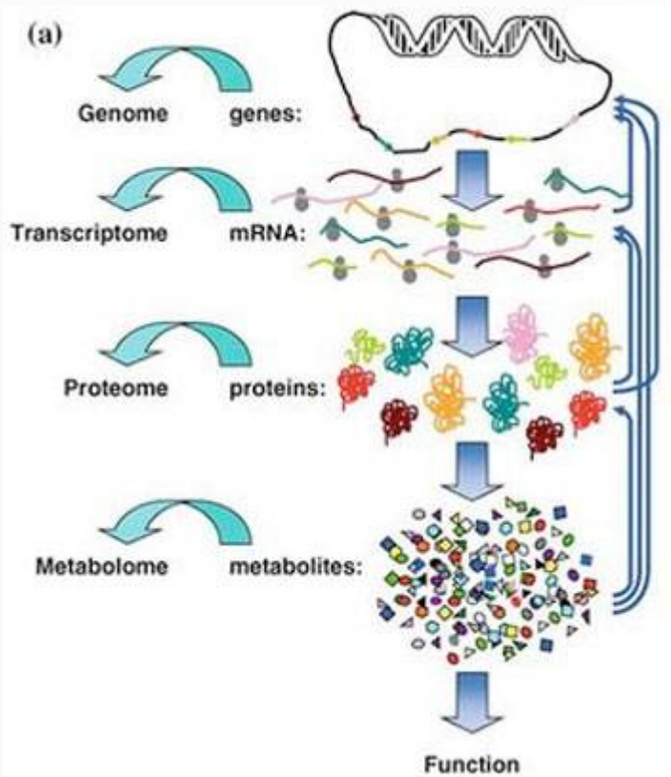
Ciências ômicas

Prof. Leonardo M. Cruz

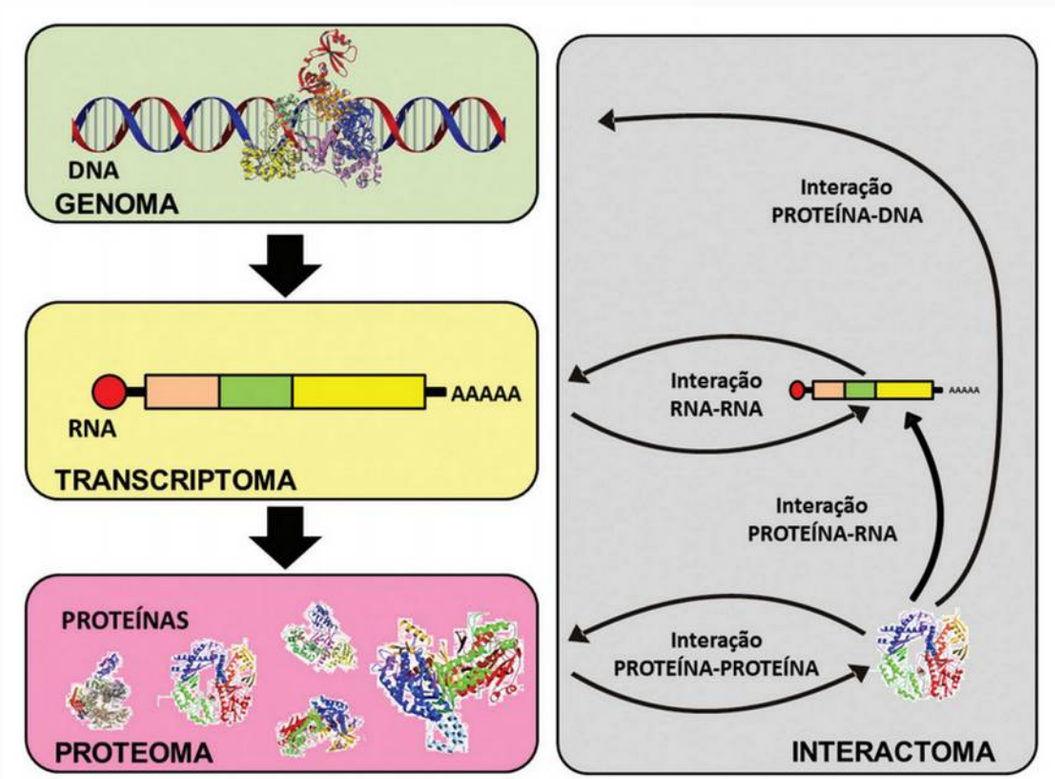
Departamento de Bioquímica e Biologia Molecular
Universidade Federal do Paraná

Dogma central e análises ômica

Análise de um único organismo → ômica
Análise de uma comunidade de organismos → metaômica

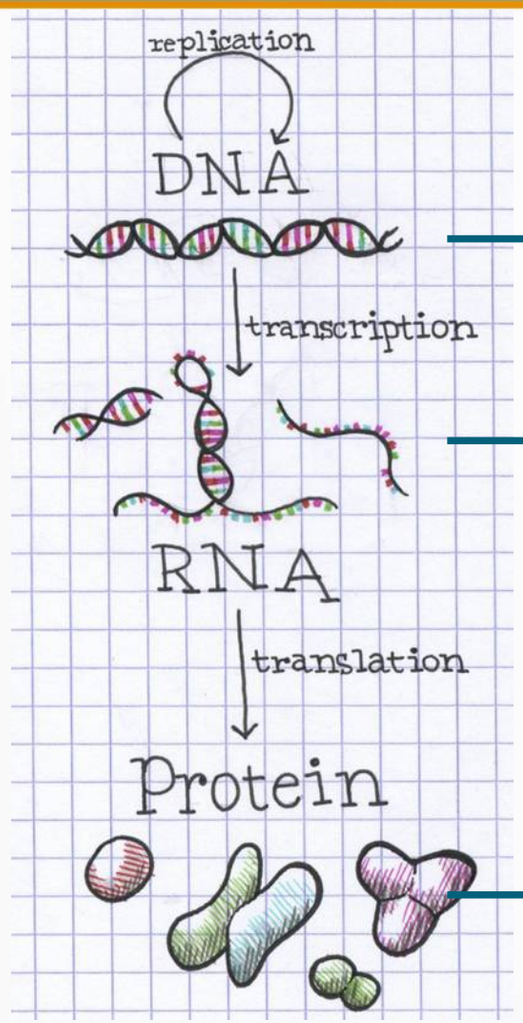


http://cisncancer.org/research/what_we_know/omics/metabolomics.html



MOREIRA, L. M. (organizador). Ciências genômicas - fundamentos e aplicações. Ribeirão Preto: Sociedade brasileira de genética. 403p., 2015

Dogma central e análises ômicas



Biologia de Sistema

Genômica

Transcriptômica

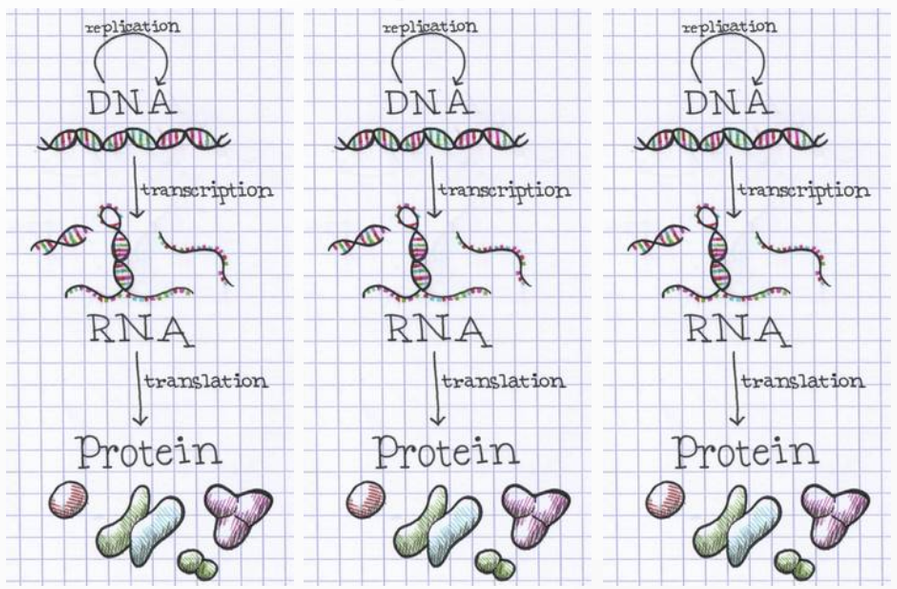
Proteômica
Interatômica

Produtos

Metabolômica

Metaômicas

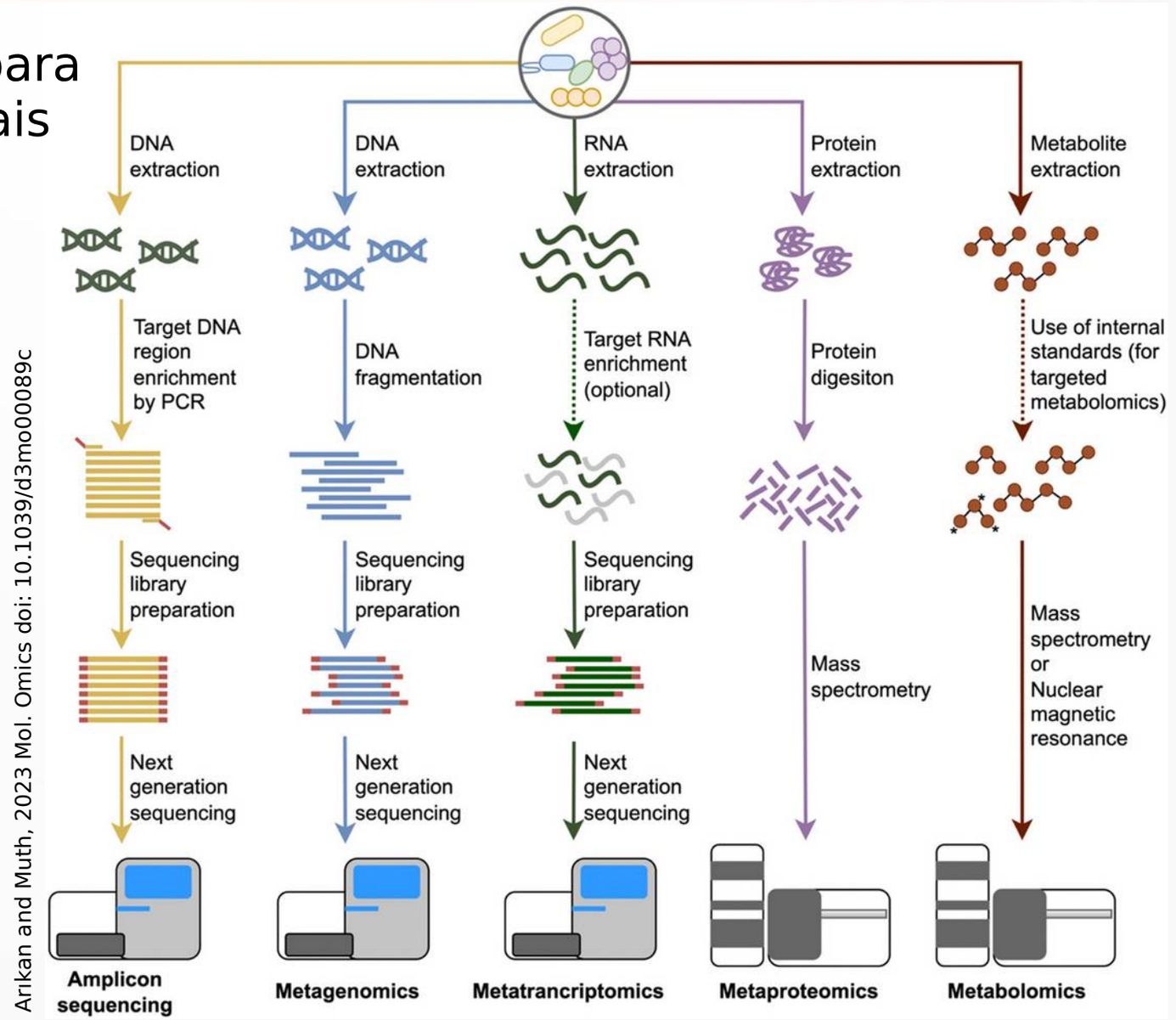
Comunidades de organismos ocupando um nicho ecológico



Tradicionalmente, a Bioinformática foi usada para descrever a ciência para armazenamento e análise de dados de sequêncais de biomoléculas.

Abordagens em metaômicas

Fluxos de trabalho experimentais para tipos de análise ômica convencionais



Estudando comunidades de microorganismos através da metagenômica

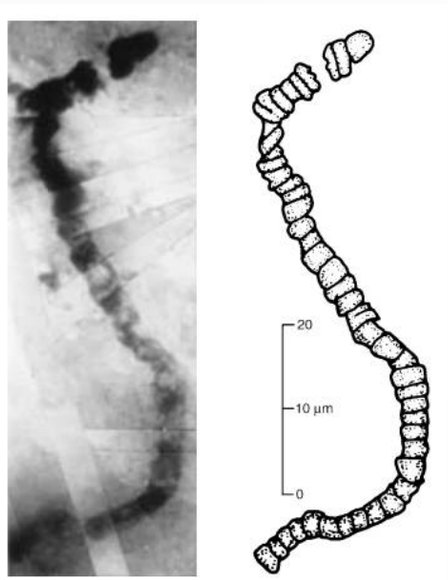
Metagenômica

O termo Metagenômica foi usado pela primeira vez em 1998 pro Jo Handelsman (Universidade de Wiscosin – EUA)

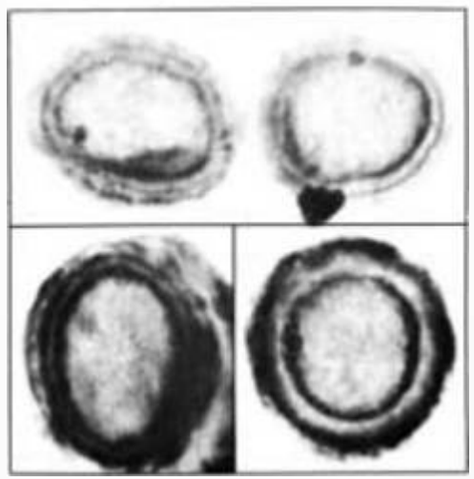
METAGENOMA
Genoma coletivo da microbiota encontrada em um habitat

Análise genômica das comunidades de microrganismos de um determinado ambiente por técnicas independentes de cultivo

Por que estudar o metagenoma?



