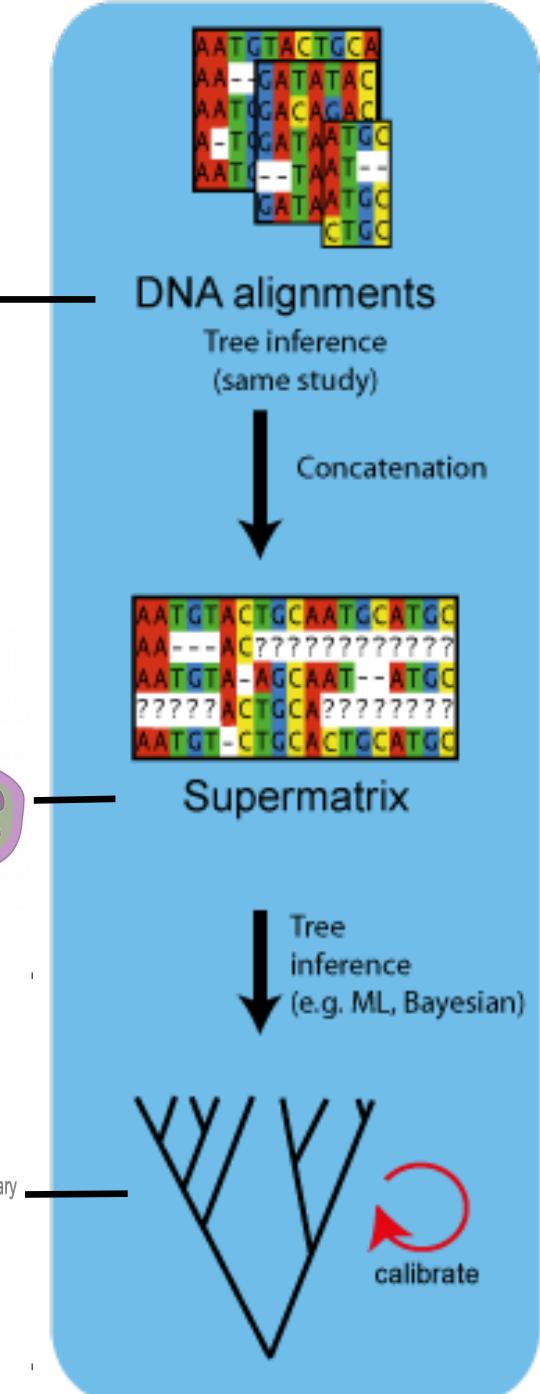
```
LOCUS KU985485          658 bp    DNA linear INV 25-OCT-2016
DEFINITION Pseudomyrmex gracilis voucher BCCISECO010109 cytochrome oxidase
subunit 1 (COI) gene, partial cds; mitochondrial.
ACCESSION KU985485
VERSION KU985485.1
KEYWORDS BARCODE.
SOURCE mitochondrial Pseudomyrmex gracilis
ORGANISM Pseudomyrmex\_gracilis
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
Pterygota; Neoptera; Holometabola; Hymenoptera; Apocrita; Aculeata;
Vespoidea; Formicidae; Pseudomyrmecinae; Pseudomyrmex.
REFERENCE 1 (bases 1 to 658)
AUTHORS Dominguez,D.F., Bustamante,M., Albuja,R.A., Castro,A., Lattke,J.E.
and Donoso,D.A.
TITLE C digos de barras (COI barcodes) para hormigas (Hymenoptera:
Formicidae) de los bosques secos del sur del Ecuador
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 658)
AUTHORS Dominguez,D.F., Bustamante,M., Albuja,R.A., Castro,A., Lattke,J.E.
and Donoso,D.A.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2016) Ciencias Agropecuarias, Universidad de
Cuenca, Av. 12 de Abril s/n, Cuenca, Azuay, Ecuador
FEATURES Location/Qualifiers
source 1..658
/organism="Pseudomyrmex gracilis"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/specimen_voucher="BCCISECO010109"
/db_xref="BOLD:DRYLO015-15\_COI-5P"
/db_xref="taxon:219809"
/country="Ecuador: Loja, Macara, Reserva Laipuna"
/lat_lon="4.21\_S 79.88\_W"
/collection_date="01-Oct-2014"
/collected_by="M. Velez, C. Gomez, M. Tuza, JE Lattke. G.
Piedra"
/identified_by="John E. Lattke"
/PCR_primers="fwd_seq: atccaaccatcataaagatattgg, rev_seq:
