

# Metagenômica: rRNA 16S

M.Sc: Pedro Ivo Neves de Almeida



[pna.ivo@gmail.com](mailto:pna.ivo@gmail.com)



[pedroalmeida@aluno.fiocruz.br](mailto:pedroalmeida@aluno.fiocruz.br)



# Microbioma

# Microbioma

Bacteria

+

Archaea

+

Fungo

+

Virus

# Microbioma

Nem todos os micróbios são ruins.

Ainda tentando entender as interações micróbios e micróbios

Abordagens tradicionais dependem de micróbios isolados

Somos incapazes de cultivar muitas espécies microbianas

→ Necessidade de sequência direta

# Microbioma

Qual a função e estrutura das comunidades microbianas

- Função = genes + vias metabólicas
- Estrutura = riqueza e distribuição de espécies

# Projeto de Microbioma Humano

Lançado 2008, Finalizado (fase 1) 2012

- A segunda fase está em andamento

Métodos independentes da cultura

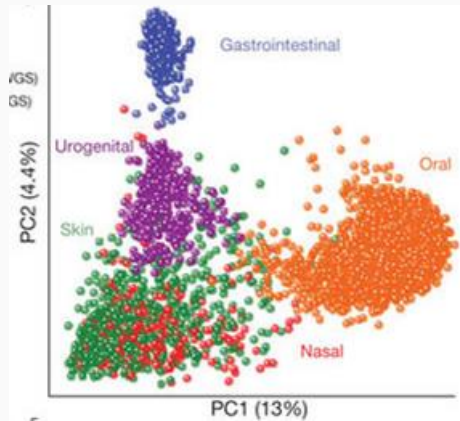
- Sequenciamento rRNA 16S
- Sequenciamento de genoma inteiro
  - Criou genomas de referência!

Não é o primeiro estudo a levantamento de microbiomas

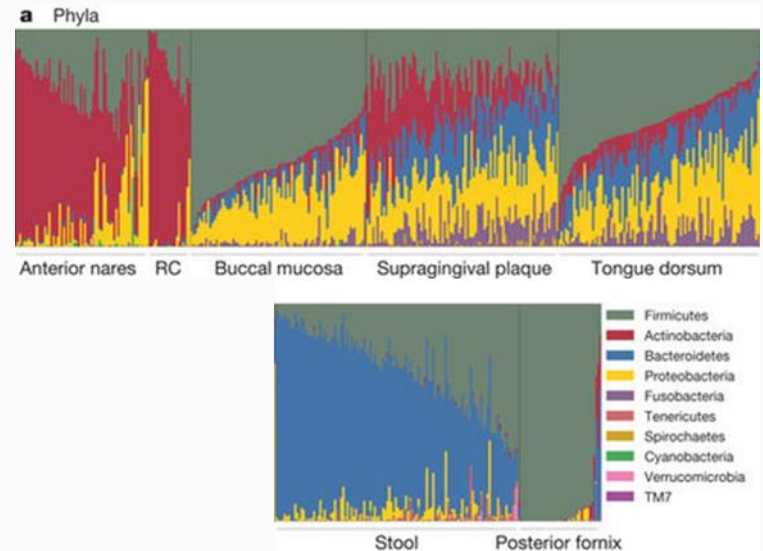
- Apenas o primeiro a examinar o microbioma humano em larga escala