1g de amostra de solo pode conter 10 bilhões de microrganismos de milhares de espécies

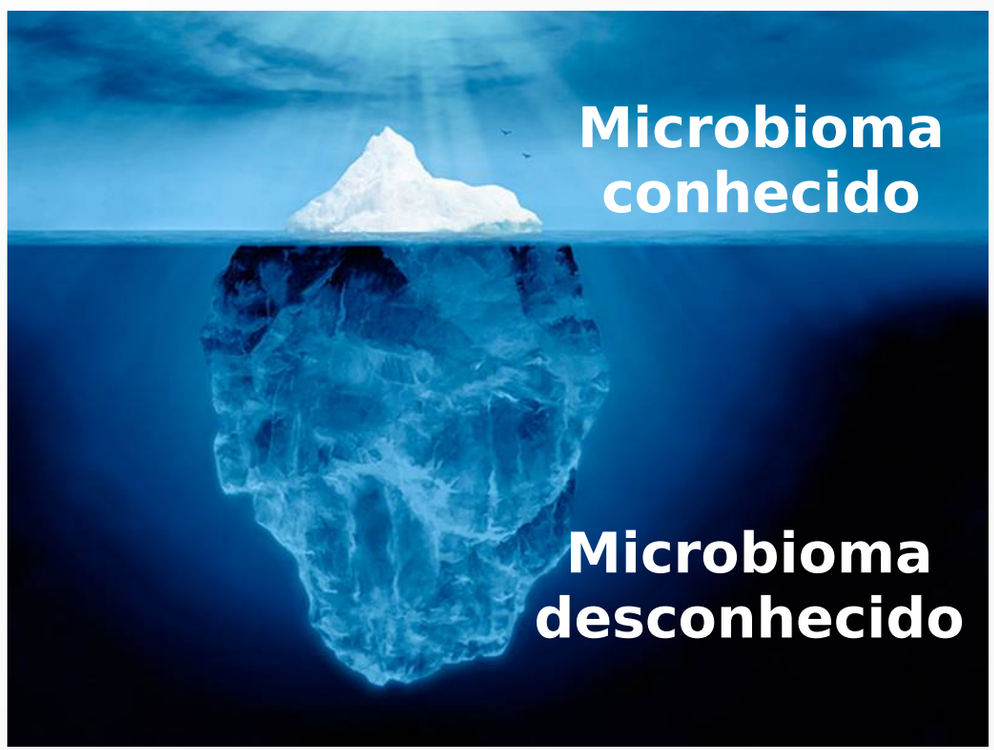


Somente uma pequena porcentagem das espécies bacterianas existentes pode ser acessada por técnicas tradicionais dependentes de cultivo

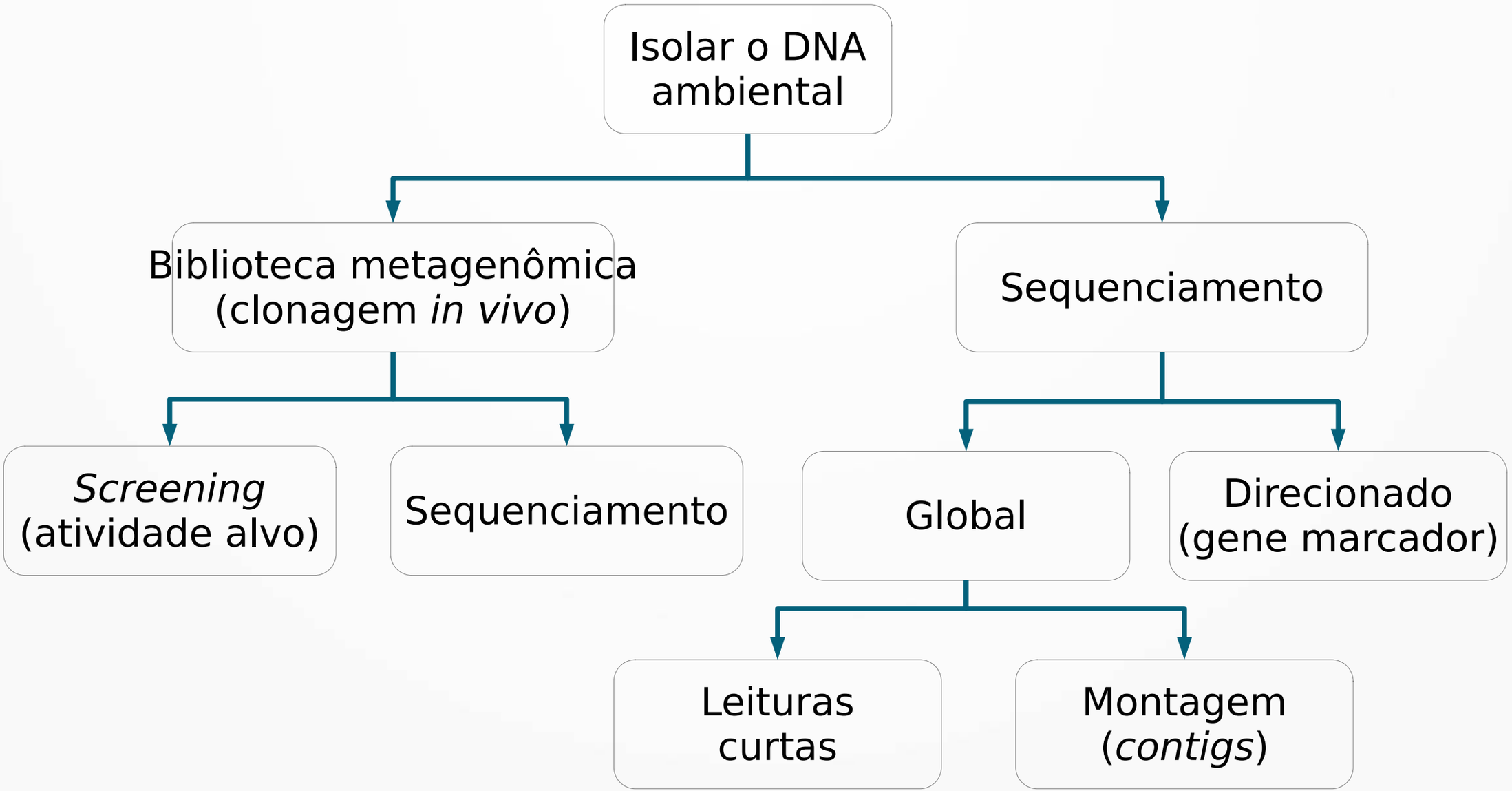
Por que estudar o metagenoma?

Staley e Konopka (1985)
→ 0,1-1,0% das bactérias presentes no solo podem ser cultivadas em laboratório
→ Em ambientes aquáticos este número é ainda menor

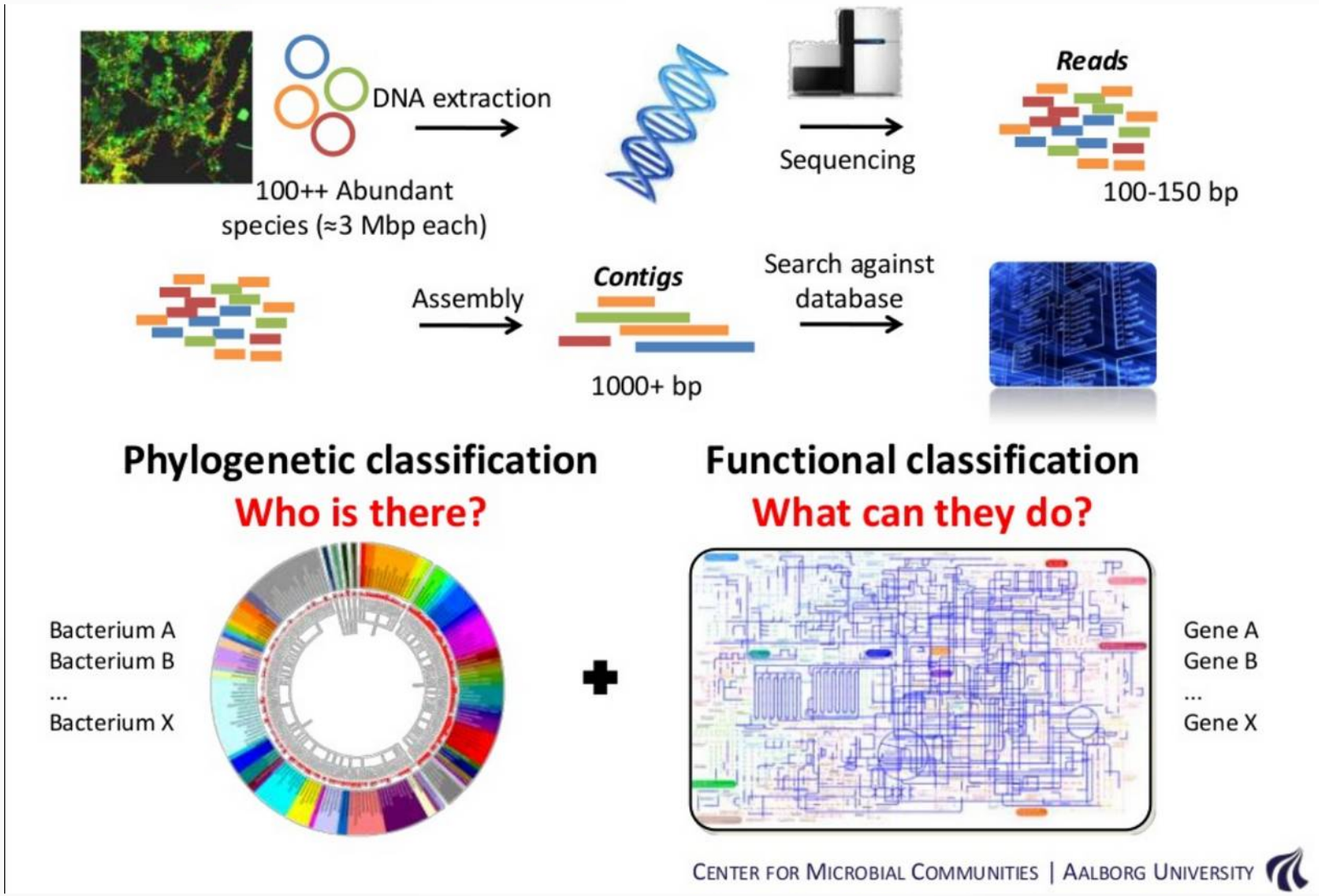
Os organismos cultiváveis não são necessariamente os dominantes ou mais influentes em um ambiente



Abordangens na metagenômica

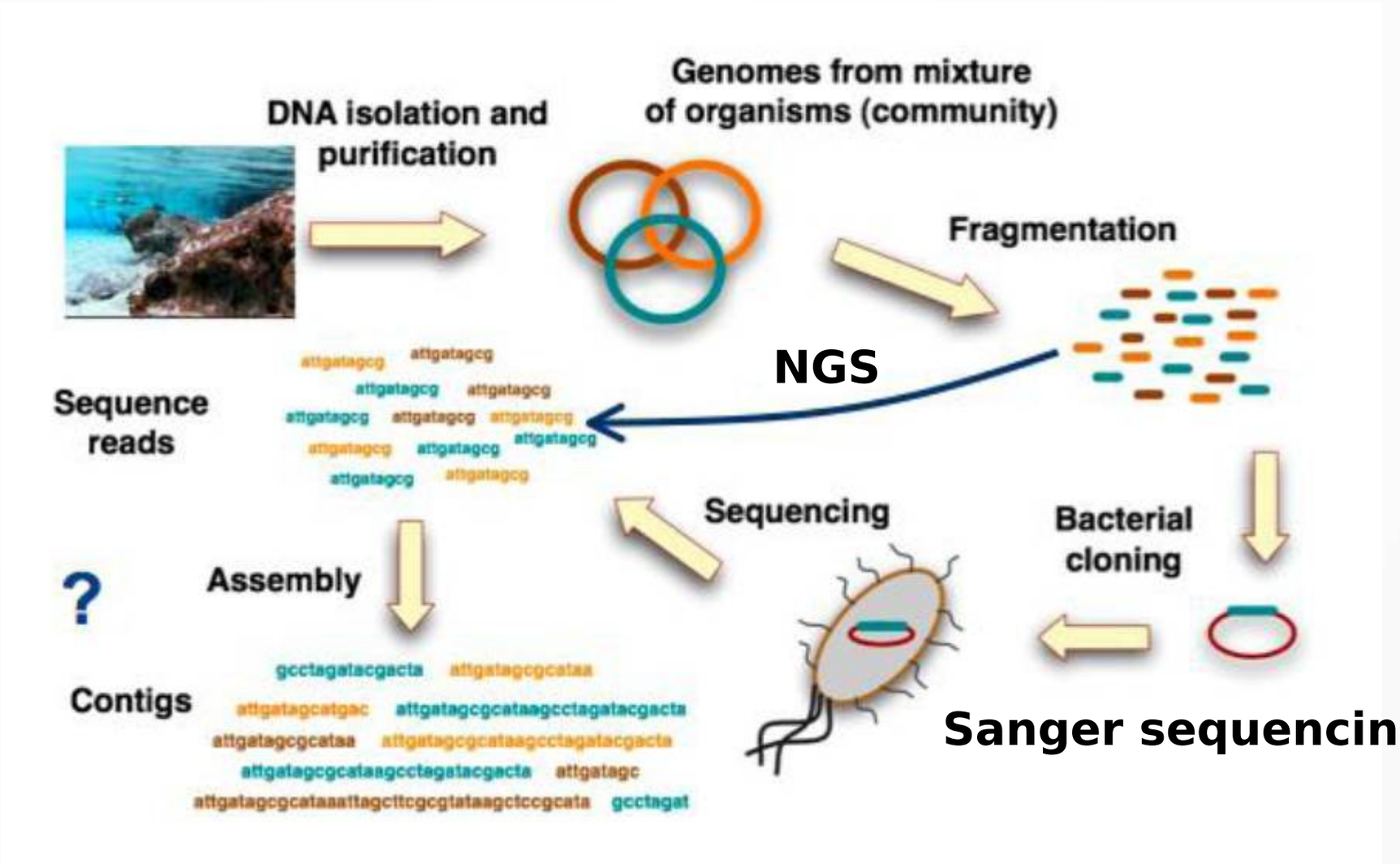


Metagenômica por sequenciamento



https://pt.slideshare.net/MadsAlbertsen/20131202-mads-albertsen-extracting-genomes-from-metagenomes

