

Metagenômica

Metagenômica

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

Por que metagenômica?

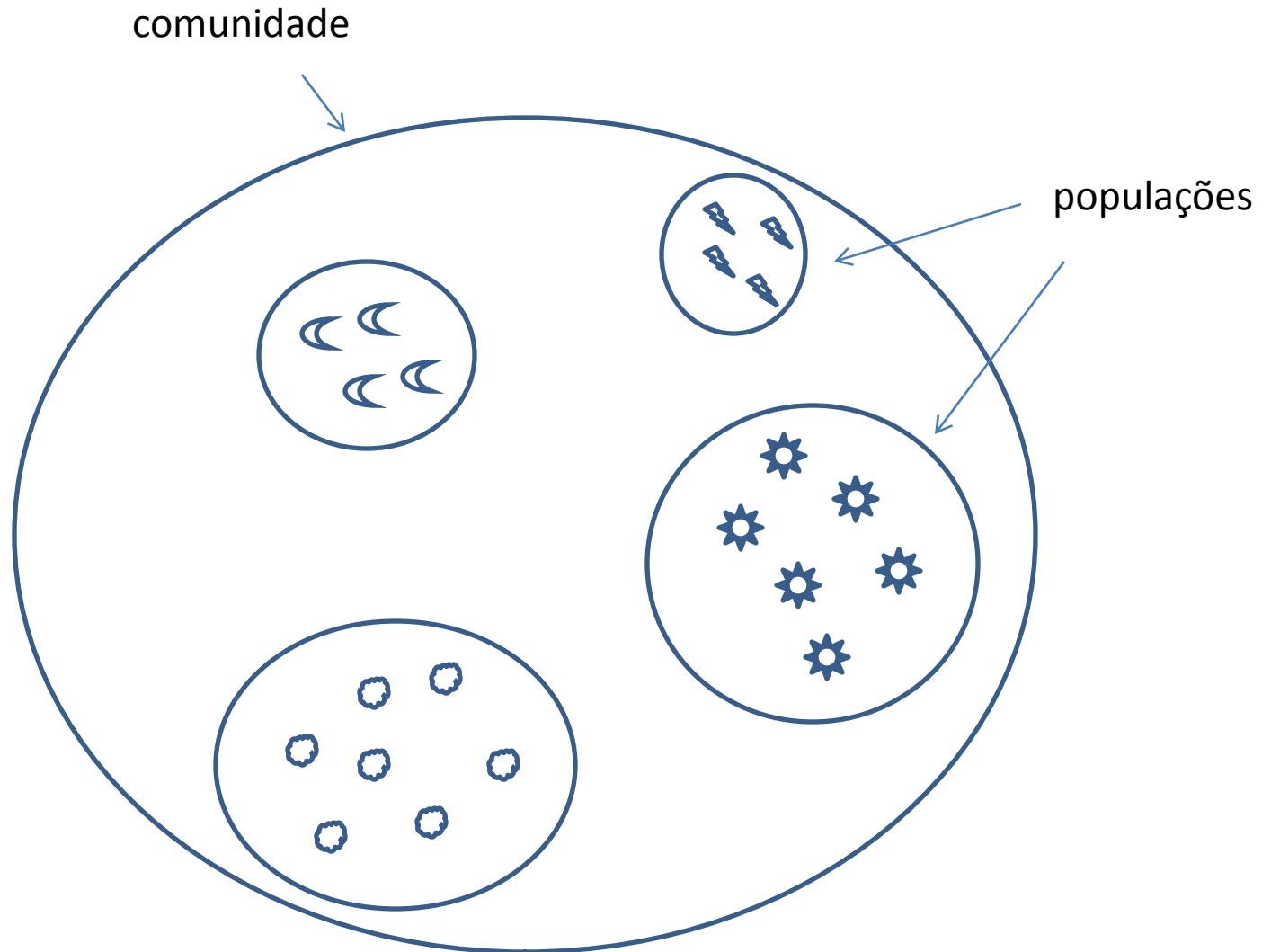
- Procariotos: “a maioria invisível”
 - Tanto carbono quanto plantas [Whitman et al, 1998]
- Estão por todo lugar
 - Da atmosfera superior ao subsolo
- Enorme diversidade genética
 - organismos
 - proteínas

Como podemos estudar os genomas dessa maioria invisível?

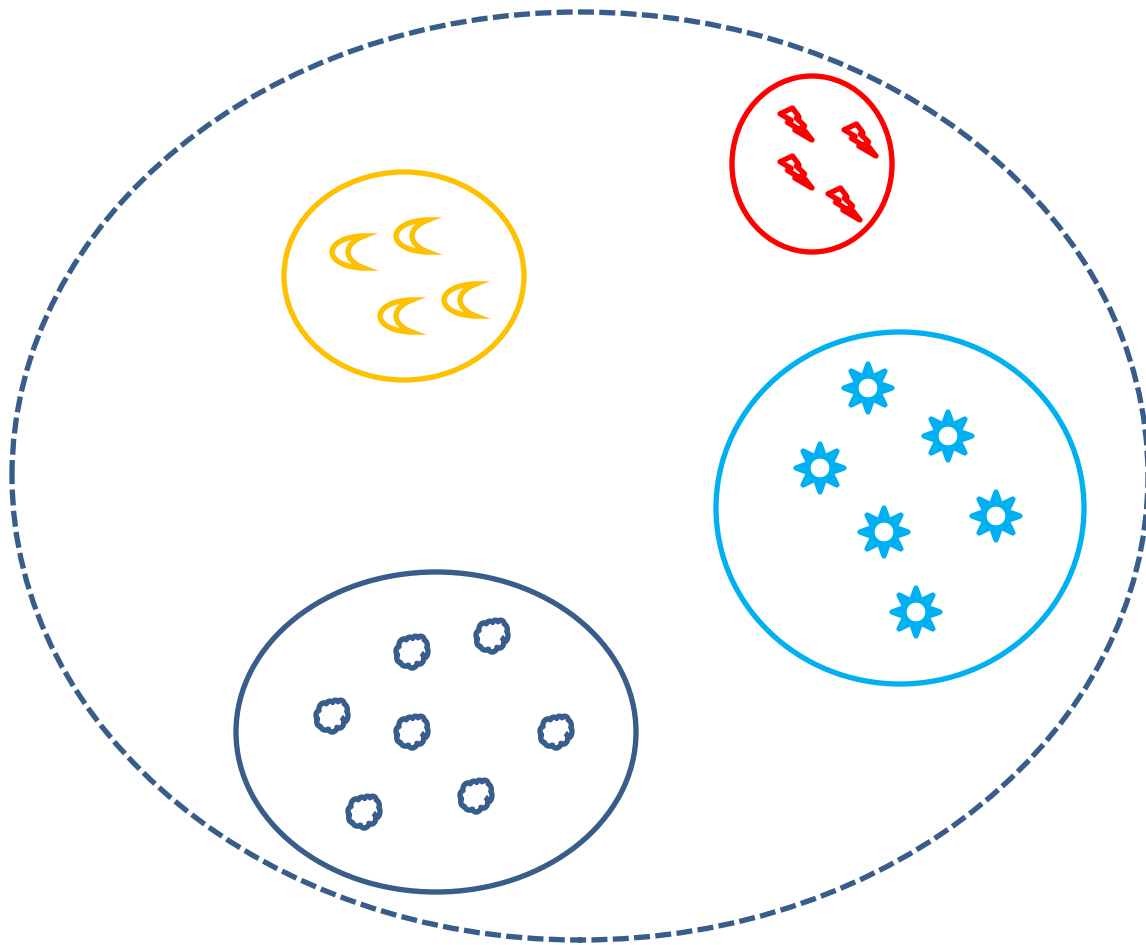


- Organismo isolado: cultura em laboratório
- Menos do que 1% dos procariotos são cultiváveis
- Metagenômica: não há necessidade de cultura

