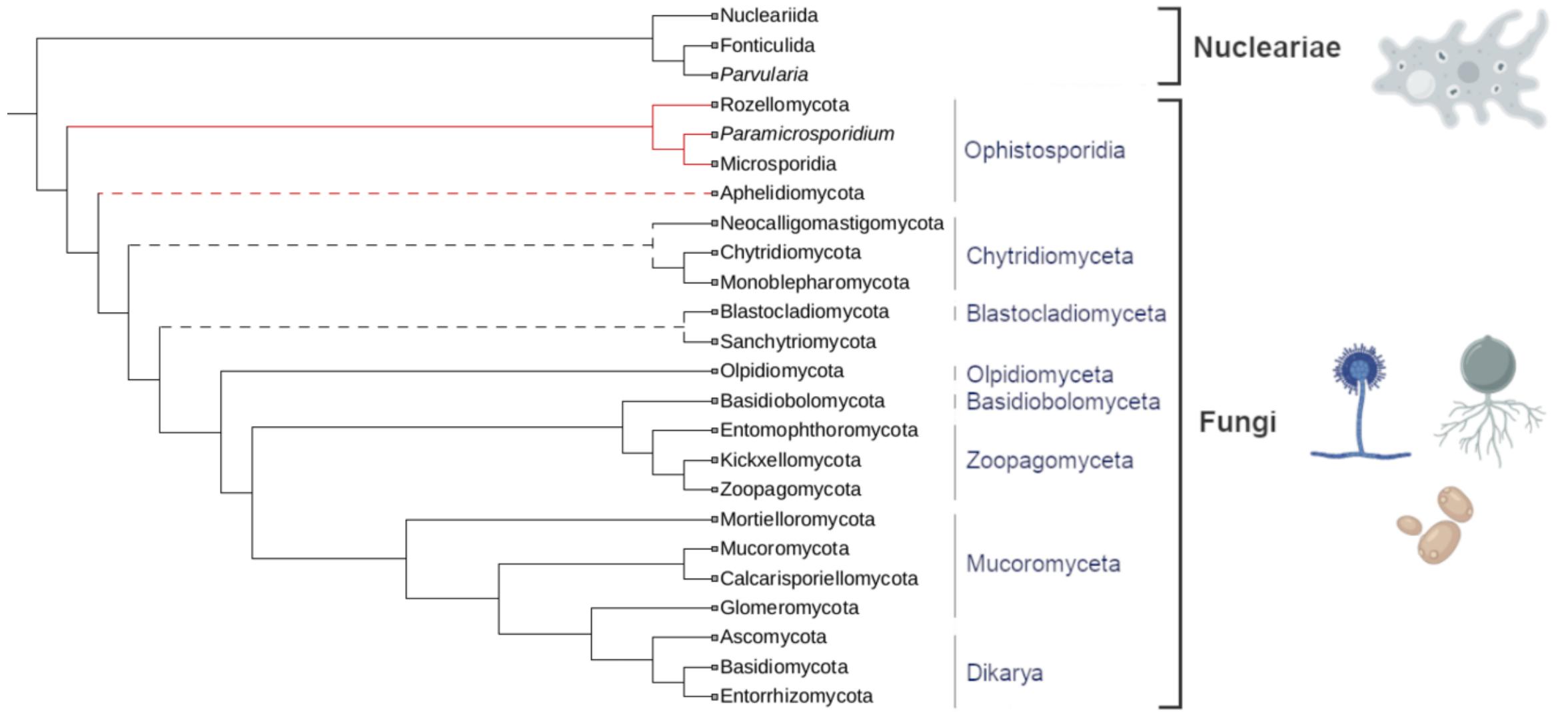


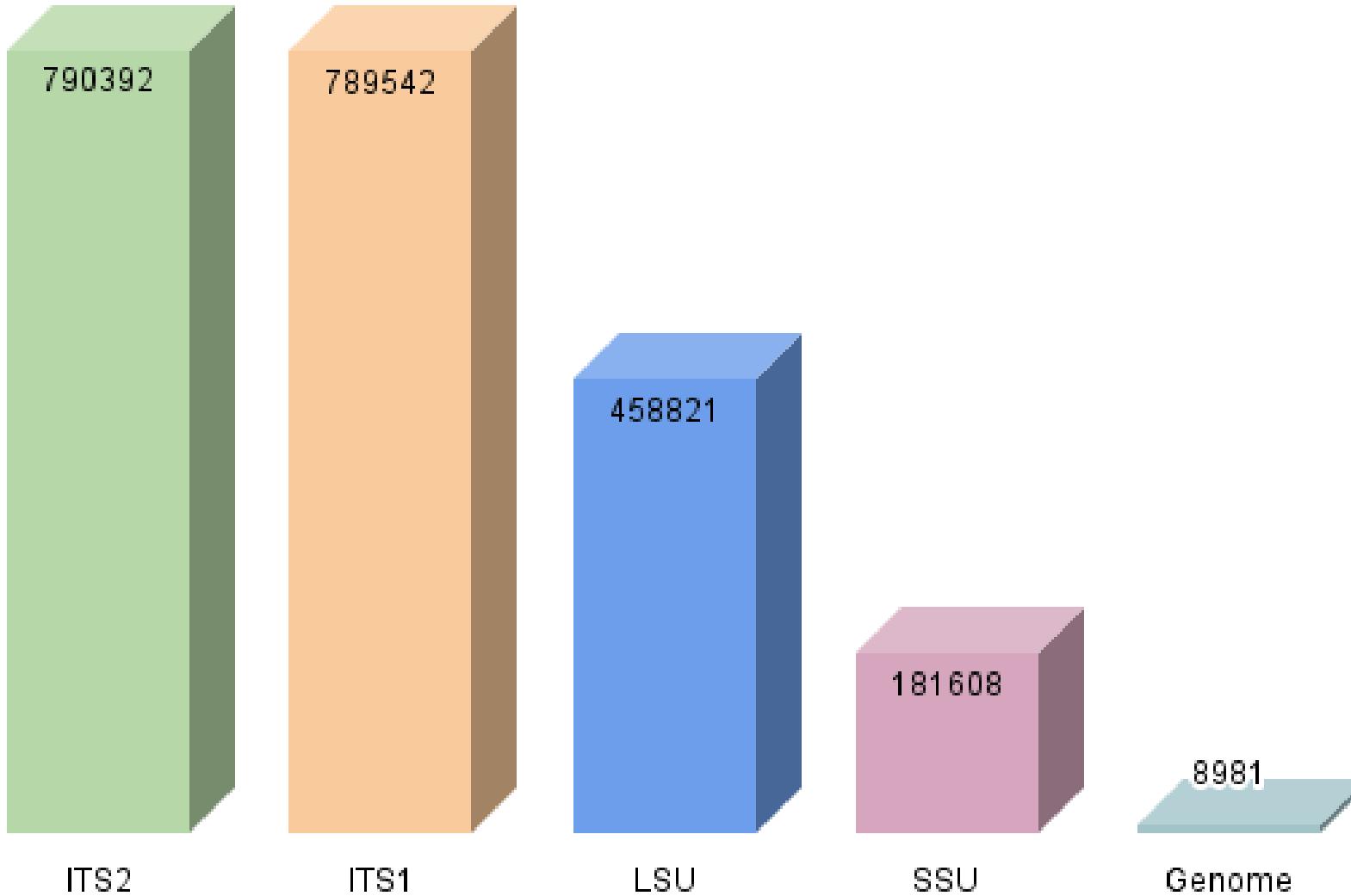
Curso sobre Genômica no Estudo de Microrganismos – Módulo Filogenoma

Professoras UFPR
Dra. Chirlei Glienke
Dra. Desirrê Alexia Petters Vandresen
GS Treinamentos e Consultoria



Ramos coloridos em **vermelho** indicam clados potencialmente parafiléticos
 Ramos pontilhados indicam filos cujas topologias ainda são contestadas

Comparação entre o número de barcordings genéticos publicamente disponíveis e de genomas



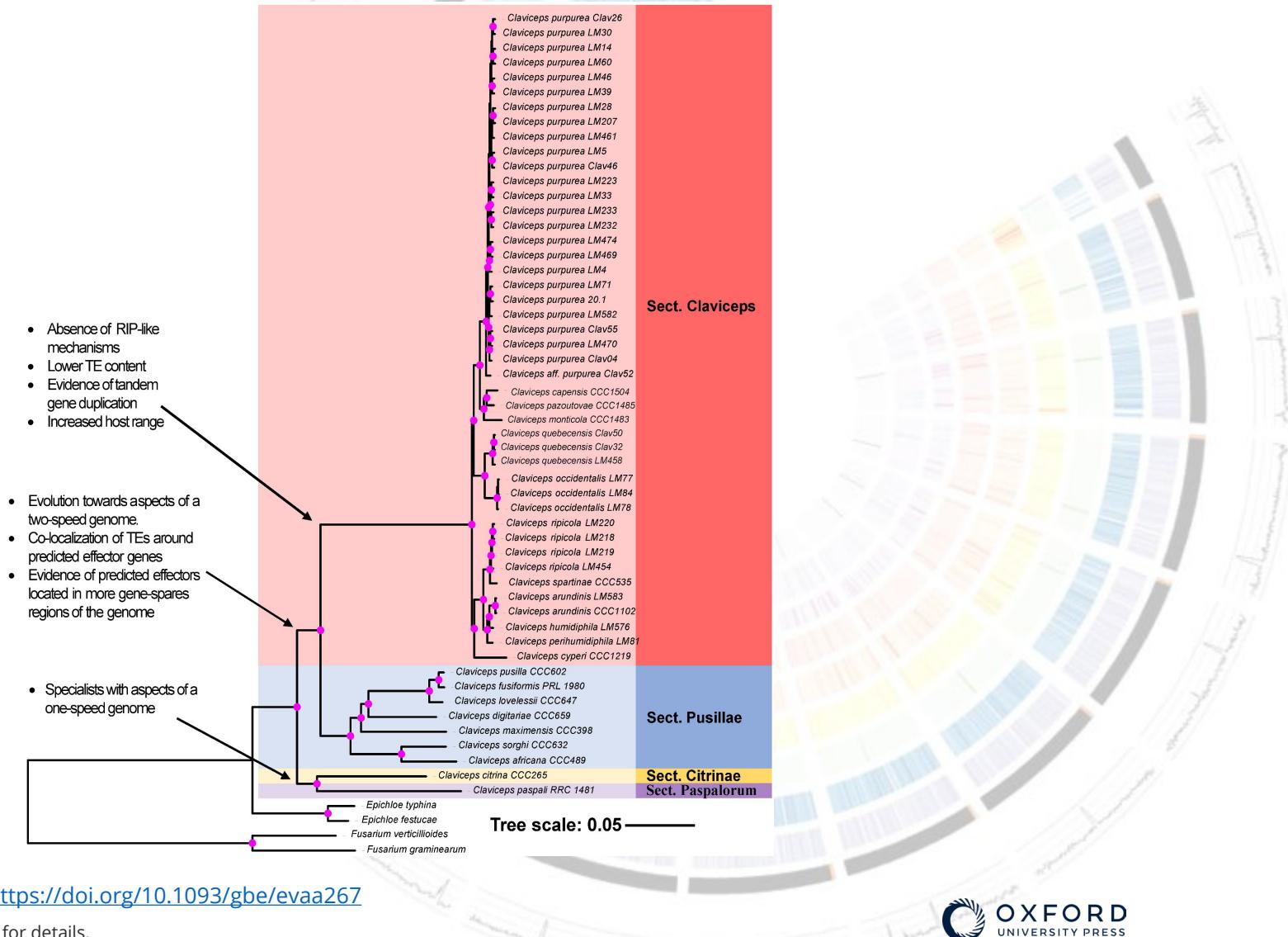
Filogenômica

- Definição: análises baseadas no genoma completo para inferir filogenia
- Na prática: não necessariamente todos os genes de um determinado conjunto de espécies: Ortólogos de Cópia Única (SCOs)
- Por que?
 - uso de parálogos, genes duplicados de qualquer tipo ou pseudogene pode causar ruído filogenético nos dados



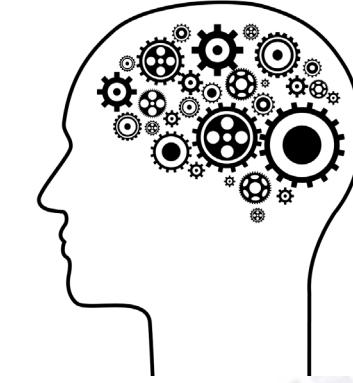
Reconstrução filogenética (baseada em ML) do gênero *Claviceps* usando sequências de aminoácidos de 2.002 ortólogos de cópia única com 1.000 réplicas de bootstrap

Pink dots at branches represent bootstrap values ≥ 95
Arrows and descriptions indicate potential changes in genomic architecture between *Claviceps* sections identified in this study



Planejando.....

- Qual o meu objetivo?
 - O que eu quero responder?
 - **Identificação** ou **delimitação** de espécies?
- Qual minha hipótese?
 - O que eu sei sobre o meu organismo?
- Quais genomas estão disponíveis?
- Quais linhagens eu devo ter?
- Onde achar?
- Como construir e interpretar uma árvore filogenética/filogenômica?



Planejando.....

- Qual o meu objetivo?
- O que eu quero responder?
- Identificação de gênero x espécie?
- Quando eu preciso primeiro fazer delimitação de espécies baseada em análise filogenética?



G

Identificação de espécies

- Quando fazer?
 - Gêneros e espécies já resolvidas (quando as espécies já estão delimitadas)
 - Identificação de uma amostra comparando à bancos de dados – *type strain, type species*
 - Uso de DNA barcoding ou genomas

Delimitação de espécies

- Quando fazer?
 - Reconstrução filogenética utilizando todas as *type strains* de todas as espécies aceitas do gênero
 - Famílias já resolvidas com gêneros aceitos - delimitação de gêneros caso necessário
 - Complexos de espécies: pode necessitar de filogenômica

Descrição de espécies

- Análise filogenética (ou filogenômica) utilizando todas as *type strains* de todas as espécies aceitas do gênero
- Descrição morfológica detalhada
- Fisiologia/associação com hospedeiro
- Depósito em bancos de dados (p.ex. Sequencias ou genomas no GenBank, Alinhamentos e árvores no Treebase, Espécies no Mycobank ou correlatos)
- Depósito de exemplar em coleção biológica

Conceitos básicos

- Análise filogenética



Análise Filogenética – conceitos básicos

- O DNA de qualquer espécie acumula mutações ao longo do tempo
- Quando duas espécies surgem a partir de um ancestral comum, deixa de ocorrer fluxo gênico entre elas (salvo em casos de transferência horizontal) e dessa forma, passam a acumular mutações distintas
- O número de mutações acumuladas tende a ser proporcional ao tempo de divergência entre as espécies

Análise Filogenética – conceitos básicos

- Consequência:

A análise destas mutações permite a inferência do processo evolutivo dos organismos que estamos comparando

Análise Filogenética – quando fazer

- Para representar as relações entre espécies;
- Para descrever a história de populações;
- Para descrever as dinâmicas evolutivas e epidemiológicas dos patógenos;
- Análise filogenética – genealogia – quanto mais próximos dois indivíduos, mais similares serão as sequências de DNA;
- Estimativa do tempo de divergência (possuem ancestral comum);

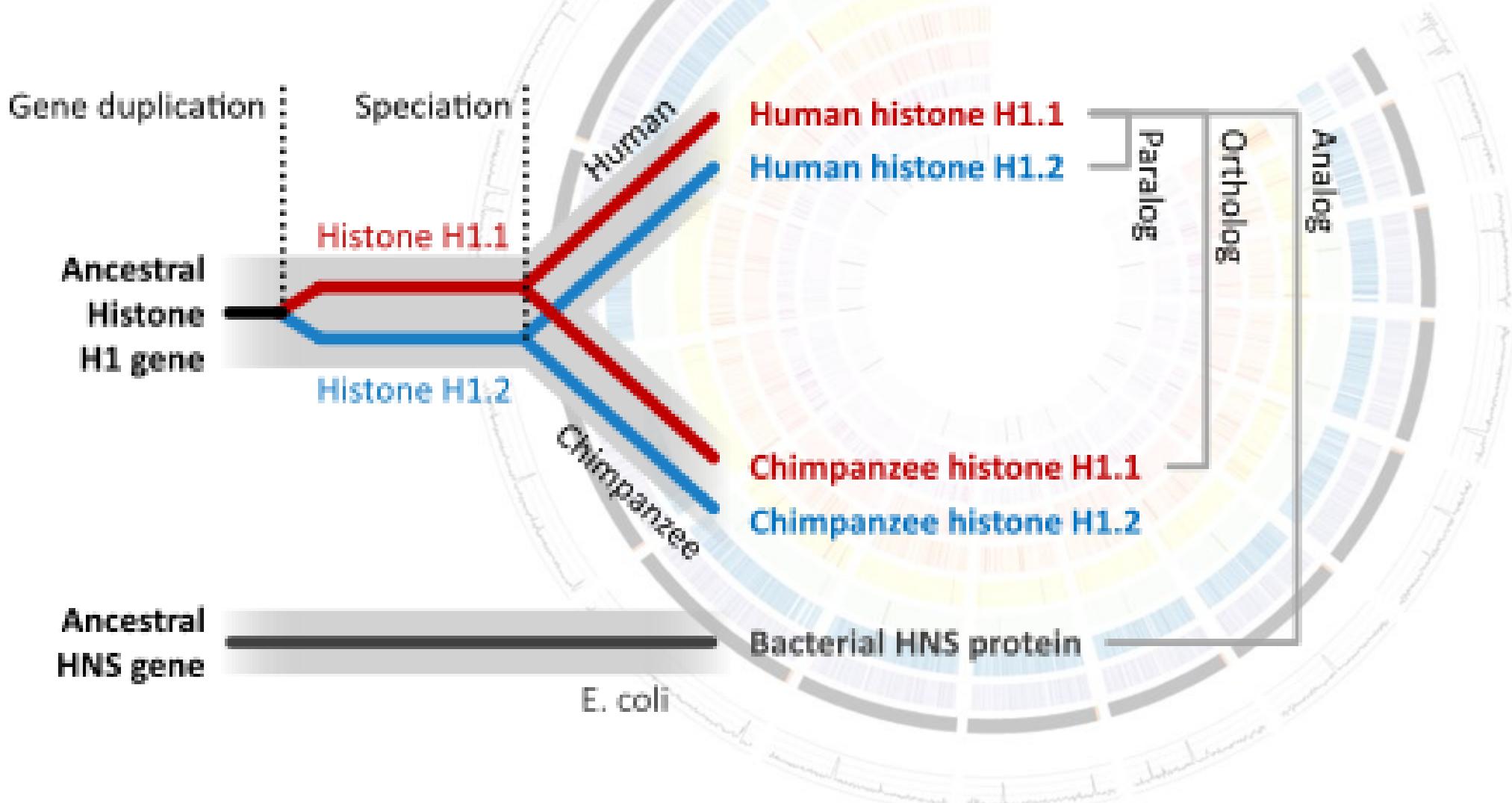
Análise Filogenética – definições

- **Filogenia:** Ajuda a inferir a história evolutiva das espécies e verifica os relacionamentos entre estas espécies, a fim de determinar possíveis ancestrais comuns entre elas
- **Árvore Filogenética:** É a representação visual da história evolutiva dos organismos, baseada em análises rigorosas.
 - É uma árvore onde as folhas representam os organismos (ou as sequencias estudadas) e os nós internos, seus supostos ancestrais. Diagrama que mostra as linhagens e relações dos organismos
- **Homologia:** A relação entre sequências que compartilham uma sequência ancestral comum

Análise Filogenética – definições

- **Homologia:** A relação entre sequências que compartilham uma sequência ancestral comum.
 - Importante: Saber se os genes são ortólogos ou parálogos
 - Parálogos: genes que se originaram de um evento recente de duplicação.
 - Se a filogenia for realizada com tais genes, a informação que iremos obter será sobre o evento de duplicação
 - Ortólogos: Genes homólogos em diferentes espécies que evoluíram independentemente por causa da especiação.
 - Se a análise filogenética for realizada com tais genes, a informação que teremos será sobre o evento de especiação

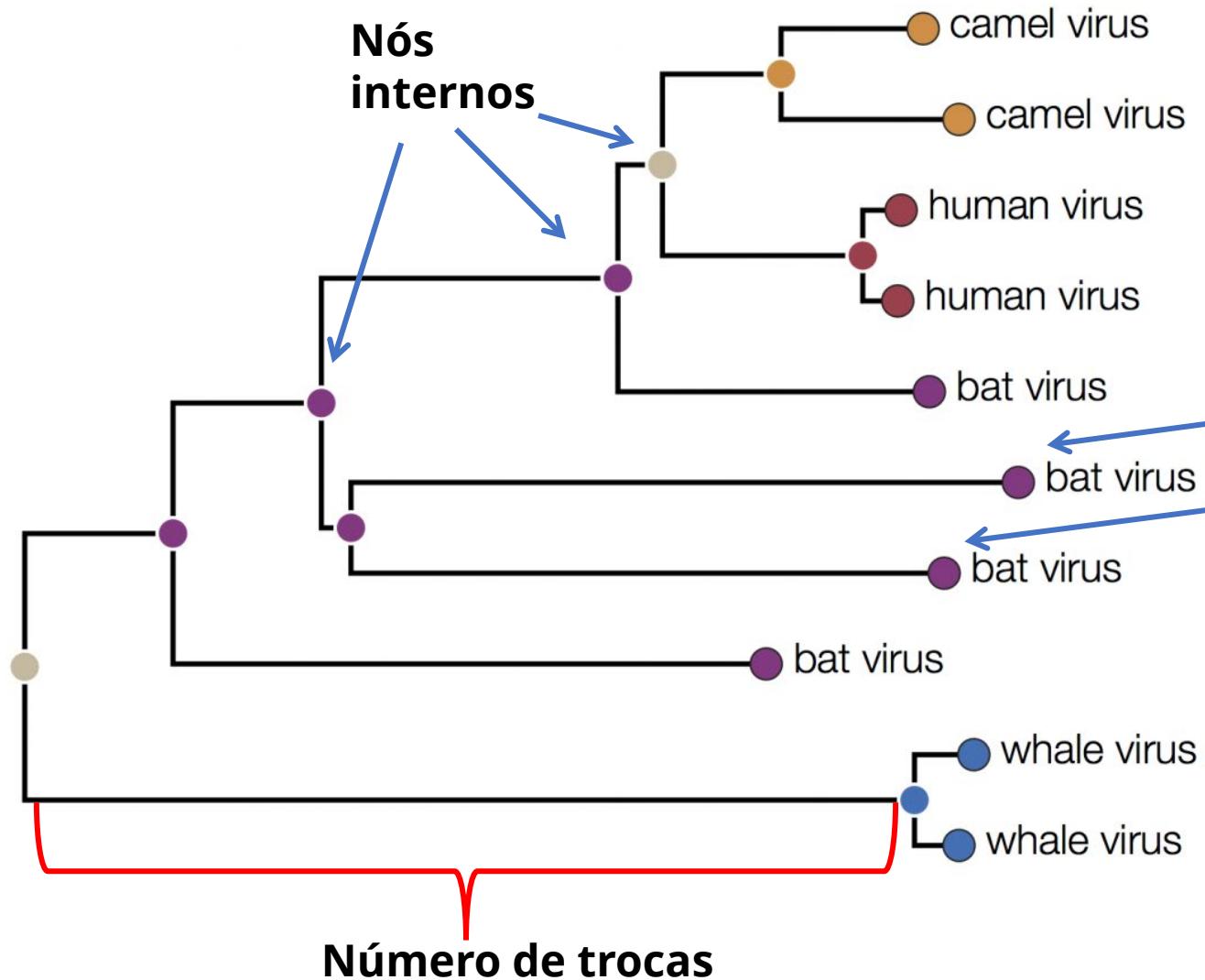
GENES CÓPIAS, PARÁLOGOS OU ORTÓLOGOS



Análise Filogenética – definições

- **Nós e Ramos** – conecta os nós e representa a quantidade de trocas genéticas entre o nó ancestral e o descendente
 - Nó externo: ponta da árvore (linhagens)
 - Nó interno: taxa ancestral (não presente)
- **Comprimento dos ramos:** número de trocas

