Sequências de DNA na identificação de espécies e análise filogenética de microrganismos

Dra. Chirlei Glienke

Dra. Desirrê Petters-Vandresen

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).

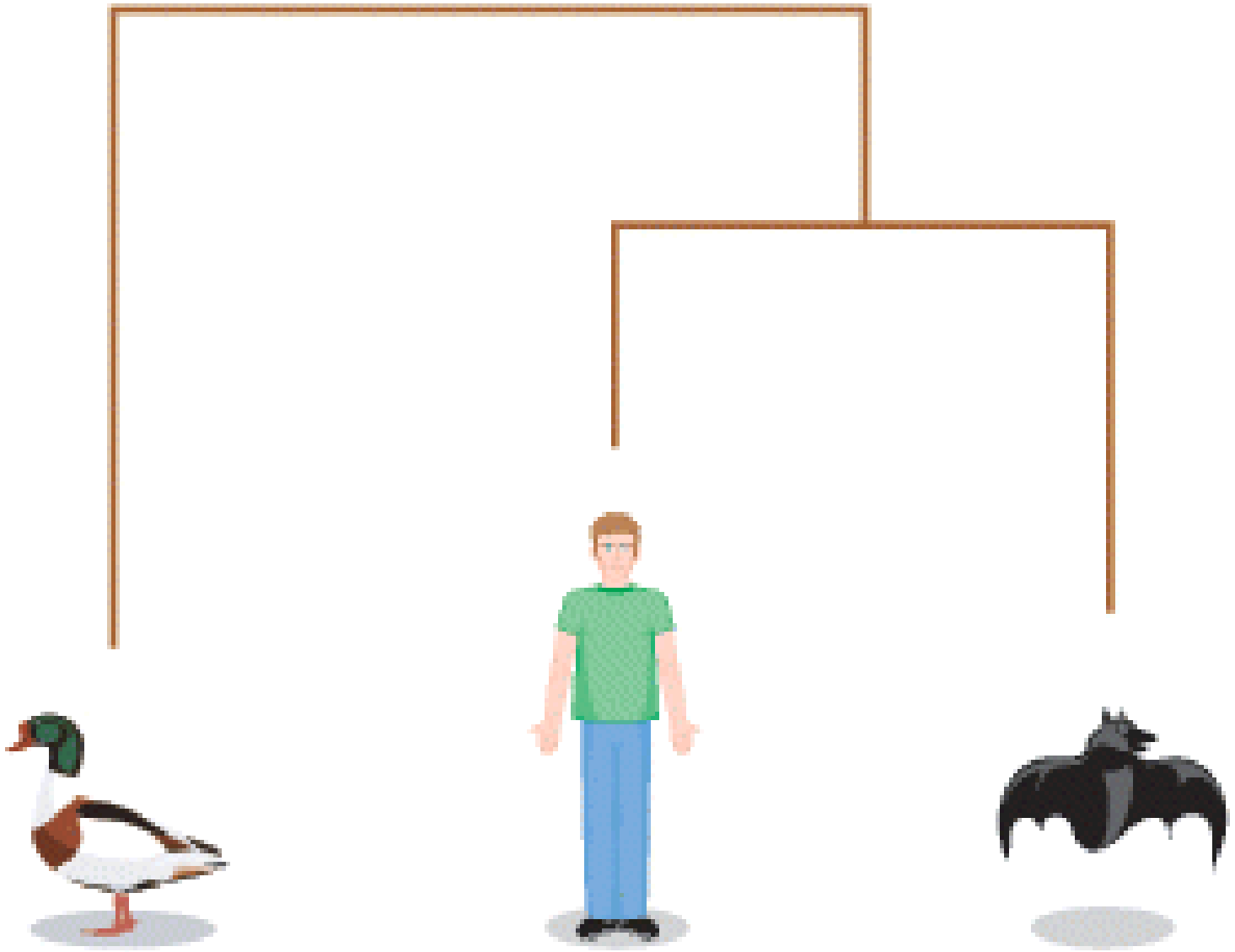
Que genes usar?

- 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
RPB1	Partial sequence of the RNA polymerase II largest subunit coding gene	253–873	620
RBP2	Partial sequence of the RNA polymerase II second largest subunit coding gene	1645–2319	680–1000
TEF1-alpha	Partial sequence of the translation elongation factor 1-alpha coding gene	64–1190	930
ACT1	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

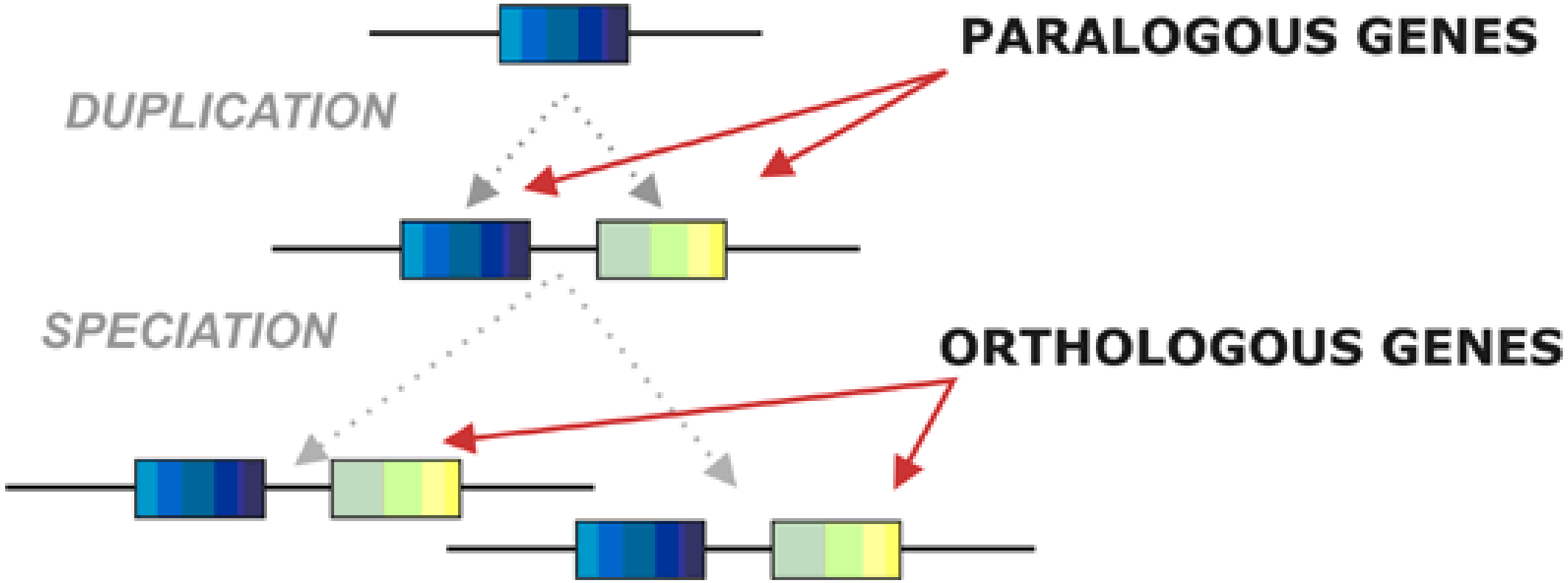
(A)



Wings

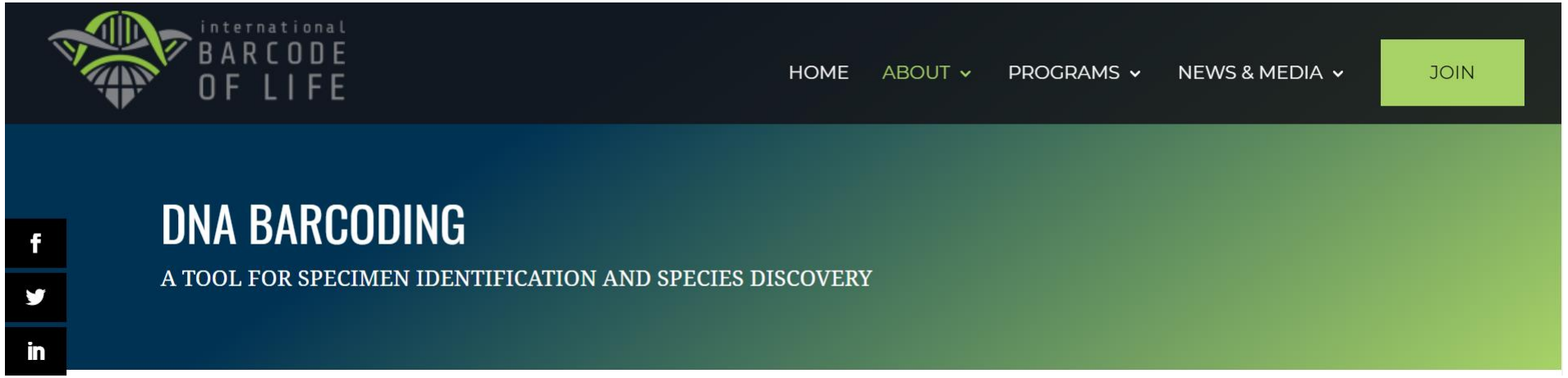
No wings

Wings


PARALOGOUS GENES
DUPLICATION
SPECIATION
ORTHOLOGOUS GENES

DNA BARCODING

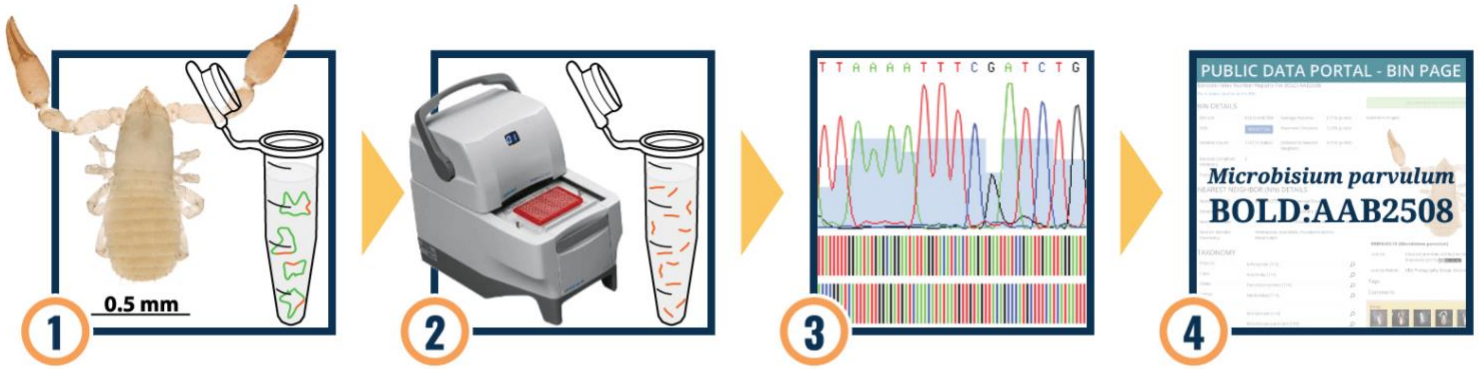
A TOOL FOR SPECIMEN IDENTIFICATION AND SPECIES DISCOVERY



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

DNA BARCODING

Como funciona?



