

BIOINFORMATIC TOOLS FOR THE MOLECULAR IDENTIFICATION OF BACTERIA

Dr. Hong Kar Wai

Bacterial Identification

Chromogenic media

- Growth on selection media and selection media
- Colony morphology

Microscopy Techniques

- Differential staining
- Bright field,
- dark field,
- Scanning Electron Microscopy (SEM)
- Transmission Electron Microscopy (TEM)
- Confocal laser scanning microscopy (CLSM)

Biochemical Techniques

Classical approaches

- Production of specific enzyme
- Utilization of specific substrates
- Antibiotic susceptibility test
- Dichotomous Identification Keys

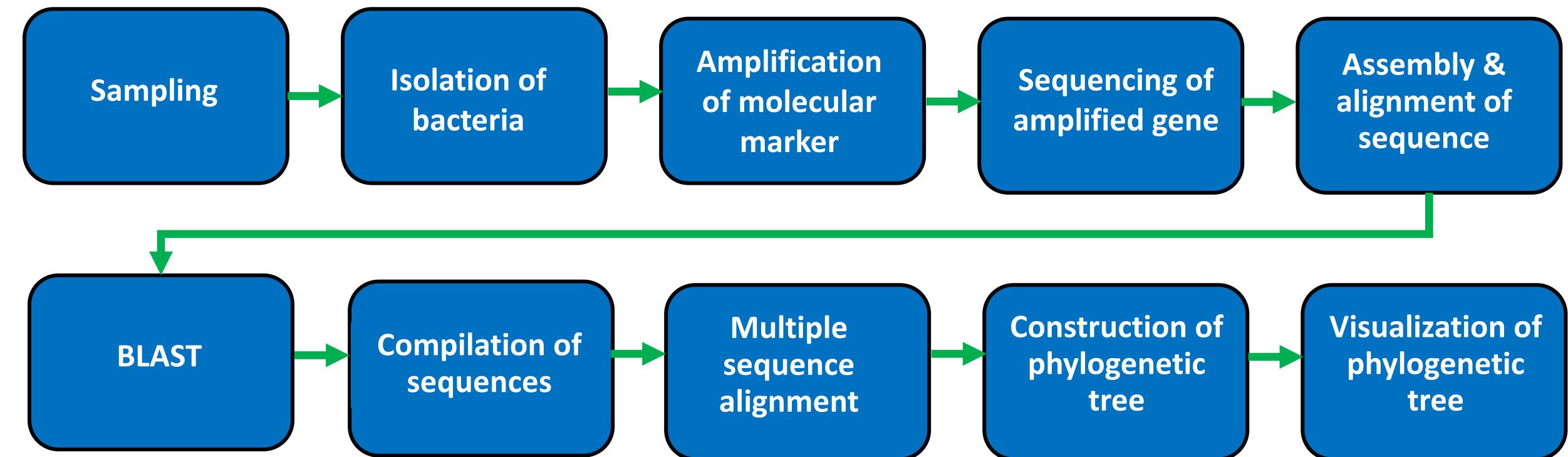
Modern approaches

- Fatty acid fingerprinting
- Serological/ Immunological Identification: ELISA-based methods
- MALDI-ToF (matrix-assisted desorption/ionization time-of-flight) mass spectrometry (MS)
- Fourier Transform Infrared (FTIR) spectroscopy
- Raman spectroscopy

Molecular Techniques

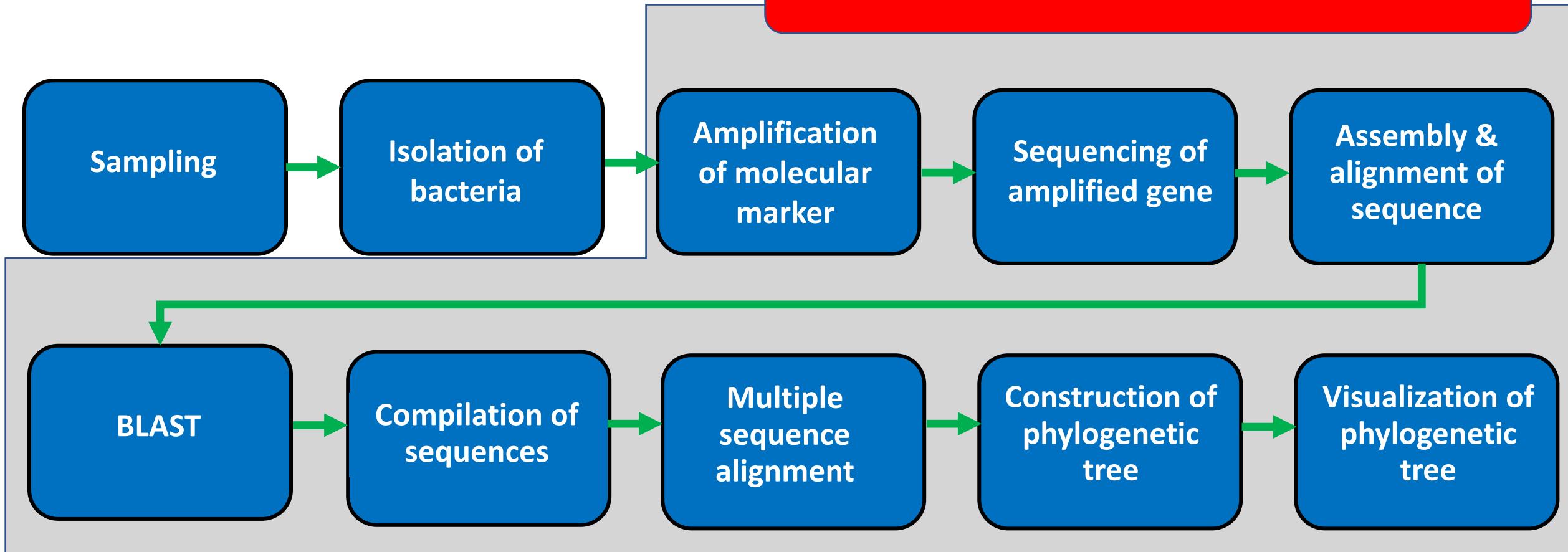
- Restriction Fragment Length Polymorphism (RFLP)
- Pulsed-field Gel Electrophoresis (PFGE)
- PCR
- Rapid Sanger sequencing
- Ribotyping
- Real time qPCR
- Whole genome sequencing (WGS)

Experimental Design

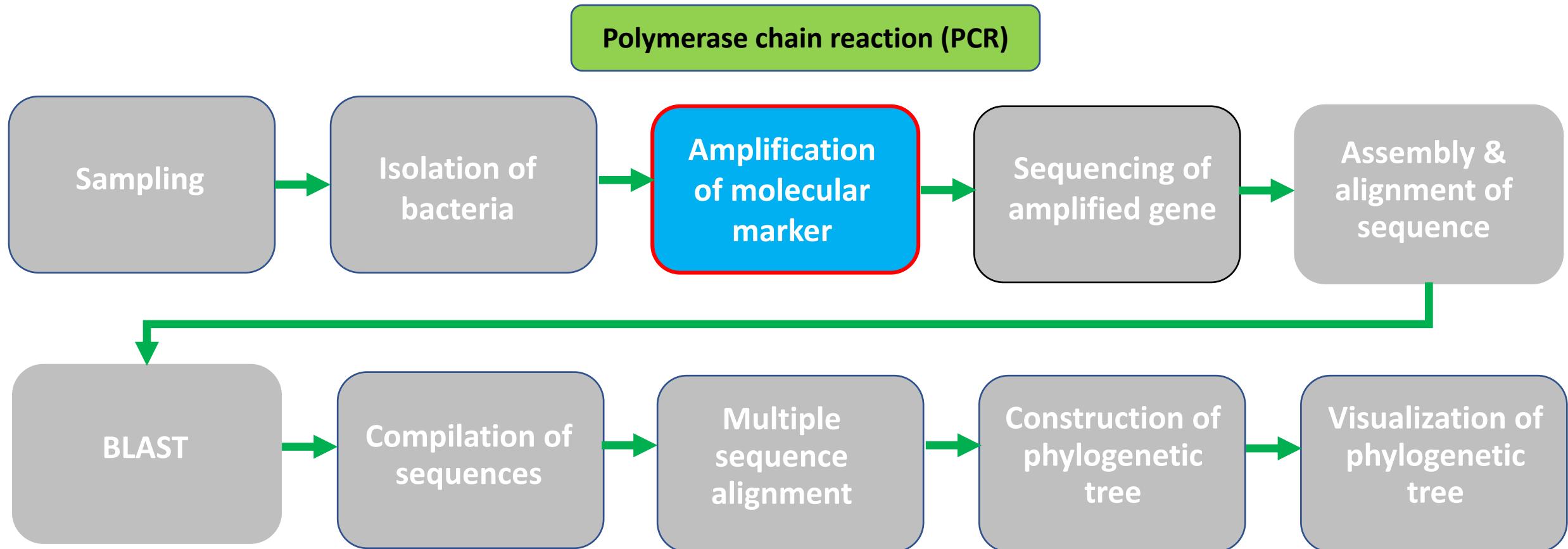


Experimental Design

Molecular identification of bacteria



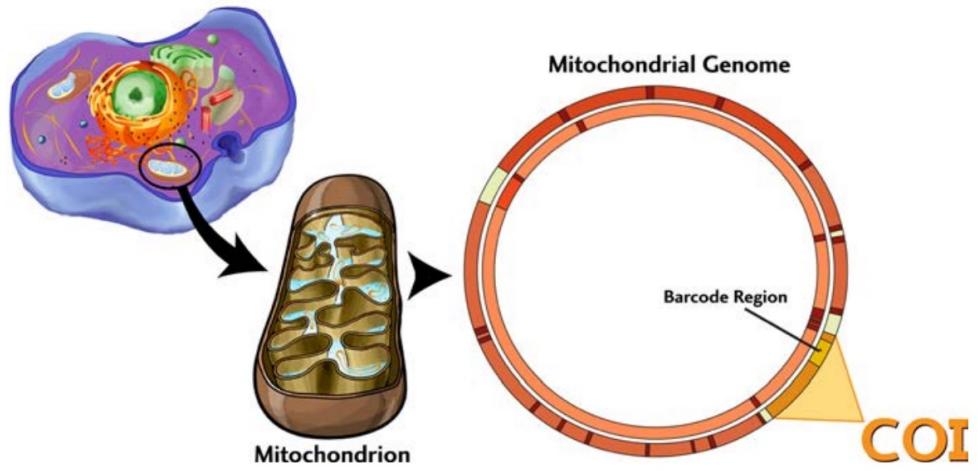
Amplification of molecular marker



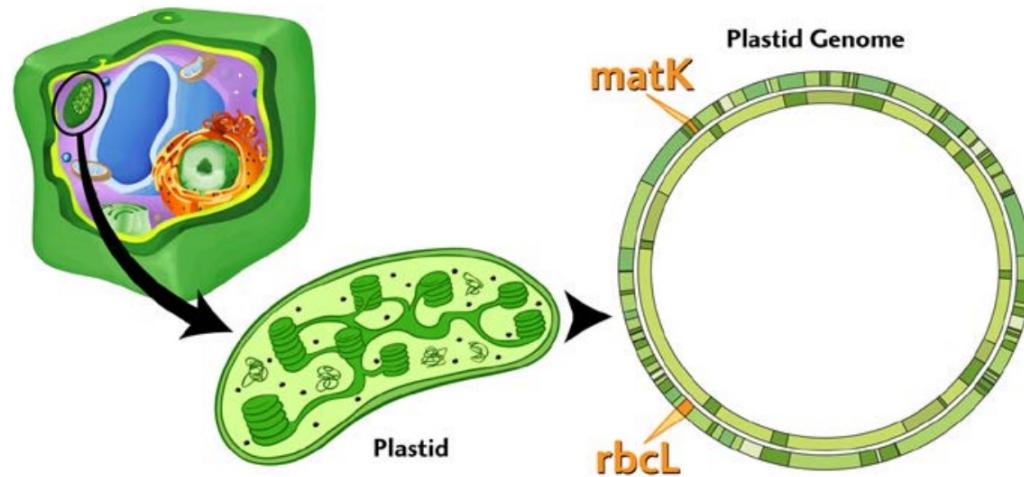
Question

- Given the molecular technique used for the bacterial identification is **polymerase chain reaction (PCR)** followed by gene sequencing and sequence analysis. What is the target gene that you are going to amplify? Justify your answer.

Animal Cells



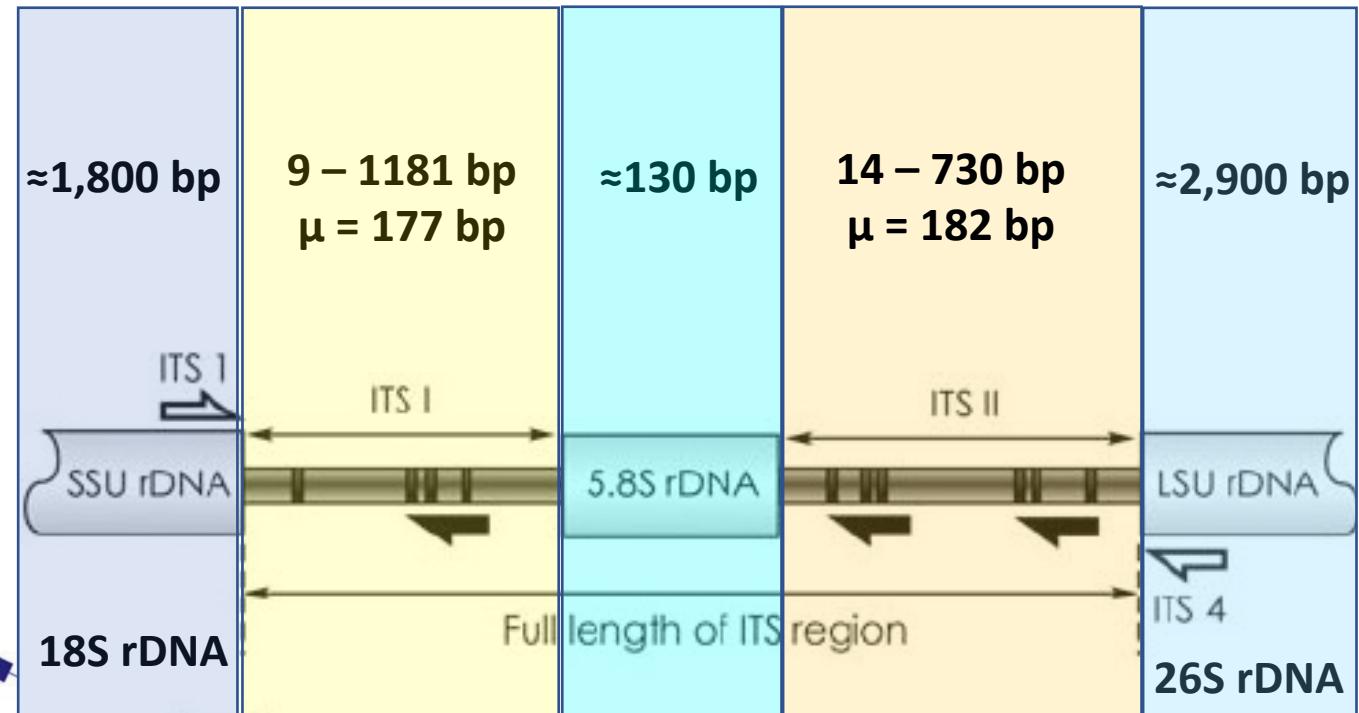
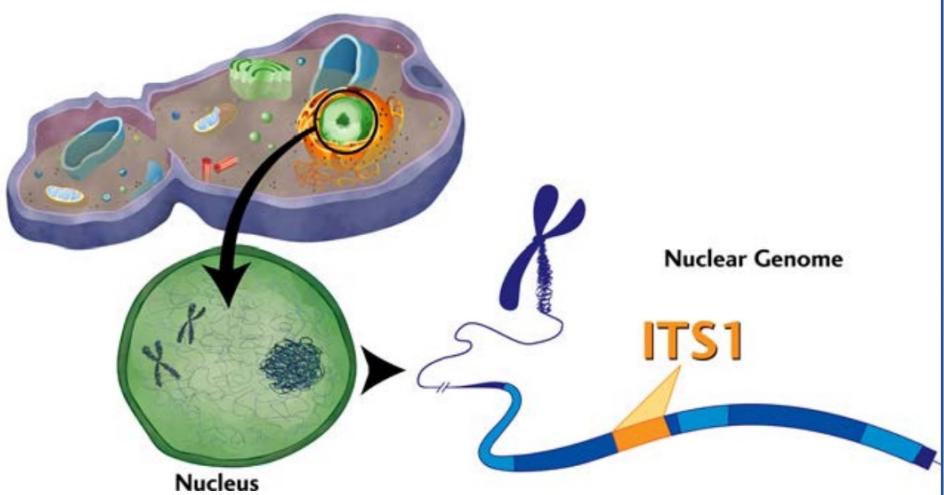
Plant Cells