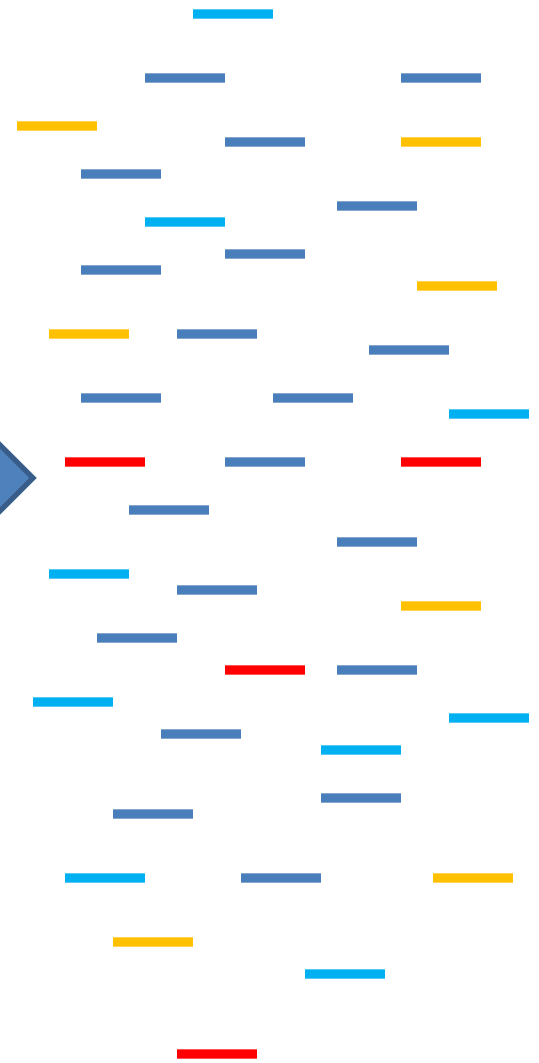
Um nicho ecológico



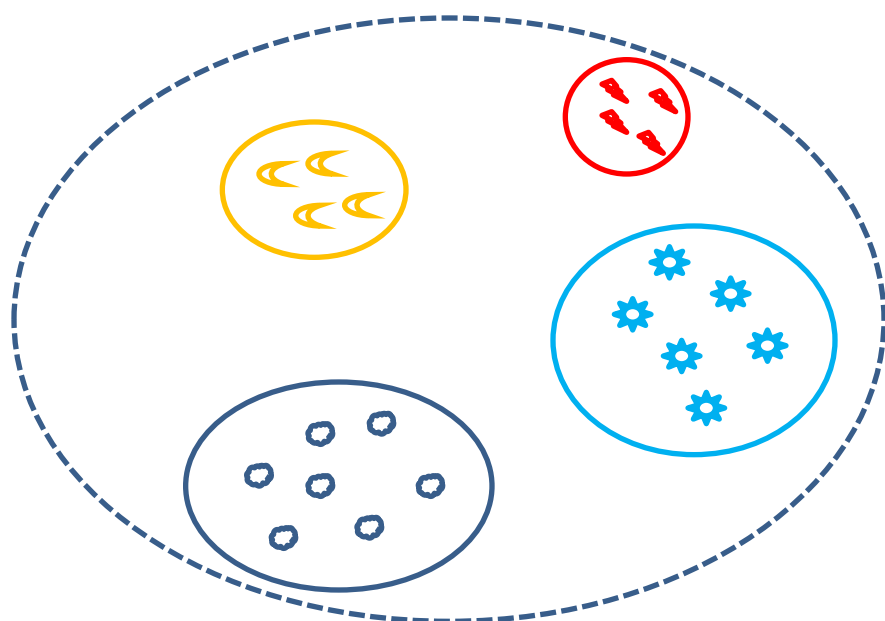
A comunidade



DNA

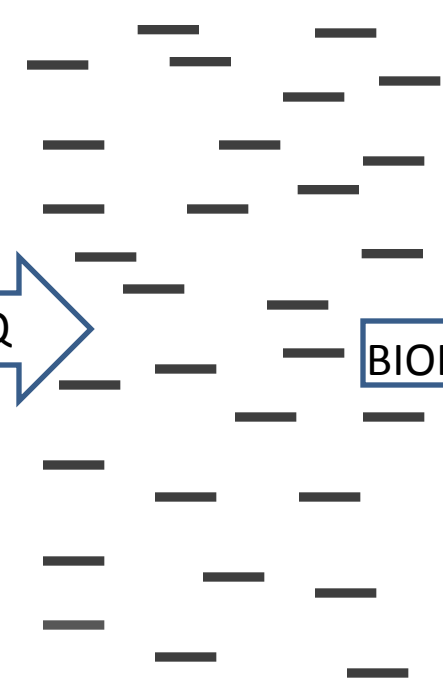


A comunidade

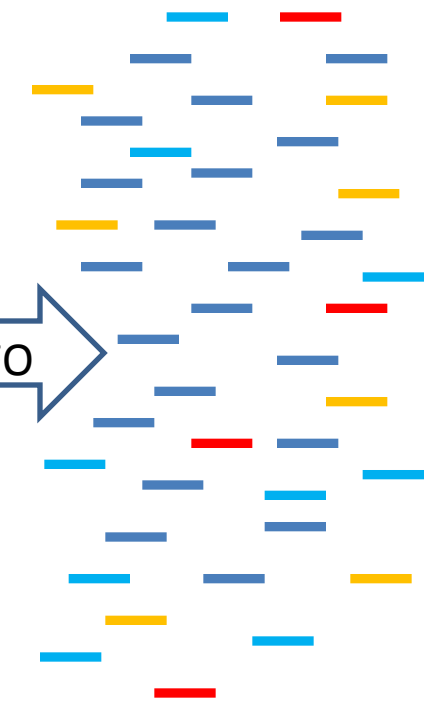


SEQ

DNA



BIOINFO



Questões básicas

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

Bioinformática é complexa

- Volume de dados: milhões de reads
- Baixa cobertura de genomas individuais
- erros de sequenciamento
- → baixa qualidade de dados
- → algoritmos precisam ser mais robustos
- Montagem
- Binning; classificação filogenética
- Metodologia padronizada para comparações

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**ATAGCCTAGATCGATC**ATTA**CC

