

Lesson 2: An overview of the DNA barcoding workflow

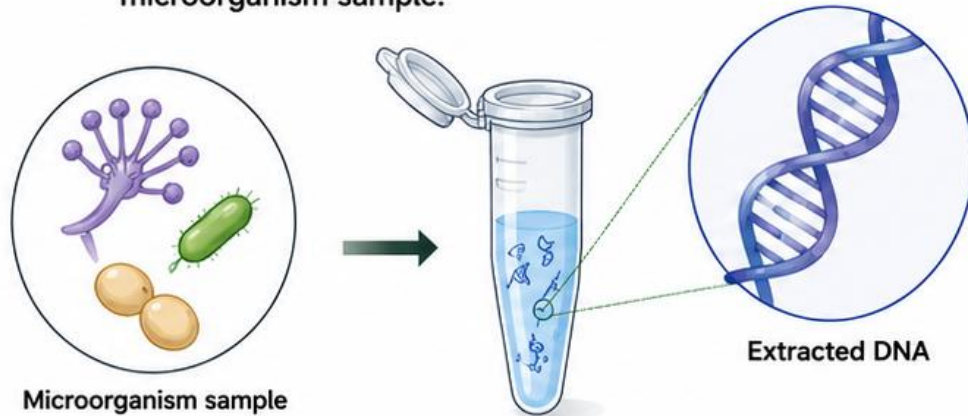
DNA BARCODING: STEPS 1–2

Using a short DNA sequence (barcode) to identify the **species** of an unknown microorganism.

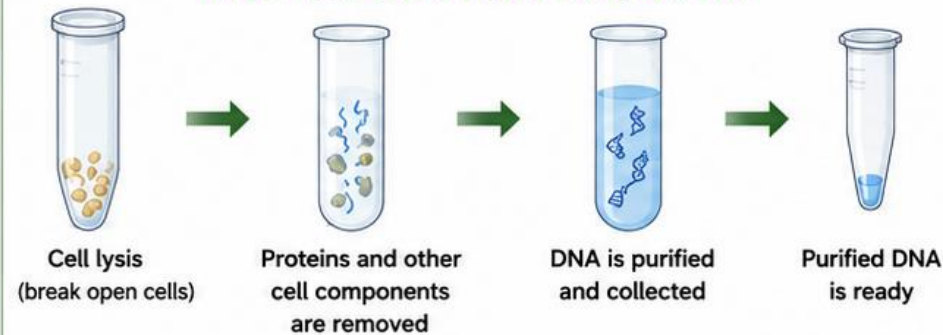


1 EXTRACT DNA.

DNA is extracted from the microorganism sample.



