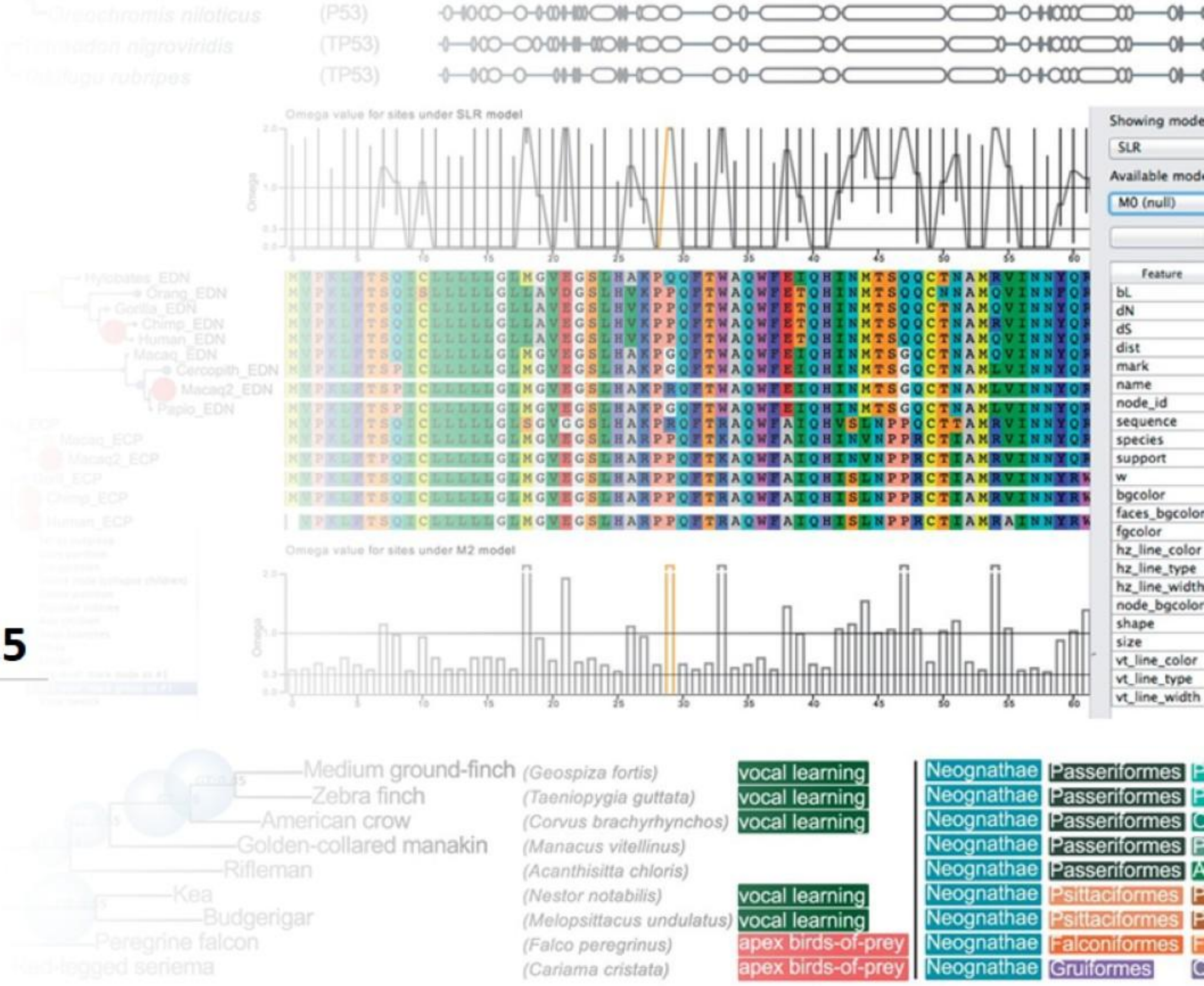
Phylogenetic tree (I)

