

# Filogenética molecular aplicada

## 3. Identificação biológica com marcadores moleculares

# Revisão

---

Boas práticas para inferir uma filogenia a partir de sequências de DNA ou proteínas:

- Máxima verossimilhança é o método de inferência mais indicado
- Certifique-se que o modelo de evolução de sequências utilizado é o mais adequado para o seu conjunto de dados (use métodos de seleção de modelos)
- Para poder avaliar o nível de confiança de cada ramo da árvore inferida, inclua uma análise de bootstrap
- Programas recomendados:
  - Visualização de alinhamentos: [Aliview](#)
  - Alinhadores: [MAFFT](#) (inclui versão Web)
  - Inferência por máxima verossimilhança: [PhyML](#) ou [IQ-TREE](#) (ambos tem versão Web)
  - Visualização e formatação de árvores: [FigTree](#) ou <https://itol.embl.de/>
  - Estudo e familiarização: [MEGA](#)

# Plano de aula

---

- Histórico e princípios da classificação biológica
- Identificação biológica tradicional (morfologia) e por DNA
- Encontrando marcadores moleculares adequados
- Bancos de dados públicos

Prática: Montando e curando bancos de sequências locais, identificação biológica de amostras usando filogenias

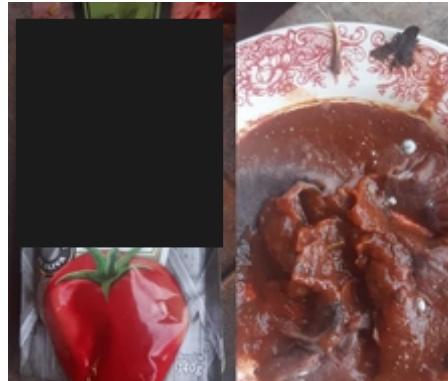
# Princípios da classificação biológica

---

- Sistemas de classificação biológica
- Taxonomia e Sistemática
- O que é uma espécie?

# Que bicho é esse?

---



# Taxonomia e classificação biológica

Atribuições da taxonomia: catalogar novas espécies, descrições morfológicas, curadoria de coleções, curadoria da classificação dos grupos (revisão taxonômica)



<https://medium.com/sci-five-university-of-basel/>



<https://herbarium.rutgers.edu/>

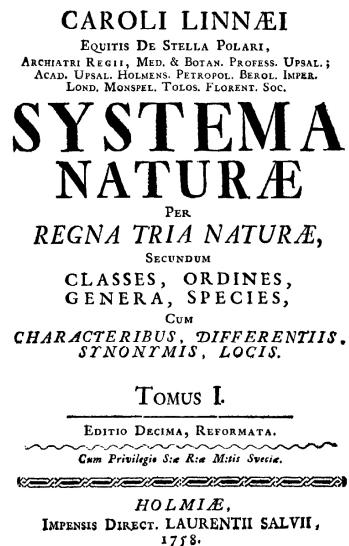
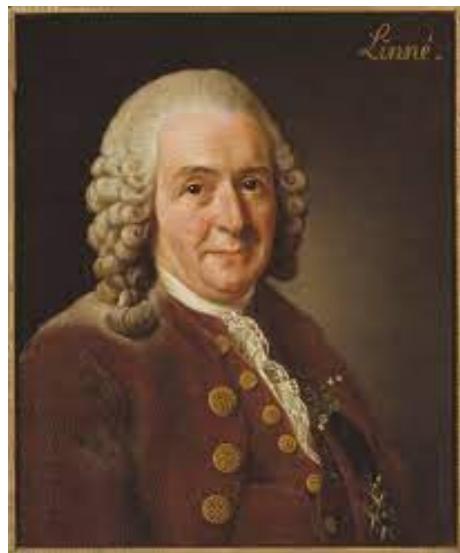


<https://abelha.org.br/>

# Taxonomia e classificação biológica

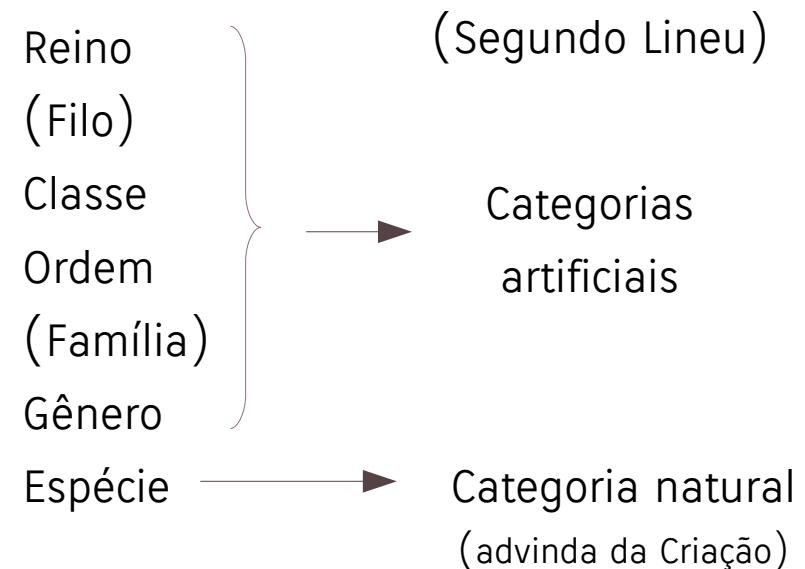
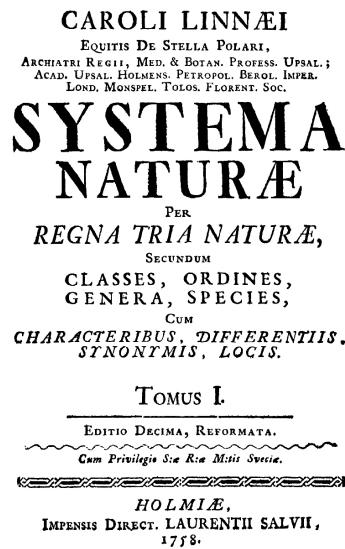
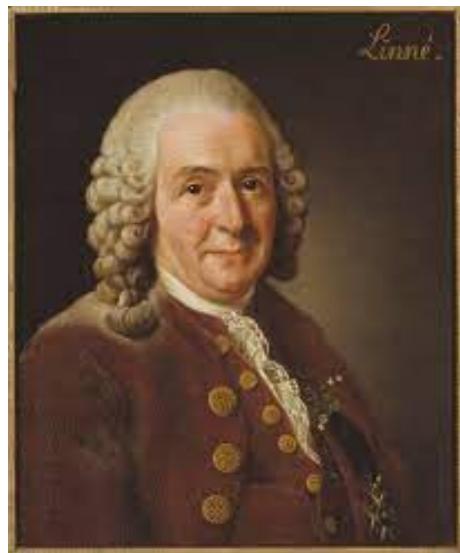
---

- Princípio básico da classificação biológica pré-Darwin: agrupamentos baseados em 'afinidades' entre organismos
- Lineu e o *Systema Naturae* (1735): base do sistema de classificação usado hoje



# Taxonomia e classificação biológica

- Princípio básico da classificação biológica pré-Darwin: agrupamentos baseados em 'afinidades' entre organismos
- Lineu e o *Systema Naturae* (1735): base do sistema de classificação usado hoje



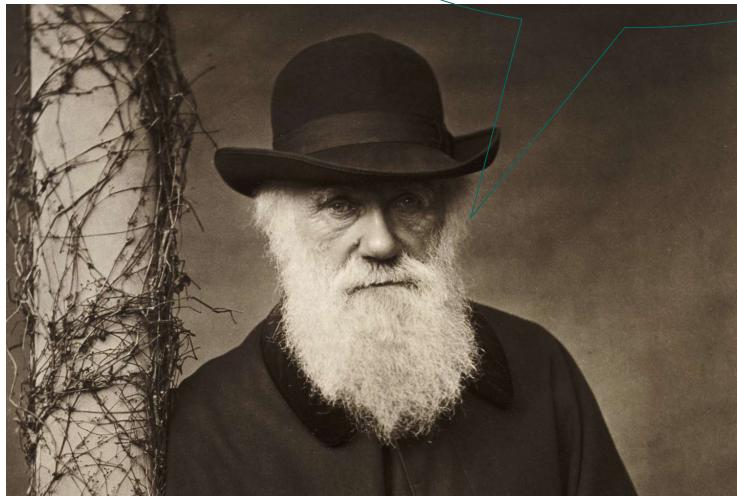
# Taxonomia e classificação biológica

---

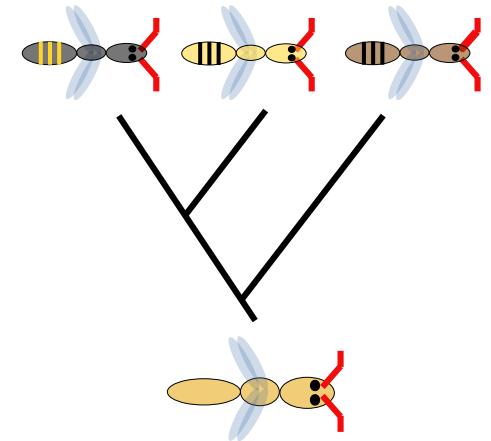
- Por que grupos de organismos compartilham características?

