

Metagenômica

João Carlos Setubal

IQ/USP

Estudo de genomas isolados

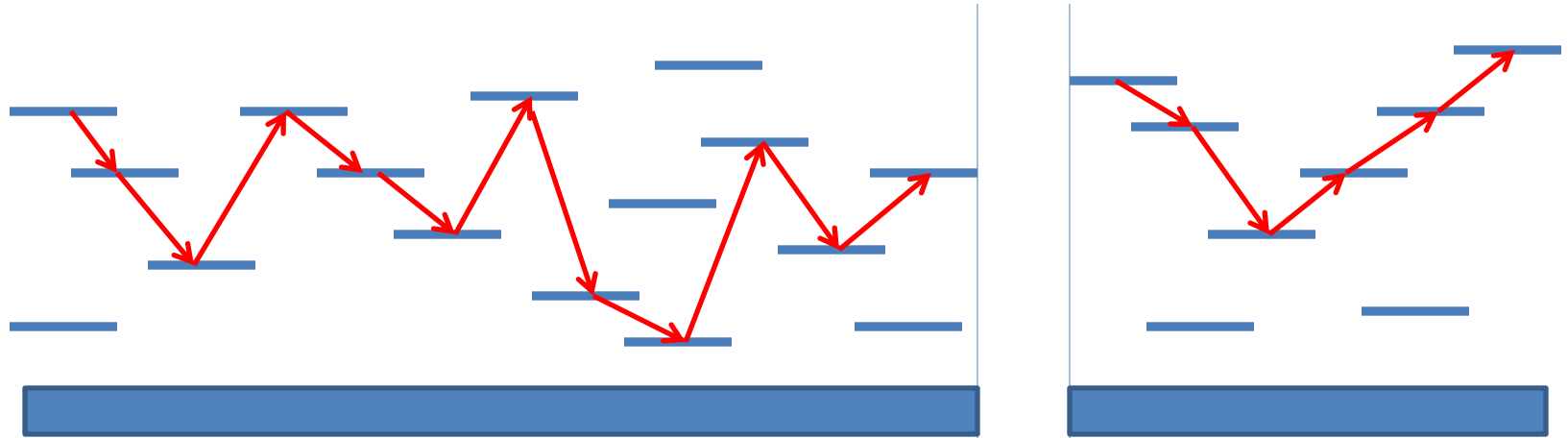
- Isolar organismo
- Cultivar
- Extrair DNA
- Fragmentar DNA
- Sequenciar os fragmentos DNA
- Montar os fragmentos → bioinformática
- Anotar a sequência final → bioinformática
- Correlacionar fenótipo com genoma
- Comparar com outros genomas

Entrada para montagem



tamanhos: 50 bp a 1000bp

Montagem



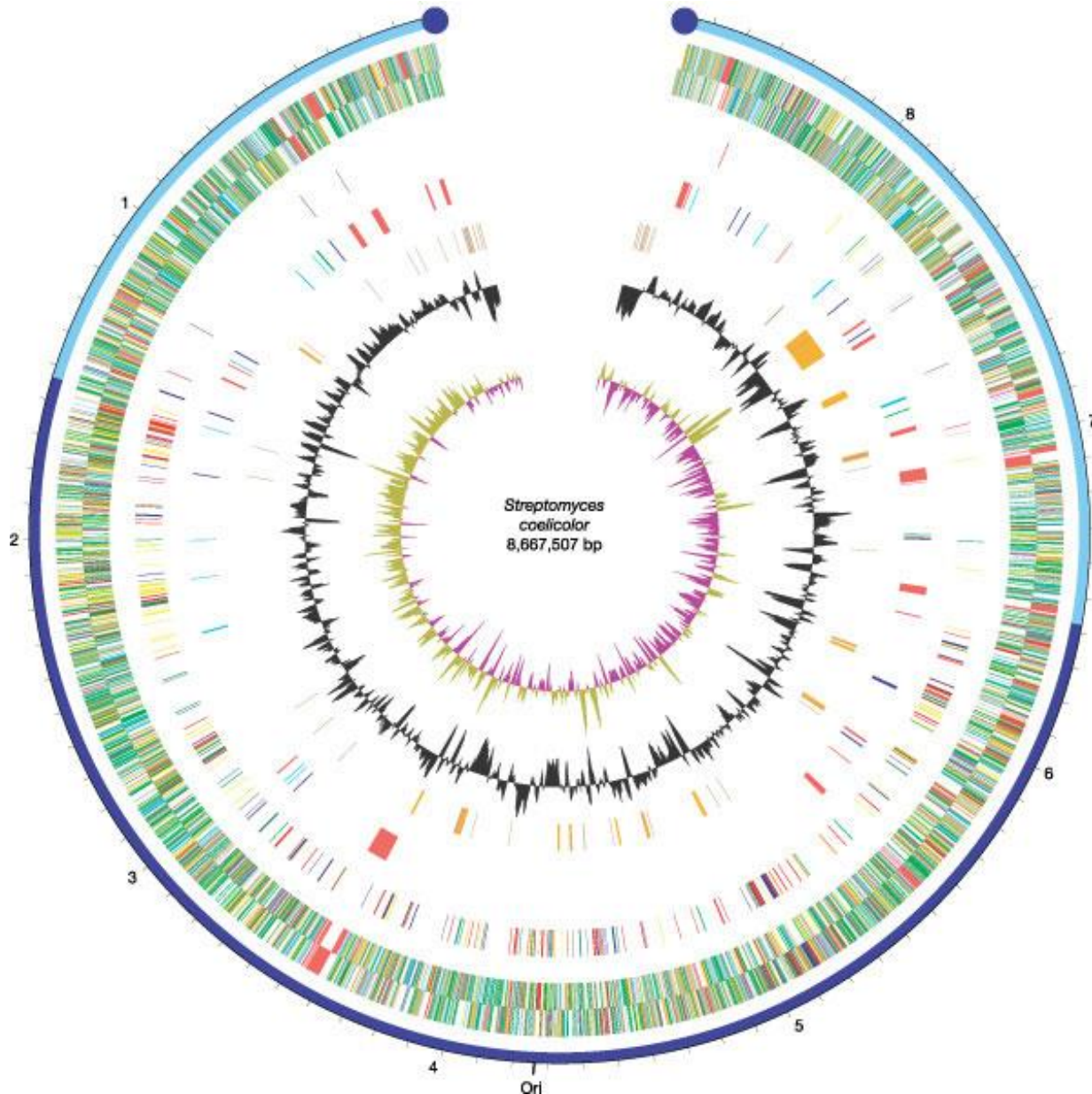
contig

```
...ACCGTAAATGGGCTGATCATGCTTAAA
      TGATCATGCTTAAACCCTGTGCATCCTACTG...
```