1. **Cytochrome c oxidase subunit I**
(COI)

1. **Ribulose-1,5-bisphosphate carboxylase-oxygenase (rbcL/**
RuBisCo/ rubisco/ RuBPCase/ RuBPco)
2. **Maturase K (matK)**

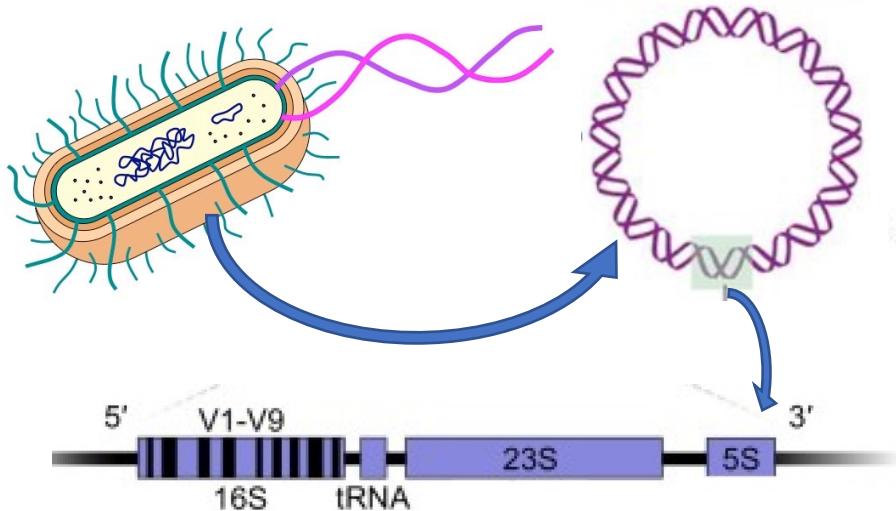
Fungal Cells



1. Internal Transcribed Spacer
1 (ITS1) region

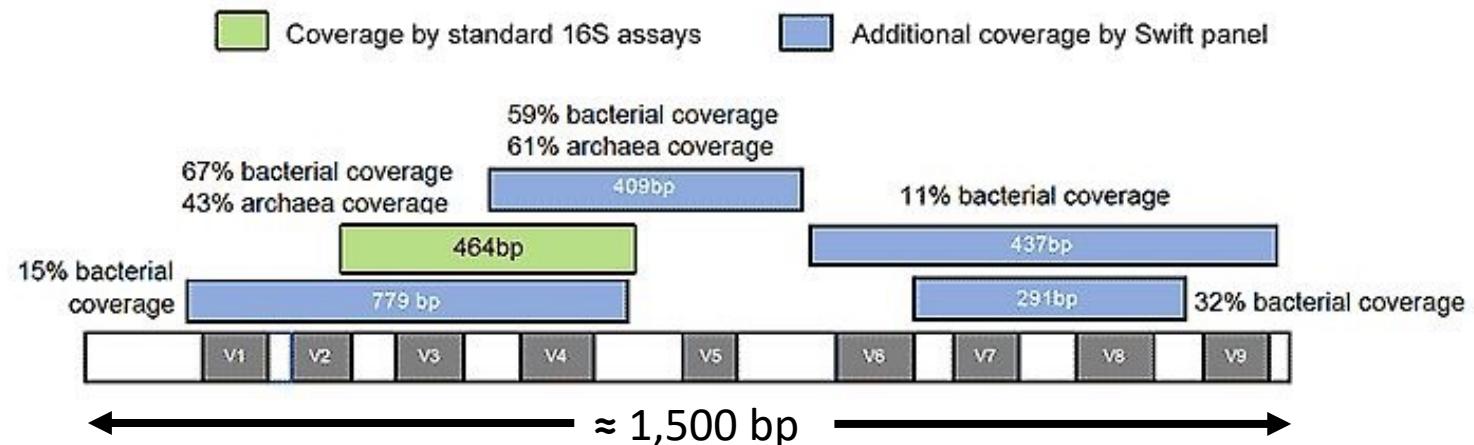
Conservative region
 Variable region
 Species-specific sequence
 Universal primer
 Species-specific primer

Bacterial Cells



1. 16S ribosomal RNA (16S rRNA)

Amplicon coverage of 16S rRNA gene variable regions



widely **conserved** across the prokaryotes

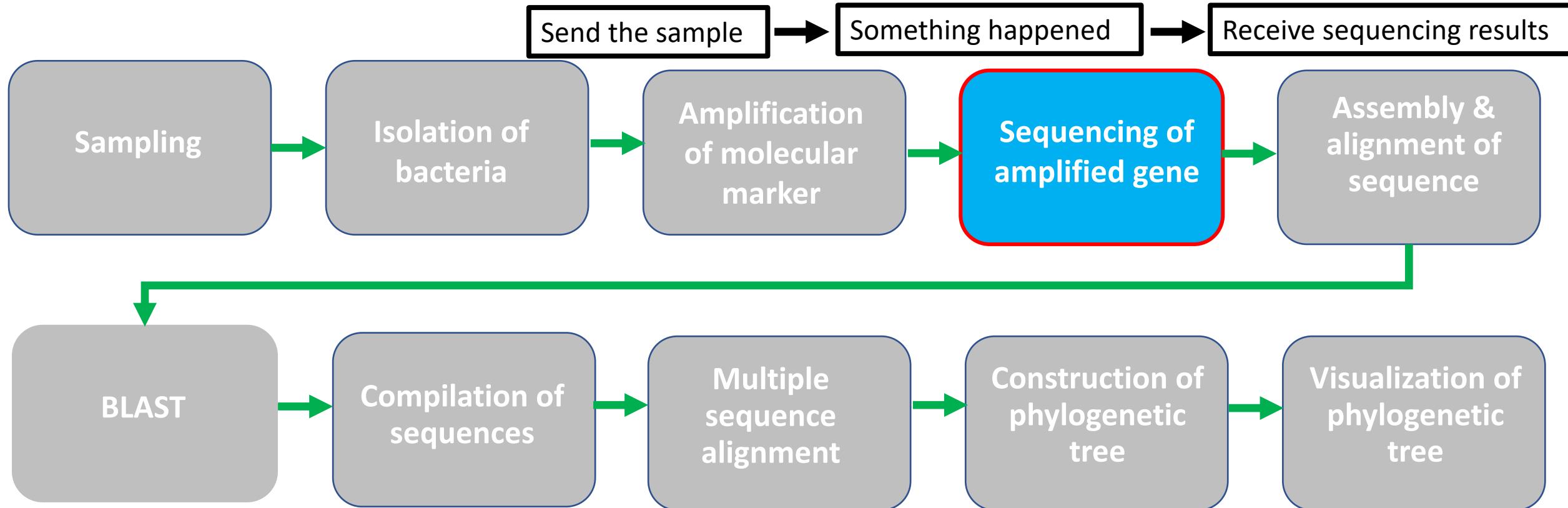
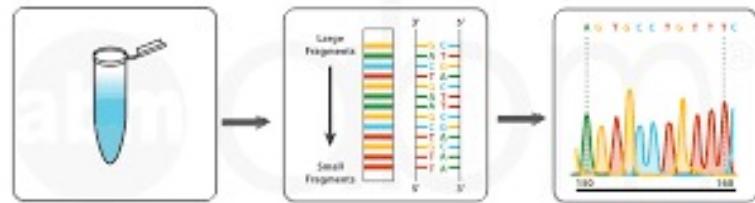
Besides being highly conserved, it also contains many **variable regions** that varies across lineages or even individual bacterial species