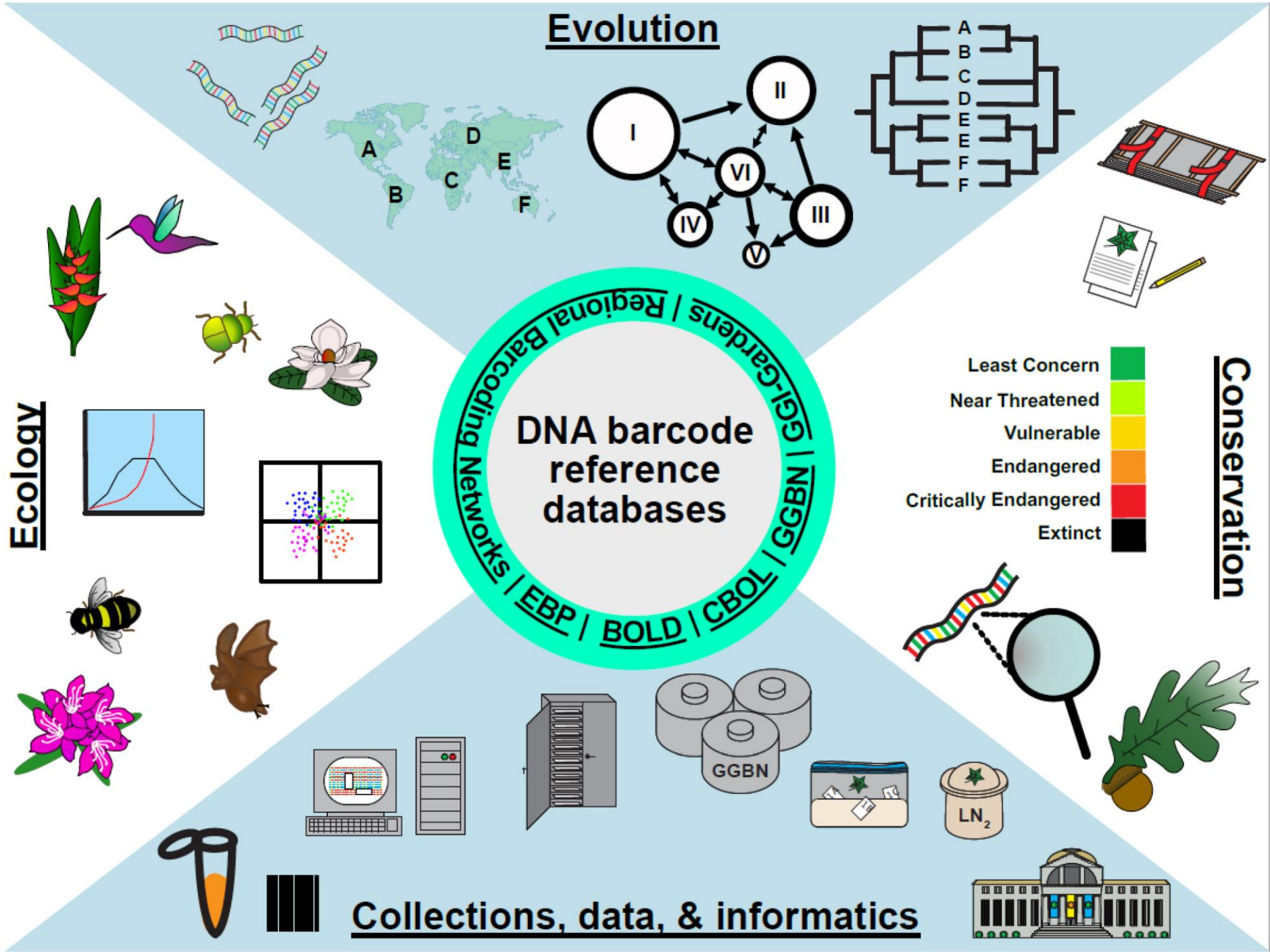
Step 1: Isolate DNA from the sample

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/


Evolution
A
B
C
D
E
F
I
II
III
IV
V
VI
A
B
C
D
E
E
F
F
Ecology
DNA barcode reference databases
Regional Barcoding Networks | EBP | BOLD | CBOL | GGBN | GGI-Gardens
Conservation
Least Concern
Near Threatened
Vulnerable
Endangered
Critically Endangered
Extinct
GGBN
LN2
Collections, data, & informatics

THE DNA BARCODE

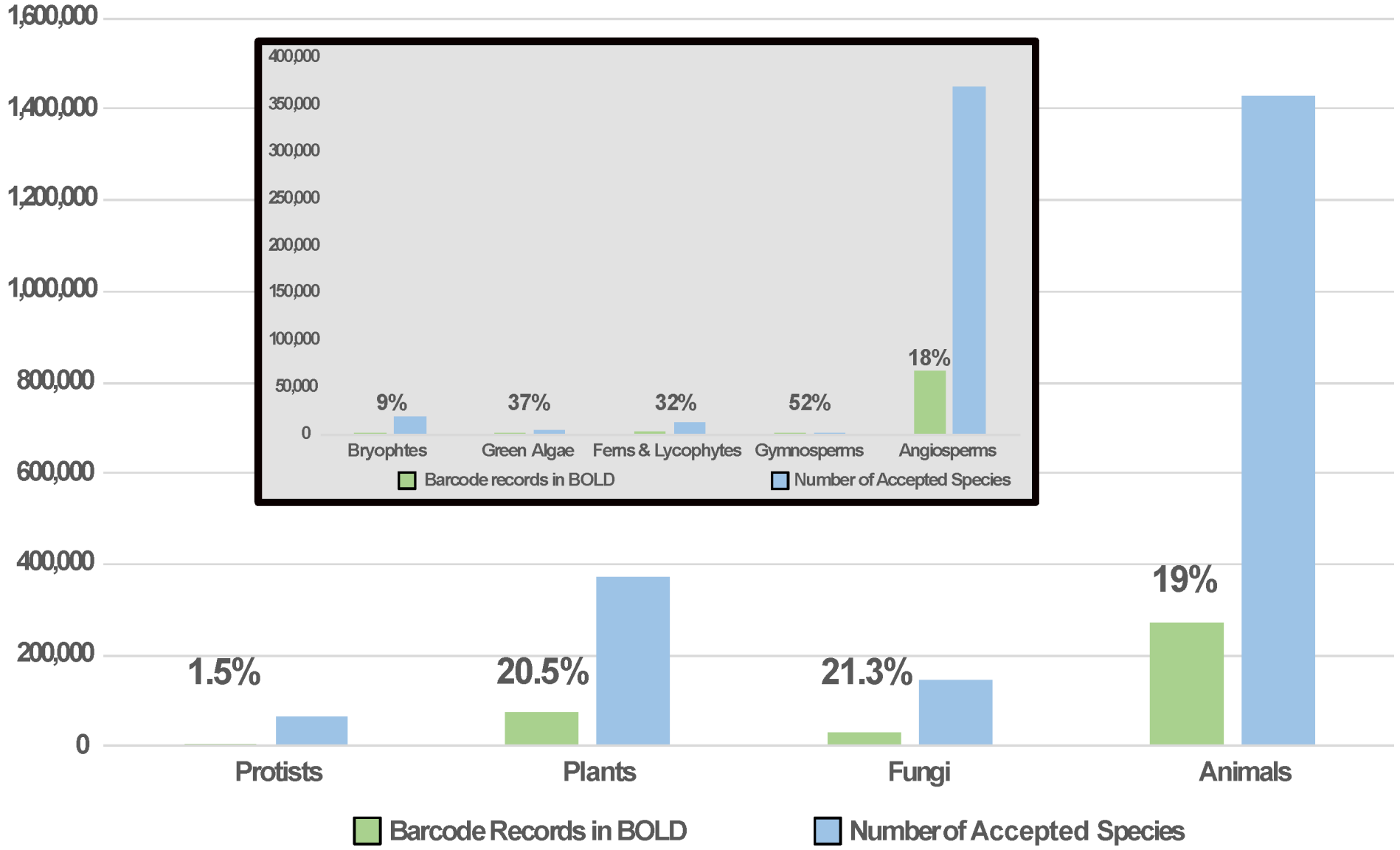
international BARCODE OF LIFE

HOME ABOUT PROGRAMS NEWS & MEDIA JOIN

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.


1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
400,000
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
9%
37%
32%
52%
18%
Bryophtes
Green Algae
Ferns & Lycophytes
Gymnosperms
Angiosperms
Barcode records in BOLD
Number of Accepted Species
1.5%
20.5%
21.3%
19%
Protists
Plants
Fungi
Animals
Barcode Records in BOLD
Number of Accepted Species

DNA BARCODING

international BARCODE OF LIFE

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.