AGAT ocorre 2 vezes

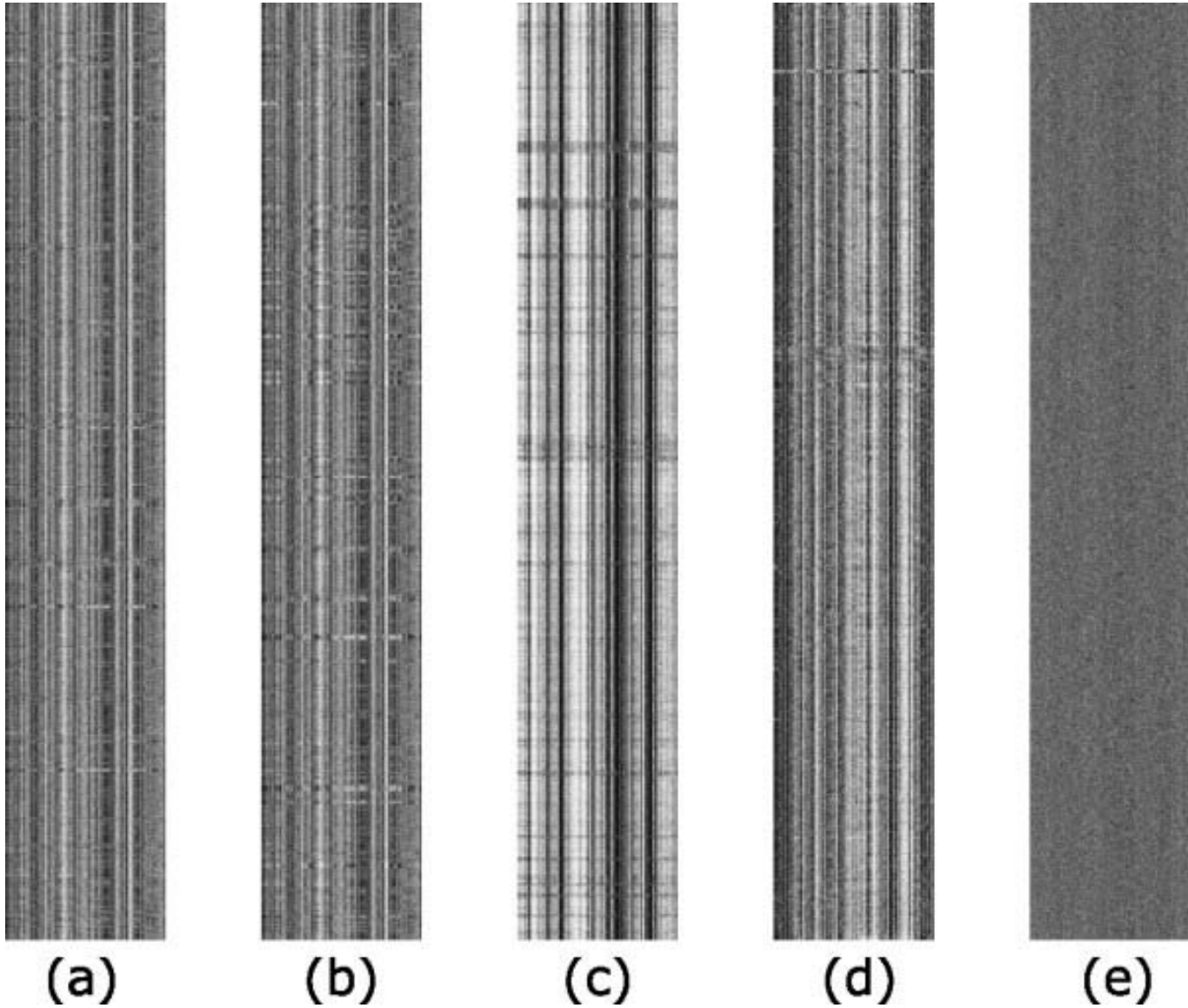
ATTA ocorre 3 vezes

etc

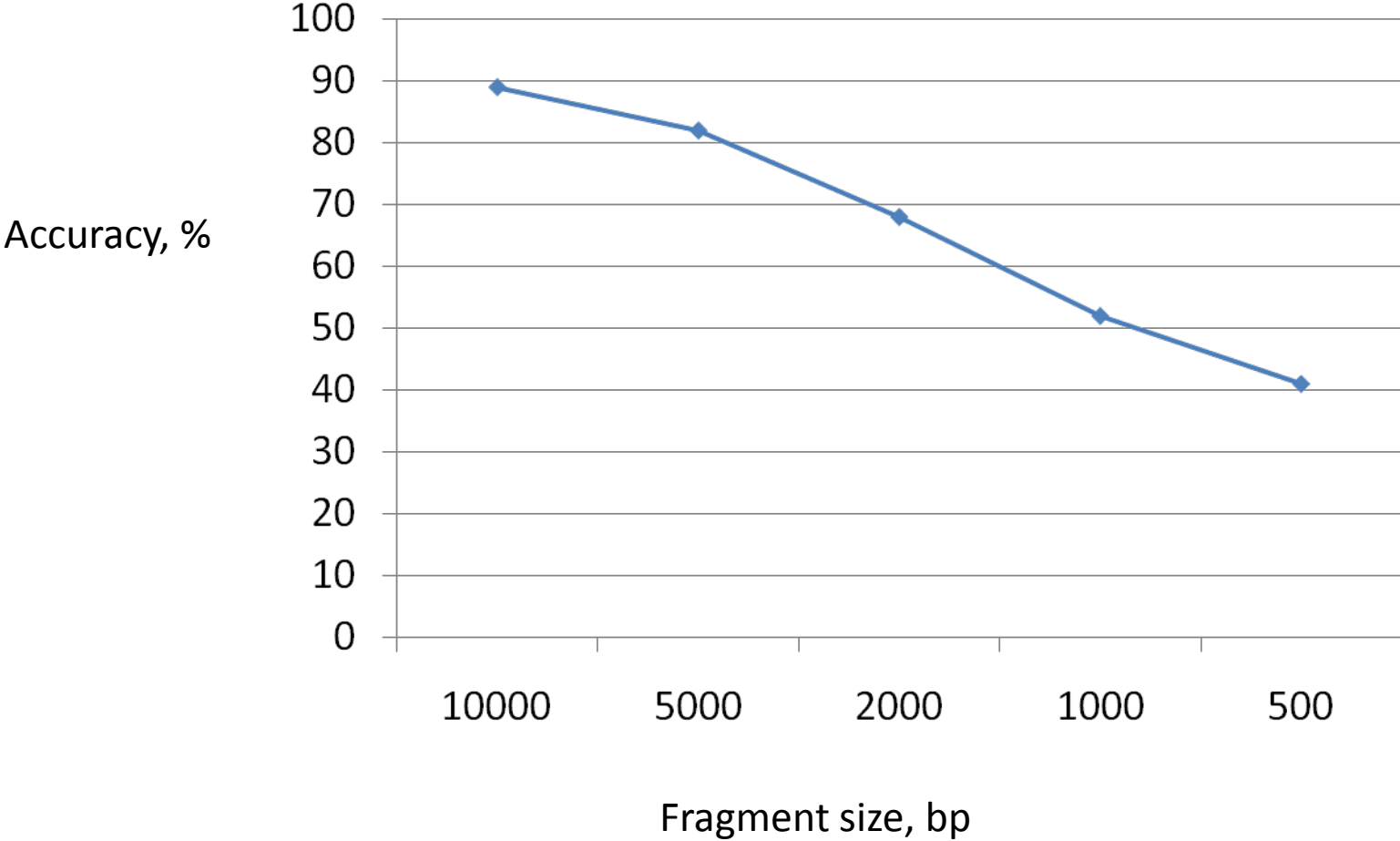
Matriz de frequências

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”

