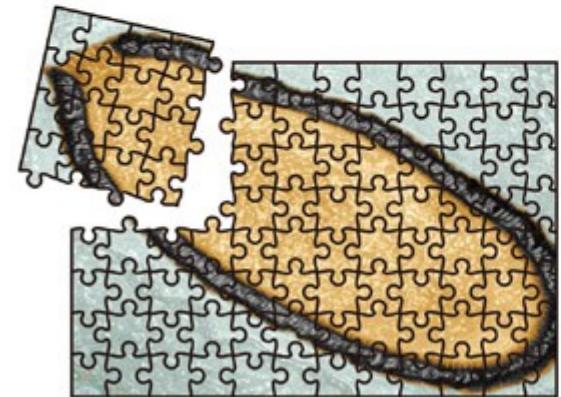
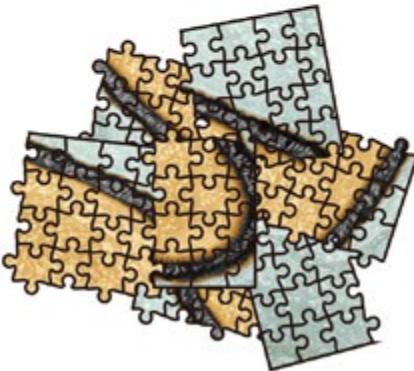
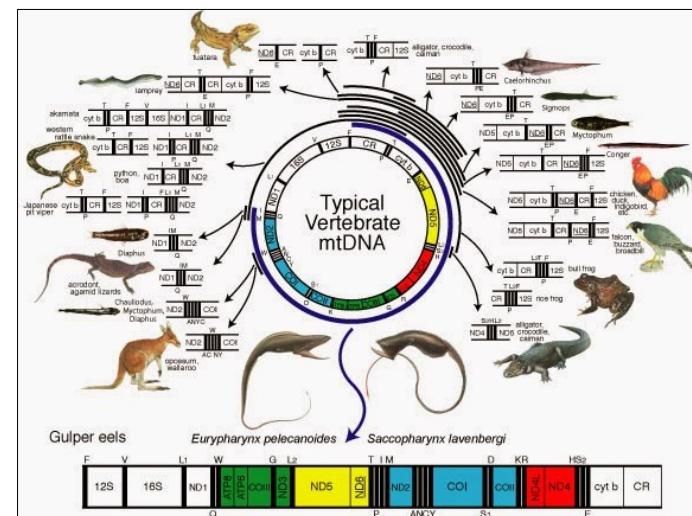


Do lixo ao luxo: o uso de dados públicos de sequenciamento para a montagem de mitogenomas