# Taxonomia e classificação biológica

- Por que grupos de organismos compartilham características?
- Darwin forneceu a teoria básica da classificação biológica



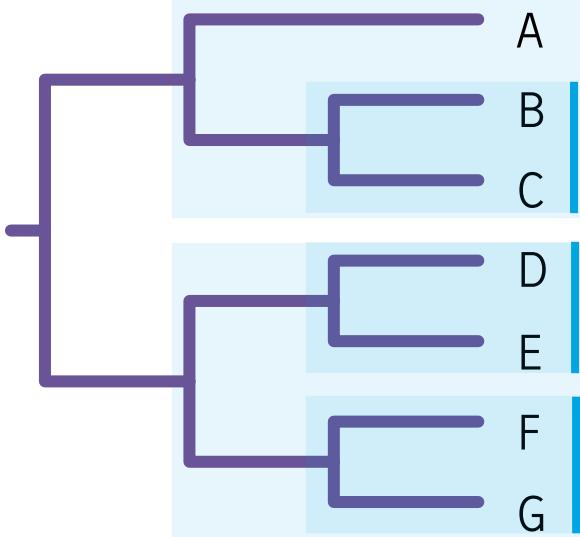
Ancestralidade comum!



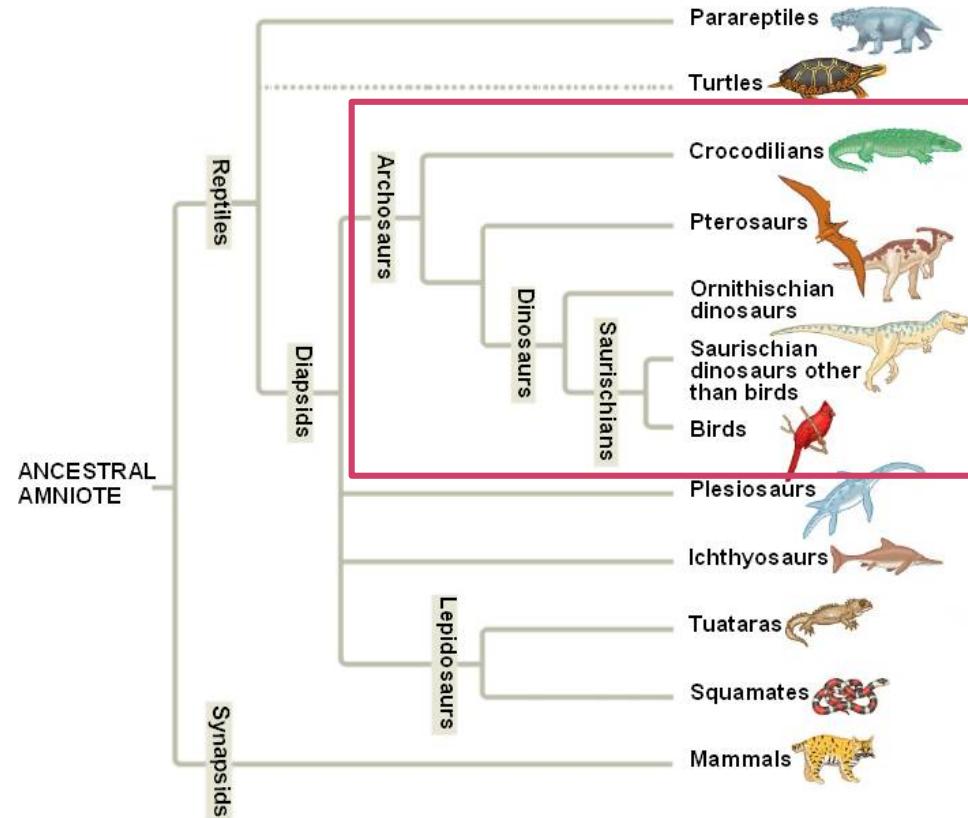
“Desde a origem da vida, todos os seres vivos parecem-se uns aos outros em graus decrescentes, de modo que podem ser classificados em grupos sob grupos [sistema hierárquico]. Essa classificação, evidentemente, não é arbitrária tal como o agrupamento de estrelas em constelações (...) Eu acredito que algo além da mera semelhança existe [no nosso sistema de classificação], e que a propensão à descendência [hereditariedade] – a única causa conhecida de similaridade entre organismos vivos – é o vínculo, escondido como é por vários graus de modificação, que é parcialmente revelado a nós pelas nossas classificações.” Charles Darwin, Origem das espécies, 1859.

# Sistemática

- Classificação biológica moderna → baseado em hipóteses filogenéticas (= árvores)
- A Sistemática busca reconstruir a história evolutiva ('natural') das linhagens

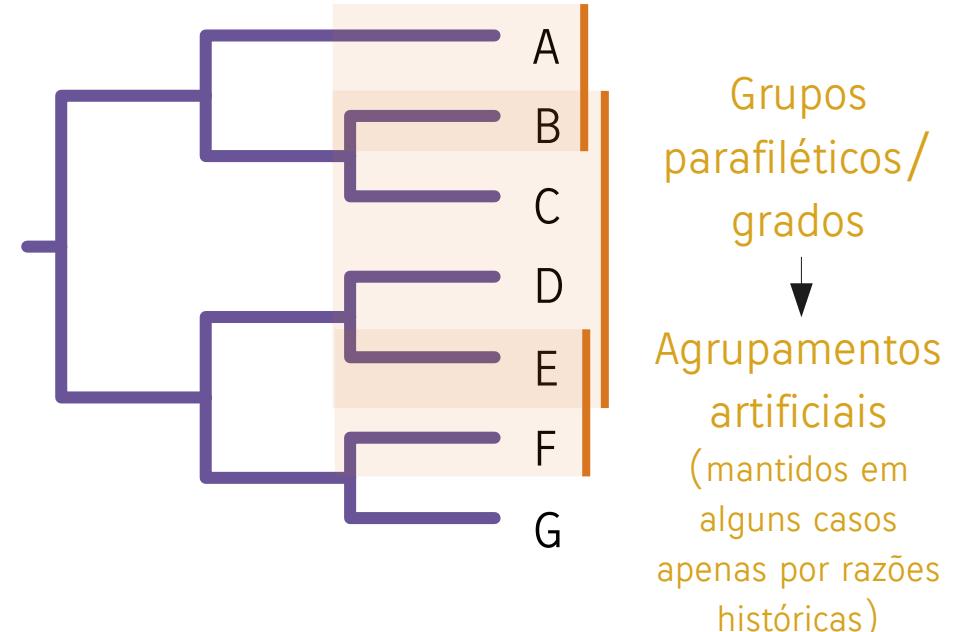
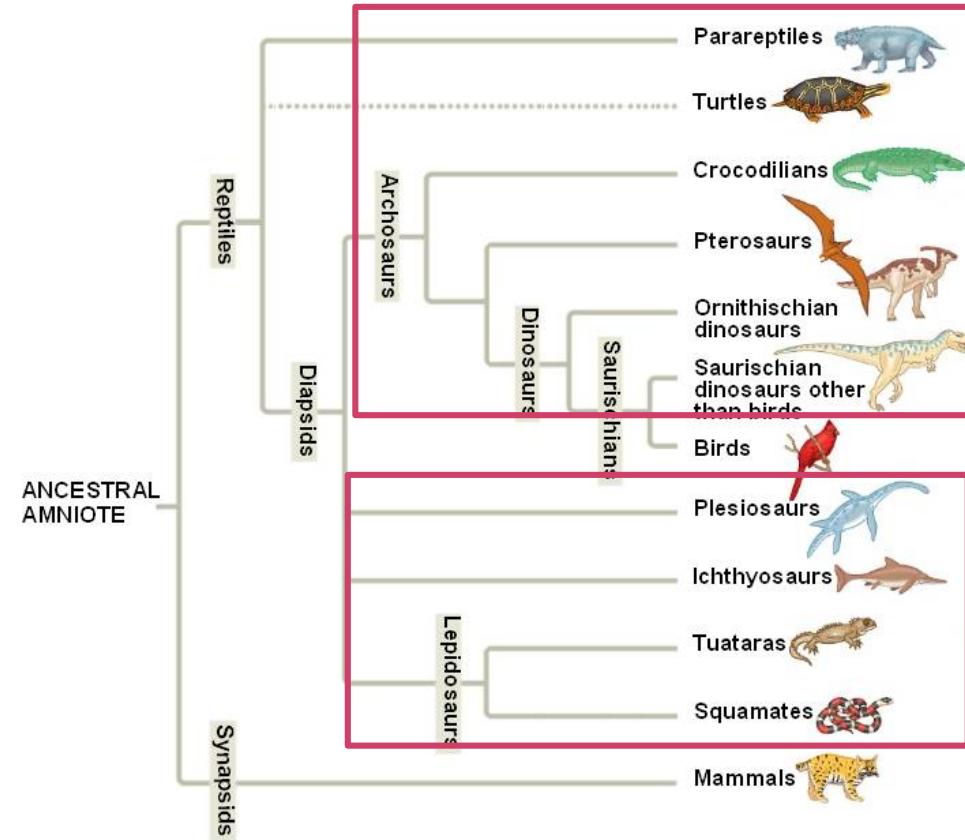