# Composição do Microbioma Humano

Diferentes micróbios colonizam diferentes áreas do corpo

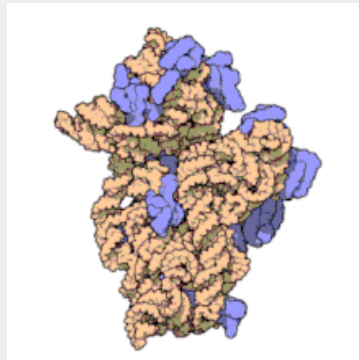


Muita variação entre indivíduos



# Sequenciamento rRNA 16S

## Biologia



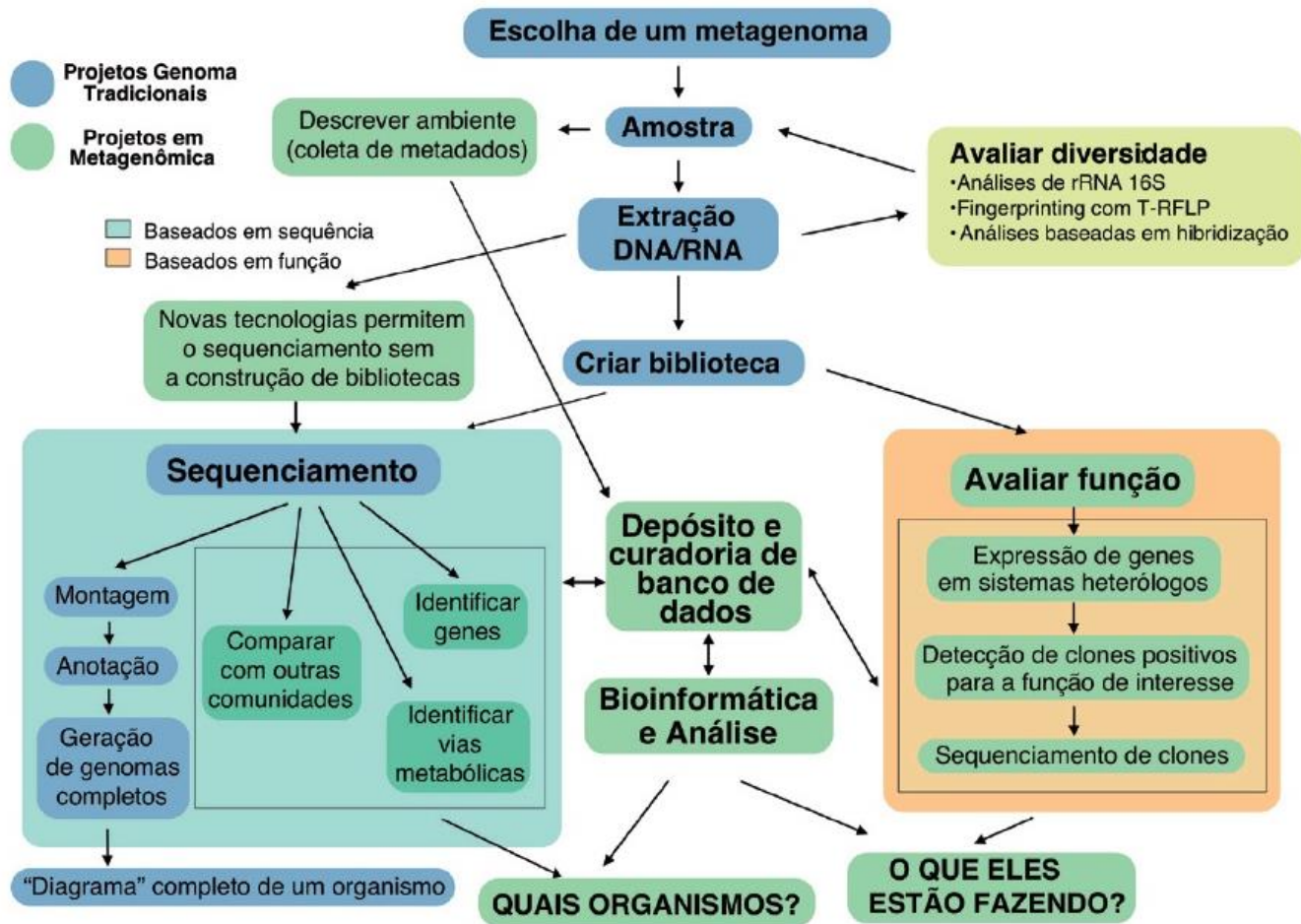


# Definição

**Genômica:** Organismo único

**Metagenômica:** Grupo de (micro)organismos

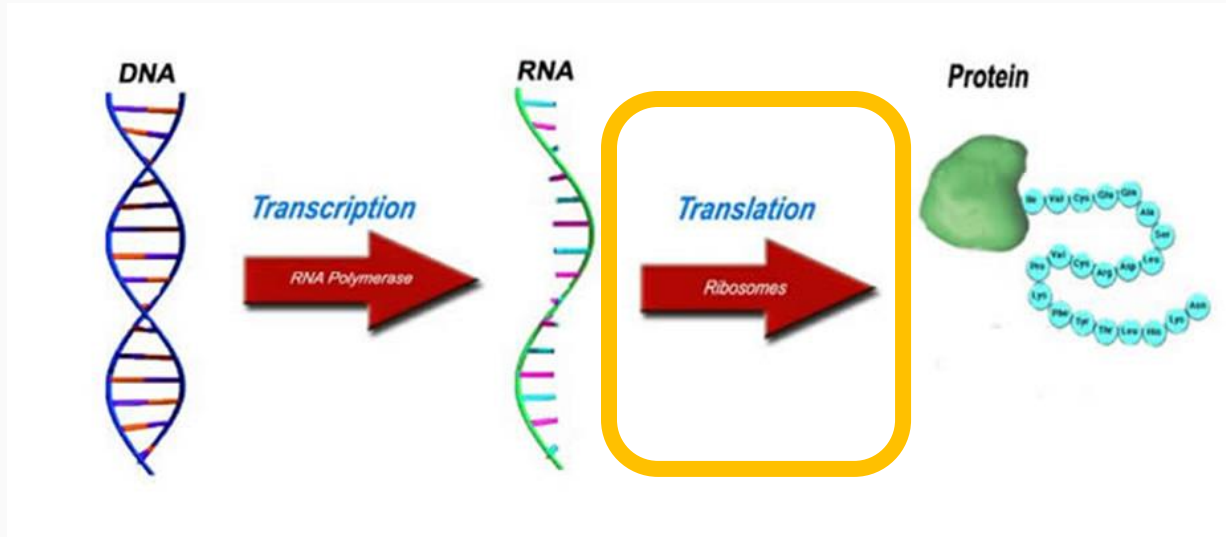
- Metagenômica: abordagem utilizada para a análise de comunidades de microrganismos de um determinado ambiente por técnicas independentes de cultivo;;
- Metagenoma: conjunto de genomas do microbioma de um determinado ambiente.



# Por que rRNA?

- Relativamente curto (~1,5 kb)
  - Curto = barato
- Altamente conservado
  - Forma um ribossomo, que é altamente traduzido, de modo que a maioria das mutações vai impactar o fitness
- Geralmente diferente entre espécies
  - Também informações suficientes em regiões menores (variáveis) == baratas