buraco

Resultado da anotação



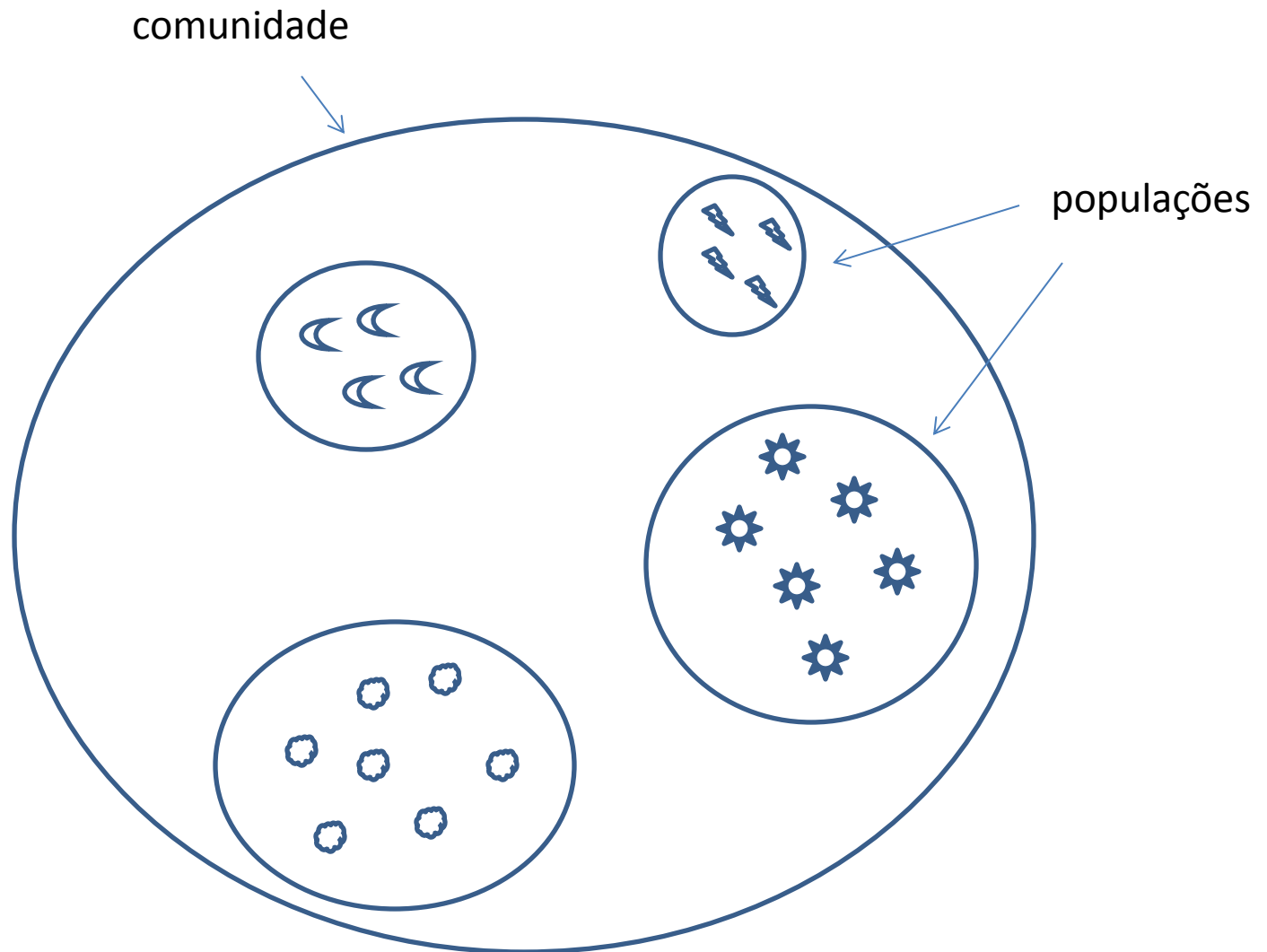
premissa

- isolamento resultou numa população de indivíduos **genomicamente idênticos**

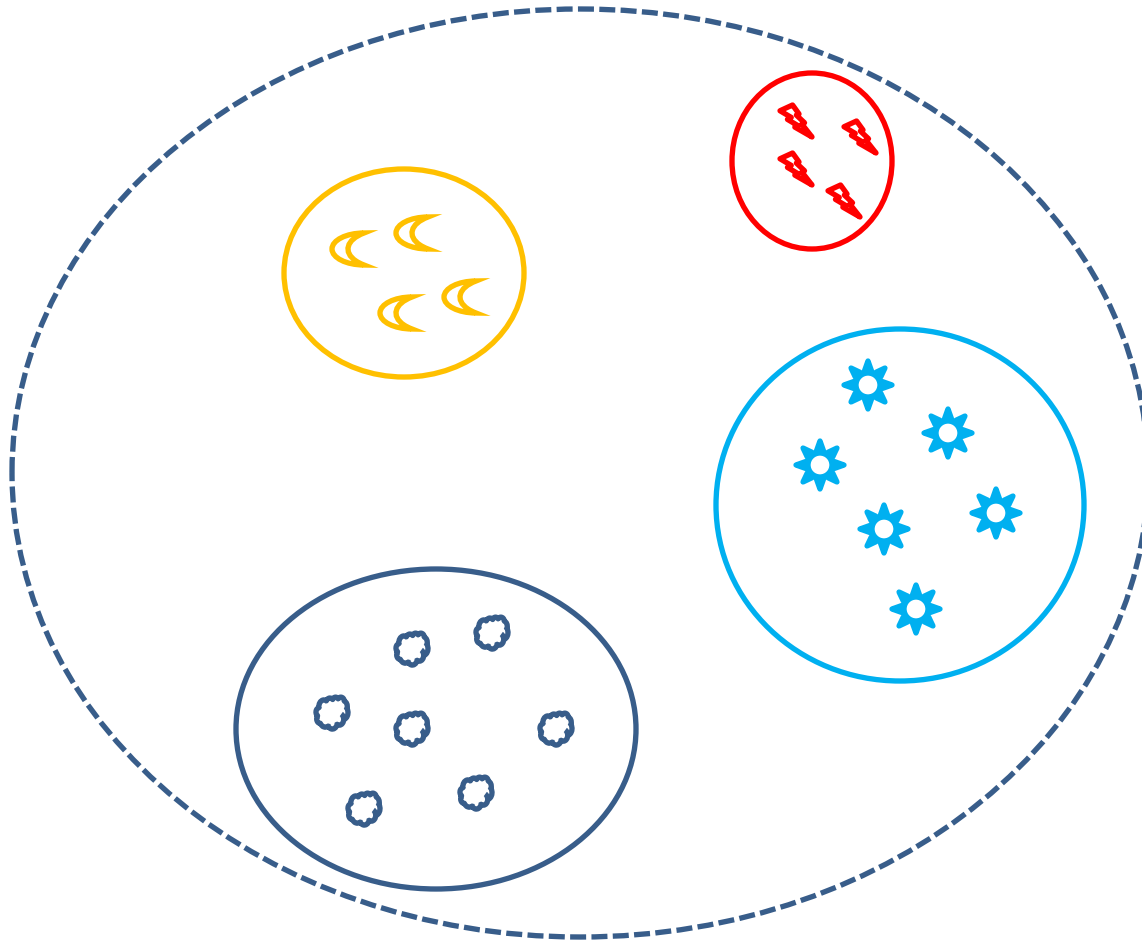
Metagenômica

- DNA é amostrado de um nicho ecológico
 - solo
 - composteira
 - pele, estômago, gengiva
 - Água do mar
- O DNA amostrado provém de vários organismos, tudo misturado

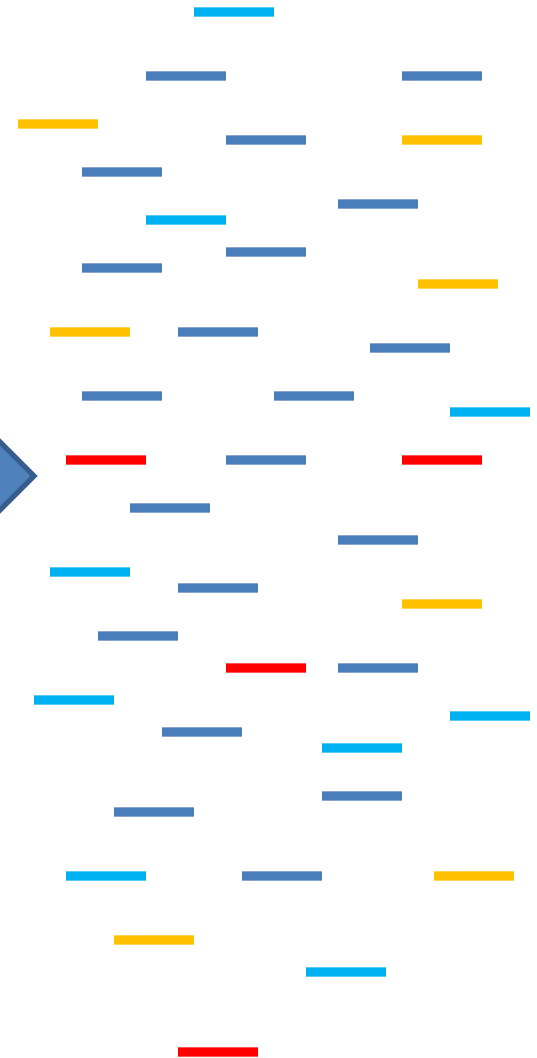
Um nicho ecológico



A comunidade



DNA



Por que metagenômica?