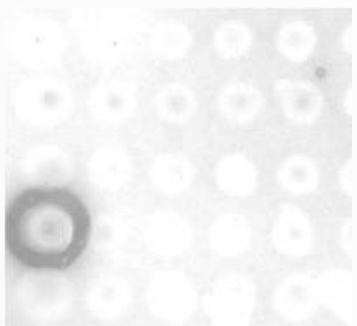
Metagenômica por sequenciamento



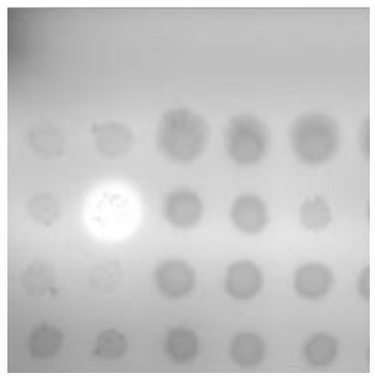
Prospecção funcional de enzimas

Seleção de clones com atividade de interesse – meio de cultivo de triagem

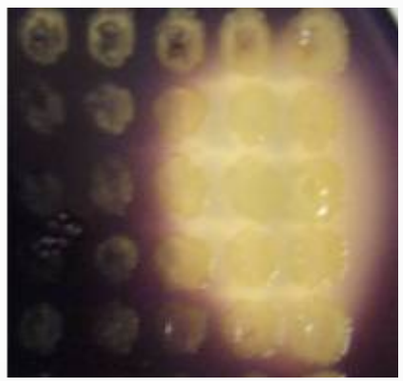
Esterase



Lipase

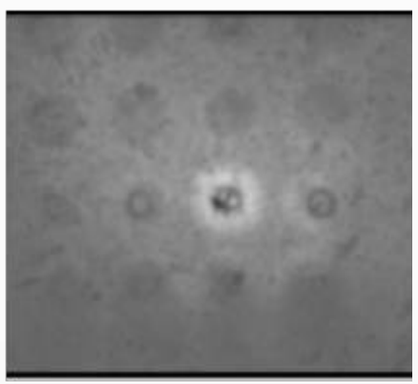


Amilase

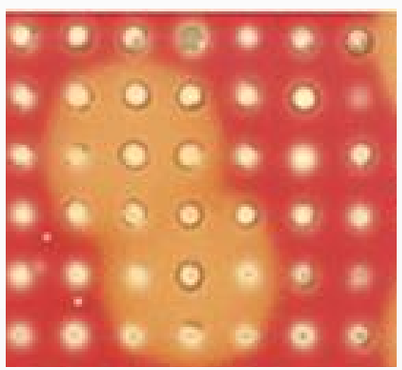


Protease

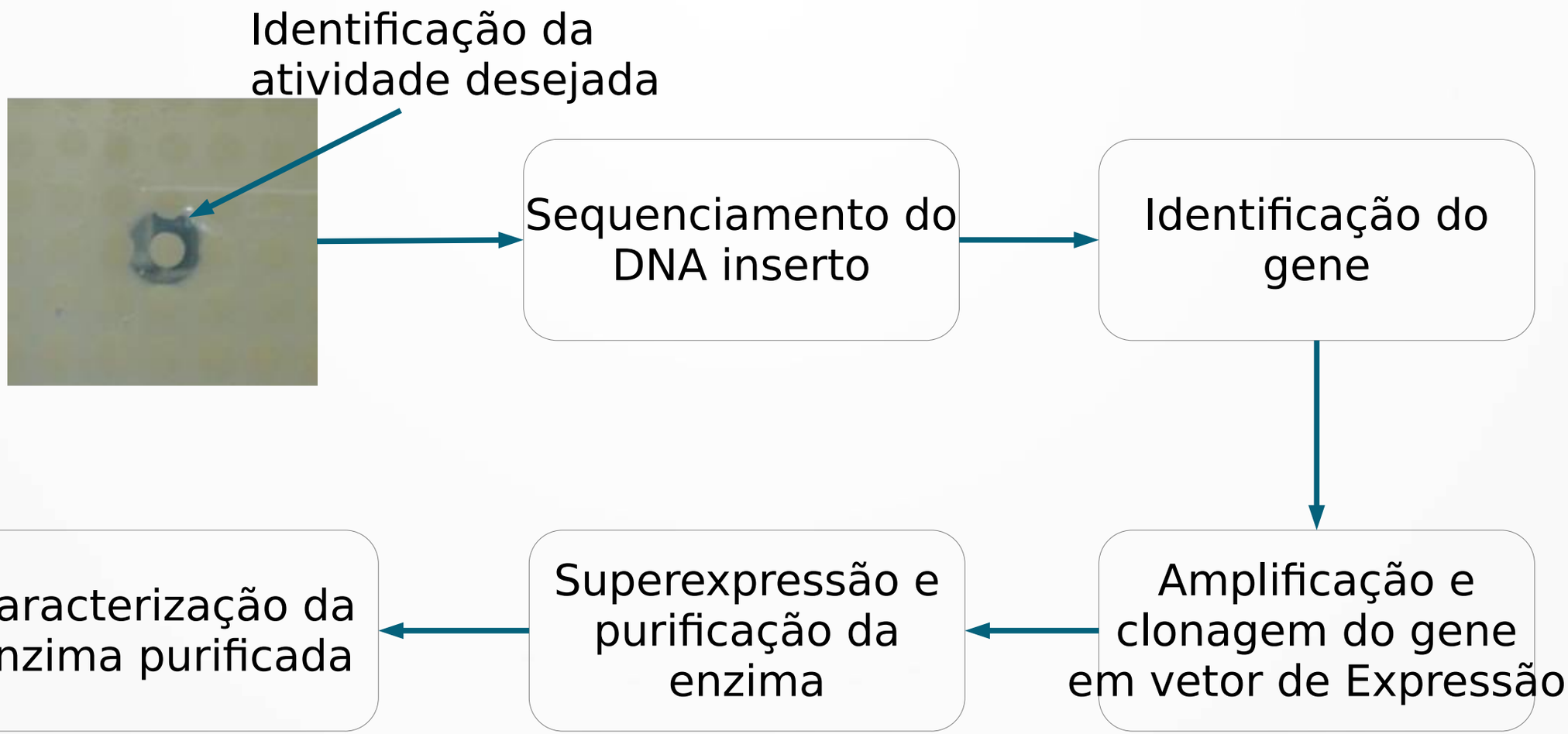
Quitinase



Celulase



Caracterização de uma enzima metagenômica

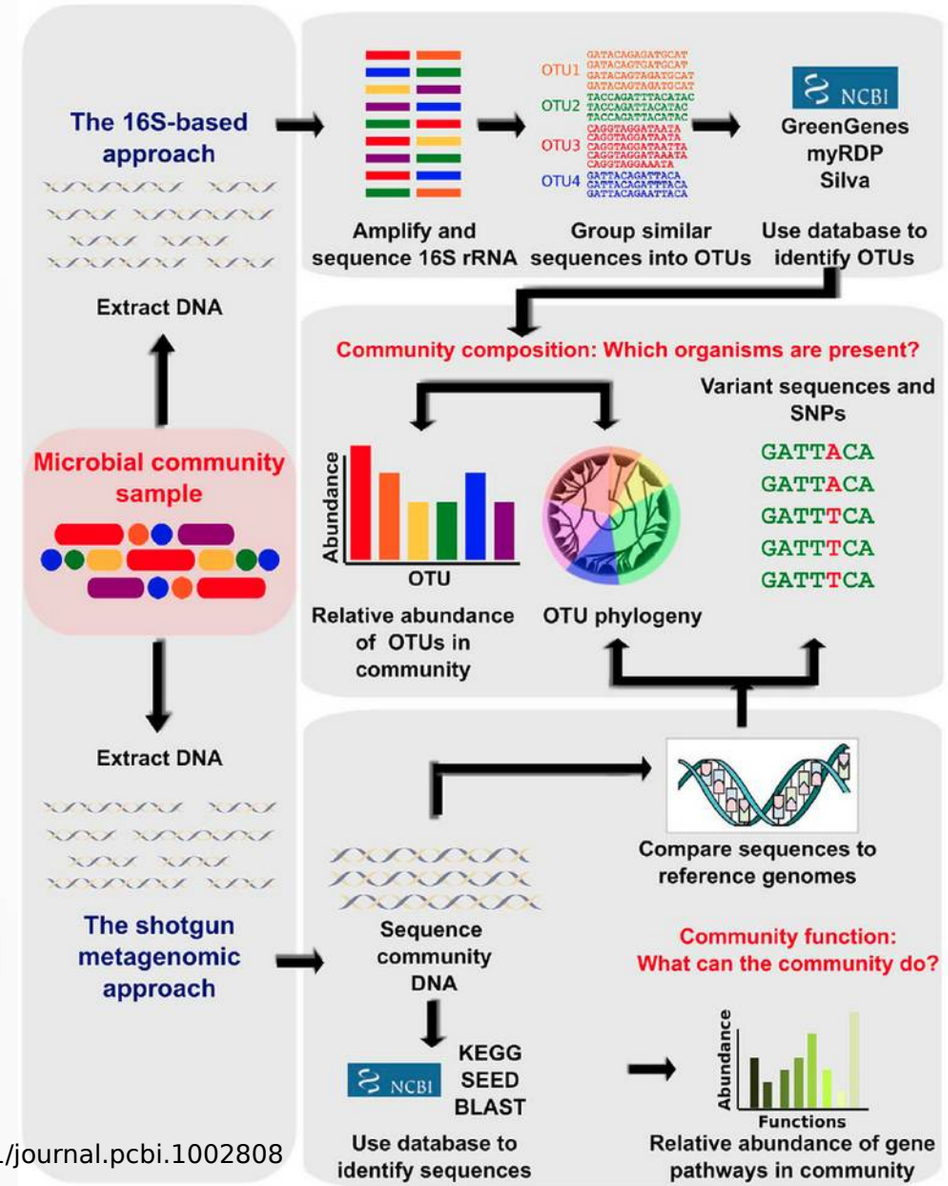


Análise metagenômica baseada em marcador genético
Gene 16S rRNA

Análise metagenômica baseada em sequenciamento

Gene alvo

Metagenoma total



doi: 10.1371/journal.pcbi.1002808

O sequenciamento metagenômico permite:

1. Identificação taxonômica
   - gene alvo
   - metagenoma total

2. Identificação funcional
   - gene alvo (aproximação)
   - metagenoma total

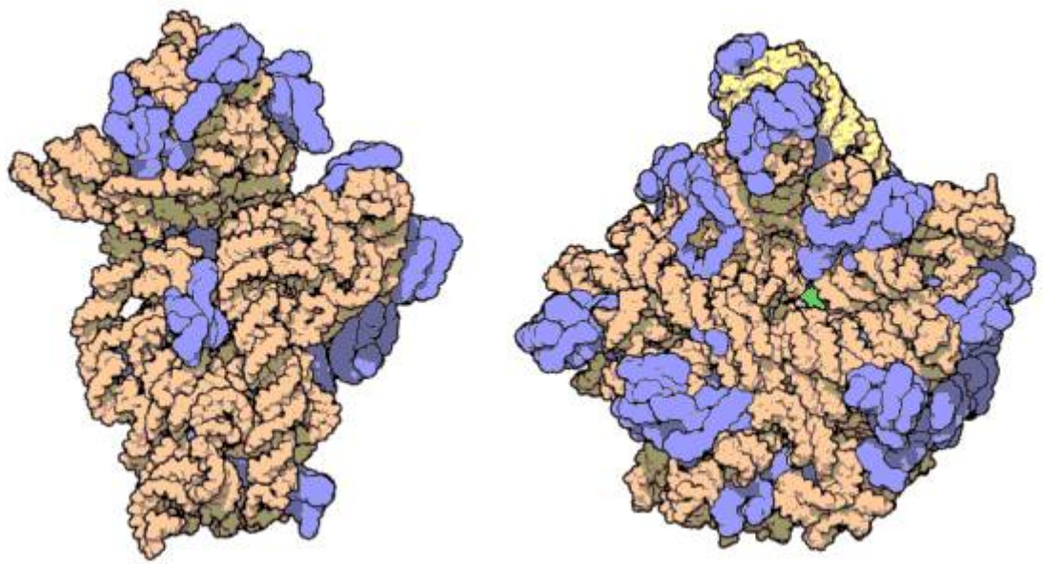
RNA ribossomais

- Encontrados em todos os organismos
  - Três em procariotos (5S rRNA, 16S rRNA e 23S rRNA)
  - Também presente nos genomas de mitocôndrias e cloroplastos
- Múltiplas cópias por genoma
- Usado na taxonomia de procariotos
- Resolução de classificação no nível de gênero ou superior
  - POUCA RESOLUÇÃO NO NÍVEL DE ESPÉCIE
- Outros genes marcadores
  - 18S rRNA – Microrganismos eucariontes
  - ITS (região inergênica para rRNA) – frequentemente usado para fungos
  - COI (citocromo c oxidase no genoma mitocondrial) – frequentemente usado para animais

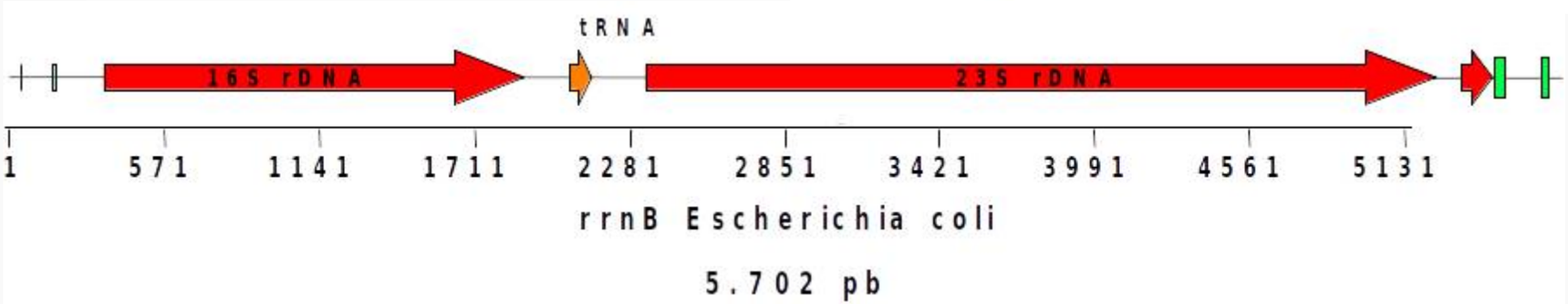
Os genes RNA ribossomais

→ RNA ribossomais (procariotos)
    5S (~120pb)
    16S (~1.500pb)
    23S (~2.900pb)
→ 16S rRNA – mais usado em filogenia/taxonomia
    Contém regiões conservadas e variáveis – identificação em vários níveis taxonômicos

rRNA – laranja e amarelo
Proteínas – azul



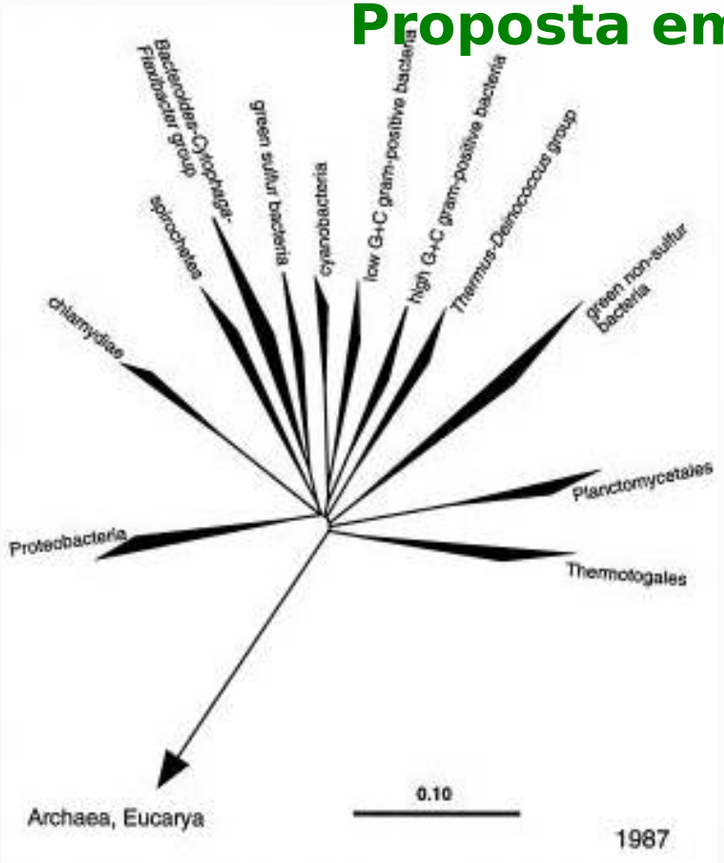
Operon rRNA



Importância do gene 16S rRNA na microbiologia

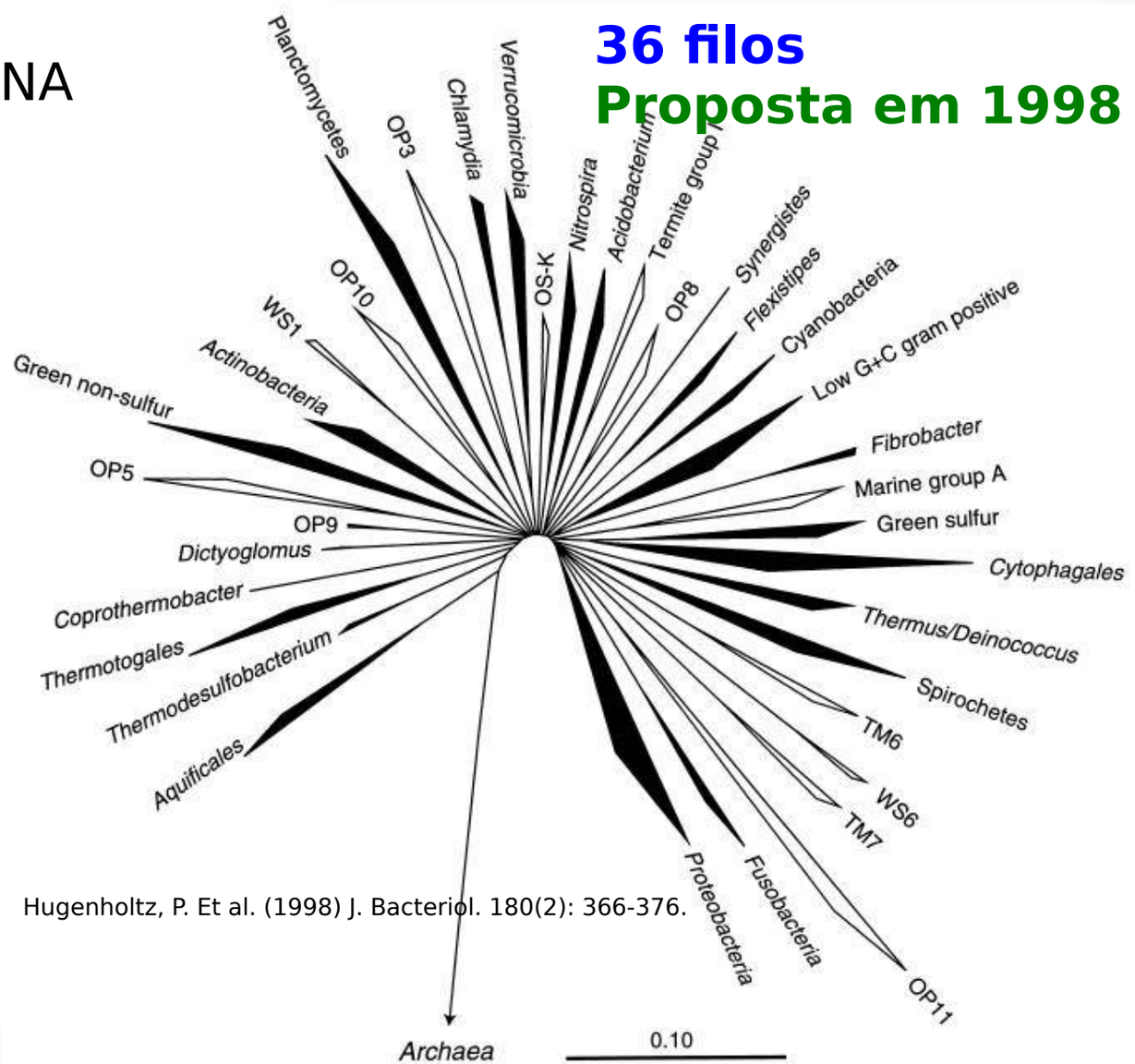
Árvore filogenética do domínio bactéria baseada em sequências do gene 16S rRNA

12 filos
Proposta em 1987



Modificado de Hugenholtz, P. Et al. (1998) J. Bacteriol. 180(18): 4765-4774.

36 filos
Proposta em 1998



Hugenholtz, P. Et al. (1998) J. Bacteriol. 180(2): 366-376.