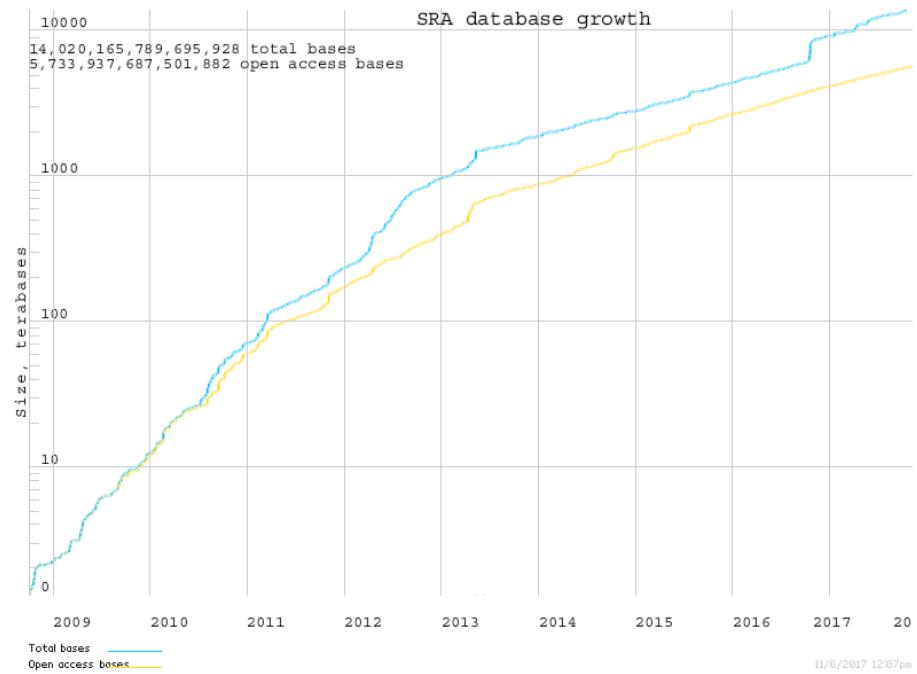


Gabriel Alves Vieira
Francisco Prosdocimi



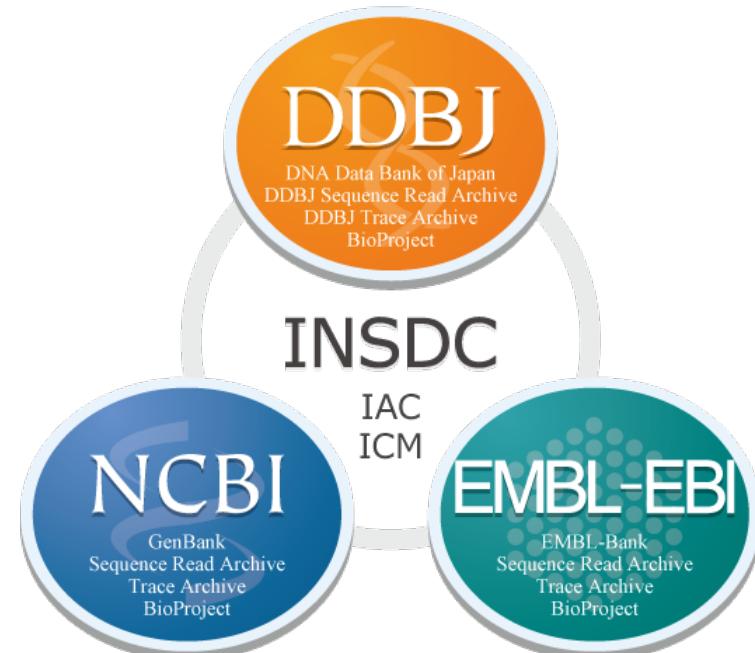
International Nucleotide Sequence Database Collaboration

- 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



O LIXO DE UNS...

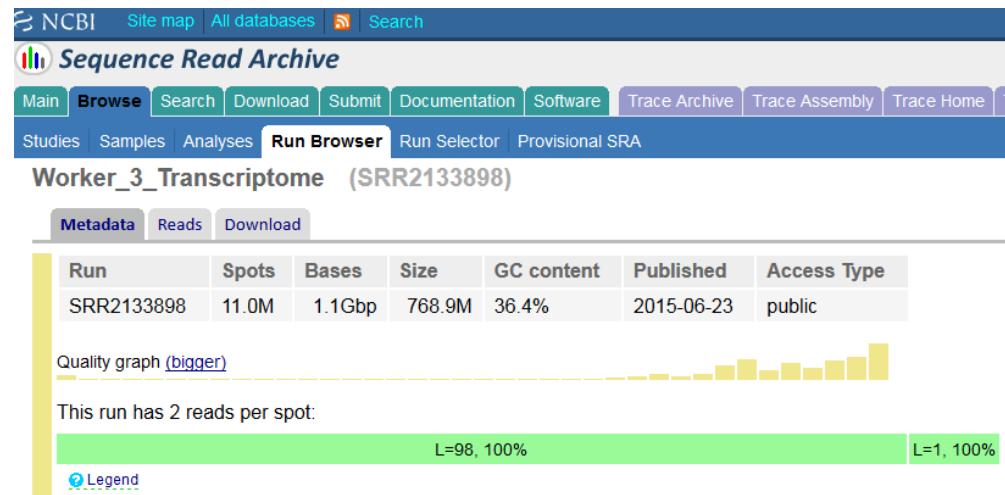
- Dados usados para responder uma pergunta