Bioinformatics lec7 /2025

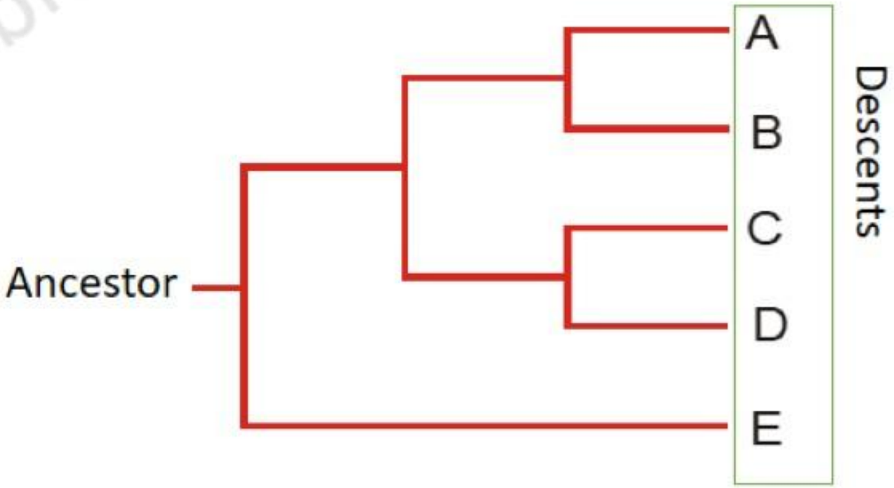
By

Dr Delveen R. Ibrahim



Phylogenetics: is the study of evolutionary relationships among organisms (**ancestor descendent** relationships).

Phylogenies can be used to study the species evolution or the evolution of genes.