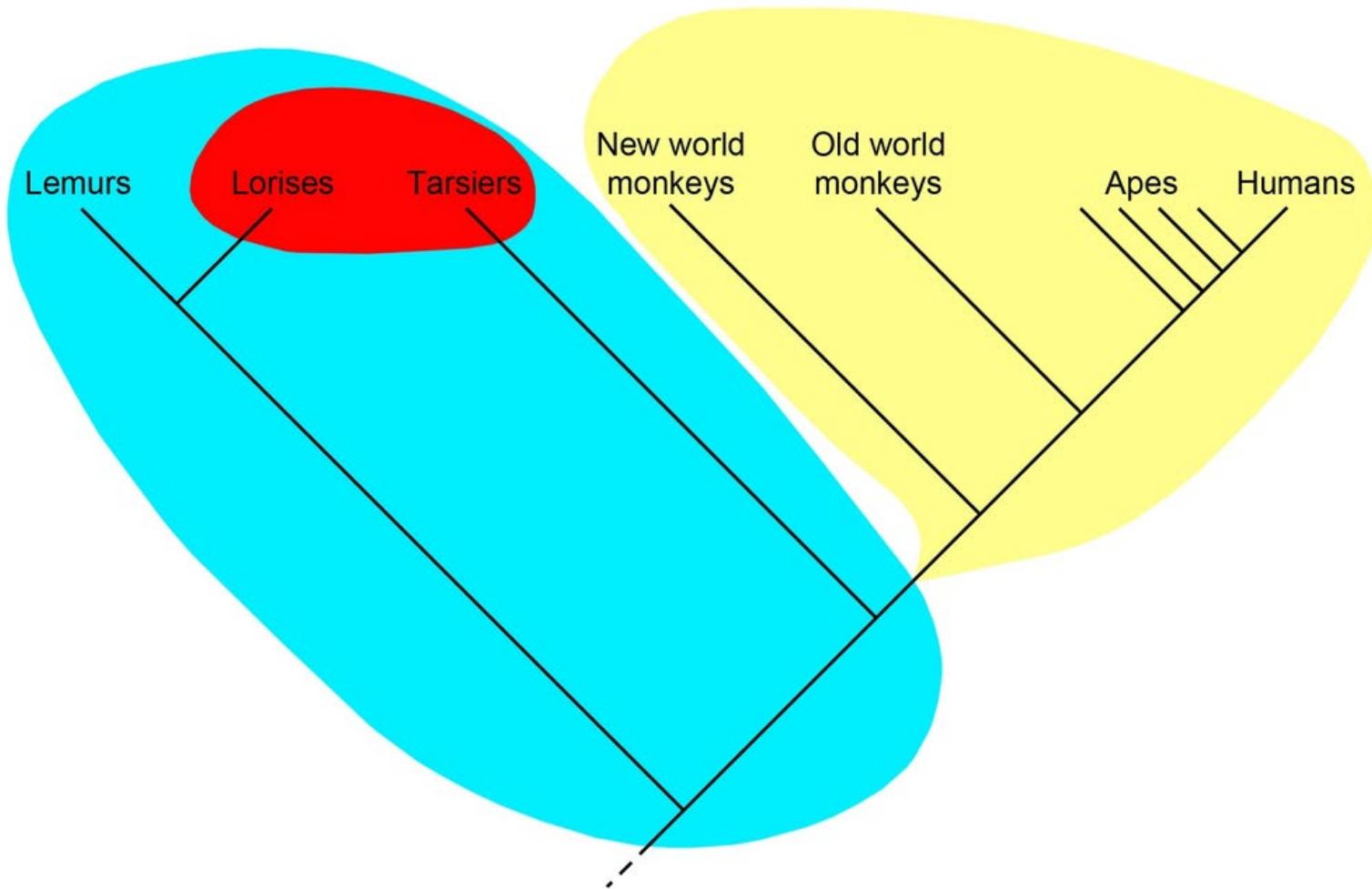
Árvore Filogenética



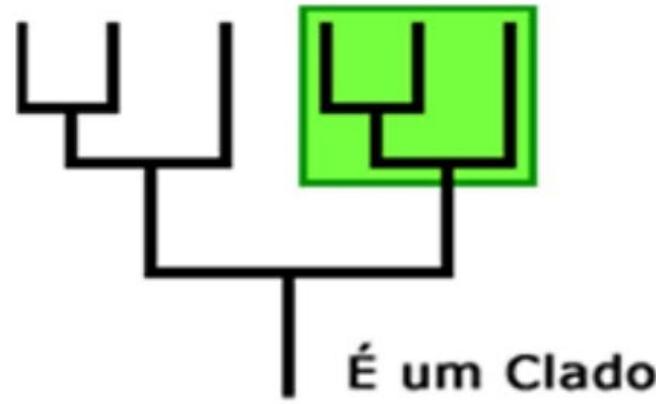
Árvore Filogenética - definições

- **Monofilético:** Um grupo de taxa (indivíduos da análise) que compartilham o mesmo braço, também chamado de cluster – inclui o ancestral e todos os descendentes
- **Parafilético:** Um grupo de taxa que não formam um cluster sem incluir linhagens adicionais – inclui os ancestrais mas não todos os descendentes

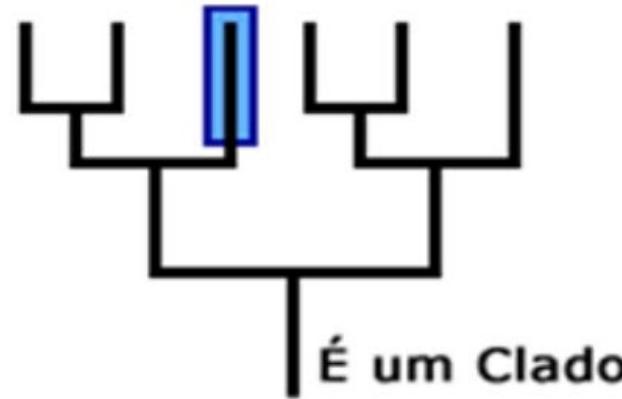
- Monofilético (ou natural)
- Parafilético
- Polifilético



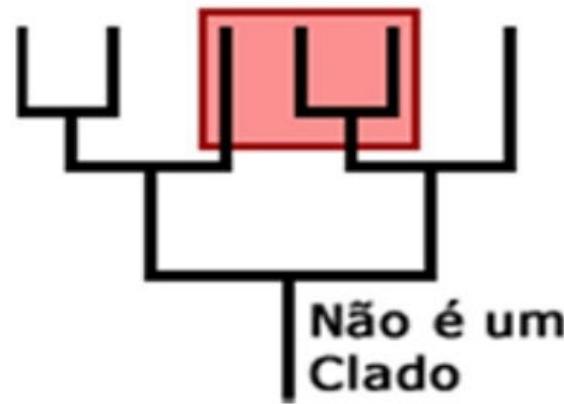
Árvore Filogenética - definições



É um Clado



É um Clado



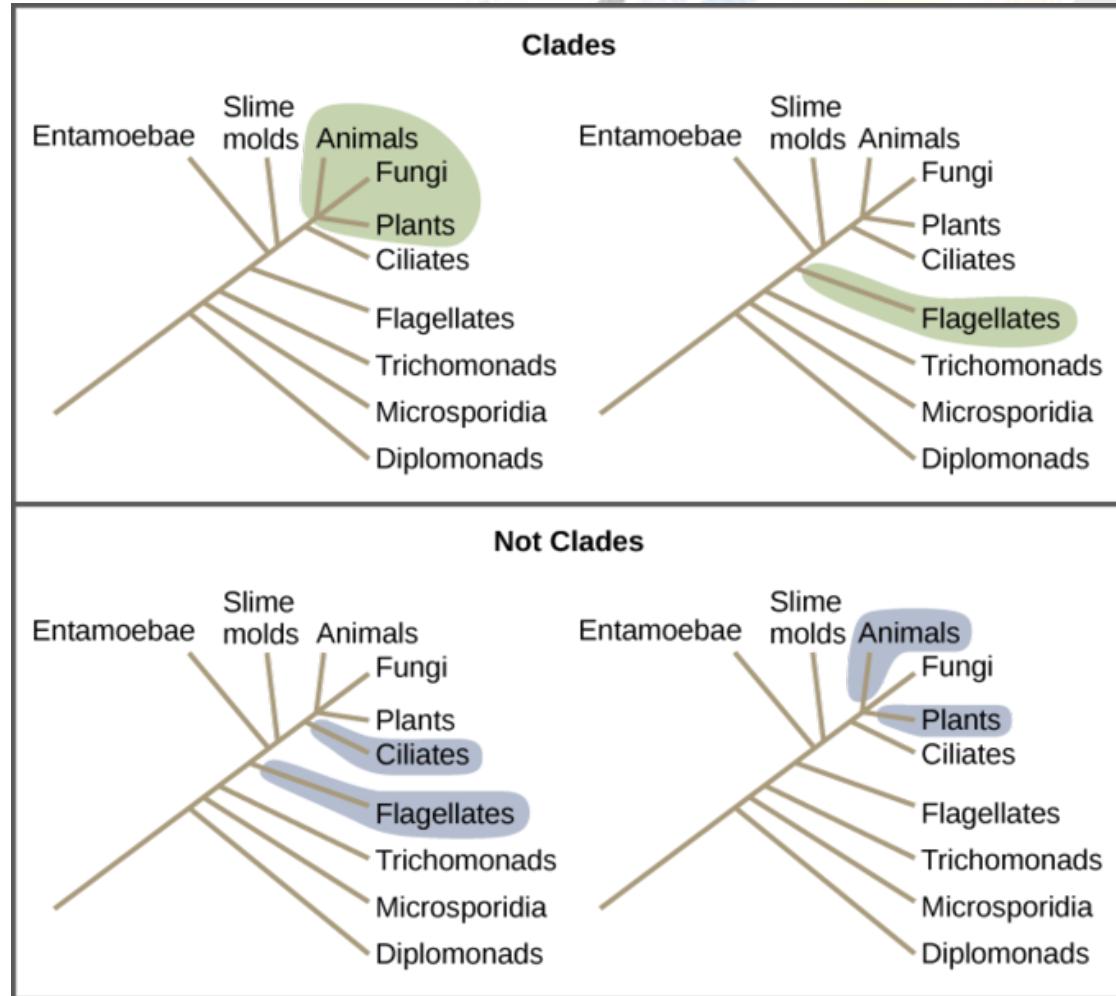
Não é um
Clado



Não é um
Clado

clados: grupos de organismos que descendem de um único ancestral

Árvore Filogenética - definições



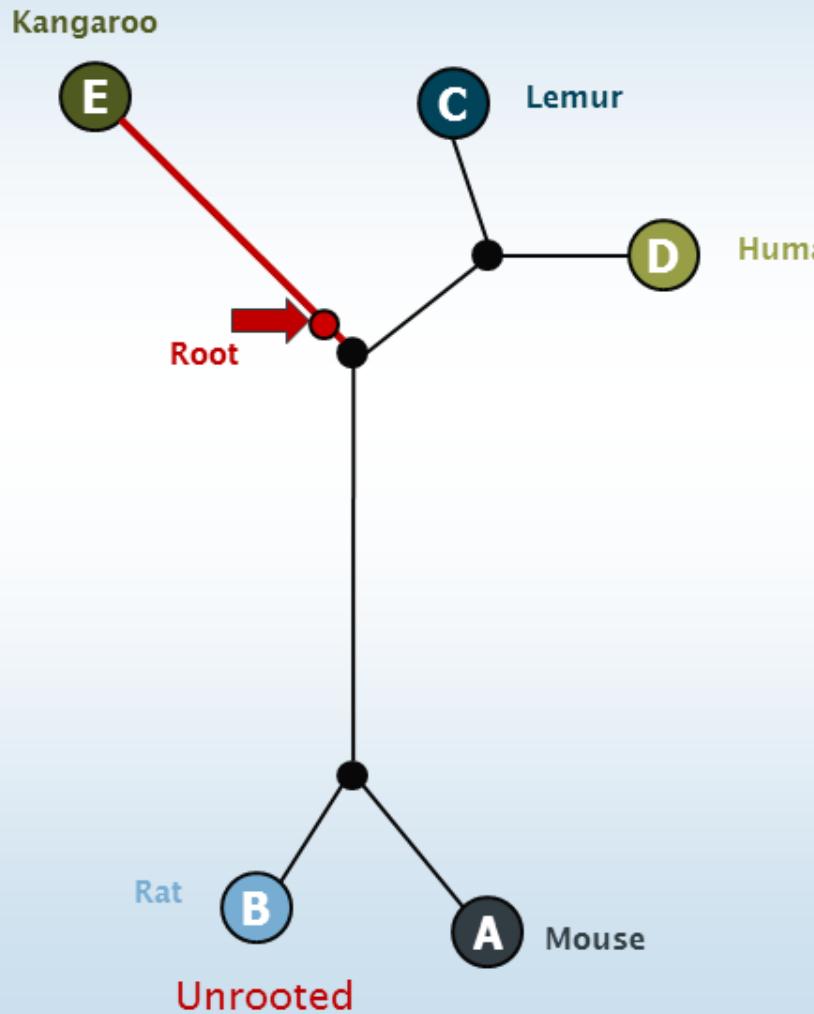
Árvore Filogenética

- Dois interesses principais:
 - * Obter a topologia da árvore - a forma como os nós internos se conectam uns com os outros e com nós da extremidade
 - * Obter as distâncias entre todos os nós da árvore

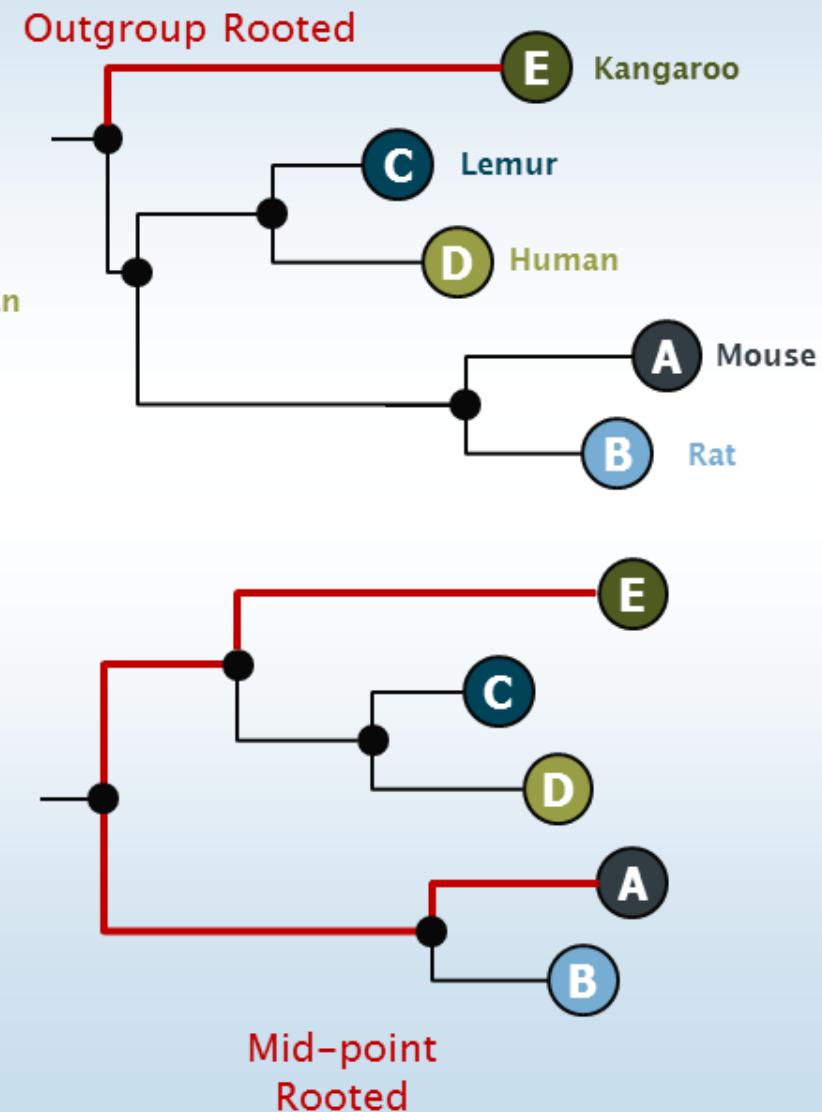
Árvore Filogenética

- A raiz de uma árvore filogenética:
 - Na árvore com raiz (ou enraizada), a raiz representa o ancestral comum a todos os nós da árvore
 - Sem informações suficientes para determinar o ancestral comum a todos os nós - árvore sem raiz
 - Especifica somente as relações entre os taxa e não define a via evolutiva