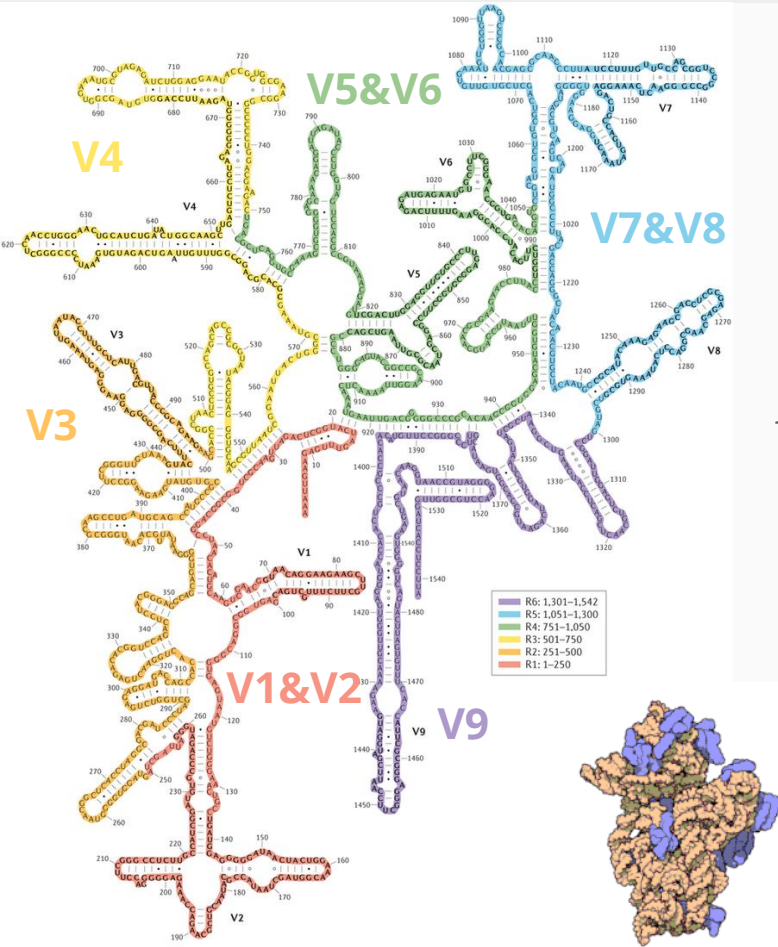
# Dogma Central



A tradução requer ribossomos...

... RNA ribossômico (rRNA) codifica ribossomos

# Gene16S rRNA



- Gene RNA ribossômico  
Codifica a pequena subunidade 30S
- 9 regiões variáveis  
Usado como assinaturas para determinar filogenia  
Escolha qual região (s) sequenciar
- Várias regiões universais  
Usado para criar primers
- S = “Svedberg”  
Svedberg = unidade de tamanho molecular