- Procariotos: “a maioria invisível”
 - Tanto carbono quanto plantas [Whitman et al, 1998]
 - 9 de cada 10 células do corpo humano são de procariotos
- Estão por todo lugar
 - Da atmosfera superior ao subsolo
- Enorme diversidade genética
 - organismos
 - proteínas

Sabemos cultivar menos do que
1% dos procariotos

- Metagenômica: não há necessidade de cultura

Metagenômica está na moda (microbioma humano)



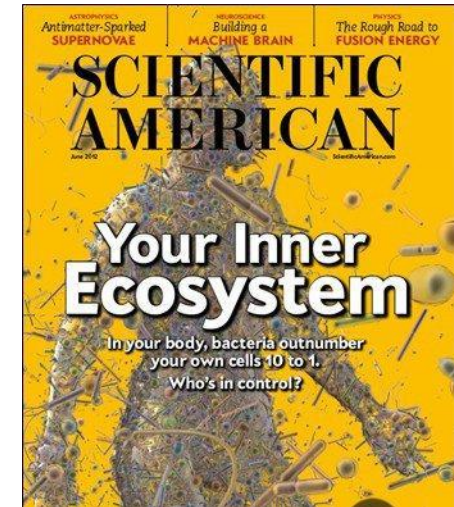
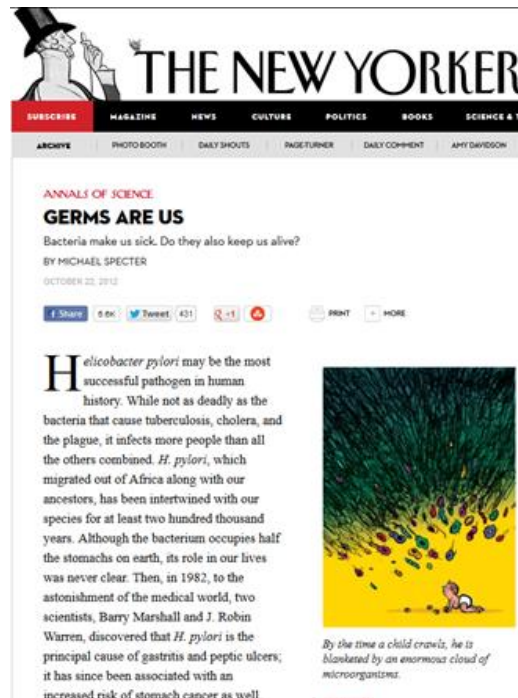
junho 2012

outubro 2012



My Microbiome and Me

Science 8 June 2012:



June 2012 Issue



maio 2013

The New York Times

Some of My Best Friends Are Germs



By MICHAEL POLLAN
Published: May 15, 2011

www.earthmicrobiome.org

[Home](#)[Defining the Tasks](#)[Getting Involved](#)[EMP Protocols and Standards](#)[Affiliations](#)[Publications](#)[Meetings](#)[EMP Logo](#)[No categories](#)

The Earth Microbiome Project is a systematic attempt to characterize the global microbial taxonomic and functional diversity for the benefit of the planet and mankind

Constructing the Microbial Biomap for Planet Earth

The Earth Microbiome Project is a proposed massively multidisciplinary effort to analyze microbial communities across the globe. The general premise is to examine microbial communities from their own perspective. Hence we propose to characterize the Earth by environmental parameter space into different biomes and then explore these using samples currently available from researchers across the globe. We will analyze 200,000 samples from these communities using metagenomics, metatranscriptomics and amplicon sequencing to produce a global Gene Atlas describing protein space, environmental metabolic models for each biome

Meetings

There are currently no EMP centric meetings planned, however we will update this space as soon as the next meeting is organized.

News

Earth Microbiome Project:
Rick Stevens at
TEDxNaperville

Uma motivação mais bioquímica

- Taq polimerase
- DNA polimerase que funciona a 70 °C
- Presente na bactéria *Thermus aquaticus*, isolada em fontes termais (Parque Yellowstone) por Thomas Brock em 1965
- Ingrediente essencial de PCR, desenvolvida em 1983

Estudo de metagenomas

- Isolar organismo 
- Cultivar 
- Extrair DNA
- Fragmentar DNA
- Sequenciar os fragmentos DNA
- Montar os fragmentos
- Anotar contigs
- Correlacionar fenótipo com metagenoma
- Comparar com outros metagenomas
- **Fazer análise de abundância**

Bioinformática é complexa

- Volume de dados: milhões de reads (NGS)
 - Baixa cobertura de genomas individuais
 - Depende da diversidade da amostra
 - erros de sequenciamento
- baixa qualidade de dados
- algoritmos precisam ser mais robustos

Questões básicas

- Que organismos estão na amostra?
- Que funções estão presentes?
- Avaliação quantitativa (abundância)
- Metagenômica comparativa
- Metadados são essenciais

Metadados



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

ALL
Genomes

Quick Genome Search:

GO



INTEGRATED MICROBIAL GENOMES
with MICROBIOME SAMPLES

IMG/M HomeFind GenomesFind GenesFind FunctionsCompare GenomesAnalysis CartMy IMG

[Home](#) > Find Genomes

[Loaded.](#)

Thermobifida fusca YX

Metadata	
Biotic Relationships	Free living
Cell Shape	Rod-shaped
Ecosystem	Host-associated
Gram Staining	Gram+
Host Name	Homo sapiens
Isolation	compost pile in the 1970's
Library Method	3Kb, 40Kb
Motility	Nonmotile
Oxygen Requirement	Aerobe
pH	4 - 10
Sporulation	Sporulating
Temperature Range	Thermophile
Temperature Optimum	50 - 55
Type Strain	No
Relevance	Bioenergy, Biofuels, Biotechnological, Human Pathogen, Medical
Phenotype	Pathogen
Habitat	Self-heated organic materials, Host
Metabolism	Cellulose degrader, Degrades organic material
Diseases	Farmer's lung, Mushroom worker's disease, Respiratory infection