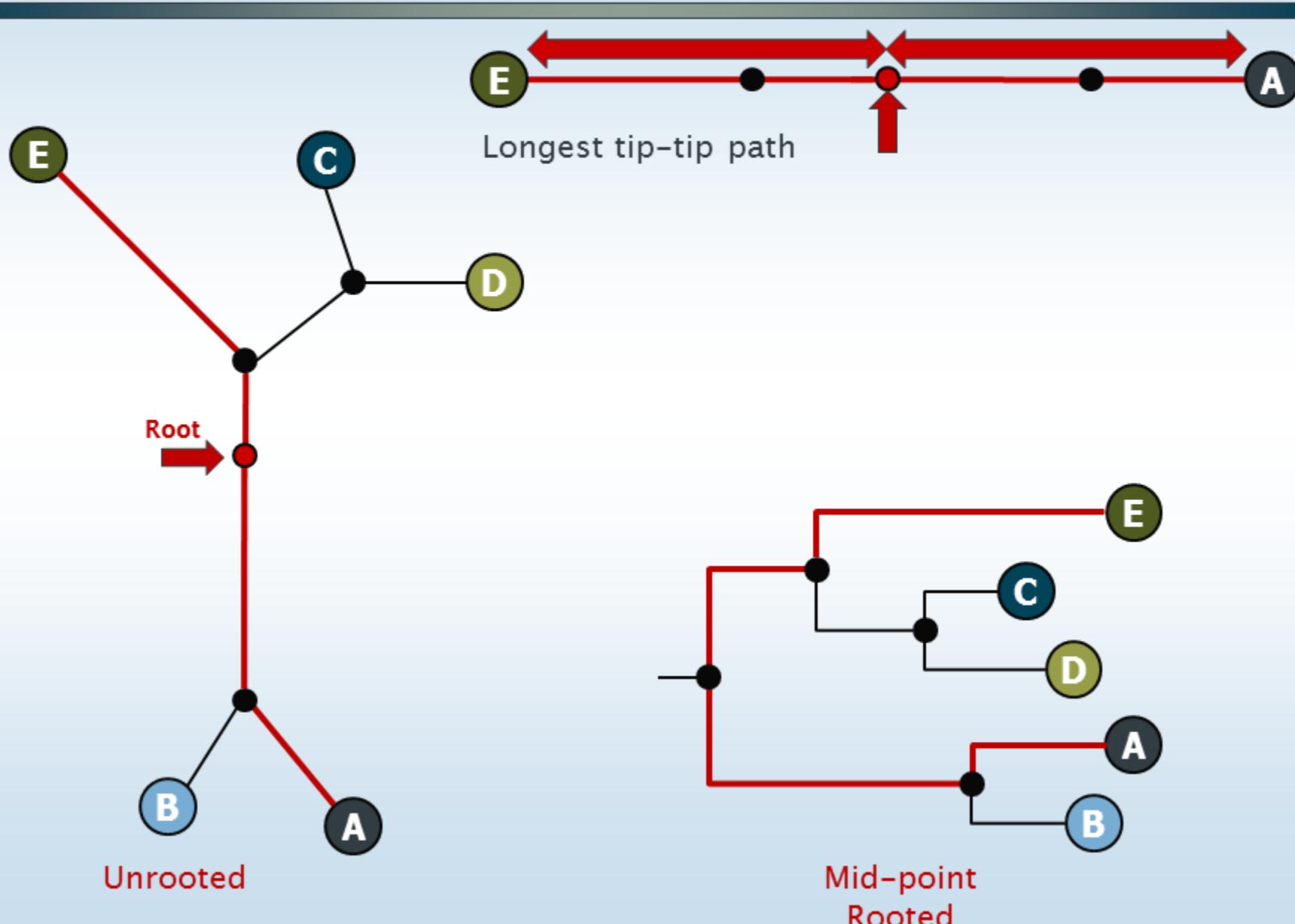
Define relações entre os
taxa



Define via evolutiva

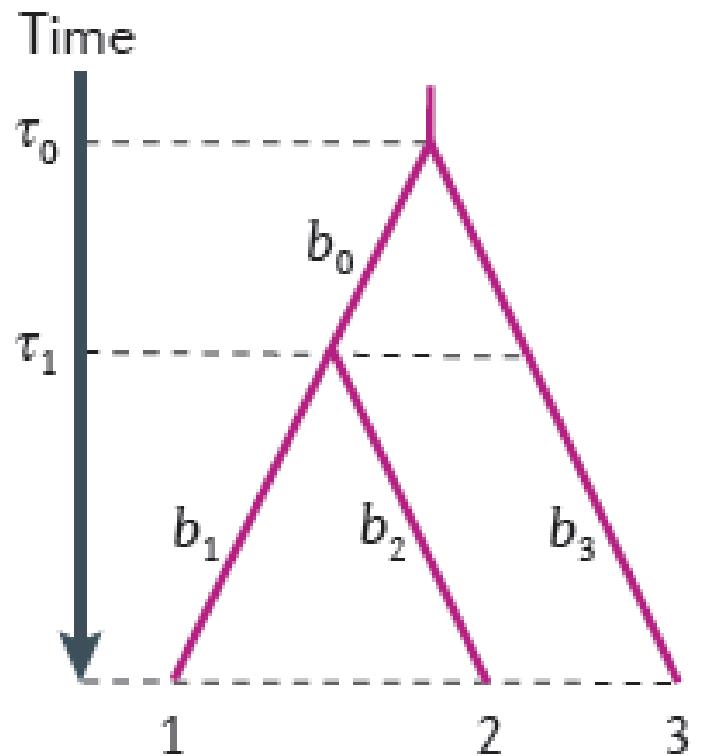


Mid-point Rooting

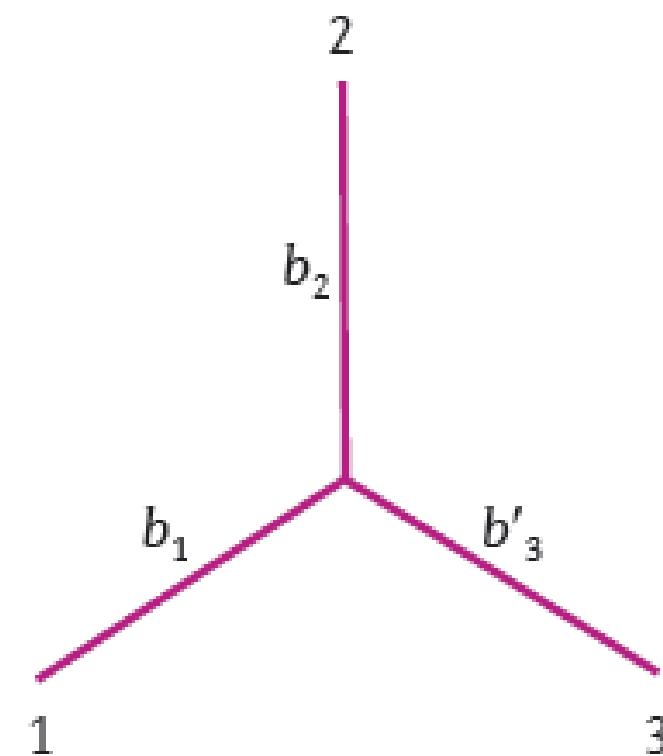


Árvore Filogenética

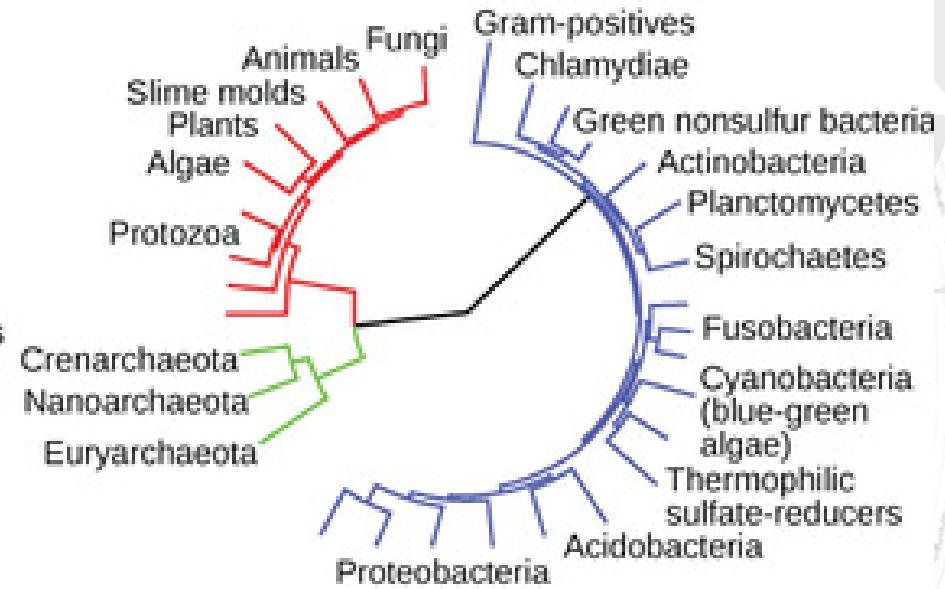
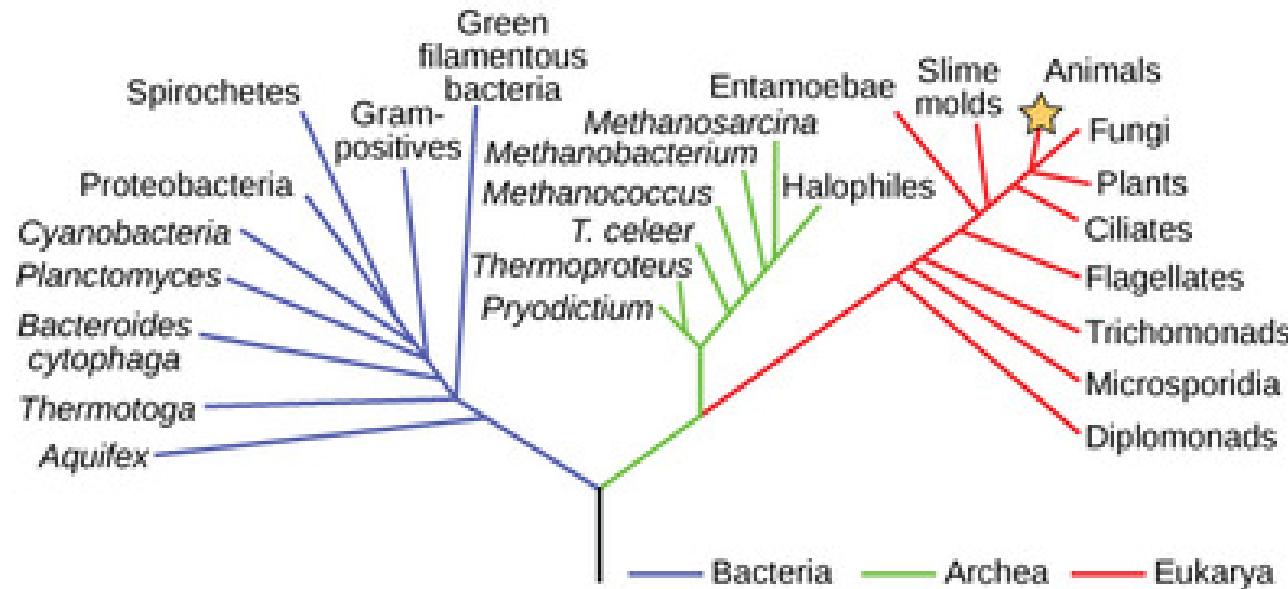
a Rooted tree



b Unrooted tree



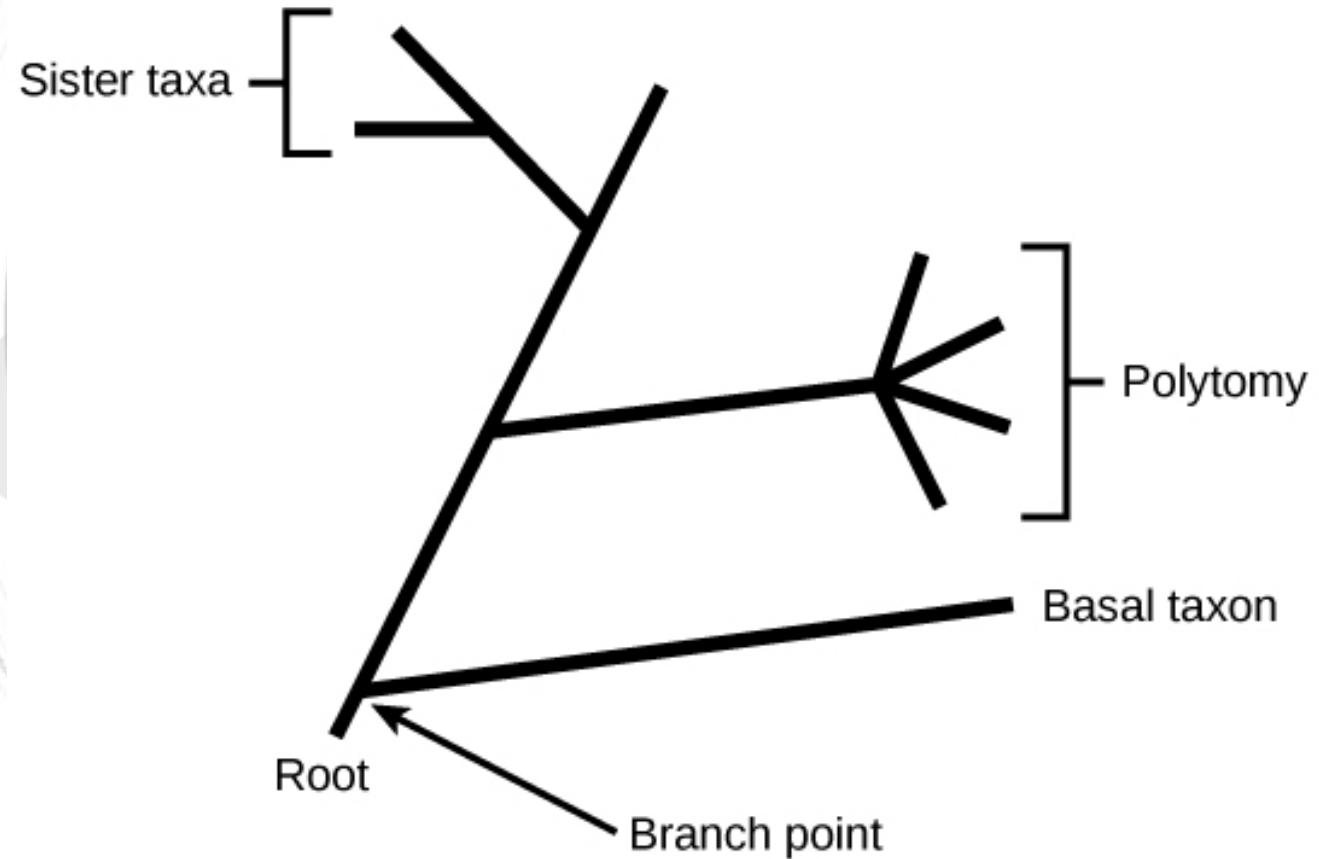
Árvore Filogenética



Ambas as árvores filogenéticas mostram a relação dos três domínios da vida (Bacteria, Archaea e Eukarya), mas a (a) árvore enraizada tenta identificar quando várias espécies divergiram de um ancestral comum, enquanto a (b) árvore não enraizada não mostra isso

Árvore Filogenética

- A **raiz** de uma árvore filogenética indica que uma linhagem ancestral deu origem a todos os organismos da árvore
- Um **ponto de ramificação** indica onde duas linhagens divergiram
- Uma linhagem que evoluiu cedo e permanece não ramificada é um **táxon basal** – sem ramificação
- Quando duas linhagens derivam do mesmo ponto de ramificação, elas são **táxons irmãos**
- Um ramo com mais de duas linhagens é uma **politomia** – relação evolutiva não resolvida



Que sequências usar?

- **Filogenia:** Barcoding
- **Filogenomas:** Geralmente sequências de aminoácidos de genes ortólogos de cópia única

Que genes usar?

- O que usar para procariotos?
 - Para gêneros: gene 16S do rRNA
 - Para espécies: Sequenciamento parcial de genes house-keeping:
 - Exemplos: genes *atpD* (ATP synthase F1, beta subunit), *gyrB* (DNA gyrase B subunit), *rpoB* (RNA polymerase beta subunit), *recA* (recombinase A) and *trpB* (tryptophan synthetase, beta subunit).

Que genes usar?

- O que usar para eucariotos?
- Para fungos:
 - Para gêneros: LSU (28S do rRNA)
 - ITS1-5.8S-ITS2 do rDNA, Segmento α1 do Fator de Elongação da Tradução; Sequencias parciais dos genes: β-tubulina, GPDH, Histona, Actina, rpb2 e Mating Type
- Para trypanosomas:
- glucose-6-phosphate isomerase (GPI), 18S ribosomal RNA (SSU rRNA), glycosomal GAPDH (gGAPDH), DNA Mitocondrial

SEQUENCIAS DE DNA PARA FILOGENIA DE LEVEDURAS - YeastIP

Marker	Description	Position in the gene*	Length (bp)
LSU	Complete sequence of the large subunit 26S ribosomal RNA gene	65–3364	3300
D1/D2 LSU	Partial sequence of the 26S ribosomal gene comprising the D1/D2 region	65–636	570
SSU	Complete sequence of the small subunit 18S ribosomal RNA gene	1–1800	1700
ITS	Ribosomal RNA region containing the intergenic region 1 (between 18S and 5.8S), the 5.8S ribosomal RNA gene and the intergenic region 2 (between 5.8S and 26S)	First base of ITS1 to last base of ITS2	400–600
mtSSU	Mitochondrial small subunit 15S ribosomal RNA gene	383–1006	400–600
<i>RPB1</i>	Partial sequence of the RNA polymerase II largest subunit coding gene	253–873	620
<i>RPB2</i>	Partial sequence of the RNA polymerase II second largest subunit coding gene	1645–2319	680–1000
<i>TEF1-alpha</i>	Partial sequence of the translation elongation factor 1-alpha coding gene	64–1190	930
<i>ACT1</i>	Partial sequence of the exon2 of the actin coding gene	405–1383	980 [†]
mtCOX II	Partial sequence of mitochondrial cytochrome C oxidase subunit 2 coding gene	121–707	590

(A)



Wings

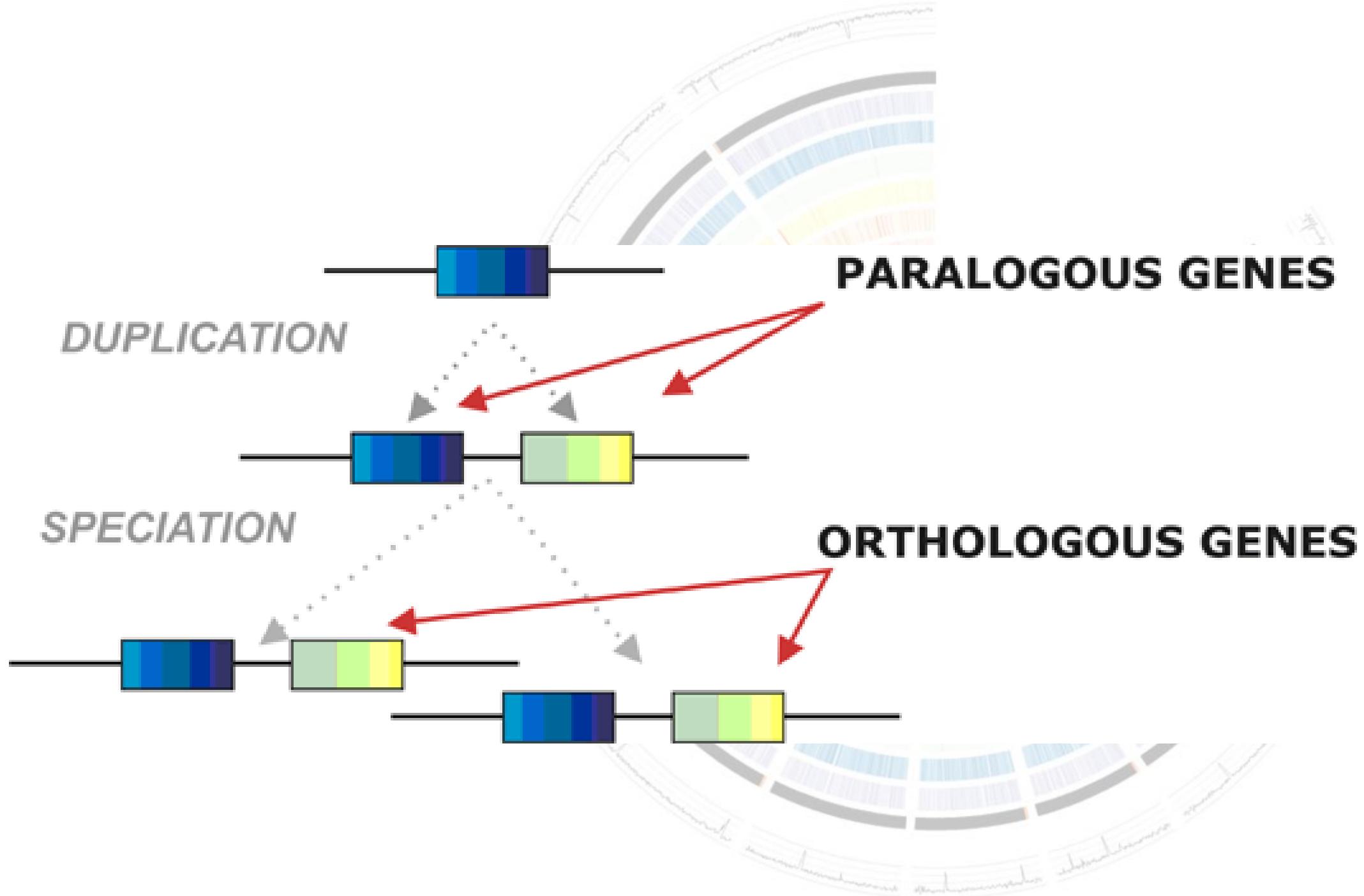


No wings



Wings





DNA BARCODING

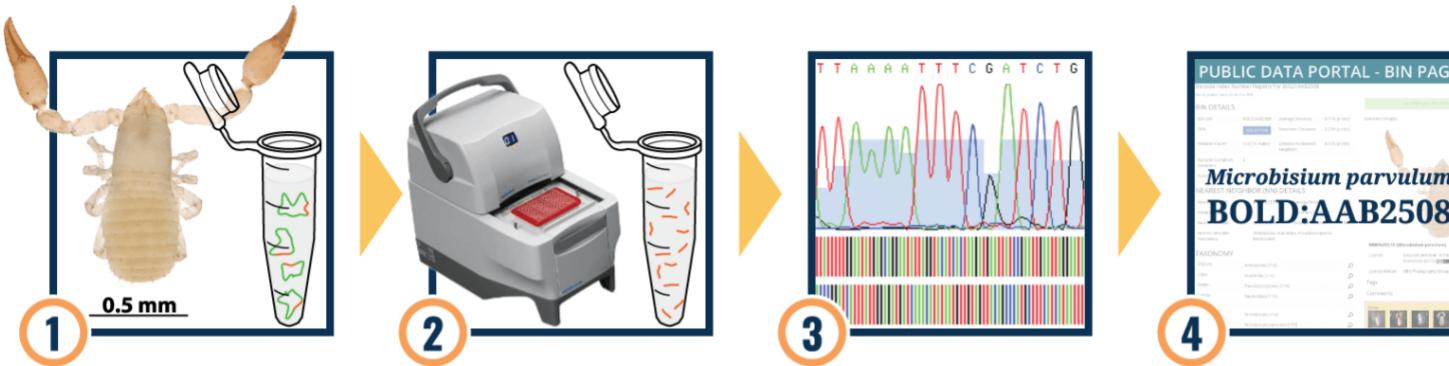
A TOOL FOR SPECIMEN IDENTIFICATION AND SPECIES DISCOVERY

The screenshot shows the homepage of the International Barcode of Life (IBOL) website. At the top left is the IBOL logo, which features a stylized green DNA double helix and the text "international BARCODE OF LIFE". To the right is a navigation bar with links for "HOME", "ABOUT", "PROGRAMS", "NEWS & MEDIA", and a prominent green "JOIN" button. Below the navigation bar is a large, semi-transparent circular graphic composed of many small, colored DNA barcode fragments. On the left side of the main content area, there is a vertical column of social media icons for Facebook ("f"), Twitter ("t"), and LinkedIn ("in"). The main content area has a dark blue-to-green gradient background and contains the text "DNA BARCODING" in large white letters, followed by "A TOOL FOR SPECIMEN IDENTIFICATION AND SPECIES DISCOVERY" in smaller white letters.

<https://ibol.org/about/dna-barcoding/>

DNA BARCODING

Como funciona?



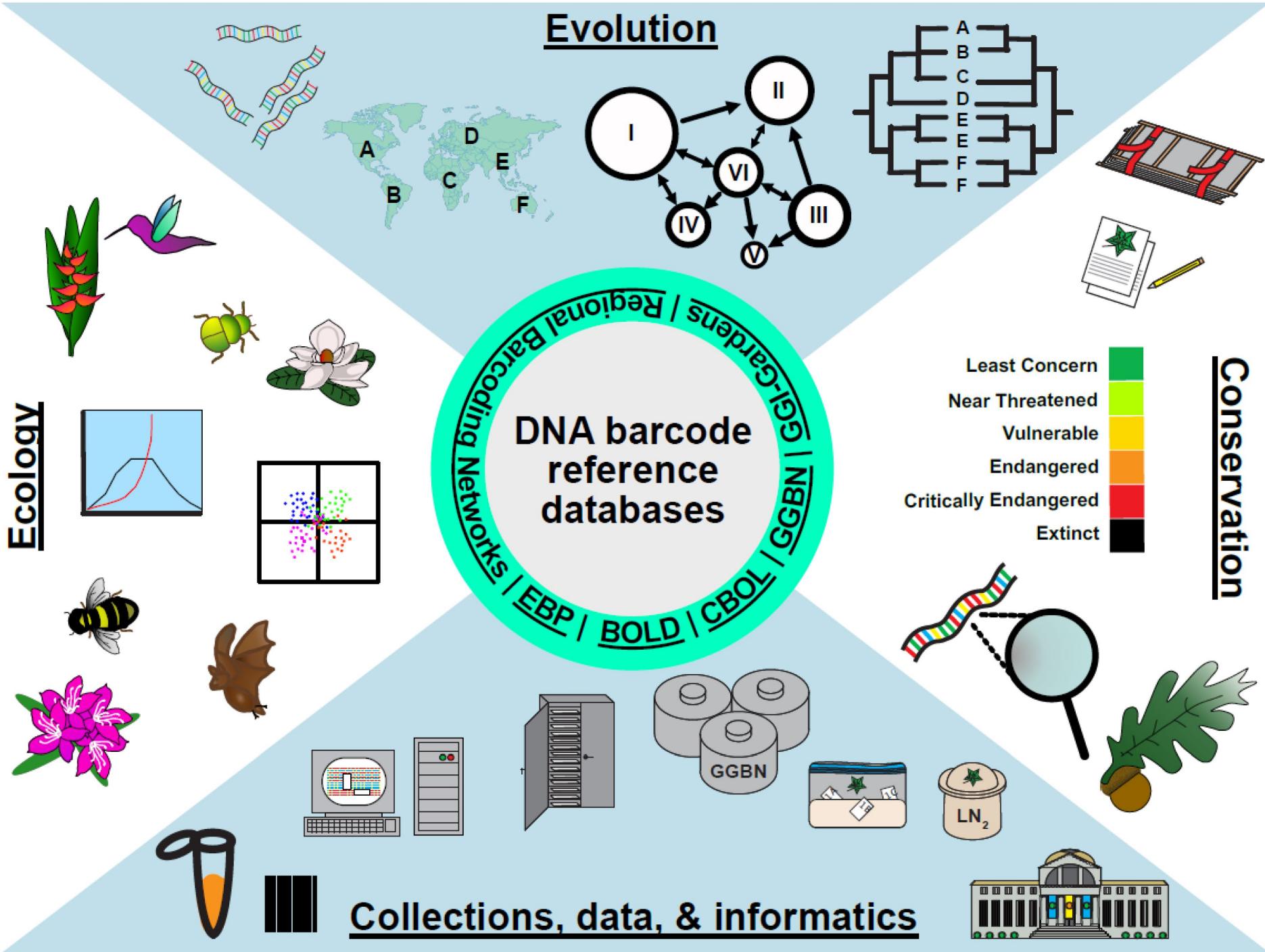
Step 1: Isolate DNA from the sample

Step 2: Amplify the target DNA barcode region using PCR

Step 3: Sequence the PCR products

Step 4: Compare the resulting sequences against reference databases to find the matching species