# Genes rRNA como marcador

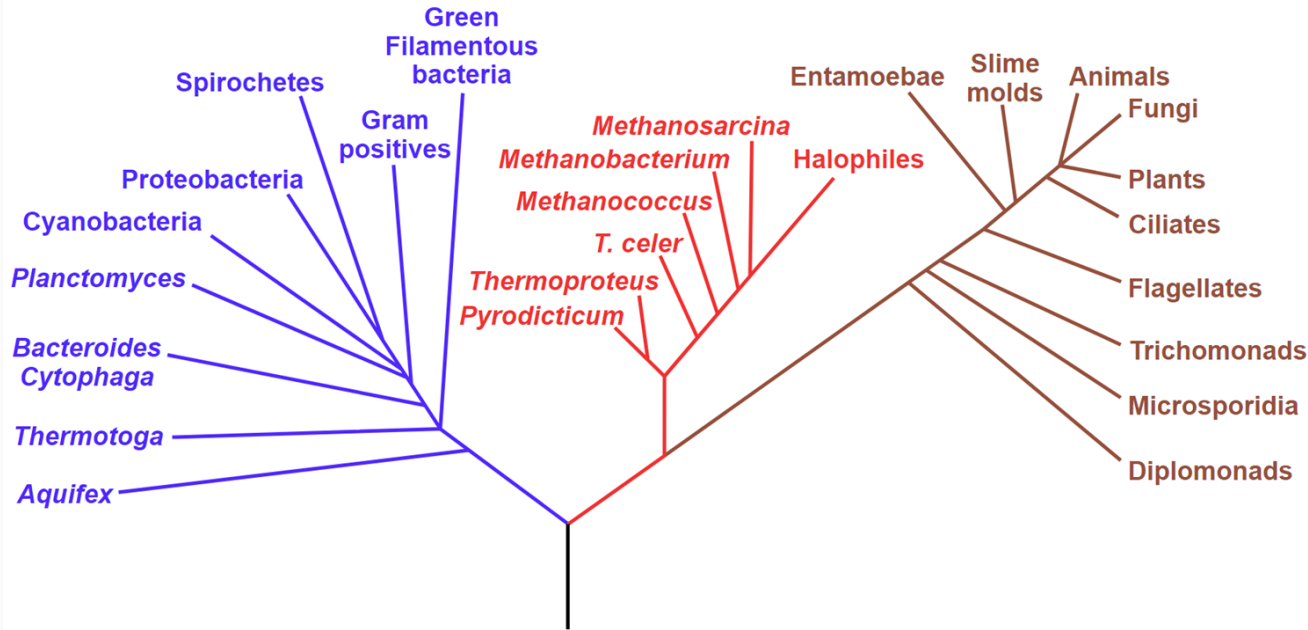
16S rRNA

Bacteria

Archaea

Eukaryota

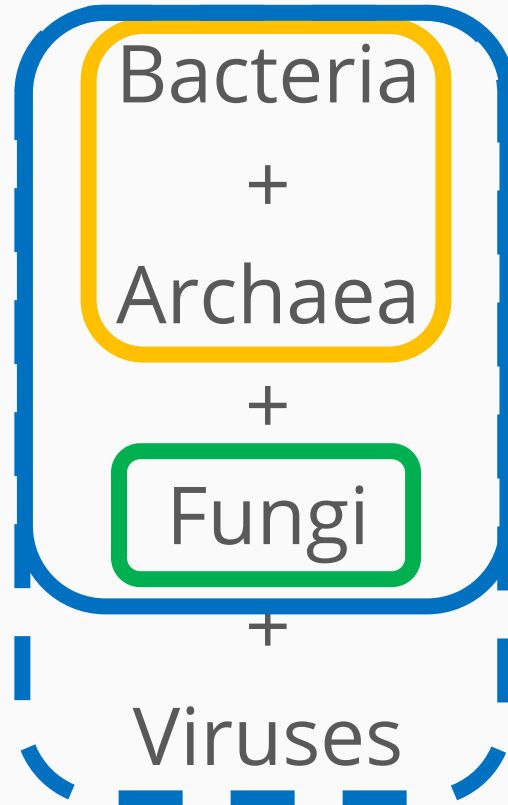
18S rRNA



rRNA usado para classificar a vida em três domínios

# Microbioma

Metagenômica/  
Genoma inteiro  
Sequenciamento  
Shotgun



Sequenciamento rRNA  
16S

Sequenciamento rRNA  
18S

# Resolução Taxonômica

**Domínio**

**Reino**

**Filo**

**Classe**

**Ordem**

**Família**

**Gênero**

**Espécie**

16S/18S rRNA não produz informações sobre nível de espécie

→ nível-**gênero**, na melhor das hipóteses, mas geralmente maior

- Espécies intimamente relacionadas têm uma alta semelhança sequencial em todo o gene 16S
- Normalmente não sequencia todo o gene, apenas 1(+) regiões variáveis



# Análise Populacional

**Diversidade Alfa:** Dentro de uma amostra

**Diversidade Beta:** Entre amostras

**Uniformidade:** Distribuição de taxons

**Riqueza:** Número de taxons

**Curvas de rarefação:** Diversidade alfa versus número de observações

- Usado para testar se um ambiente foi suficientemente sequenciado para observar toda a taxa

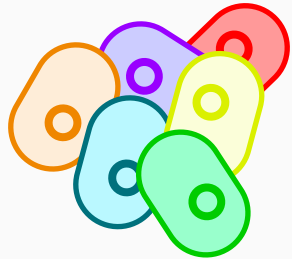
**PCA/PCoA:** Análise principal do componente/coordenada

**NMDS:** Dimensionamento multidimensional não métrico

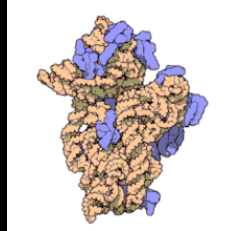
Técnicas de redução de dimensionalidade para visualização