- Nada a ver com a mitocôndria

- Sequenciamento

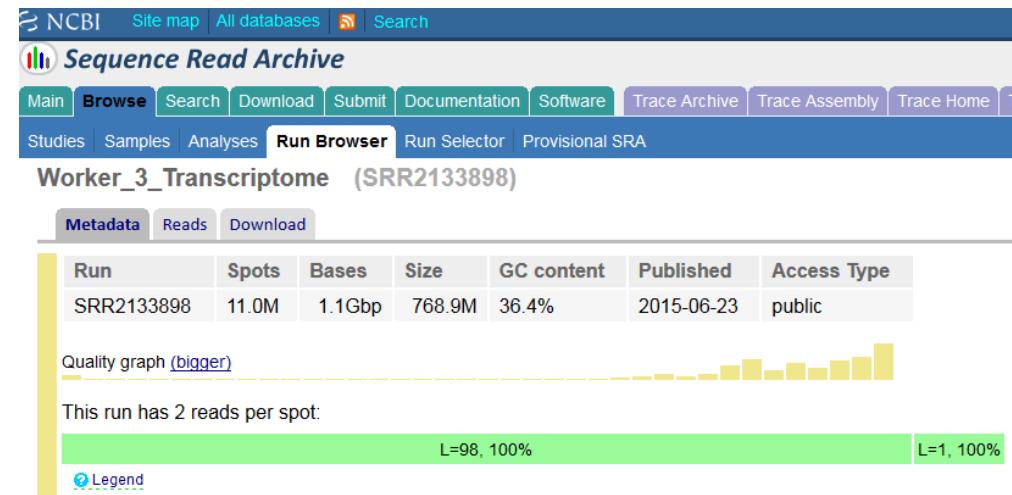
- Não separa a mitocôndria
 - 0,05% dos dados é mitocondrial
 - Espécies com dados públicos e sem mitogenoma descrito

- Dá pra usar esses dados para obter o mitogenoma?



O LIXO DE UNS...

- 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
- Dá pra usar esses dados para obter o mitogenoma?

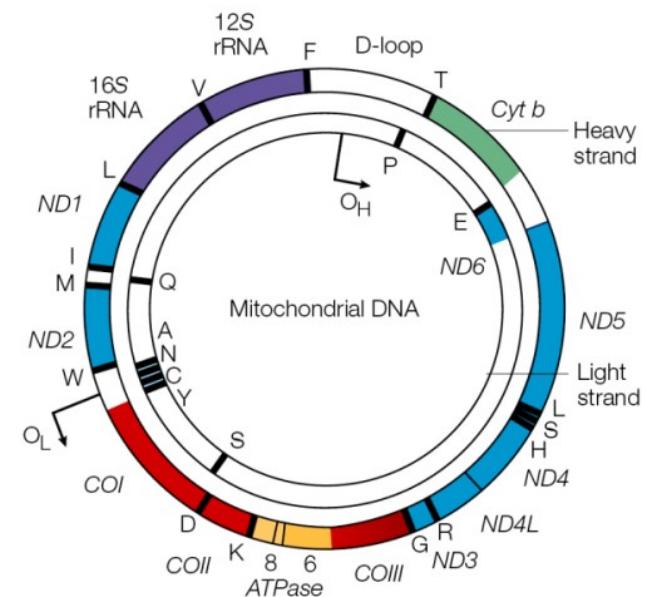


...O LUXO DE OUTROS

- Mitogenomas
 - Sem recombinação
 - Herança materna

Filogenia e
Filogeografia

- LAMPADA
 - Filogenômica mitocondrial
 - 13 genes codificadores de proteínas
 - Elaboração de ferramentas específicas
 - ↓ tempo gasto - montagem e análise
 - 30+ mitogenomas
 - Aves, moluscos, insetos, dentre outros
 - Vários projetos No-Budget



Mitochondrial Phylogenomics in Animal Models

Francisco Prosdocimi, Marcela Uliano-Silva, Igor Rodrigues da Costa, Nicholas Costa Barroso Lima

mitoMaker assembler



Automatic assembly and annotation of whole mitochondria

\$> DeNovo assembly using 5 different kmers
\$> Reference-based assembly
\$> Automatic annotation
\$> GenBank annotated output

mt-dowloader



Download all mitochondria from a taxID

\$> Finds a taxID on NCBI
\$> Searches for whole mitochondria on RefSeq
\$> Download all mitochondria in GB format