DNA BARCODING

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

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

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

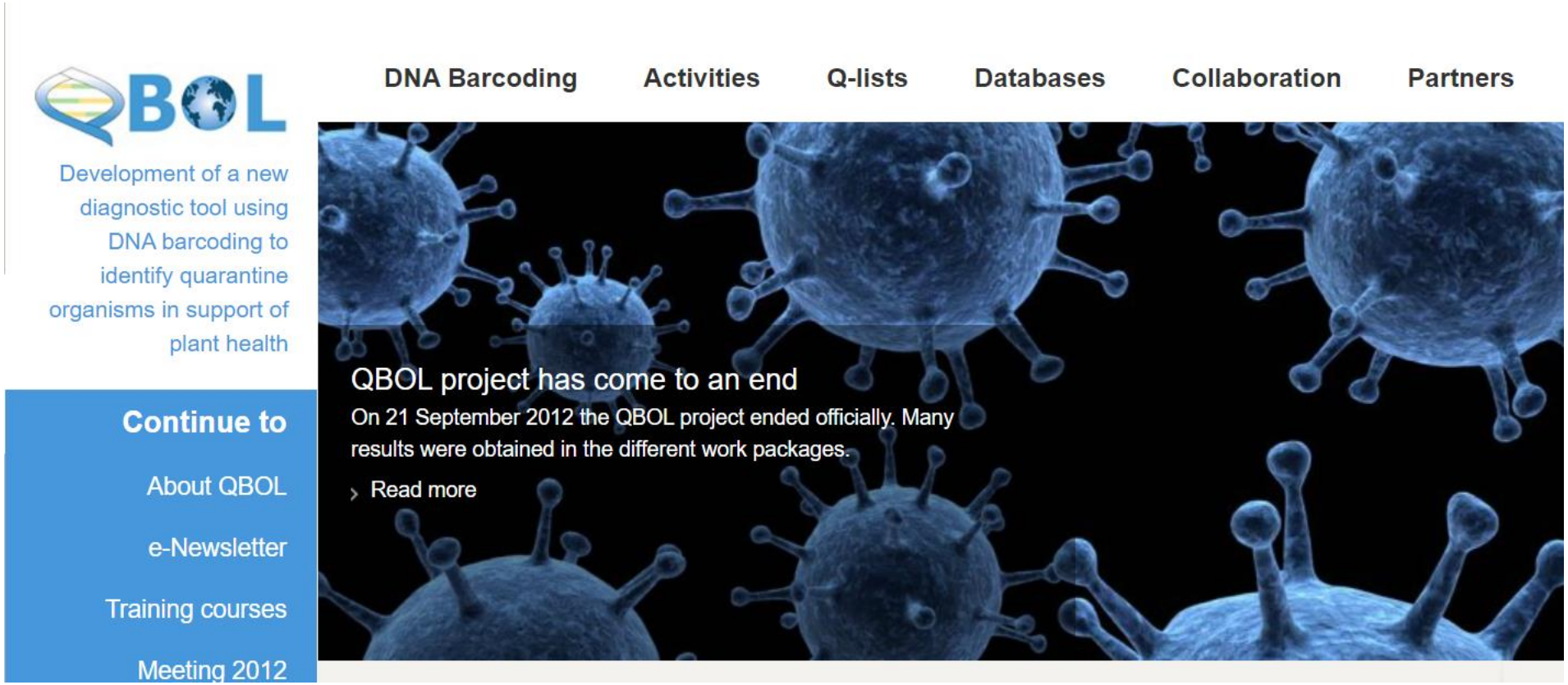
Oomycota: mitochondrial cytochrome oxidase c subunits (COX1 and COX2).

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



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

Barcoding of Arthropods

Barcoding of Bacteria

Barcoding of Fungi

Barcoding of Nematodes

Barcoding of Phytoplasmas

Barcoding of Viruses

DNA Banks

Library/Database/Informat

Validation/Evaluation

Dissemination

Q-BOL Activities

QBOL Activities:

DNA Barcoding de fungos – CBS na Holanda

QBOL

DNA Barcoding | Activities | Q-lists | Databases | Collaboration | Partners

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

Barcoding of Fungi

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

Contact

Ewald Groenewald
View profile >>

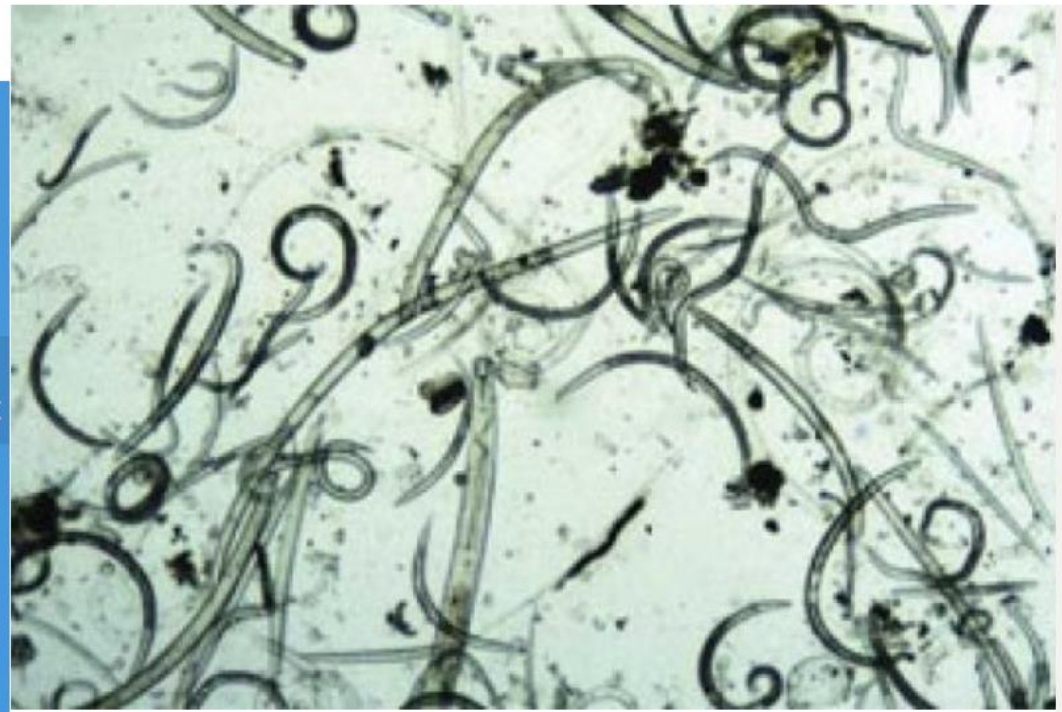
WP Coordinator: KNAW-CBS (Partner 9)