well-studied and characterized genes

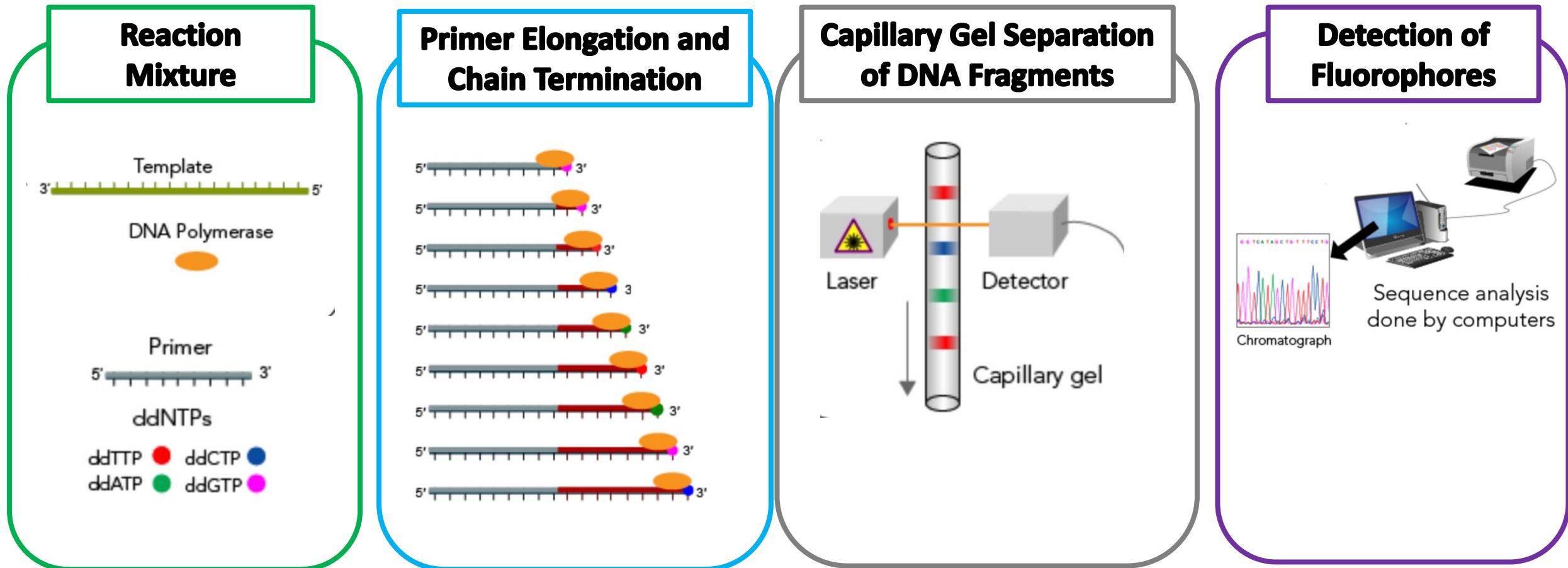
Universal primers

Databases

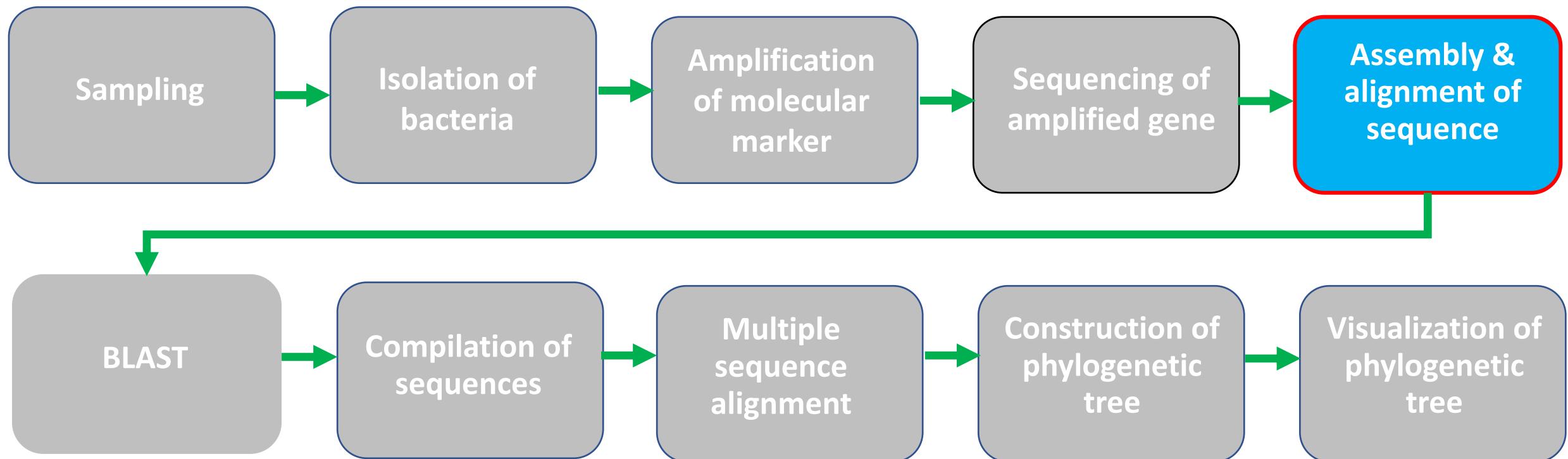
Sequencing of amplified gene



Sequencing of Amplified Genes



Assembly & alignment of sequence

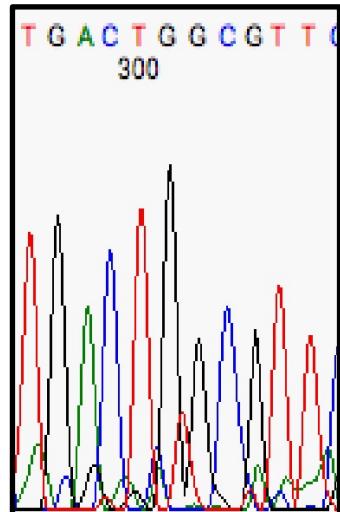


Interpretation of Sequencing Chromatograms

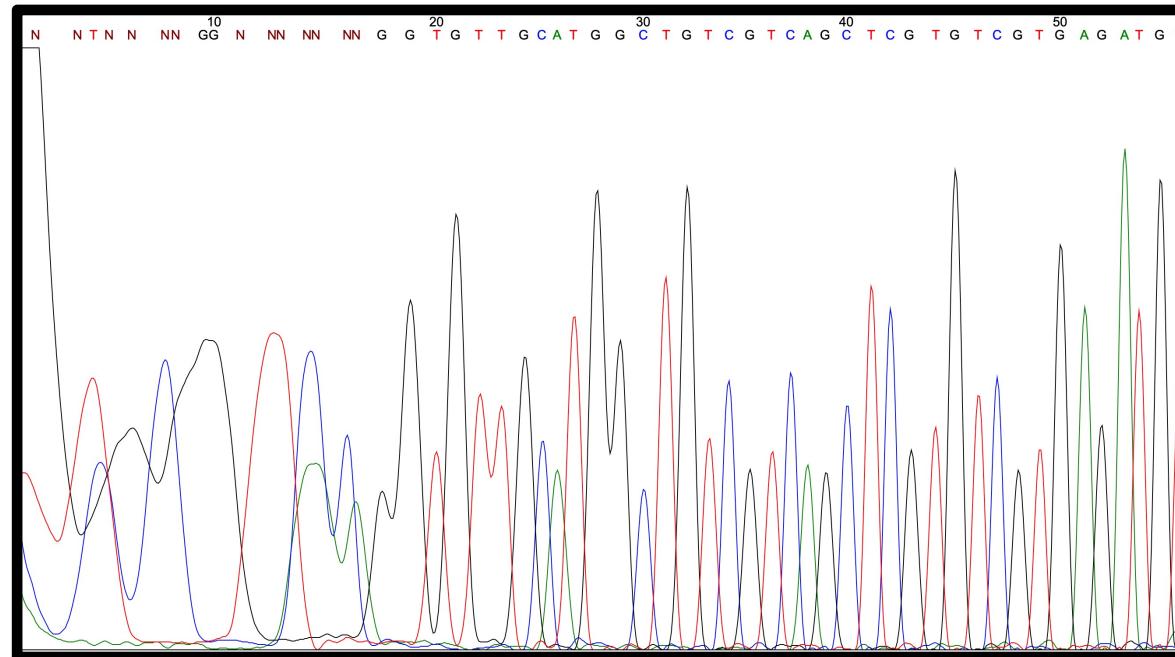
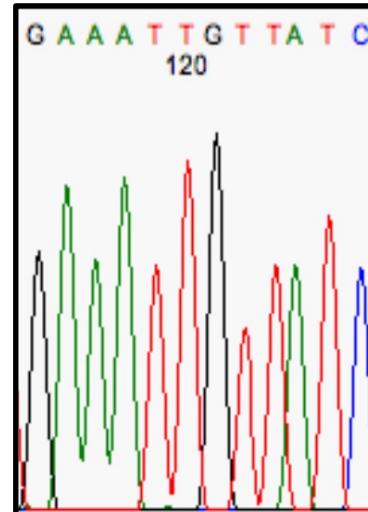
Low Quality Base Calling

Normally, 20-30 bases of a DNA sequencing read will be discarded due to low quality of base calling

(a)



(b)



Peaks

individual, sharp and evenly spaced peaks

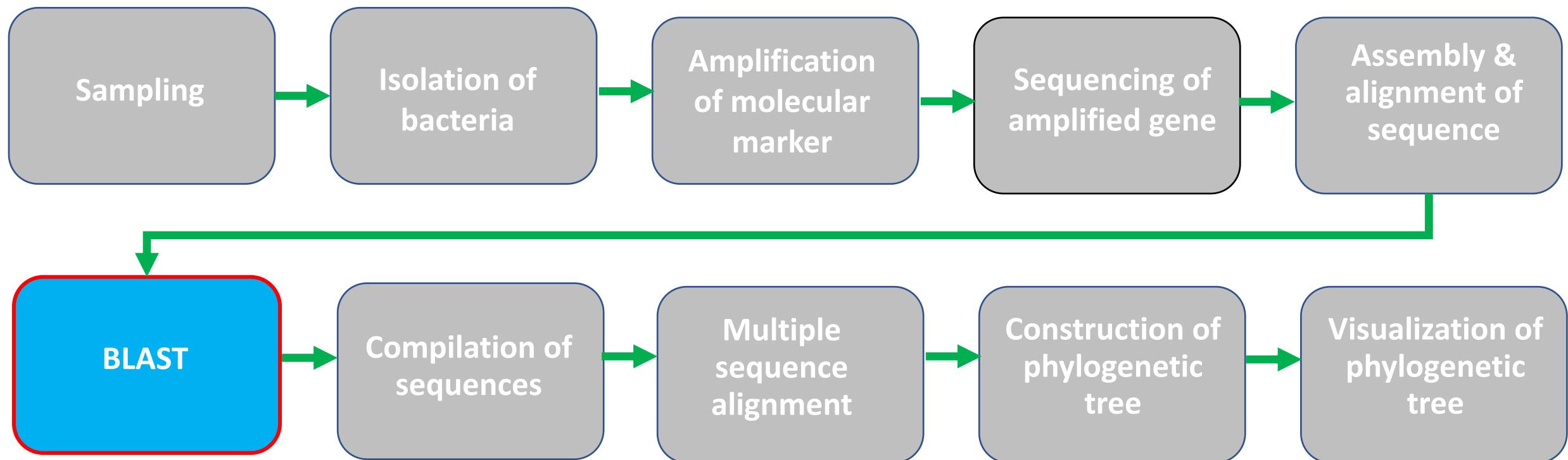
Length

The DNA sequence should be slightly short or similar with the length of amplified genes

Background Noise

Low level of background noise

Basic Local Alignment Search Tool (BLAST)



Basic Local Alignment Search Tool (BLAST)

Program	Database (Subject)	Query	Typical uses
BLASTN	Nucleotide	Nucleotide	<ul style="list-style-type: none">• Mapping oligonucleotides, cDNAs, and PCR products to a genome• screening repetitive elements• cross species sequence exploration• annotating genomic DNA• clustering sequencing reads• vector clipping
BLASTP	Protein	Protein	<ul style="list-style-type: none">• Identifying common regions between proteins• collecting related proteins for phylogenetic analyses
BLASTX	Protein	Nucleotide translated into protein	<ul style="list-style-type: none">• Finding protein-coding genes in genomic DNA• determining if a cDNA corresponds to a known protein
TBLASTN	Nucleotide translated into protein	Protein	<ul style="list-style-type: none">• Identifying transcripts, potentially from multiple organisms, similar to a given protein• mapping a protein to genomic DNA
TBLASTX	Nucleotide translated into protein	Nucleotide translated into protein	<ul style="list-style-type: none">• Cross-species gene prediction at the genome or transcript level• searching for genes missed by traditional methods or not yet in protein databases

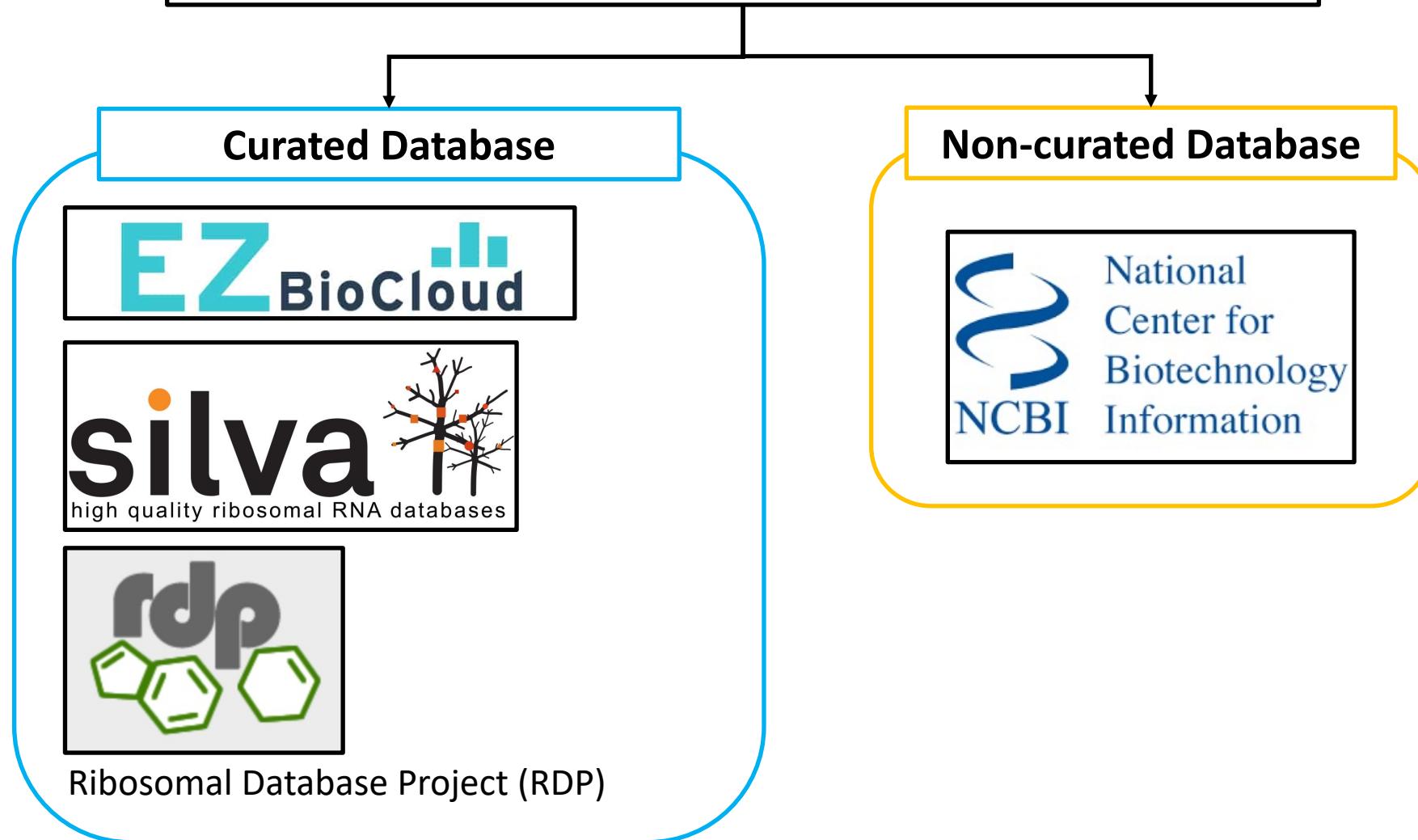
BLAST Databases



Database	Content
nt	The nucleotide sequence database contains entries from traditional divisions of GenBank, EMBL and DDBJ.
nr	A collection of protein sequences with entries from: <ul style="list-style-type: none">• GenPept,• Swiss-Prot,• Protein Data Bank (PDB),• Protein Research Foundation (PRF),• Protein Information Resource (PIR) and• NCBI Reference Sequence (RefSeq) project.
16S_ribosomal_RNA	Microbial 16S RNA sequences from the RefSeq Targeted Loci project
ITS	Databases with collection fungal or eukaryotic Internal Transcribed Spacer sequences.
LSU_rRNA	Database with large submit rRNA sequences for prokaryotes and eukaryotes.
SSU_rRNA	A database with sequences small from fungi and eukaryotes
human_genome	Current refseq human genome assembly (GRCh) with various database
refseq_euk_rep_genomes	Eukaryotic representative genomes from NCBI RefSeq project

BLAST Databases

Databases for 16S Ribosomal RNA Sequences



Interpreting BLAST Results

Job Title Query

RID 50TA379D013 Search expires on 03-17 15:58 pm Download All

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID lcl|Query_19725

Description Query

Molecule type dna

Query Length 1422

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity [] to [] E value [] to [] Query Coverage [] to []

Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download New Select columns Show 100

select all 100 sequences selected

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Uncultured bacterium clone ER-EC-12_06H 16S ribosomal RNA gene, partial sequence	NA	2617	2617	99%	0.0	99.93%	1496	MW686508.1
<input checked="" type="checkbox"/>	Uncultured bacterium clone ER-EC-11_07A 16S ribosomal RNA gene, partial sequence	NA	2612	2612	99%	0.0	99.86%	1517	MW686432.1
<input checked="" type="checkbox"/>	Uncultured bacterium clone ER-EC-13_upper Bact_05G 16S ribosomal RNA gene, partial sequence	NA	2612	2612	99%	0.0	99.86%	1497	MW652349.1
<input checked="" type="checkbox"/>	Uncultured Xanthomonadaceae bacterium clone DR938CH110701SACH44 16S ribosomal RNA gene, partial sequence	NA	2612	2612	99%	0.0	99.86%	1500	DQ230964.1
<input checked="" type="checkbox"/>	Uncultured bacterium clone 7-166 16S ribosomal RNA gene, partial sequence	NA	2606	2606	99%	0.0	99.79%	1500	KC170377.1
<input checked="" type="checkbox"/>	Silanimonas sp. JK12 16S ribosomal RNA gene, partial sequence	NA	2549	2549	97%	0.0	99.86%	1387	KF20636

Expect value:

The measure of likeliness that sequence similarity is not by random chance
(The smaller the better)

All scores and percentages, the greater the better

Maximum score:

The highest alignment score (bit-score) between the query sequence and the database segments

Total score:

The sum of the alignment scores of all sequences from the same database

Percentage of query coverage:

The percent of the query length that is included in the aligned segments

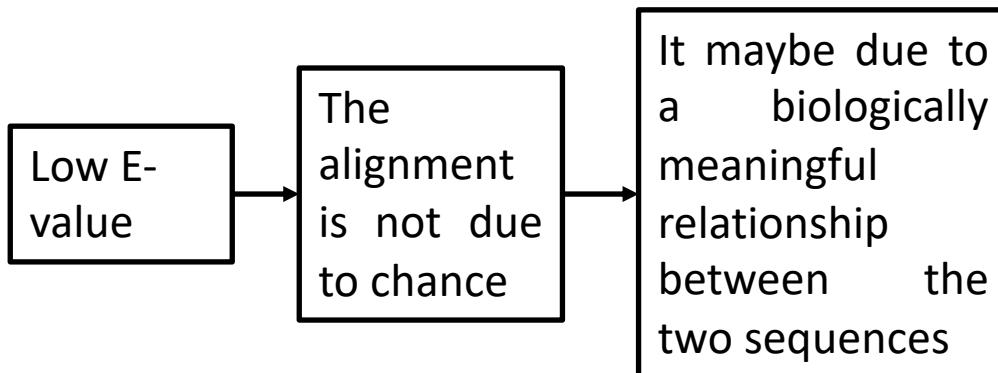
Percentage of Identity:

The degree of similarity measured in percentage between the query and the aligned sequences (subject).

Interpreting BLAST Results

5. Expect Value

Is this alignment meaningful?



The higher the score, the lower the E-value

How low the E-value is needed in order to conclude both sequences are homologous?

No single answer. E-value depends on query and the length of subjects in the database

How are results sorted when multiple hits have the same E-value and score?

The sorting is random.

What is the reported E-value of 0.0?

Calculated E-value < 1e-180

Does low E-value indicate both the sequences are homologous?