Importância do gene 16S rRNA na microbiologia

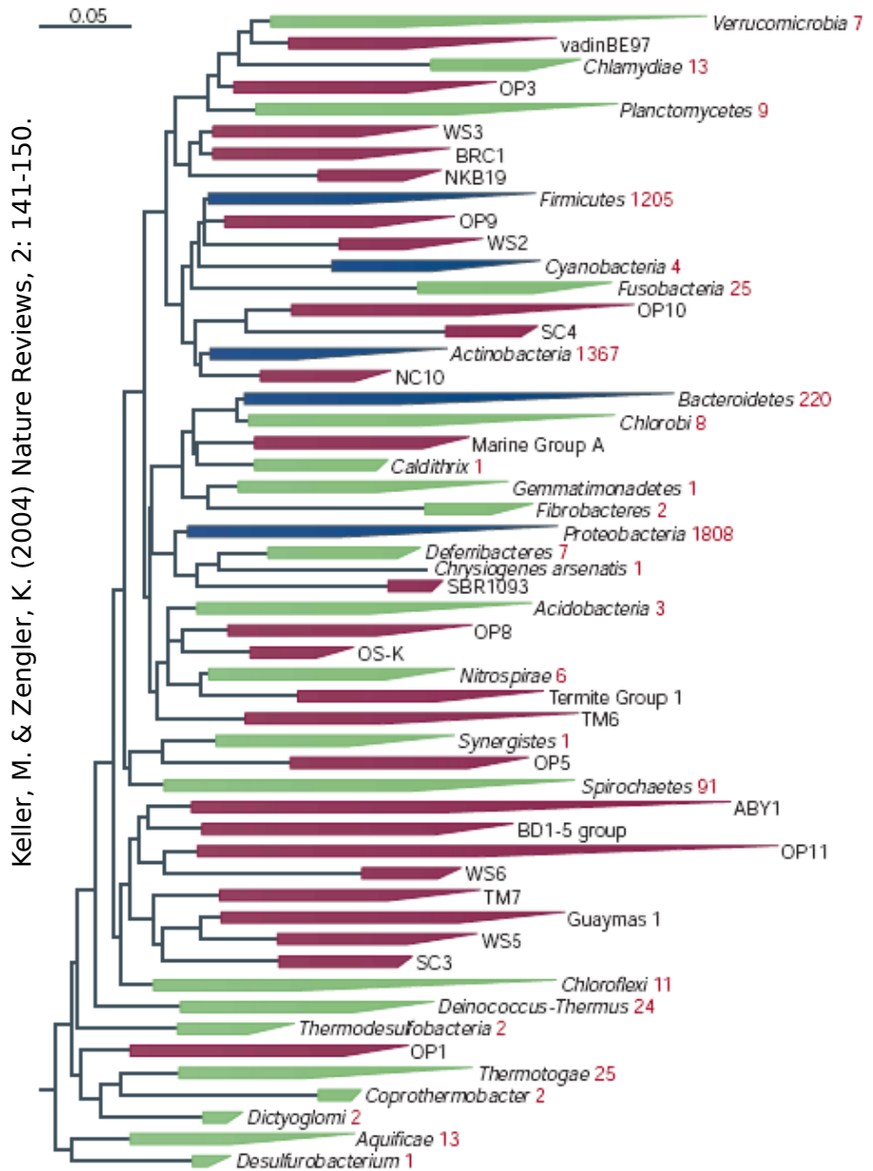
Filos de bactérias com representantes cultiváveis (azul e verde)

Os números em vermelho representam o número de espécies publicadas para um filo

53 filos
26 sem representantes cultivados
Proposta em 2003

Maioria de organismos NÃO cultivados!

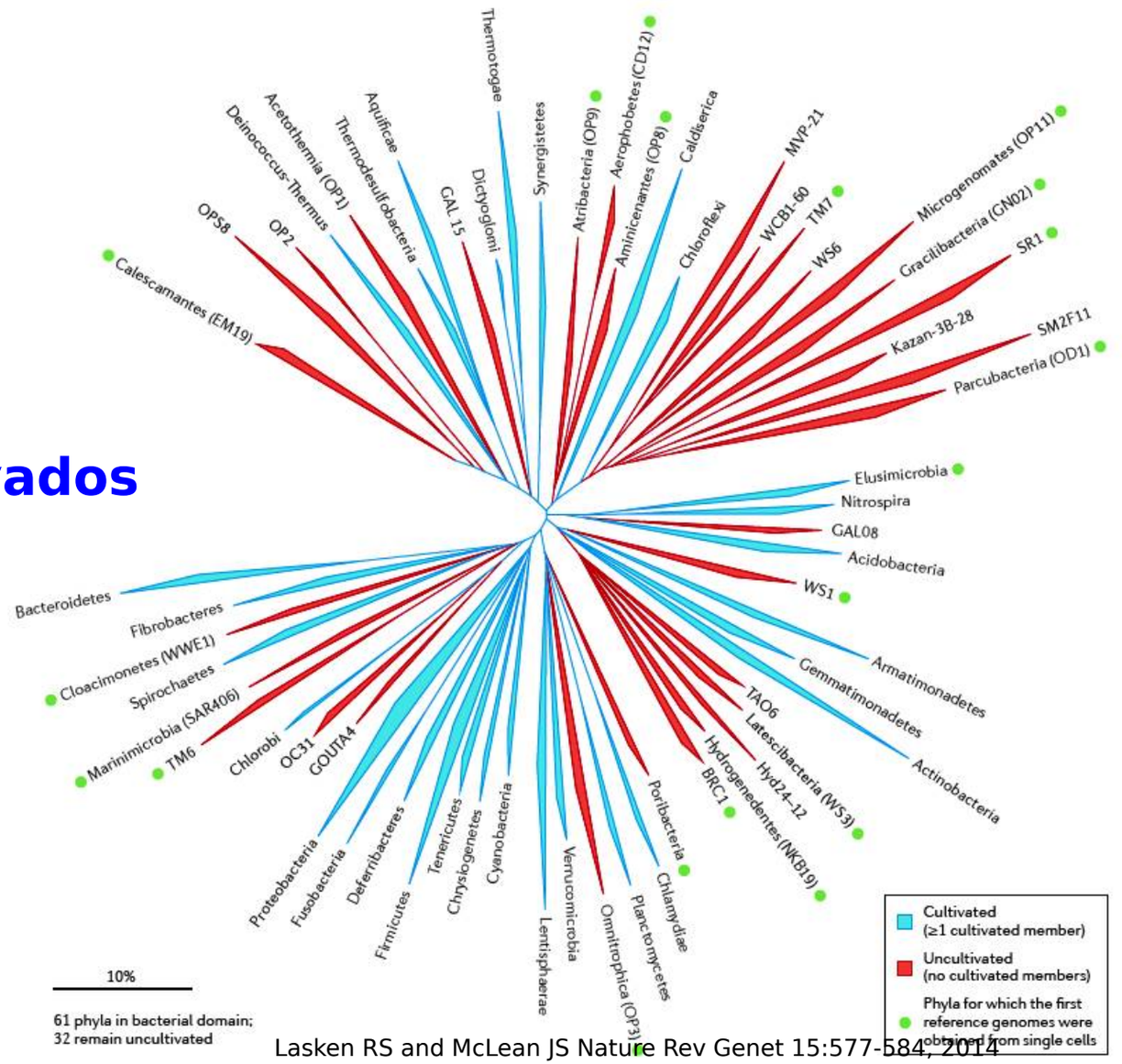
Árvore filogenética baseada em sequências do gene 16S rRNA



Importância do gene 16S rRNA na microbiologia

Árvore filogenética do domínio bactéria baseada no gene 16S rRNA

61 filos
32 sem representantes cultivados
19 com sequência genômica
Proposta em 2014



Lasken RS and McLean JS Nature Rev Genet 15:577-584, 2014

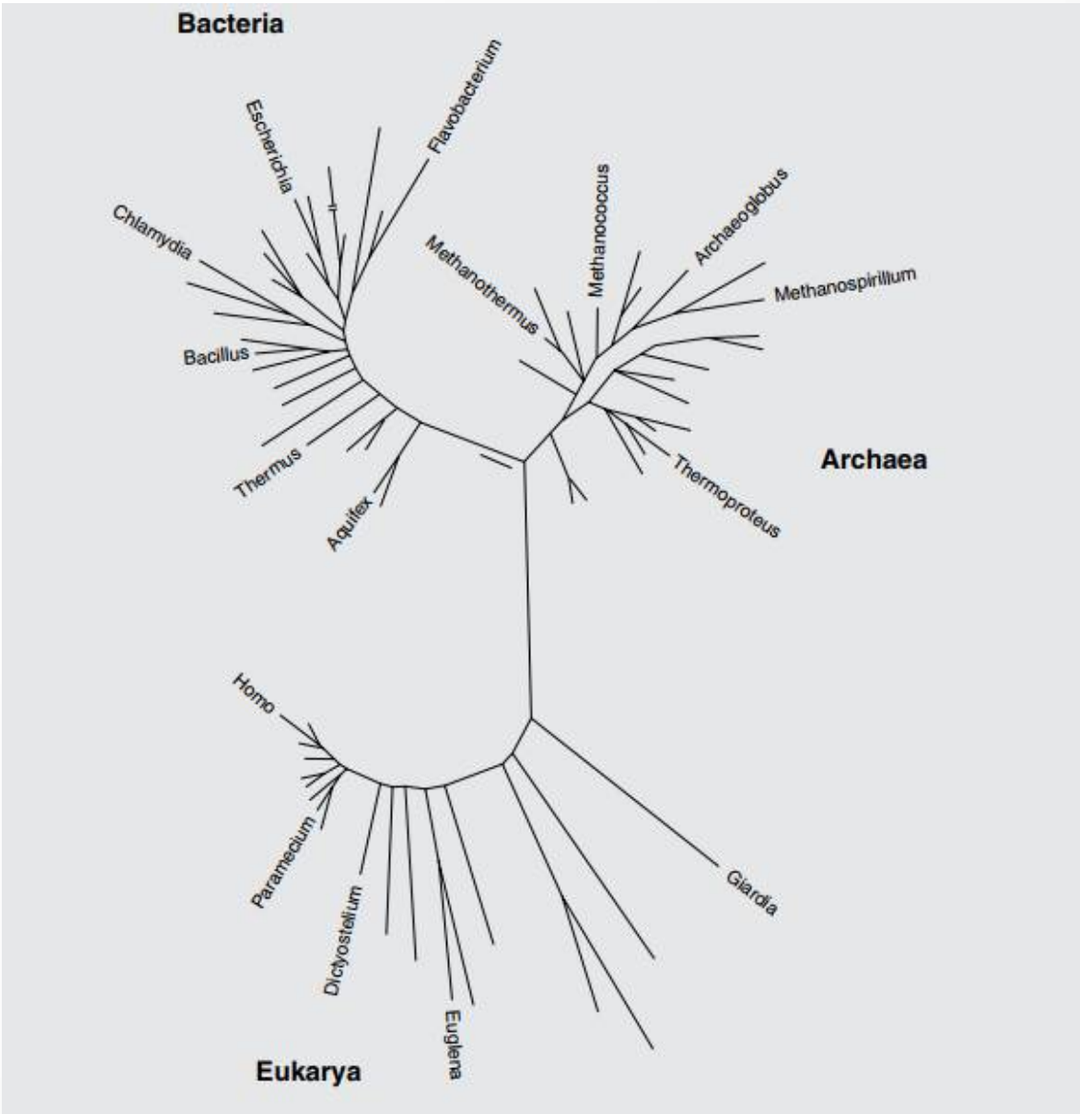
Importância do gene 16S rRNA na microbiologia

→ Compõe o ribossoma de procariotos

→ Dois organismos próximos terão sequências similares, enquanto organismos distantes terão sequências pouco similares

→ *primers* universais para PCR e sequenciamento

→ Permite a identificação taxonômica de grupos conhecidos e estimativa da diversidade de espécies



Análise metagenômica com o gene 16S rRNA

A abundância relativa de grupos taxonômicos é comparada em diferentes amostras

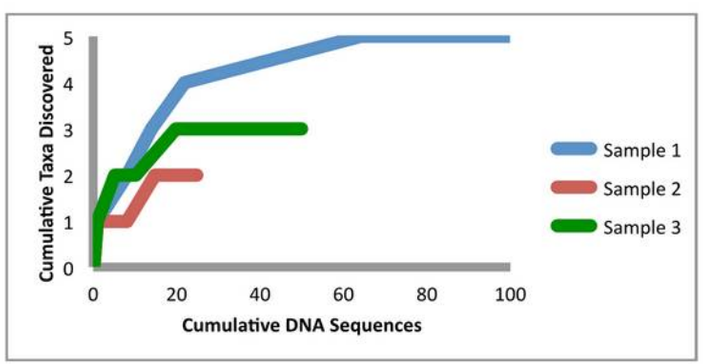
A) Sequence Abundance

OTU	Sample 1	Sample 2	Sample 3
A	60	0	35
B	24	5	5
C	10	0	0
D	5	0	0
E	1	0	0
F	0	20	10
Total	100	25	50

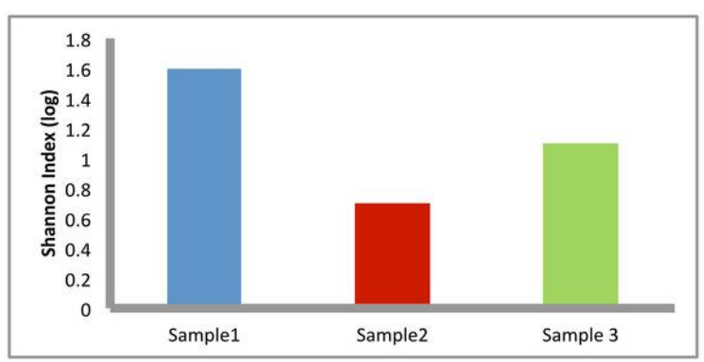
B) Sequence Relative Abundance

OTU	Sample 1	Sample 2	Sample 3
A	0.60	0	0.70
B	0.24	0.20	0.10
C	0.10	0	0
D	0.05	0	0
E	0.01	0	0
F	0	0.80	0.20
Total	1.0	1.0	1.0

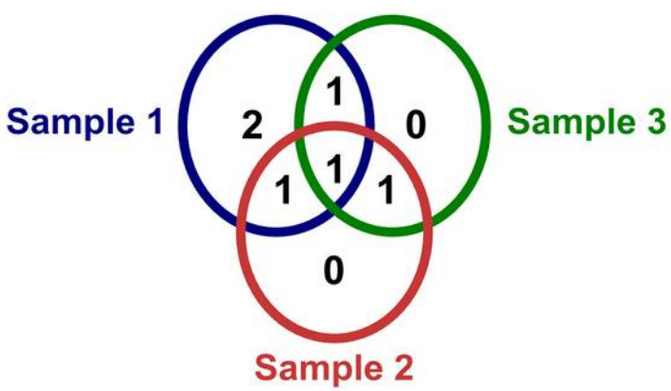
C) Collector's Curve of Sample Richness



D) Within-Sample Alpha Diversity



E) Between-Sample Beta Diversity



doi: 10.1371/journal.pcbi.1002808

Medidas de diversidade

Medidas de diversidade em escala espacial

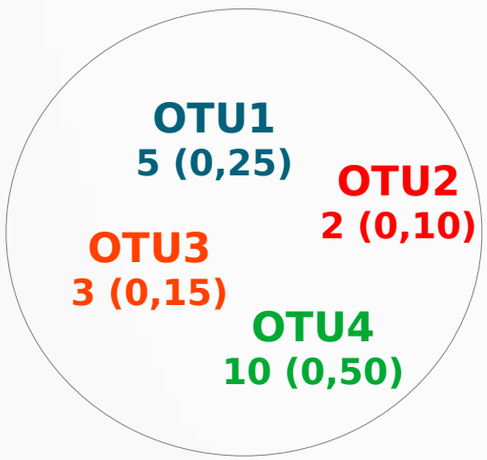
- **Alfa diversidade**
  - Dentro de uma área/ecossistema particular
  - Considera número de espécies e/ou homogeneidade
  - Gera valores por amostra
- **Beta diversidade**
  - Comparação ente as diversidades de duas áreas/ecossistemas
  - Considera alterações nas quantidades de espécies (distâncias)
  - Gera valores por par de amostra (matriz)
- **Gama diversidade**
  - Medida da diversidade global de uma grande região

Hypothetical species	Woodland habitat	Hedgerow habitat	Open field habitat
A	X		
B	X		
C	X		
D	X		
E	X		
F	X	X	
G	X	X	
H	X	X	
I	X	X	
J	X	X	
K		X	
L		X	X
M			X
N			X
Alpha diversity	10	7	3
Beta diversity	Woodland vs. hedgerow: 7	Hedgerow vs. open field: 8	Woodland vs. open field: 13
Gamma diversity	14		