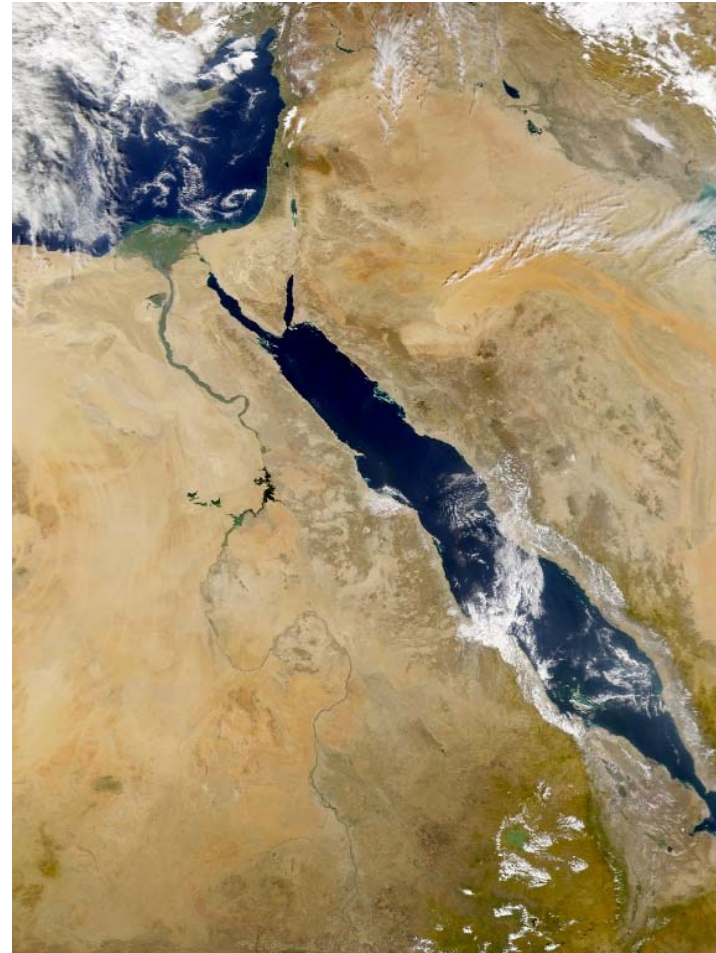


Não funciona bem com fragmentos curtos

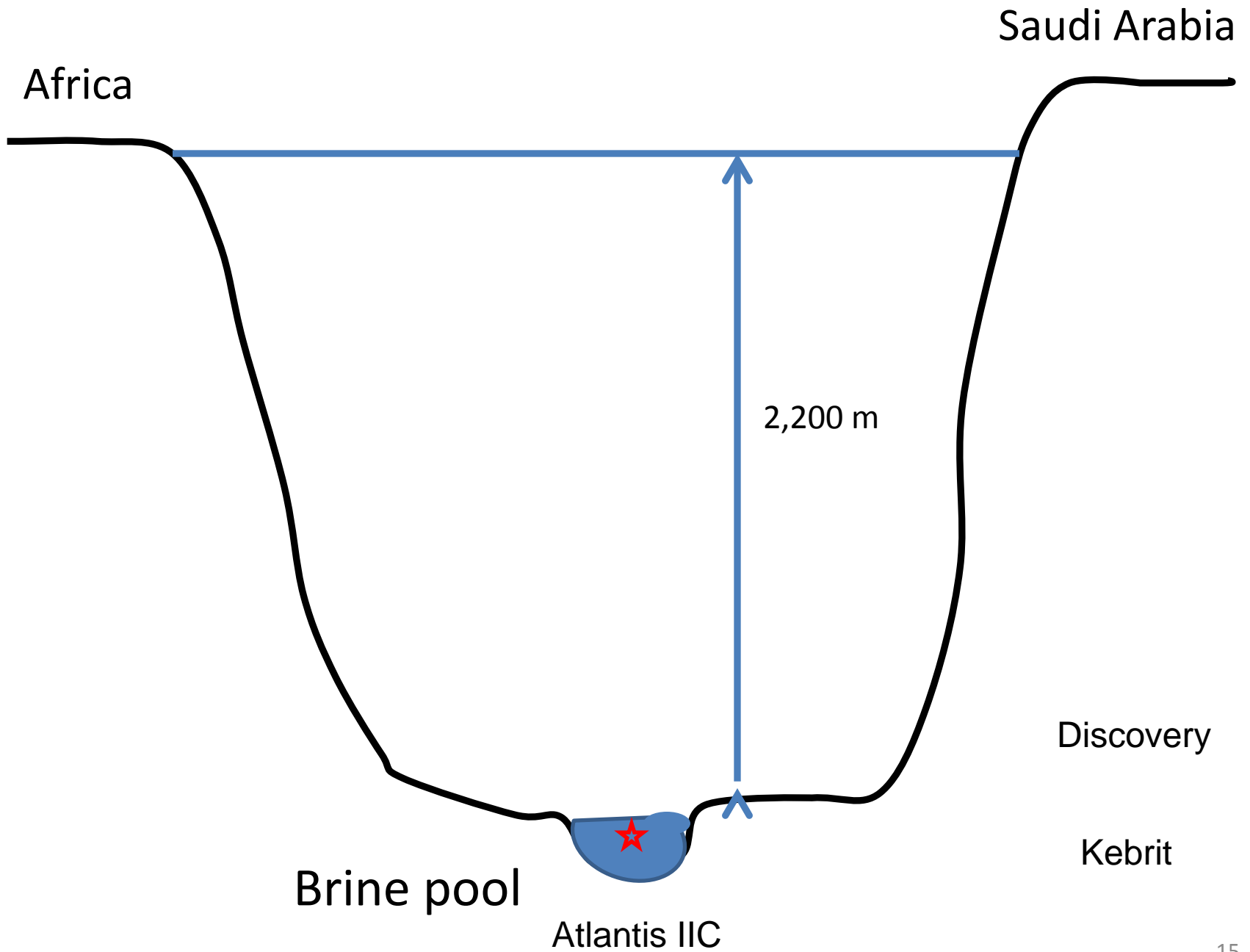


Zhou et al, 2009 simulated data

Red Sea Project

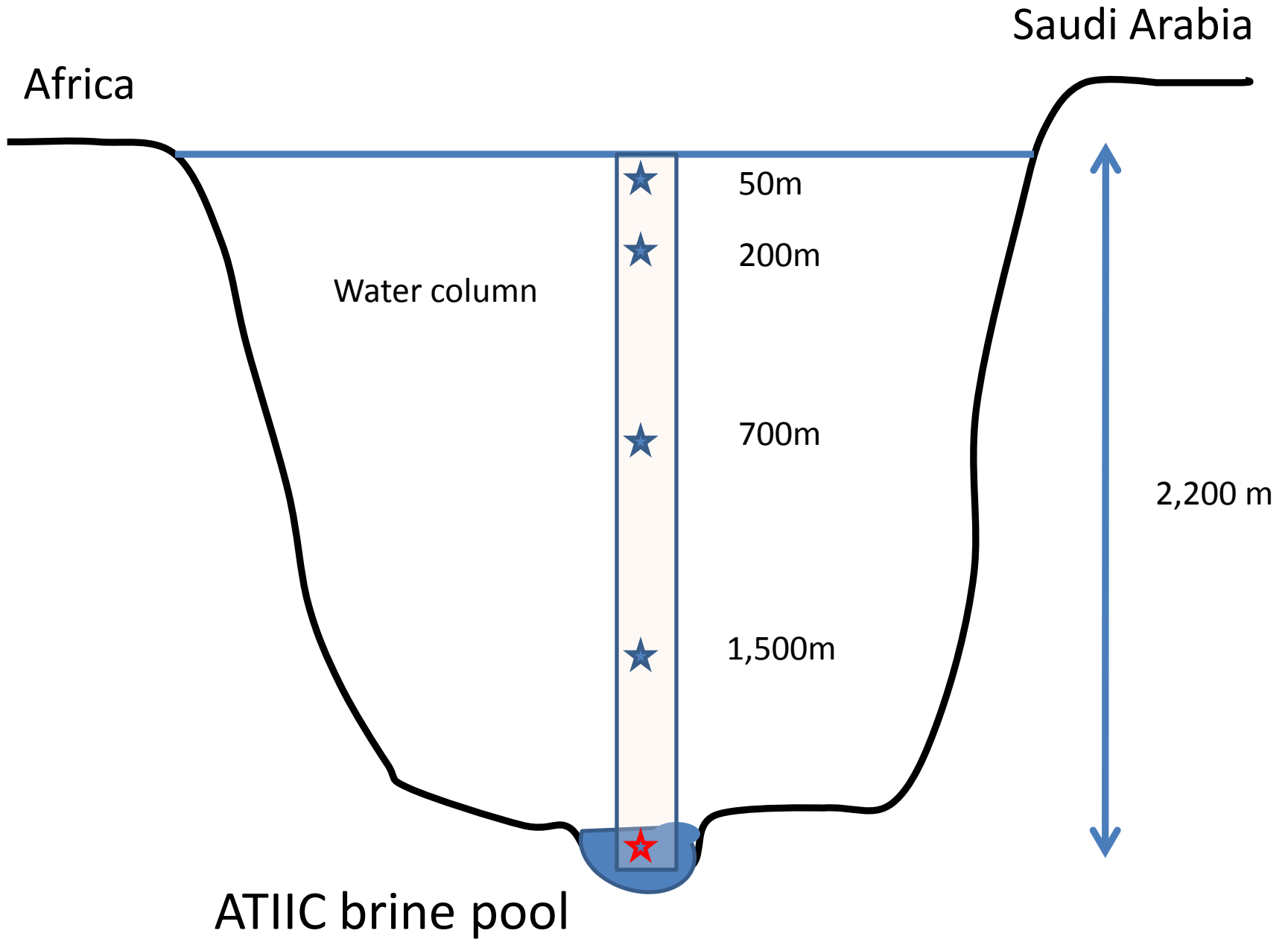


- American University in Cairo



The brine pools are **extreme environments**

- **High salinity** (10X more than surface water)
- **Enriched with heavy metals:** iron, manganese, copper, zinc (1000X more concentrated than normal water)
- **High temperatures** (70 °C)
- **High pressure**
- **No light**



Location of the 11 sites and number of sequences of the 24 data sets

Figure S1



Depth (m)	North Pacific Gyre ALOHA Station, Hawaii, USA	Sargasso Sea BATS Station, Bermudas	Red Sea Atlantis II Basin	Southern Pacific Station 3, Coast of Iquique, Chile	Mediterranean Sea Coast of Alicante, Spain	Sea of Marmara Central Basin, Turkey	Atlantic Ocean, Puerto Rico Trench North of Puerto Rico
20		334					
25	581						
50		410	1100	292	1204		
75	645						
85				529			
100		484					
110	454			334			
200			655	453			
500	954	158					
700			654			253	
1000							
1500			1138				
4000	116						
6000							545