BLAST does not measure homology, and E-value does not represent the level of homology. However, one can infer homology from an alignment with low E-value

Interpreting BLAST Results

CCCCGGGC
||| |||
CCCCGGGC

Match

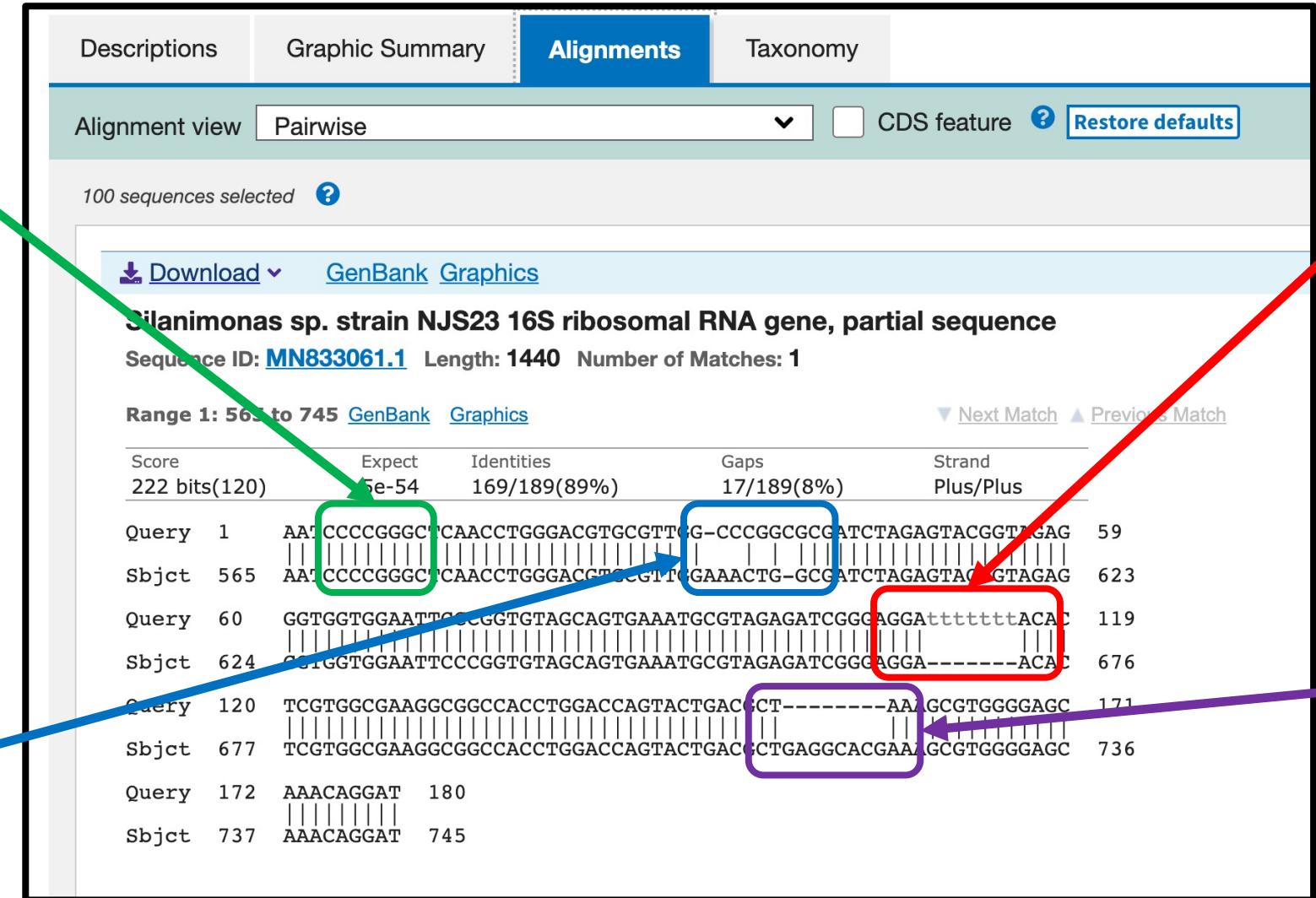
↑ Identity

TGG-CCCGGCGCG
||| |||
TGGAAACTG-CCG

Mismatch

Substitution

↓ Identity



GAtttttttACAC
|||
GA-----ACAC

Insertion

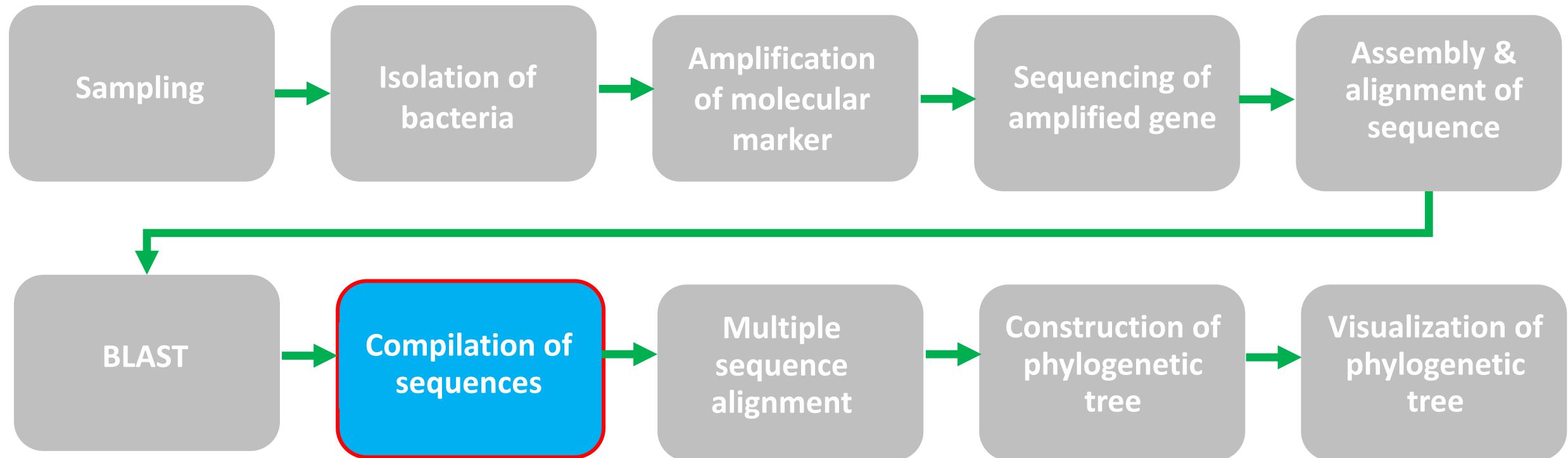
↑ Gap penalty

CT-----AAA
|||
CTGAGGCACGAAA

Deletion

↑ Gap penalty

Compilation of sequences



Compilation of Sequences

BLAST Results from NCBI nt database

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

select all 100 sequences selected

	Description
<input checked="" type="checkbox"/>	Uncultured bacterium clone ER-EC-12_06H 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone ER-EC-11_07A 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone ER-EC-13_upper_Bact_05G 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured Xanthomonadaceae bacterium clone DR938CH110701SACH44 16S ribosomal RNA gene.
<input checked="" type="checkbox"/>	Uncultured bacterium clone 7-166 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas sp. JK12 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas lenta strain 25-4 16S ribosomal RNA., partial sequence
<input checked="" type="checkbox"/>	Uncultured gamma proteobacterium gene for 16S rRNA., partial sequence, isolate: sd-jx78
<input checked="" type="checkbox"/>	Uncultured bacterium clone XT1 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas sp. JK13 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone AW18 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone Wat148 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone AW55 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas sp. strain NJS23 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone GBI-16 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas lenta strain EGK12 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas algicola strain M23 16S ribosomal RNA., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone 16S-27F&1492R-C12-clone4 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Silanimonas sp. YT6 16S ribosomal RNA gene., partial sequence
<input checked="" type="checkbox"/>	Uncultured bacterium clone FMSB6 16S ribosomal RNA gene., partial sequence

Download New Select columns Show 100 ?

BLAST Results from EzBioCloud

List of hits from EzBioCloud 16S database

Select hits by database

Tasks	Hit taxon name	Hit strain name	Accession	Similarity	Variation ratio	Hit taxonomy	Completeness (%)
<input checked="" type="checkbox"/>	Silanimonas lenta	DSM 16282(T)	AUBD01000017	99.79	3/1422	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Silanimonas	100.0
<input checked="" type="checkbox"/>	Silanimonas algicola	M23(T)	KY363638	98.56	20/1392	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Silanimonas	95.5
<input checked="" type="checkbox"/>	Silanimonas mangrovi	AK13(T)	HE573746	95.29	67/1422	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Silanimonas	100.0
<input checked="" type="checkbox"/>	Lysobacter yangpyeongensis	GH19-3(T)	DQ191179	94.73	75/1422	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Lysobacter	99.8
<input checked="" type="checkbox"/>	VOHE_s	YD-1	VOHE01000003	94.59	77/1422	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Luteimonas	100.0
<input checked="" type="checkbox"/>	Lysobacter oryzae	YC6269(T)	EU376963	94.54	77/1410	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Lysobacter	97.0
<input checked="" type="checkbox"/>	AWZR_s	J29	AWZR01000002	94.51	78/1422	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Luteimonas	100.0
<input checked="" type="checkbox"/>	Lysobacter spongiae	119BY6-57(T)	KY451771	94.51	74/1349	Bacteria;Proteobacteria;Gammaproteobacteria;Lysobacterales;Lysobacteraceae;Lysobacter	94.5

NA 2451 2451 99% 0.0 97.82% 1503 KX348537.1

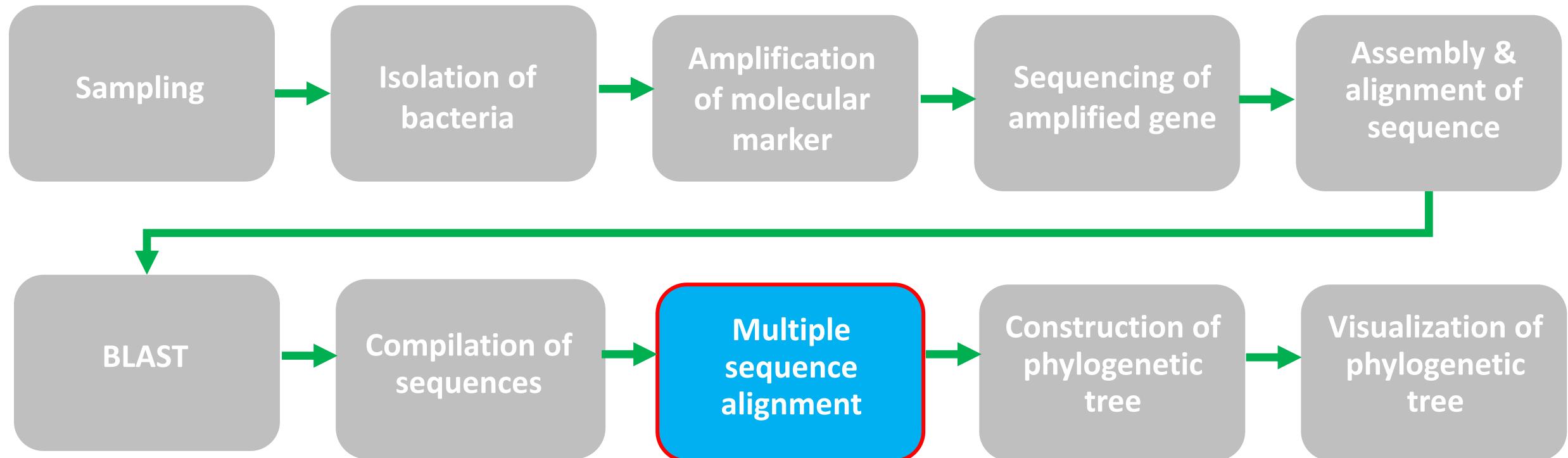
NA 2447 2447 96% 0.0 98.83% 1413 KF206370.1

NA 2446 2446 99% 0.0 97.75% 1495 KF97554

Feedback

Objective:
Molecular identification
of bacteria

Multiple sequence alignment



Multiple Sequence Alignment (MSA)

Function	Alignment of ≥3 sequences (DNA/ amino acid), of similar length
Goal	Align sequences so that homologies are in line and all sequences are the same length with gaps.
Purpose	Gather information about the sequences as a whole (mutation points, level of divergence, etc.) and using these information for downstream analysis (e.g. inferring phylogenetic relationship)
Applications	Inferred and the evolutionary relationships between the sequences studied.

Multiple Sequence Alignment (MSA)

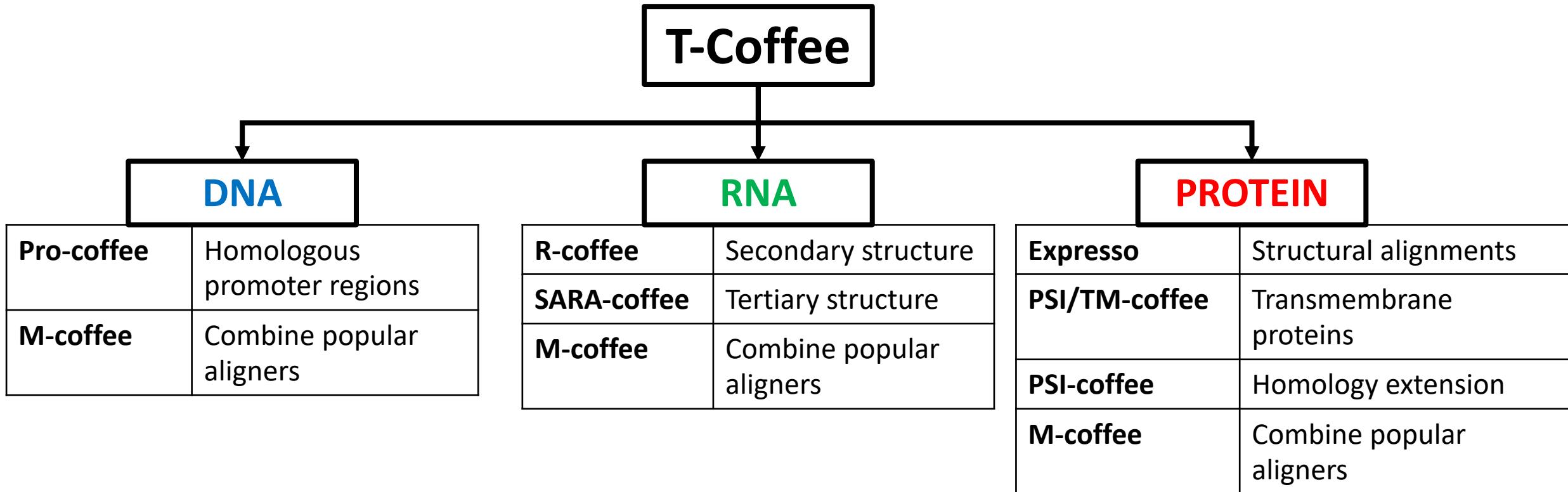
Biologically ideal MSA

1	Maximum similarity
2	Minimum number of gaps
3	Retain conserved motifs and patterns
4	Retain functionally important alignments
5	Able to recapitulate the actual phylogeny
6	Attention will be given on alignable regions instead of gapped regions
7	The limitation imposed by the 3D structure will be considered