Grupos  
monofiléticos /  
clados  
↓  
Agrupamento  
reflete  
ancestralidade  
comum



# Sistemática

- Classificação biológica moderna → baseado em hipóteses filogenéticas (= árvores)
- A Sistemática busca reconstruir a história evolutiva ('natural') das linhagens

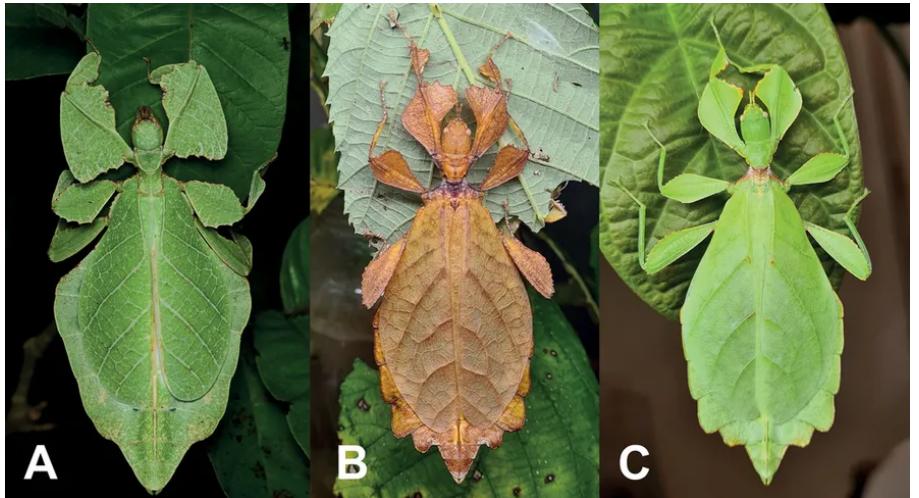


# O que é uma espécie?

---



<https://rcannon992.com/2019/12/01/stingless-bees-fascinating-little-builders/>



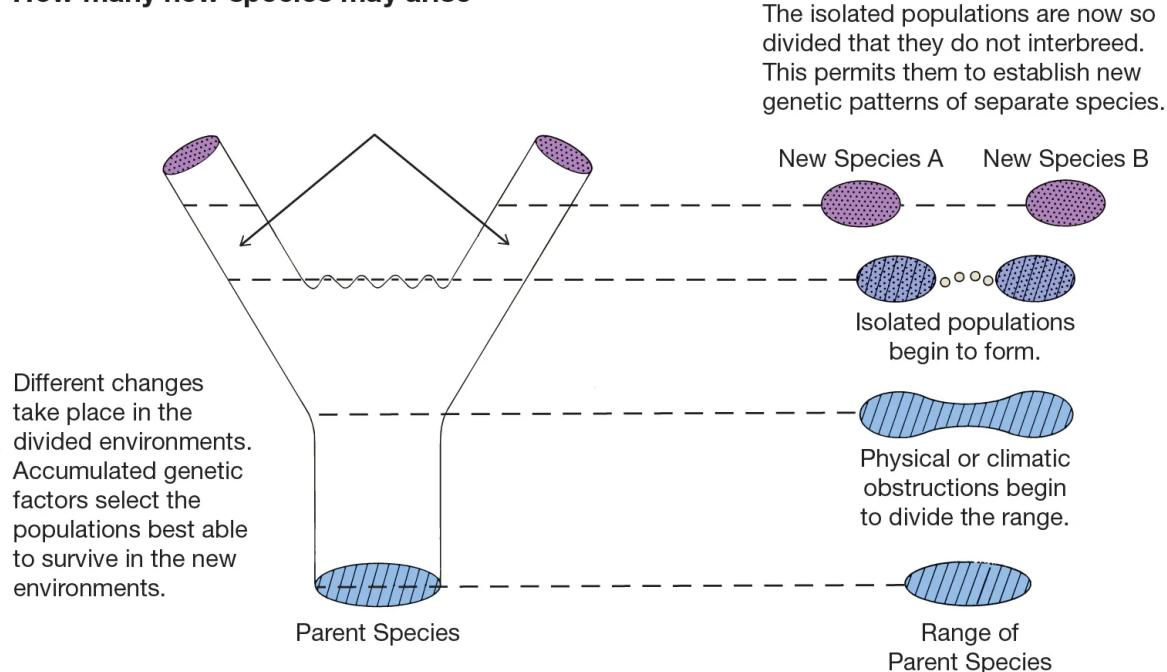
<https://revistagalileu.globo.com/>



# O que é uma espécie?

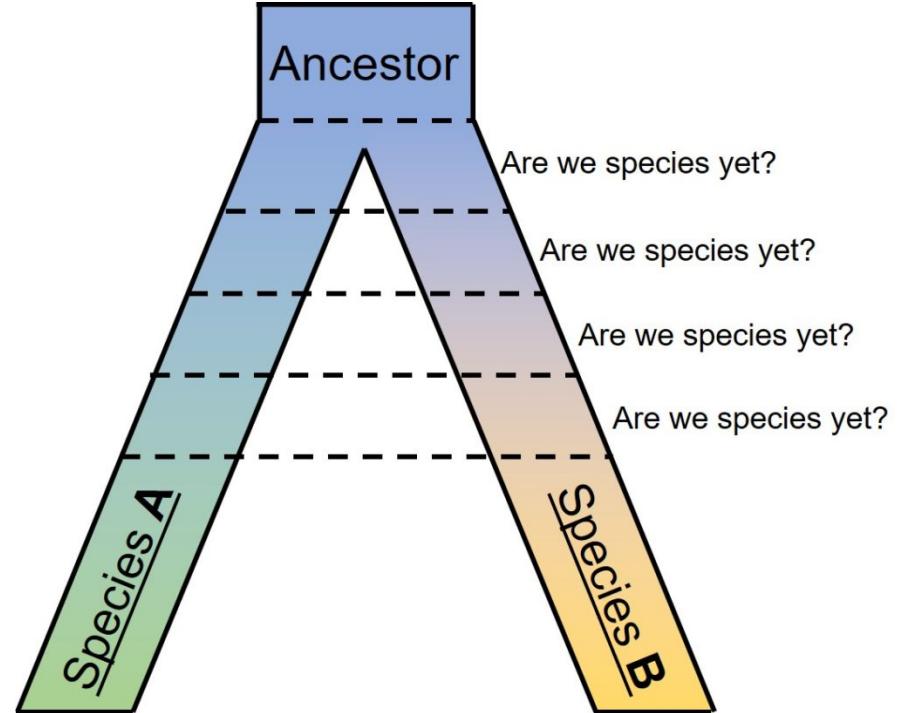
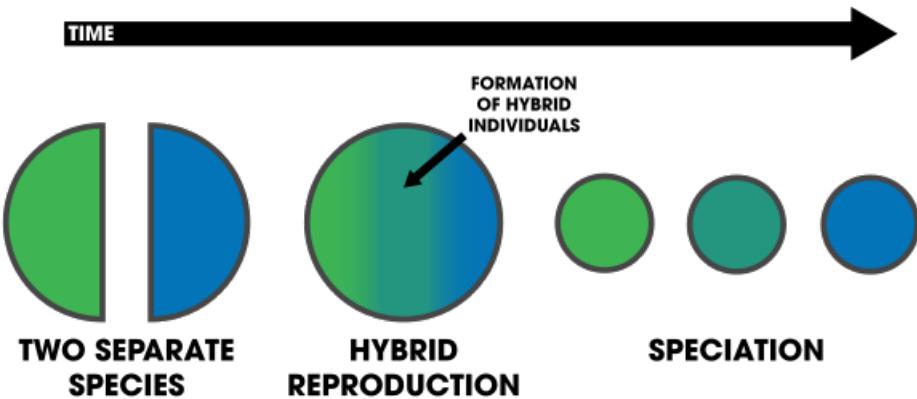
- Conjunto de indivíduos/populações que podem cruzar entre si e produzir descendentes férteis
- Para espécies se consolidarem como entidades independentes, é necessário o cessamento de fluxo gênico entre populações

## How many new species may arise



# O que é uma espécie?

- Nem sempre há cessamento de fluxo gênico
  - Zonas híbridas
  - Populações em processo de especiação
- Híbridos podem ‘fundar’ uma nova espécie



# O que é uma espécie?

---

Na prática:

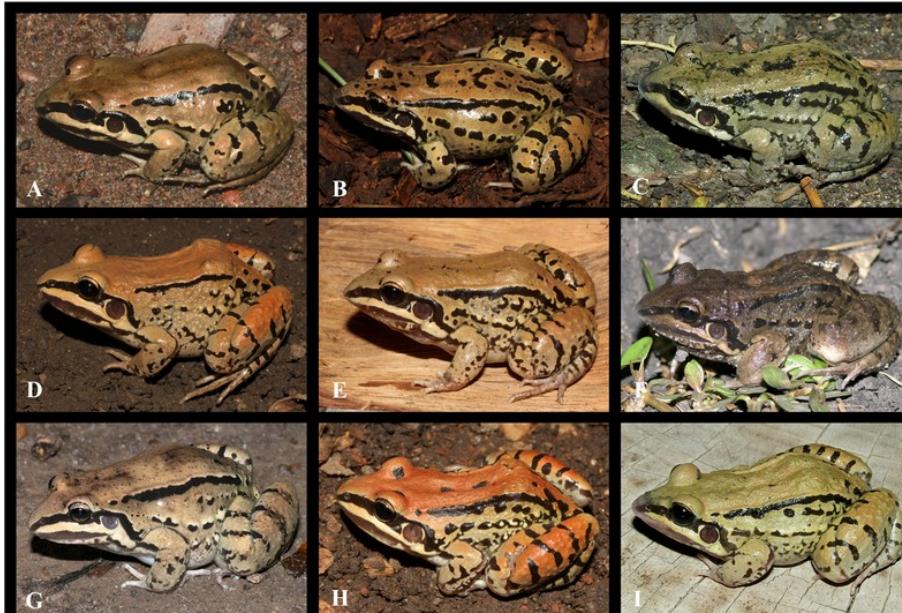
- Conceito morfológico: Conjuntos de indivíduos/populações com características morfológicas mais ou menos homogêneas que as distinguem de outro grupo

# O que é uma espécie?

---

Na prática:

- Conceito morfológico: Conjuntos de indivíduos/populações com características morfológicas mais ou menos homogêneas que as distinguem de outro grupo
  - Diversidade intraspecífica

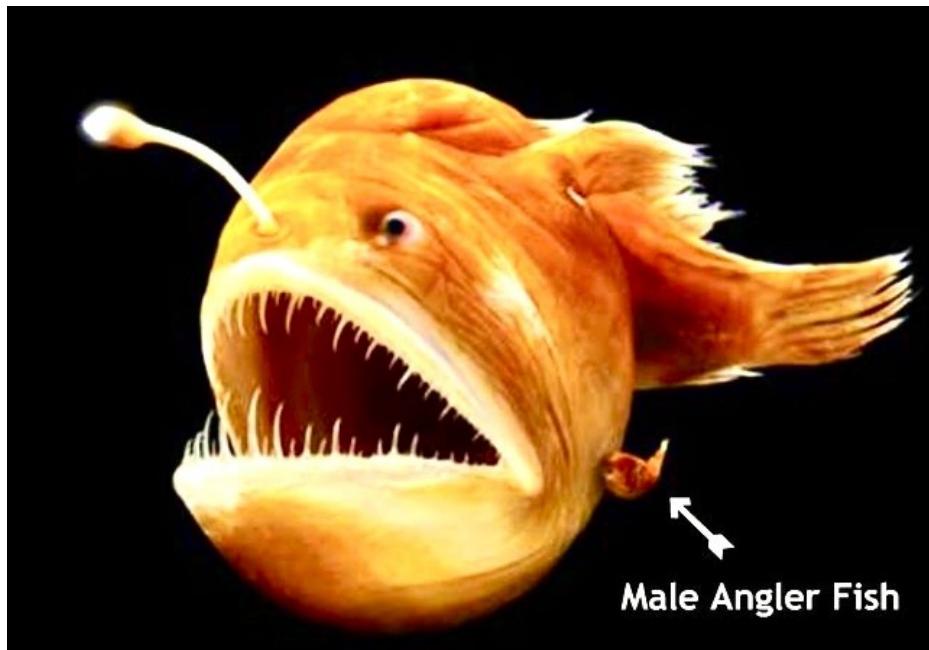


# O que é uma espécie?

---

Na prática:

- Conceito morfológico: Conjuntos de indivíduos/populações com características morfológicas mais ou menos homogêneas que as distinguem de outro grupo
  - Dimorfismo sexual

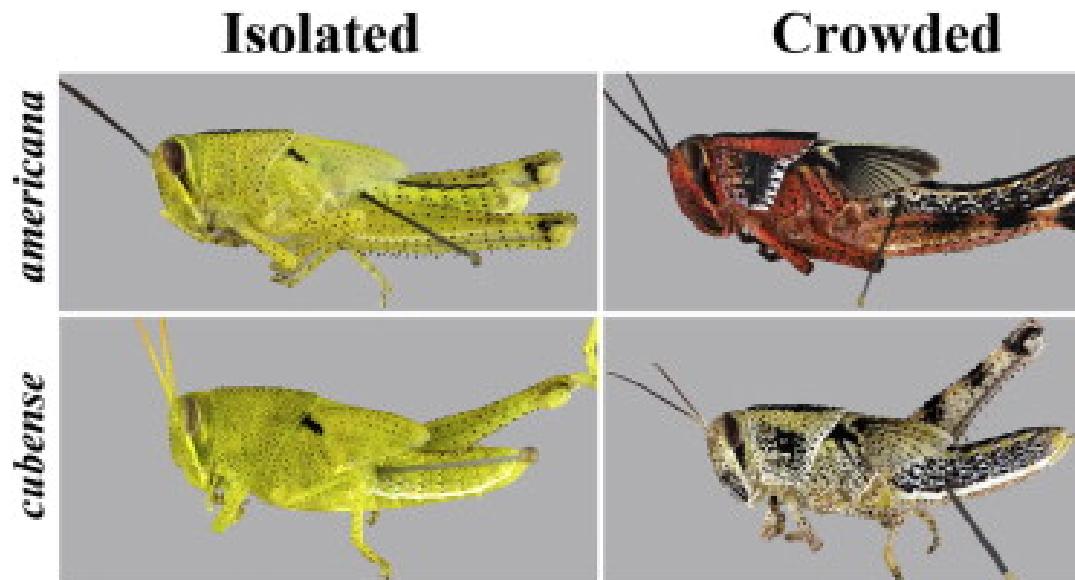


# O que é uma espécie?

---

Na prática:

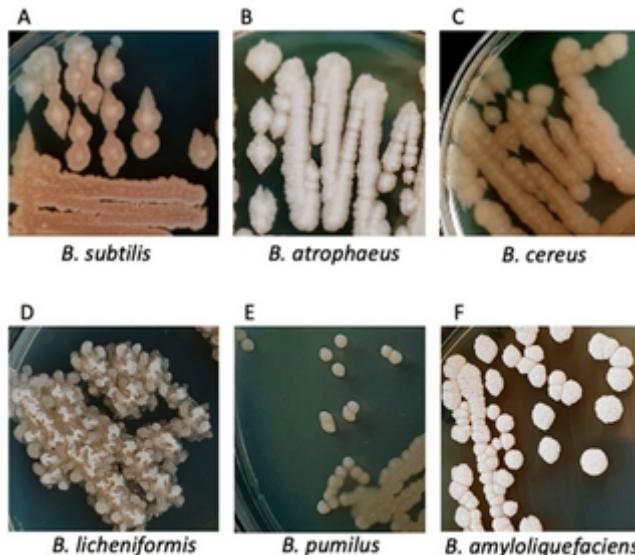
- Conceito morfológico: Conjuntos de indivíduos/populações com características morfológicas mais ou menos homogêneas que as distinguem de outro grupo
  - Plasticidade fenotípica



# O que é uma espécie?

---

- E em procariotos?
  - Não há um processo biológico relevante que distingua os diferentes grupos (como o fluxo gênico, no caso de organismos sexuados)
  - A distinção entre espécies geralmente é baseada em características fisiológicas e ecológicas compartilhadas entre grupos, e mais recentemente na comparação de genomas