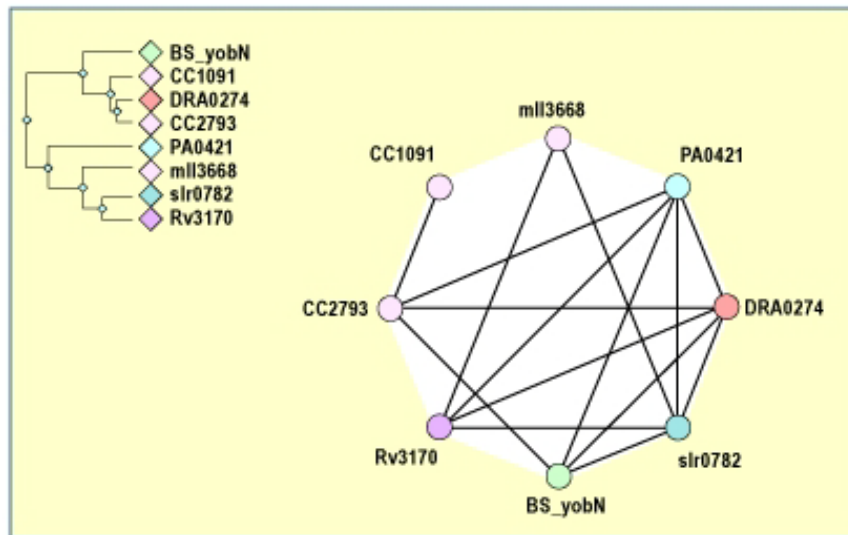
Sequences, x10³

GOS (surface water): Caribbean Sea (GS018) Eastern Tropical Pacific (GS023) Galapagos Islands (GS034) Indian Ocean (GS114)

Sequences x10³ 143 133 134 349

Bioinformatics

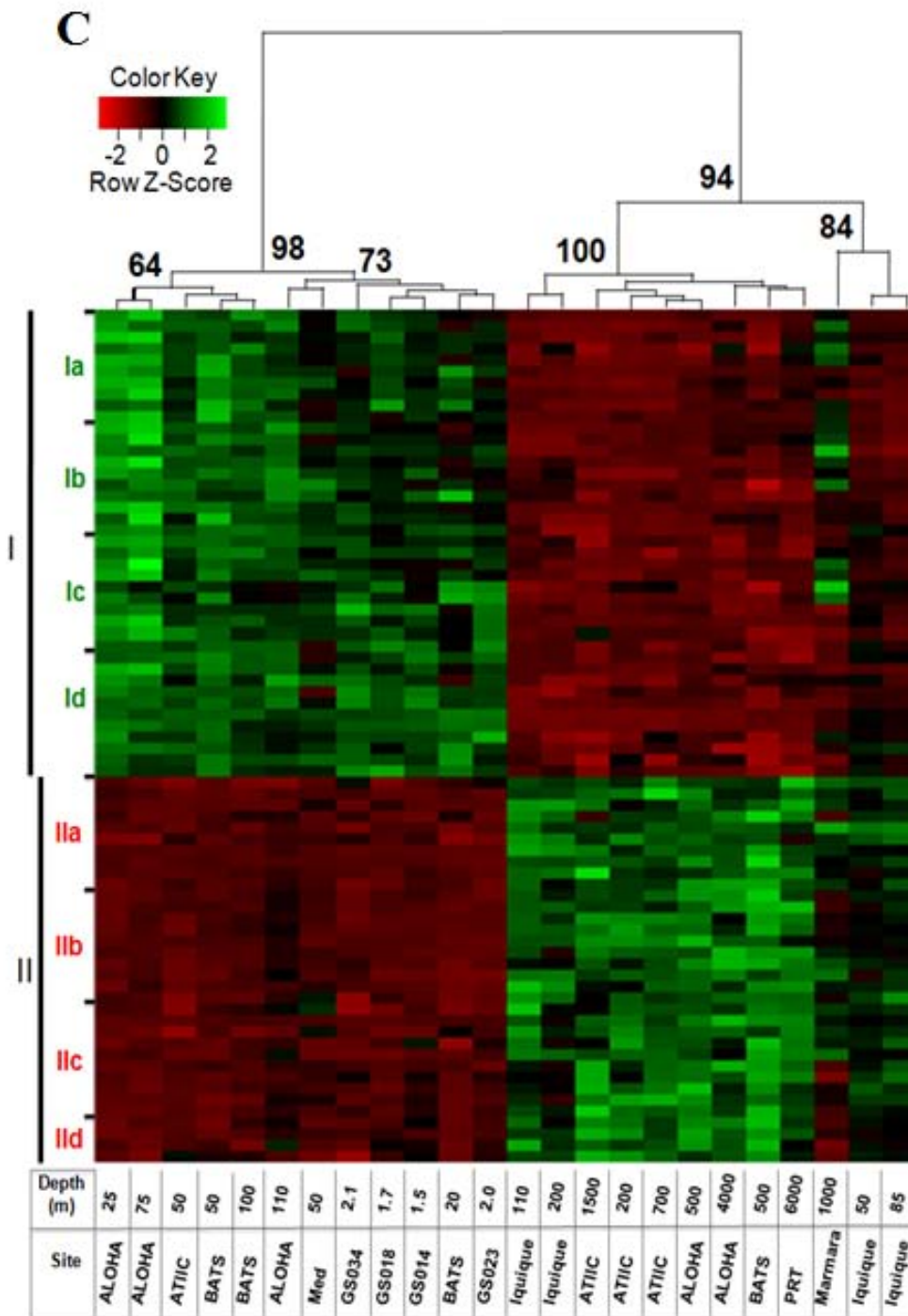
- BLASTX of reads against a **COG database**
- **Cluster of Orthologous Groups**



Example of a COG: monoamine oxidase

COGs diferencialmente representados

- Quais genes (representados pelos COGs ao qual foram associados) são mais ou menos abundantes em diferentes amostras



41 COGs with higher abundance in photic zones

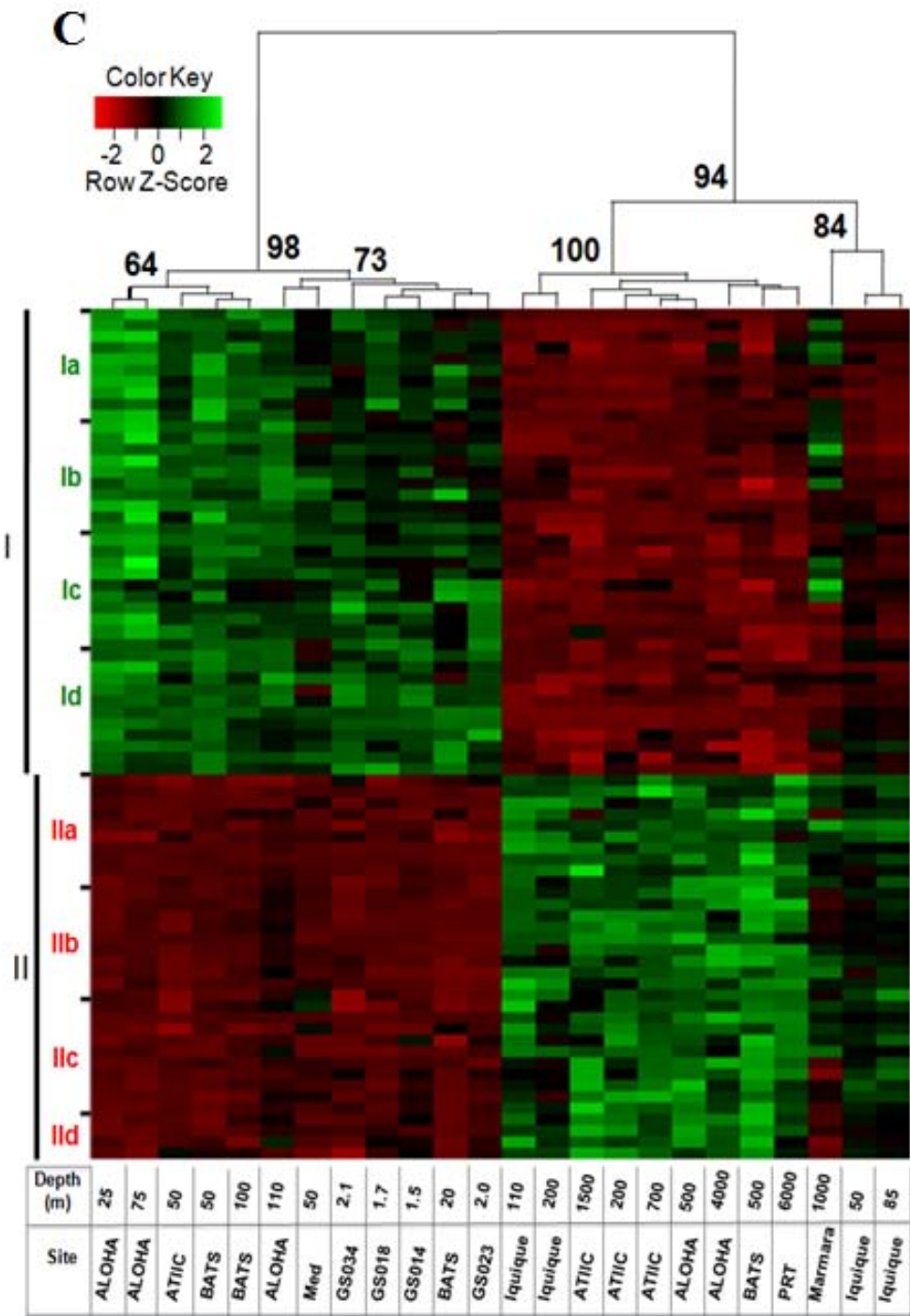
34 COGs with higher abundance in aphotic zones