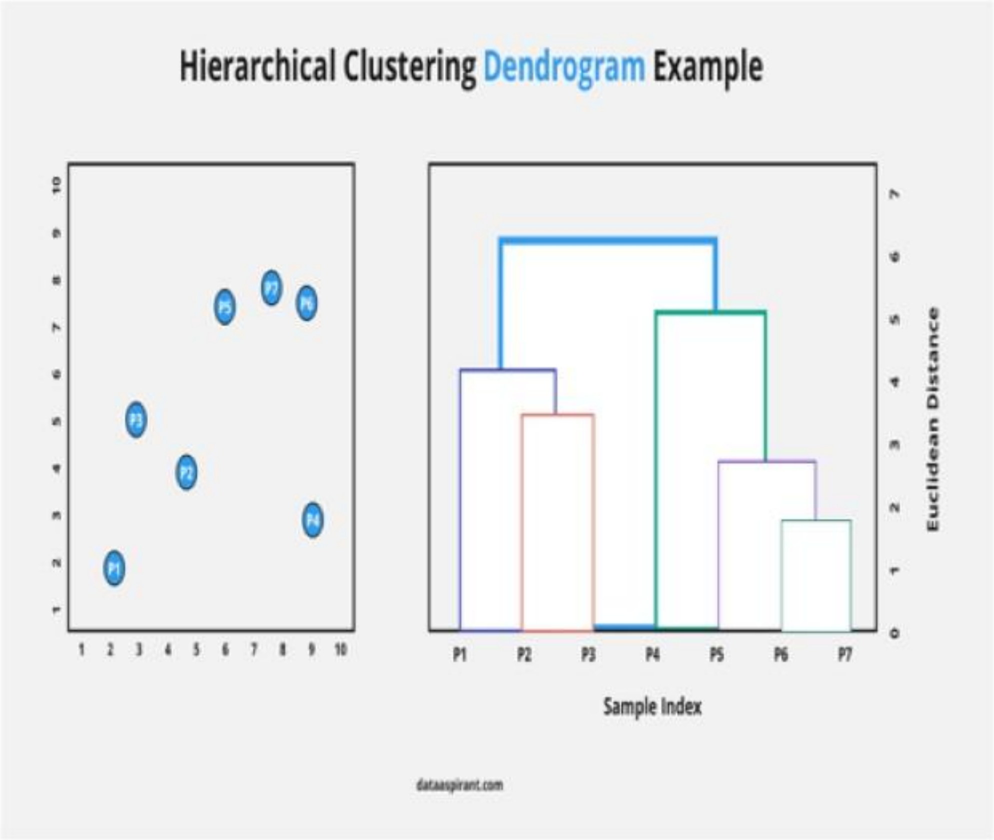


Phylogenetic tree

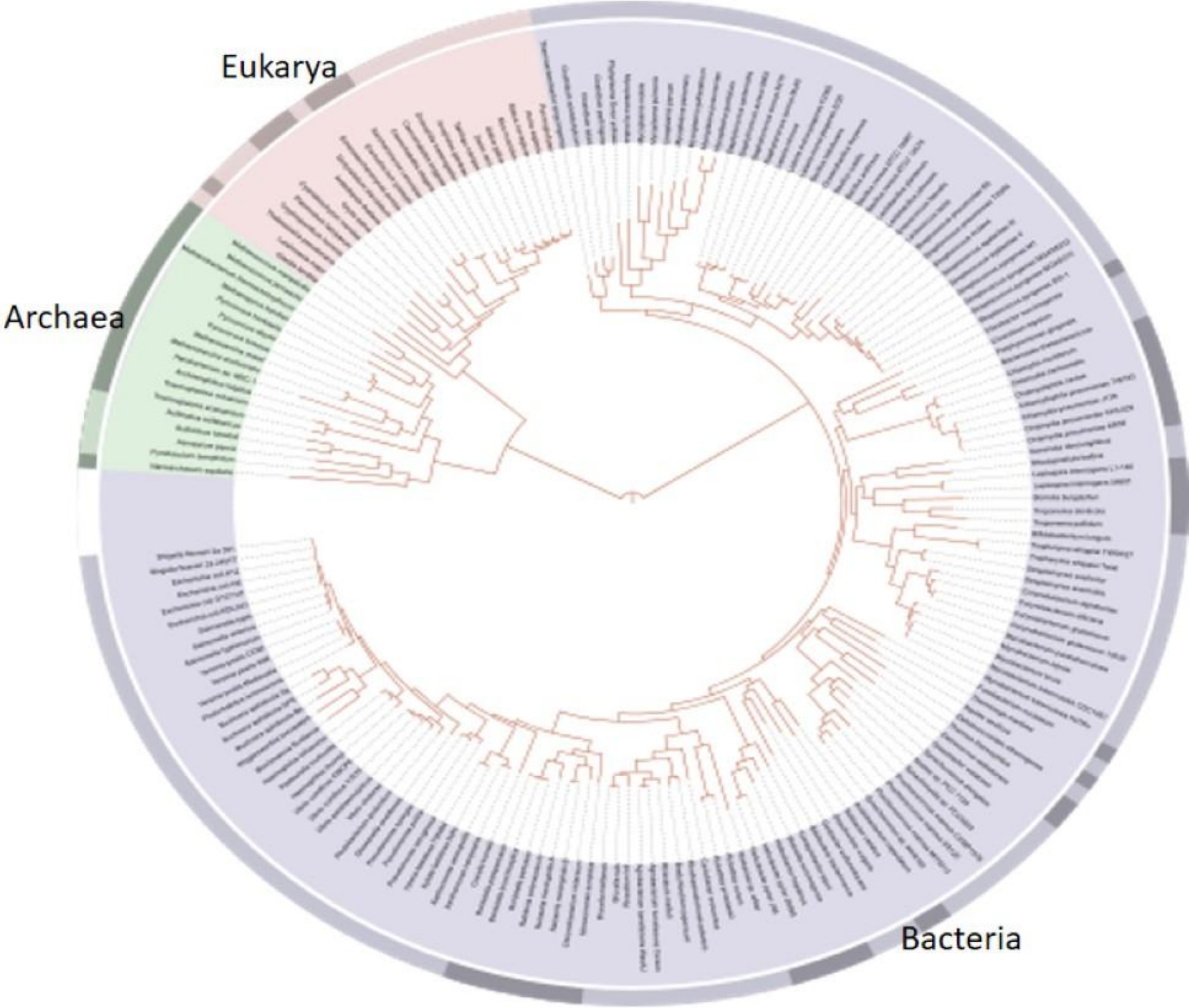
- **A phylogenetic tree** is a diagrammatic representation of the evolutionary relationship among various taxa, here by using gene/protein sequences.
- In the context of molecular phylogenetics, the expressions phylogenetic tree, phylogram, cladogram, and dendrogram are used interchangeably to mean the same thing— but there are some differences.
- **Thus**, it is important to be aware that usage of the vocabulary is not always consistent in the literature, although the context is the same.