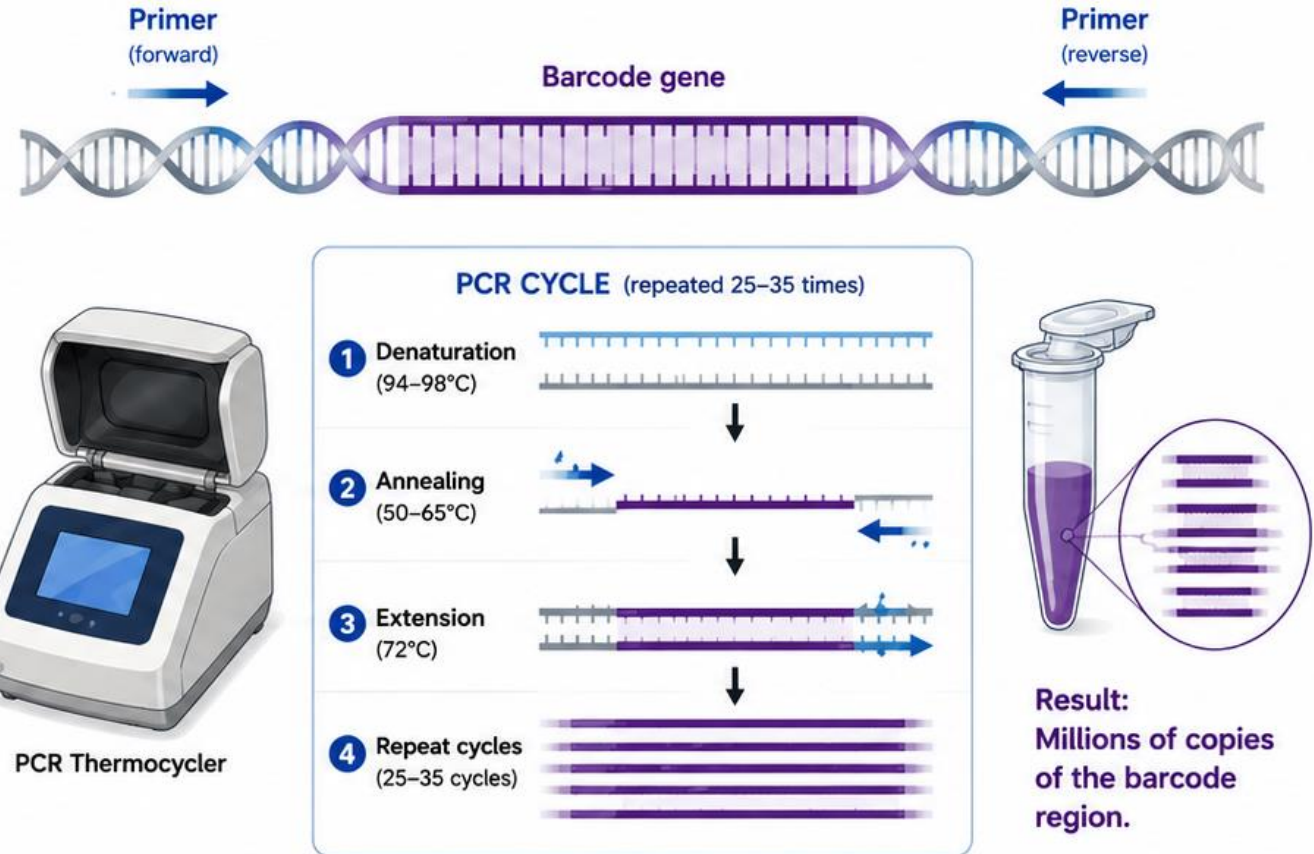
WHAT HAPPENS DURING EXTRACTION?



✓ Result: High-quality DNA free of inhibitors and contaminants.

2 AMPLIFY A STANDARD BARCODE GENE.

The barcode region is amplified using PCR (Polymerase Chain Reaction).



KEY POINT

Primers are short DNA sequences that bind to conserved regions flanking the barcode gene. PCR selectively copies only the barcode region, generating millions of copies for analysis.

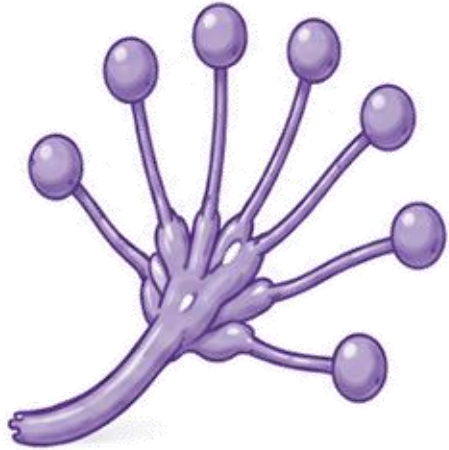
The Internal Transcribed Spacer (ITS) region is the most widely used barcode for identifying fungi.



A GOOD BARCODE GENE IS:

- ① Found in all members of the group
- ② Easy to amplify
- ③ Has conserved regions for primer binding
- ④ Has variable regions to distinguish species