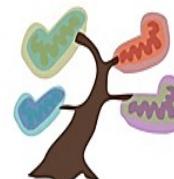
mitoCompare



Mitochondria comparison and reports

\$> Multiple sequence alignment
\$> Definition of a query sequence
\$> Amino acid mutations tabulated

phyloMito



Mitochondria phylogenomics analyses

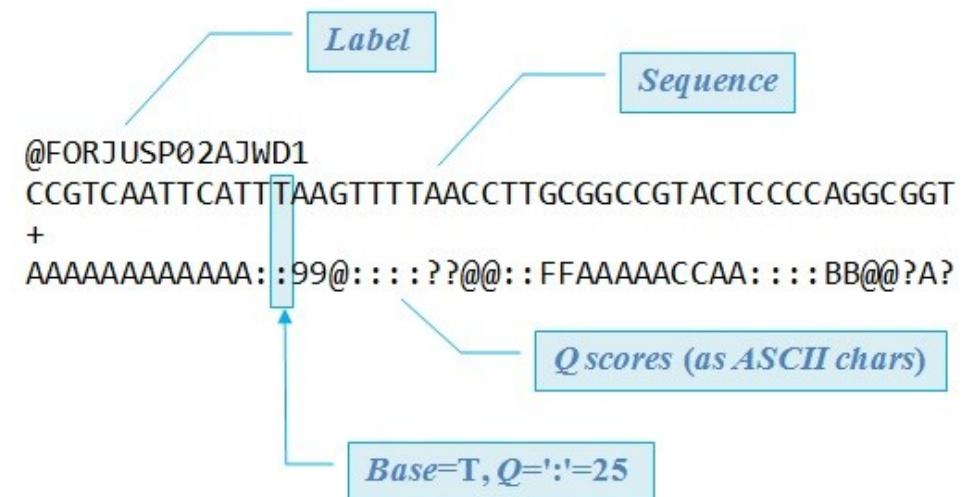
\$> Alignment of amino acids from each gene
\$> Retrotranslation to nucleotides
\$> Concatenation of gene sequences
\$> Maximum likelihood analysis by supermatrix

Obtenção dos dados

- Arquivos *.sra*
 - Utilizado pelos 3 bancos
 - Disponíveis para download direto (ERA e DRA)
 - Precisa ser convertido (SRA Toolkit) para *.fastq*