Photic COGs

- Photosynthesis
- biosynthesis of light-harvesting pigments
- assimilation of CO₂ by photosynthetic bacteria
- Light-induced DNA repair
- oxidative stress response
- N₂ fixation
- phosphate metabolism

Aphotic COGs

- Catabolism of proteins and aminoacids
- Methane oxidation
- sulfate assimilation and metabolism
- selenocysteine metabolism
- terpenoid biosynthesis

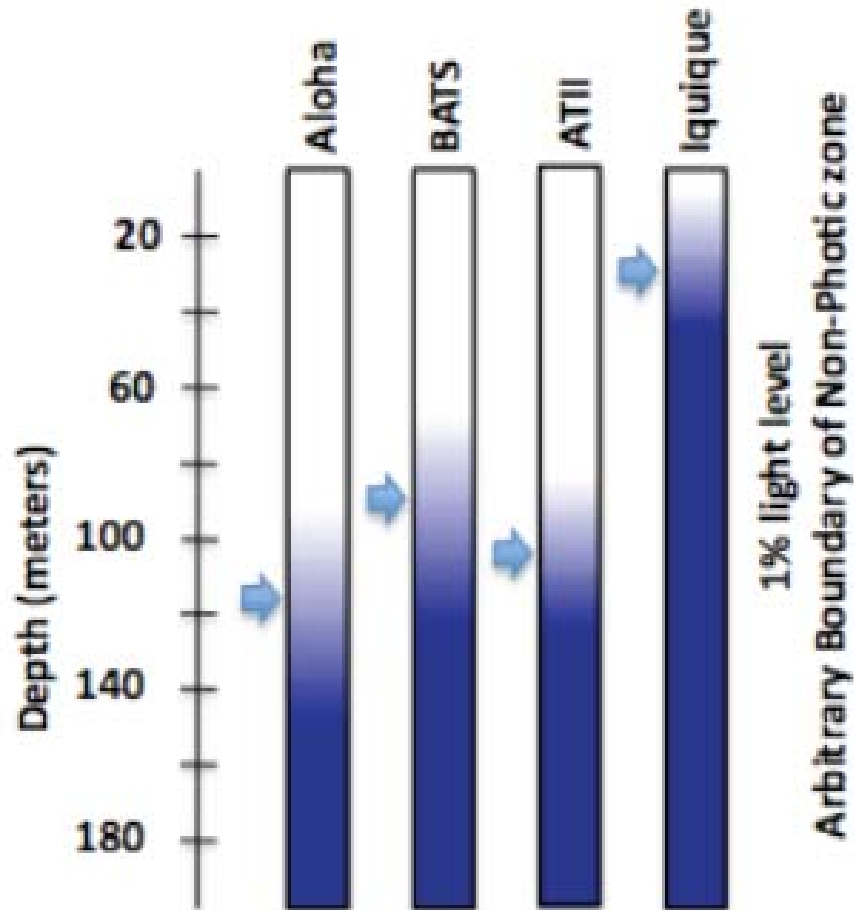


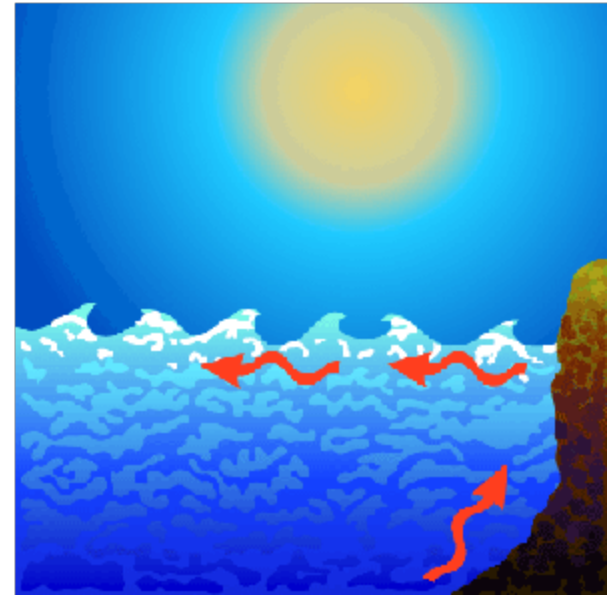
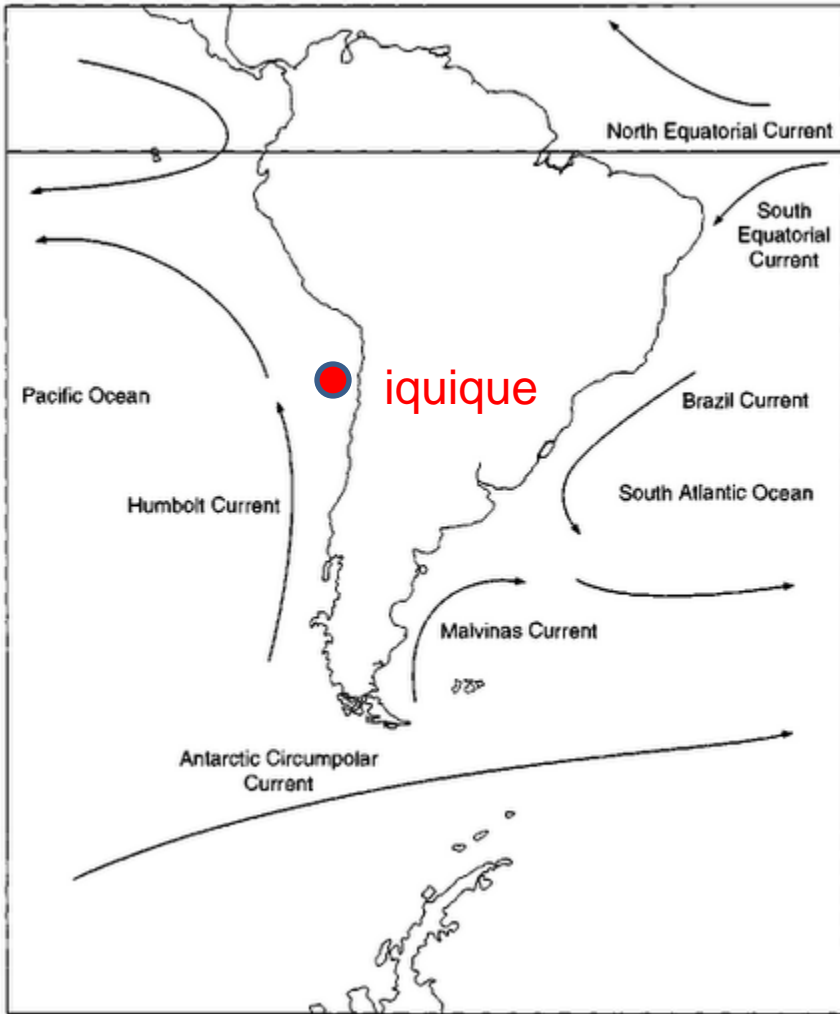
41 COGs with higher abundance in photic zones