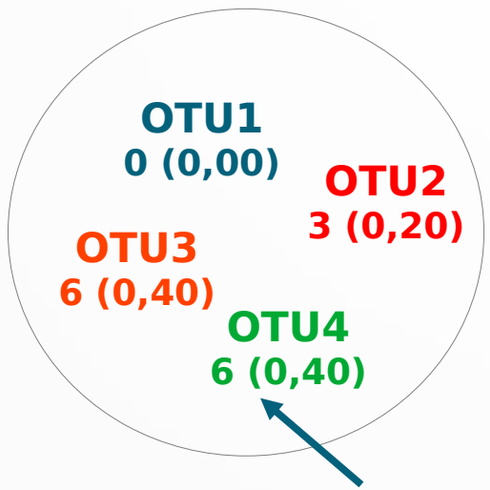
https://bio.libretexts.org/Bookshelves/Ecology/Biodiversity_(Bynum)/7%3A_Alpha_Beta_and_Gamma_Diversity
https://bio.libretexts.org/Courses/Gettysburg_College/01%3A_Ecology_for_All/22%3A_Biodiversity/22.02%3A_Diversity_Indices

Níveis de diversidade

Amostra 1



Amostra 2



Diversidade alfa

Diversidade dentro das amostras
(quantifica riqueza e equitabilidade)

Amostra 1
Riqueza = 4 OTU

Amostra 2
Riqueza = 3 OTU

Diversidade beta

Diversidade entre amostras (quantifica (dis)similaridade)

Índice de Jaccard
OTU compartilhadas / total de OTU observadas

Amostra 1 = OTU1, OTU2, OTU3, OTU4
Amostra 2 = OTU2, OTU3, OTU4

Jaccard = 3 / 4 = 0,75

https://www.statology.org/jaccard-similarity

Agrupamento de OTU vs denoising

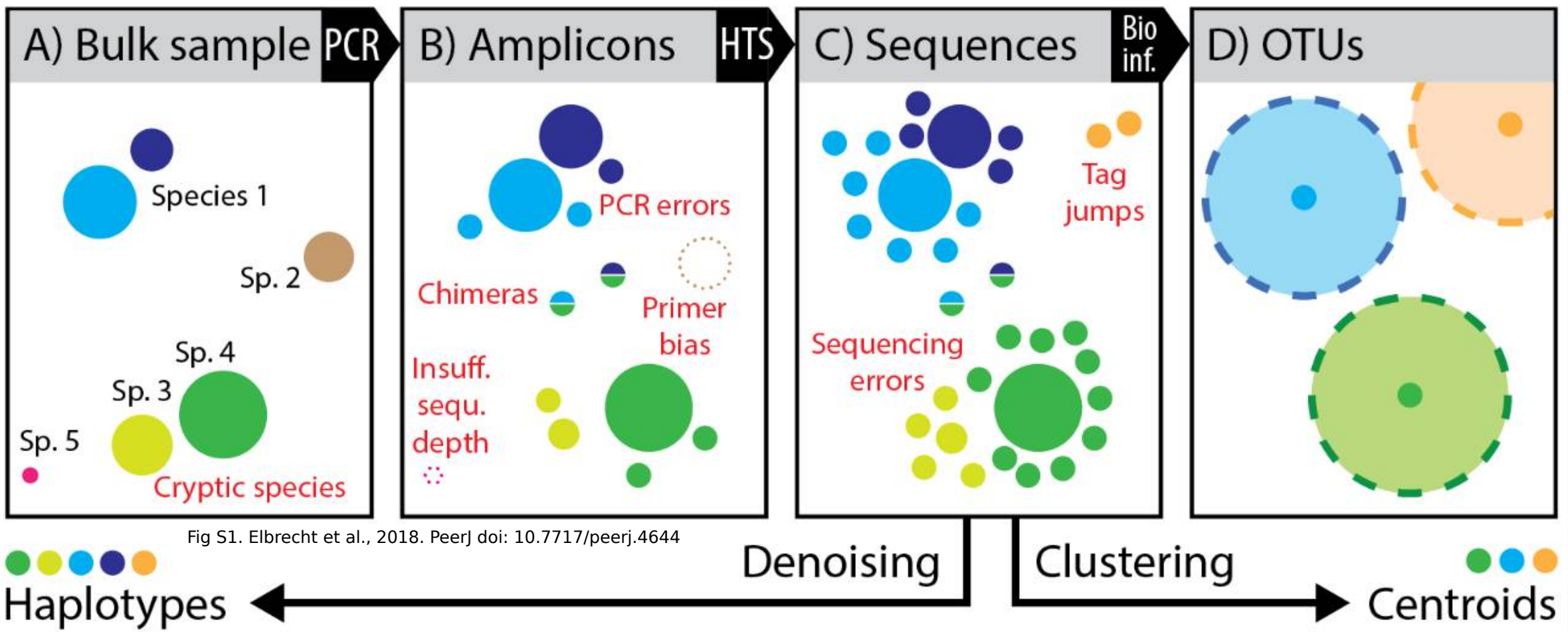


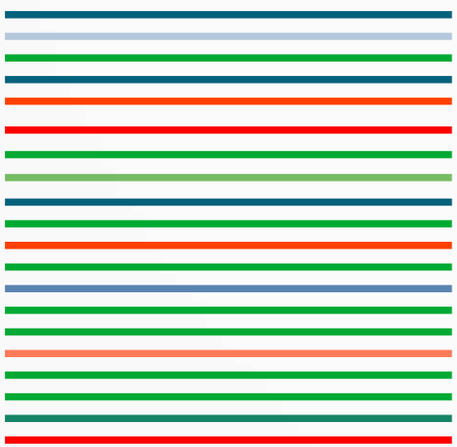
Fig S1. Elbrecht et al., 2018. PeerJ doi: 10.7717/peerj.4644

Metagenômica com o gene 16S rRNA

DNA metagenômico → Amplificação 16S rRNA → Agrupamento das sequências → Abundância OTU absoluta (relativa)

Identificação taxonômica

Análise de diversidade



OTU1 — 5 (5/20=0,25)

OTU2 — 2 (2/20=0,10)

OTU3 — 3 (3/20=0,15)

OTU4 — 10 (10/20=0,50)

OTU (*Operational Taxonomic Unit*) = Unidade Taxonômica Operacional

Grupos de sequências formados por graus de identidade para representar um nível taxonômico (ex., grupos com 97% de identidade representam espécies)

Métodos para agrupamento de OTU

- ***Closed reference OTU picking***
  - Usa um banco de dados de sequências referências para agrupamento
  - Variantes biológicas podem ficar de fora do agrupamento
  - Sujeito a tendências ou erros nos bancos de dados
  - Permite comparar estudos se os mesmos bancos de dados forem usados
- ***De novo OTU picking***
  - Sequências são agrupadas somente por comparação dentro do conjunto de dados
  - Não dependente de banco de dados
  - Não permite comparação entre estudos

- ***Open reference OTU picking***
  - combinação dos anteriores
  - Primeiro a comparação é feita com referências (banco de dados)
  - Sequências não agrupadas com referência são agrupadas com método *de novo*
  - Não permite comparação entre estudos

Identificação e distribuição taxonômica

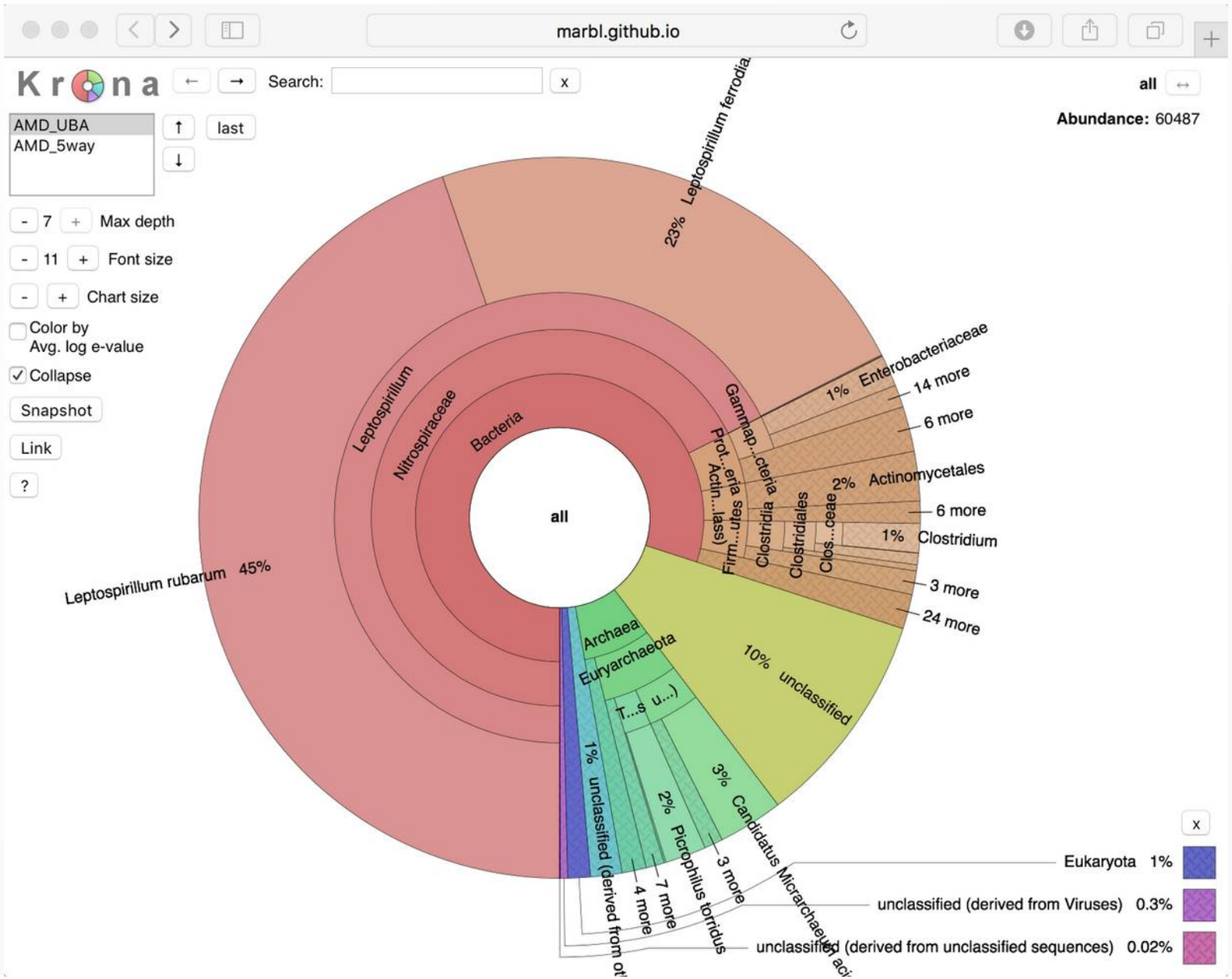
Sequências de amostras metagenômicas

Sequências de espécies conhecidas

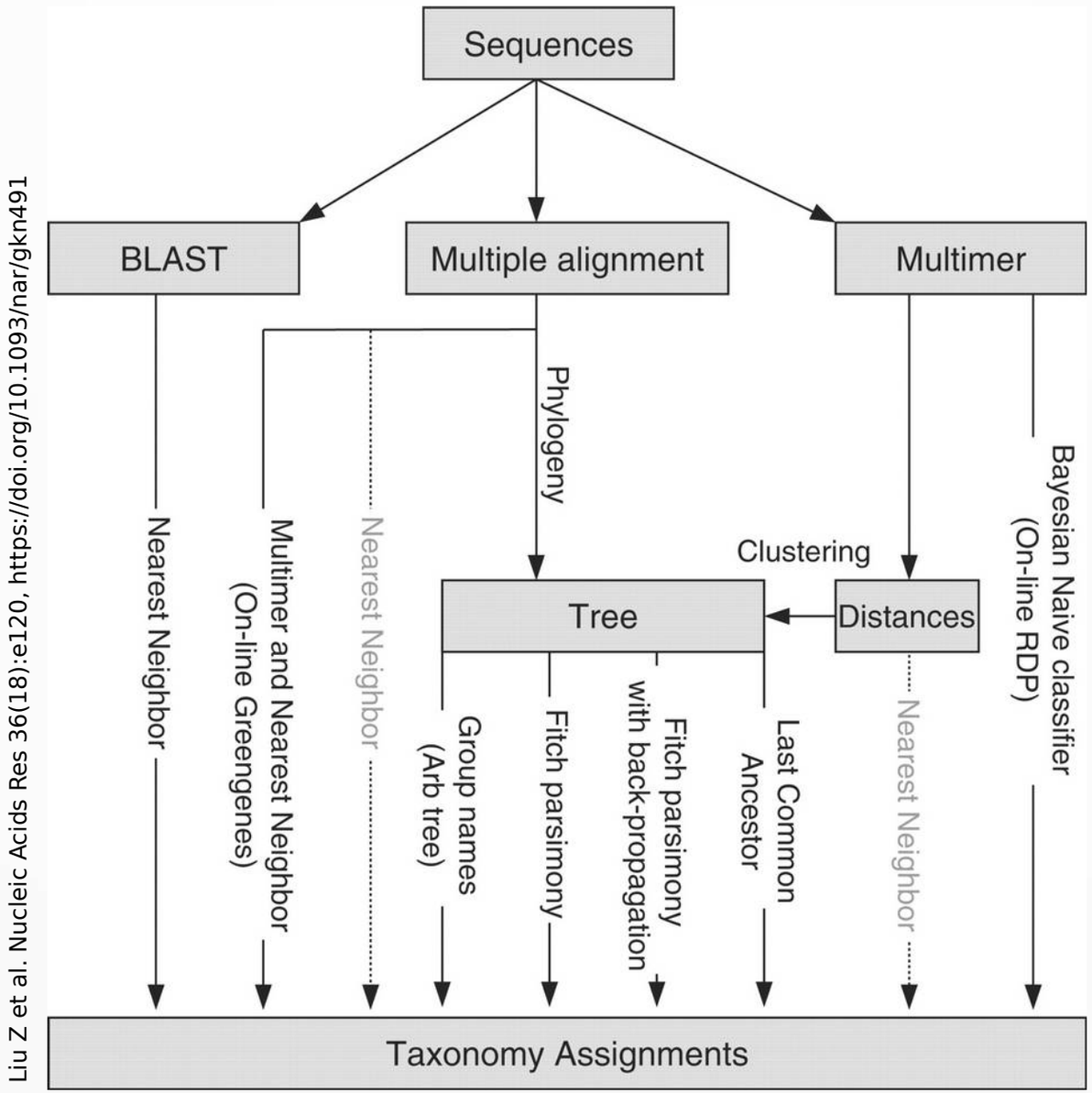
Banco de dados

OTU

Identificação taxonômica



Identificação taxonômica



Liu Z et al. Nucleic Acids Res 36(18):e120, https://doi.org/10.1093/nar/gkn491

Técnicas que usam kmer

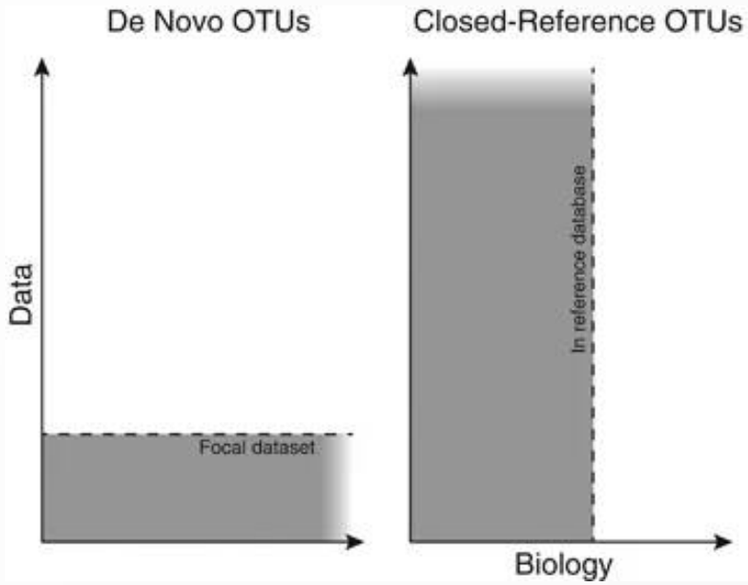
Denoising

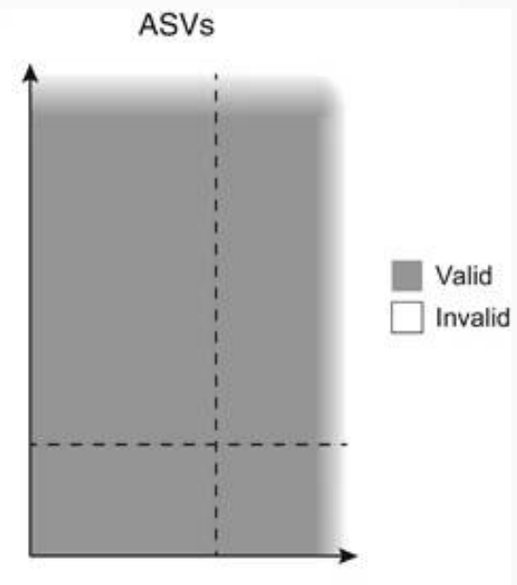
- Nesta abordagem a sequência biológica exata é inferida e erros são removidos
- Um modelo para os erros é aplicado, dependente das informações de qualidade das sequências
- Diferntes métodos:
  - DADA2, Deblur, MED e UNOISE
  - Adaptados para plataforma Illumina