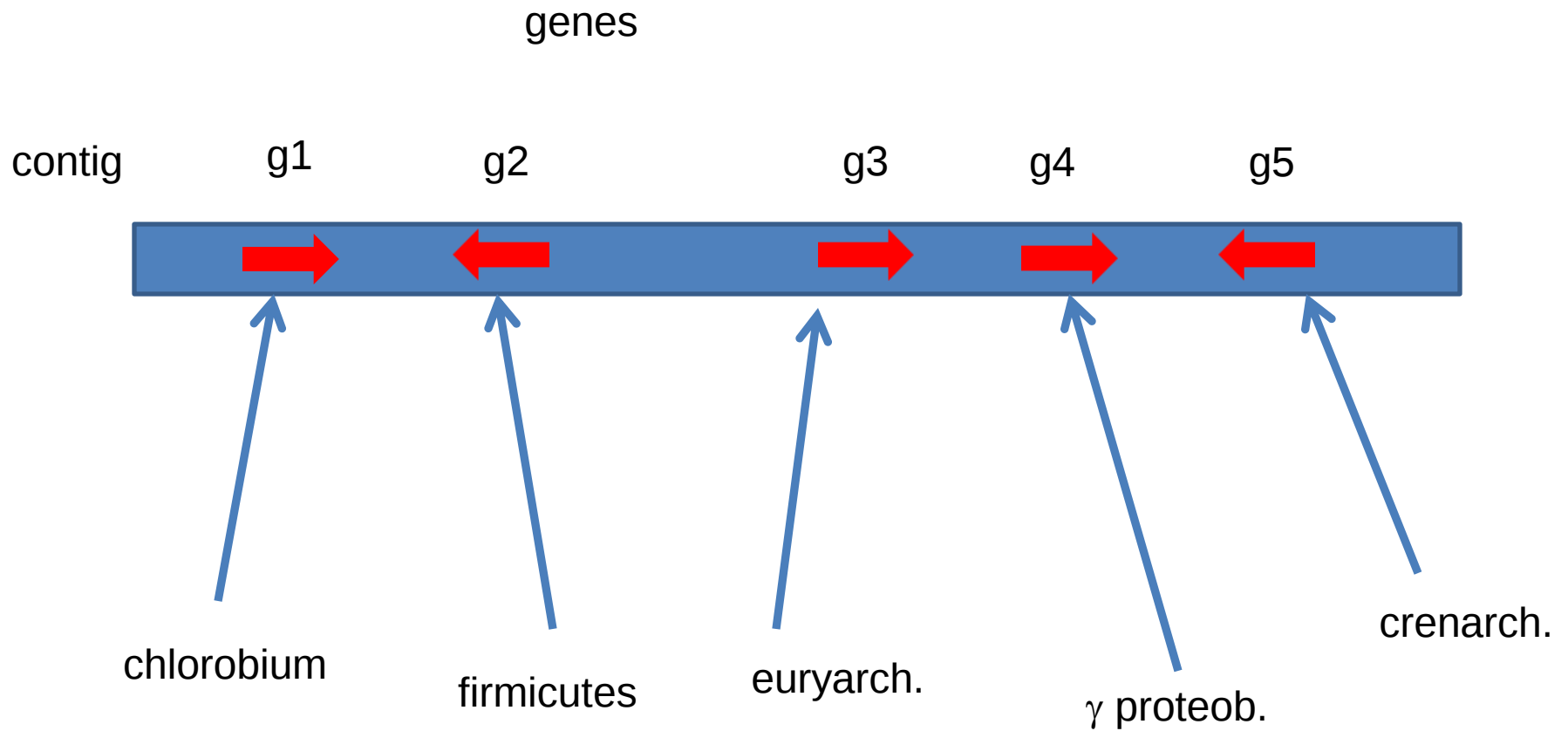
Metadados

- A **tirania da sequência** diminuiu a importância dos metadados
- Precisam ser revitalizados
- Metadados precisam ser **computáveis!**
 - Ontologia
 - Informações mínimas
- **Environmental Microbiology** [Volume 12, Issue 7](#), pages 1803–1805, July 2010
- **Nature Biotechnology** 29, 415–420 (2011) doi:10.1038/nbt.1823

Montagem

- Em genomas bacterianos isolados, é um processo razoavelmente bem compreendido
- Em metagenomas há velhas e novas dificuldades
 - Mistura de organismos
 - Quimeras
 - Transferência lateral
 - Repetições
 - Tamanho dos conjuntos de dados

Exemplo de quimerismo



Reads vs. contigs

- **Reads**: o que sai da máquina sequenciadora
- **Contigs**: resultado da montagem

Reads

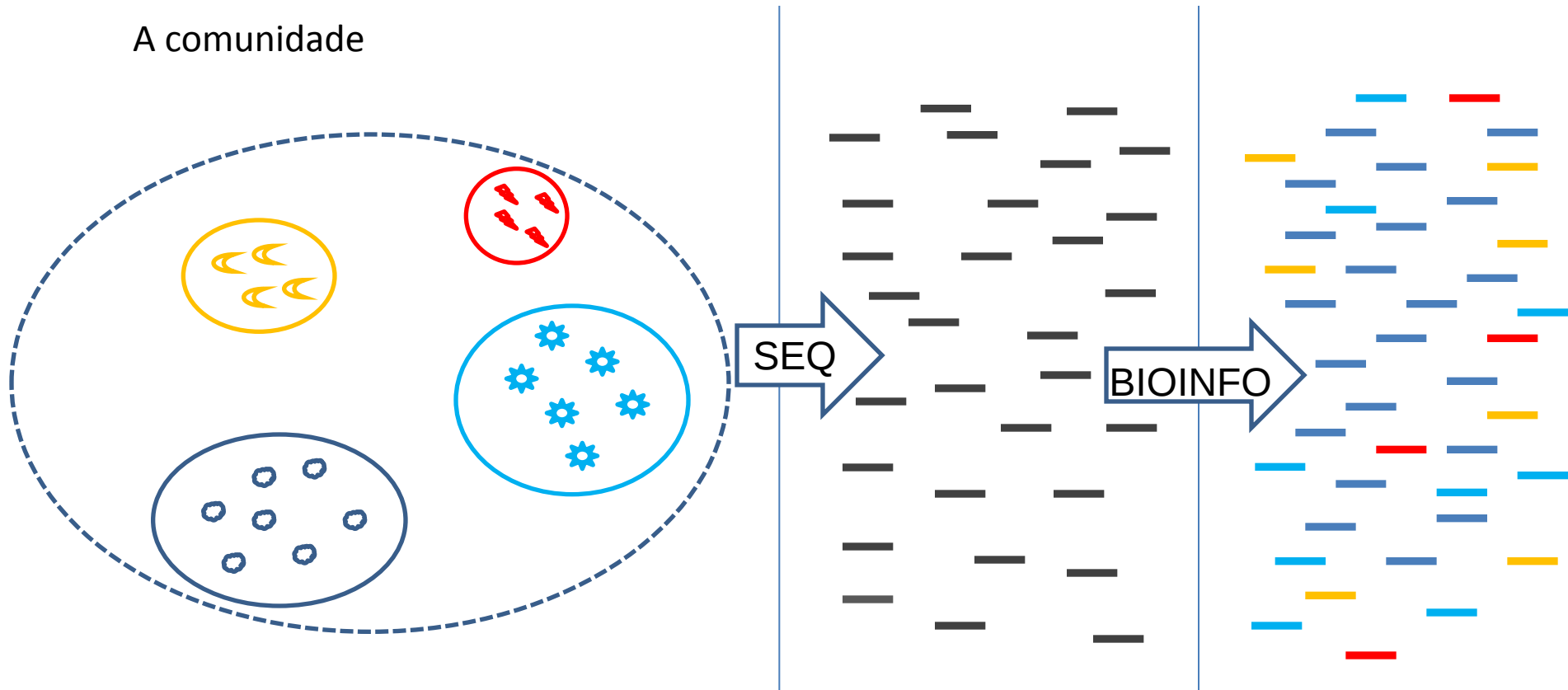
- Essencial usar reads para análises de abundância
- Também é melhor usar reads para classificação filogenética por causa do possível quimerismo

Contigs

- Bons para achar genes
 - Mais provável achar ORFs completas em contigs
- Para simples presença de genes **quimerismo não é um problema sério**

DNA

A comunidade



Classificação filogenética

- 16S
 - Primers específicos
 - Ribosomal Database Project (RDP)
 - Inclui muito mais espécies do que GenBank
- Dados de DNA total
 - Maior parte das sequências vem de genes codificadores de proteínas, mas também tem 16S
- Qual escolher depende dos objetivos do projeto

Classificação de reads de DNA total

- **Similaridade** com sequências de origem conhecida
 - BLAST
- Propriedades intrínsecas de cada sequência
 - **Assinaturas genômicas**

Classificação com base na frequência de palavras de k bases

$k = 4$: AAAA, AAAC, AAAG, AAAT, CAAA, etc...