taaacttctggatgtccaaaaatca"
<1..>658
/gene="COI"
gene <1..>658
/gene="COI"
/codon_start=2
/transl_table=5
/product="cytochrome oxidase subunit 1"
/protein_id="AOX21864.1"
/translation="ILYFMFAMWAGMIGSSMSMIIRIELGSCGSIINNDQLYNISVTG
HAFIMIFFMVMPFMIGGFGNFLVPLMIGSPDMAYPRMNNMSFWLLPPSIMLLTLSSFI
NSGAGTGTWVYPPLSSSIFHSGASVDSLAIHAGISSIMGAINFISTIINMTHKNF
SMDKTPLMWWSILITAVLLLLSLPVLAGAITMLTDRNLNTSFFDPAGGGDPILYQHL
F"
CDS <1..>658
/gene="COI"
/codon_start=2
/transl_table=5
/product="cytochrome oxidase subunit 1"
/protein_id="AOX21864.1"
/translation="ILYFMFAMWAGMIGSSMSMIIRIELGSCGSIINNDQLYNISVTG
HAFIMIFFMVMPFMIGGFGNFLVPLMIGSPDMAYPRMNNMSFWLLPPSIMLLTLSSFI
NSGAGTGTWVYPPLSSSIFHSGASVDSLAIHAGISSIMGAINFISTIINMTHKNF
SMDKTPLMWWSILITAVLLLLSLPVLAGAITMLTDRNLNTSFFDPAGGGDPILYQHL
F"
ORIGIN
1 aattctatac ttatattttg ctatatggc aggtataatc ggatcatca taagaataat
61 tattcgaaatt gagtttagat catggatc tattataat aacgaccaac tataacaatc
121 tatcgtaaaca ggacatgcat ttatataat ttctttata gtatataat ttataatcg
181 aggattttgt aactttctag ttccactaat aattggatca ccagatataat cttacccctg
241 tataaataac aaagatttt gacttttacc cccatcaatt atacttc a ctttaagaag
301 atttataatc tcaggagctg gaactggctg aacagtatat cctcccttat ttcaagaat
361 ttttcatagg ggagcctcag tagattttg aattttctt ctacatattt caggaattt
421 atcaatcattt ggacgttata acttcatctc tacaattt aatataaccc ataaaaattt
481 ttaaatagac aaaactctt taatagtctg atcttattttt attacagcag ttctactact
541 tcttccccctt ctgttcttag ccggagcaat tacaatacta ttaacagatc gaaatcttaa
601 tacttcattt ttgaccacag caggtggagg ggacccaatt ctttaccaac actttattt
//
```