https://www.zymoresearch.com/blogs/blog/microbiome-informatics-otu-vs-asv

OTU vs. ASV

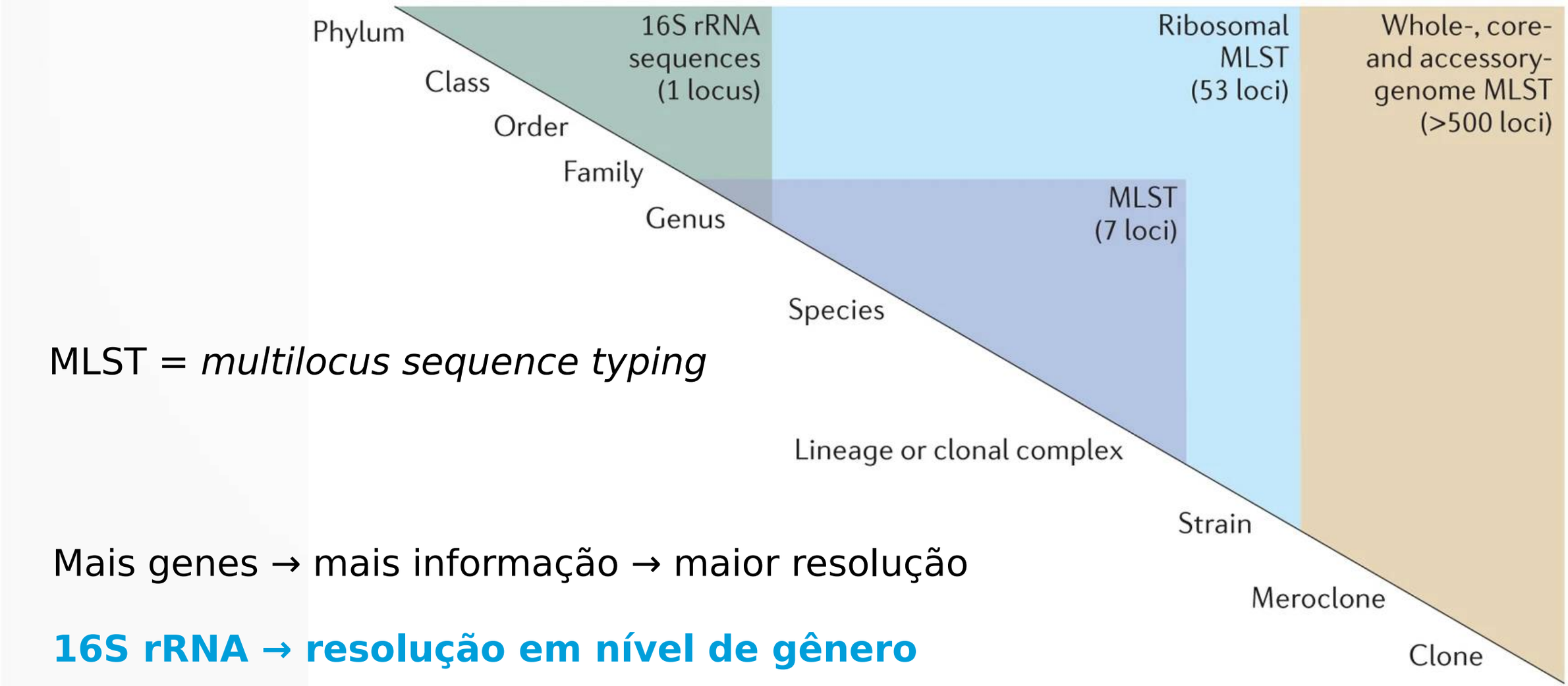
- ***operational taxonomic units* (OTUs)**
  - Agrupamentos de leituras de sequências que diferem por menos de um limite de dissimilaridade fixado (normalmente 3%)
  - Limitada ao conjunto de dados (*de novo*)
  - Limitada ao banco de dados (*closed-reference*)
- ***amplicon sequence variants* (ASVs)**
  - Sequências biológicas são inferidas corrigindo erros de sequenciamento
  - Variantes por diferença de até uma base





Callahan et al. ISME J 11, 2639–2643 (2017) doi: 10.1038/ismej.2017.119

Resolução para identificação taxonômica



MLST = *multilocus sequence typing*

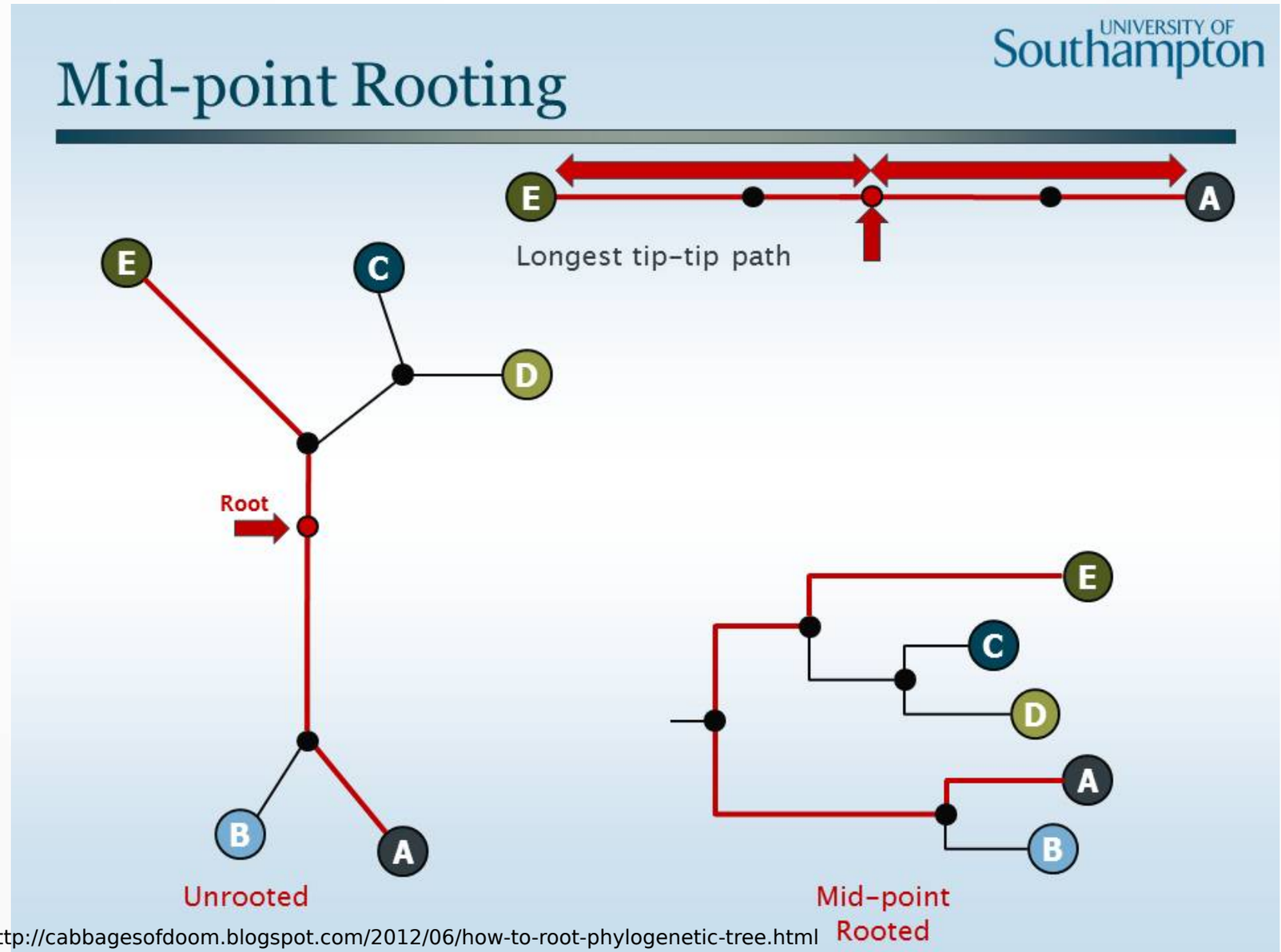
Mais genes → mais informação → maior resolução

16S rRNA → resolução em nível de gênero

Maiden et al. Nat Rev Microbiol 11, 728–736 (2013) doi: 10.1038/nrmicro3093
https://www.genome.gov/sites/default/files/media/files/2021-07/Sean_Conlan_Josh_Farr_Lesson_Plan.pdf

Nature Reviews | Microbiology

Árvore filogenética



http://cabbagesofdoom.blogspot.com/2012/06/how-to-root-phylogenetic-tree.html

Iniciadores e regiões amplificadas

Table 8.1 Primers for 16S rRNA metagenomics[a]

Target	Forward 5′-3′	Reverse 5′-3′	Size[c]	Reference
V3[b]	CCTACGGGAGGCAGCAG	ATTACCGCGGCTGCTGG	194	Lane (1991); Danzeisen et al. (2011)
V1–V3	AGAGTTTGATCCTGG	ATTACCGCGGCTGCTGG	527	Lane (1991);
				Danzeisen et al. (2011)
V3–V4	GGAGGCAGCAGTRRGGAAT	CTACCTGGGTATCTAATCC	457	Nossa et al. (2010)
V4–V5	GTGYCAGCMGCCGCGGTA	CCCGYCAATTCMTTTRAGT	413	Tang et al. (2014)

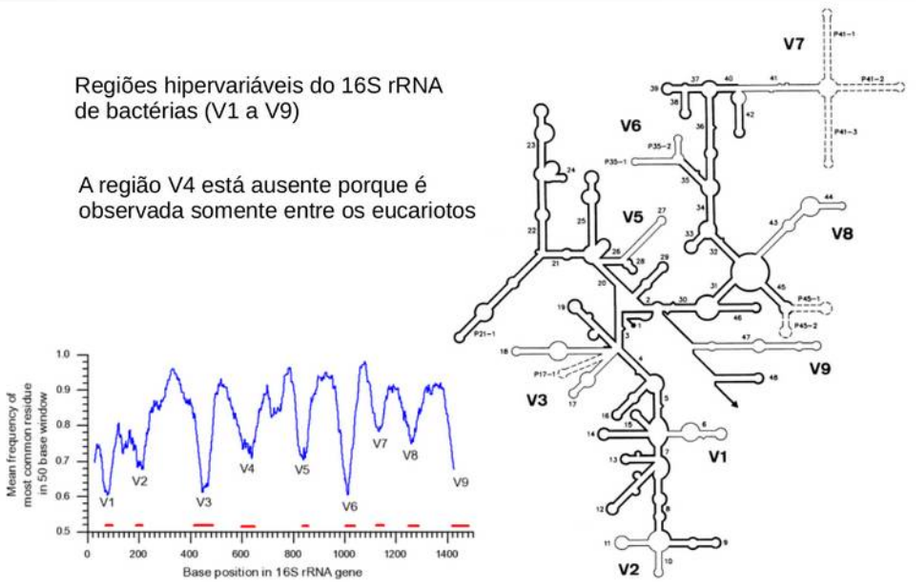
[a]The paper of Soergel et al. (2012) presents an exhaustive list of primer combinations and comparisons
[b]Information on V1-V9 (Ashelford et al. 2005)
[c]Referring to acc. no. J01695 of *E. coli rrnB*

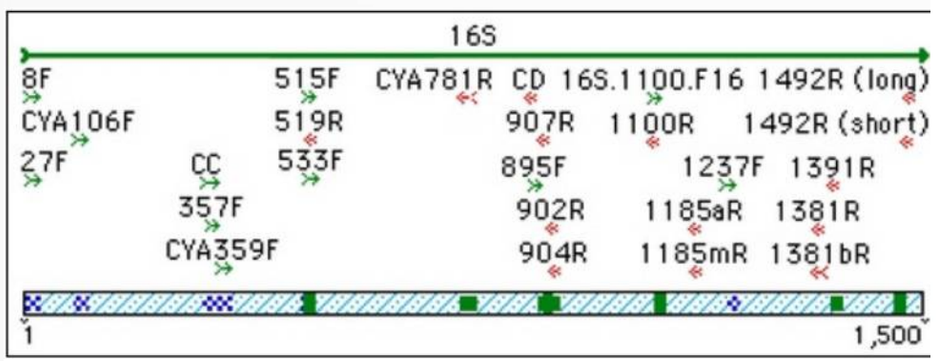
Christensen et al. 2018 doi: 10.1007/978-3-319-99280-8_8

Análise metagenômica com o gene 16S rRNA

Regiões hipervariáveis do 16S rRNA de bactérias (V1 a V9)

A região V4 está ausente porque é observada somente entre os eucariotos





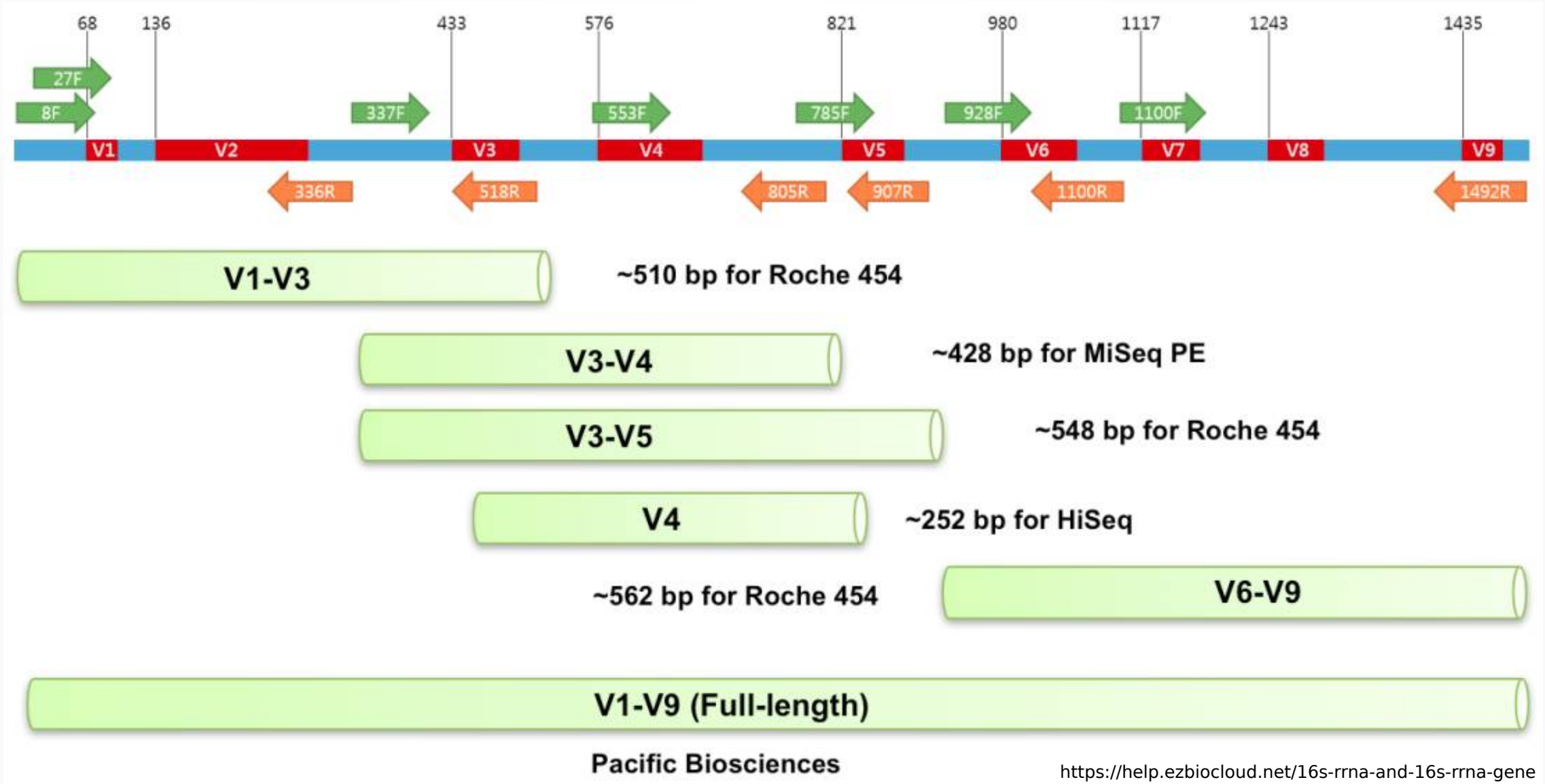
http://www.lutzonilab.net/primers/page604.shtml