Dada uma janela de x kb, podemos contar as ocorrências de cada uma dessas palavras dentro da janela

Exemplo:

AG**ATTA**GCGACT**ATT**A**T**AGCCTAGATCGATC**ATTA**CC

AGAT ocorre 2 vezes

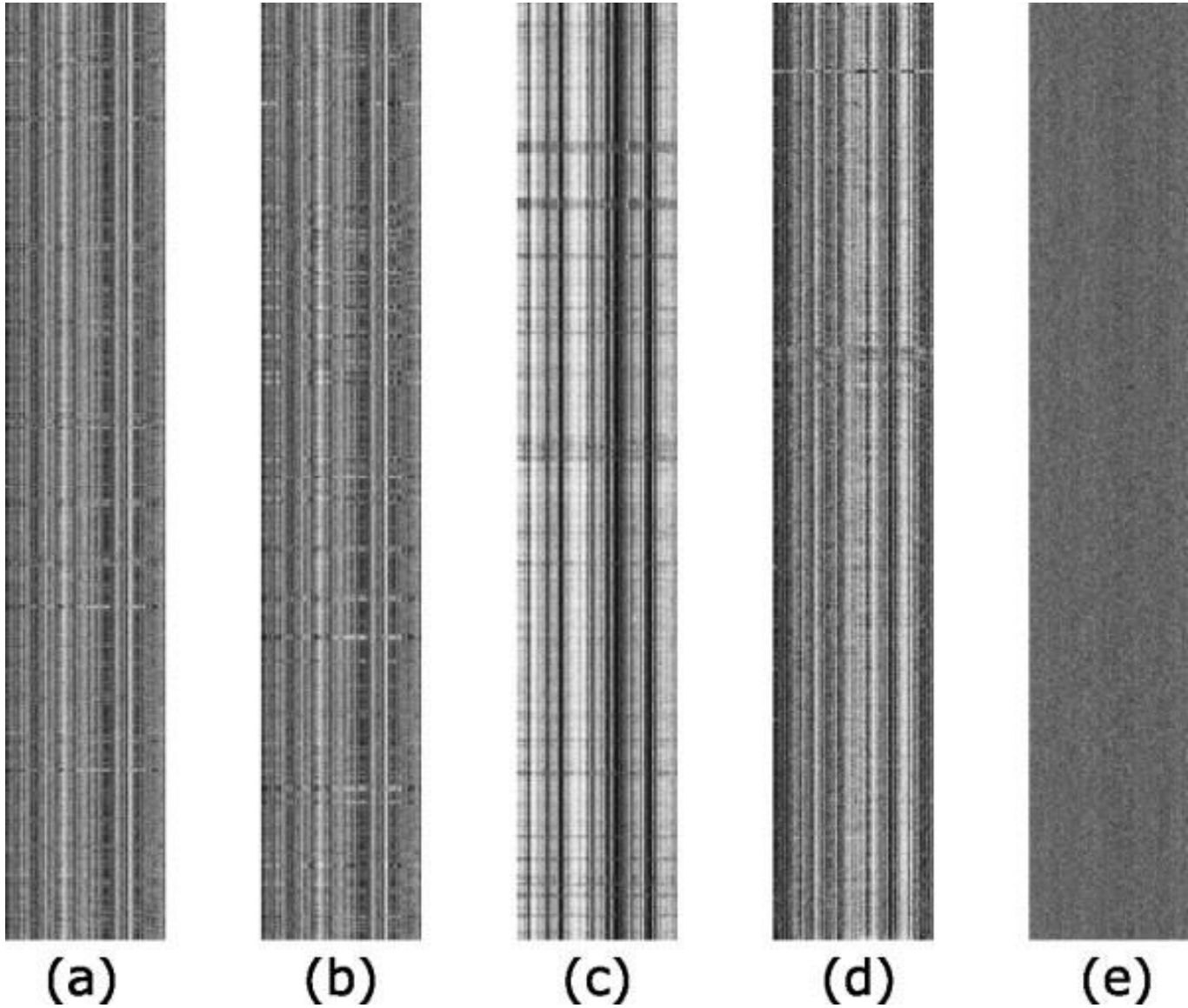
ATTA ocorre 3 vezes

etc

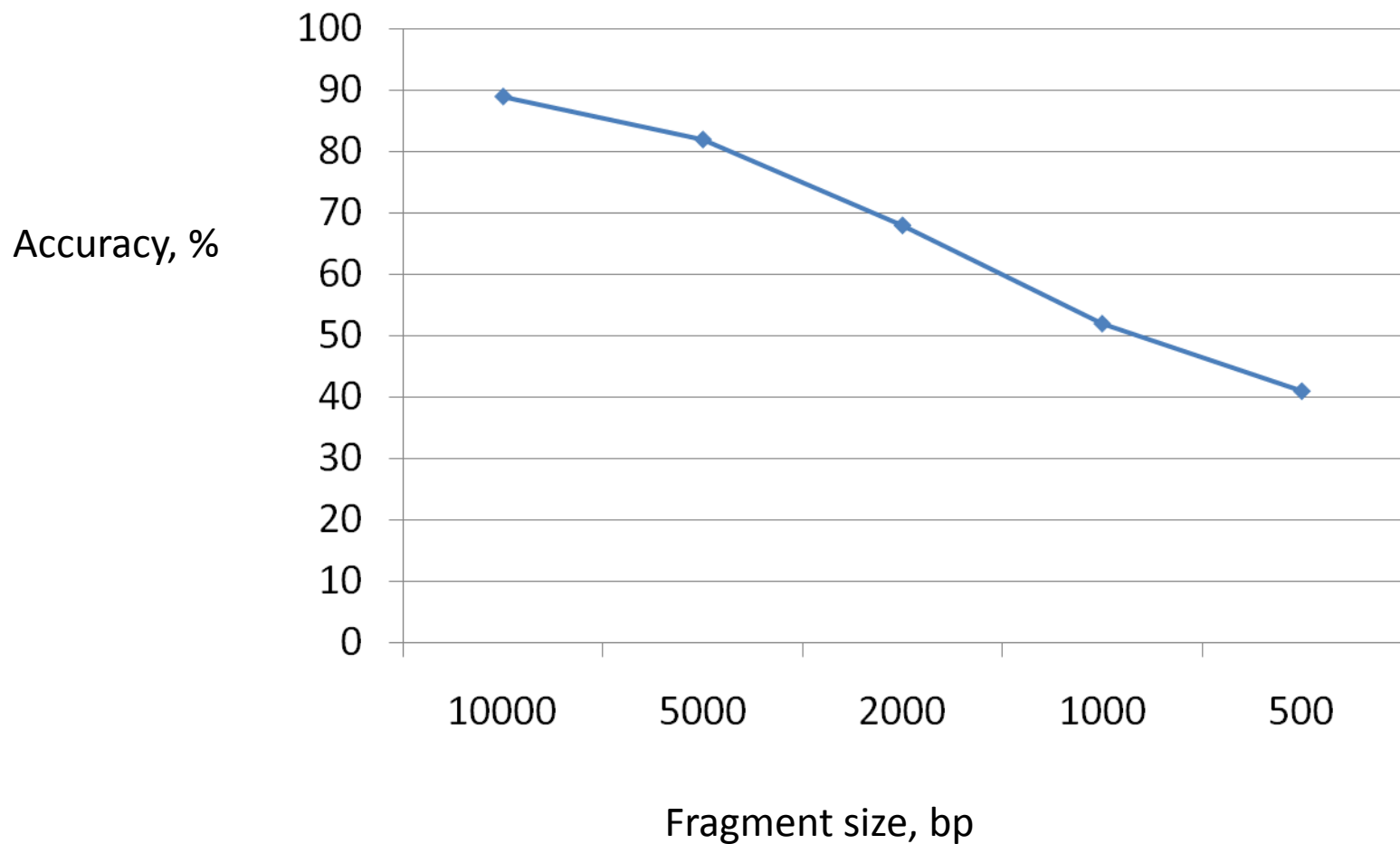
Matriz de frequências

janela	AAAA	AAAC	AAAG	AAAT	ACAA	ACAC	ACAG	ACAT
1	15	2						
2	16	3						
3	14	0						
4	13	2						
5	15	4						
6	12	0						
7	18	1						
8	17	3						
9	16	1						