# O que é uma espécie?

---

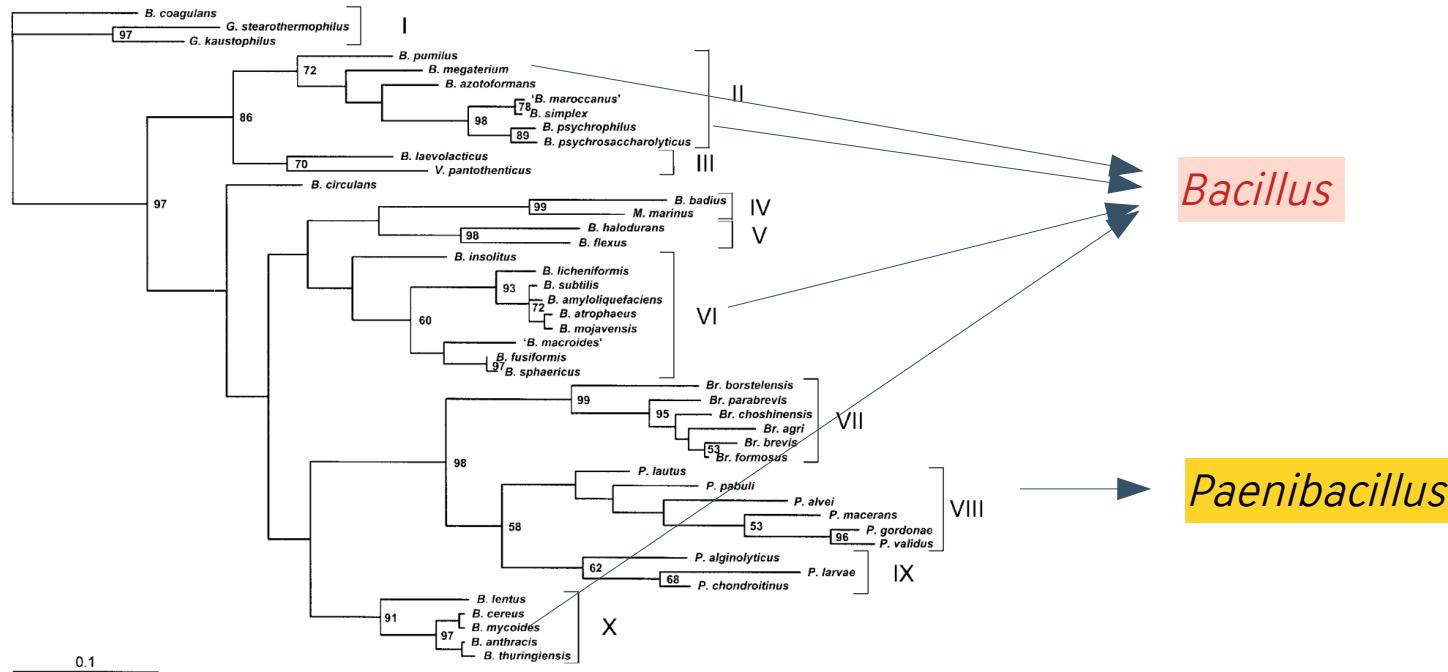
Problemas:

- Nomenclaturas em geral são anteriores à era filogenética (grupos monofiléticos)
- Revisões taxonômicas não são infrequeentes

# O que é uma espécie?

Problemas:

- Nomenclaturas em geral são anteriores à era filogenética (grupos monofiléticos)
- Revisões taxonômicas não são infrequentes



# Estratégias para identificação biológica

---

- Problemas reais que demandam identificação biológica
- Quando e por que usar filogenias

# A identificação biológica em problemas reais

---

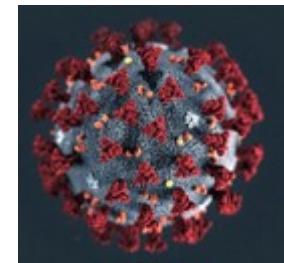
- Saúde
- Indústria
- Agricultura
- Meio ambiente

# A identificação biológica em problemas reais

---

## Saúde

- Infecções por bactérias, fungos, vírus e protozoários
- Macroparasitas



- Protocolos biomédicos já bem estabelecidos

Microbiologia: Cultura em placas → identificação por métodos químicos ou morfologia

Imunoreações → Antígenos para proteínas específicas do parasita

Microscópio → identificação por morfologia e/ou coloração (esporos, ovos, larvas, adultos)

Biologia molecular → PCR/qPCR com marcadores espécie/variedade-específicos

- Filogenias são essenciais na revisão dos sistemas de classificação de grupos de interesse na saúde, mas no dia-a-dia tem uso limitado a casos mais raros e epidemiologia

# A identificação biológica em problemas reais

---

## Indústria

- Controle sanitário (contaminação por micro/macrorganismos)
- Controle de qualidade
- P&D



- Protocolos em geral bem estabelecidos, como na saúde
- Controle de qualidade: em transição dos métodos clássicos para moleculares → maior resolução (nível de espécie/cepa) e assertividade
- P&D: identificação precisa de novas cepas/variedades exigem métodos moleculares (sequenciamento e montagem de genomas)
- Filogenias aumentam a resolução da identificação → distinção entre cepas/variedades

# A identificação biológica em problemas reais

---

## Agricultura

- Controle fitossanitário (macro e micro fitopatógenos)
  - Distinção entre cultivares (incluindo transgênicos)
  - Microbiologia do solo
  - P&D (bionsumos)
- 
- Protocolos fitossanitários em geral bem estabelecidos: uso de métodos tradicionais (sintomatologia, morfologia e eventualmente métodos moleculares)
  - Cultivares/variedades nem sempre podem ser distinguidos por morfologia → desenvolvimento de estratégias moleculares próprias para cada caso (análise de polimorfismos com múltiplos marcadores)

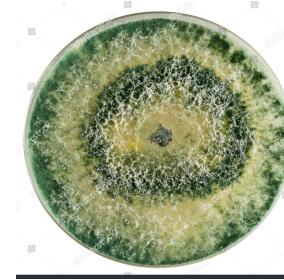


# A identificação biológica em problemas reais

---

## Agricultura

- Controle fitossanitário (macro e micro fitopatógenos)
  - Distinção entre cultivares (incluindo transgênicos)
  - Microbiologia do solo
  - P&D (bionsumos)
- 
- Microbiologia do solo → metataxonomia, metagenômica
  - P&D: identificação precisa de novas cepas/variedades exigem métodos moleculares (sequenciamento e montagem de genomas)
- 
- Filogenias aumentam a resolução da identificação → distinção entre cepas/variantes/raças de patógenos e bioinsumos



# A identificação biológica em problemas reais

---

## Meio ambiente

- Estudos de impacto ambiental
  - Monitoramento da fauna
  - Monitoramento da qualidade da água
  - Biodiversidade e prioridades de conservação
- 
- Levantamentos faunísticos e florísticos: em geral, baseados em morfologia (chaves dicotômicas de identificação)
  - Monitoramento da fauna: identificação através de morfologia (imagens), pode usar métodos moleculares
- 
- Filogenias podem ser usadas como ferramentas para estimar a biodiversidade de uma área; podem ser usadas para aumentar a resolução da identificação, quando necessário

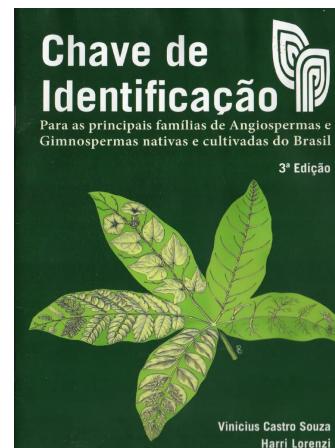
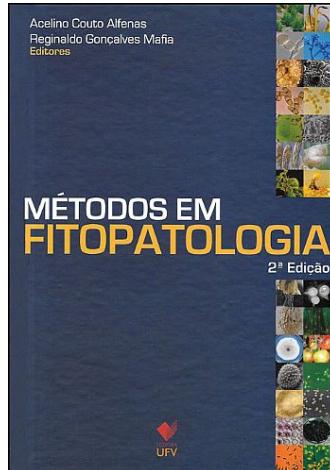


# Métodos morfológicos

- Treinamento técnico em morfologia
- Auxílio de 'autoridades' de cada grupo
- Chaves dicotômicas de identificação
- Identificação em nível de espécie nem sempre é possível  
(tipo de material, estado de conservação, estágio, chaves ruins, grupos pouco estudados...)



[www.ib.usp.br](http://www.ib.usp.br)



- 2(1') – Cabeça mais longa que larga (Fig 10-2); mandíbulas com dentes marginais proeminentes ou sem eles.....3
- 2' – Cabeça curta, côncava; mandíbulas sem dentes marginais.....Kalotermitidae
- 3(2) – Mandíbulas com ao menos 1 dente marginal proeminente.....4
- 3' – Mandíbulas sem dentes marginais (Fig 10-2) .....Rhinotermitidae
- 4(3) – Mandíbulas com apenas um dente marginal proeminente; cabeça estreitada anteriormente.....Termitidae
- 4' – Mandíbulas com mais de um dente marginal proeminente; cabeça não estreitada anteriormente.....Kalotermitidae

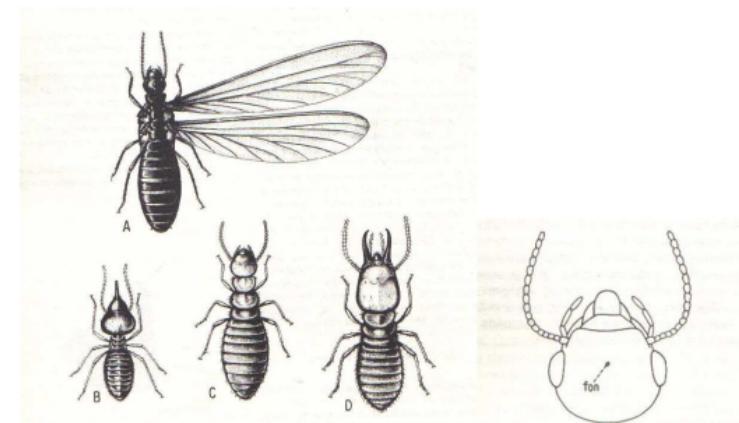


Fig. 10-2. As castas das termitas. A, adulto sexualizado de *Amitermes tuberculatus* (Buckley), 10 ×.; B, náuplio de *Constrictotermes tenuirostris* (Desnoux), 15 ×.; C, operário e D, soldado de *Prorhinotermes simplex* (Hagen), 10 ×. (Cortesia de Banks e Snyder e do U. S. National Museum.)