34 COGs with higher abundance in aphotic zones

Figure 2B

PAR values (Photosynthetically active radiation)



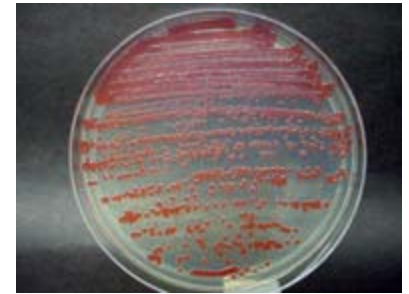


upwelling

<http://mydasdata.larc.nasa.gov/glossary.php?&word=upwelling>

Compostagem no Zoo-SP

- Escala “industrial”
- Entram todos os resíduos orgânicos disponíveis
 - Galhos e folhas da mata atlântica
 - Resíduos dos animais
 - Carcassas de animais mortos
- Resultado: adubo para a fazenda do zoo



Objetivo do projeto

- Estudar a diversidade microbiana da compostagem
- Estudar a diversidade proteica da compostagem (biotecnologia)
- Resultados preliminares: duas amostras de duas diferentes compostagens

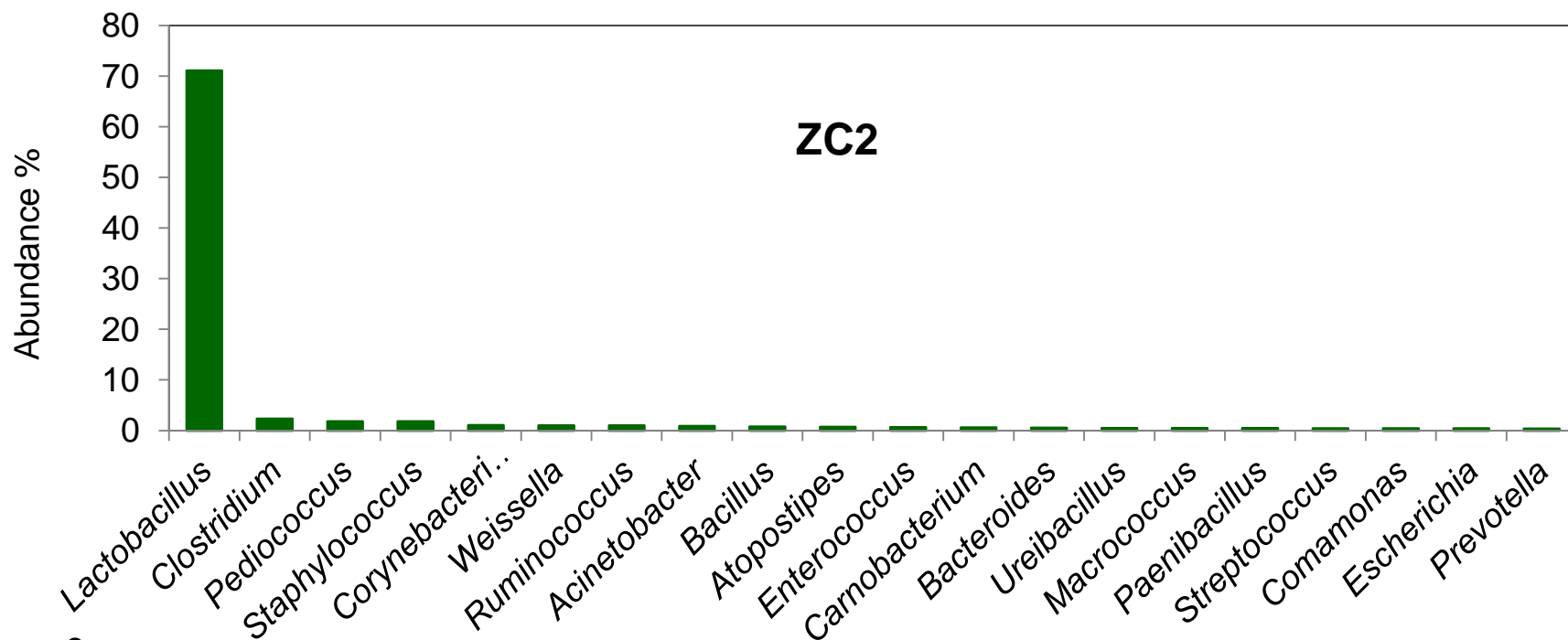
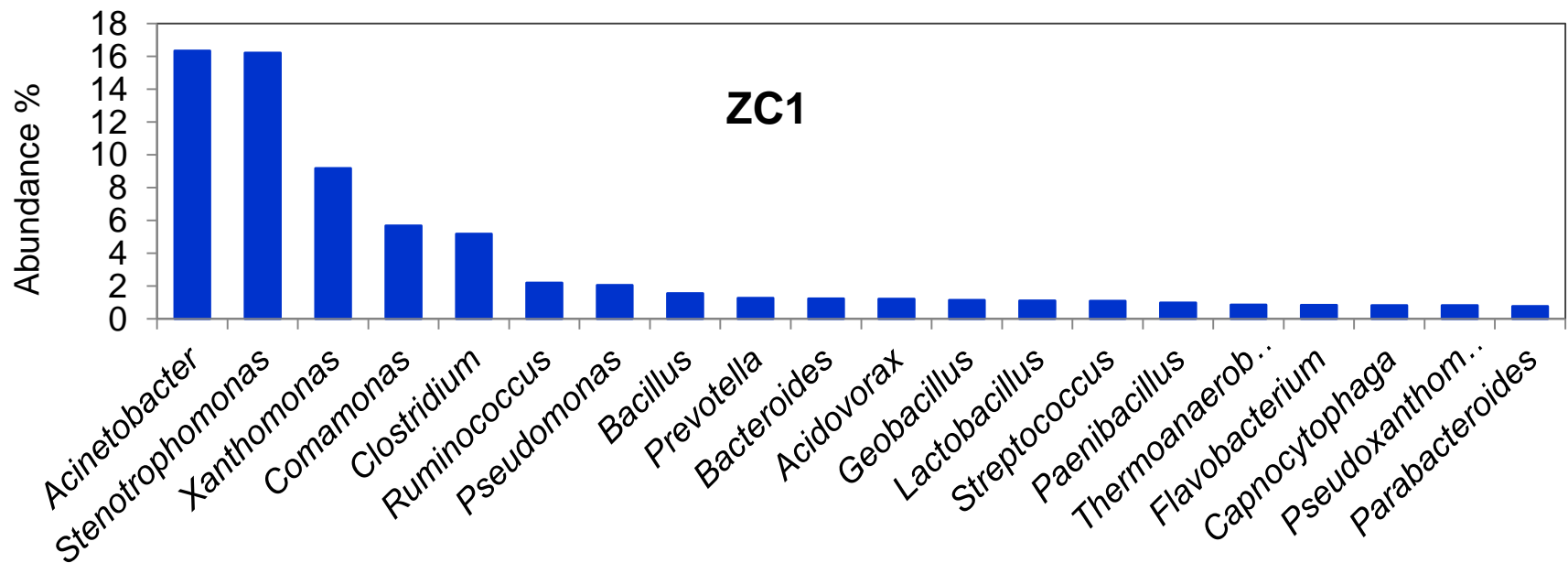