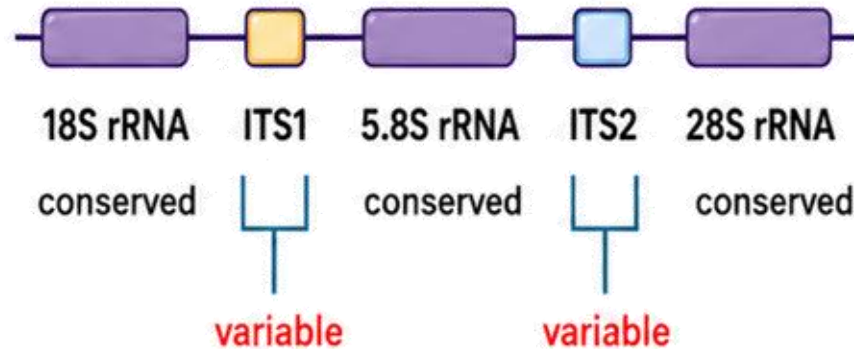
FUNGI



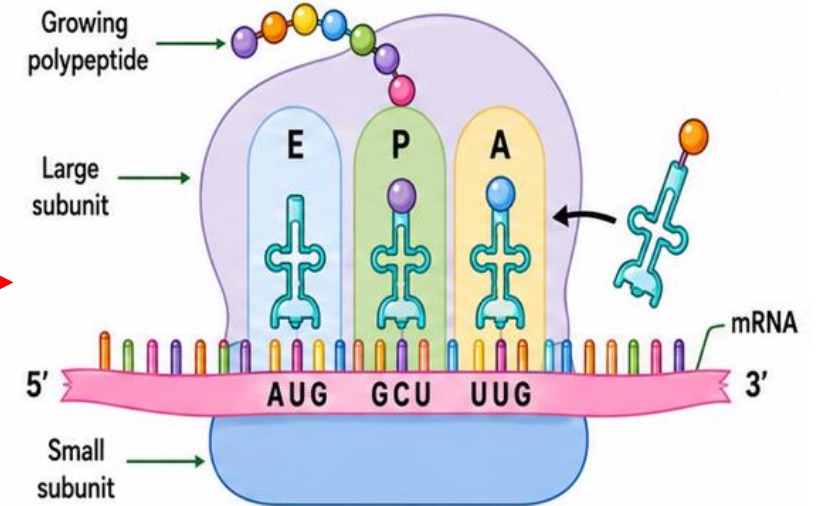
FUNGAL rRNA GENE CLUSTER



ITS (Internal Transcribed Spacer)



RIBOSOME: MAKING PROTEINS



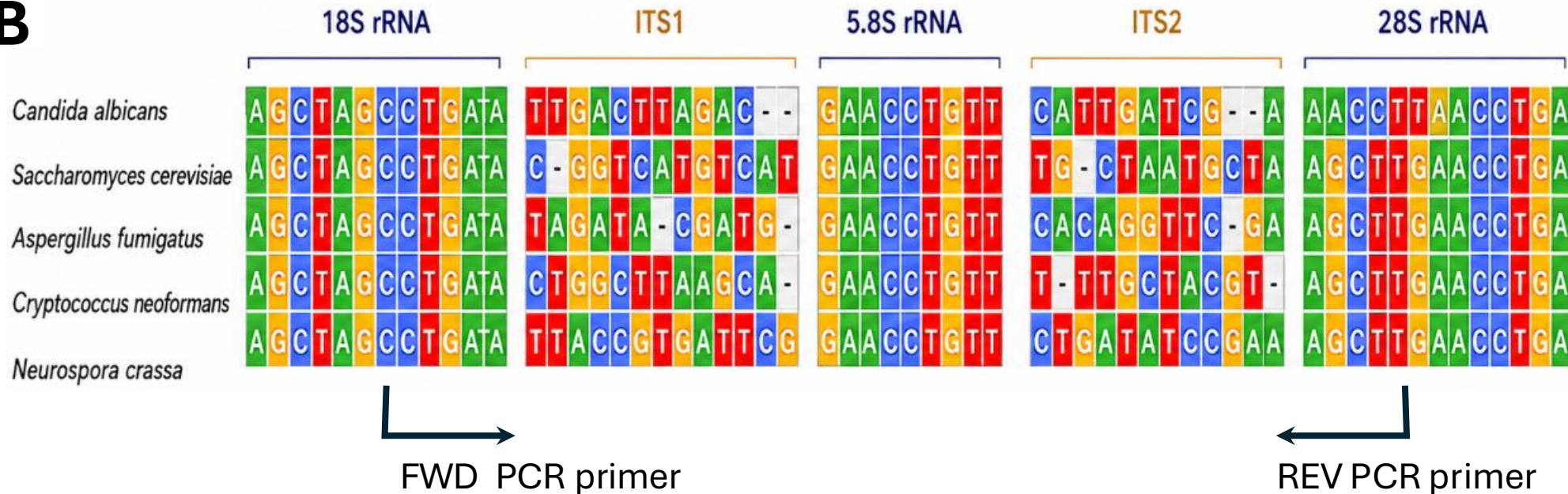
Structural Organization of Ribosomal RNA Genes in Fungi