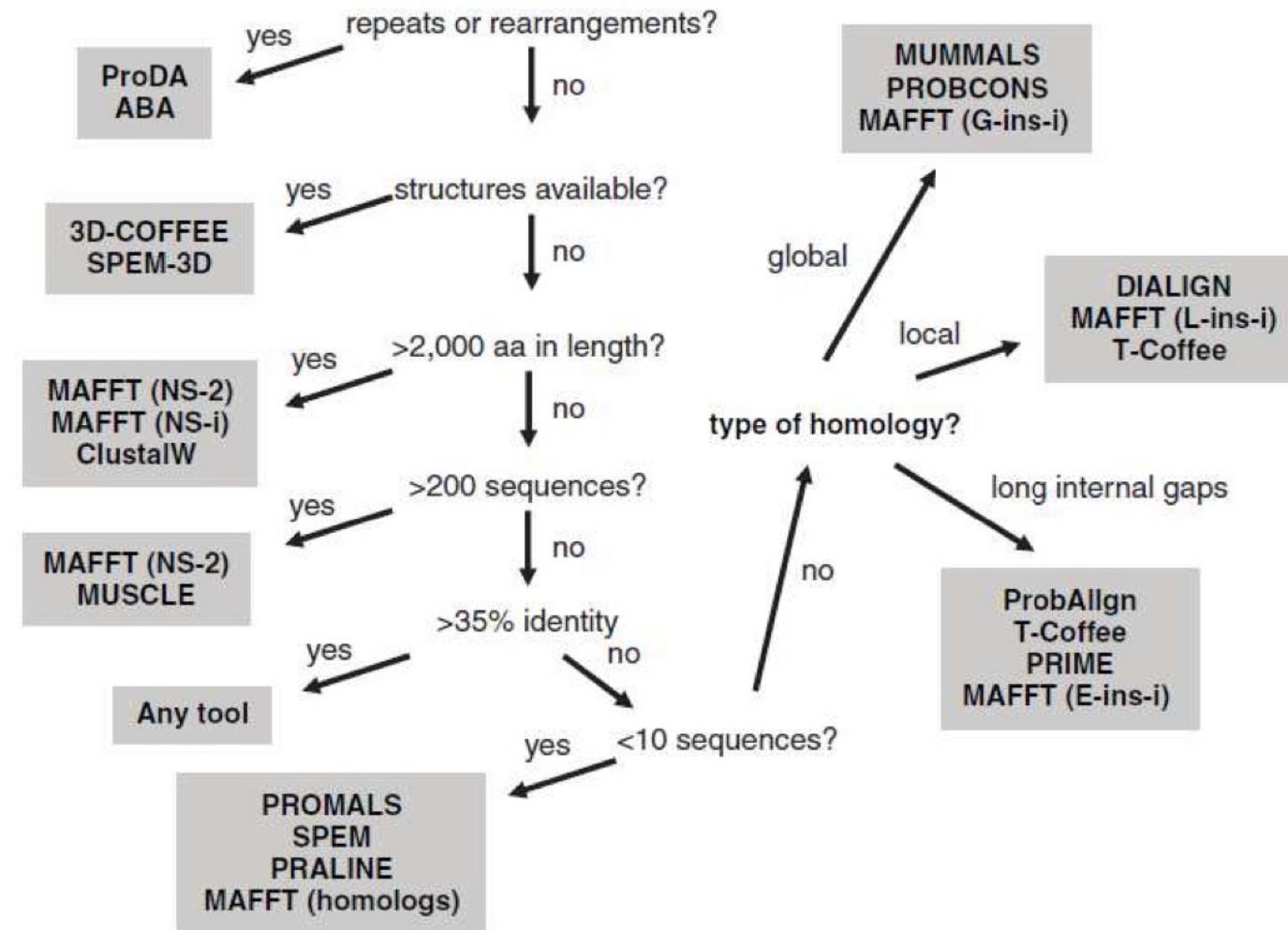
Multiple Sequence Alignment (MSA)

Programme	MSA Tools	Installation/ Web-based	URL
MEGA (version 11)	ClustalW ; MUSCLE		https://www.megasoftware.net/
Jalview	T-Coffee; Probcons; MUSCLE; MAFFT; MSAProbs; Glprobs; Clustal; Clustal Omega	Installation required	jalview.org/getdown/release/
Multiple Sequence Alignment Tools of EMBL-EBI	Clustal Omega; MAFFT; MUSCLE; T-Coffee; EMBOSS Cons; Kalign; WebPRANK; MView	Web-based application	https://www.ebi.ac.uk/Tools/msa/

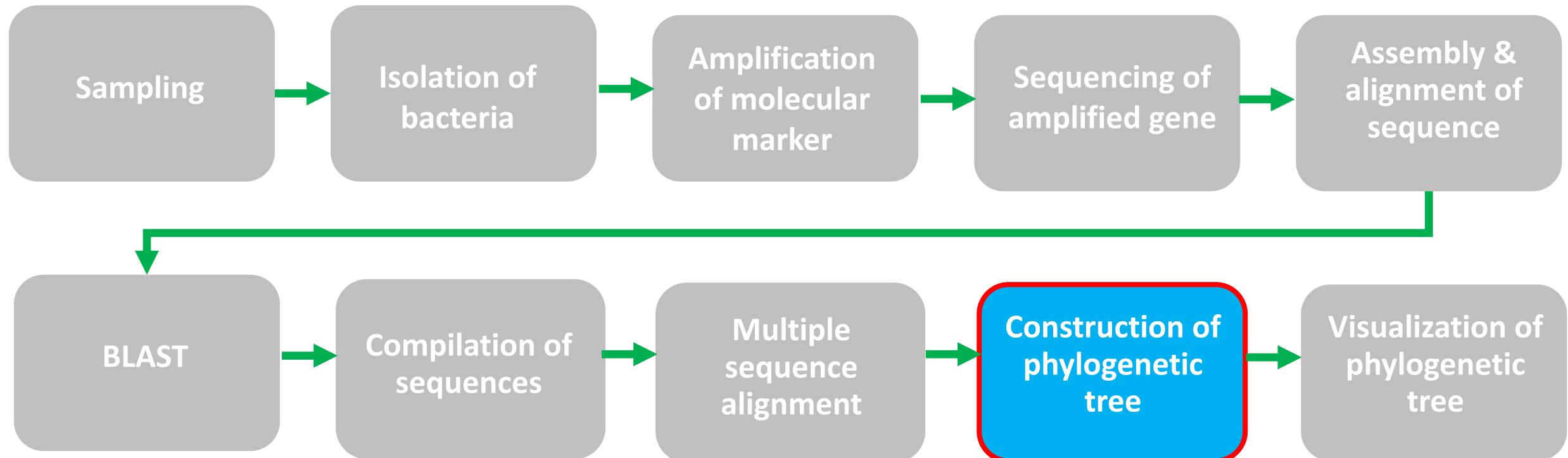
Multiple Sequence Alignment (MSA)



Choosing MSA Programme



Construction of Phylogenetic Tree



Why Construct Phylogenetic Trees?

To infer the evolutionary relationships amongst different species, classes, subclasses, strans, or organisms

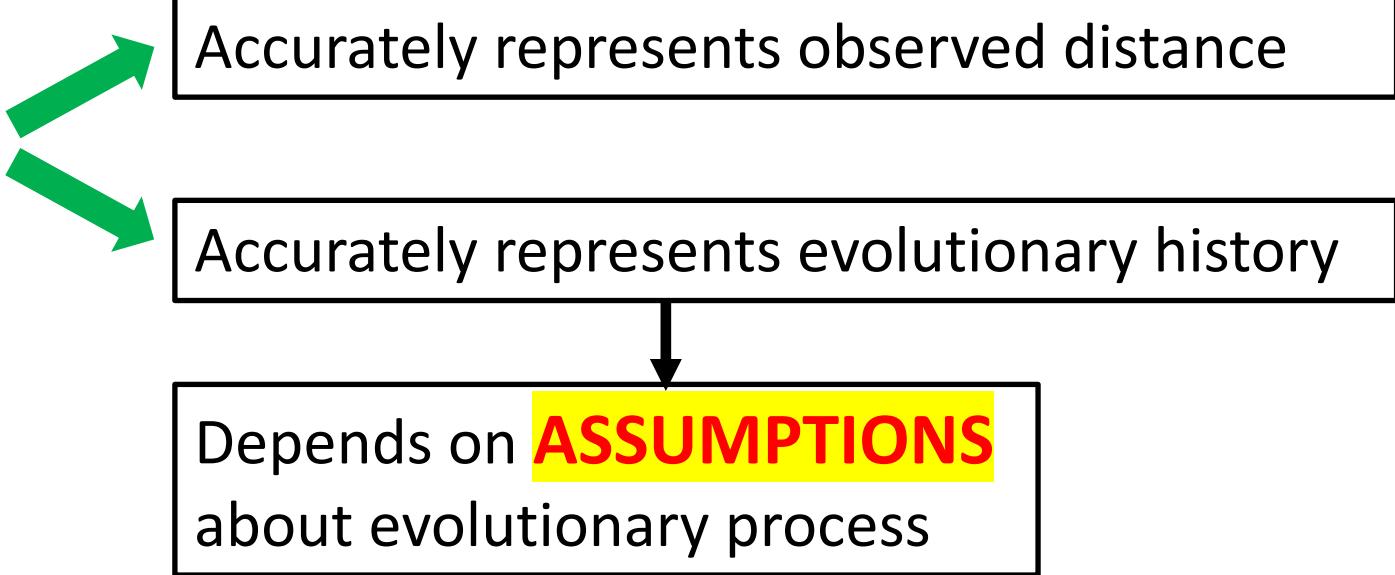
To infer the functions of genes or proteins and to understand how various functions evolved

To use the resulting tree as basis for additional analysis, e.g. to inform multiple alignments

To identify what is most conserved/important in some class of sequences

A Good Phylogenetic Trees

A Good Phylogenetic Tree



Similar Sequence
Arrangement

Similar 3D Fold
= Similar Protein Structure

Similar Biological Function

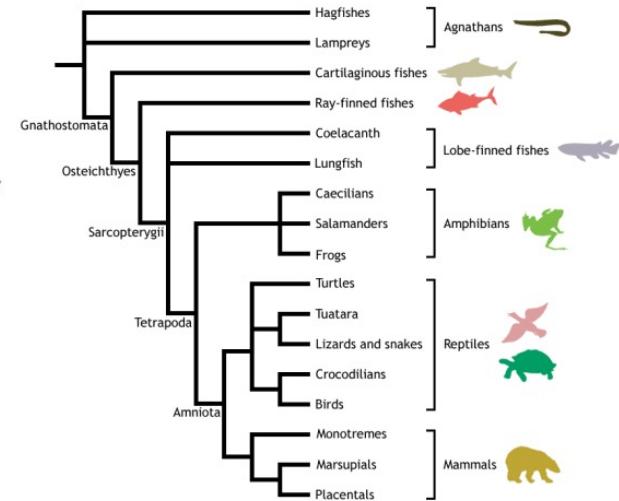
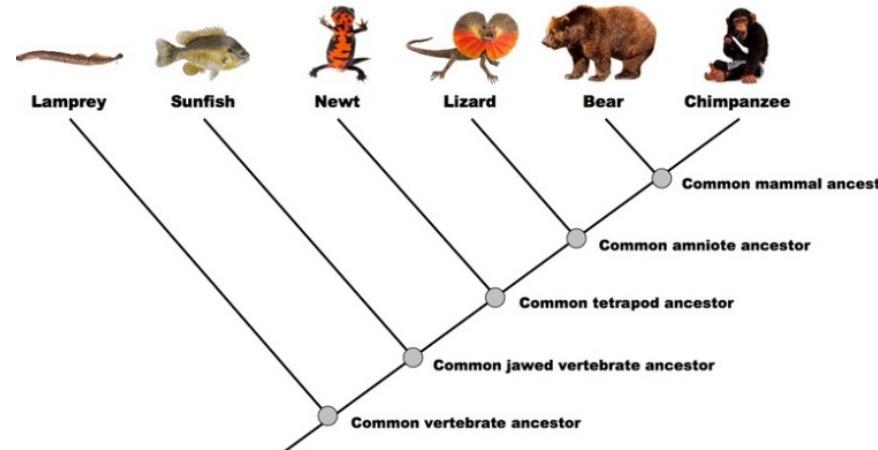
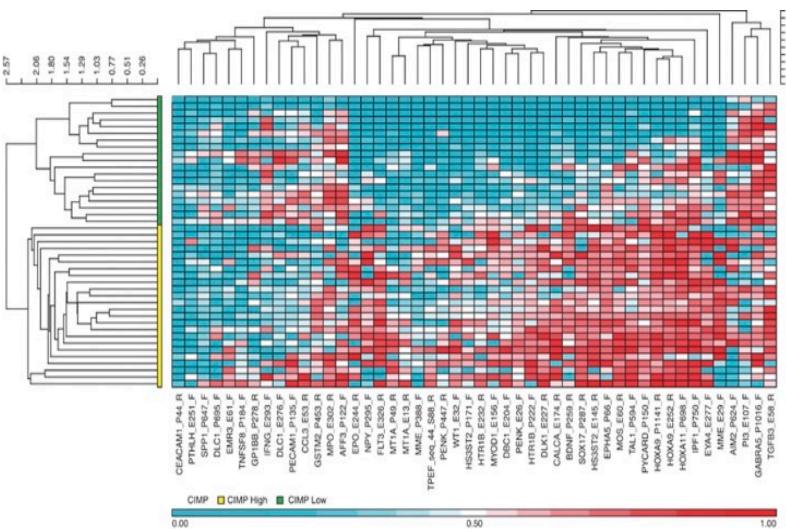
Similar / Same
Ancestor / Origin

Types of Trees

Dendrogram

visually describe the relationships between objects

shows the hierarchical relationship between objects



Types of Trees

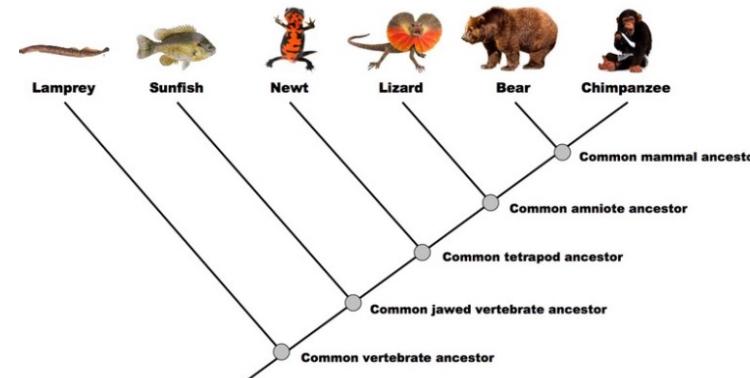
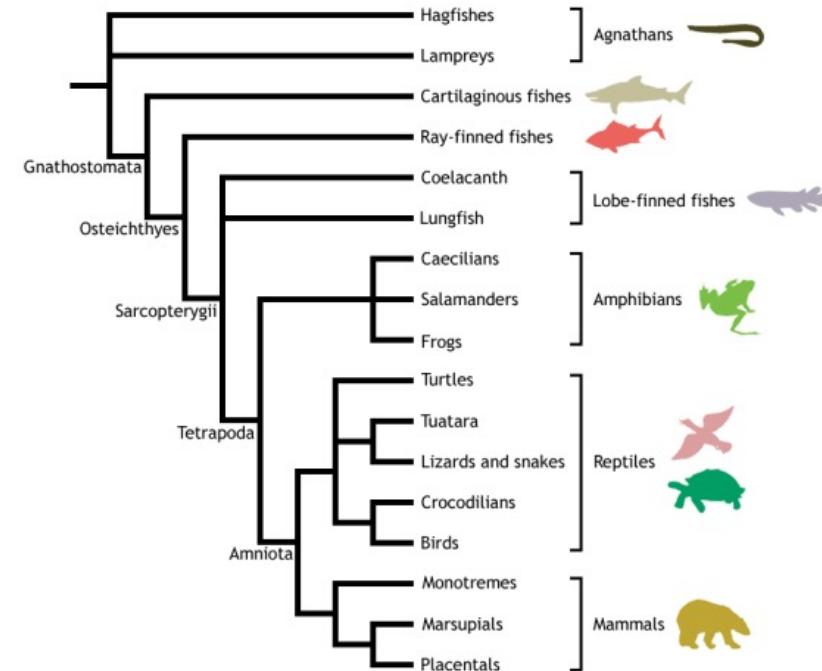
Cladogram

represents a hypothesis about the actual evolutionary history

tree topology (the shape of the tree) indicates the phylogeny relationship between organisms

Shows **branching order**, but branch lengths meaningless.