Primer*	Sequence (5'-3')	Target Group	Reference
8F	AGAGTTTGATCCTGGCTCAG	Universal	Turner et al. 1999
27F	AGAGTTTGATCMTGGCTCAG	Universal	Lane et al. 1991
CYA106F	CGGACGGGTGAGTAACGCGTGA	Cyanobacteria	Nübel et al. 1997
CC [F]	CCAGACTCCTACGGGAGGCAGC	Universal	Rudi et al. 1997
357F	CTCCTACGGGAGGCAGCAG	Universal	Turner et al. 1999
CYA359F	GGGGAATYTTCCGCAATGGG	Cyanobacteria	Nübel et al. 1997
515F	GTGCCAGCMGCCGCGGTAA	Universal	Turner et al. 1999
533F	GTGCCAGCAGCCGCGGTAA	Universal	Weisburg et al. 1991
895F	CRCCTGGGGAGTRCRG	Bacteria exc. plastids & Cyanobacteria	Hodkinson & Lutzoni 2009
16S.1100.F16	CAACGAGCGCAACCCT	Universal	Turner et al. 1999
1237F	GGGCTACACACGYGCWAC	Universal	Turner et al. 1999
519R	GWATTACCGCGGCKGCTG	Universal	Turner et al. 1999
CYA781R	GACTACWGGGGTATCTAATCCCWTT	Cyanobacteria	Nübel et al. 1997
CD [R]	CTTGTGCGGGCCCCCGTCAATTC	Universal	Rudi et al. 1997
902R	GTCAATTCITTTGAGTTTYARYC	Bacteria exc. plastids & Cyanobacteria	Hodkinson & Lutzoni 2009
904R	CCCCGTCAATTCITTTGAGTTTYAR	Bacteria exc. plastids & Cyanobacteria	Hodkinson & Lutzoni 2009
907R	CCGTCAATTCMTTTRAGTTT	Universal	Lane et al. 1991
1100R	AGGGTTGCGCTCGTTG	Bacteria	Turner et al. 1999
1185mR	GAYTTGACGTCATCCM	Bacteria exc. plastids & Cyanobacteria	Hodkinson & Lutzoni 2009
1185aR	GAYTTGACGTCATCCA	Lichen-associated Rhizobiales	Hodkinson & Lutzoni 2009
1381R	CGGTGTGTACAAGRCCYGRGA	Bacteria exc. *Asterochloris* sp. plastids	Hodkinson & Lutzoni 2009
1381bR	CGGGCGGTGTGTACAAGRCCYGRGA	Bacteria exc. *Asterochloris* sp. plastids	Hodkinson & Lutzoni 2009
1391R	GACGGGCGGTGTGTRCA	Universal	Turner et al. 1999
1492R (l)	GGTTACCTTGTTACGACTT	Universal	Turner et al. 1999
1492R (s)	ACCTTGTTACGACTT	Universal	Lane et al. 1991

*Numbered primers are named for the approximate position on the *E. coli* 16S rRNA molecule. For each degenerate primer, an equimolar mix of all of the constituent primers implied by the degenerate sequence is recommended, since machine mixes are generally not guaranteed to approximate equimolarity. Primers developed by members of the Lutzoni Lab are in bold.

http://www.lutzonilab.net/primers/page604.shtml

Iniciadores e regiões amplificadas



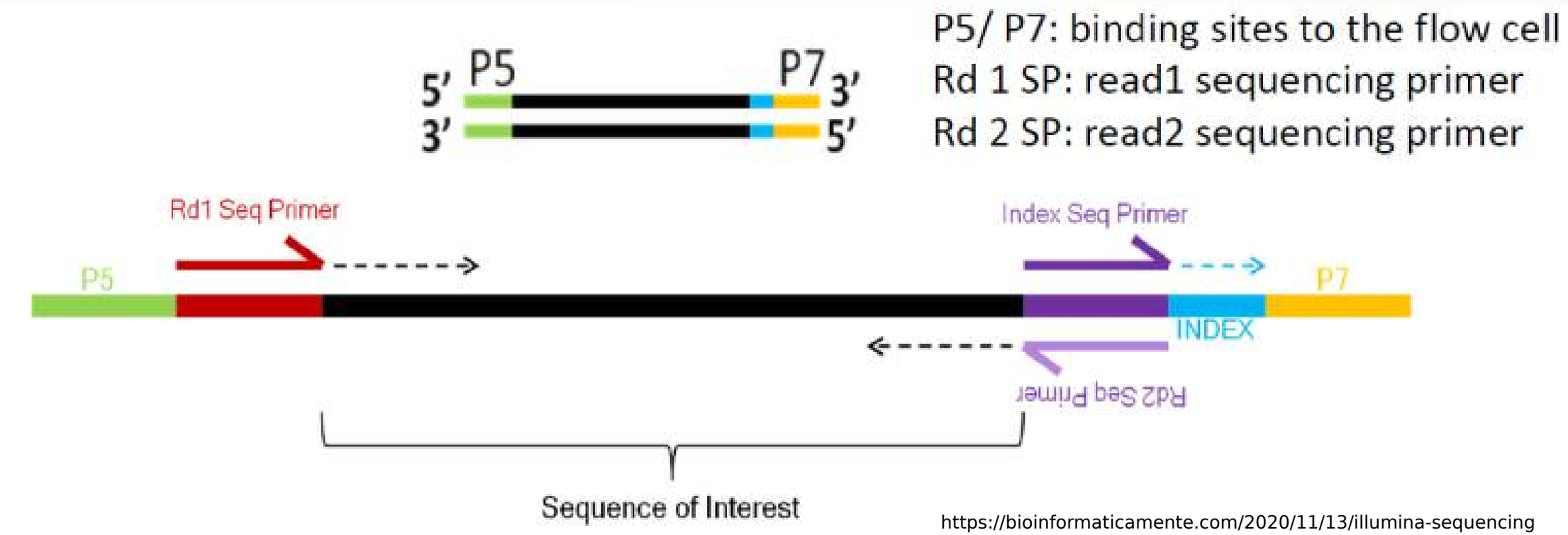
https://help.ezbiocloud.net/16s-rrna-and-16s-rrna-gene

Formatos de sequenciamentos recomendados

NGS systems	16S region	PCR primers	Estimated insert size (E. coli)	Sequencing format
Illumina MiSeq	V3V4	341F & 805R	428 bp	250 x 2
Illumina iSeq 100 [Learn more]	V4	515FB & 806RB	252 bp	300 x 1
Illumina HiSeq	V4	515FB & 806RB	252 bp	150 x 2

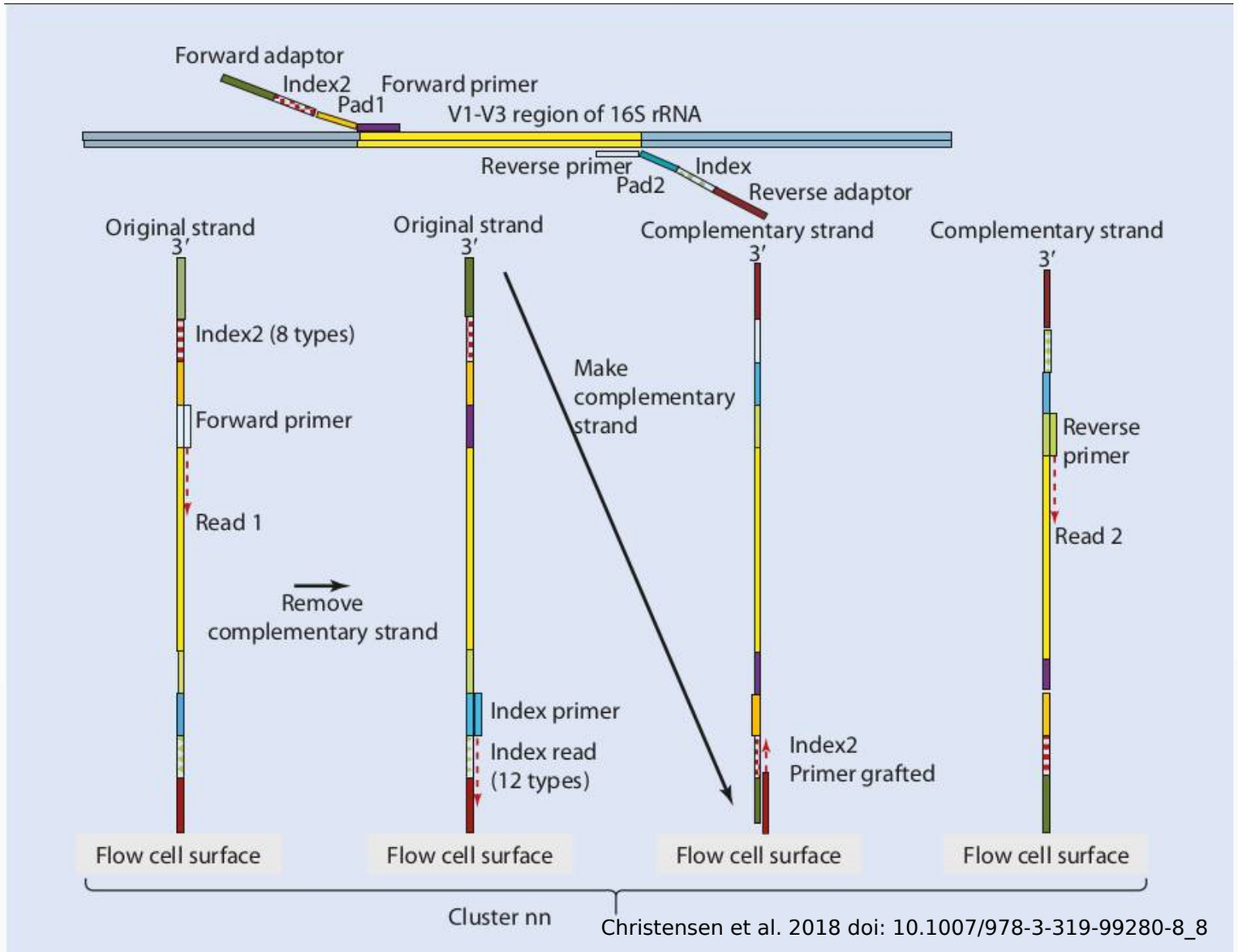
https://help.ezbiocloud.net/16s-rrna-and-16s-rrna-gene

Estrutura dos iniciadores da Illumina



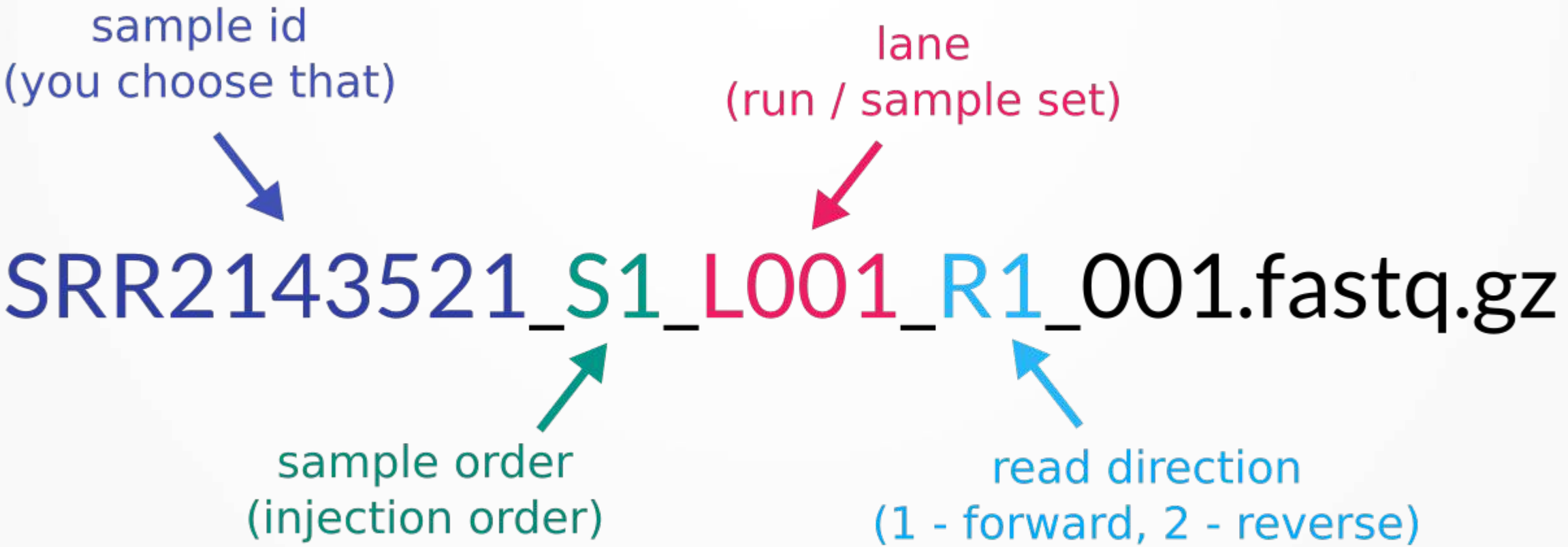
https://bioinformaticamente.com/2020/11/13/illumina-sequencing

Sequenciamento Illumina



Christensen et al. 2018 doi: 10.1007/978-3-319-99280-8_8

Arquivos de Illumina



https://gibbons-lab.github.io/isb_course_2020/16S

Formato Fastq

```
@SRR2143527.13917 13917 length=251
TACGTAGGTGGCGAGCGTTATCCGGAATTATTGGGCGTAAA...
+
BBBBAF?A@D2BEEEGGGFGGGHGGGCFGFHHCFHCEFGGH...
```

1) Identificador da sequência (Inicia com @)
2) Sequência nucleotídica
3) Identificador da sequência (opcional) (Inicia com +)
4) Qualidade em Phred score - caracteres ASCII

Formato Fastq

$$P = 10^{(-Q/10)}$$

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

ASCII_BASE=33 Illumina, Ion Torrent, PacBio and Sanger

Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII
0	1.00000	33 !	11	0.07943	44 ,	22	0.00631	55 7	33	0.00050	66 B
1	0.79433	34 "	12	0.06310	45 -	23	0.00501	56 8	34	0.00040	67 C
2	0.63096	35 #	13	0.05012	46 .	24	0.00398	57 9	35	0.00032	68 D
3	0.50119	36 $	14	0.03981	47 /	25	0.00316	58 :	36	0.00025	69 E
4	0.39811	37 %	15	0.03162	48 0	26	0.00251	59 ;	37	0.00020	70 F
5	0.31623	38 &	16	0.02512	49 1	27	0.00200	60 <	38	0.00016	71 G
6	0.25119	39 '	17	0.01995	50 2	28	0.00158	61 =	39	0.00013	72 H
7	0.19953	40 (	18	0.01585	51 3	29	0.00126	62 >	40	0.00010	73 I
8	0.15849	41)	19	0.01259	52 4	30	0.00100	63 ?	41	0.00008	74 J
9	0.12589	42 *	20	0.01000	53 5	31	0.00079	64 @	42	0.00006	75 K
10	0.10000	43 +	21	0.00794	54 6	32	0.00063	65 A			

ASCII_BASE=64 Old Illumina

Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII
0	1.00000	64 @	11	0.07943	75 K	22	0.00631	86 V	33	0.00050	97 a
1	0.79433	65 A	12	0.06310	76 L	23	0.00501	87 W	34	0.00040	98 b
2	0.63096	66 B	13	0.05012	77 M	24	0.00398	88 X	35	0.00032	99 c
3	0.50119	67 C	14	0.03981	78 N	25	0.00316	89 Y	36	0.00025	100 d
4	0.39811	68 D	15	0.03162	79 O	26	0.00251	90 Z	37	0.00020	101 e
5	0.31623	69 E	16	0.02512	80 P	27	0.00200	91 [	38	0.00016	102 f
6	0.25119	70 F	17	0.01995	81 Q	28	0.00158	92 \	39	0.00013	103 g
7	0.19953	71 G	18	0.01585	82 R	29	0.00126	93]	40	0.00010	104 h
8	0.15849	72 H	19	0.01259	83 S	30	0.00100	94 ^	41	0.00008	105 i
9	0.12589	73 I	20	0.01000	84 T	31	0.00079	95 _	42	0.00006	106 j
10	0.10000	74 J	21	0.00794	85 U	32	0.00063	96 `			

https://www.drive5.com/usearch/manual/quality_score.html

Bancos de dados de sequências amplicon do gene 16S rRNA

Bancos de dados

- Três principais bancos de dados são usados
  - **SILVA**
    - https://www.arb-silva.de
    - Quast et al., 2013 doi: 10.1093/nar/gks1219
  - **Greengenes**
    - https://greengenes2.ucsd.edu
    - McDonald et al., 2024 doi: 10.1038/s41587-023-01845-1
    - McDonald et al., 2012 doi: 10.1038/ismej.2011.139)
    - DeSantis et al., 2006 doi: 10.1128/AEM.03006-05
  - **RDP (???)**
    - http://rdp.cme.msu.edu
    - Cole et al., 2014 doi: 10.1093/nar/gkt1244
    - Cole et al., 2009 doi: 10.1093/nar/gkn879