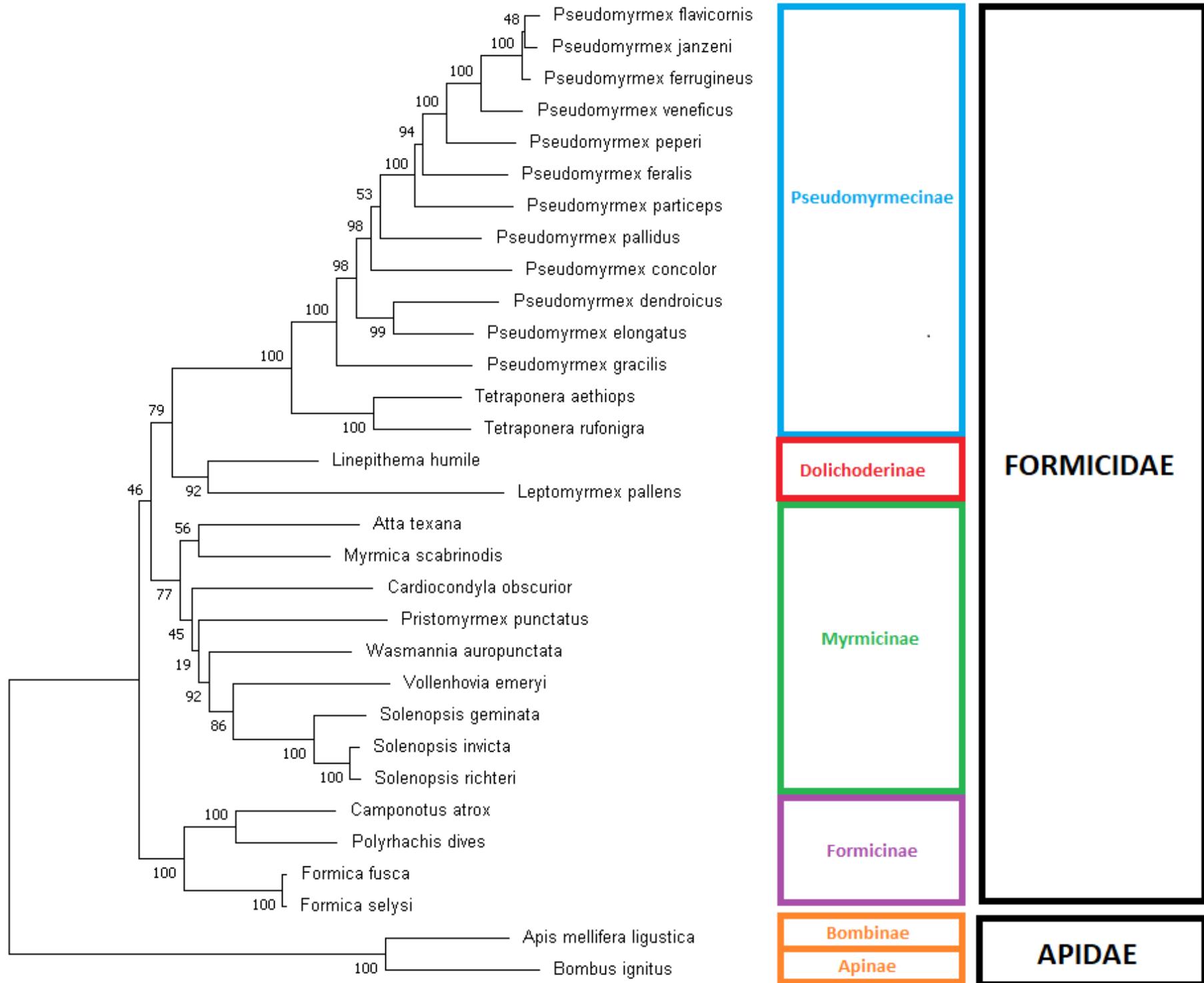
Metodologia

- 29 mitogenomas de formiga
 - 14 novos
 - 15 já descritos
 - +2 outgroups (abelhas)
- Supermatriz:
 - 13 genes mitocondriais
 - Alinhados e concatenados
- Inferência filogenômica
 - Máxima verossimilhança
 - Reamostragem: Bootstrap (500 repetições)



M E G A
Molecular Evolutionary
Genetics Analysis





Pseudomyrmecinae



- Grande variação do tamanho da sequência:
 - 15704 (*Pseudomyrmex gracilis*) até 18835 (*Pseudomyrmex feralis*)
- Mesma sintenia da subfamília Dolichoderinae:
 - Clados evolutivamente próximos



100 mitocôndrias – Classe Insecta

- Conversa informal – é factível?
- Muitos dados para o grupo
- Diptera + Coleoptera
 - Dados para 155 espécies