Branch length does not represent evolutionary time or the genetic distance



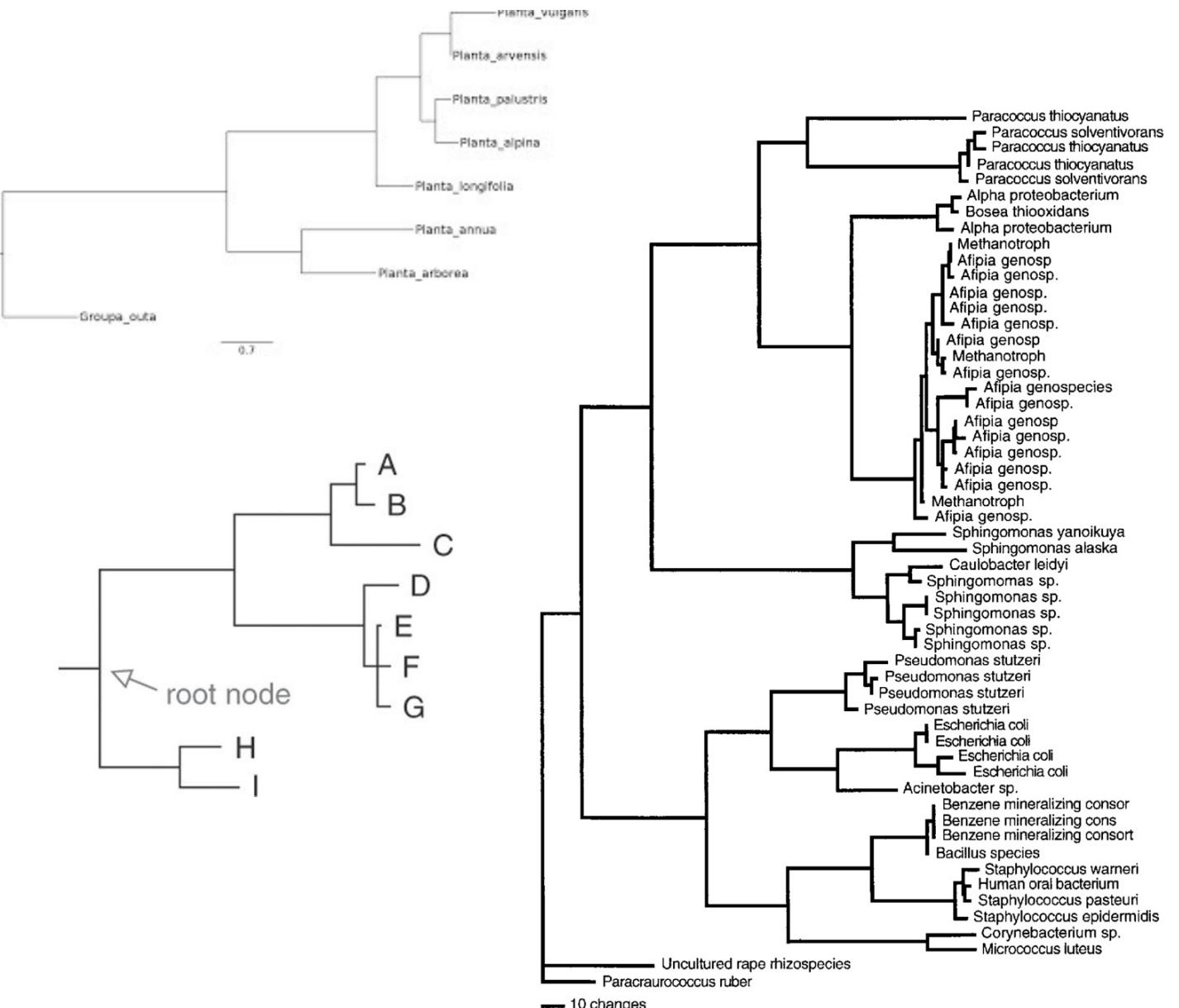
Types of Trees

Phylogram

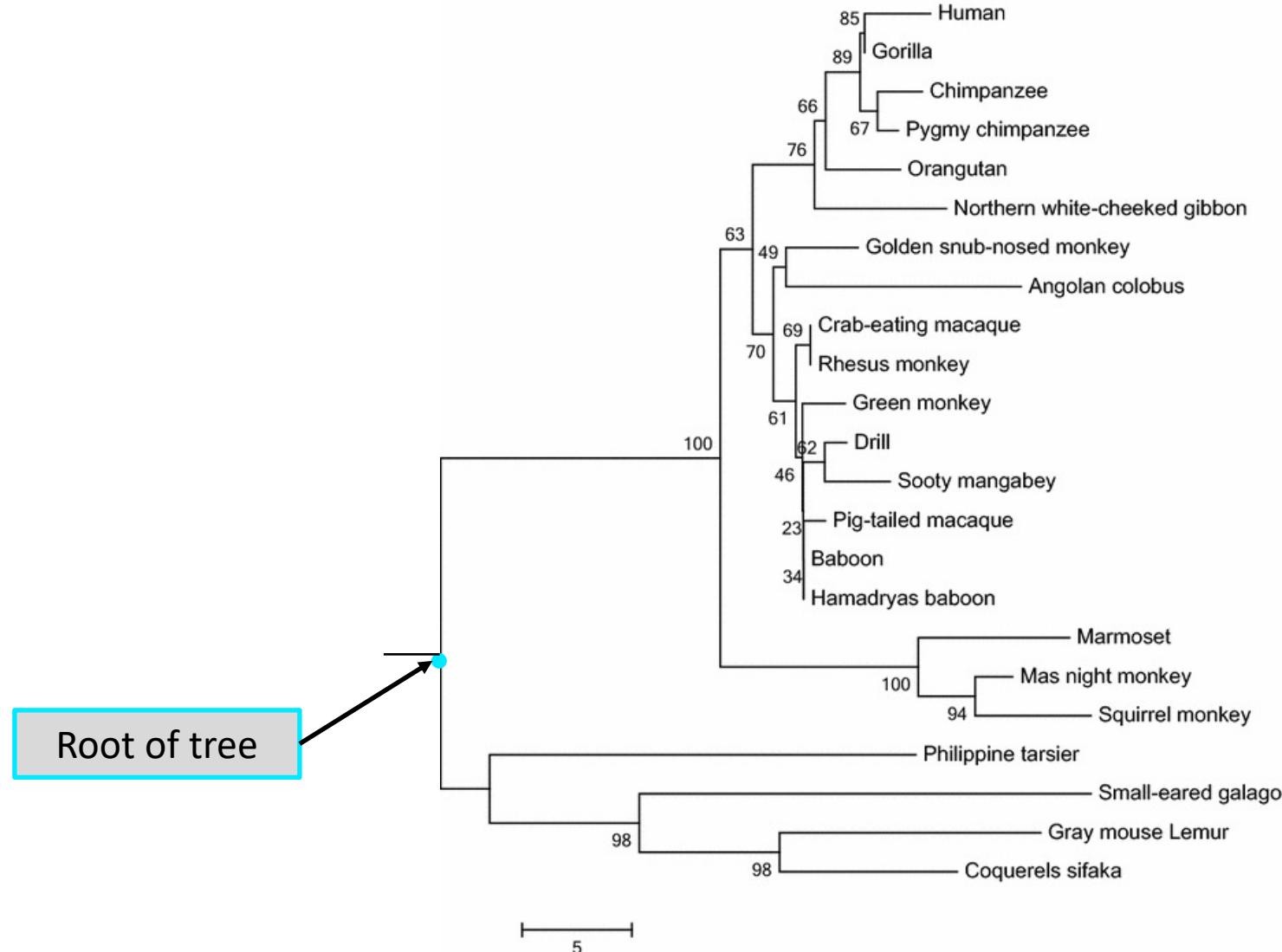
The **distance of the branch** depends on the amount of inferred **evolutionary changes**

Represents the **evolutionary time** and the **genetic distance** between the group of organisms

branch lengths are proportional to the **number of substitutions** along the branches and root-to-tip path lengths are usually unequal.

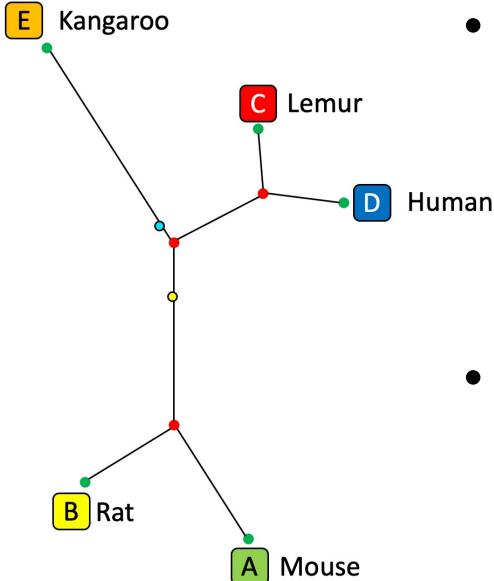


Understanding Phylogenetic Trees



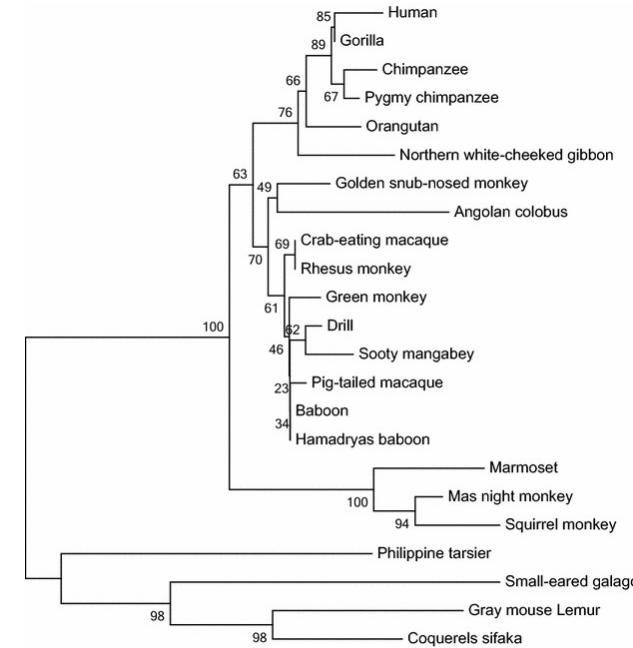
Rooted and Unrooted Trees

Unrooted Tree



- Represent **branching** **order** which specifies relationships among objects
- Do not indicate the root of the location of the last common ancestor (**no ancestral root**)

Rooted Tree



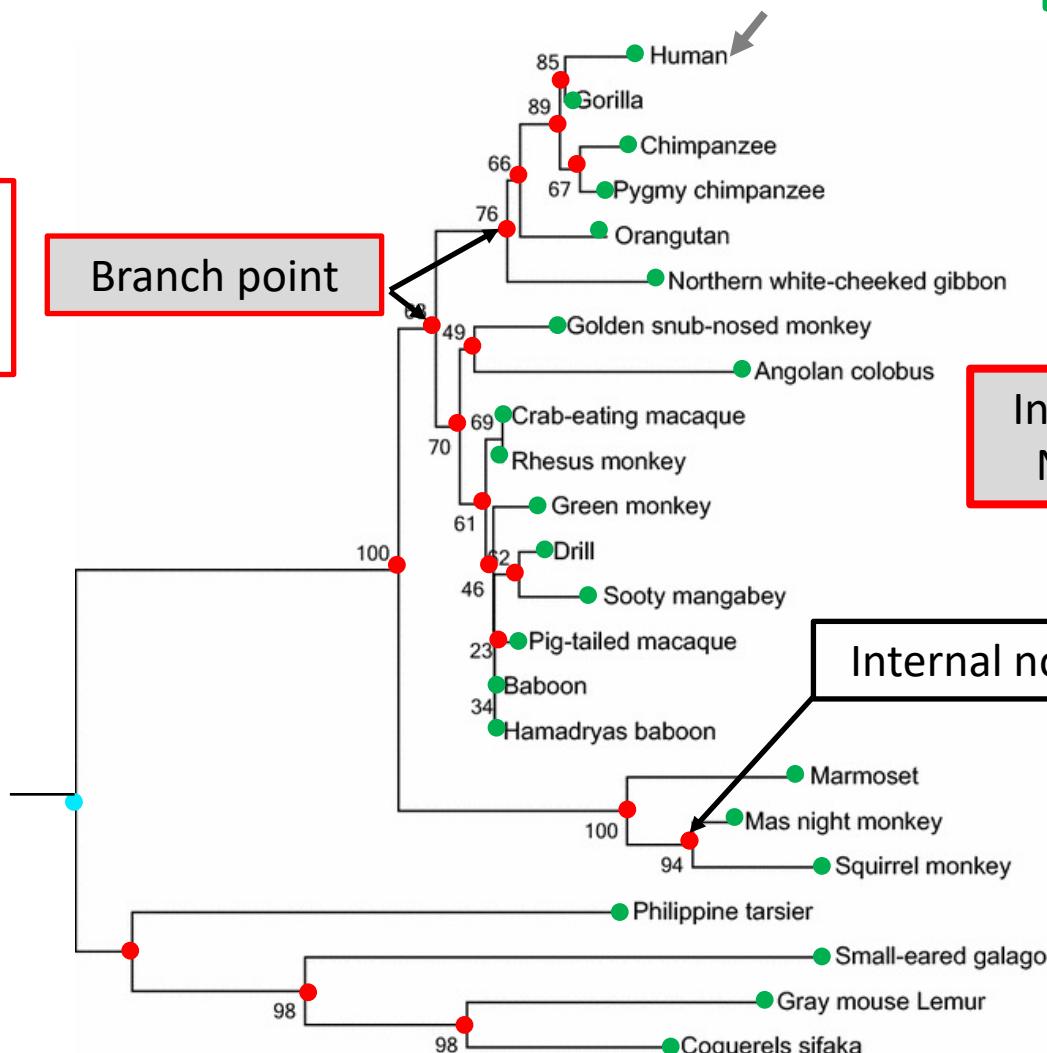
- The root represents the **common ancestor**
- Path from root to a node represents an **evolutionary path**.