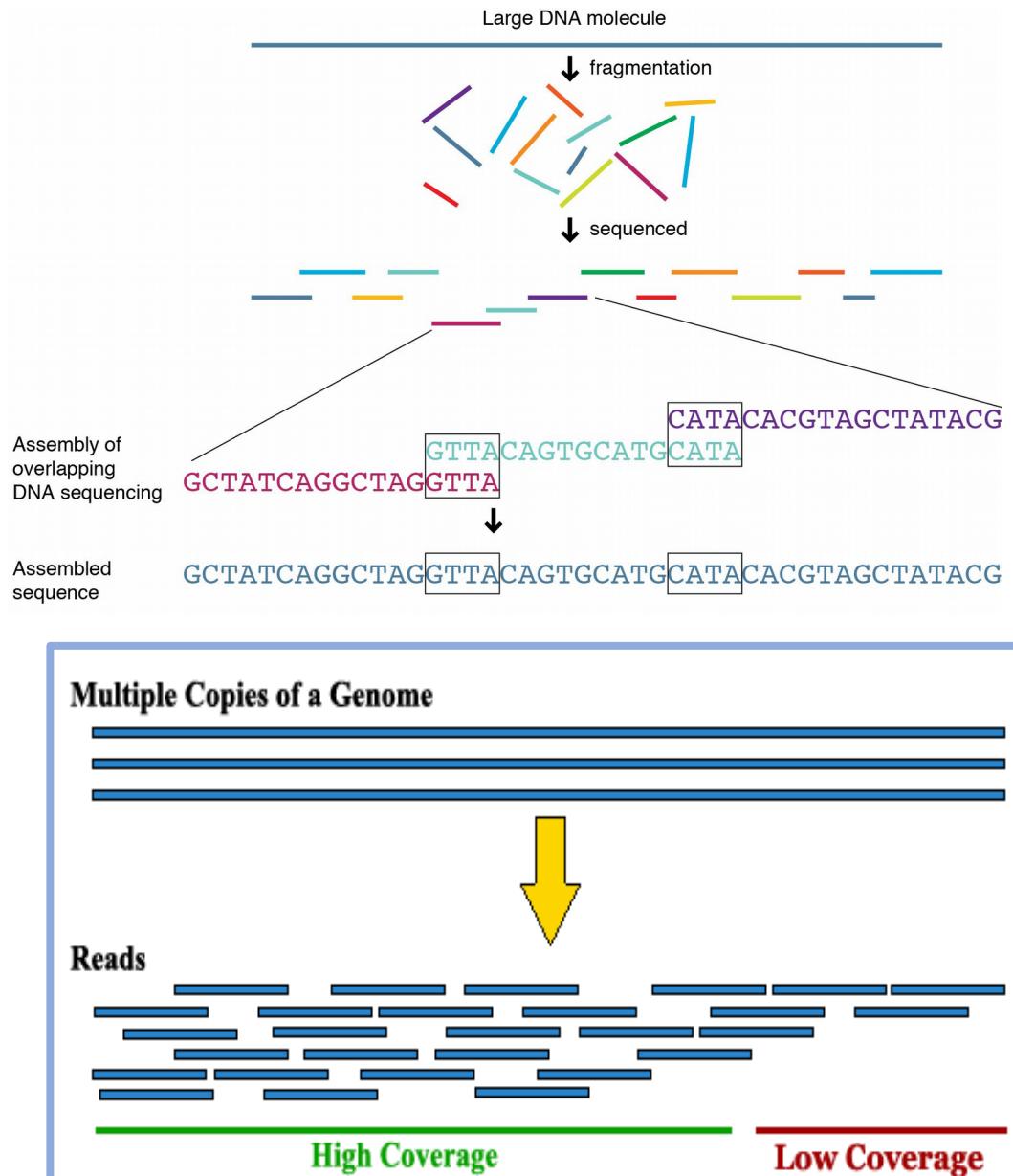
Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	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*
 - Cabeçalho, sequência e qualidade
 - Aceitos pela maioria dos programas



MONTAGEM

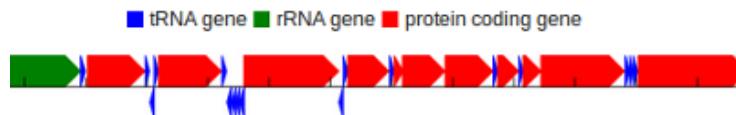
- Quebra da molécula de DNA
- Obtenção de pequenas sequências
- Quebra-cabeça – Ferramentas computacionais
 - NOVOPlasty, MIRA, MITObim
- Checagem (TABLET)
 - Cobertura
 - Quantas sequências corroboram a montagem
 - Circularização