# Sequenciamento rRNA 16S

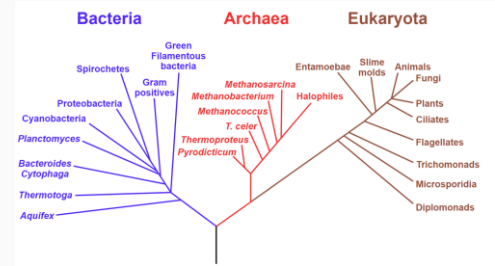
**Amostra**



**rRNA 16S  
Sequenciamento**



## Classificação Taxonômica



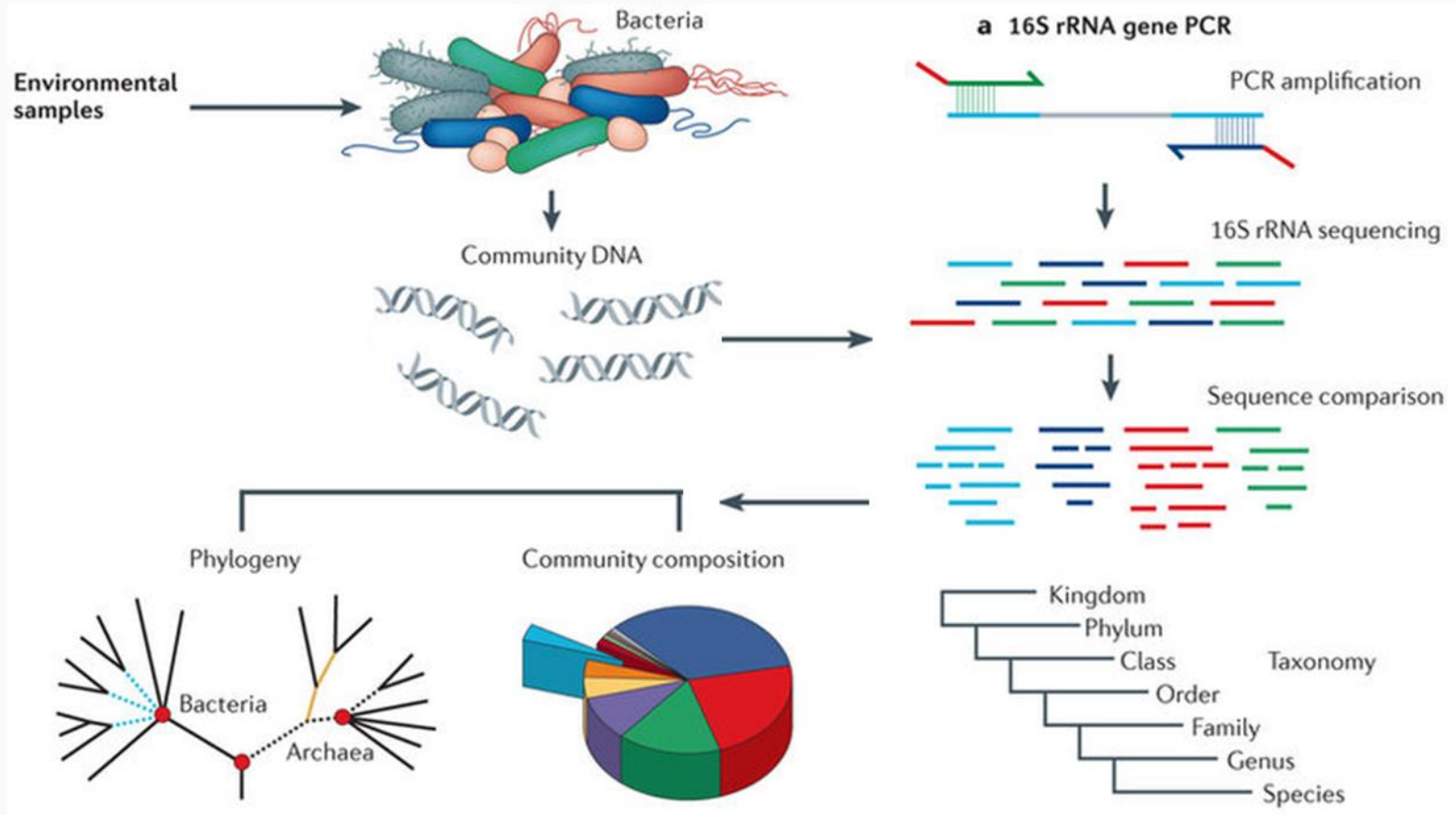
## Análise Populacional

- Diversidade alfa
- Diversidade beta
- Mais/sub-representação

# **Sequenciamento rRNA 16S**

## **Bioinformática**

# Fluxo de trabalho básico



# Coisas para Ter em Mente

- Plataforma de sequenciamento

Taxa de erro, vieses, comprimento de leitura, ruído

- Escolha de região variável(s)

- Processo de amplificação

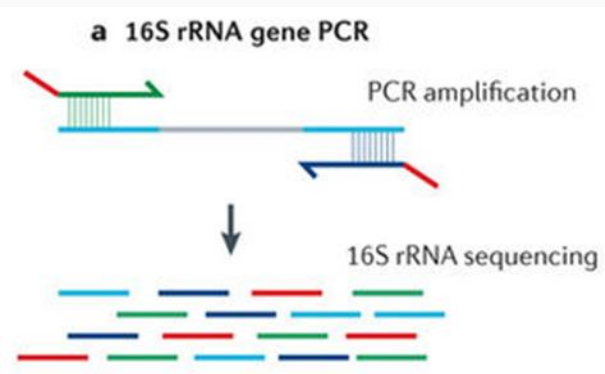
Taxa de erro, vieses, escolha do primer, concentração de modelo de DNA, número do ciclo PCR, introdução de quimeras

- Cobertura/Profundidade