✓

NCBI Taxonomy Browser

Search for: Insecta

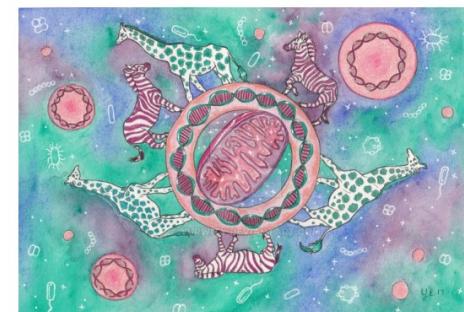
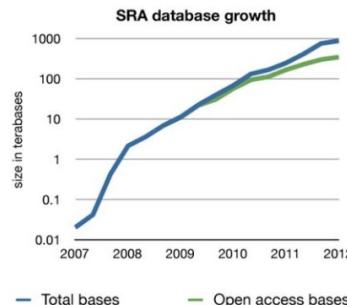
Display: 1 levels using filter: none

Lineage (full): cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Insecta (true insects) **110,704** Click on organism name to get more information.

- Insecta (true insects) **110,704**
- Dicondylia **110,628**
- Monocondylia **16**
- unclassified Insecta
- environmental samples **54**

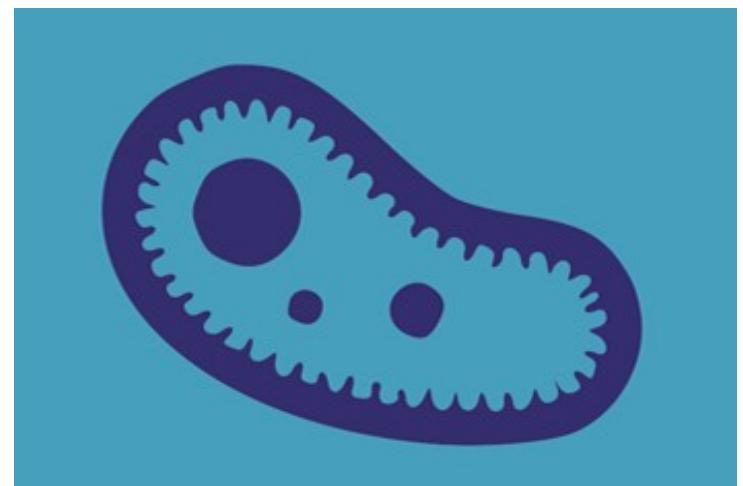
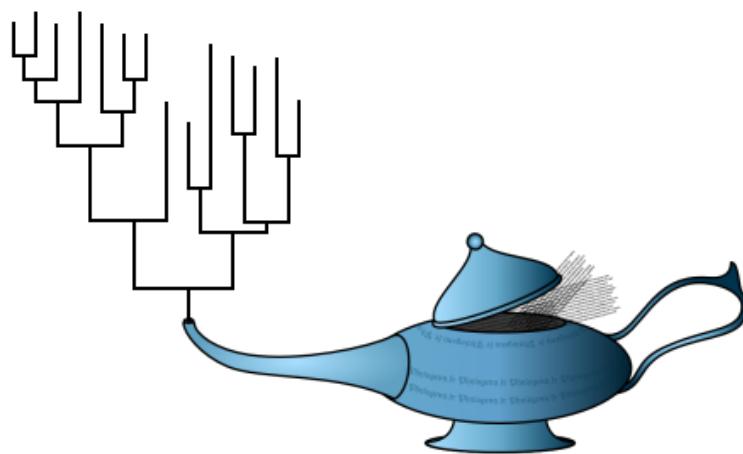
- 6 alunos de graduação
- 5 mitogenomas montados
 - 2 formigas
 - 2 gafanhotos
 - 1 gafanphoto

TUTORIAL: MONTAGEM DE MITOGENOMAS USANDO DADOS PUBLICOS



Perspectivas

- Realizar análises adicionais com as Pseudomyrmecinae
 - SNPs, heteroplasmia, mudanças estruturais, dentre outras
- Escrever manuscrito relatando nossos resultados
- Avançar com o projeto das 100 mitocôndrias
- MitoFree: programa que automatize o processo



OBRIGADO!