<https://ibol.org/about/dna-barcoding/>



THE DNA BARCODE



international
BARCODE
OF LIFE

HOME ABOUT ▾ PROGRAMS ▾ NEWS & MEDIA ▾ JOIN

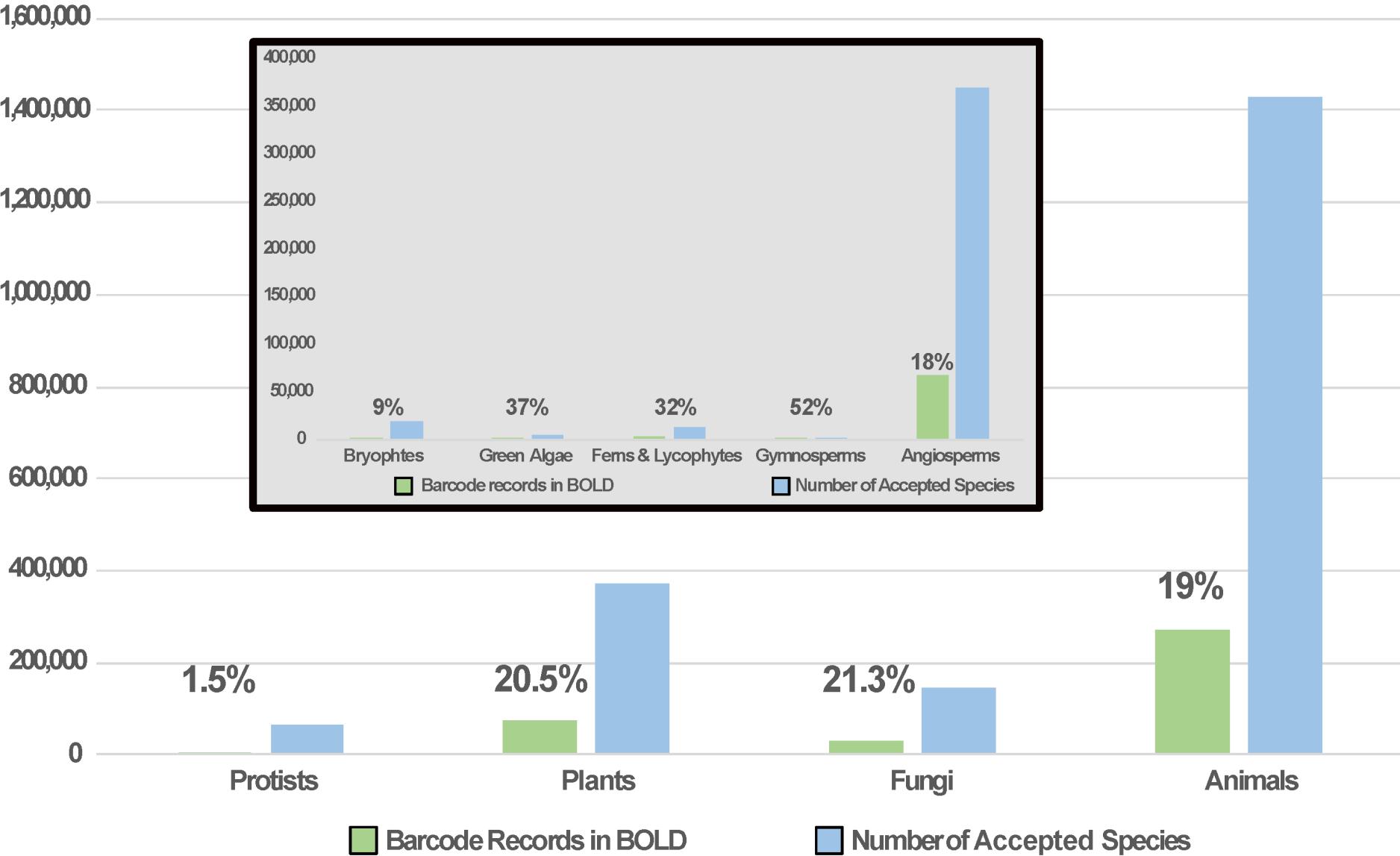
f
t
in



Plant barcoding studies use one or a few plastid regions (e.g. *rbcL* and *matK*, and the non-coding spacer *trnH-psbA*) and the internal transcribed spacer (ITS) region of nuclear ribosomal DNA.

Animal barcoding studies use a region in the mitochondrial cytochrome *c* oxidase 1 gene ("CO1").

Fungal barcoding studies use the internal transcribed spacer (ITS) region in the nuclear ribosomal cistron. This region shows reasonable discriminatory power at the species level in many groups.



DNA BARCODING



HOME ABOUT ▾ PROGRAMS ▾ NEWS & MEDIA ▾

JOIN

THE LIBRARY

Barcode sequences are placed in the Barcode of Life Data Systems (BOLD) database – an online workbench that includes a reference library of DNA barcodes that can be used to assign identities to sequences of unknown origin.



BOLD is a searchable repository for barcode records, storing specimen data and images as well as sequences and trace files. It provides an identification engine based on the current barcode library and monitors the number of barcode sequence records and species coverage.



The screenshot shows the homepage of the BOLD Systems website. At the top, there's a navigation bar with links for HOME, ABOUT ▾, PROGRAMS ▾, NEWS & MEDIA ▾, and JOIN. Below the navigation is a large banner with a world map and the text "BARCODE OF LIFE DATA SYSTEM v4". The banner also includes the tagline "Advancing biodiversity science through DNA-based species identification." and a "EXPLORE THE DATA" button. The main content area is titled "DESIGNED TO SUPPORT THE GENERATION & APPLICATION OF DNA BARCODE DATA". It describes BOLD as a cloud-based data storage and analysis platform developed at the Centre for Biodiversity Genomics in Canada. It consists of four main modules: a data portal, an educational portal, a registry of BOLD putative species, and a data collection and analysis workbench. A note states that this version is beta and contains bugs, with users encouraged to report them. Below this, there are four circular icons with labels: DATA PORTAL, EDUCATION PORTAL, BIN DATABASE, and WORKBENCH. Each icon has a brief description underneath.

DATA PORTAL
A data retrieval interface that allows for searching over 1.7M public records in BOLD using multiple search criteria including, but not limited to, geography, taxonomy, and depository.

EDUCATION PORTAL
A custom platform for educators and students to explore barcode data and contribute novel barcodes to the BOLD database.

BIN DATABASE
A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.

WORKBENCH
A data collection and analysis environment that supports the assembly and validation of DNA barcodes and other sequences.

DNA BARCODING

Fungos: nuclear internal transcribed spacer (ITS) do RNA ribossomal – Barcoding primário

Fusarium* e *Trichoderma necessidade de translation elongation factor 1 alpha (TEF-1 α) como barcode secundário

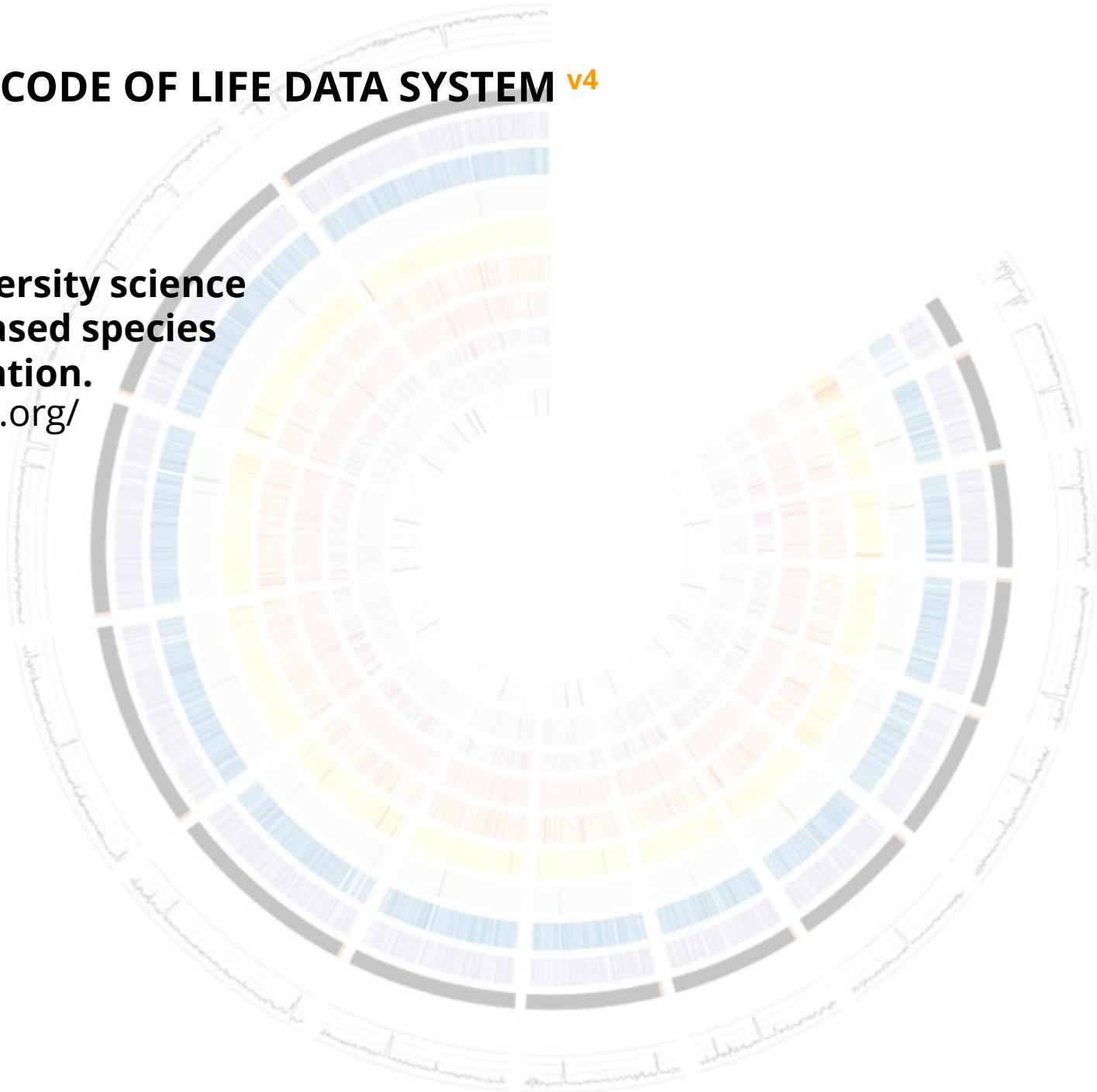
Arbuscular mycorrhizal (AM) and in rust fungi: small subunit (nuSSU) and the large subunit (nuLSU) do RNA ribossomal RNA

Oomycota: mitochondrial cytochrome oxidase c subunits (COX1 and COX2).

BARCODE OF LIFE DATA SYSTEM ^{v4}

**Advancing biodiversity science
through DNA-based species
identification.**

<https://www.boldsystems.org/>



QBOL – DNA Barcoding para pragas quarentenárias vegetais



Development of a new diagnostic tool using DNA barcoding to identify quarantine organisms in support of plant health

DNA Barcoding

Activities

Q-lists

Databases

Collaboration

Partners



Continue to

[About QBOL](#)

[e-Newsletter](#)

[Training courses](#)

[Meeting 2012](#)

Centralizado na Europa
<https://www.qbol.org/en/qbol.htm>



DNA Barcoding

Activities

Q-lists

Databases

C



Q-BOL Activities

QBOL Activities:

- Barcoding of Arthropods
- Barcoding of Bacteria
- Barcoding of Fungi
- Barcoding of Nematodes
- Barcoding of Phytoplasmas
- Barcoding of Viruses
- DNA Banks
- Library/Database/Information
- Validation/Evaluation
- Dissemination

DNA Barcoding de fungos – CBS na Holanda



DNA Barcoding

Activities

Q-lists

Databases

Collaboration

Partners



Barcodeing of Fungi

In this work package a short list of 19 Q-species were selected for barcodeing.

Contact

Ewald Groenewald

[View profile >>](#)

WP Coordinator: KNAW-CBS
(Partner 9)

KNAW-CBS, PRI, FERA and CIP



DNA Barcoding

Activities

Q-lists

Databases

Collaboration

Partners

Barcode of
Arthropods

Barcode of Bacteria

Barcode of Fungi

**Barcode of
Nematodes**

Barcode of
Phytoplasmas

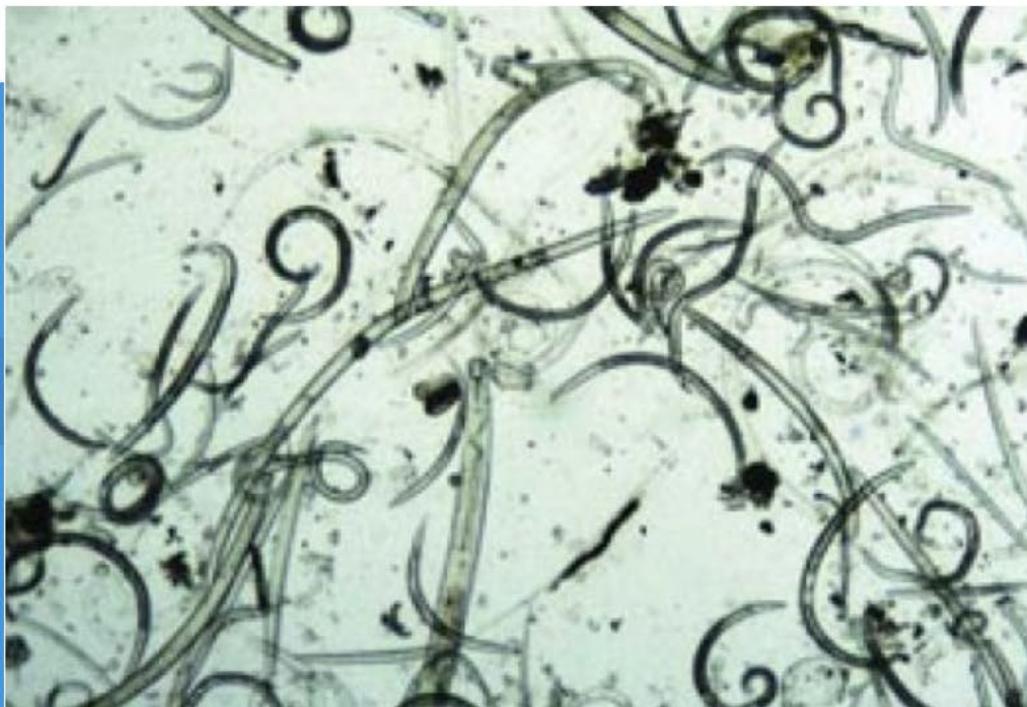
Barcode of Viruses

DNA Banks

Library/Database/Information

Validation/Evaluation

Dissemination



Barcode of Nematodes

In this work package a base list of 32 nematode species was created for which barcodes needed to be collected

Contact

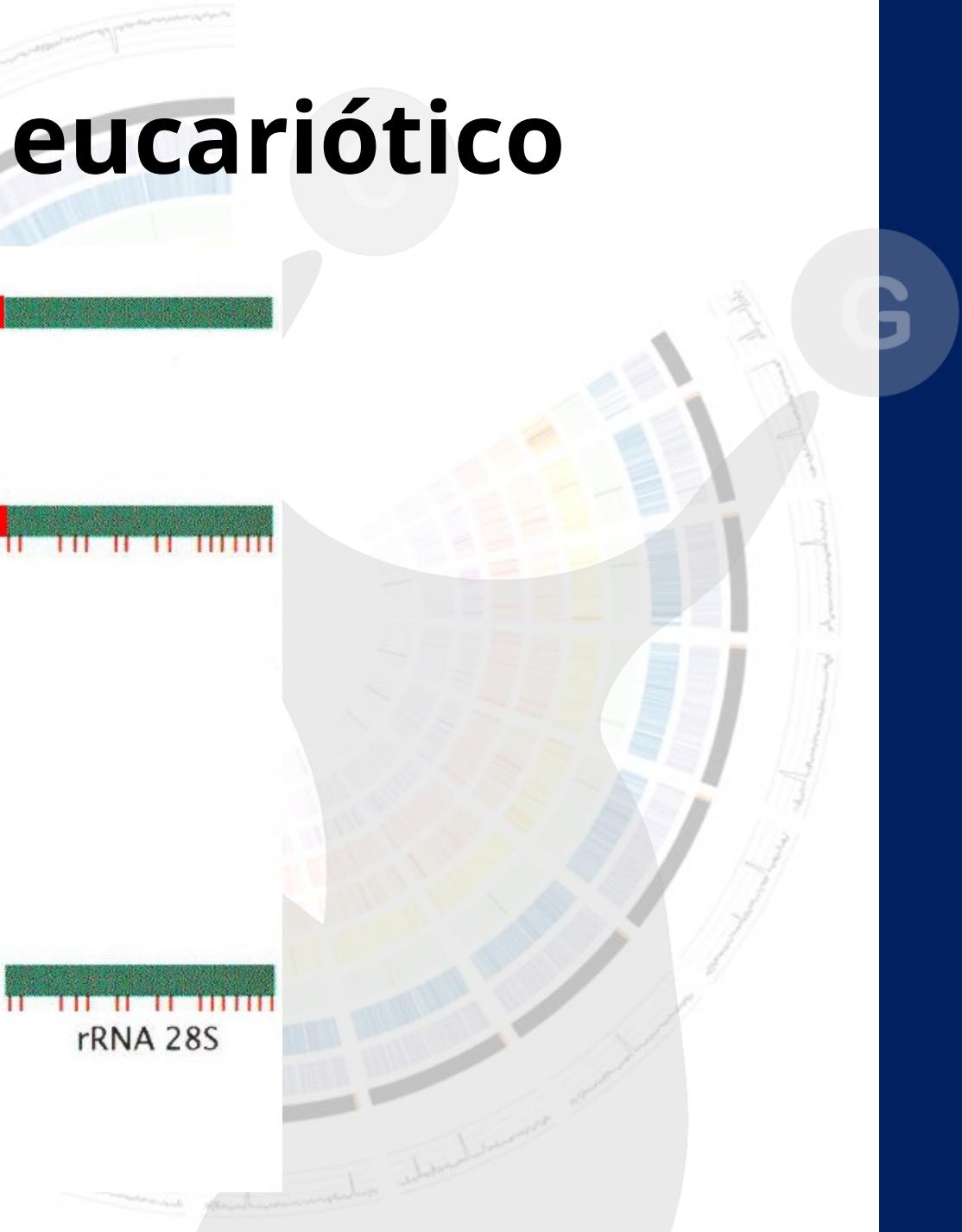
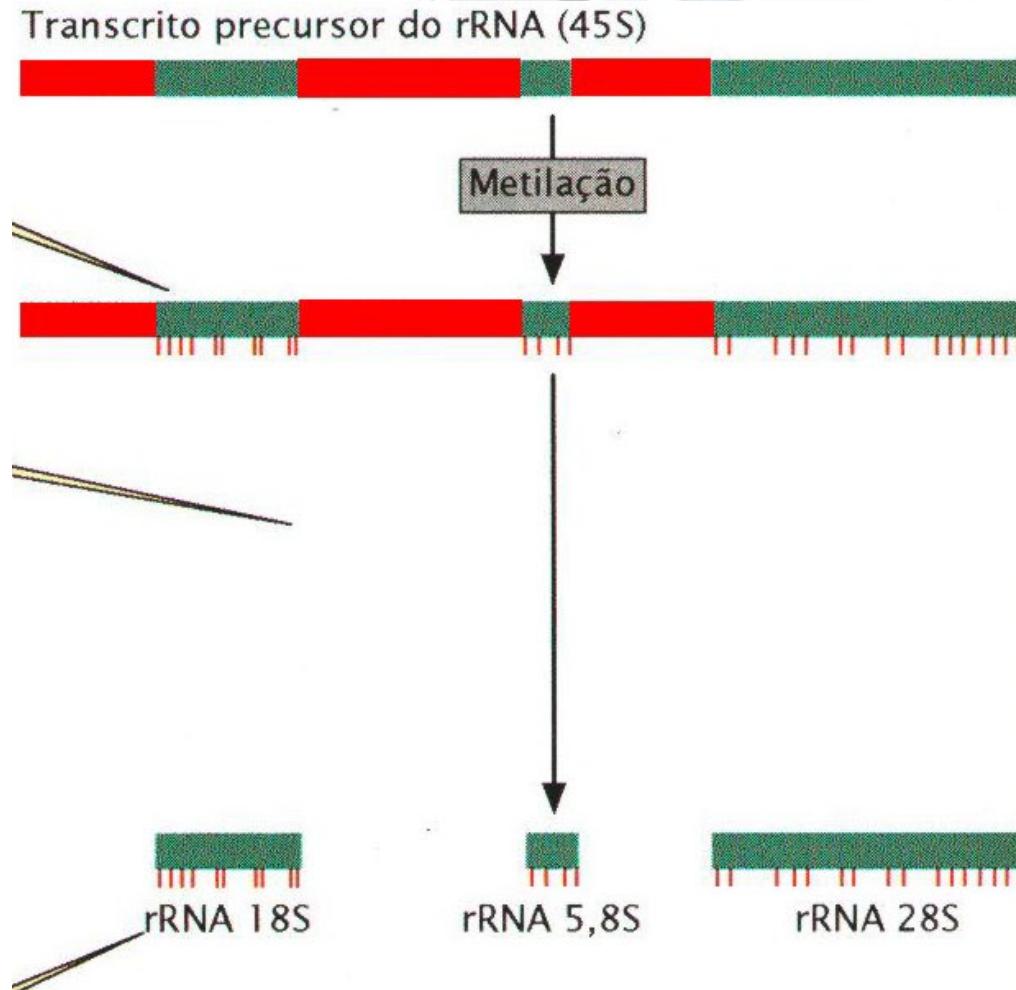
Juerg Frey

[View profile >>](#)