Conserved regions are very similar across species; variable regions show differences

A



B



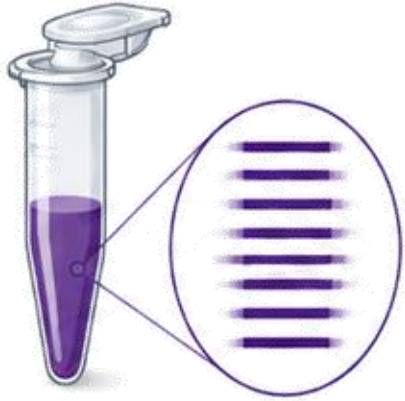
A Adenine T Thymine C Cytosine G Guanine - Gap (insertion or deletion)

DNA BARCODING: STEPS 2-5

Using a short DNA sequence (barcode) to identify the species of an unknown microorganism.



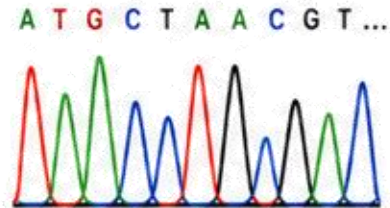
2 EXTRACT & AMPLIFY THE BARCODE REGION



PCR amplifies the ITS region from the fungal DNA.

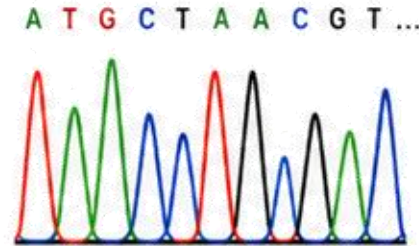
Result:
Millions of copies of the barcode region.

3 SEQUENCE THE BARCODE



The amplified DNA is sequenced to determine the order of nucleotides.

4 OBTAIN THE DNA SEQUENCE



DNA sequence

The chromatogram is analyzed to generate the DNA sequence (barcode).

5 COMPARE THE SEQUENCE TO A REFERENCE DATABASE

GenBank	
Species	Match (%)
<i>Aspergillus flavus</i>	99.8%
<i>Aspergillus oryzae</i>	95.1%
<i>Aspergillus niger</i>	93.0%
<i>Aspergillus terreus</i>	91.2%
Others	< 90%

The sequence is compared with sequences in databases such as GenBank.