Genome “barcodes”



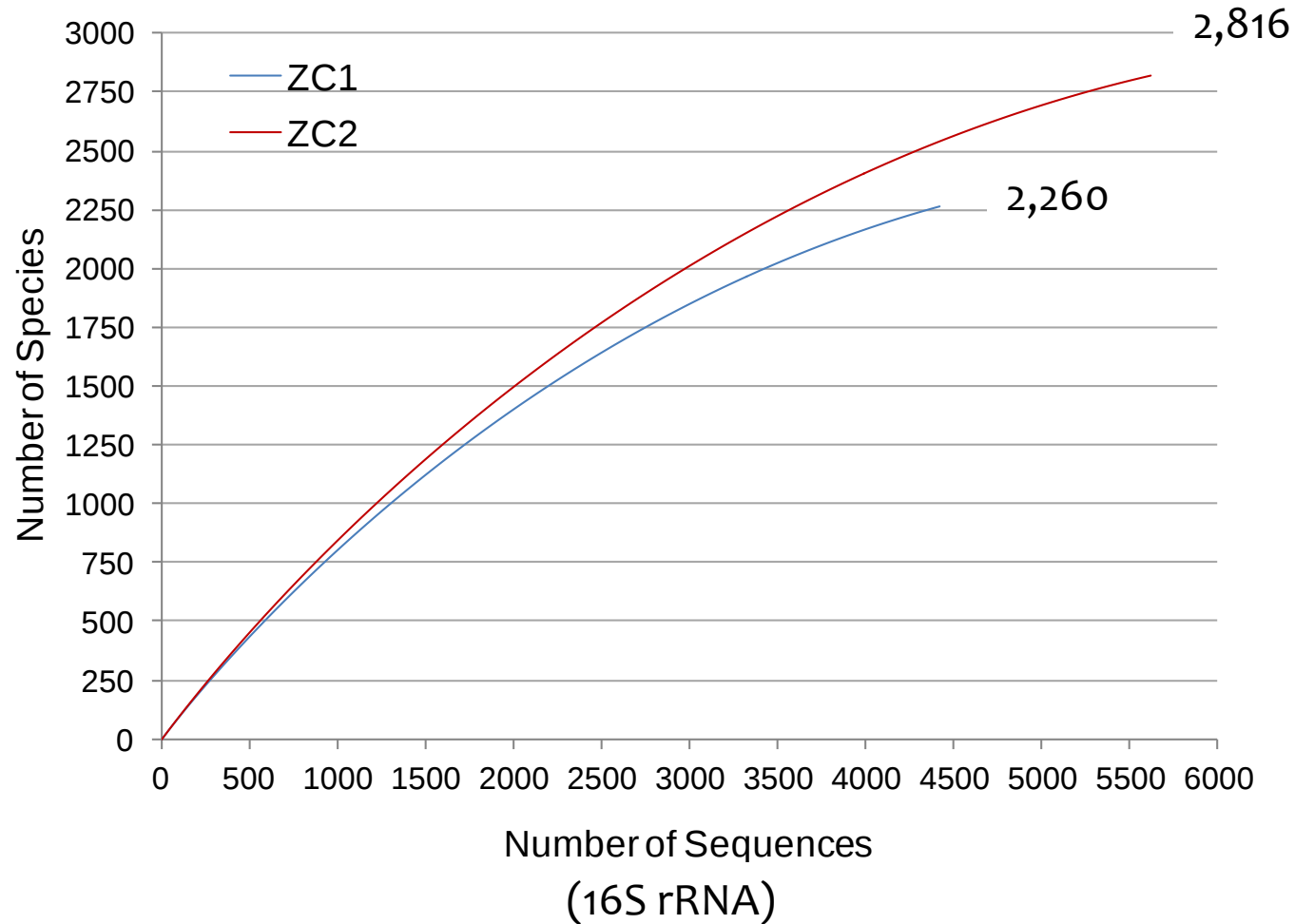
Não funciona bem com fragmentos curtos



Análise de abundância

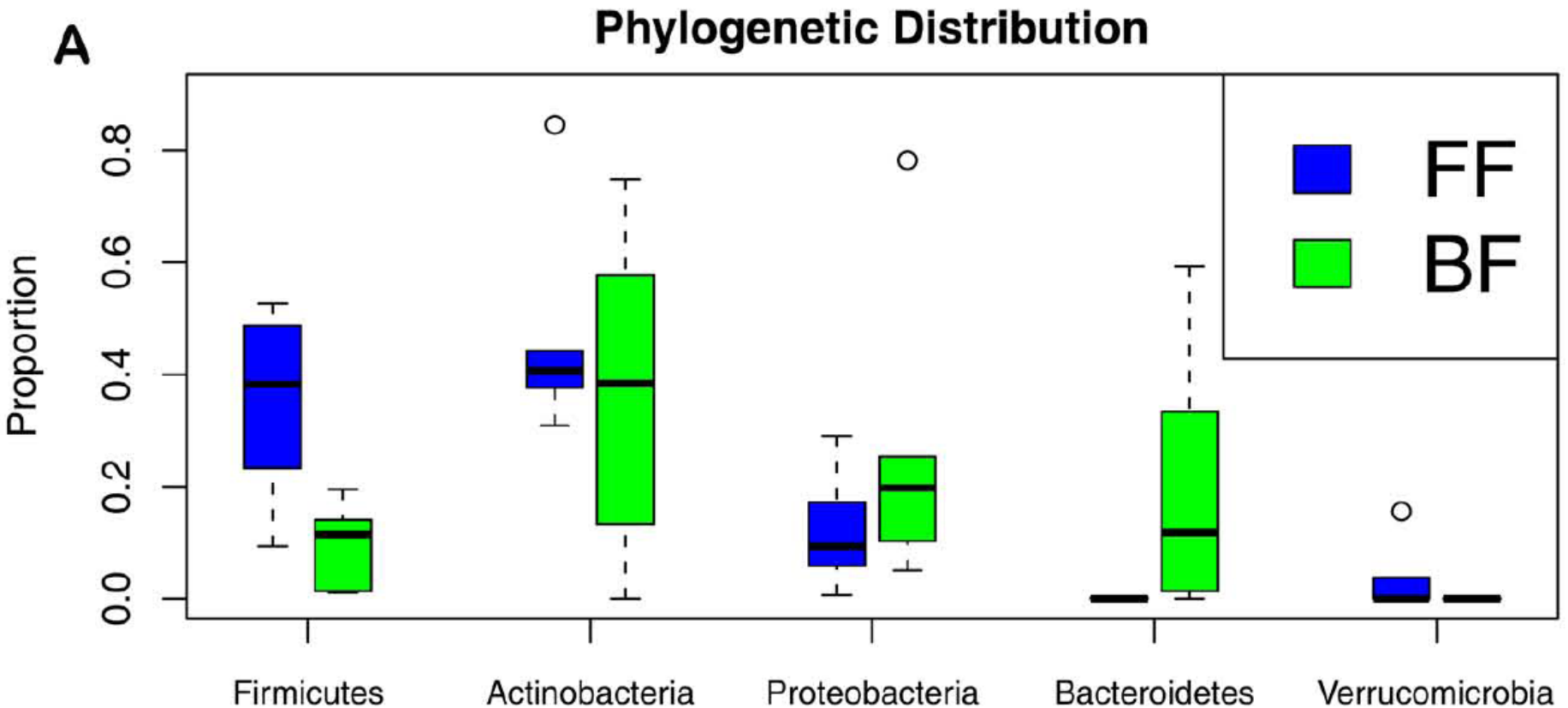
- Que organismos ou funções são mais abundantes num nicho?
- Abundância relativa
 - Curvas de rarefação
 - Variação temporal
 - Variação espacial
- Usar **contagem de reads** como indicador de abundância