Phylogenetic Trees

Leaf = Taxon = Tip = Terminal Node = Leaf Node = Operational taxonomic Unit (OTU)

Internal nodes represents the putative ancestor of the investigated terminal nodes

Branch point

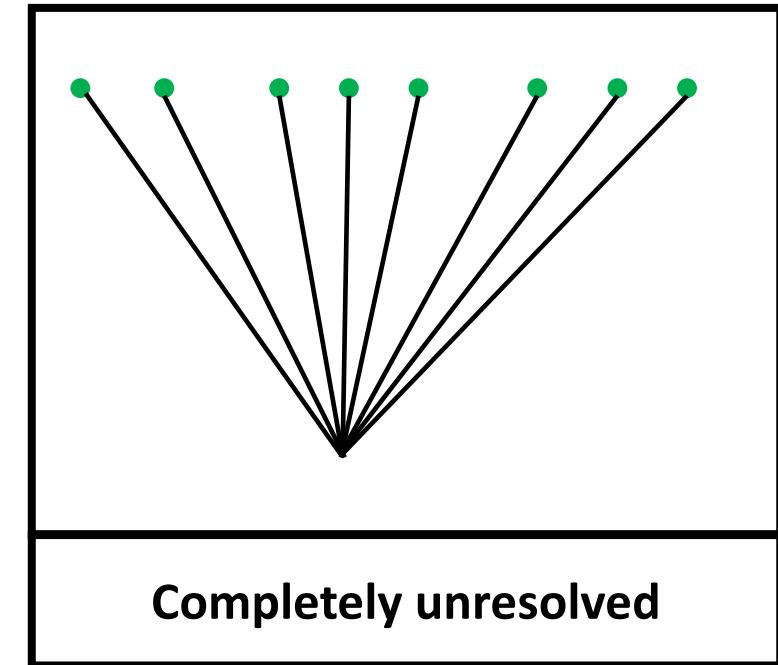
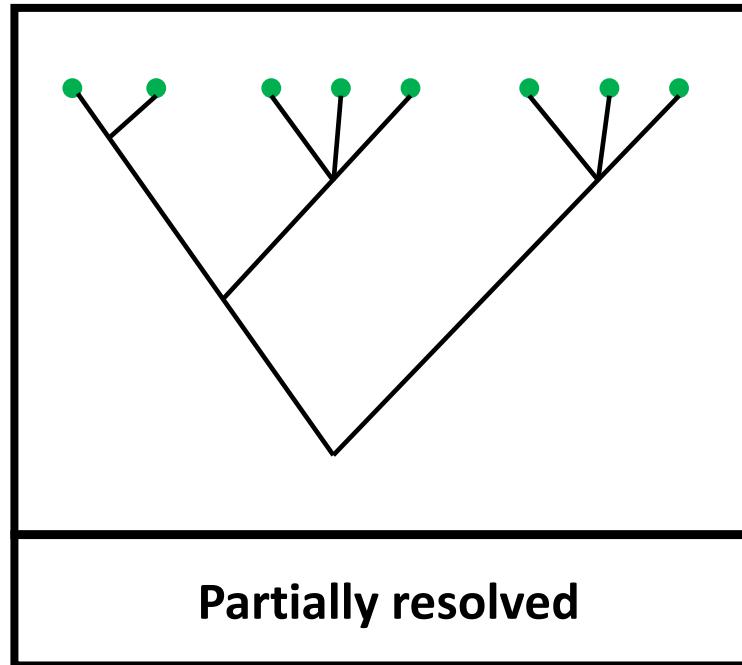
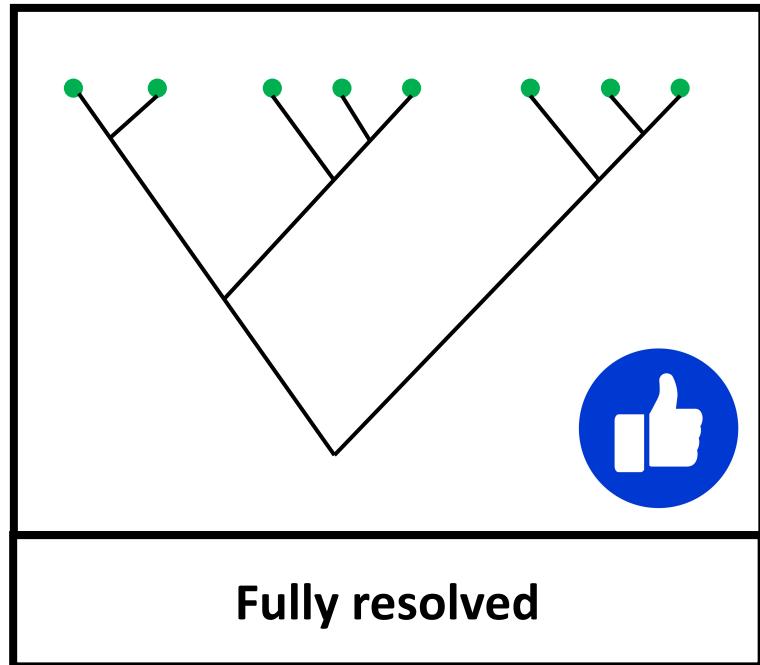


● Leaf/ Terminal Node
 ● Internal Node

Internal Node = Branching point = Hypothetical taxonomic Unit (HTU)

Internal node X exists prior to mas night monkey and squirrel monkey. X is the ancestor of these two primates

Phylogenetic Trees



Bifurcating Tree

Phylogenetic Trees

The **topology (branching order)** of a phylogenetic tree shows the phylogenetic history of sequences (or by inference, of species)

The units of branch length are usually **nucleotide substitutions per site**. In some cases, branch length may be denoted in percentage of change, thus the **number of changes per 100 nucleotides sites**.

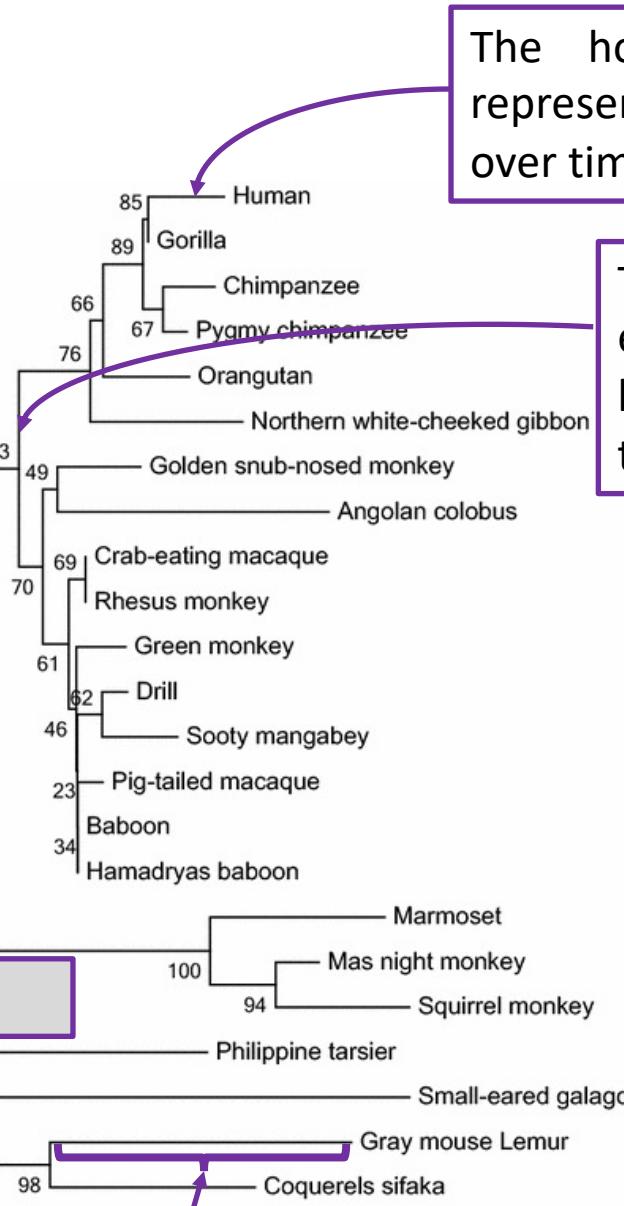
Distance scale

Branch length

Branch

Internal branch

External branch



The horizontal lines are branches and represent evolutionary lineages changing over time.

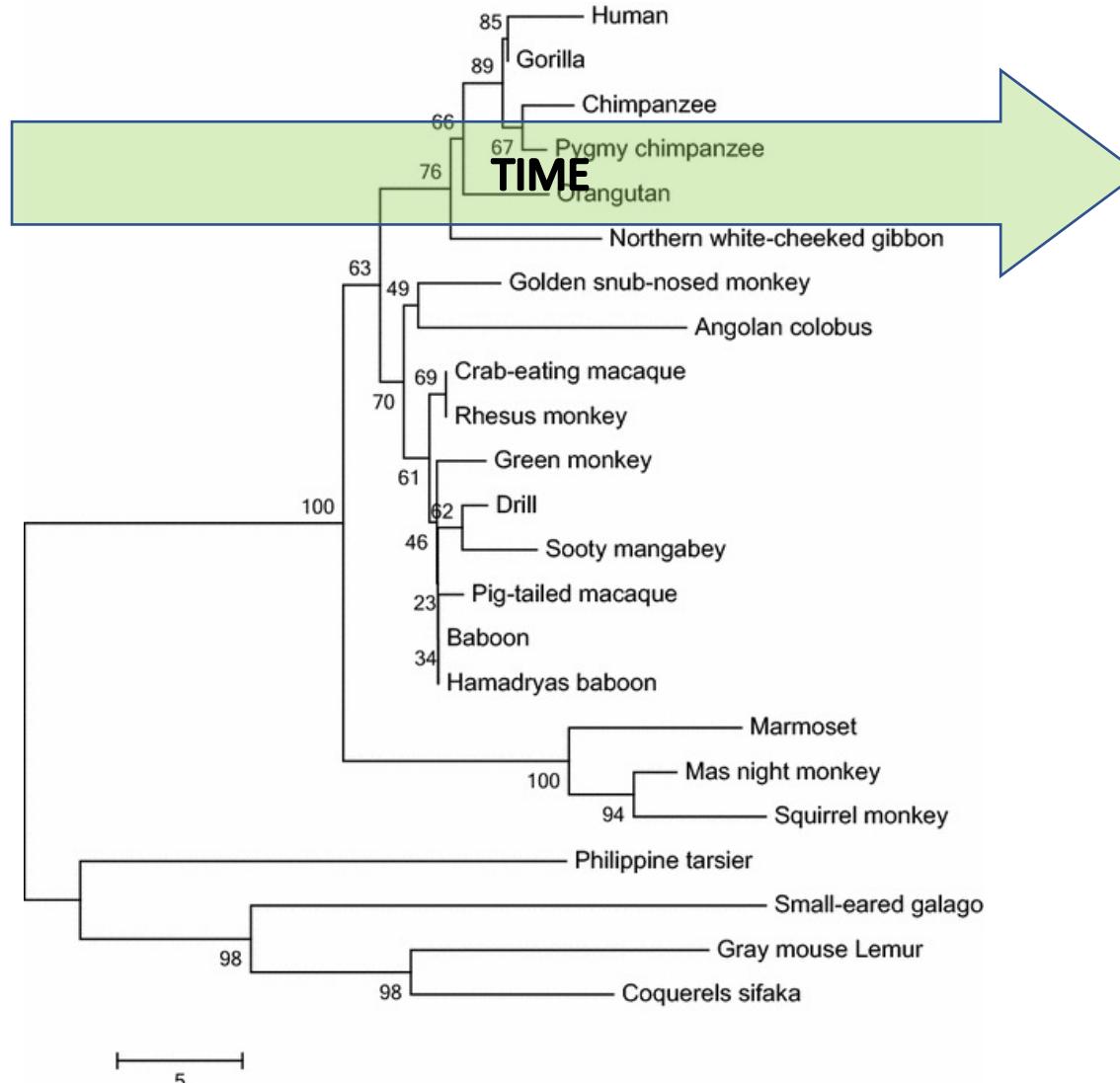
The vertical lines represent nodes or evolutionary splits, and show which branches are connected. The length of the vertical line has no meaning

The longer the branch, the larger the amount of change

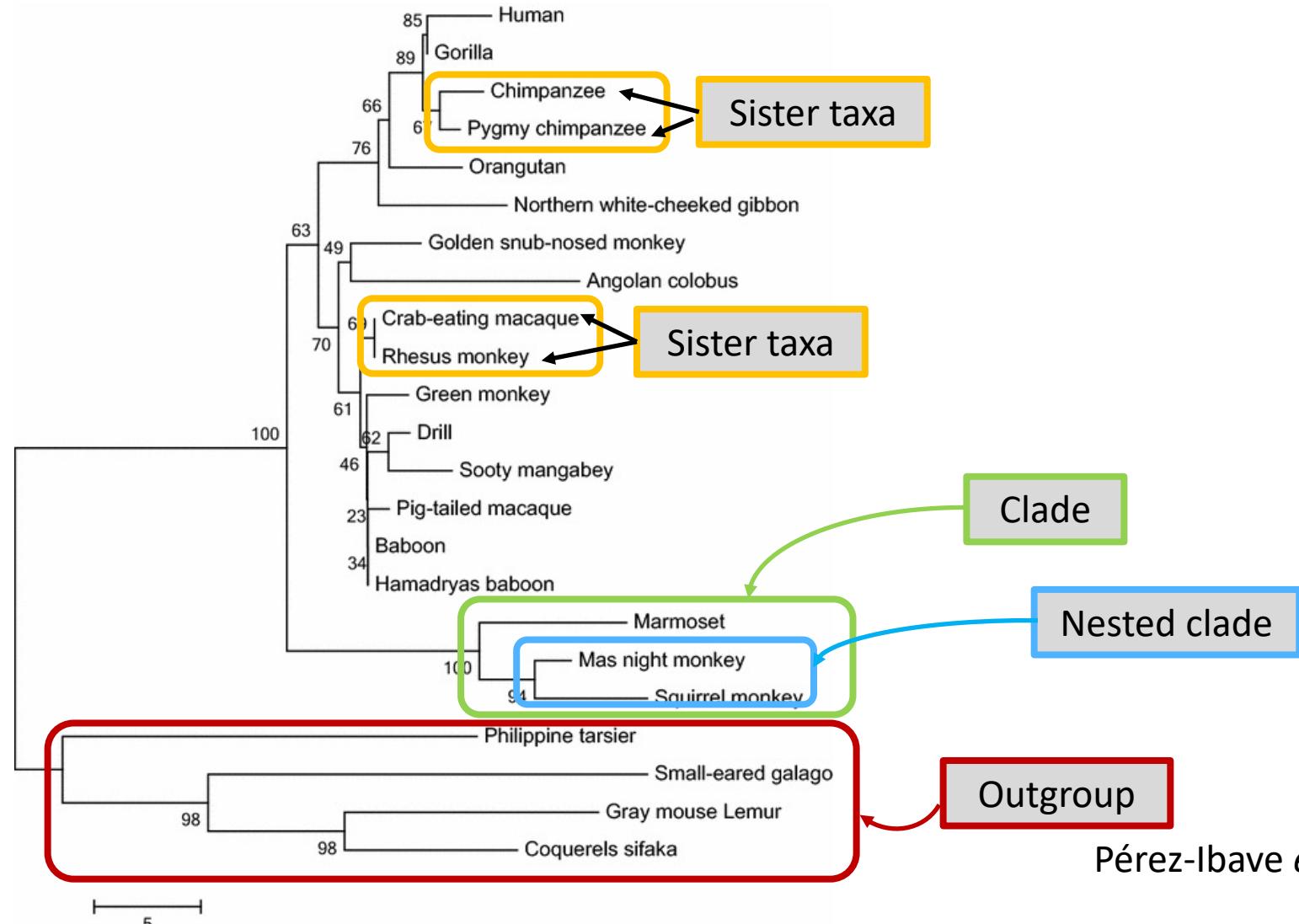
Branch lengths show the number of mutations occurred in the evolutionary time between lineages