Fig. 10-3. Cabeça de *Prorhinotermes*, vista dorsal, mostrando a fontanela (fontanelle). (Modificada de Banks e Snyder.)

# Métodos moleculares

---

- Identificação genética/molecular geralmente tem maior resolução (distinção rápida entre raças/variedades/cepas)
- Exige laboratórios equipados → extração de DNA, PCR e sequenciamento
- Diferentes processos demandam treinamento técnico (técnicas de uso variado vs treinamento em grupos específicos para identificação baseada em morfologia)
- Reduz subjetividade e permite maior automatização dos processos (pipelines p/ identificação): Promessa dos códigos de barra de DNA ('DNA barcodes')

# Identificação biológica a partir de dados moleculares

---

- Marcadores moleculares
- Principais marcadores utilizados em procariotos e eucariotos
- Como abordar a identificação usando dados moleculares

# Marcadores moleculares

- Marcadores: qualquer região do genoma onde haja variação (suficiente) entre as amostras/grupos de interesse.

Exemplos:

- Polimorfismos de tamanho (RFLPs, AFLPs, etc)
- Polimorfismos de nucleotídeo único (SNPs)
- Sequências

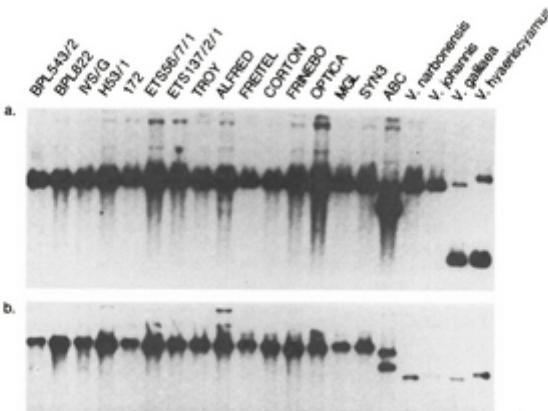
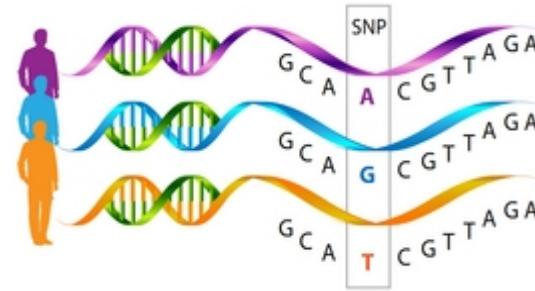
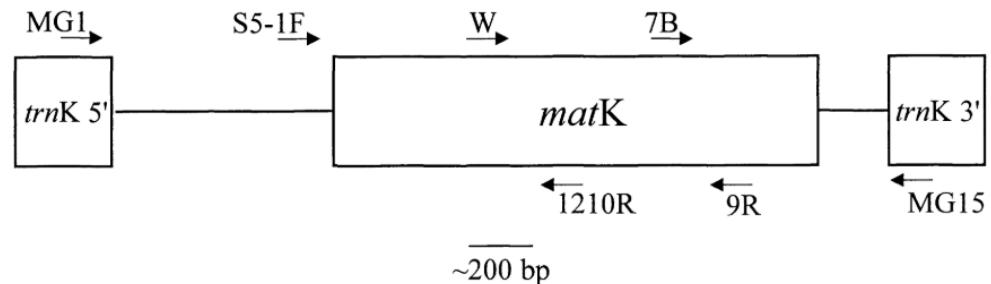


Figure 1 Polymorphism on the 20 *Vicia* accessions. (a) Restricted with enzyme *Hind*III and probed with cDNA probe 7-14. (b) Restricted with enzyme *Eco*RV and probed with cDNA probe 5-6.

De Ven et al. (1990)



[www.nutrigeneticsspecialists.com/](http://www.nutrigeneticsspecialists.com/)

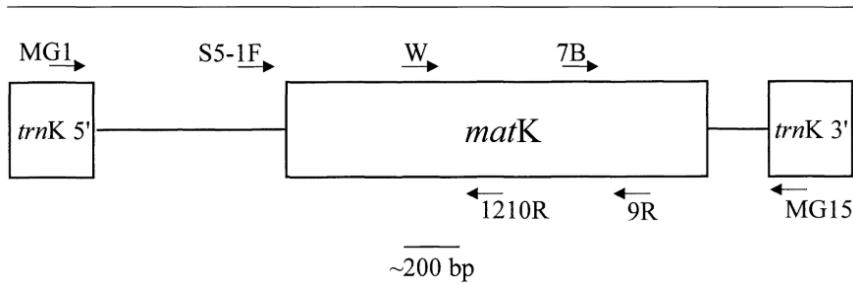


Hilu et al. (1999)

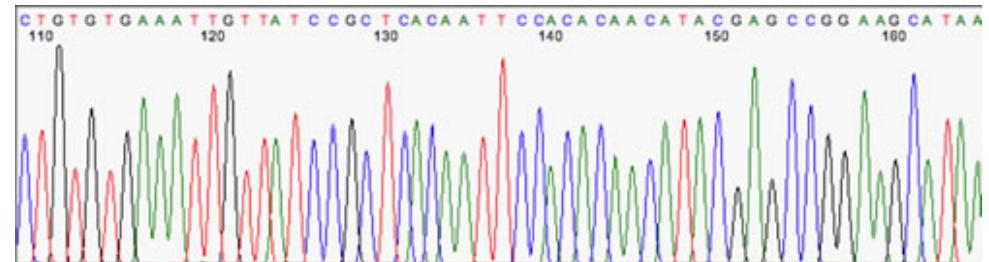
# Usando sequências como marcadores

Procedimento geral:

1. Extração de DNA
2. Amplificação via PCR da região de interesse (com primers que flanqueiam a região)
3. Sequenciamento (Sanger/Shotgun)



Hilu et al. (1999)



biology.unt.edu

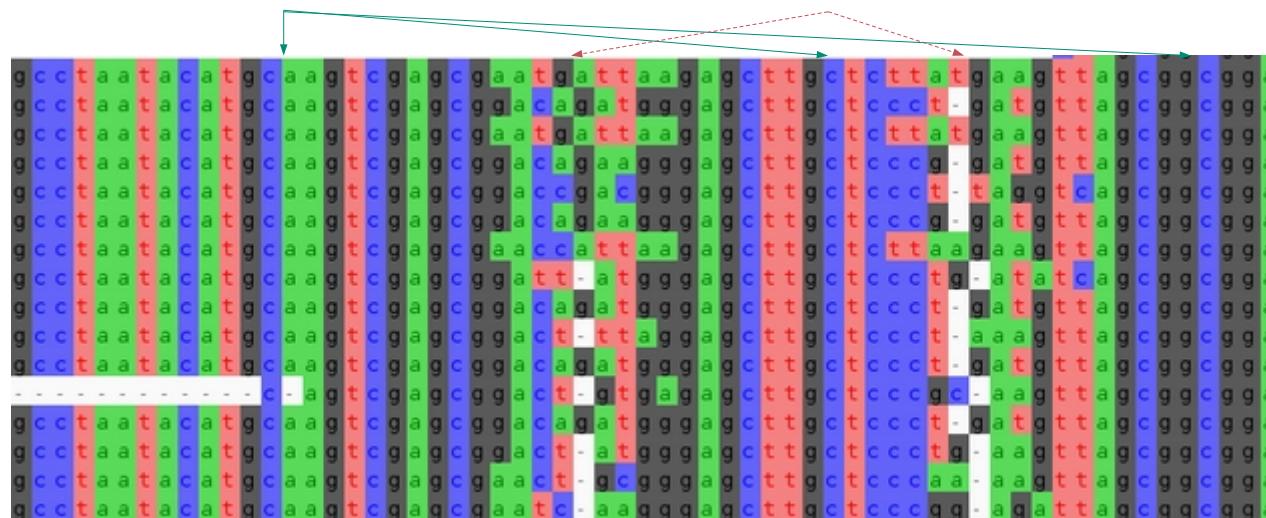
# Usando sequências como marcadores

## Características importantes de um marcador:

- Sequência suficientemente conservada, para existirem regiões homólogas entre sequências
  - Sequência suficientemente variável para poder diferenciar amostras

## Regiões conservadas

## Regiões variáveis



# Usando sequências como marcadores

---

Características importantes de um marcador:

- Evolução neutra

# Usando sequências como marcadores

---

Características importantes de um marcador:

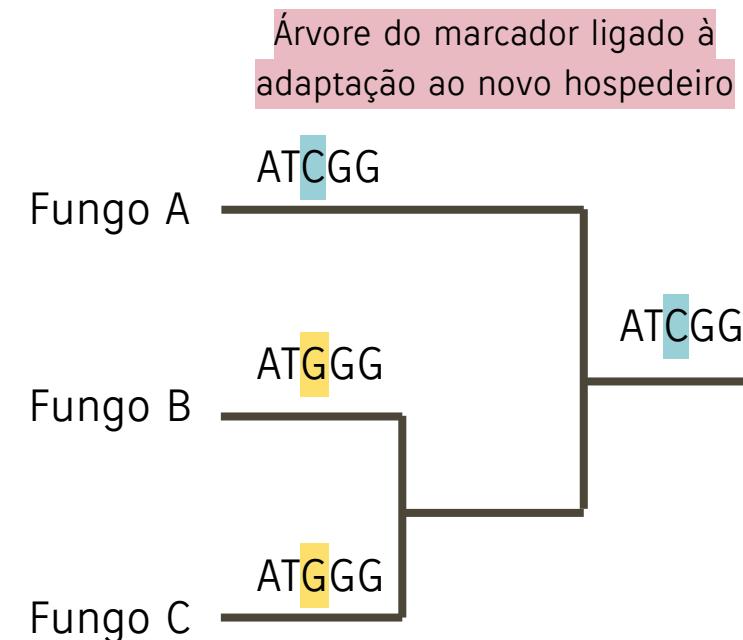
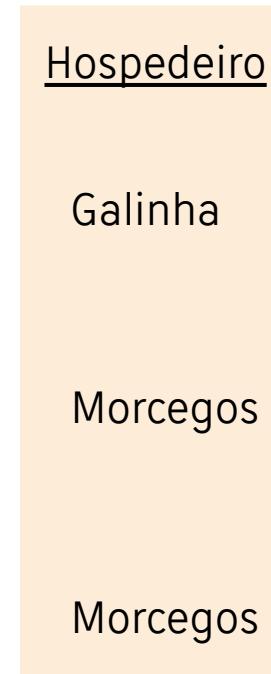
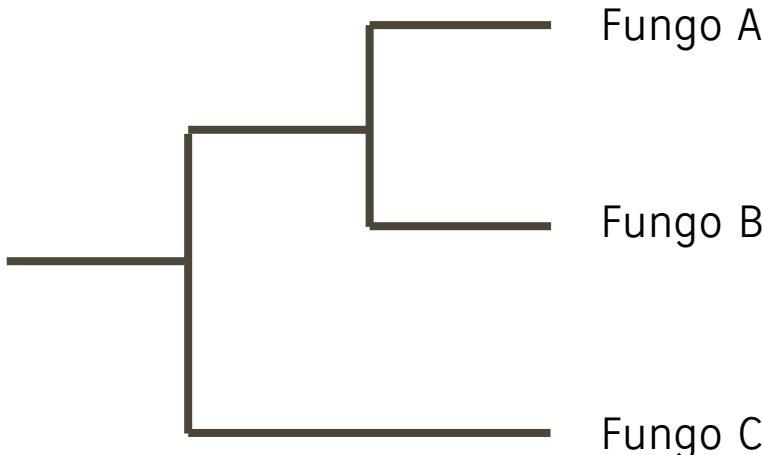
- Evolução neutra  
→ genes/regiões ligados a características adaptativas podem ter falso sinal filogenético!  
(convergência evolutiva)

# Usando sequências como marcadores

Características importantes de um marcador:

- Evolução neutra  
→ genes/regiões ligados a características adaptativas podem ter falso sinal filogenético!  
(convergência evolutiva)

Árvore verdadeira  
(maioria do genoma)

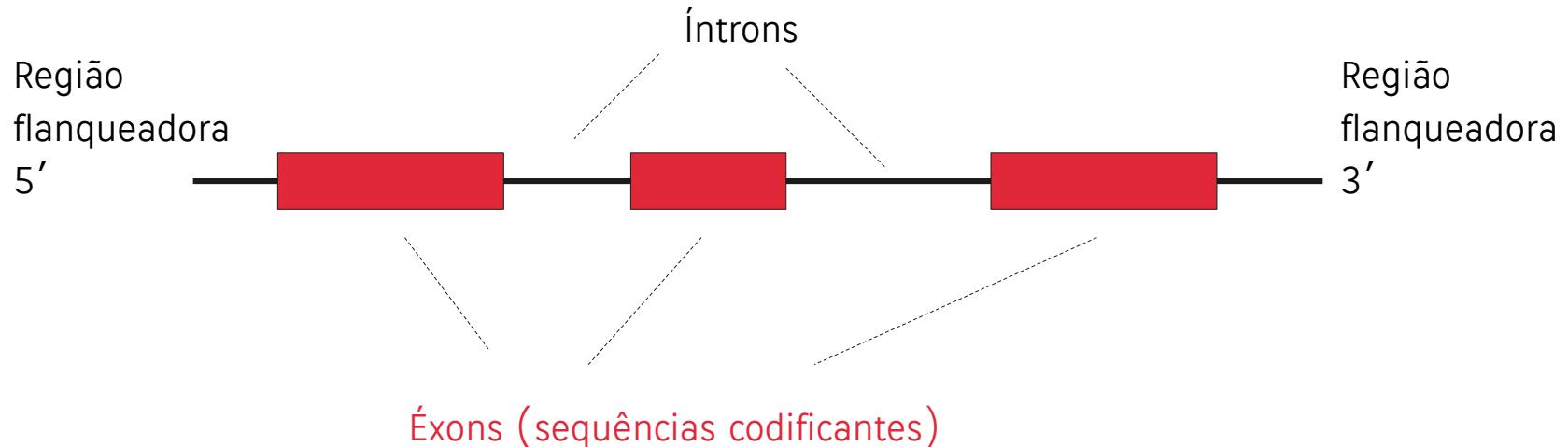


# Usando sequências como marcadores

---

Possíveis saídas quando suspeita-se que o marcador não é neutro:

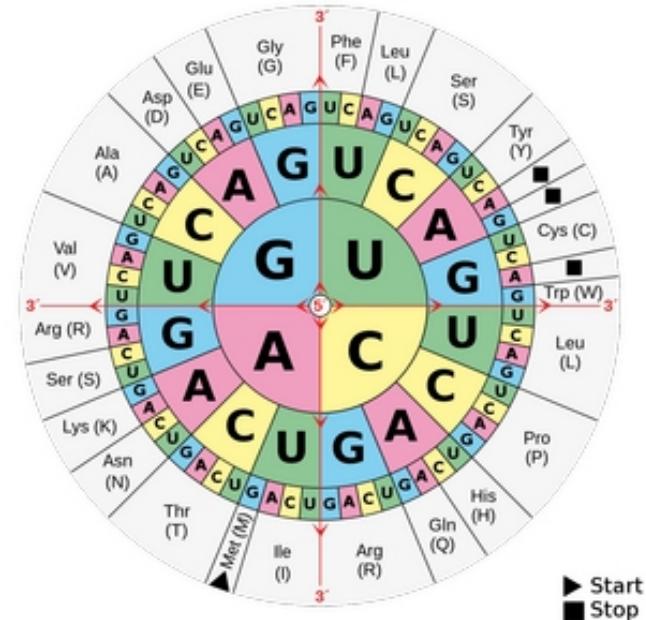
- Uso de regiões de evolução neutra → mudanças genéticas são ‘invisíveis’ para a seleção natural  
Íntrons, regiões flanqueadoras (problema: difícil identificar regiões homólogas; problema aumenta com a distância filogenética)



# Usando sequências como marcadores

Possíveis saídas quando suspeita-se que o marcador não é neutro:

- Uso de regiões de evolução neutra → mudanças genéticas são ‘invisíveis’ para a seleção natural  
Íntrons, regiões flankeadoras (problema: difícil identificar regiões homólogas; problema aumenta com a distância filogenética)
- Uso da terceira posição dos códons  
→ mudanças genéticas geralmente são sinônimas



# Usando sequências como marcadores

Qual seu objetivo?

- Análise filogenética entre domínios  
→ Marcadores ultraconservados

Como encontrar tais marcadores?

*Proc. Natl. Acad. Sci. USA*  
Vol. 74, No. 11, pp. 5088–5090, November 1977  
Evolution

## Phylogenetic structure of the prokaryotic domain: The primary kingdoms

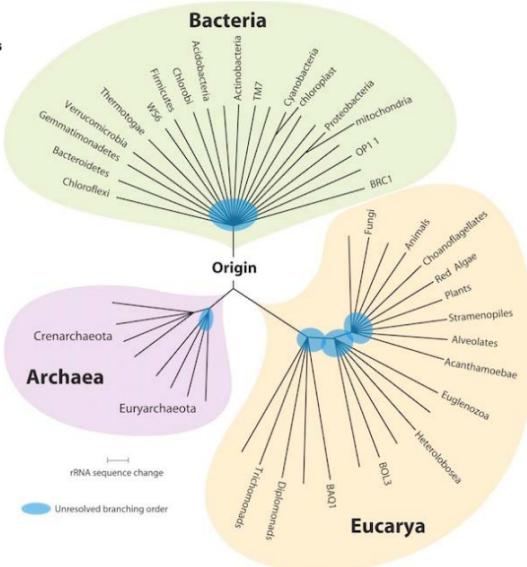
(archaeabacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX\*

Department of Genetics and Development, University of Illinois, Urbana, Illinois

Communicated by T. M. Sonneborn, August 18, 1977

**ABSTRACT** A phylogenetic analysis based upon ribosomal RNA sequence characterization reveals that living systems represent one of three aboriginal lines of descent: (*i*) the eubacteria, comprising all typical bacteria; (*ii*) the archaeabacteria, containing methanogenic bacteria; and (*iii*) the urkaryotes, now represented in the cytoplasmic component of eukaryotic cells.