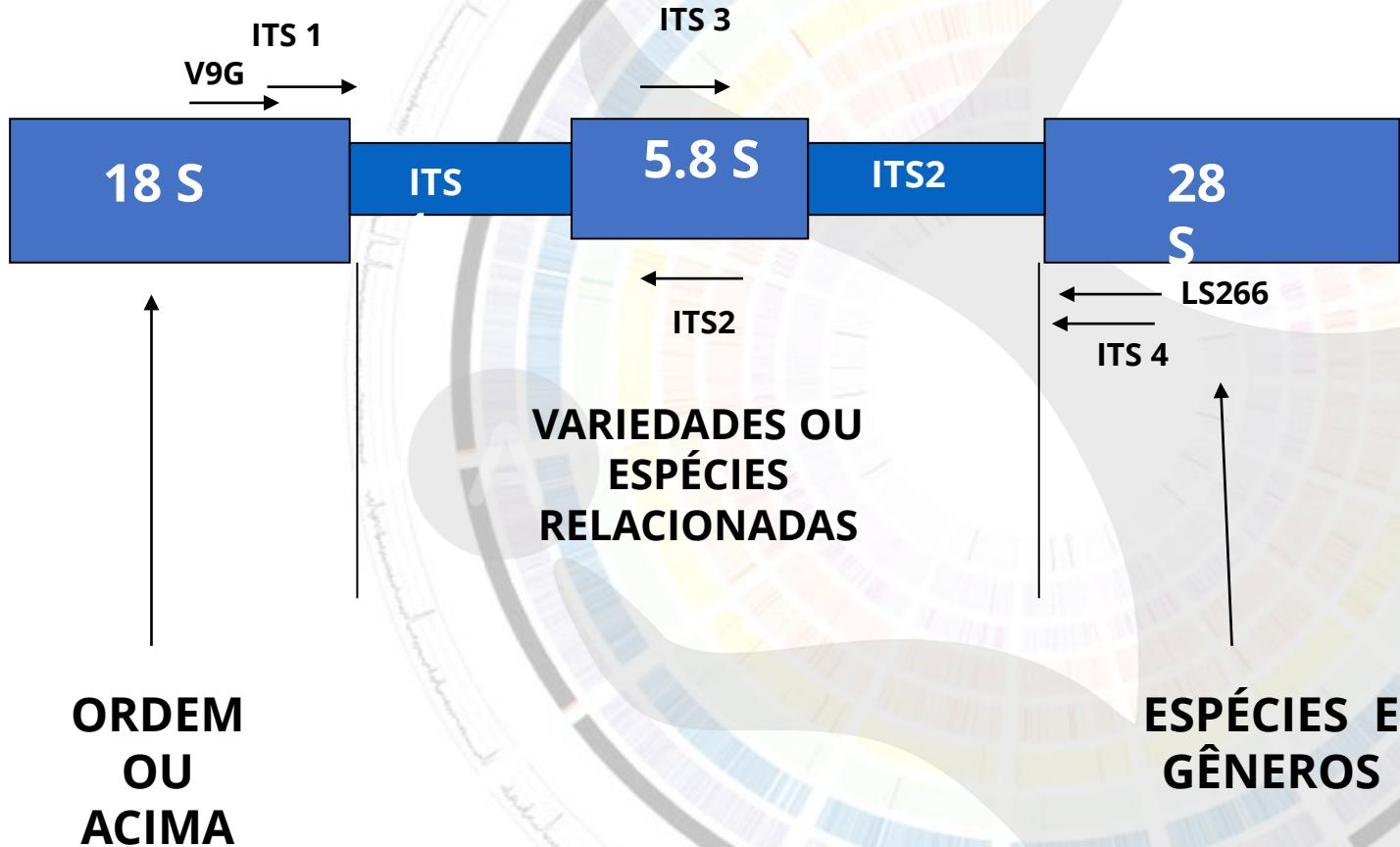
WP Coordinator: ACW (Partner 5)

ACW, PRI, VLAGEW (ILVO).
INRA and CIP

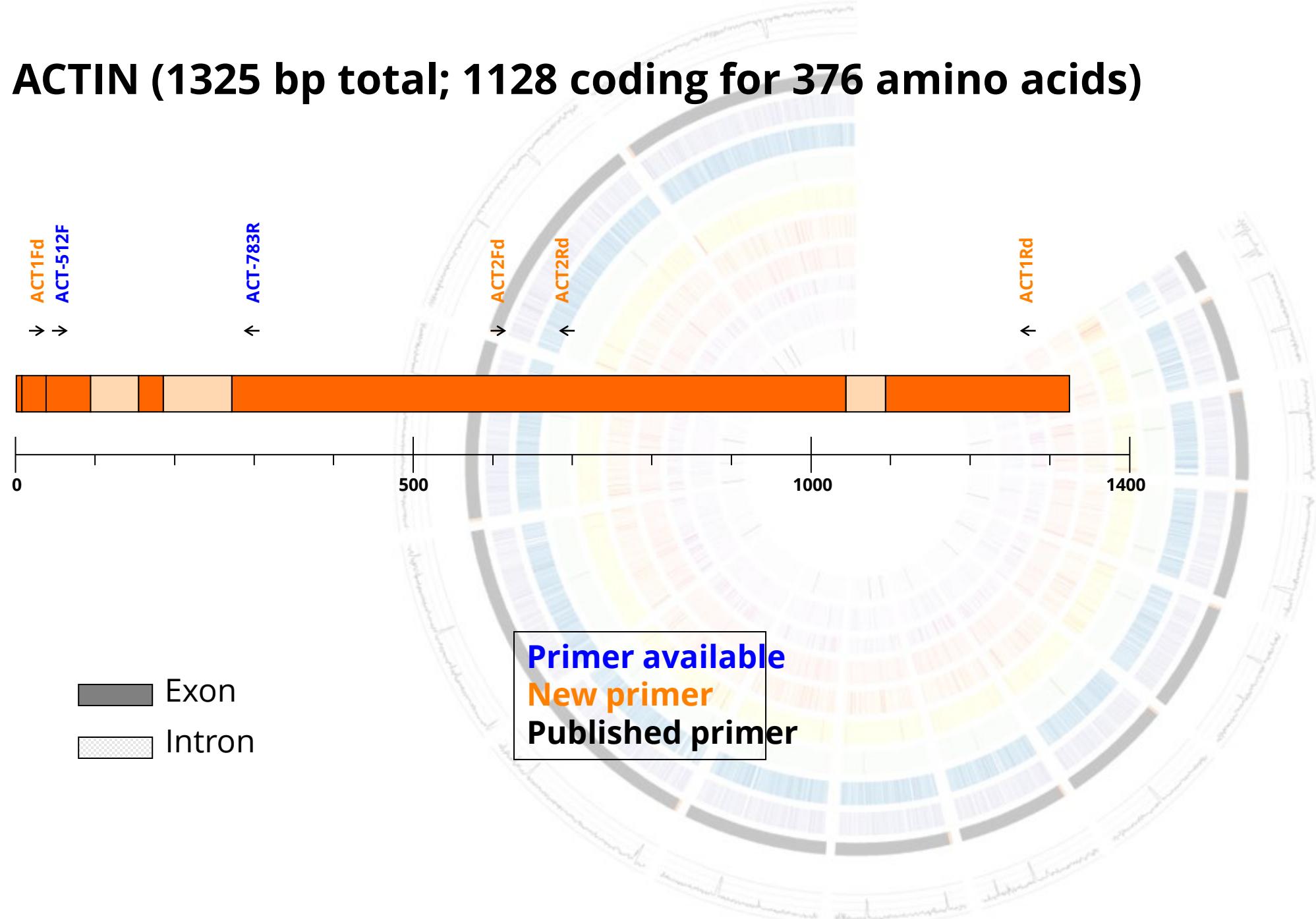
DNA Ribossomal eucariótico



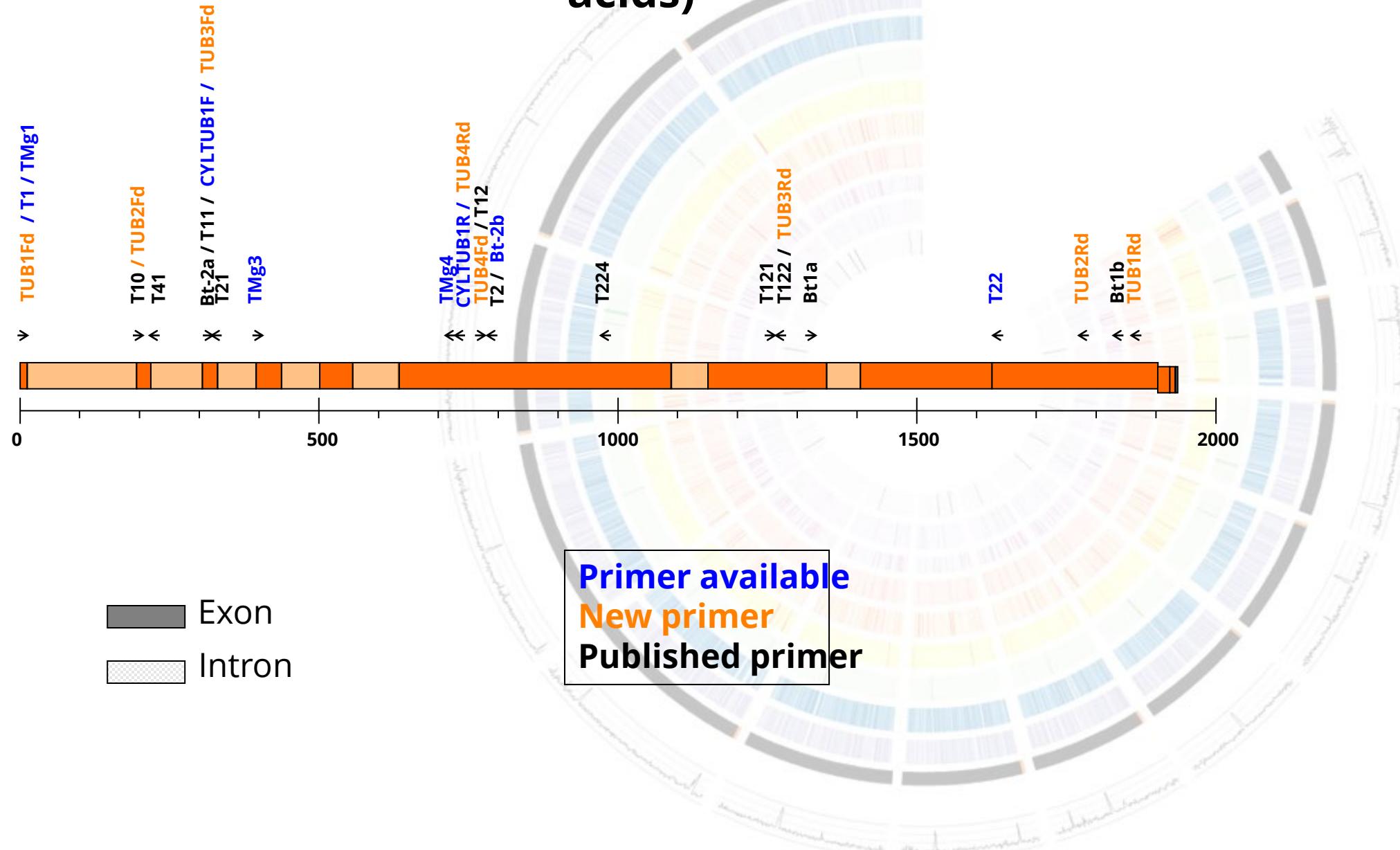
DNA Ribossomal eucariótico - região ITS



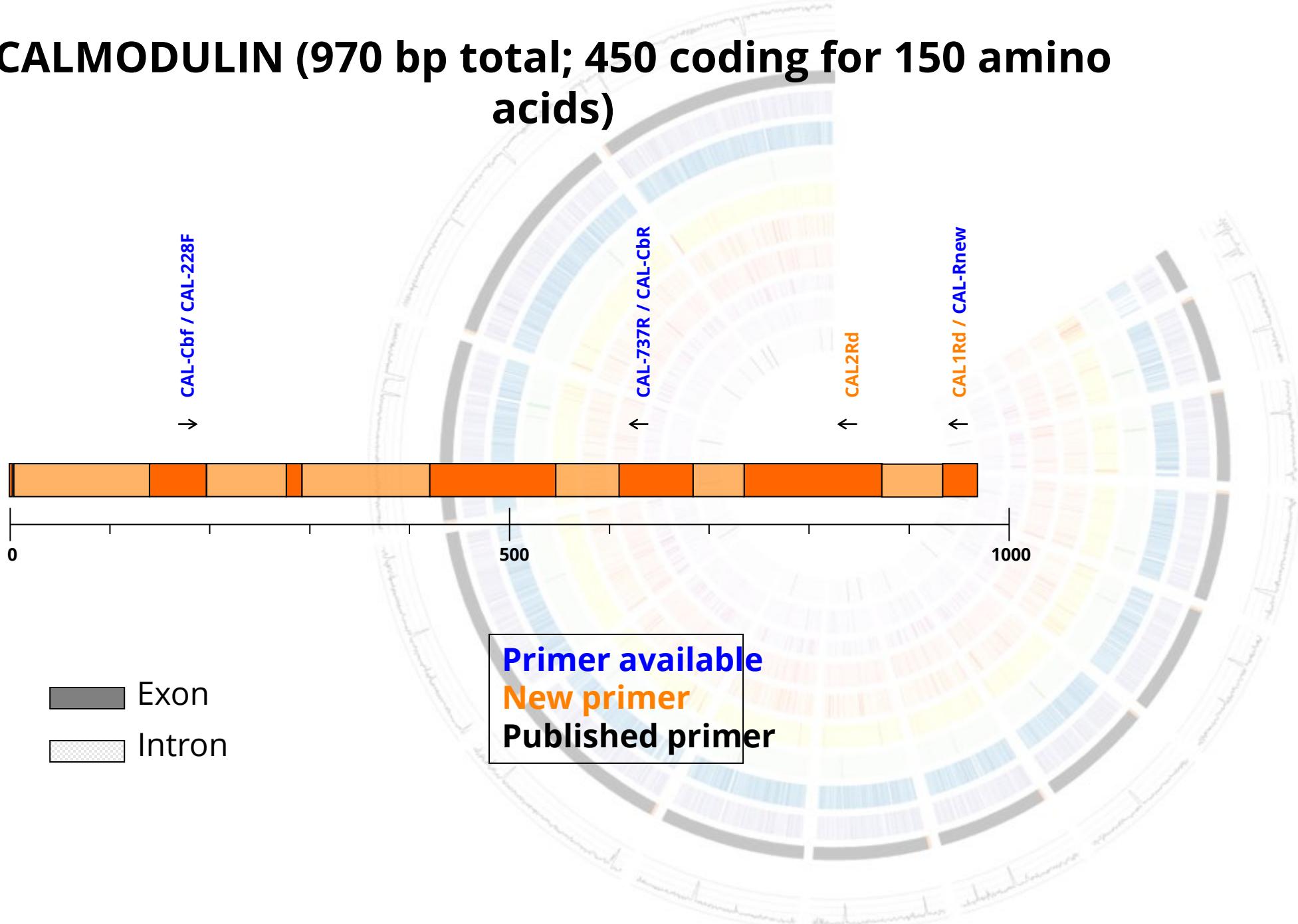
ACTIN (1325 bp total; 1128 coding for 376 amino acids)



BETA-TUBULIN (1933 bp total; 1344 coding for 448 amino acids)

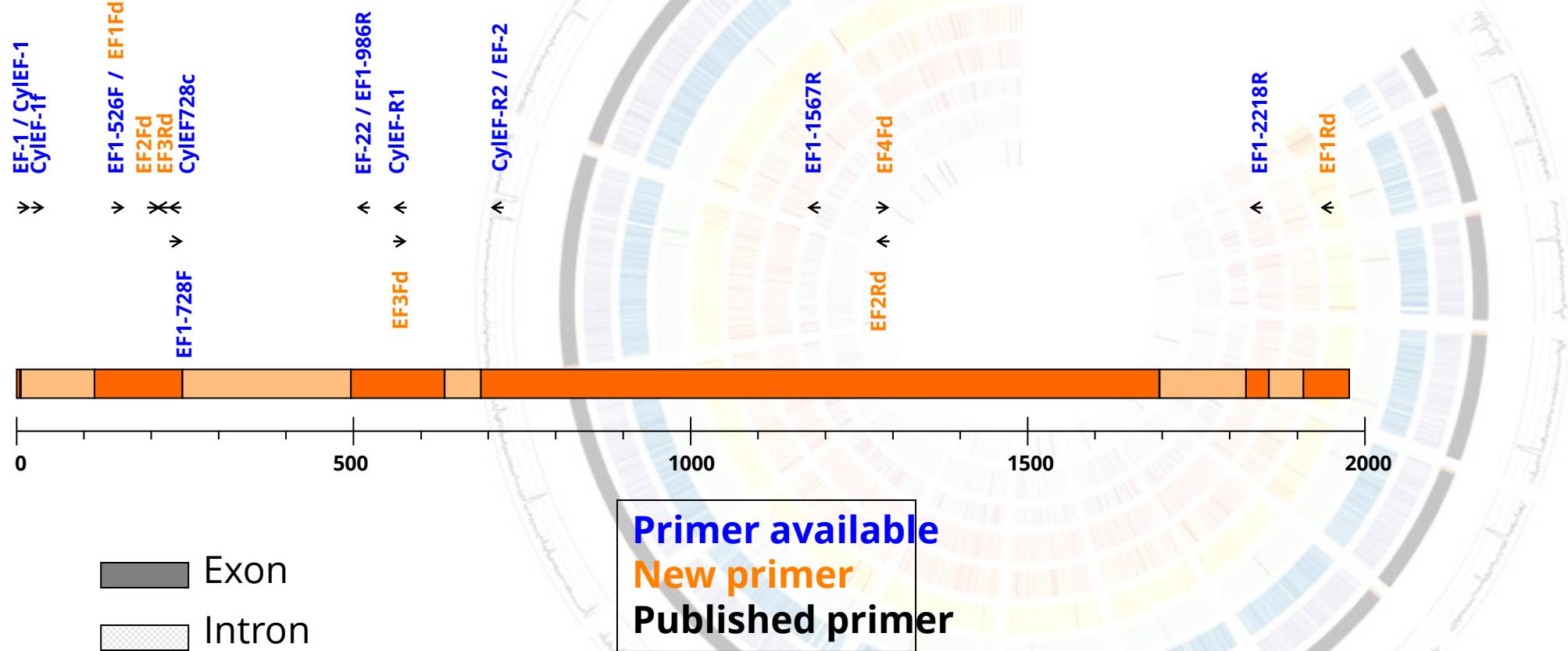


CALMODULIN (970 bp total; 450 coding for 150 amino acids)

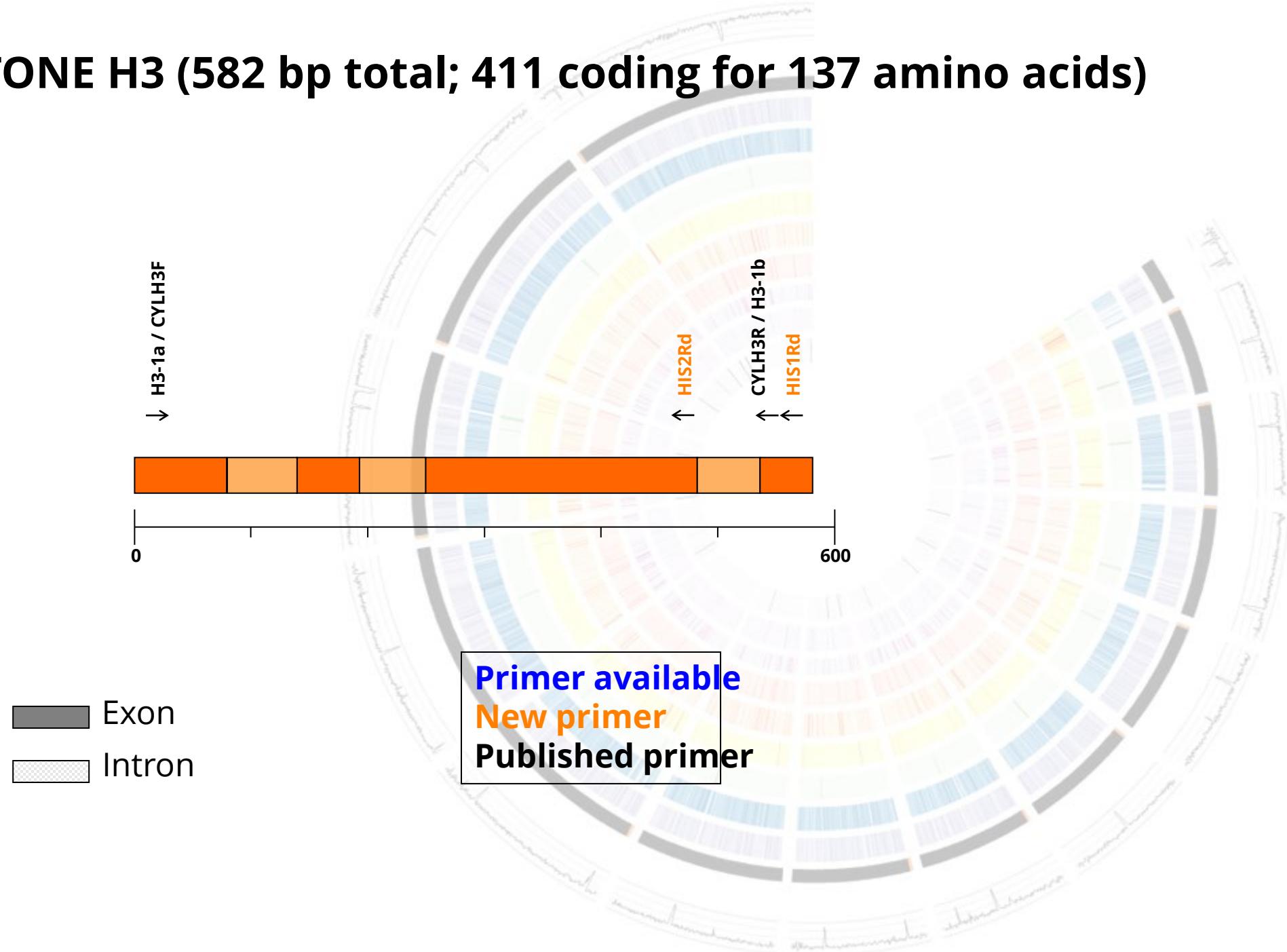


ELONGATION FACTOR 1-ALPHA

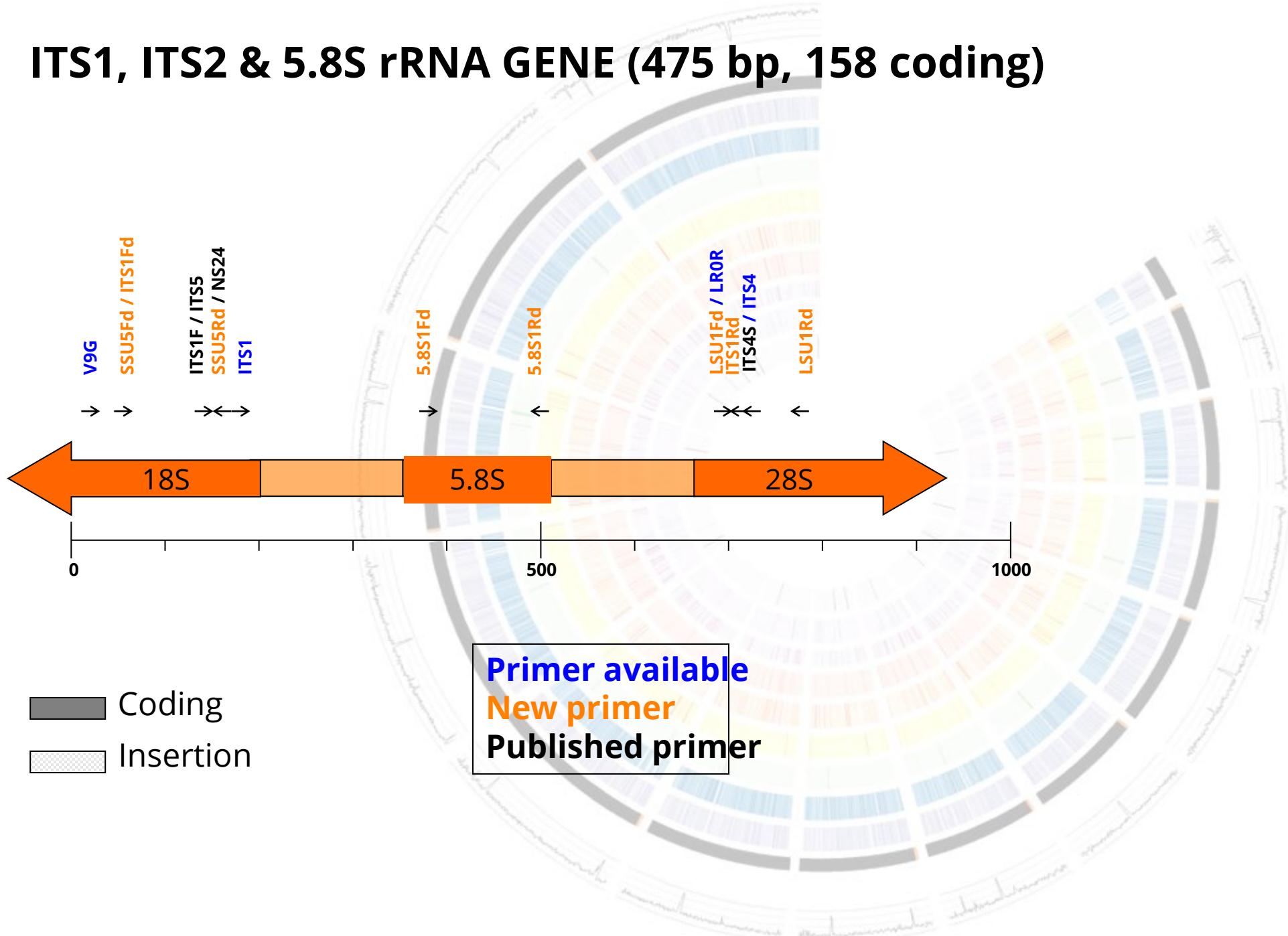
(1975 bp total; 1383 coding for 461 amino acids)