Result

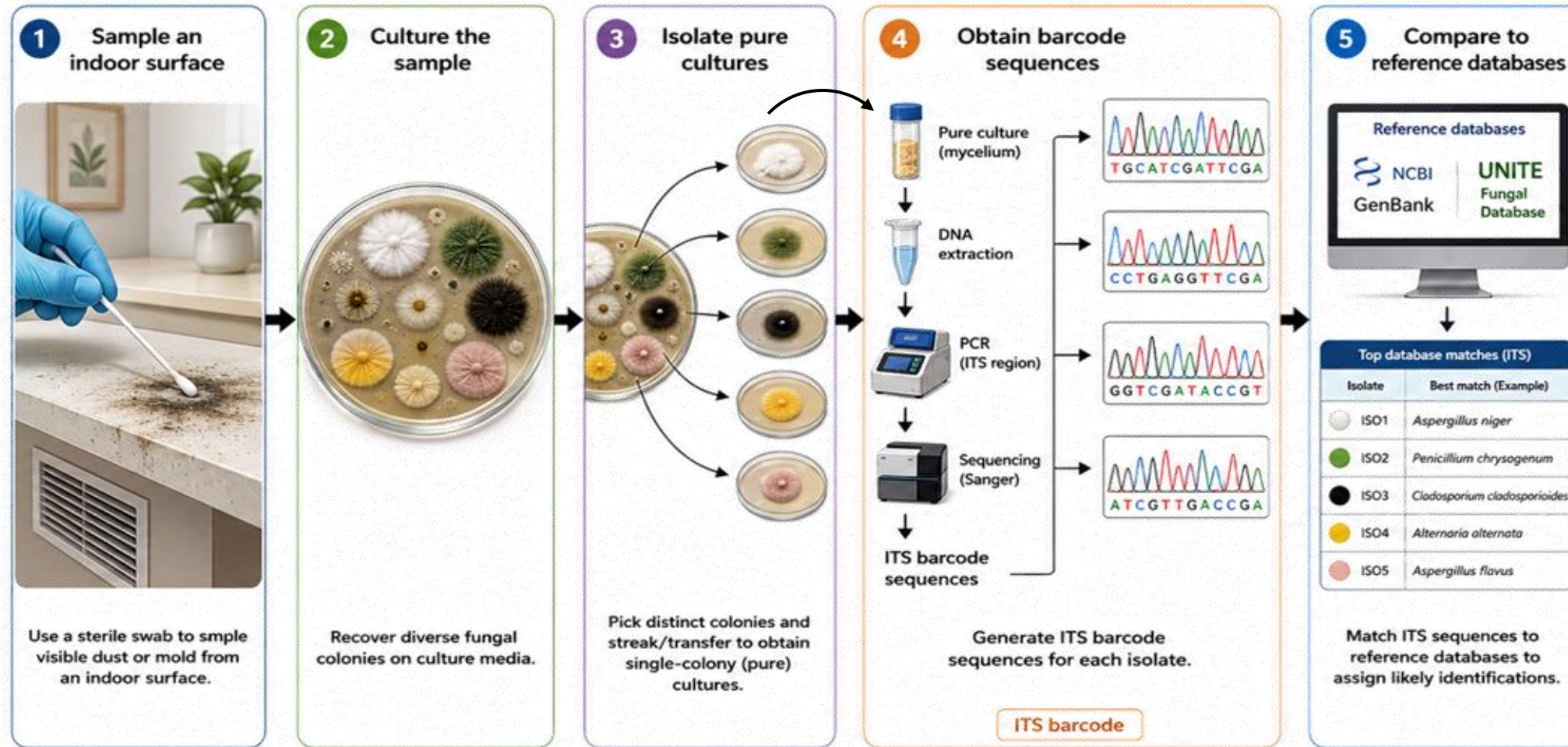


99.8% match to
Aspergillus flavus

Species identified

The ITS sequence from the unknown fungus matches 99.8% with *Aspergillus flavus* in the database. Therefore, the unknown organism is identified as *Aspergillus flavus*.

From indoor surface sampling to a broad but summarized description of the detected fungal isolates.



- Database searches identify the closest known relatives for each unknown isolate, generating **a large and complex sequence dataset**.
- Multiple sequence alignment (MSA)** comprehensively evaluate the dataset, enabling the construction of **phylogenetic trees** that summarize genetic relationships among fungal isolates and simplify data interpretation.

How DNA Sequence Alignment Works

1 Multiple sequence alignments (MSAs) line up DNA sequences from many organisms.

This puts similar positions (letters) in the same columns so they can be compared.

2 MSAs help identify matching sequences in databases.

An unknown sequence is aligned with database sequences. The best match(es) indicate the most likely identity or closest relatives.

3 The alignment shows conserved (same) and variable (different) positions.

This information reveals similarities and differences across sequences.