# Usando sequências como marcadores

---

Qual seu objetivo?

- Identificação de espécies em larga escala  
(e.g. DNA ambiental, microbioma)
    - Marcadores conservados, pode ser grupo-específico
- Bactérias e árqueas: 16S
- Fungos e microeucariotos: ITS

# Usando sequências como marcadores

---

Qual seu objetivo?

- Análise filogenética de grupos de espécies
  - Marcadores podem ser grupo-específicos
  - Usar vários marcadores (árvore gênica nem sempre segue a árvore da espécie – próxima aula)

# Usando sequências como marcadores

---

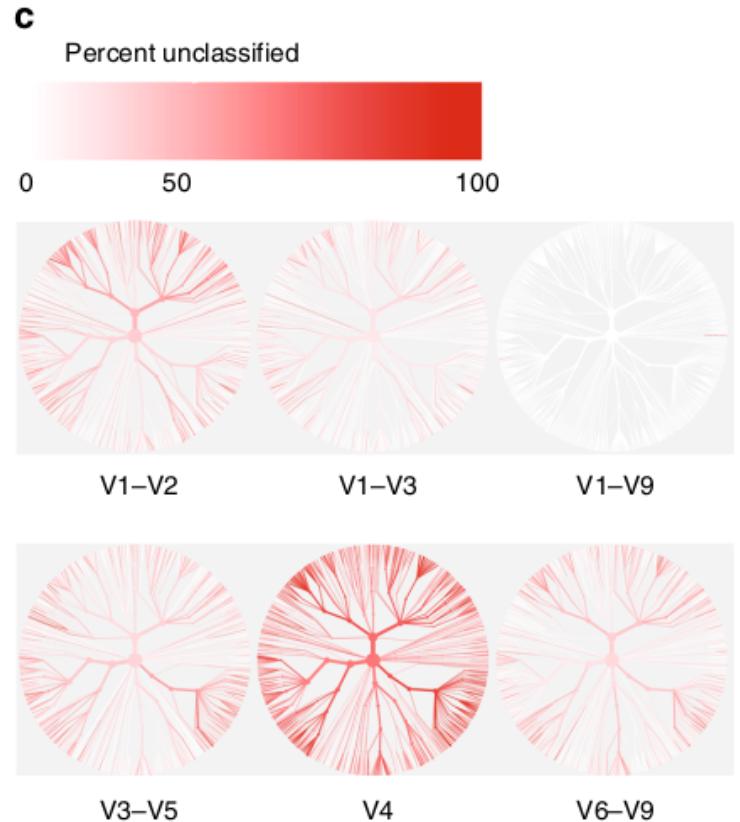
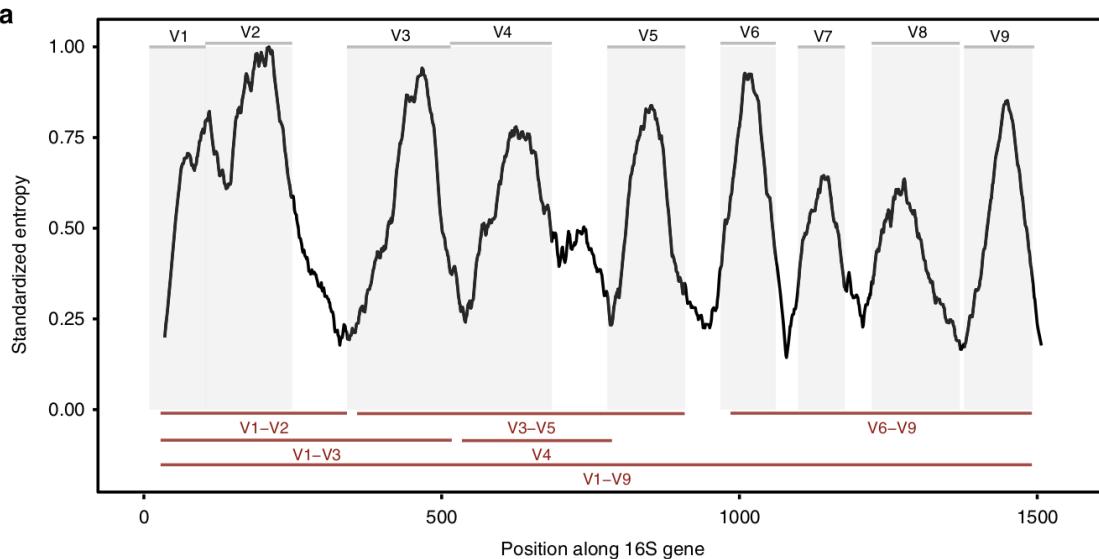
Qual seu objetivo?

- Distinguir populações dentro de uma espécie  
(espécie mais antigas)
  - Marcadores grupo-específicos
  - Usar muitos marcadores (centenas/milhares)
- (espécies mais recentes)
  - Varredura do genoma por marcadores grupo-específicos

# Marcadores comumente usados

## Bactérias e árqueas

- 16S rRNA
  - 9 regiões ‘hipervariáveis’ (V1–V9)
  - Resolução em nível de espécie: Depende do grupo, tamanho da sequência e das regiões usadas
  - Bancos de dados: SILVA, GreenGenes, NCBI RefSeq

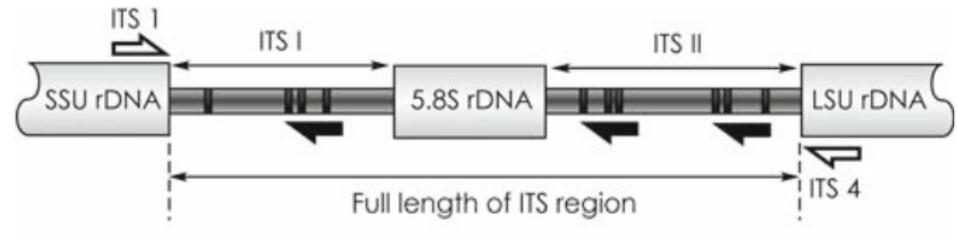


# Marcadores comumente usados

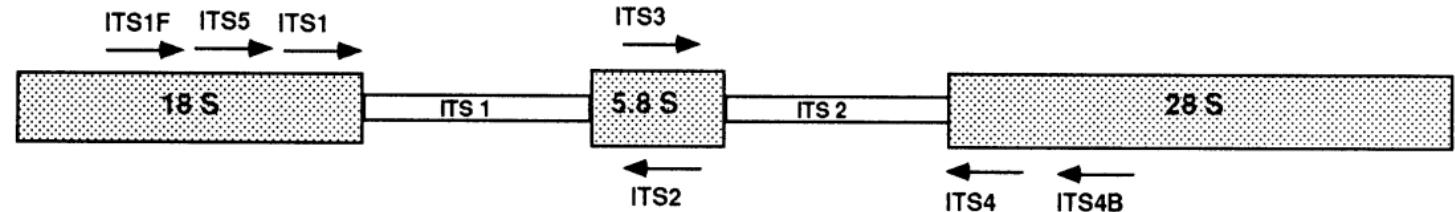
## Fungos

- ITS (Internal Transcribed Spacer)

- Resolução em nível de espécie: Depende do grupo, tamanho da sequência e das regiões usadas
- Bancos de dados: UNITE, NCBI RefSeq



Legend:  
■ Conservative region  
■ Variable region  
|| Species-specific sequence  
→ Universal primer  
↘ Species-specific primer



# Marcadores comumente usados

---

Fungos, Metazoa

- 18S rRNA
- COI (citocromo oxidase I) (mitocondrial)
- 12S rRNA (mitocondrial)
  - Bancos de dados: SILVA, NCBI RefSeq, PR2, bancos grupo-específicos



Plantas

- Maturase k (matk) (cloroplasto)
- rbcL (cloroplasto)
- ndhF (cloroplasto)
- ITS (nuclear)
  - Bancos de dados: NCBI RefSeq, bancos grupo-específicos

<https://academic.oup.com/database>

# Algumas considerações sobre marcadores

---

- Marcadores para análise filogenética nem sempre serão adequados para análises em larga escala (DNA ambiental) e vice-versa
- A promessa do código de barras de DNA (i.e. um gene para identificar espécies) funciona para alguns objetivos, para outros não; a identificação em nível de espécie é por vezes limitada
- Árvores gênicas nem sempre seguem a árvore da espécie (próxima aula)
- Na maioria dos casos, a identificação biológica em estudos ambientais é feita por similaridade de sequência (e.g. BLAST, ferramentas de machine learning) e não análise filogenética
  - 97%, 99% de identidade = mesma espécie? Funciona para todos os grupos?
  - Boa amostragem é fundamental (aula prática)

# Algumas considerações sobre marcadores

---

- Futuro próximo: Sequenciamento de genomas completos como rotina
- Ponto crítico: bancos de dados públicos e bem curados!

# Bases de dados públicas: demonstração NCBI

---

# Prática

---