ANOTAÇÃO E FILOGENÔMICA

- ANOTAÇÃO

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



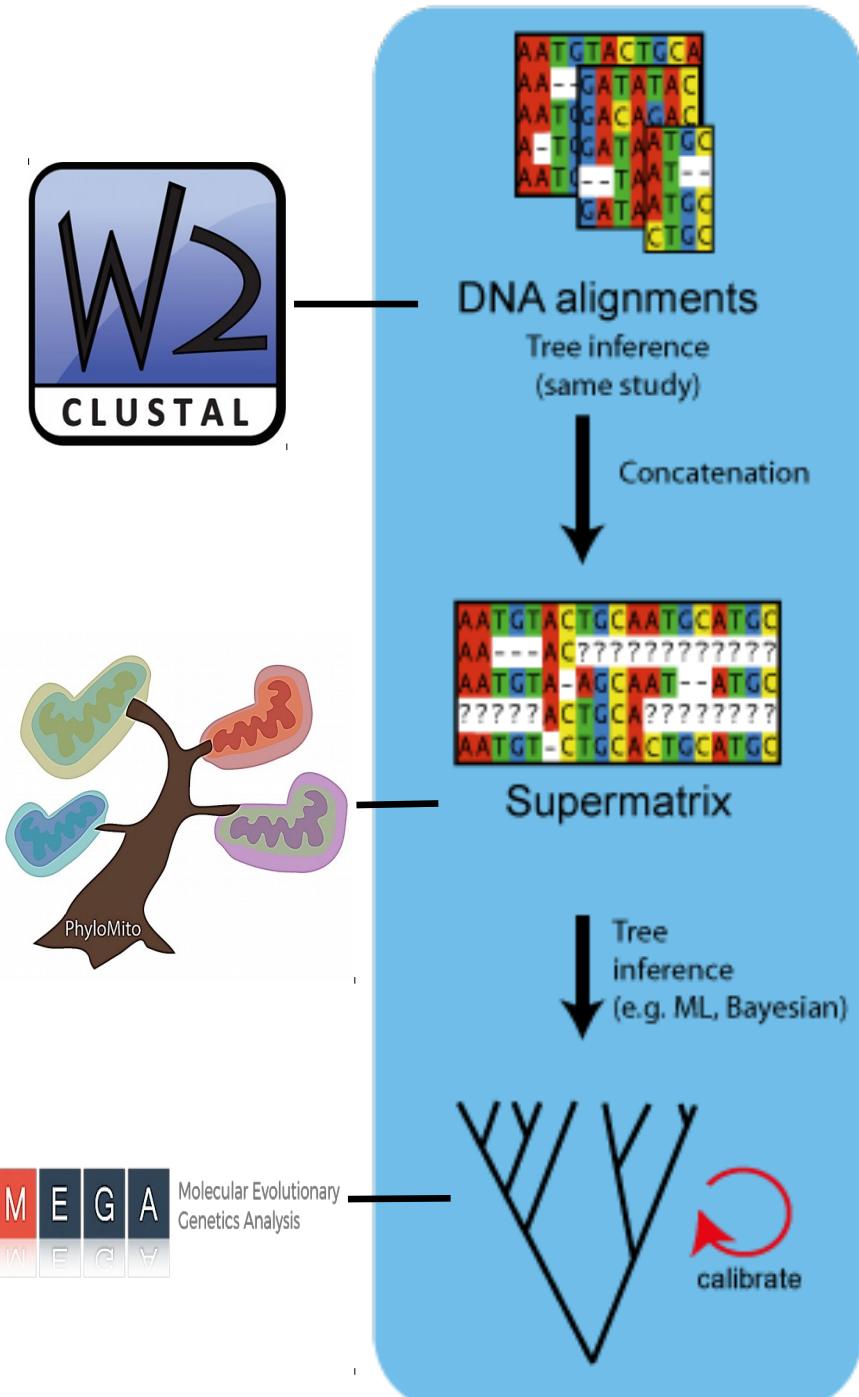
- Submissão ao GenBank

- BankIt (Plataforma Online)