# Pré-processamento

## Avaliação de Qualidade & Cutoffs

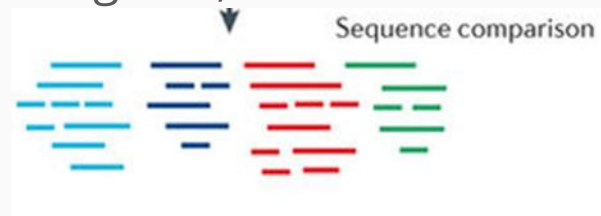
- Remova adaptadores, primers PCR e bases de baixa qualidade
- Demultiplexando usando códigos de barras, descarte leituras sem código de barras



# Binning

## OTU: Operational Taxonomy Unit

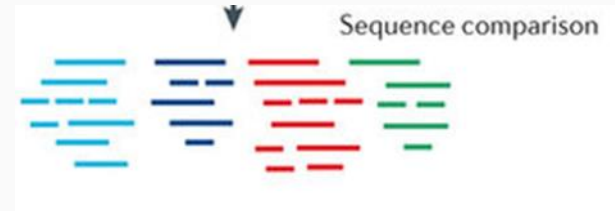
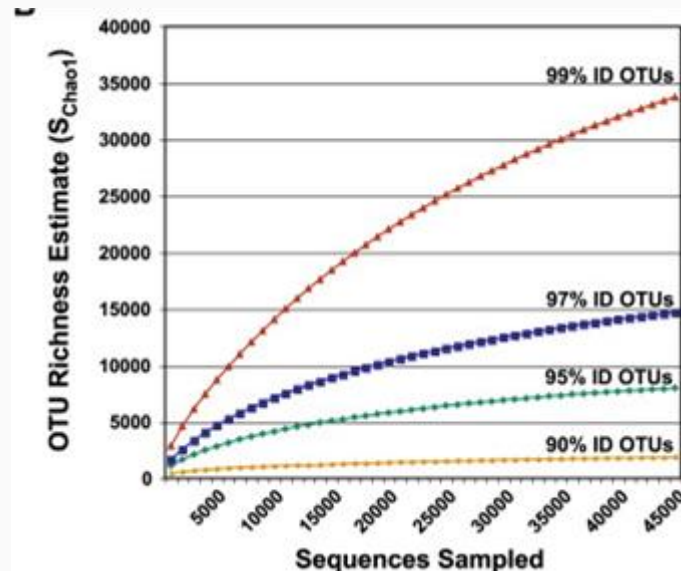
- Um grupo de sequências agrupadas baseadas puramente na similaridade e um limiar arbitrário
- Pode ou não ser equivalente a entidades taxonômicas (espécies, gêneros, etc.)
- Pode gerar cluster com base na similaridade a um banco de dados de referência ou de novo (comparado entre si)
- Também pode gerar cluster *de novo* e, em seguida, atribuir taxonomia