Fluxo de análise de dados

Principais etapas na análise

- Desmultiplexação das sequências
- Controle de qualidade
- Agrupamento em OTU
- Identificação taxonômica
- Alfa diversidade
- Beta diversidade
- Análise de associação a metadados
- Análise de abundância diferencial

Fluxo de análise

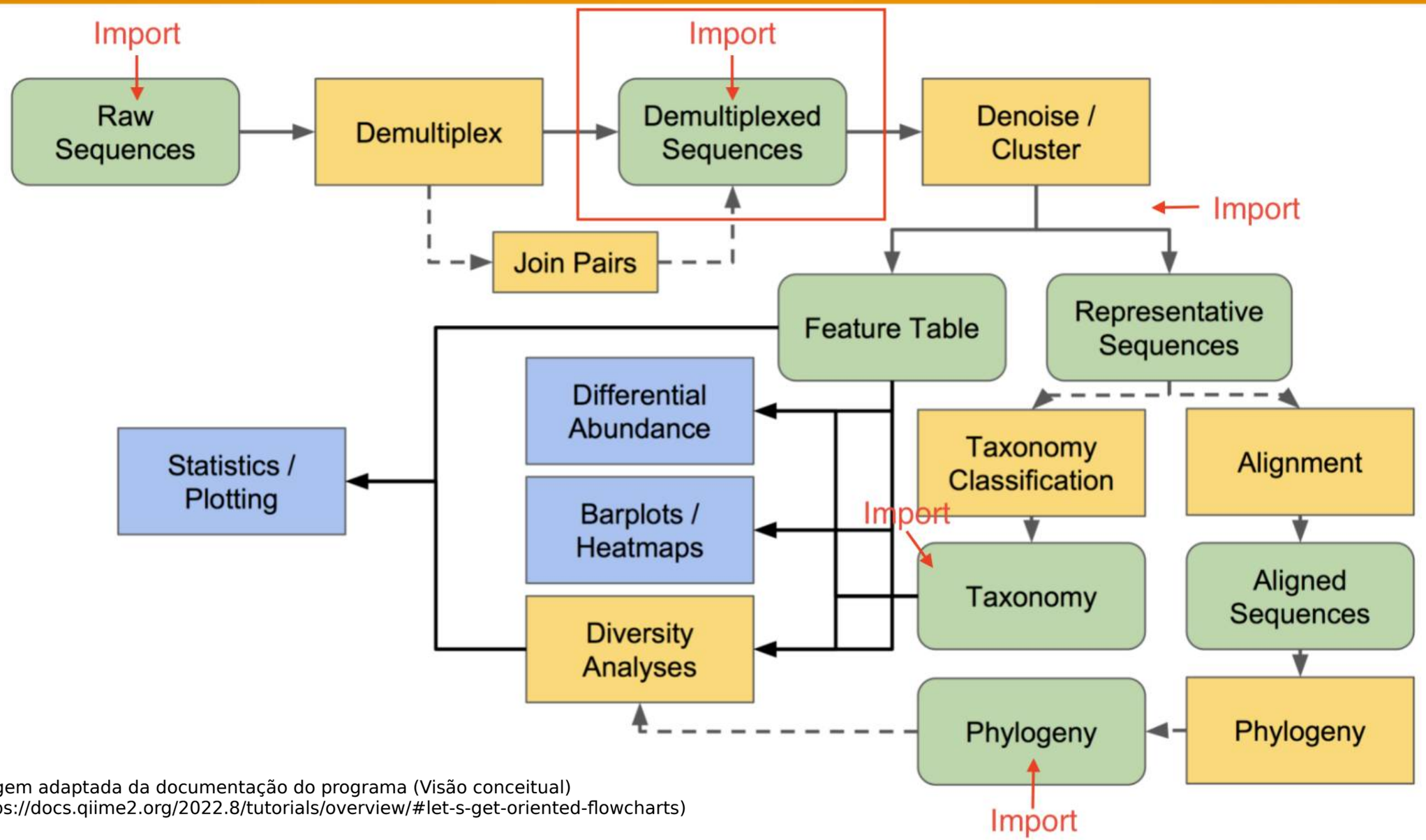
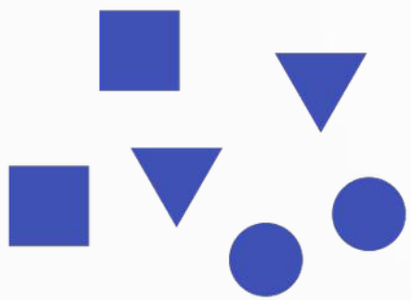
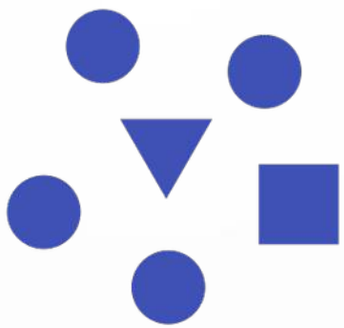


Imagem adaptada da documentação do programa (Visão conceitual)
(https://docs.qiime2.org/2022.8/tutorials/overview/#let-s-get-oriented-flowcharts)

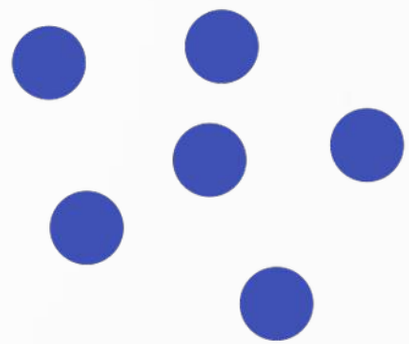
Diversidade alfa



very diverse



somewhat diverse



not diverse

- Riqueza (richness): quantidade de “espécies” (*taxa*) observadas?
  - Número de *taxa* observados, índice de Simpson
- Uniformidade (*evenness*): quão uniformemente as abundâncias são distribuídas entre os *taxa*
  - Índice evenness
- Misto: métricas que combinam ambos, riqueza e uniformidade
  - Índice de Shannon

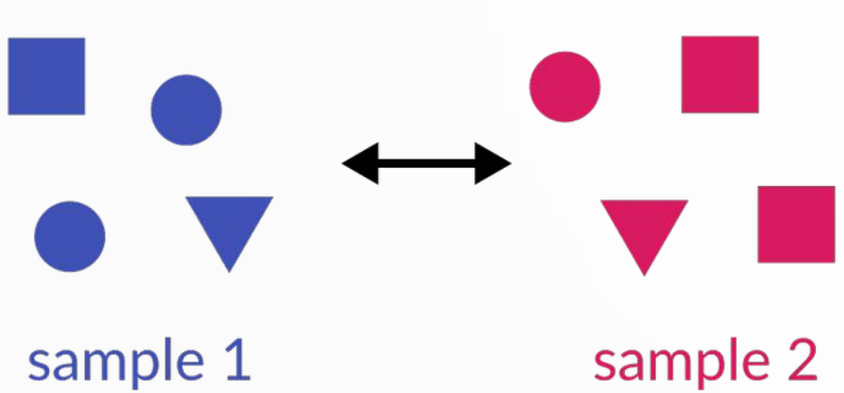
https://gibbons-lab.github.io/isb_course_2020/16S

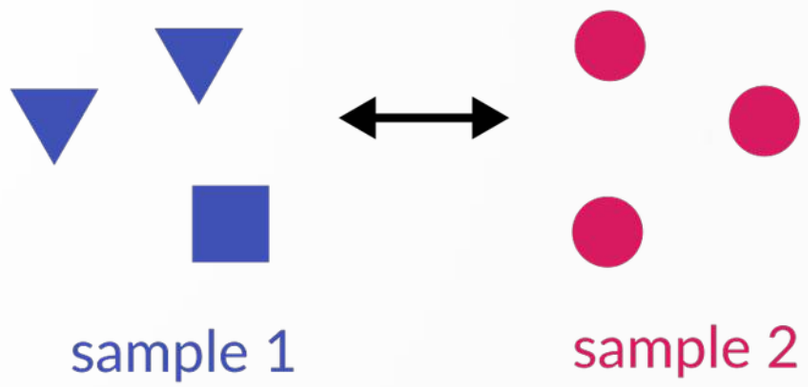
Testes estatísticos para diversidade alfa

- A diversidade alfa pode fornecer um valor único para cada amostra
- Pode ser tratado como qualquer outra medida de amostra e pode ser usado em testes univariados clássicos (teste-t, Mann-Whitney U)

https://gibbons-lab.github.io/isb_course_2020/16S

Diversidade beta

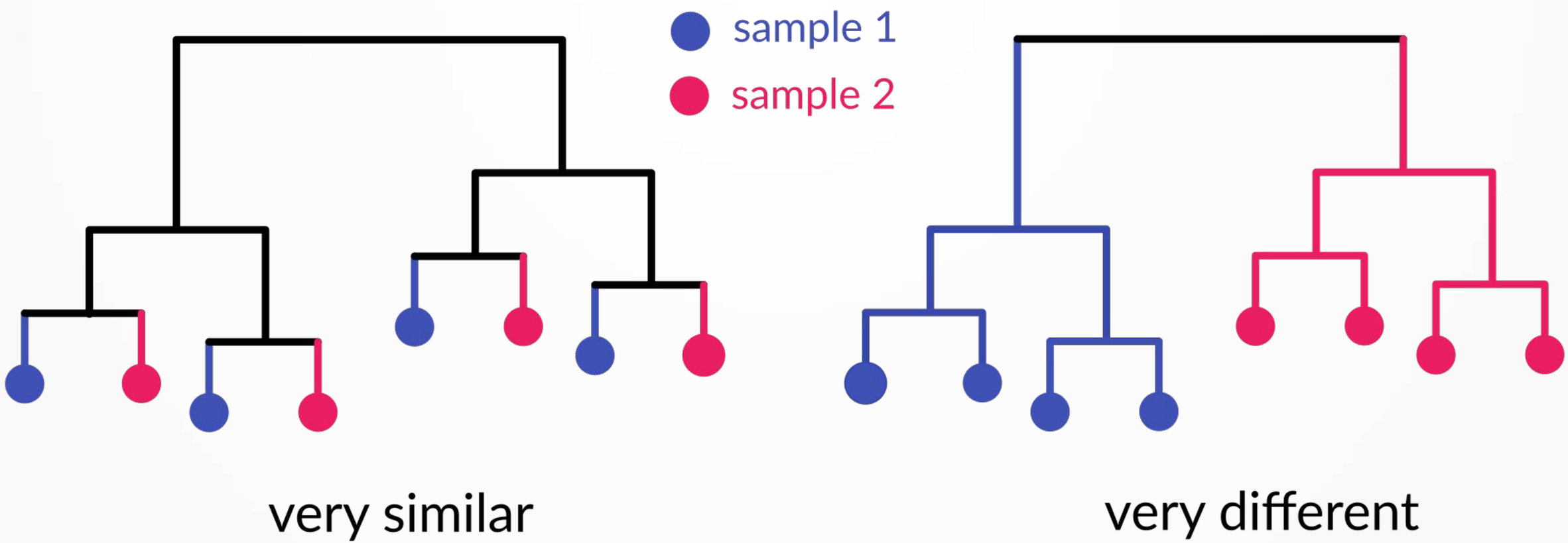




- Quão diferentes são duas ou mais amostras?
- Não ponderada (*unweighted*) = presença/ausência: quantos *taxa* são compartilhados entre amostras?
  - Índice de Jaccard, unweighted UniFrac
- Ponderada (*weighted*) = abundância como peso: os *taxa* compartilhados apresentam abundância semelhante?
  - Distâncias de Bray-Curtis, weighted UniFrac

https://gibbons-lab.github.io/isb_course_2020/16S

UniFrac



As amostras compartilham taxa geneticamente similares?
O índice Weighted UniFrac escalona os ramos por abundância

https://gibbons-lab.github.io/isb_course_2020/16S

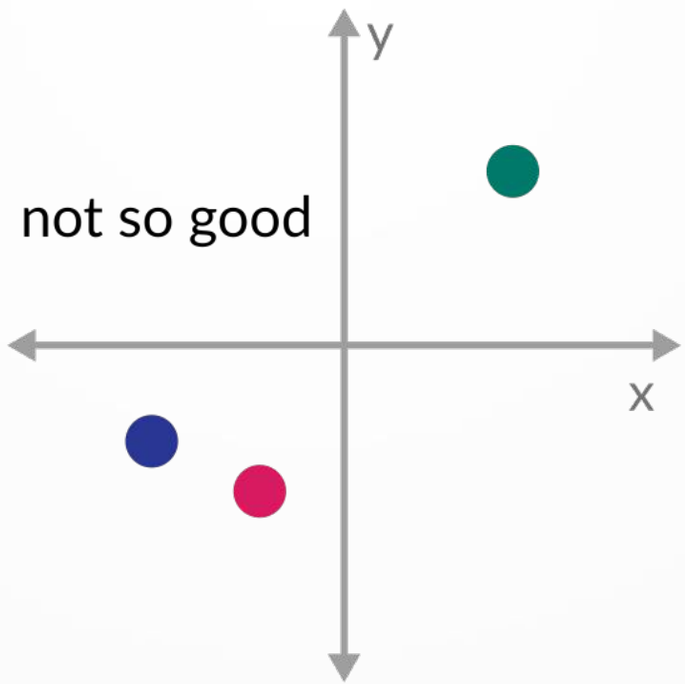
Análise de diversidade beta

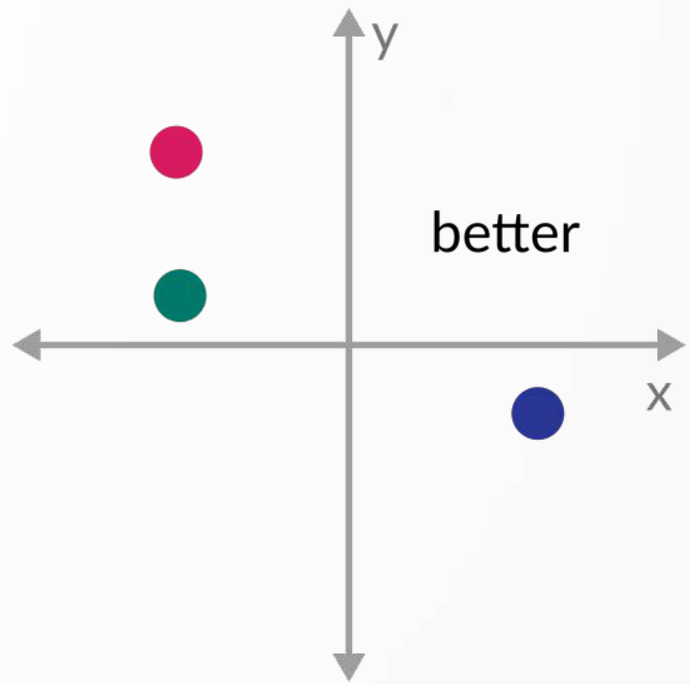
- Podemos comparar a dissimilaridade (ou distância) das amostras entre si
- Para isso, usamos técnicas de redução de dimensionalidade como:
  - PCoA = análise de coordenadas principais
    - Análise multidimensional para visualizar quais as fontes principais de variabilidade dos dados
    - Duas ou três dimensões (maior contribuição para a variabilidade dos dados)
    - Amostras são vistas como pontos num espaço e podemos calcular a distância entre elas.

Suzana Eiko Sato Guima. Análise de microbioma a partir do
gene rRNA 16S. Doutoranda pelo Programa Interunidades de Pós-Graduação em Bioinformática da USP
Orientador: Prof. Dr. João Carlos Setubal. Monitor: Robson Pontes IBI5035 – Biologia Molecular Computacional

Análise de coordenada principal

β div.	sample 1	sample 2	sample 3
sample 1	0		
sample 2	0.8	0	
sample 3	0.2	0.1	0





https://gibbons-lab.github.io/isb_course_2020/16S

Análise de Coordenadas Principais (PCoA)

- Teste estatístico para beta diversidade
  - Mais complicado
  - Geralmente não é normal e é heterogêneo
  - PERMANOVA pode lidar com isso