KNAW-CBS, PRI, FERA and CIP

QBOL

DNA Barcoding | Activities | Q-lists | Databases | Collaboration | Partners

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



Barcoding of Nematodes

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

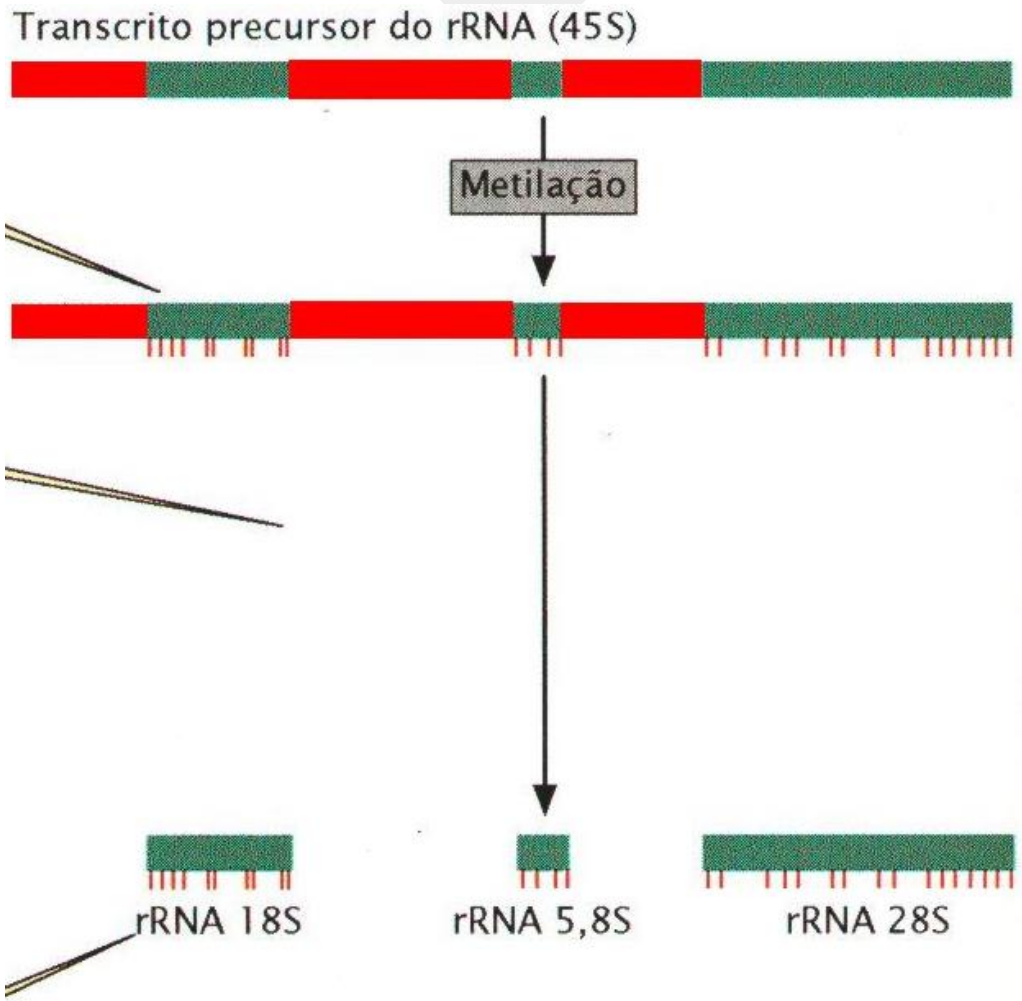
Contact

Juerg Frey
View profile >>

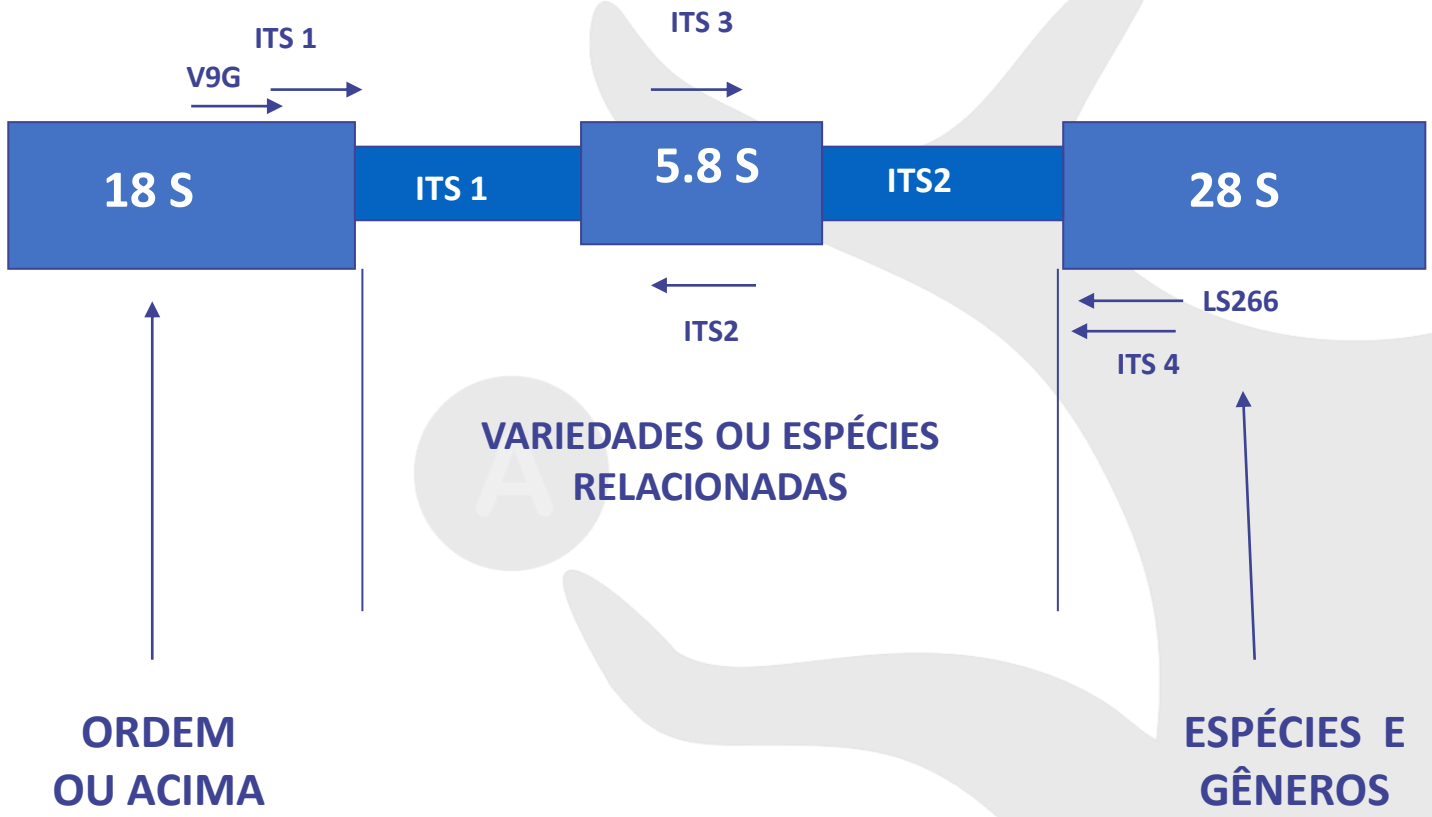
WP Coordinator: ACW (Partner 5)