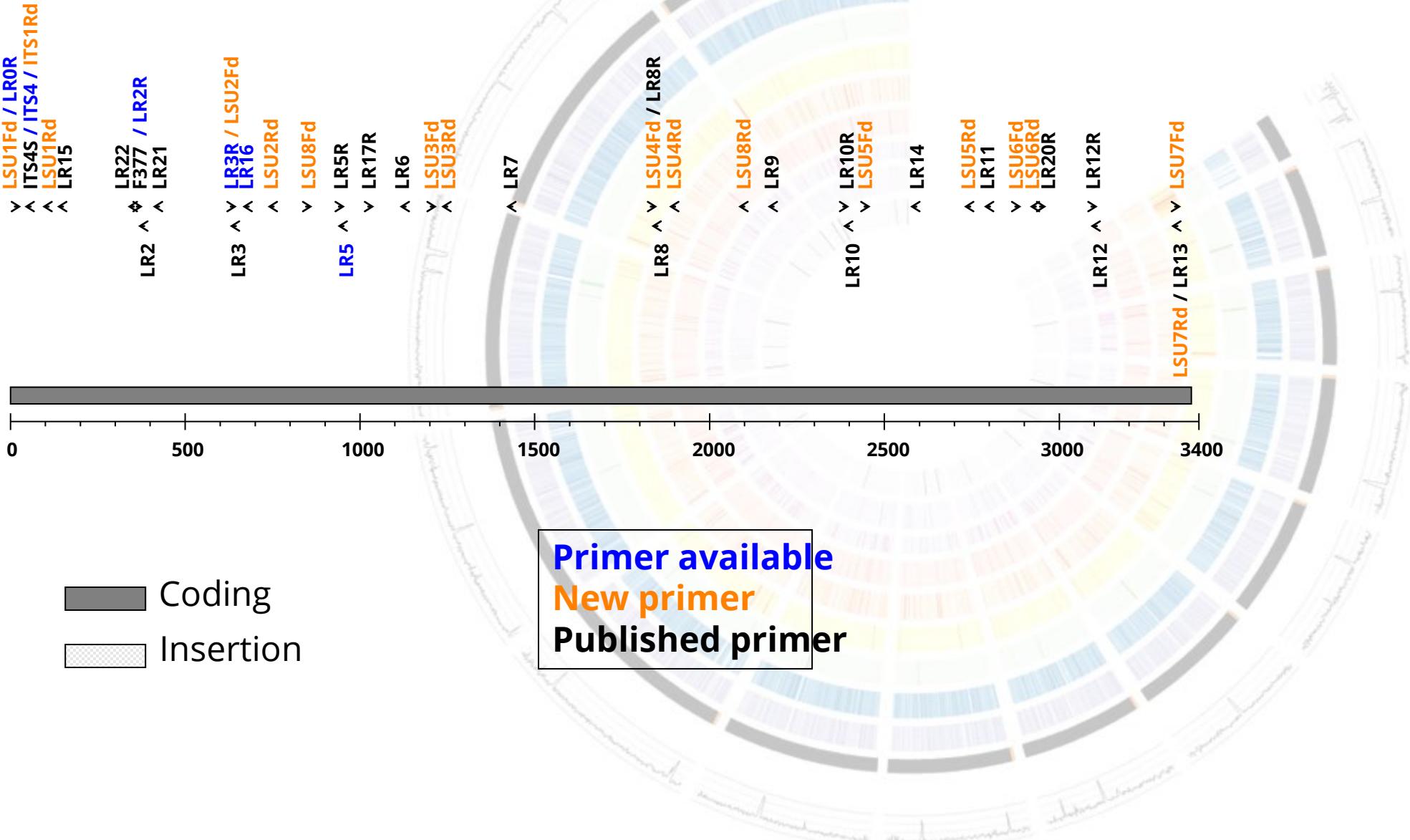
HISTONE H3 (582 bp total; 411 coding for 137 amino acids)



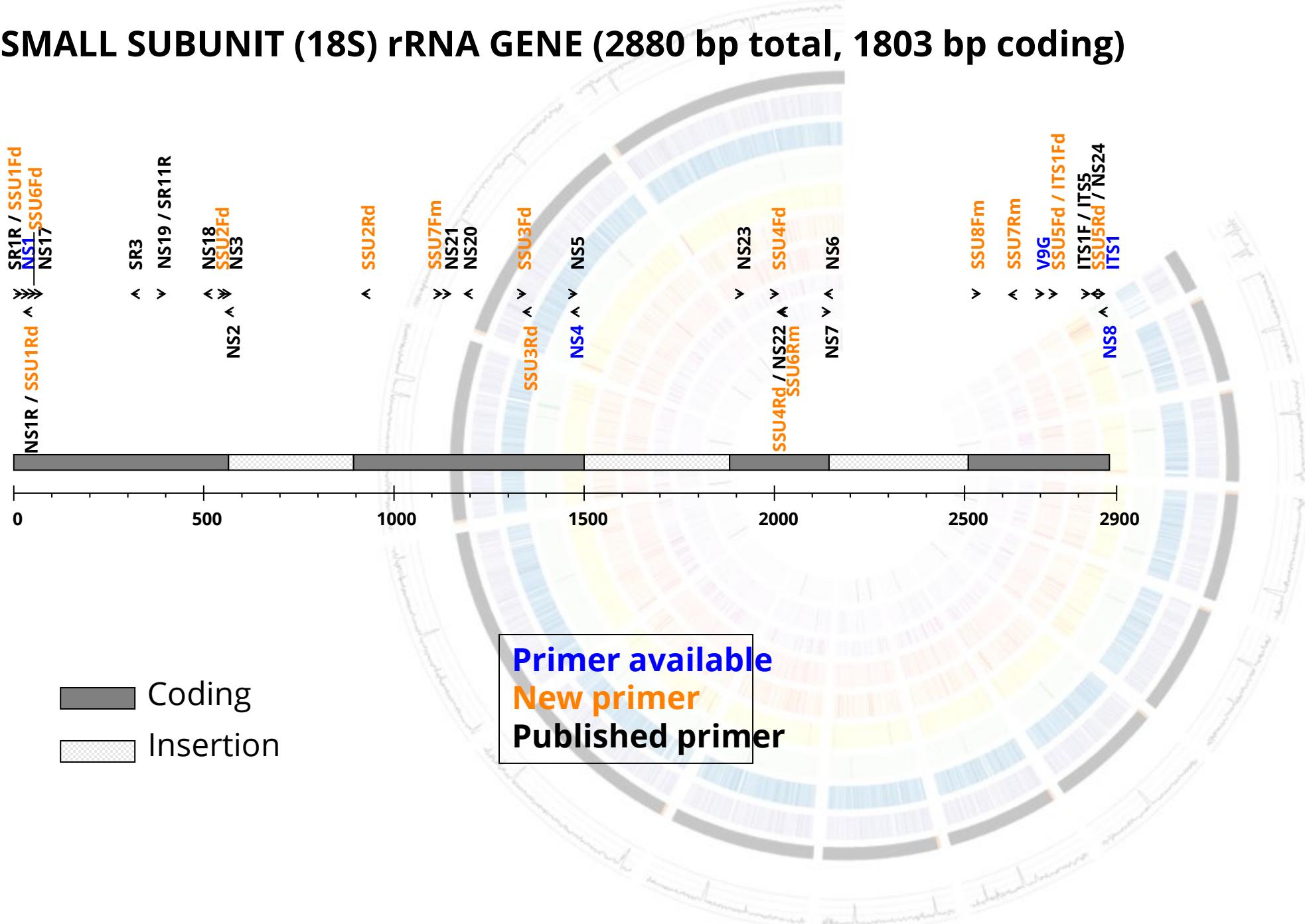
ITS1, ITS2 & 5.8S rRNA GENE (475 bp, 158 coding)



LARGE SUBUNIT (28S) rRNA GENE (3378 bp)



SMALL SUBUNIT (18S) rRNA GENE (2880 bp total, 1803 bp coding)



**Então, qual
gene vou usar?**



**Depende da
pergunta/objetivo/
problema a
resolver**



Problemas

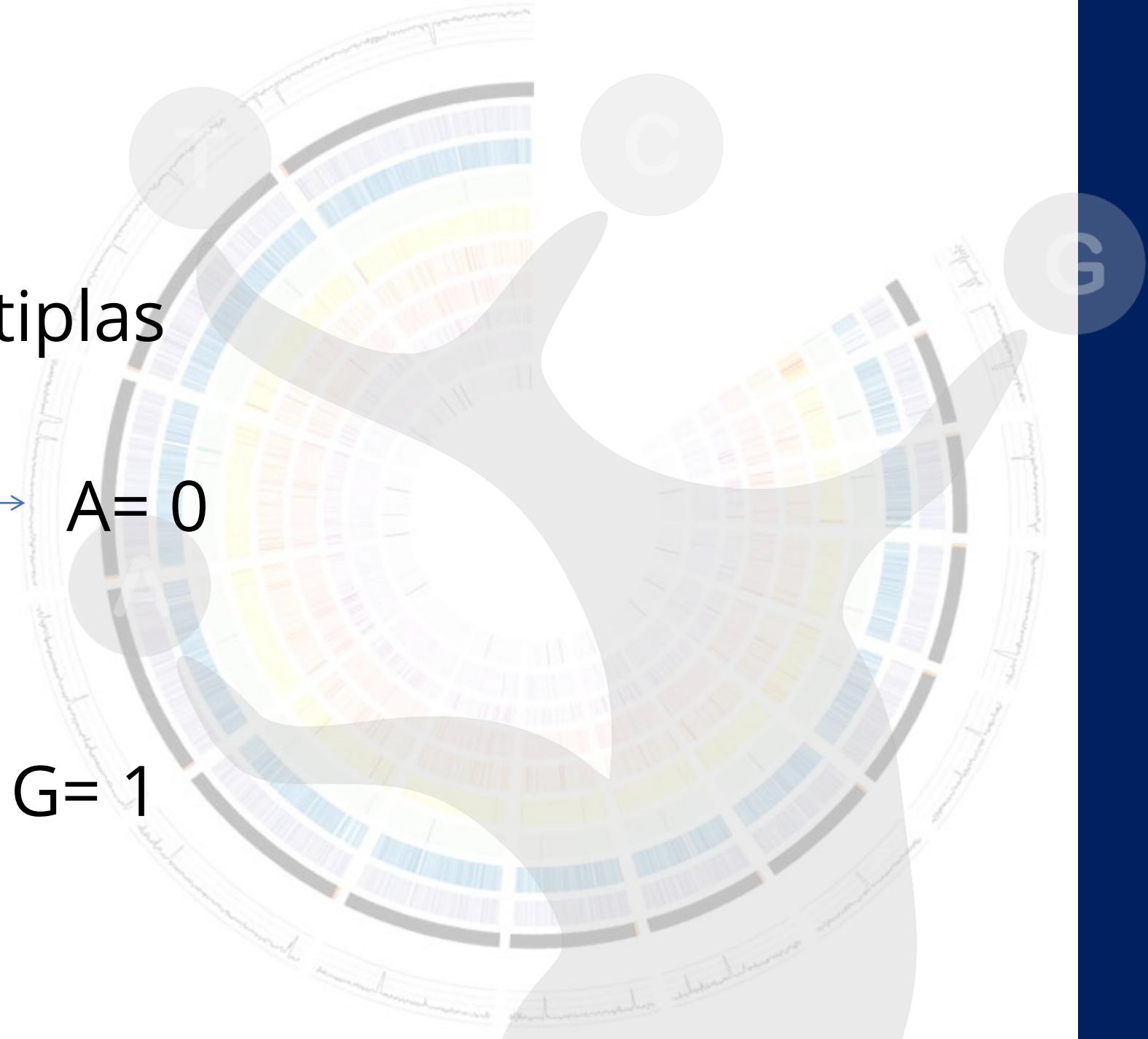
- Substituições múltiplas

- 1) A \longrightarrow A = 0

- 2) A \longrightarrow C \longrightarrow A = 0

- 1) A \longrightarrow A = 0

- 2) A \longrightarrow C \longrightarrow G = 1



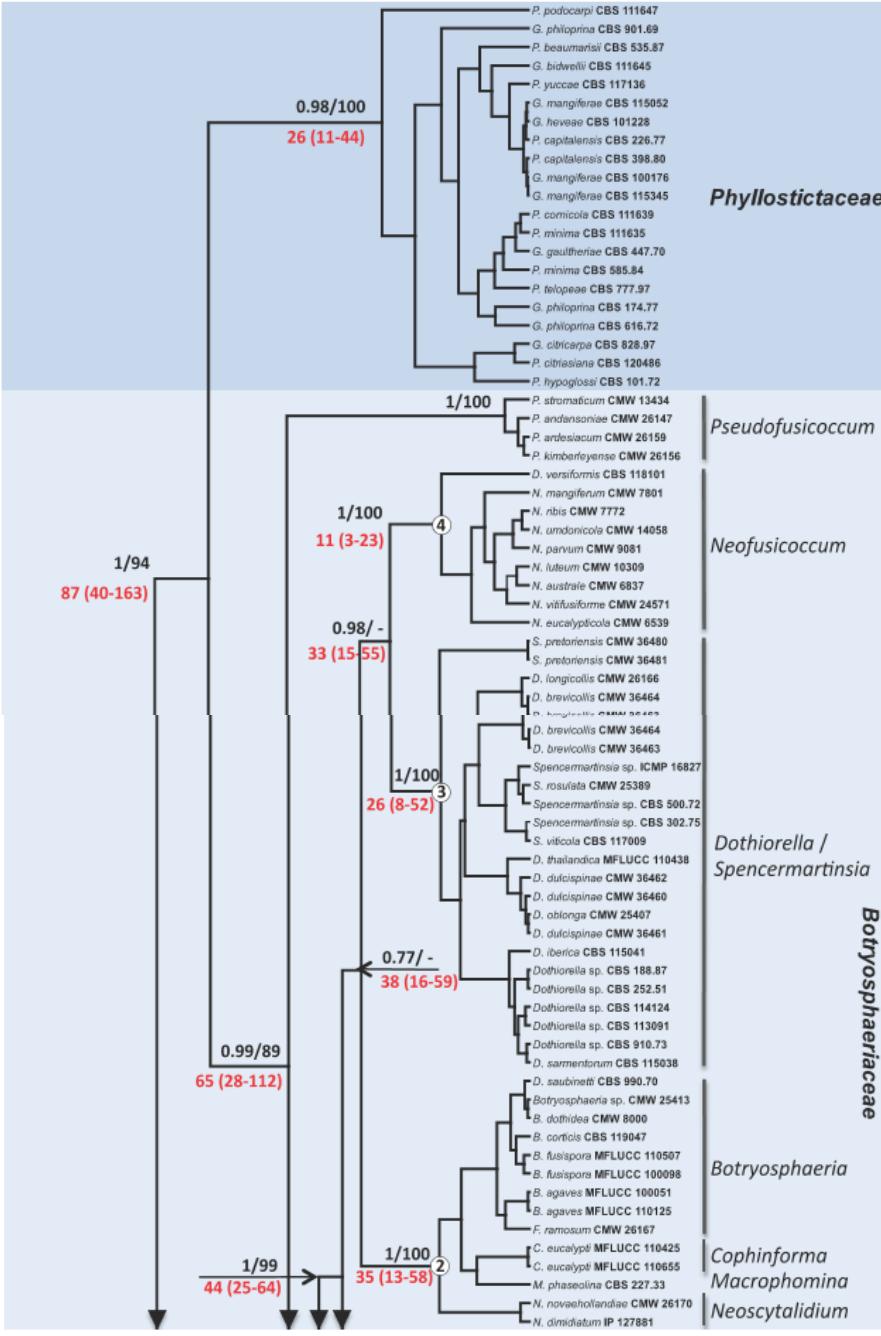
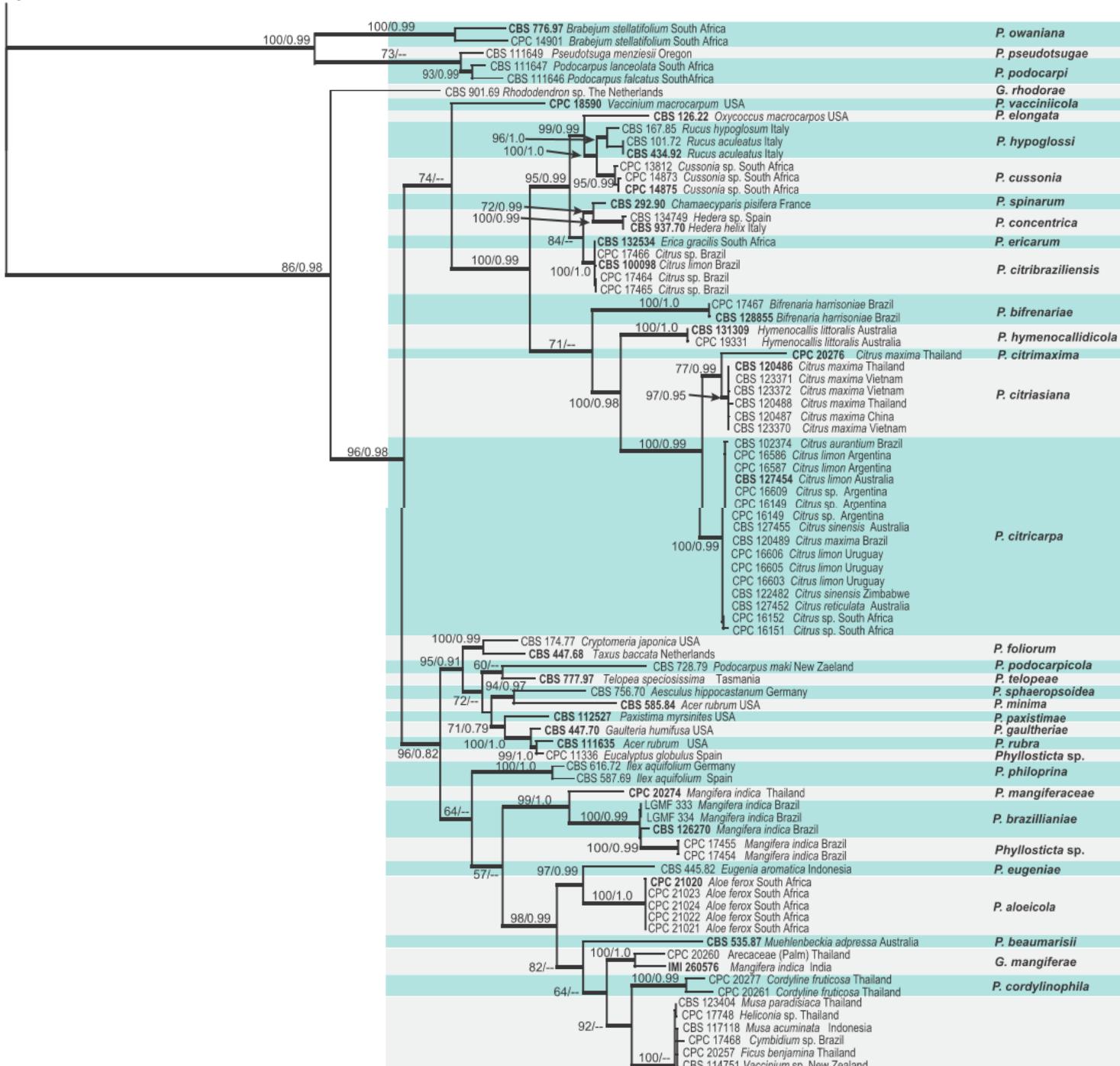
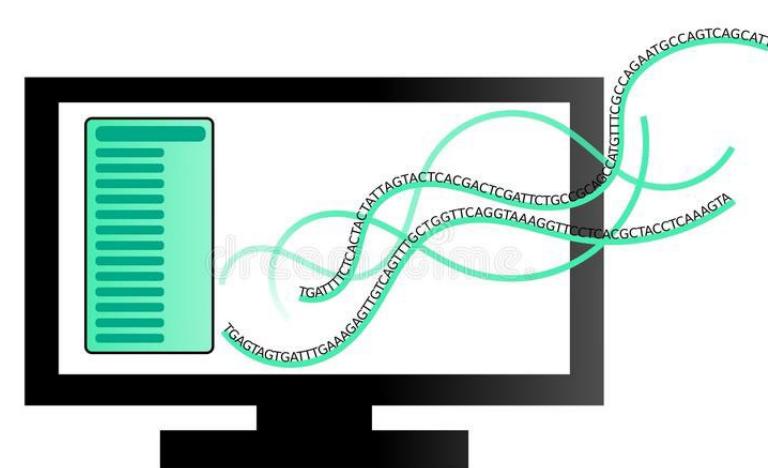


Fig. 1. Phylogenetic relationships of the Botryosphaeriales using Bayesian reconstruction and six gene portions (LSU, SSU, ITS, EF1, BT and mtSSU). Numbers above branches indicate bootstrap values/posterior probabilities. Numbers highlighted in red below branches indicate estimated dates in million years with the 95 % Highest Posterior Density interval given in brackets. Clades 1–4 in the Botryosphaeraceae are indicated by a circled number on the corresponding node.