4 MSAs detect mutations and evolutionary changes.

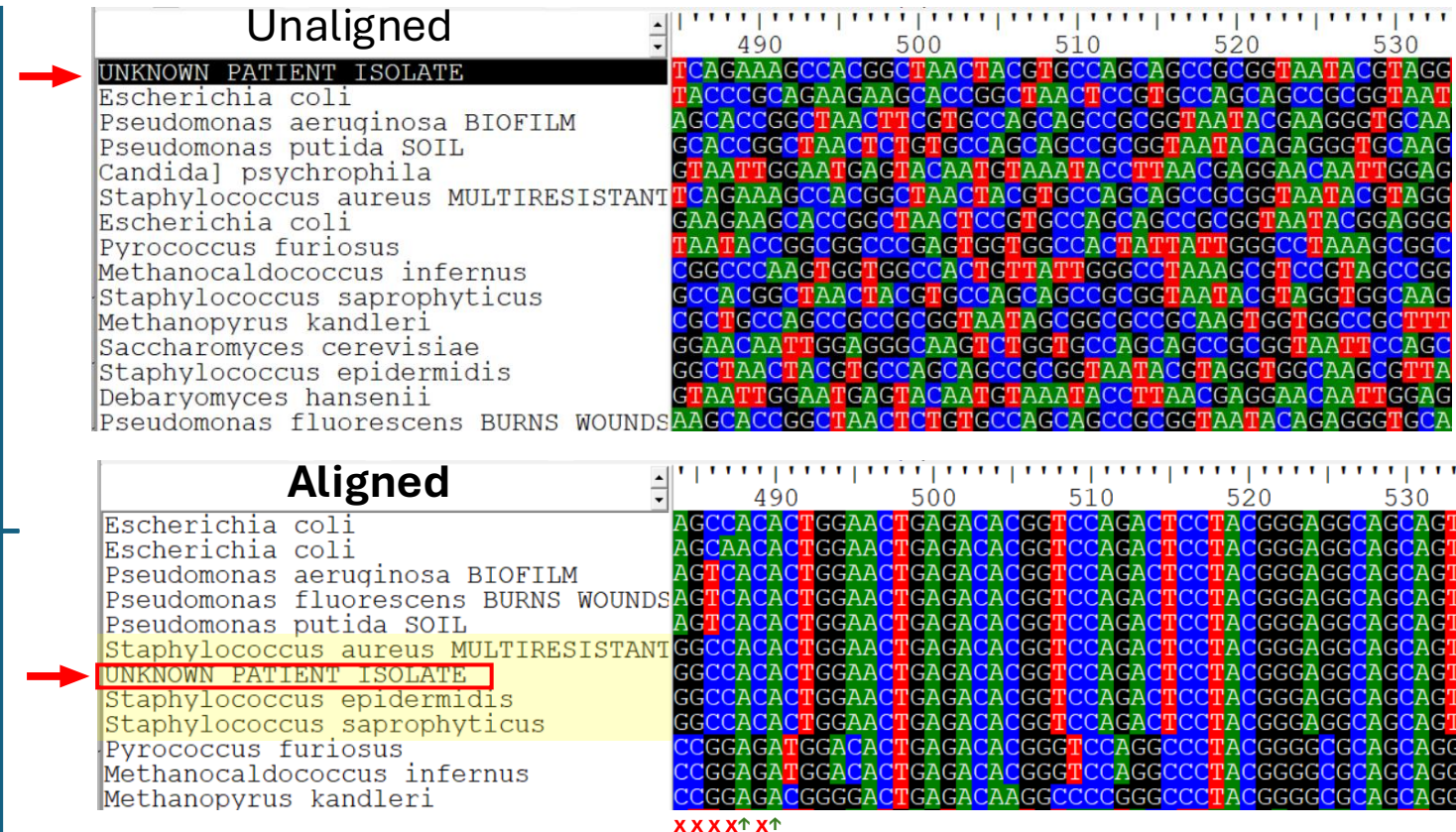
Different letters in a column represent mutations or natural variation over time.

5 Aligned sequences can be used to build phylogenetic trees.

The alignment provides the data needed to calculate evolutionary relationships.

6 Phylogenetic trees translate alignment information into a tree diagram.

This makes it easy to visualize how organisms are related and to communicate the results.



↑ = Conserved columns (similar)

Same letter in most or all sequences.
Often indicates important function.

X = Variable columns (different)

Different letters indicate mutations
or natural variation.

Analysis of genetic
relationships