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

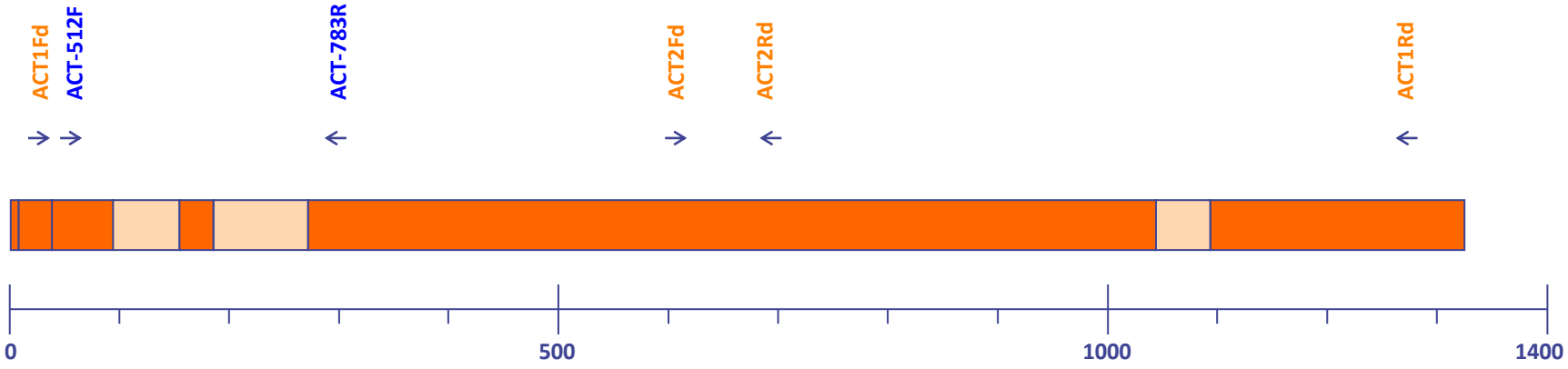
RNA RIBOSSOMAL EUCARIÓTICO



REGIÃO ITS1-5.8S-ITS2 DO rRNA



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



Exon

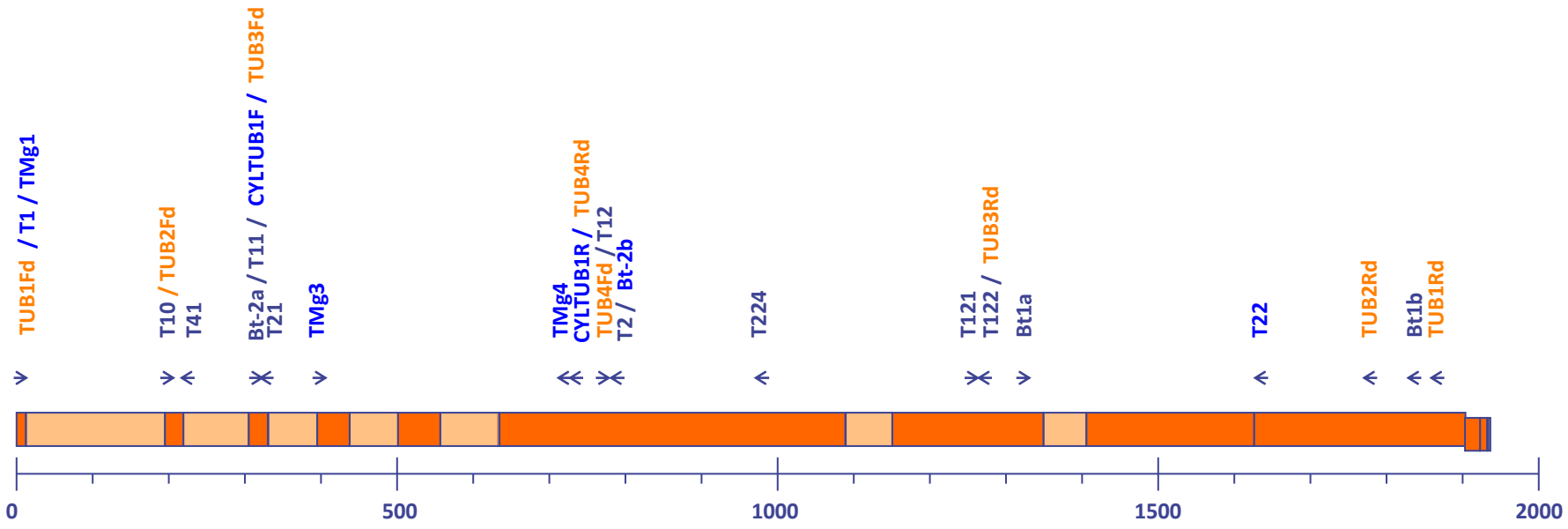
Intron

Primer available

New primer

Published primer

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



Exon

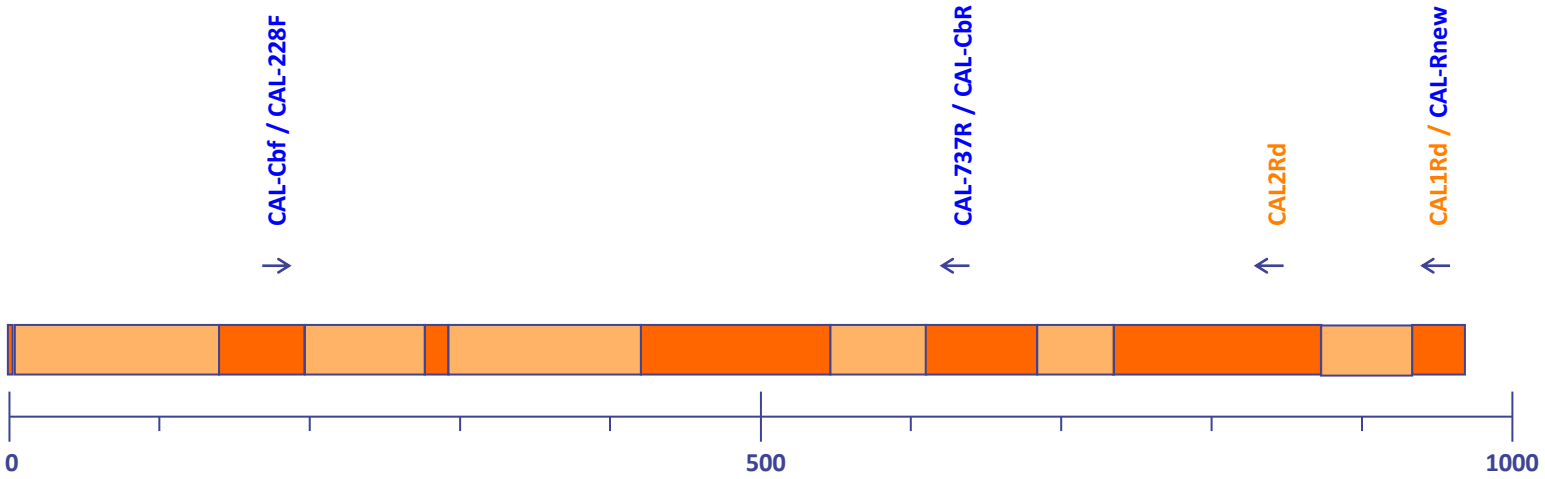
Intron

Primer available

New primer

Published primer

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



Exon

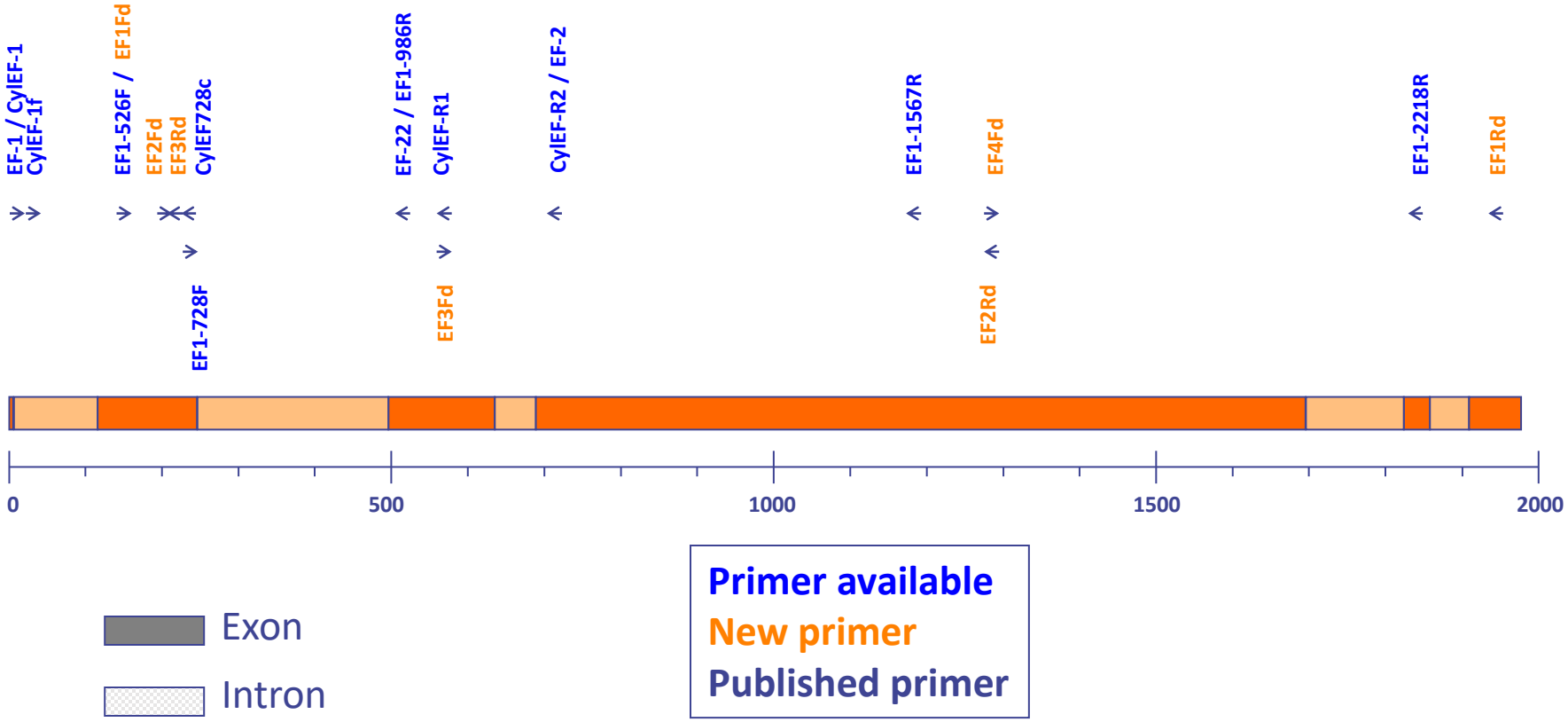
Intron

Primer available

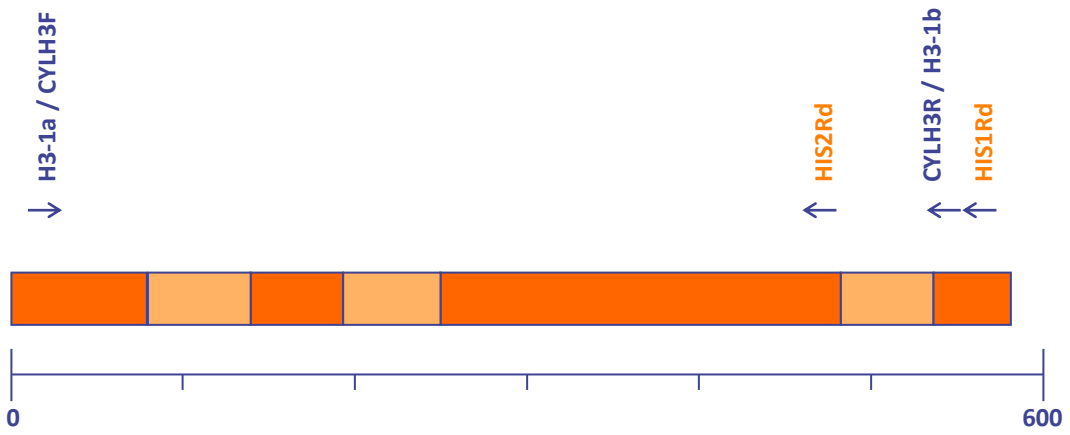
New primer

Published primer

ELONGATION FACTOR 1-ALPHA
(1975 bp total; 1383 coding for 461 amino acids)