Figure 3

Invisible Residents

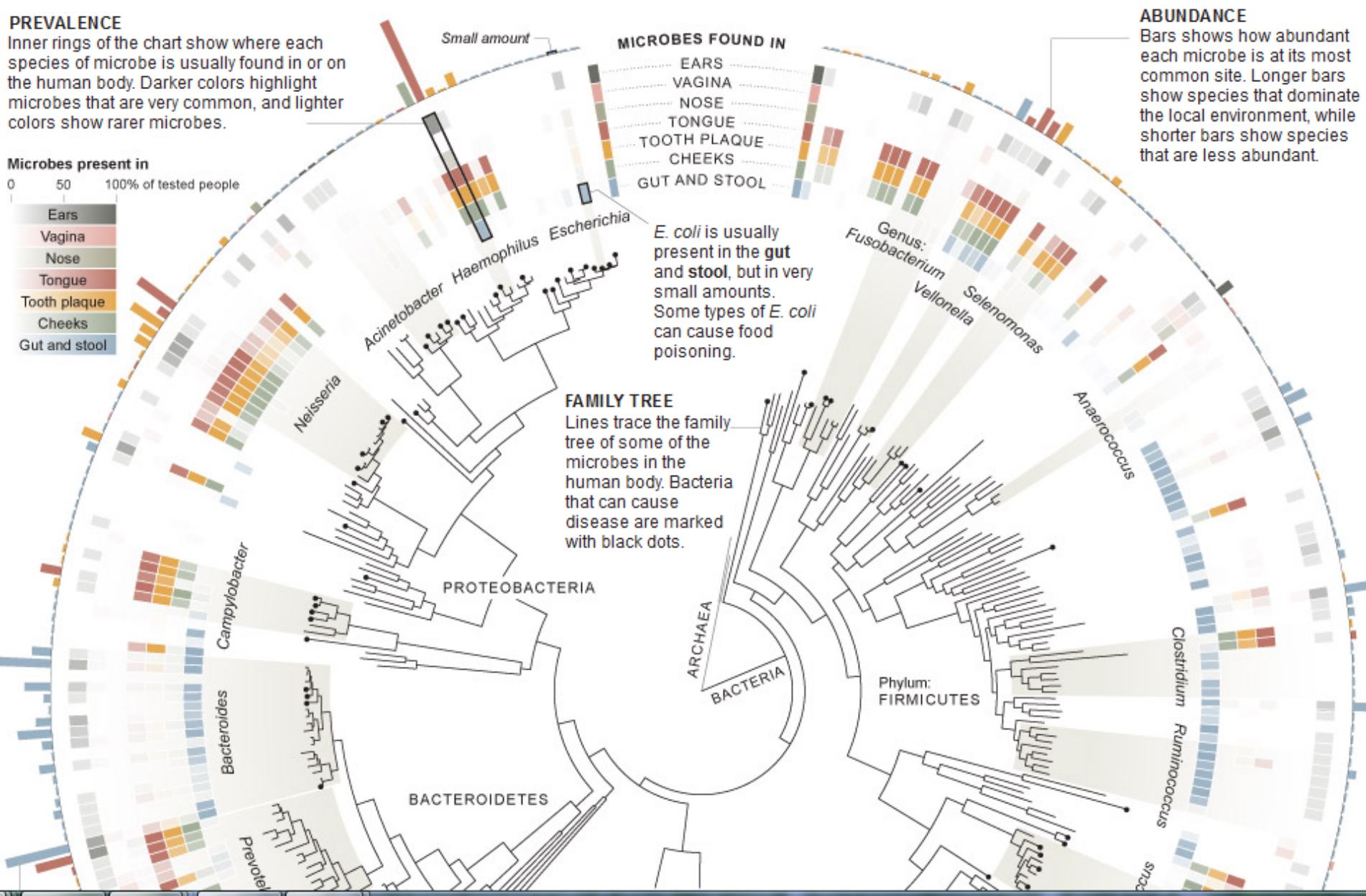
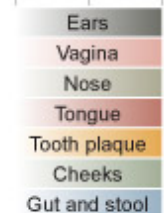
The Human Microbiome Project has spent two years surveying bacteria and other microbes at different sites on 242 healthy people. The chart below hints at the complex combinations of microbes living in and on the human body. [Related Article »](#)

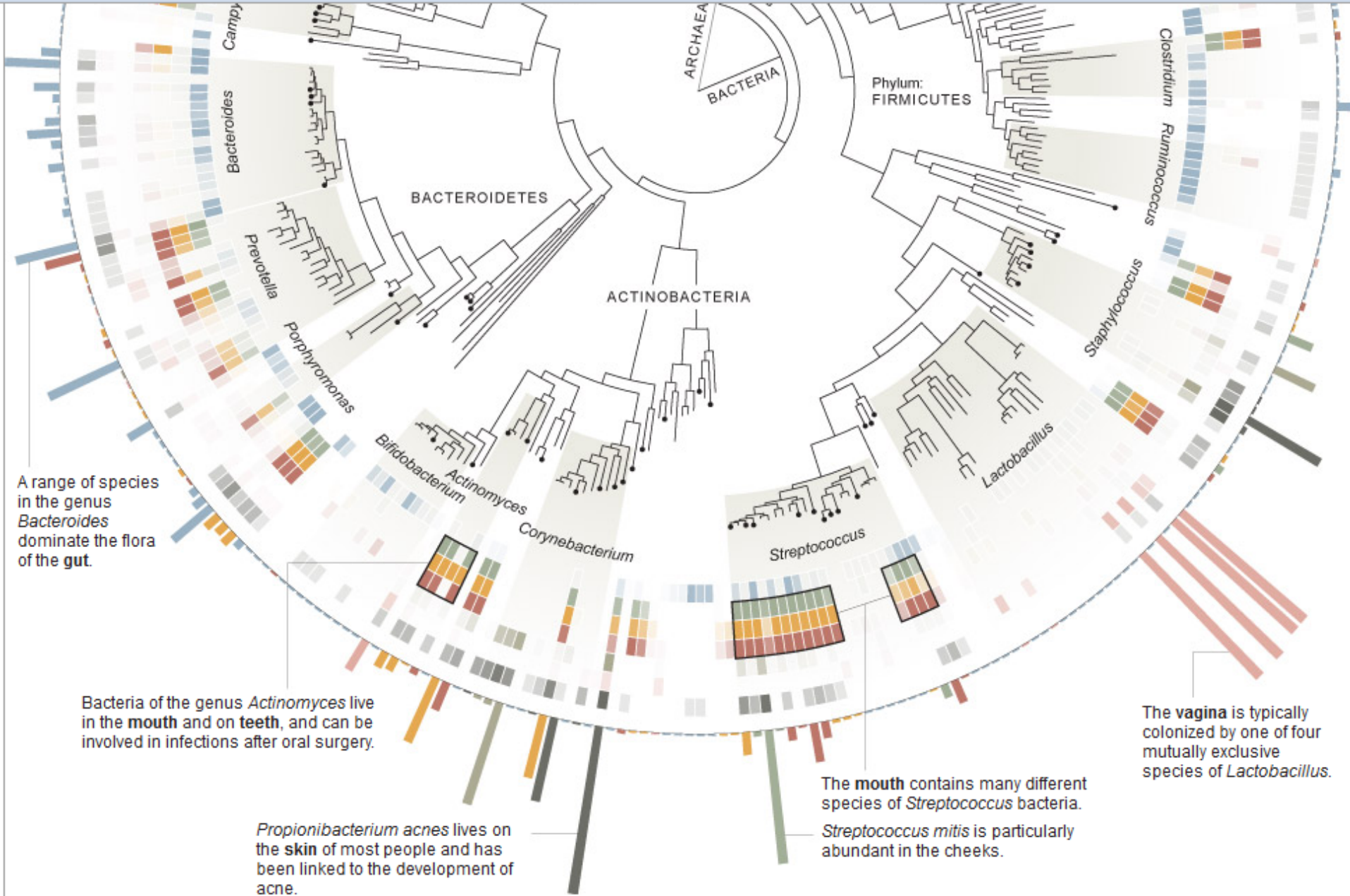
PREVALENCE

Inner rings of the chart show where each species of microbe is usually found in or on the human body. Darker colors highlight microbes that are very common, and lighter colors show rarer microbes.

Microbes present in

0 50 100% of tested people





[Send Feedback](#)

Sources: Curtis Huttenhower and Nicola Segata, Harvard School of Public Health; National Institutes of Health Human Microbiome Project