Curvas de rarefação

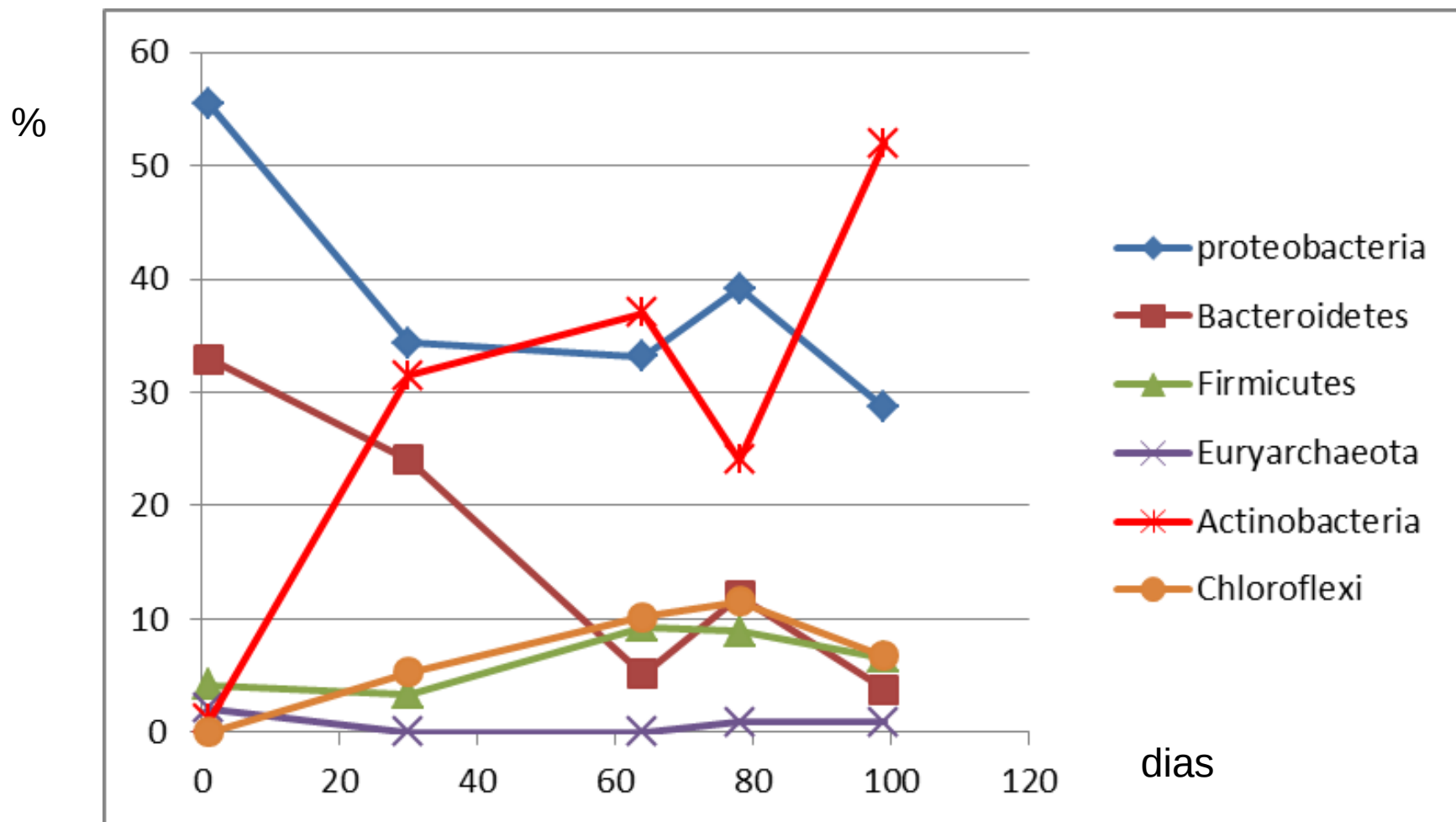


Exemplos de análise de abundância

Variação da microbiota intestinal entre bebês amamentados no seio e com mamadeira



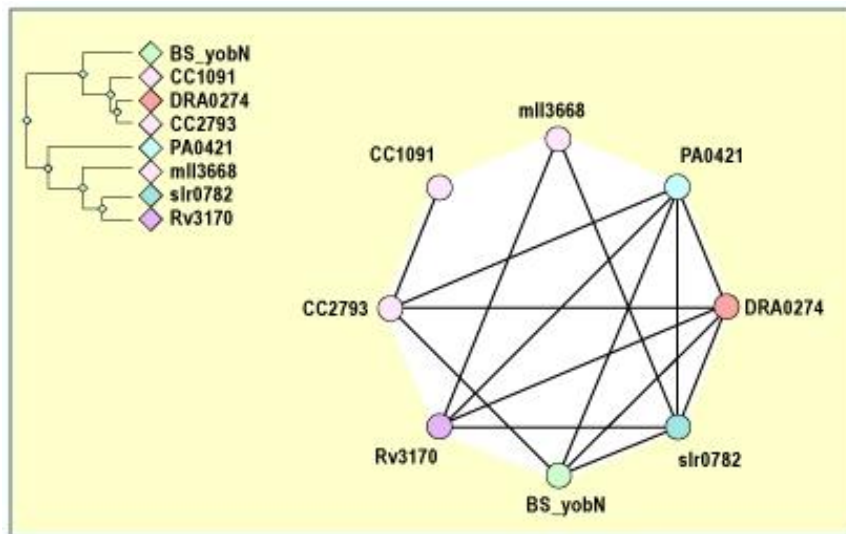
Variação da abundância relativa no tempo