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



Exon

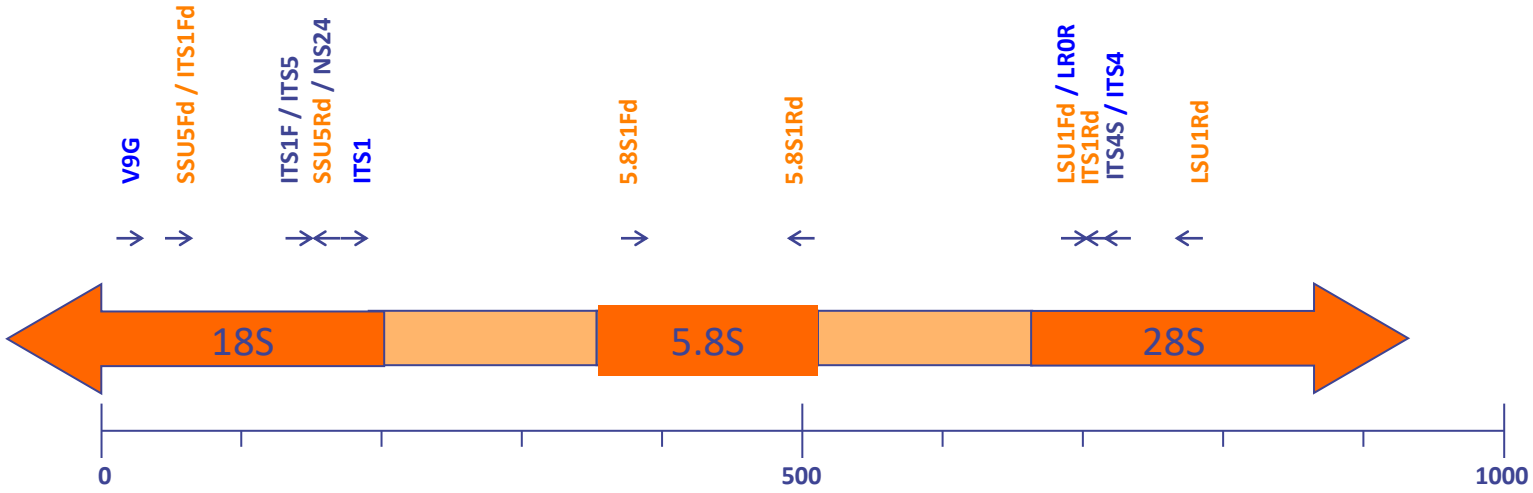
Intron

Primer available

New primer

Published primer

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



Coding

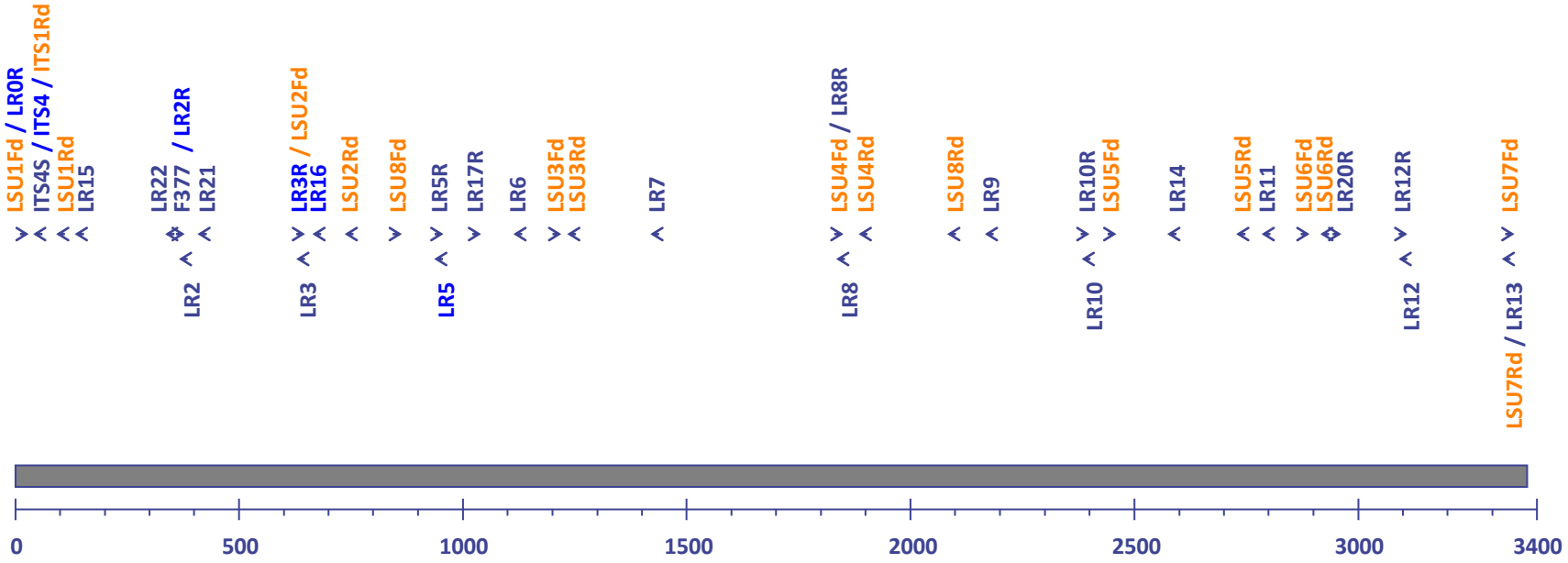
Insertion

Primer available

New primer

Published primer

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



Coding

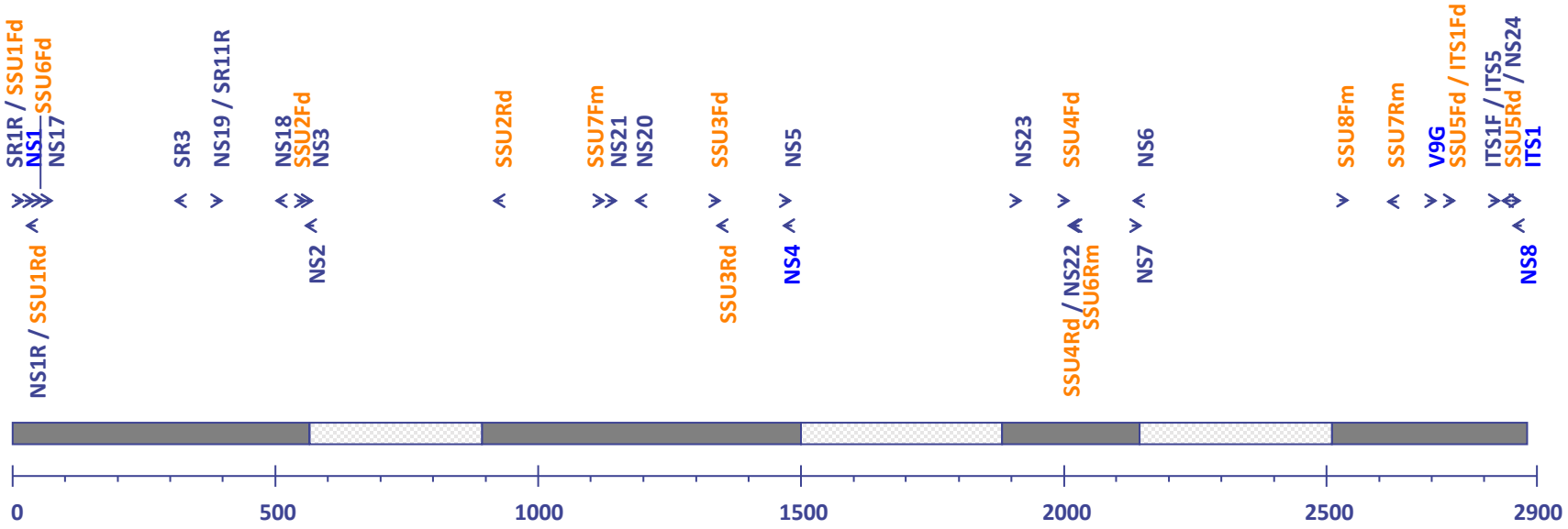
Insertion

Primer available

New primer

Published primer

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



Coding

Insertion

Primer available

New primer

Published primer

Então, qual gene vou usar?

Depende da pergunta/objetivo/ problema a resolver

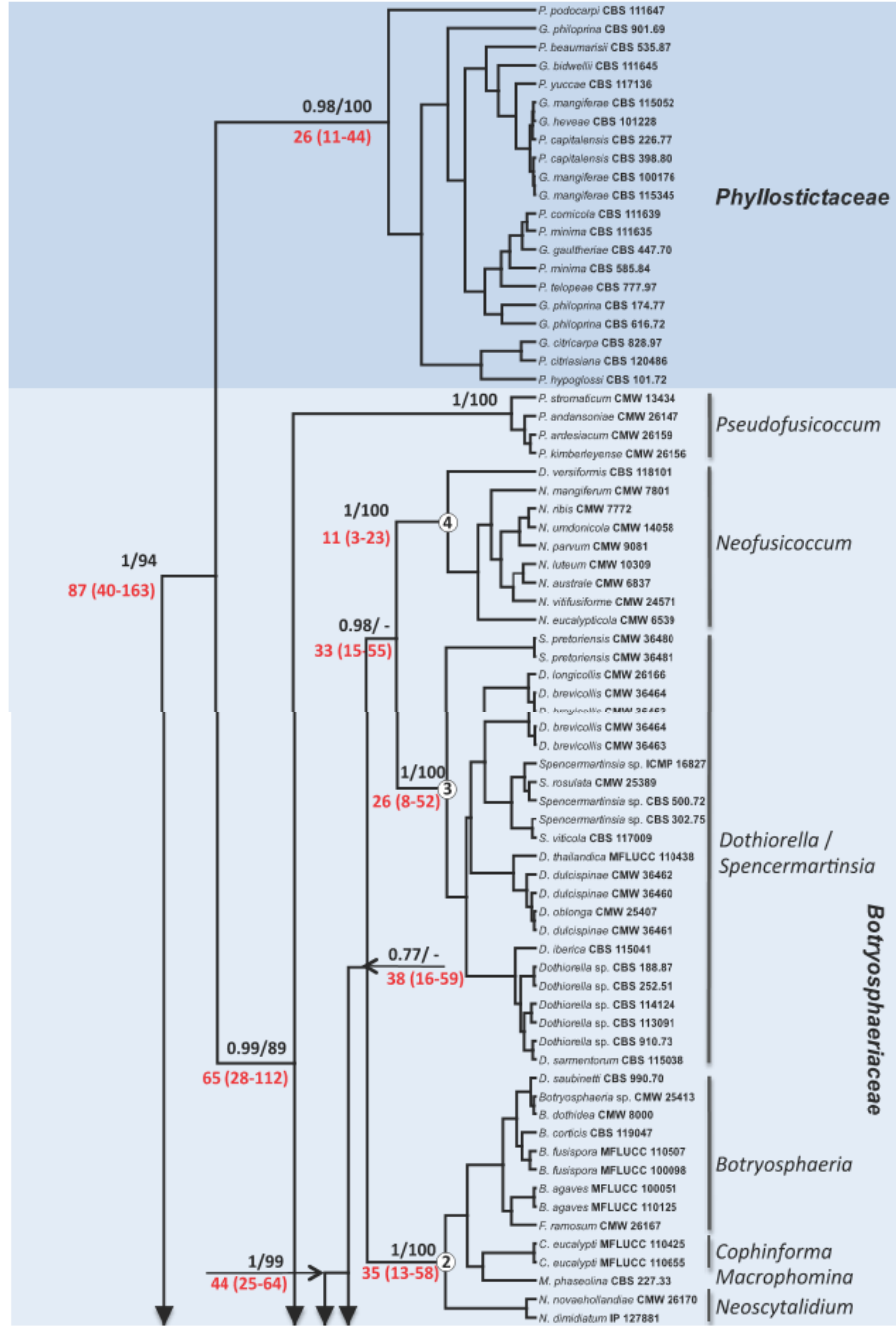
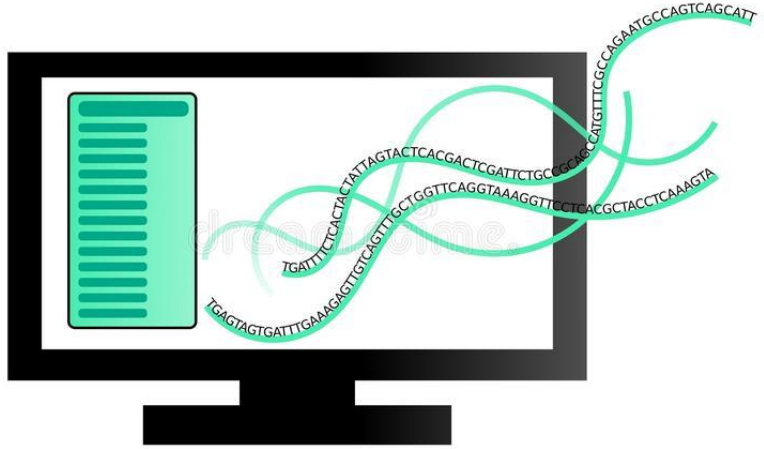


Fig. 1. Phylogenetic relationships of the *Botryosphaeriales* using Bayesian reconstruction and six gene portions (LSU, SSU, ITS, EF1, BT and mtSSU). Numbers above branches indicate bootstrap values/posterior probabilities. Numbers highlighted in red below branches indicate estimated dates in million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