# Binning

Cutoffs: Qual porcentagem da sequência identificada deve-se usar?

→ Vai depender da taxa de erro, *etc.*





# Binning

## **ASV: Amplicon Sequences Variant**

- Busca encontrar sequências únicas e exatas de um determinado organismo;
- Sendo uma sequência exata, uma determinada sequência deve sempre gerar o mesmo ASV;
- O algoritmo por trás desta abordagem extrai as sequências únicas encontradas nas amostras e conta quantas vezes esta mesma sequência apareceu;

# Qual a melhor abordagem a ser utilizada: ASV ou OTU?

- Abordagem por OTU ainda é mais utilizada pela comunidade, diversos grupos já estão adotando a abordagem por ASV;
- A abordagem com OTU, ainda apresenta uma maior vantagem em relação ao poder computacional;
- As abordagens com ASV fornecem vantagens significativas, aumentando a sensibilidade nas análises e a precisão para identificação de organismos raros;
- ASV apresenta maior reprodutibilidade e facilidade na comparação entre os estudos, por não apresentar agrupamentos de sequências e sim sequências únicas;

# Software

[mothur](#)

[QIIME](#) (“chime”)

[bioBakery](#) (PICRUSt)

[CloVR](#)

# Bancos de dados taxonômicos

[Greengenes](#)

[MG-RAST](#)

[NCBI](#)

[RDP](#)

[SILVA](#)

\*Um banco de dados diferente pode lhe dar diferentes resultados taxonômicos

# Notas

16S rRNA A análise de sequenciamento é qualitativa

- Pesquisando quem está lá

PCR depende sobre a suposição de conhecimento a priori de primers universais

- Pode render uma estimativa alterada/incompleta da diversidade
- Também, problemas no prime, amplificação desigual, *etc...*

Estão convertendo dados binários (presença/ausência) ou normalizando (abundância relativa)

Certifique-se de que está usando as ferramentas estatísticas/análises apropriadas para dados binários e normalizados!

# Notas

## Classificação taxonômica dependerá

- A resolução que região variável do gene 16S rRNA é usada
- Os primers usados para PCR
- Qual banco de dados é usado
- Qual pacote de software é usado
- Qual cutoff é usado
- Cobertura/profundidade da sequência
- Plataforma de sequenciamento
- A composição das espécies na comunidade que está sendo analisada
- Quando você realiza a análise
- ...

# Considerações

Defina a questão o mais precisamente possível.

- Que controles você precisa?

Que plataforma de sequenciamento você usará?

- A illumina é a plataforma típica (agora)

Que região do gene 16S rRNA você irá amplificar?

- V4 geralmente rende nível de gênero

Quantas reads (leituras), você precisa por amostra?

- Cobertura/Profundidade

Quais são os problemas técnicos ocultos?

- Exemplo: Quimeras

Qual ferramenta de análise você usará? Como você exibirá seus dados? Como você vai comparar seus resultados com outros estudos publicados?