Prejeto metagenoma zoológico - compostagem

Abundância de funções

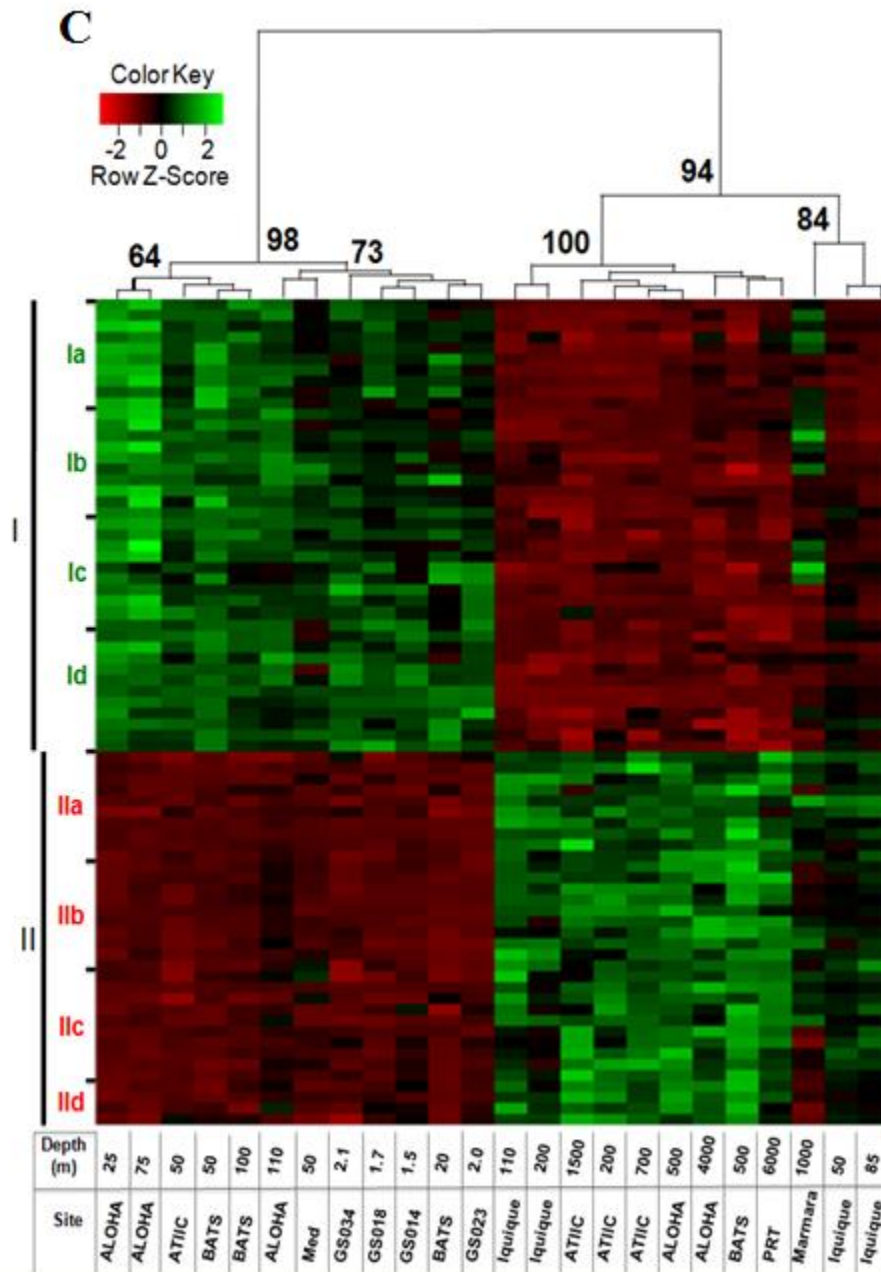
- BLASTX de reads contra um banco de **COGs**
Cluster of Orthologous Groups



Example of a COG: monoamine oxidase

Abundância relativa espacial

- **COGs diferencialmente representados**
- Semelhante a genes diferencialmente expressos
- Heat maps, clusterização hierárquica



41 COGs with higher abundance in photic zones

34 COGs with higher abundance in aphotic zones

Plataformas web de processamento

- Laboratórios governamentais
- Serviços padronizados de processamento

MG-RAST

metagenomics analysis server

LOGIN



[Browse Metagenomes](#)

search for metagenomes



[Register](#)



[Contact](#)



[Help](#)



[Upload](#)

*



[News](#)

About

MG-RAST (the Metagenomics RAST) server is an automated analysis platform for metagenomes providing quantitative insights into microbial populations based on sequence data.

# of metagenomes	77,307
# base pairs	25.81 Tbp
# of sequences	236.94 billion
# of public metagenomes	12,527

The server primarily provides upload, quality control, automated annotation and analysis for prokaryotic metagenomic shotgun samples. MG-RAST was launched in 2007 and has over 8000 registered users and 77,307 data sets. The current server version is 3.3.3.3. We suggest users take a look at [MG-RAST for the impatient](#).

[Updates](#)

[MG-RAST 3.2.4 release notes \[October 2012\]](#)

* login required

This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900040C.

This work was supported in part by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

cite MG-RAST

[Home](#) > Find Genomes
 loaded.

Microbiome Details (Assembled Data)

Add to Genome Cart

 Browse Genome

 BLAST Genome

About Genome

- [Overview](#)
- [Statistics](#)
- [Genes](#)

Overview

Proposal Name	Sao Paulo Zoo Compost
Sample Name	Sample C4
Taxon Object ID	2156126000
IMG Submission ID	2671
GOLD ID in IMG Database	Project Id: Gm0002180
External Links	
Genome type	metagenome
Sequencing Status	Draft
IMG Release	
Comment	
Sample Information	
Sample Site	Sao Paulo Zoo composting operation
Sample Collection Date	January 26, 2011
Isolation Country	Brazil
Sampling Strategy	8 days after composting started
Sample Isolation	done 8 days after composting started
Temperature Range	Thermophile
Sample Assembly Method	newbler
Sample Geographic Location	Sao Pulo Zoo
Longitude	-46.62
Latitude	-23.65



EBI Metagenomics

[Home](#)
[Submit data](#)
[Projects](#)
[Samples](#)
[About Metagenomics](#)
[Contact](#)

Not logged in [Log in](#)

Easy submission



Manually supported submission process, with help available for meta-data provision. Accepted data formats include SFF (454) and FASTQ (Illumina and IonTorrent).

[Find out more](#)

Powerful analysis



Functional analysis of metagenomic sequences using InterPro - a powerful and sophisticated alternative to BLAST-based analyses. Taxonomy diversity analysis is performed using Qiime.

[Find out more](#)

Data archiving



Data automatically archived at the Sequence Read Archive (SRA), ensuring accession numbers are supplied - a prerequisite for publication in many journals.

[Find out more](#)



Projects

Latest public projects (Total: 37)

Metatranscriptomics of the marine sponge *Geodia barretti*: Tackling phylogeny and function of its microbial community.

Geodia barretti is a marine cold-water sponge harbouring high numbers of microorganisms. ...

[View more](#) - 1 sample

A metagenomics transect into the deepest point of the Baltic Sea reveals clear stratification of microbial functional capacities

The Baltic Sea is characterized by hyposaline surface waters, hypoxic and anoxic deep waters and ...

[View more](#) - 6 samples

Gut metagenome in European women with normal, impaired and diabetic glucose control

Type 2 diabetes (T2D) is a result of complex gene-environment interactions, and several risk ...

[View more](#) - 147 samples

Samples

Latest public samples (Total: 1053)

Fecal sample from Crohn's patient 1

Fecal sample from Crohn's patient 1 ...

[View more](#) - Taxonomy | Function results |

Fecal sample from Crohn's patient 10

Fecal sample from Crohn's patient 10 ...

[View more](#) - Taxonomy | Function results |

Fecal sample from Crohn's patient 2

Fecal sample from Crohn's patient 2 ...

[View more](#) - Taxonomy | Function results |

Fecal sample from Crohn's patient 3

Fecal sample from Crohn's patient 3 ...

[View more](#) - Taxonomy | Function results |

Fecal sample from Crohn's patient 4

Fecal sample from Crohn's patient 4 ...

[View more](#) - Taxonomy | Function results |

Data content

1053 public samples (37 public projects)

191 private samples (13 private projects)

News & events

Tweets

Follow @EBImetagenomics



EBI Metagenomics @EBImetagenomics 30 Sep

Check out our new analysis page, using improved data visualisation (Google & Krona charts), and with taxonomic info: ebi.ac.uk/metagenomics/

Expand



EBI Metagenomics @EBImetagenomics 8 Aug

The poster we presented at #SMBECCB is now available at F1000 posters and describes the EBI metagenomics pipeline: [F1000 research paper](#)