Diplodia seriata CMW 8232
CBS 776.97 Brabejum stellatifolium South Africa
CPC 14901 Brabejum stellatifolium South Africa
CBS 111649 Pseudotsuga menziesii Oregon
CBS 111647 Podocarpus lanceolata South Africa
CBS 111646 Podocarpus falcatus SouthAfrica
CBS 901.69 Rhododendron sp. The Netherlands
CPC 18590 Vaccinium macrocarpum USA
CBS 126.22 Oxycoccus macrocarpos USA
CBS 167.85 Rucus hypoglosum Italy
CBS 101.72 Rucus aculeatus Italy
CBS 434.92 Rucus aculeatus Italy
CPC 13812 Cussonia sp. South Africa
CPC 14873 Cussonia sp. South Africa
CPC 14875 Cussonia sp. South Africa
CBS 292.90 Chamaecyparis pisifera France
CBS 134749 Hedera sp. Spain
CBS 937.70 Hedera helix Italy
CBS 132534 Erica gracilis South Africa
CPC 17466 Citrus sp. Brazil
CBS 100098 Citrus limon Brazil
CPC 17464 Citrus sp. Brazil
CPC 17465 Citrus sp. Brazil
CPC 17467 Bifrenaria harrisoniae Brazil
CBS 128855 Bifrenaria harrisoniae Brazil
CBS 131309 Hymenocallis littoralis Australia
CPC 19331 Hymenocallis littoralis Australia
CPC 20276 Citrus maxima Thailand
CBS 120486 Citrus maxima Thailand
CBS 123371 Citrus maxima Vietnam
CBS 123372 Citrus maxima Vietnam
CBS 120488 Citrus maxima Thailand
CBS 120487 Citrus maxima China
CBS 123370 Citrus maxima Vietnam
CBS 102374 Citrus aurantium Brazil
CPC 16586 Citrus limon Argentina
CPC 16587 Citrus limon Argentina
CBS 127454 Citrus limon Australia
CPC 16609 Citrus sp. Argentina
CPC 16149 Citrus sp. Argentina
CPC 16149 Citrus sp. Argentina
CBS 127455 Citrus sinensis Australia
CBS 120489 Citrus maxima Brazil
CPC 16606 Citrus limon Uruguay
CPC 16605 Citrus limon Uruguay
CPC 16603 Citrus limon Uruguay
CBS 122482 Citrus sinensis Zimbabwe
CBS 127452 Citrus reticulata Australia
CPC 16152 Citrus sp. South Africa
CPC 16151 Citrus sp. South Africa
CBS 174.77 Cryptomeria japonica USA
CBS 447.68 Taxus baccata Netherlands
CBS 728.79 Podocarpus maki New Zaeland
CBS 777.97 Telopea speciosissima Tasmania
CBS 756.70 Aesculus hippocastanum Germany
CBS 585.84 Acer rubrum USA
CBS 112527 Paxistima myrsinites USA
CBS 447.70 Gaulteria humifusa USA
CBS 111635 Acer rubrum USA
CPC 11336 Eucalyptus globulus Spain
CBS 616.72 Ilex aquifolium Germany
CBS 587.69 Ilex aquifolium Spain
CPC 20274 Mangifera indica Thailand
LGMF 333 Mangifera indica Brazil
LGMF 334 Mangifera indica Brazil
CBS 126270 Mangifera indica Brazil
CPC 17455 Mangifera indica Brazil
CPC 17454 Mangifera indica Brazil
CBS 445.82 Eugenia aromatica Indonesia
CPC 21020 Aloe ferox South Africa
CPC 21023 Aloe ferox South Africa
CPC 21024 Aloe ferox South Africa
CPC 21022 Aloe ferox South Africa
CPC 21021 Aloe ferox South Africa
CBS 535.87 Muehlenbeckia adpressa Australia
CPC 20260 Arecaceae (Palm) Thailand
IMI 260576 Mangifera indica India
CPC 20277 Cordyline fruticosa Thailand
CPC 20261 Cordyline fruticosa Thailand
CBS 123404 Musa paradisiaca Thailand
CPC 17748 Heliconia sp. Thailand
CBS 117118 Musa acuminata Indonesia
CPC 17468 Cymbidium sp. Brazil
CPC 20257 Ficus benjamina Thailand
P. owaniana
P. pseudotsugae
P. podocarpi
G. rhodorae
P. vacciniicola
P. elongata
P. hypoglossi
P. cussonia
P. spinarum
P. concentrica
P. ericarum
P. citribraziliensis
P. bifrenariae
P. hymenocallidicola
P. citrimaxima
P. citriasiana
P. citricarpa
P. foliorum
P. podocarpicola
P. telopeae
P. sphaeropsoidea
P. minima
P. paxistimae
P. gaultheriae
P. rubra
Phyllosticta sp.
P. philoprina
P. mangiferaceae
P. brazillianiae
Phyllosticta sp.
P. eugeniae
P. aloeicola
P. beaumarisii
G. mangiferae
P. cordylinophila

Sequenciando….



COMO FUNCIONA O SEQUENCIAMENTO?

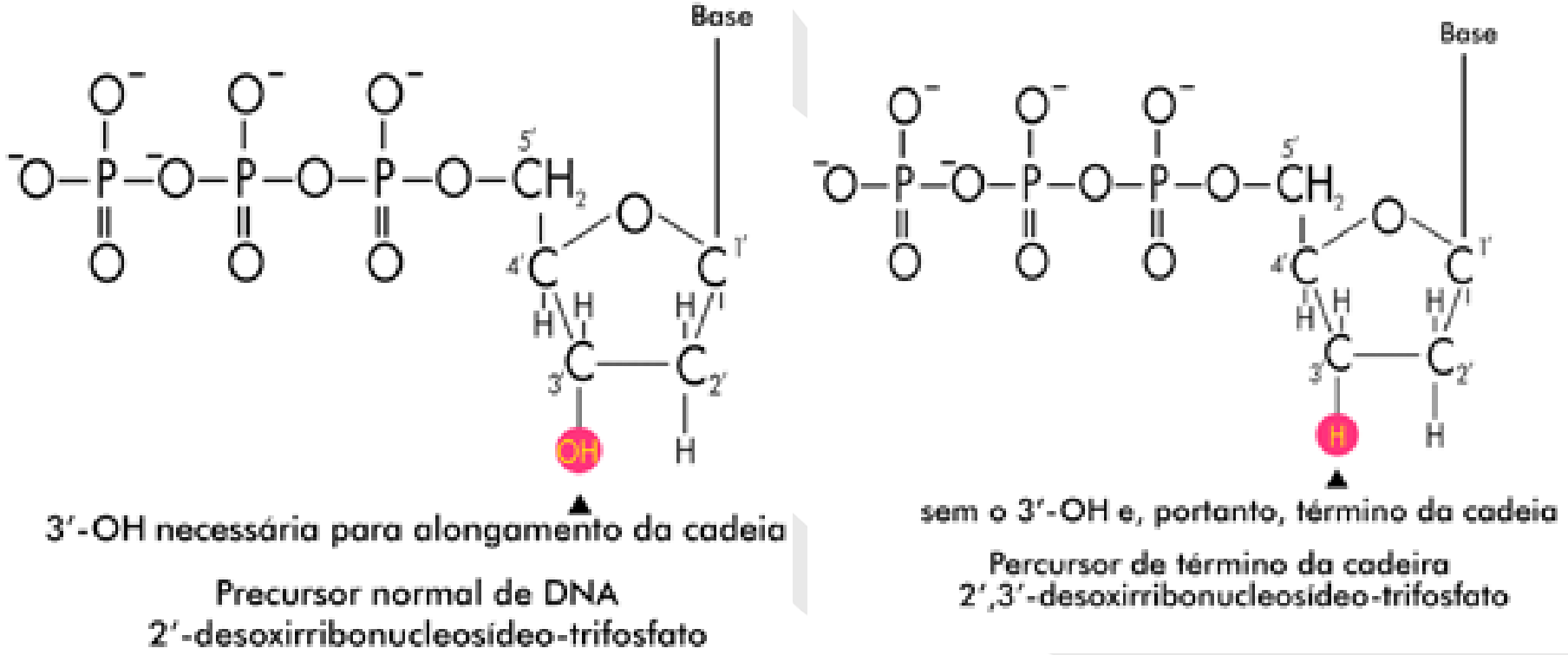
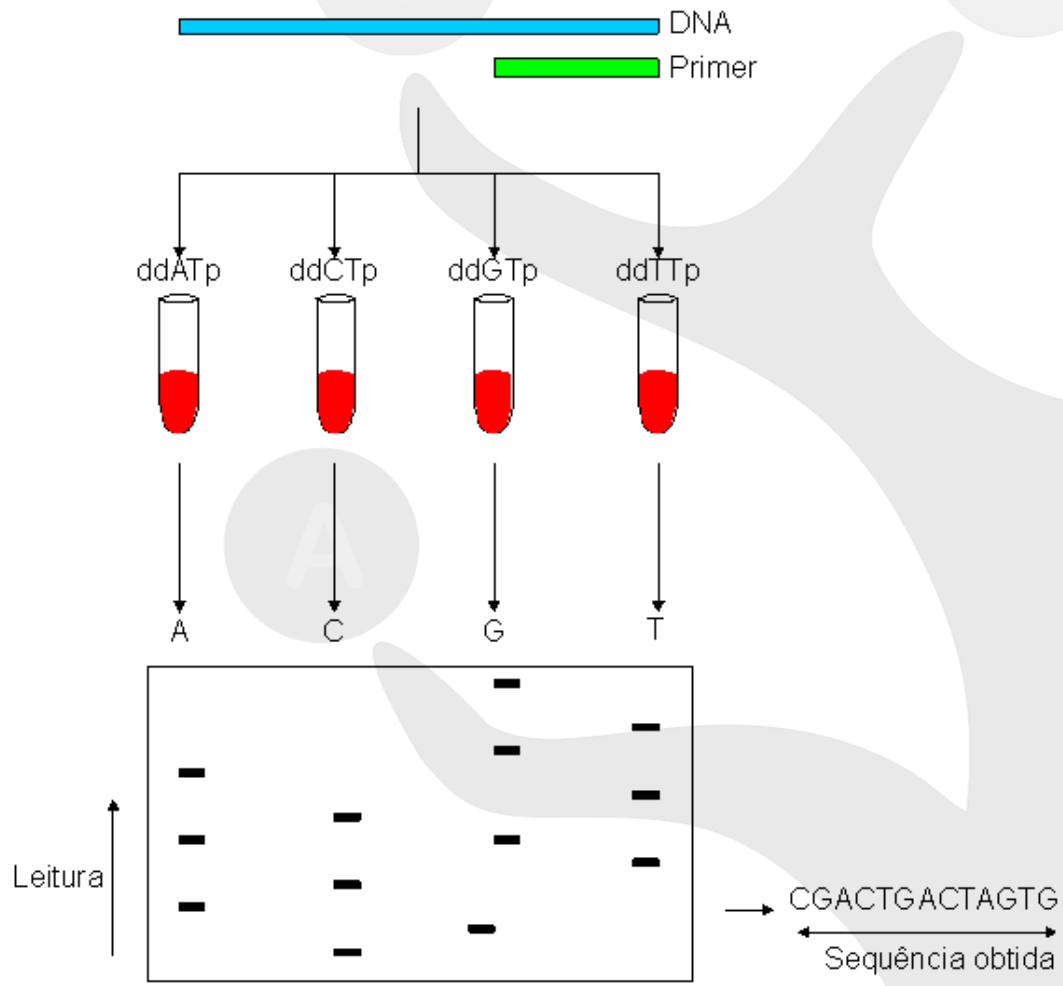