Phylogenetic Trees

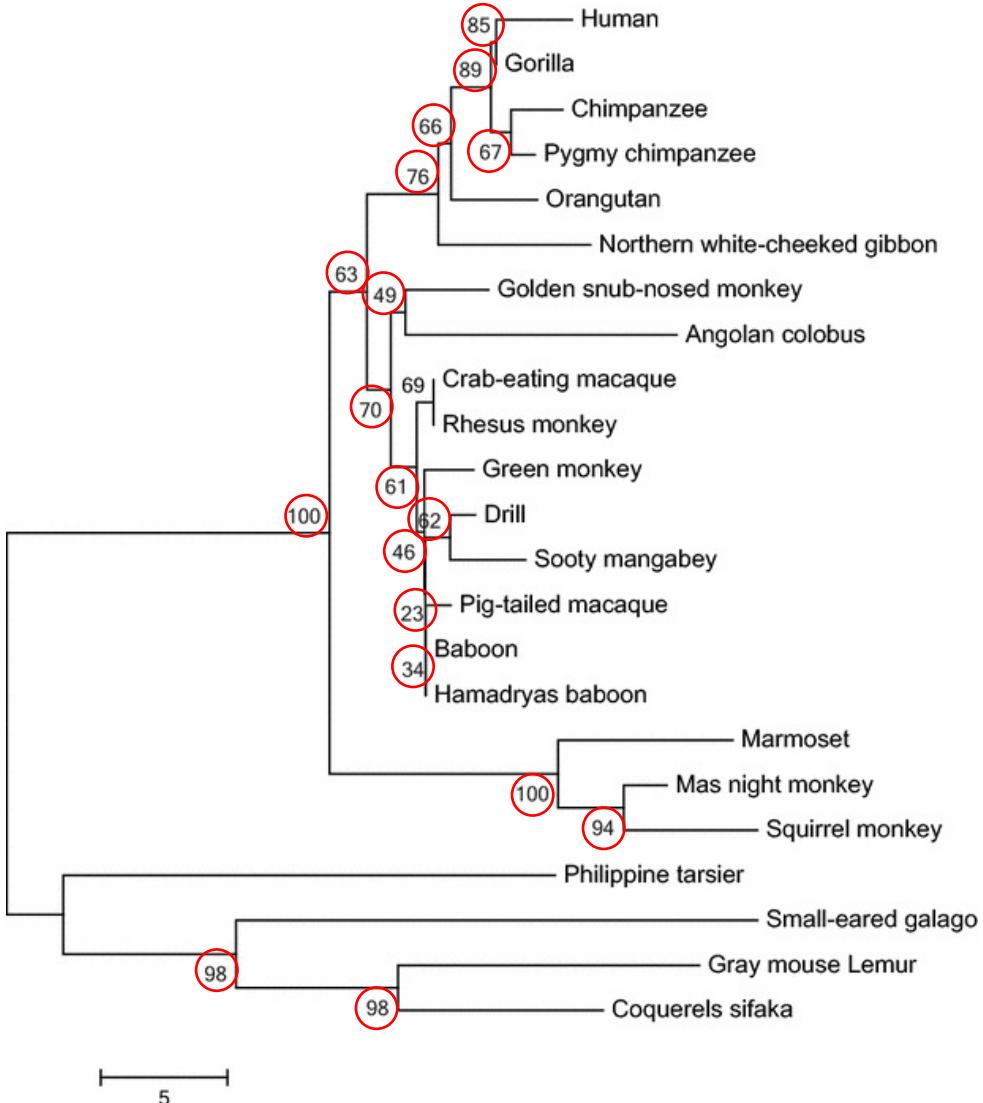
Time is **approximately** flowing from left to right. The word 'approximately' is used because in this tree the horizontal axis is measured as **genetic change**, and to convert this into actual time, **assumptions** about the relationship between genetic change and time needed to be made.



Phylogenetic Trees



Phylogenetic Trees



Support values show you how reliable a branching split is

Usually between 0 and 1, or between 0% and 100%

Branches that are not well supported might be collapsed

Phylogenetic Reconstruction Methods

	1	2	3	4	5	6	7
1. Ath	1						
2. Bra	0.027						
3. Gma	0.239	0.244					
4. Vvi	0.216	0.205	0.164				
5. Osa	0.452	0.444	0.492	0.399			
6. Mtr	0.227	0.227	0.063	0.134	0.475		
7. Ptr	0.239	0.250	0.139	0.154	0.475	0.116	

Species/Abbrv	Gr	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		
1. Ath	E	M	M	N	S	L	S	H	I	F	R	W	E	L	L	V	G	E	R	Y	G
2. Bra	D	M	M	N	S	L	S	H	I	F	R	W	E	L	L	V	G	E	R	Y	G
3. Gma	D	M	M	N	S	L	S	Q	I	F	R	W	D	L	L	V	G	E	R	Y	G
4. Mtr	D	M	M	N	S	L	S	Q	I	F	R	W	D	L	L	V	G	E	R	Y	G
5. Osa	D	M	M	A	A	L	A	G	L	F	R	W	D	L	L	G	E	R	F	G	
6. Ptr	E	M	M	N	S	L	S	Q	I	F	R	W	D	L	L	V	G	E	R	Y	G
7. Wi	D	M	M	N	S	L	C	Q	I	F	R	W	D	L	L	V	G	E	R	Y	G

Distance matrix method

Phenetics method

Suitable for **continuous characters**

Compute a **matrix of pairwise distances** between sequences that approximate evolutionary distance

Calculating the **percent difference** between each pair of sequences

- **Neighbour-Joining Method**

Discrete-data method

Cladistics method

Suitable for **discrete characters**

Examine each column of a **multiple sequence alignment** dataset separately and search for the tree that best represents all this information

Using the **raw data**. (Thus, more powerful than distance methods)

- **maximum parsimony**
- **maximum likelihood**

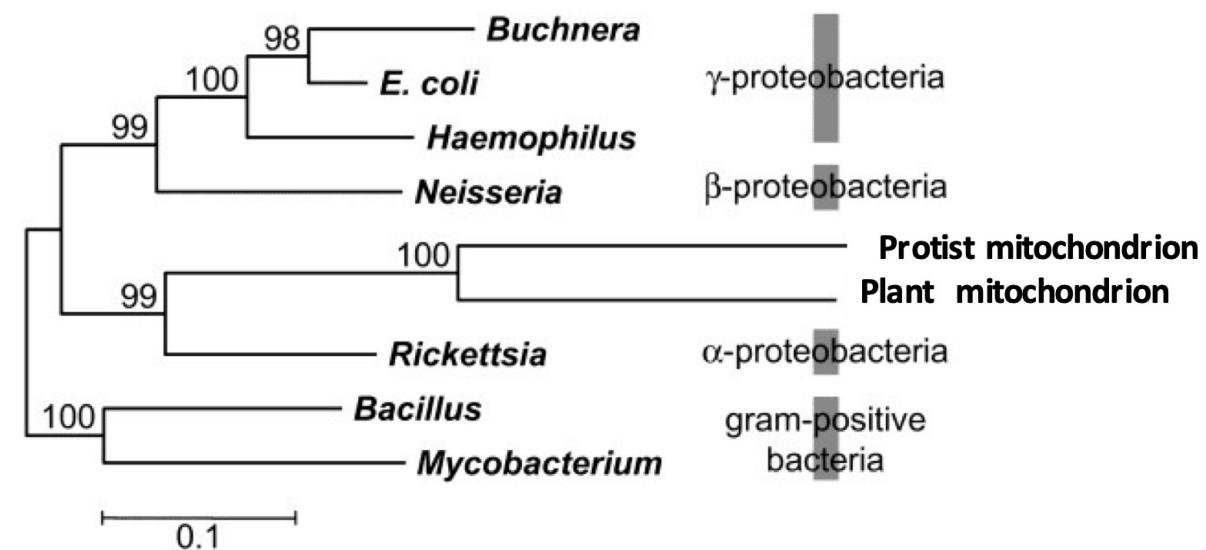
Neighbour-Joining Method

Principle

Find direct ancestor of two species, join them, iterate.

Properties

- Designed to account for a non-uniform molecular clock (**Unequal rates of evolution**)
- Most reliable when the branch lengths of trees are allowed to vary
- Goal is to find a tree that minimize the square errors of pairwise distances.



Maximum Parsimony

Principle

Identifying the potential phylogenetic tree that requires the **smallest total number of evolutionary events** or **minimal number of changes** to explain the observed sequence data

Intuitive score for tree

Number of changes along edges

Minimizing this score is called **parsimony**

Parsimony score is defined as the total number of times the value of some character changes along some edge.

Properties

- Appropriate for **very similar sequences** and a **small number of sequences**
- Calculate for all possible trees and find the tree that represents the minimum number of substitutions at each informative site
- **Time Consuming**
- Only informative sites are used
- Does not correct for multiple mutations
- Does not provide information on the branch lengths
- The most parsimonious tree is not always the correct one

Maximum Likelihood

Principle

Evaluate all possible trees (topology and branch lengths) and substitution model parameters (Transitions vs transversions, base frequency, rate heterogeneity etc.). Choose the one that maximizes the likelihood of your data (the alignment)

Properties

- Use probability calculations to find a tree that best accounts for variation in a set of sequences.
- Analysis is performed on **each column** of a multiple sequence alignment.
- All possible trees are considered. Only can be used for a small number of sequences

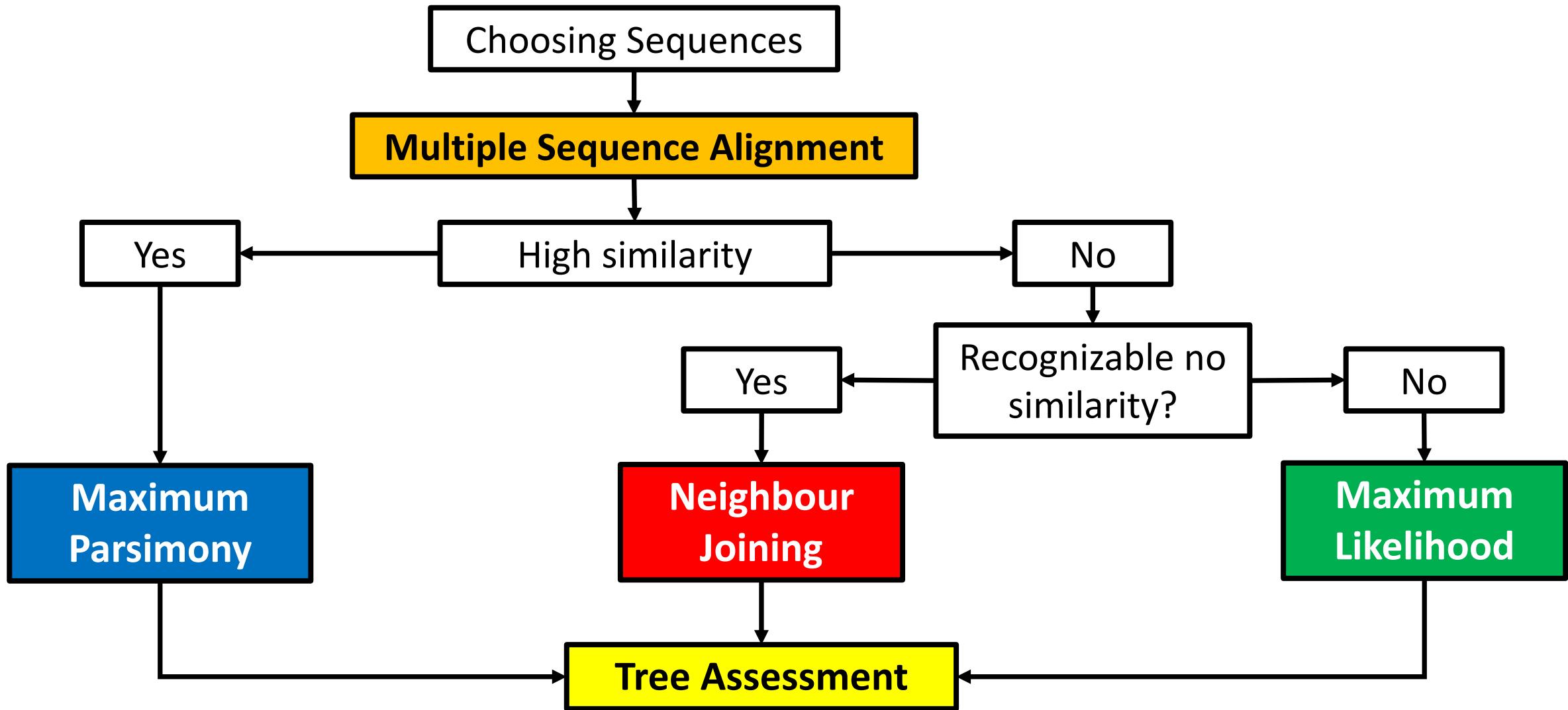
Assumptions

- Character (nucleotide) evolved independently
- Variation in mutation rate
- The molecular clock posits a constant rate of genetic change among lineages

Comparison of Phylogenetic Reconstruction Methods

Methods	Advantages	Disadvantages
Neighbour Joining	Fast	<ul style="list-style-type: none">Information is lost in compressing sequences into distancesreliable estimates of pairwise distances can be hard to obtain for divergent sequences
Maximum Parsimony	<ul style="list-style-type: none">Fast enough for the analysis of hundreds of sequencesrobust if branches are short (closely related sequences or dense sampling)	Can perform poorly if there is substantial variation in branch lengths
Maximum Likelihood	The likelihood fully captures what the data tell us about the phylogeny under a given model	Can be prohibitively slow (depending on the thoroughness of the search and access to computational resources)

Which Methods to Use?



Bootstrapping

Objective

- tests for the **reliability** of the tree **topology**
- to test whether the whole dataset supports the tree

Properties

- The bootstrap values give no indication of the robustness of these features to changing the model or method

Measurement

The bootstrap measures the **degree of support** within the data for the particular **branch**, given the evolutionary model and tree reconstruction method

Bootstrapping

Basic steps in any bootstrap analysis.

1 Sample datasets are automatically generated from an original dataset.

2 Trees are then estimated from each sample dataset.

3 The results are compiled and compared to determine a bootstrap consensus tree

Original Dataset

0123456789
seqA ACCGTTCGGT
seqB ATGGTTCAGA
seqC ATCGATCGGA

Sample Dataset 1

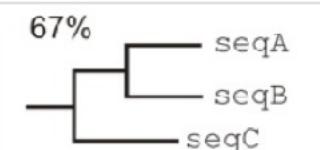
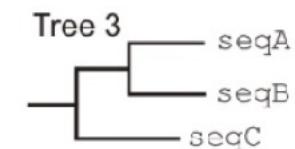
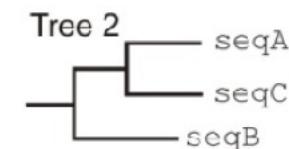
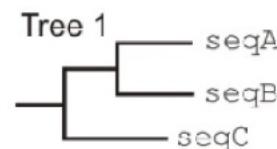
1562314951
seqA CTCCGCTTTC
seqB TTCGGTTATT
seqC TTCCGTAATT

Sample Dataset 2

5234924418
seqA TCGTTCTTCG
seqB TGGTAGTTTG
seqC TCGAACAAATG

Sample Dataset 3

5607718907
seqA TCAGGCGTAG
seqB TCAAATGAAA
seqC TCAGGTGAAG



Bootstrap Consensus Tree

Example

<https://github.com/KarwaiHong/karwaihong>

Question

<https://github.com/KarwaiHong/karwaihong>