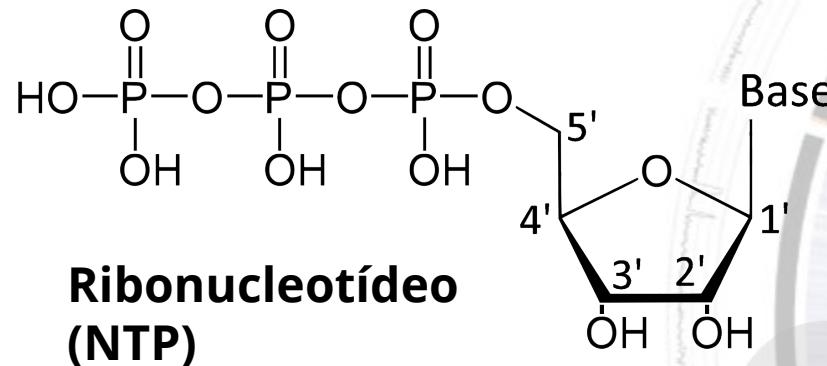
Diplodia seriata CMW 8232



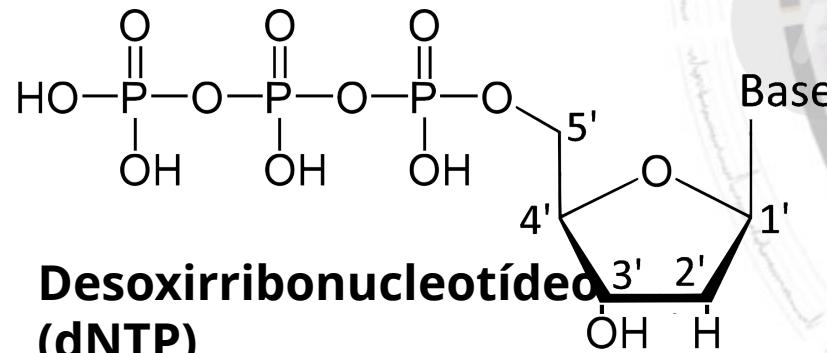
Sequenciando....



Relembrando...

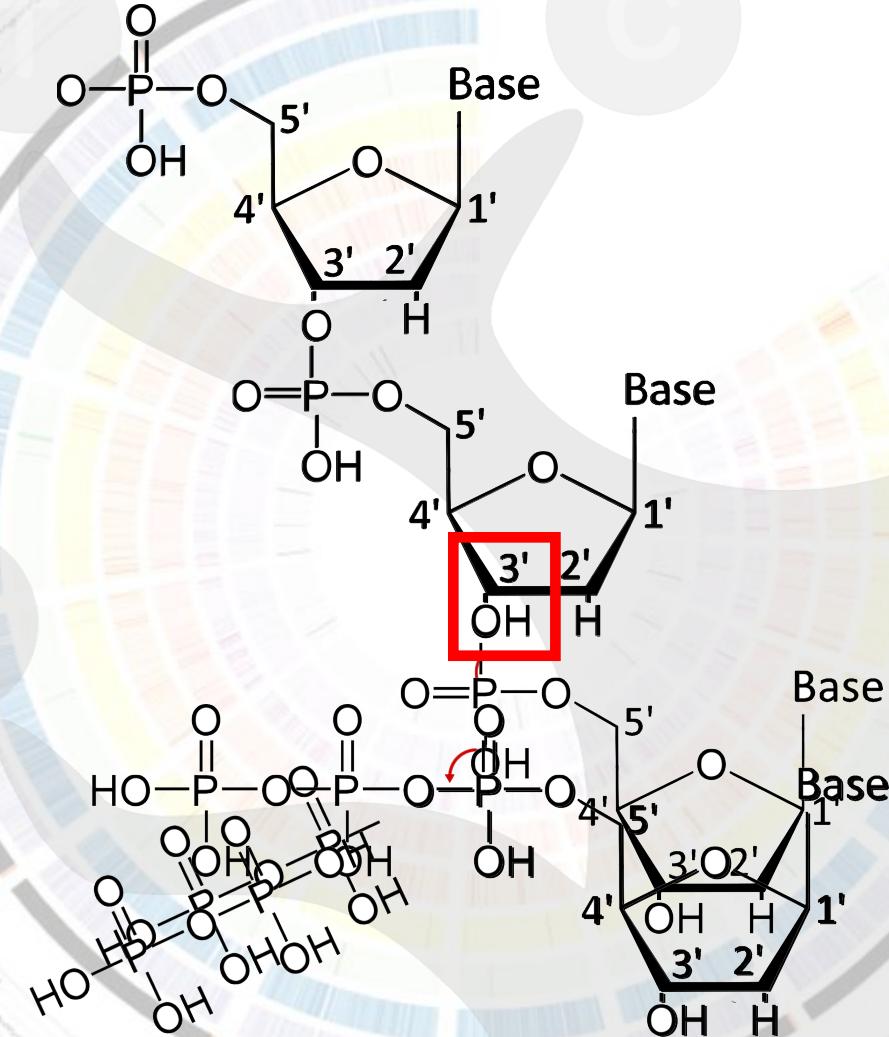


Ribonucleotídeo
(NTP)



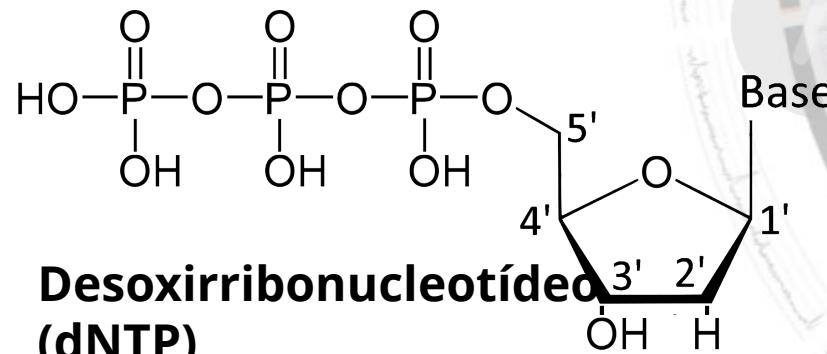
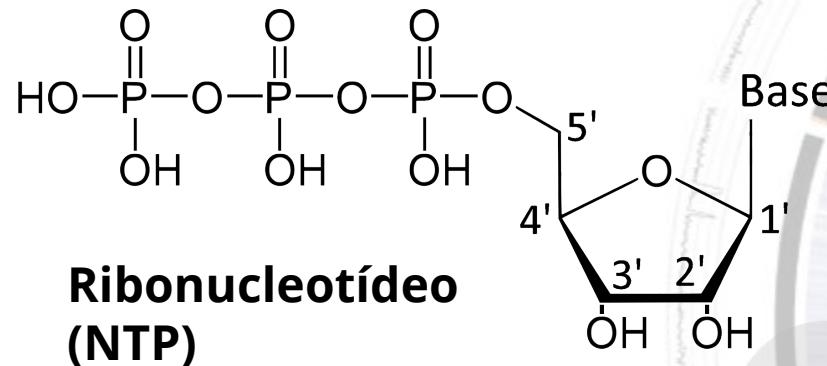
Desoxirribonucleotídeo
(dNTP)

(Brown, 2010)

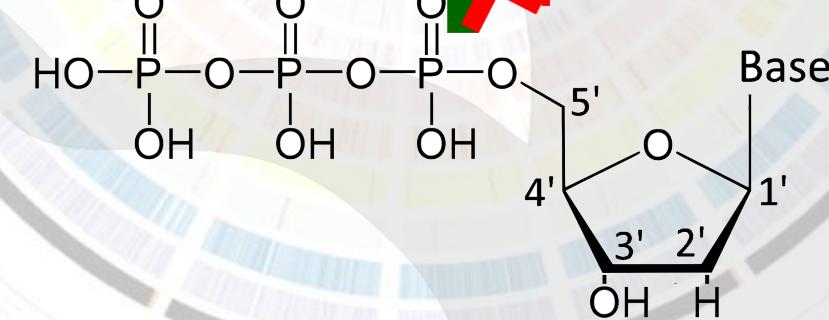
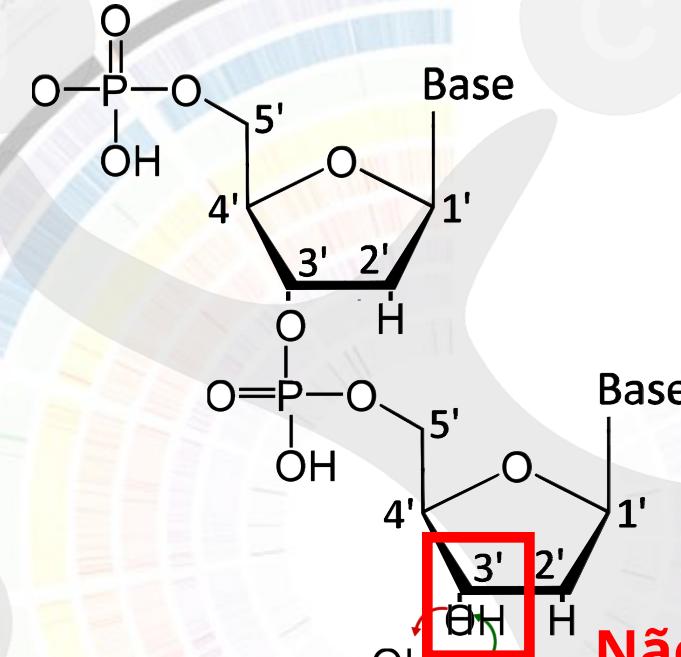


65

Relembrando...



(Brown, 2010)



**Não ocorre a
polimerização!**

Sequenciamento de DNA

Método de Sanger / Método de terminação em cadeia

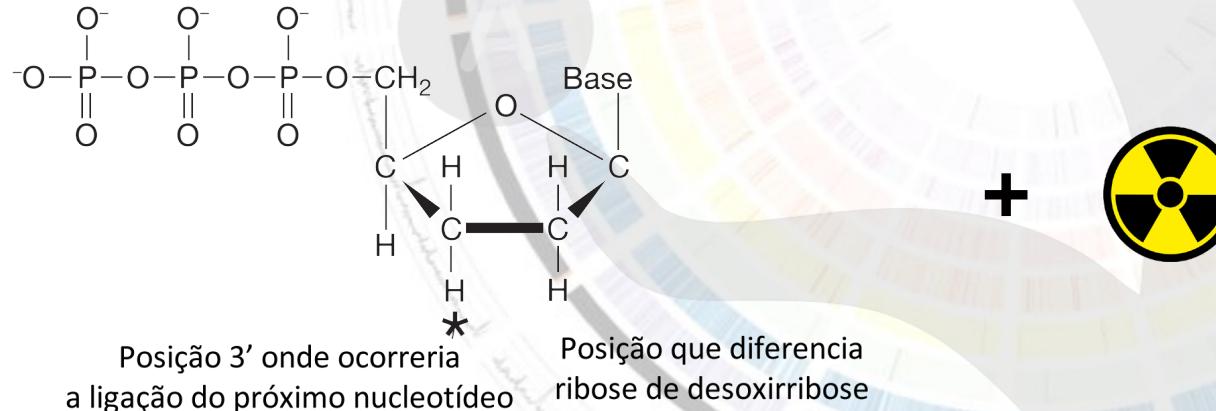
Sequenciamento associado com polimerização



Frederick Sanger

Popularizou a técnica e permite hoje sequenciamento de até cerca de 1000 pb

2',3'-didesoxirribonucleosídeo trifosfatado (ddNTP)



Método de Sanger

- Muitos Desoxirribonucleotídeos (dNTP)
- Poucos Didesoxirribonucleotídeos com uma das bases nitrogenadas (ddATP)

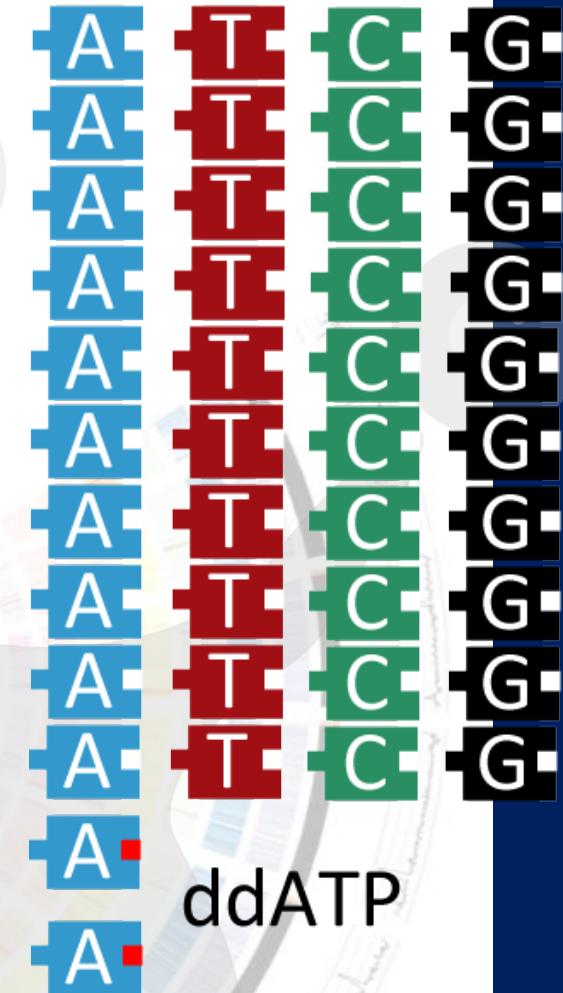
Parada da
síntese!



Método de Sanger

- Muitos Desoxirribonucleotídeos (dNTP)
- Poucos Didesoxirribonucleotídeos com uma das bases nitrogenadas (ddATP)

T A A G C T A G C T C C A -

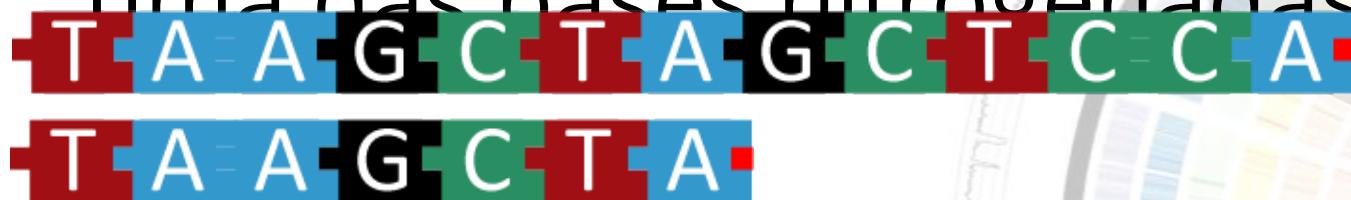


Parada da
síntese!

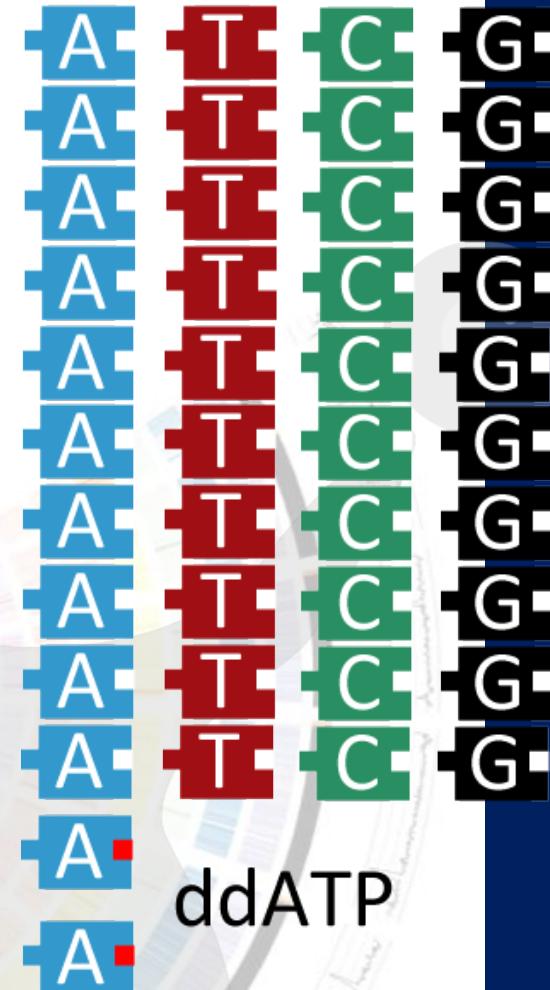


Método de Sanger

- Muitos Desoxirribonucleotídeos (dNTP)
- Poucos Didesoxirribonucleotídeos com uma das bases nitrogenadas (ddATP)

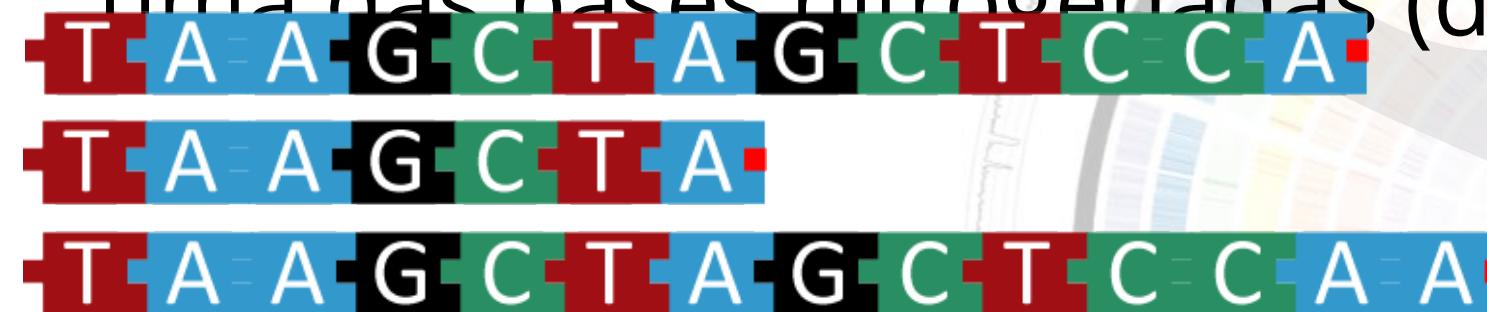


Parada da
síntese!



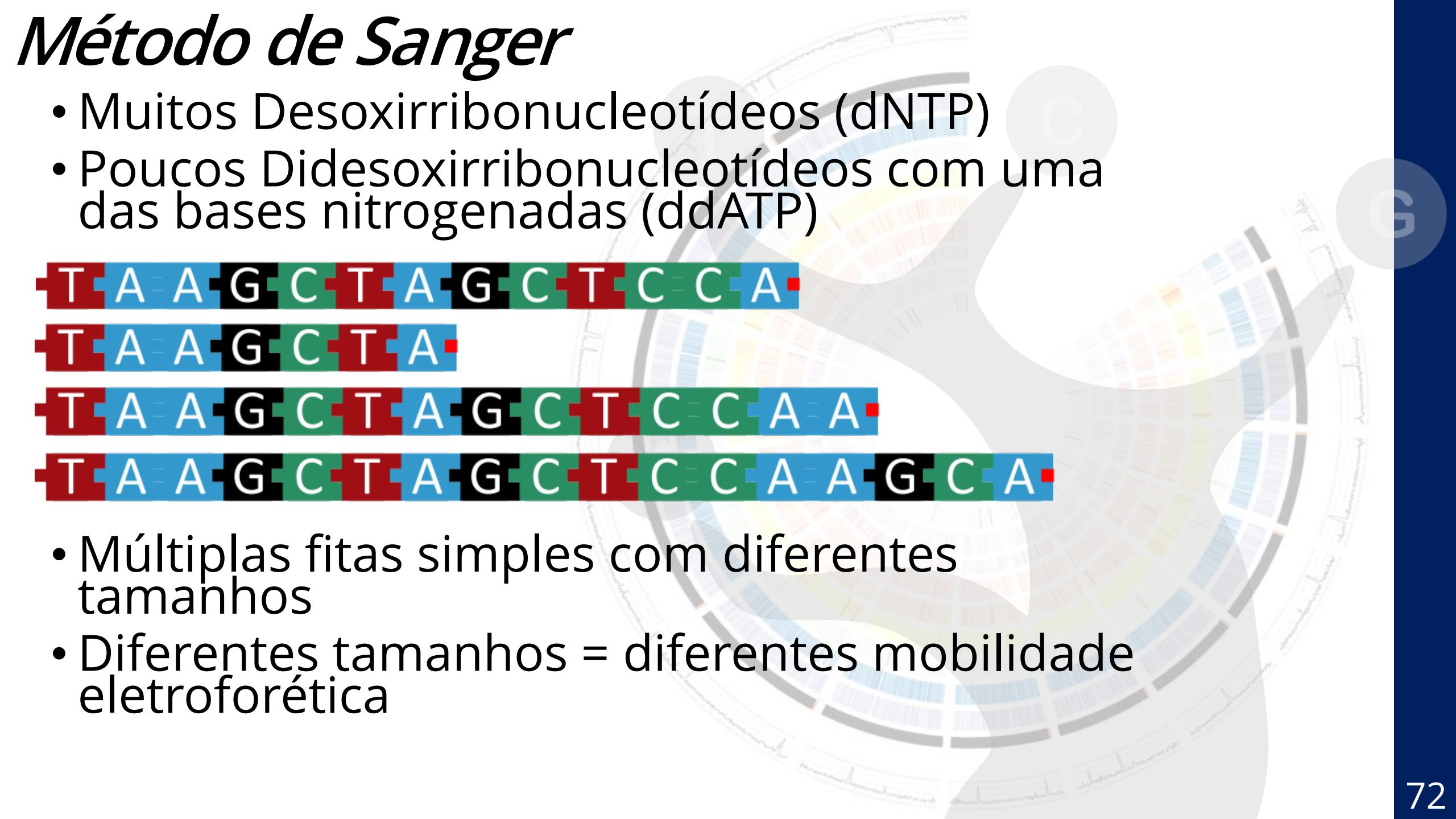
Método de Sanger

- Muitos Desoxirribonucleotídeos (dNTP)
- Poucos Didesoxirribonucleotídeos com uma das bases nitrogenadas (ddATP)



Método de Sanger

- Muitos Desoxirribonucleotídeos (dNTP)
- Poucos Didesoxirribonucleotídeos com uma das bases nitrogenadas (ddATP)



The background features a stylized DNA double helix on the right and a circular gel electrophoresis pattern with colored bands (red, blue, green, yellow) on the left.