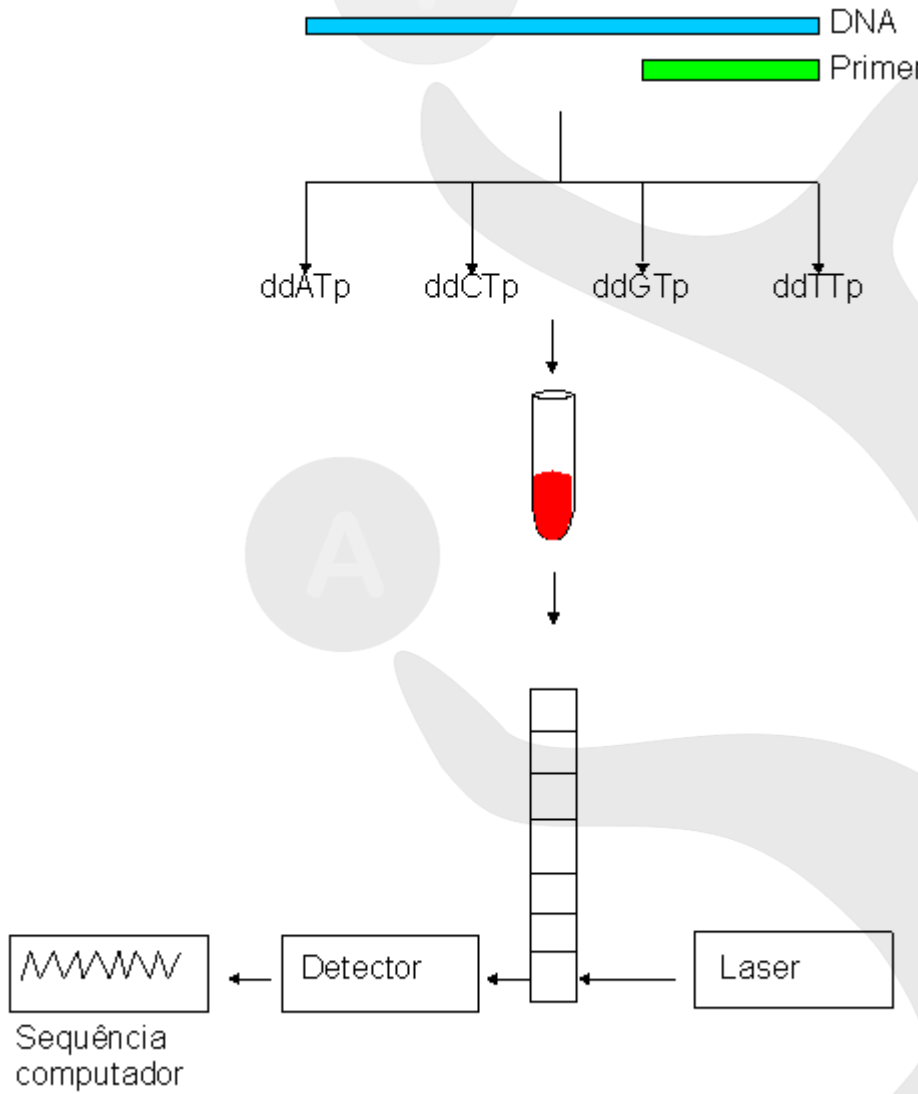
FIGURA 1. Deoxiribonucleotídeos (dNTP) e dideoxiribonucleotídeos trifosfatados (ddNTP).

SEQUENCIAMENTO



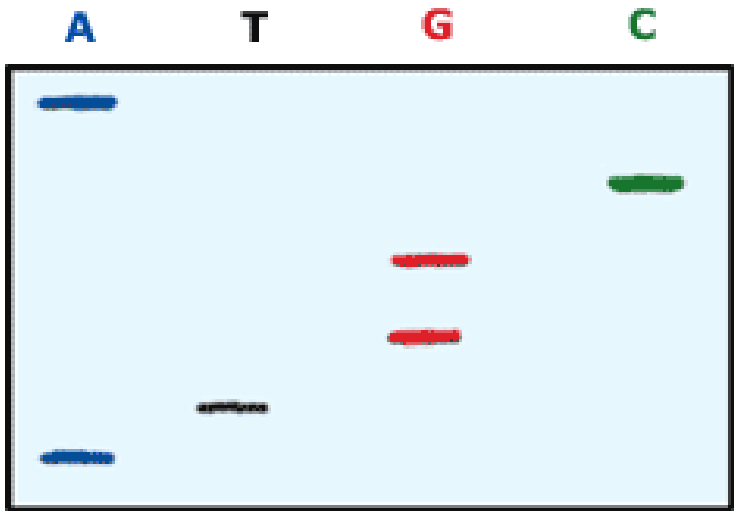
SEQUENCIAMENTO

SEQUENCIAMENTO



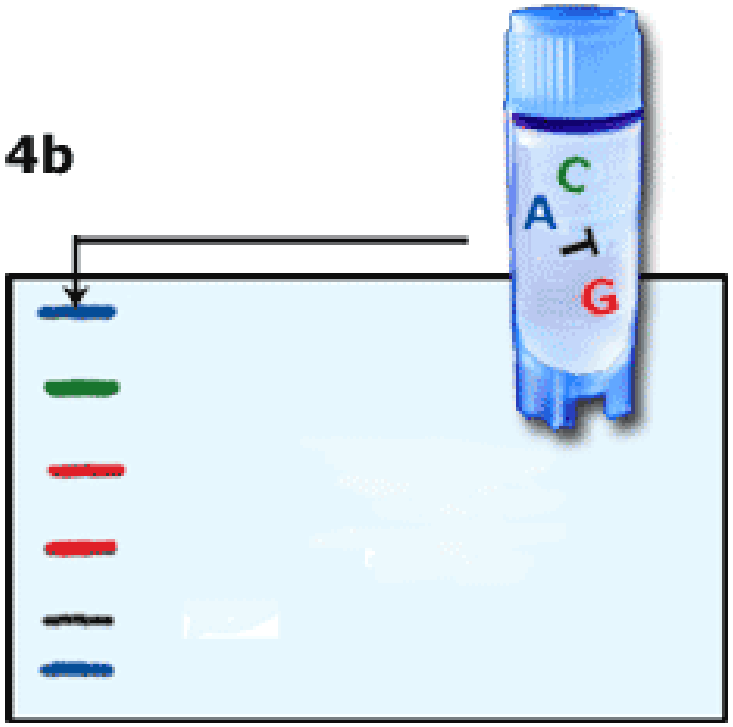
SEQUENCIAMENTO

4a



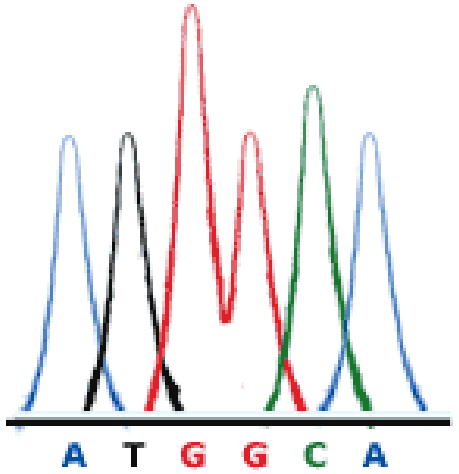
Análise dos produtos fluorecentes de sequenciamento em canais separados do gel de poliacrilamida.

4b



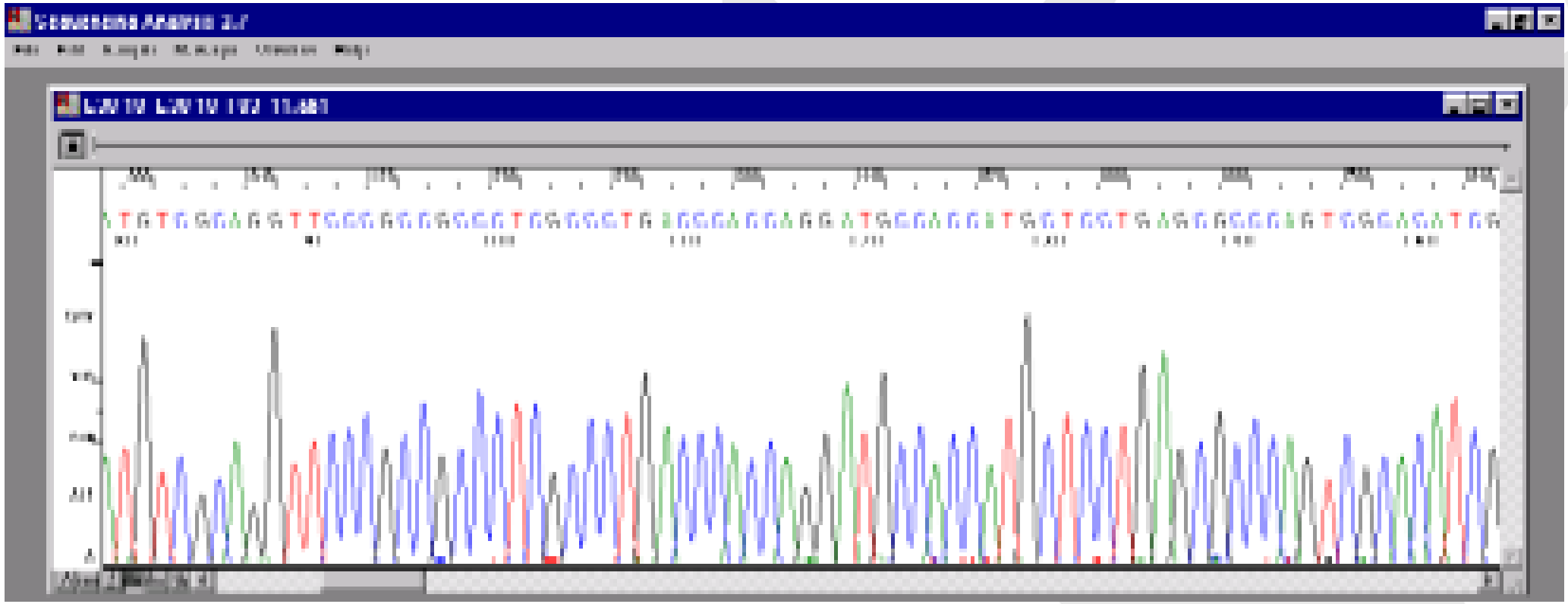
Análise dos produtos fluorecentes de sequenciamento em um único canal do gel de poliacrilamida.

4c



Eletroforegrama do gel

SEQUENCIAMENTO



SEQUENCIAMENTO

```
   1 gtgcccagtg agcgaggact gca
  61 ggggacccga acaacatgga tag
 121 acgttccatg gggccaagga ggt
 181 atgggcctca tatacaacag gat
 241 tgtgccactt gtgaacagat tgc
 301 accaccaatc cactaatcag gca
 361 gctatggaac agatggctgg atc
 421 cagactaggc agatggtaca tgc
 481 ggtctgaaag atgaccttct tga
 541 atgcagcgat tcaagtgatc 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
AGCTATGGACAGACATTAATG
***  *****  ******  *****
```

(B) A more difficult sequence alignment

```
GACGACCATAGACCAGCATAG
GACTACCATAGA-CTGCAAAG
***  *********  ***  **

GACGACCATAGACCAGCATAG
GACTACCATAGACT-GCAAAG
***  *********  ***  **
```