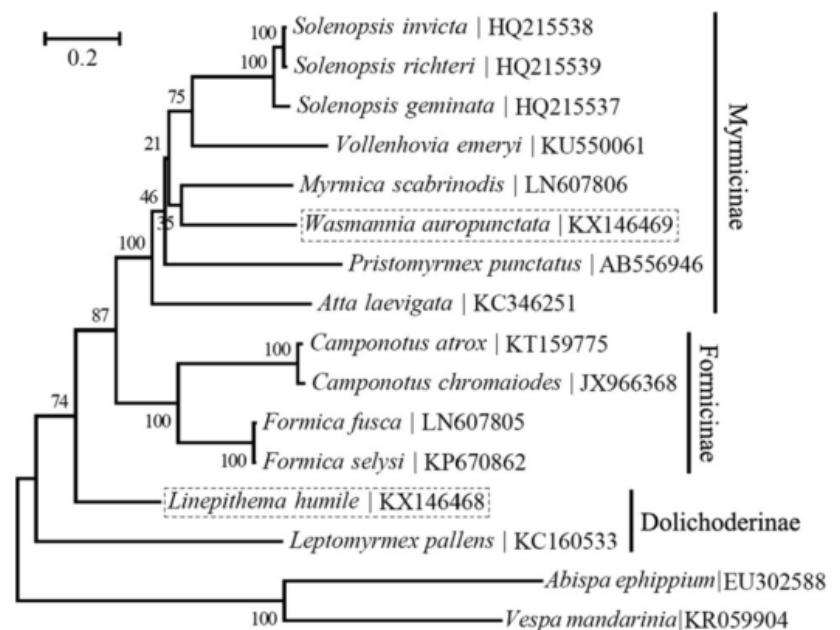
- Filogenômica

- Abordagem de supermatriz



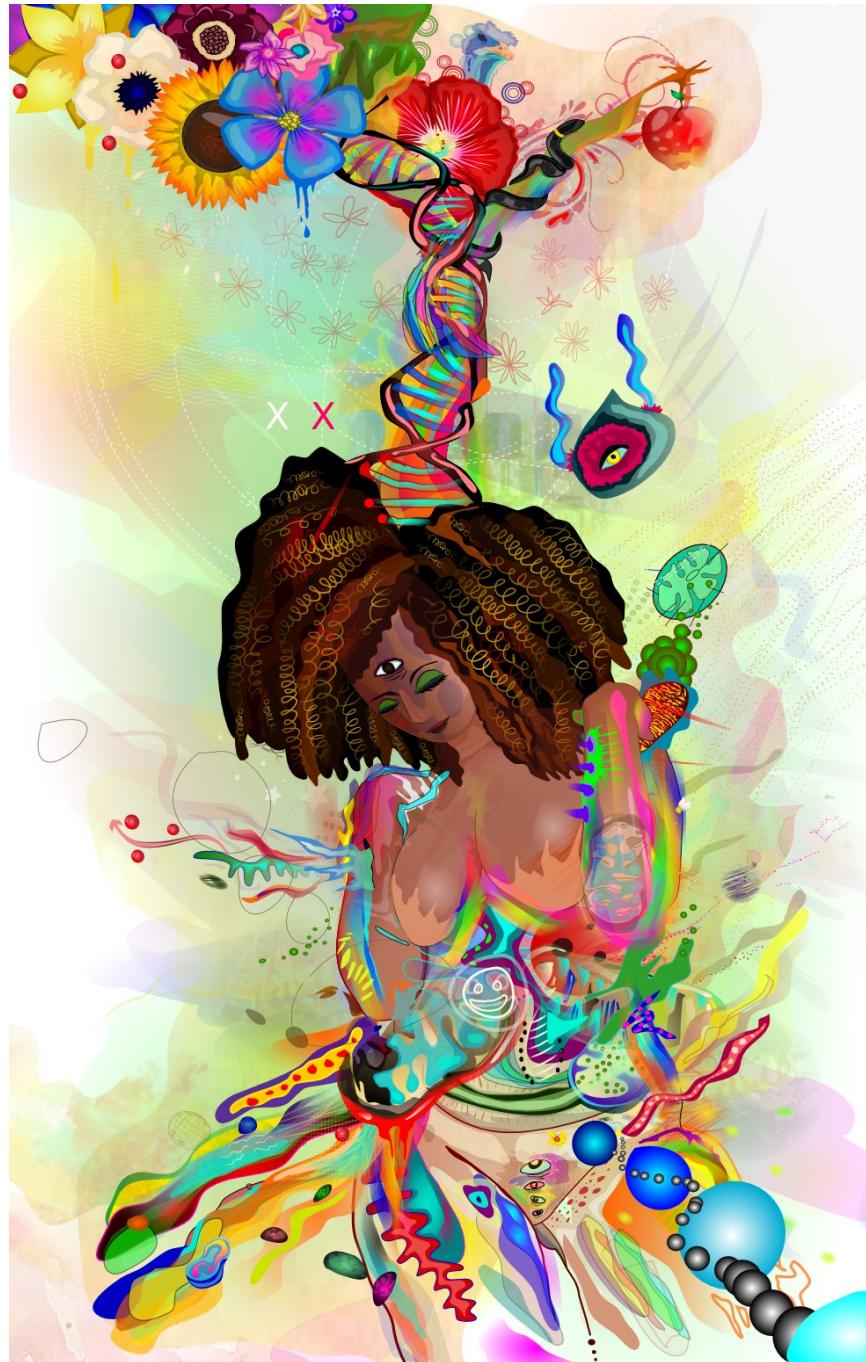
FORMIGAS

- 13262 espécies
- ≈ 3.9 Terabytes
- 14 mitogenomas
- 13 mitogenomas submetidos
 - Softwares de código aberto
 - PC mediano
 - Gasto zero
- Vários outros grupos – Informação mitogenômica subaproveitada



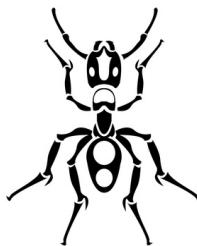
PROJETO MITOGENOMAS

- Obter genomas mitocondriais usando dados públicos
- Descobrir a história evolutiva contada pela mitocôndria



PROJETO MITOGENOMAS

- FORMIGAS



Stephanie

Juan

Vitor

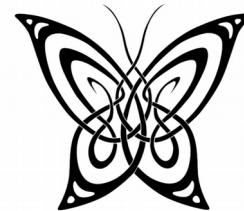
Pedro

- LAGARTOS



Gabriela

- BORBOLETAS



Vanessa

- TARTARUGAS



Ana

Deise

PROJETO MITOGENOMAS

- Muitos dados
- Poucas pessoas



- WE WANT YOU!
 - Dados públicos
 - Seus próprios dados

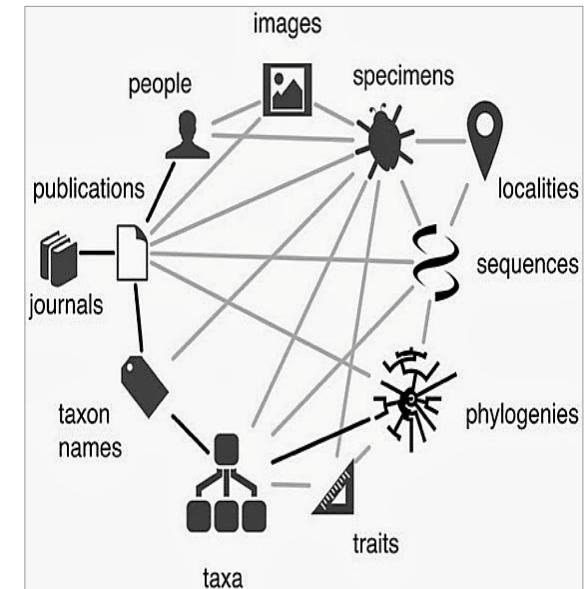


- PROJETO CLOROPLASTOS



EM SUMA

- Muitos dados disponíveis
- Potencial enorme para responder novas perguntas
- Sempre disponibilizem seus dados
 - Reprodutibilidade
 - Avanço da ciência



SRA Submission Quick Start

- [Accepted data](#)
- [Submission steps](#)
- [Human data](#)
- [Metagenomic data](#)
- [Options for depositing and updating the SRA data](#)
- [Submission Portal \(SP\) vs. Sequence Read Archive \(SRA\)](#)
- [Start submitting](#)
- [Updating the SRA data \(study or submission\)](#)
- [Contact SRA staff](#)

Accepted data

The SRA accepts **genetic data** and the associated quality scores produced by next generation sequencing technologies. Please refer to our [File Format Guide](#).

Submission steps

1. [Login to or sign up for an NCBI account](#)
2. Register your project and biological samples:
 - Register your **project** at [BioProject database](#)
 - Register your **biological samples** at [BioSample database](#)
3. Create your SRA data submission and upload sequence data files:
 - Submit **SRA metadata** - information that will link your project, samples/experiments and file names
 - Upload **sequence data files** in [Submission Portal](#) ([Basic FTP and Aspera usage help](#))

If you are making your submission via dbGaP or GEO databases please refer to our [dbGaP submission Guide](#) and [GEO submission Guide](#) respectively.

OBRIGADO!