TAAGCTAGCTCCA-

TAAGCTA-

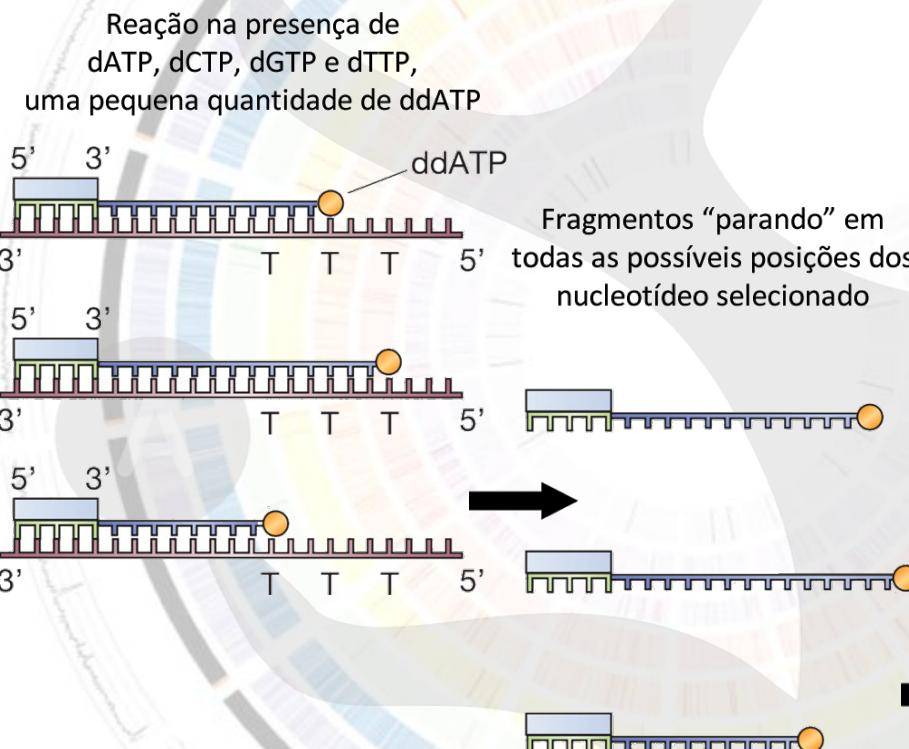
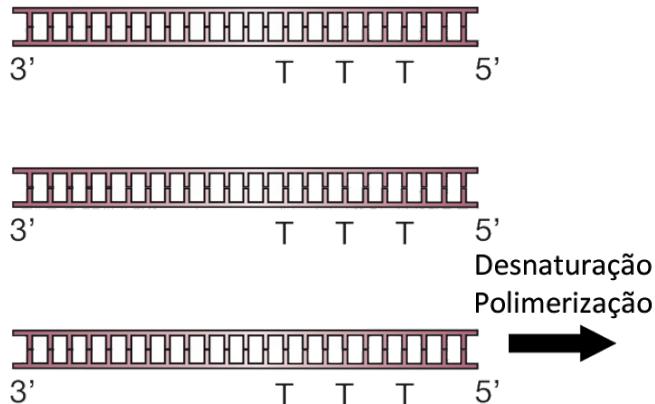
TAAGCTAGCTCCA A-

TAAGCTAGCTCCA AGCA-

- Múltiplas fitas simples com diferentes tamanhos
- Diferentes tamanhos = diferentes mobilidade eletroforética

Método de Sanger

Grande abundância de DNA molde



(Brown, 2010)

Método de Sanger

STEP

1

Montar 4 reações de polimerização de DNA com os seguintes componentes:

Fita molde 3' – GCATGATCGG – 5'

Primer (1) 5' OH 3'

DNA Polimerase
dGTP, dATP, dTTP, ^{32}P -dCTP

STEP

2

Adicione um dos quatro 2',3'-didesoxirribonucleotídeos trifosfatados (terminadores de cadeia) a cada uma das reações

Reação 1 :ddGTP Reação 2 :ddATP Reação 3 :ddCTP Reação 4 :ddTTP

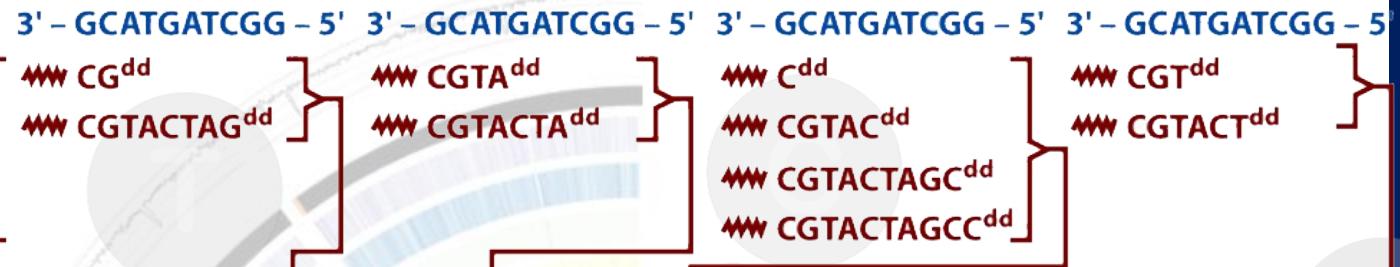
Produtos

$3' - \text{GCATGATCGG} - 5'$	$3' - \text{GCATGATCGG} - 5'$	$3' - \text{GCATGATCGG} - 5'$	$3' - \text{GCATGATCGG} - 5'$
$\text{*** CG}^{\text{dd}}$	$\text{*** CGTA}^{\text{dd}}$	*** C^{dd}	$\text{*** CGT}^{\text{dd}}$
$\text{*** CGTACTAG}^{\text{dd}}$	$\text{*** CGTACTA}^{\text{dd}}$	$\text{*** CGTAC}^{\text{dd}}$	$\text{*** CGTACT}^{\text{dd}}$
		$\text{*** CGTACTAGC}^{\text{dd}}$	$\text{*** CGTACTAGC}^{\text{dd}}$
		$\text{*** CGTACTAGCC}^{\text{dd}}$	$\text{*** CGTACTAGCC}^{\text{dd}}$

Método de Sanger

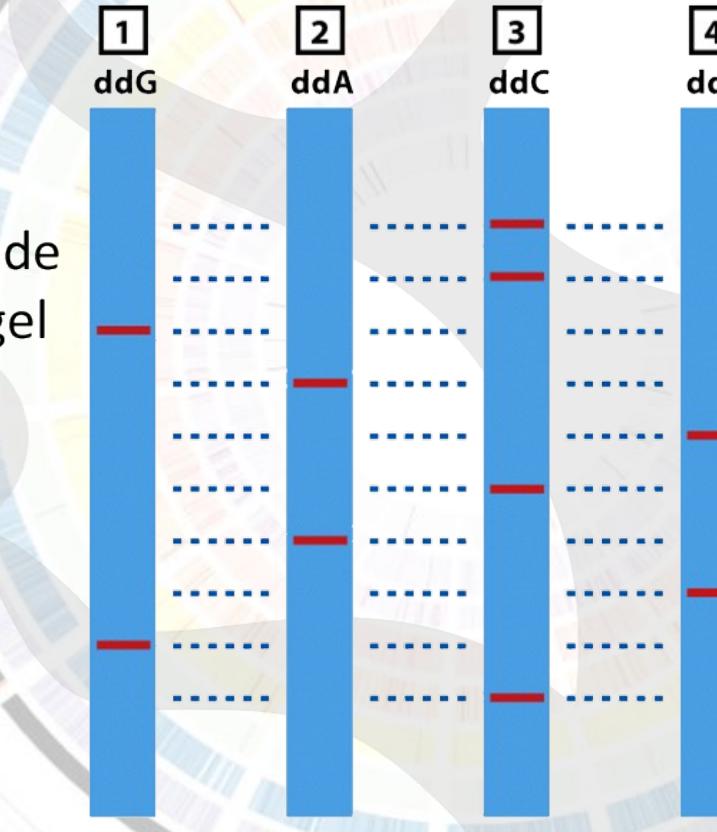
- Eletroforese em poliacrilamida (alta resolução para pequenos fragmentos)
- Corrida simultânea de todas as quatro reações

Produtos



STEP
3-6

Desnaturar os produtos de reação, aplicar em um gel de poliacrilamida, separar baseado no tamanho e expor a um filme de raios-x



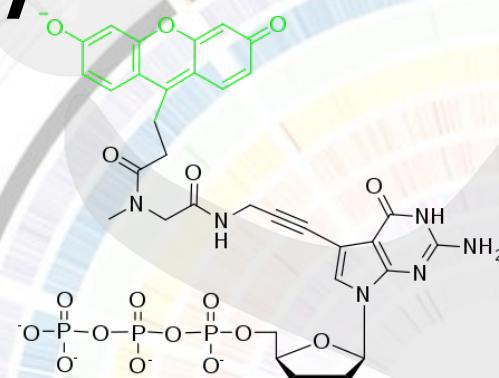
Autoradiograma do gel de sequenciamento

Sequencia
da fita nascente

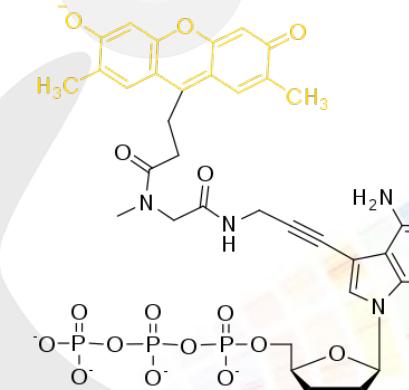
Sequencia
da fita complementar

Método de Sanger

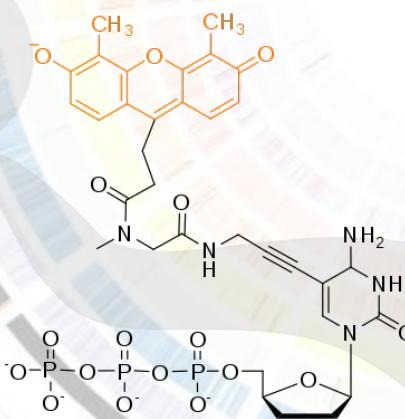
- Utilização de nucleotídeos com modificações fluorescentes
 - Permitiu realizar uma reação apenas e não quatro separadas
 - Permitiu automatizar o processo aquisição do dado e análise



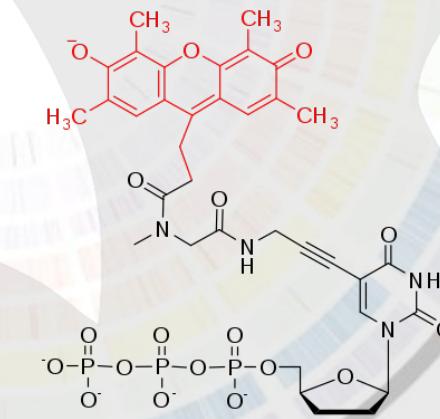
G-505



A-51



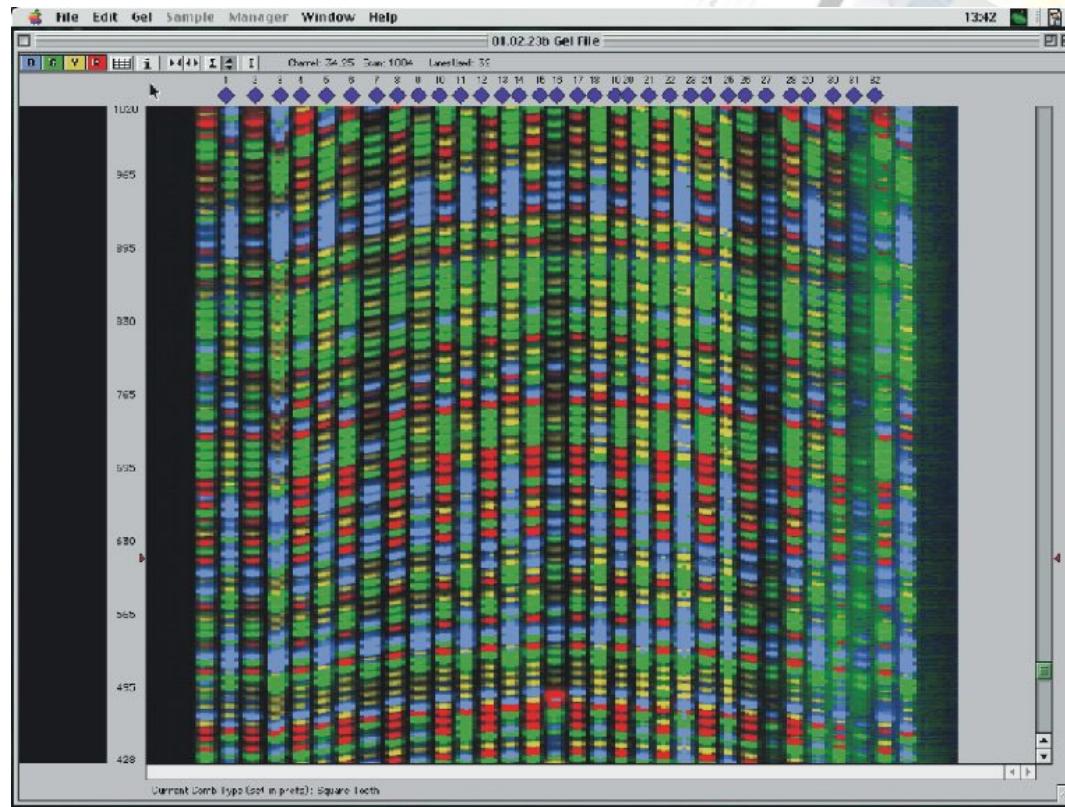
C-519



T-526

Método de Sanger

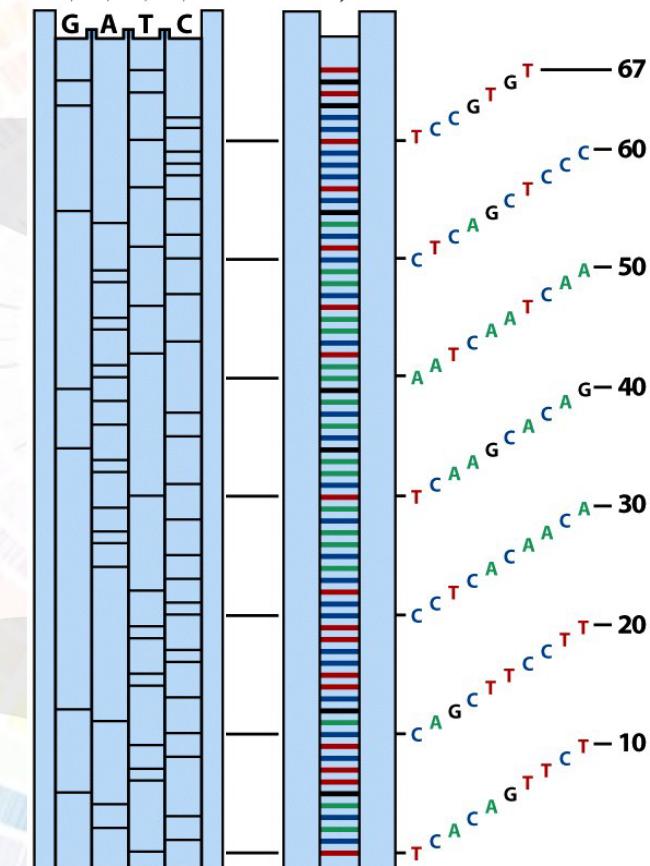
T C A G



Cada reação
em uma linha
separada

G A T C

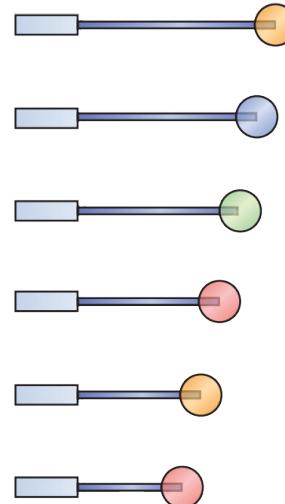
As 4 reações
na mesma
linha



Método de Sanger

Eletroforese em capilar e automatização

Separação dos fragmentos
por tamanho



Detector

Sistema de imagem

Legenda
ddATP
ddTTP
ddCTP
ddGTP

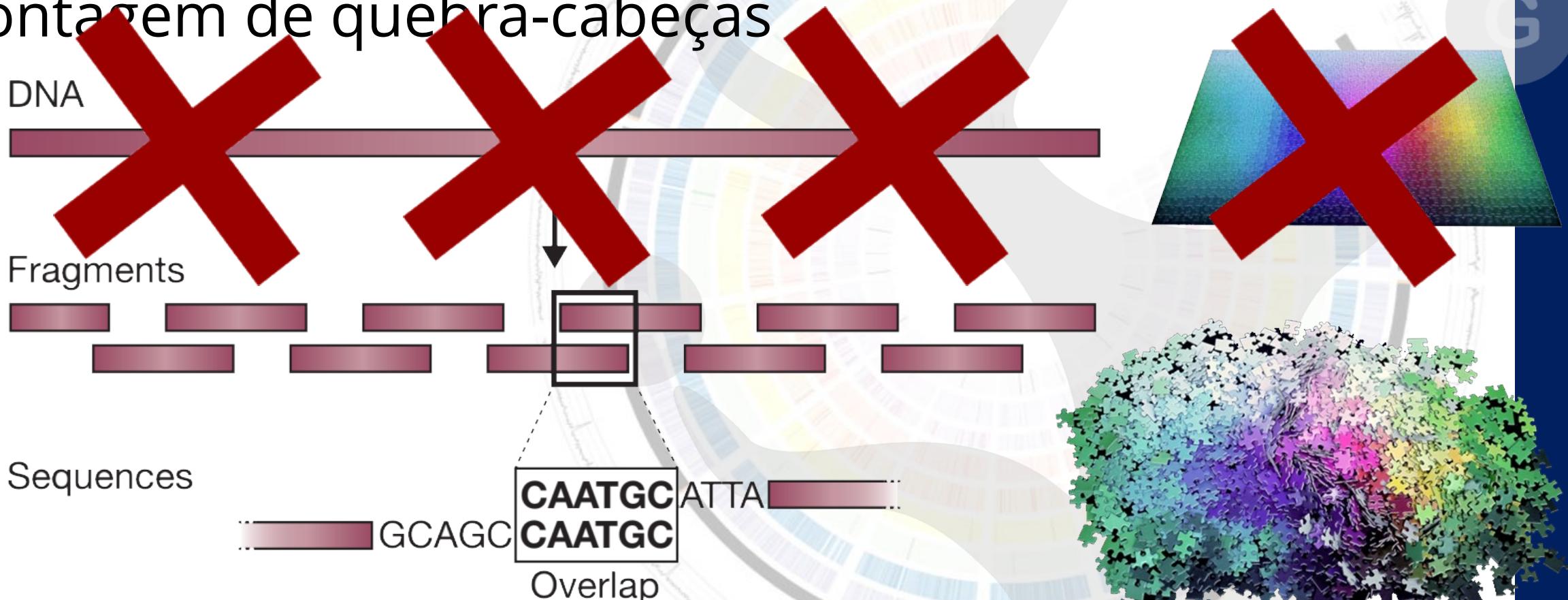
CACCGCATCGAAATTAACTTCCAAAGTTAAGCTTGG

10 20 30

Eletroforetograma

Sequenciamento de Genomas

- Promessa de resolução de todos os problemas biotecnológicos e de saúde
- Montagem de quebra-cabeças



- Sem a foto da caixa do quebra-cabeças

SEQUENCIAMENTO



SEQUENCIAMENTO

1 gtgcccaagtg agcgaggact gca
61 ggggaccgcga acaacatgga tag
121 acgtttccatg gggccaagga ggt
181 atggggcctca tataacaacag gat
241 tgtgccactt gtgaacagat tgc
301 accaccaatc cactaatcag gca
361 gctatggAAC agatggctgg atc
421 cagactaggc agatggtaCA tgc
481 ggtctgaaAG atgaccttct tga
541 atgcAGCgt tcaagtgtac ctc



Comparison of BLAST-based (pairwise alignment) vs. tree-based (multiple alignment) identification of a target fungal ITS sequence (DB42771, Vietnam; see Lücking et al. 2020).

(A) A simple sequence alignment

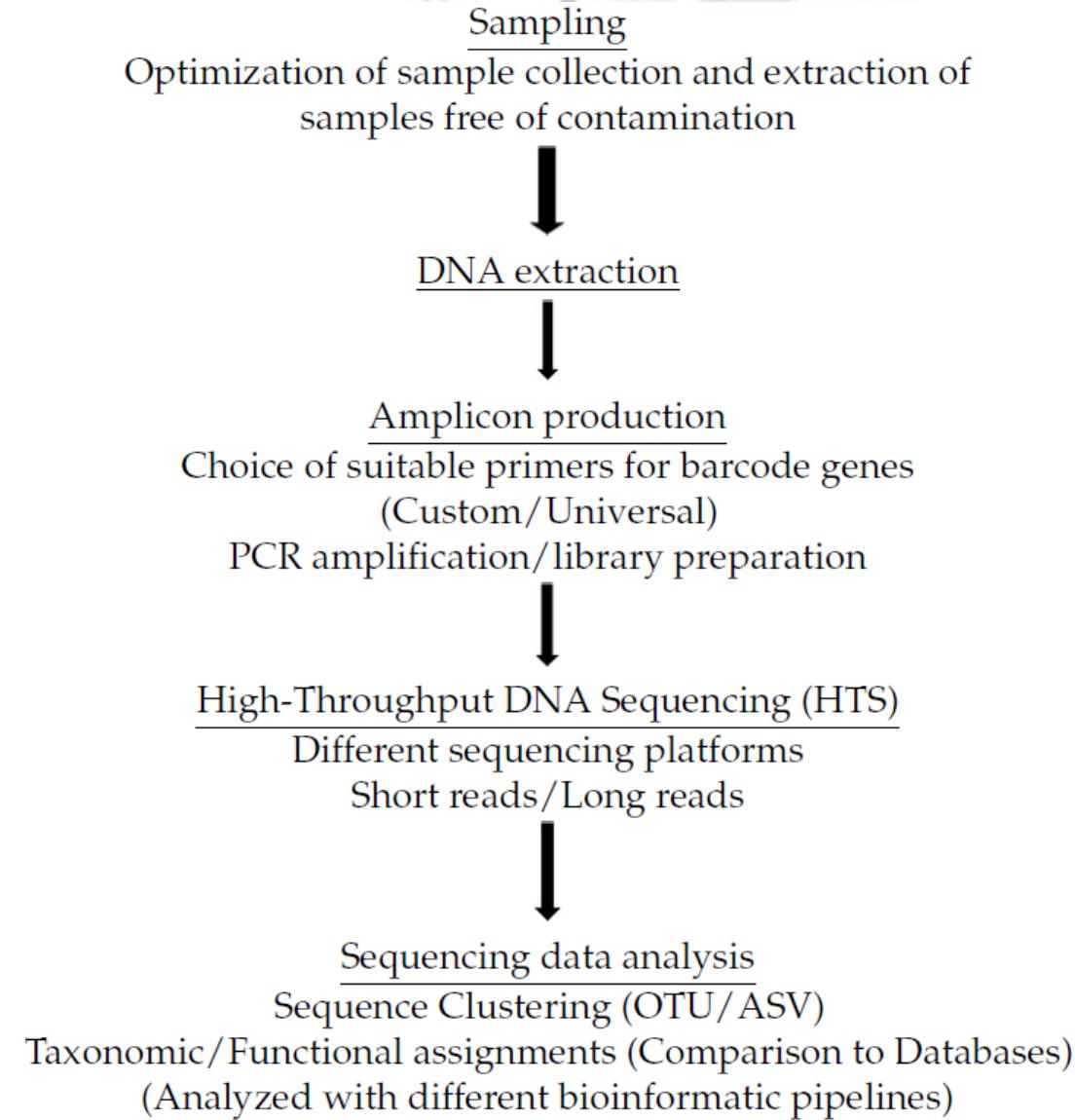
AGCAATGGCCAGACAATAATG
AGCTATGGACAGACATTAAATG
*** * *** * *** * *** *

(B) A more difficult sequence alignment

GACGACCATAGACCAGCAGTAG
GACTACCATA - CTGCAAAG
*** * *** * *** * * *

GACGACCATAGACCAGCAGTAG
GACTACCATA - GCAAAG
*** * *** * *** * *** * *** *

Two possible positions
for the indel



Importante

Estudo filogenético e filogenômico:

Uso de um conjunto de dados **bem curado** e balanceado
por táxons

Quando fazer Filogenia e quando fazer filogenômica?