Terminologies

- **Phylogram**: A tree where **branch lengths show evolutionary change.** Longer branches mean more genetic differences (scaled).
- **Cladogram** A tree that **only shows relationships** between species but **does not** include time or amount of change.
- **Dendrogram:** A general term for any tree-like diagram, not just for evolution. It can be used in other fields too, like grouping similar things in data science



A highly resolved, automatically generated tree of life (phylogram), based on completely sequenced genomes

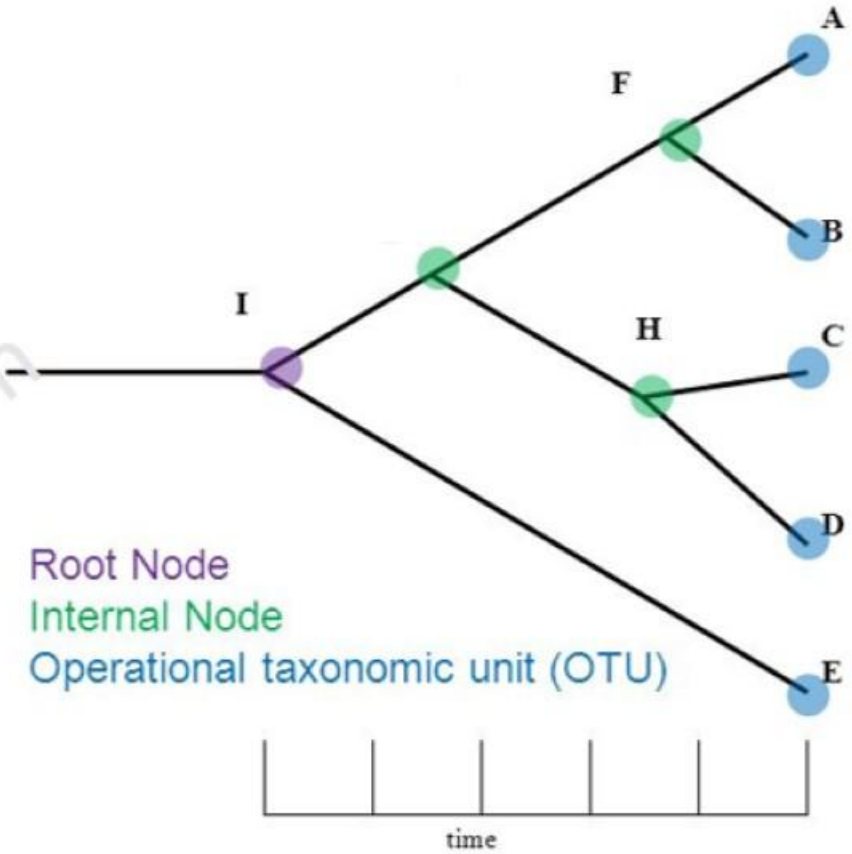


Some important terminologies related to phylogenetic tree

Node: a branchpoint in a tree (internal, terminal and root)

Operational Taxonomic Unit (OTU): It represents a unit of comparison in evolutionary studies. Such as DNA and protein sequences (also referred as TAXA if the species is defined)

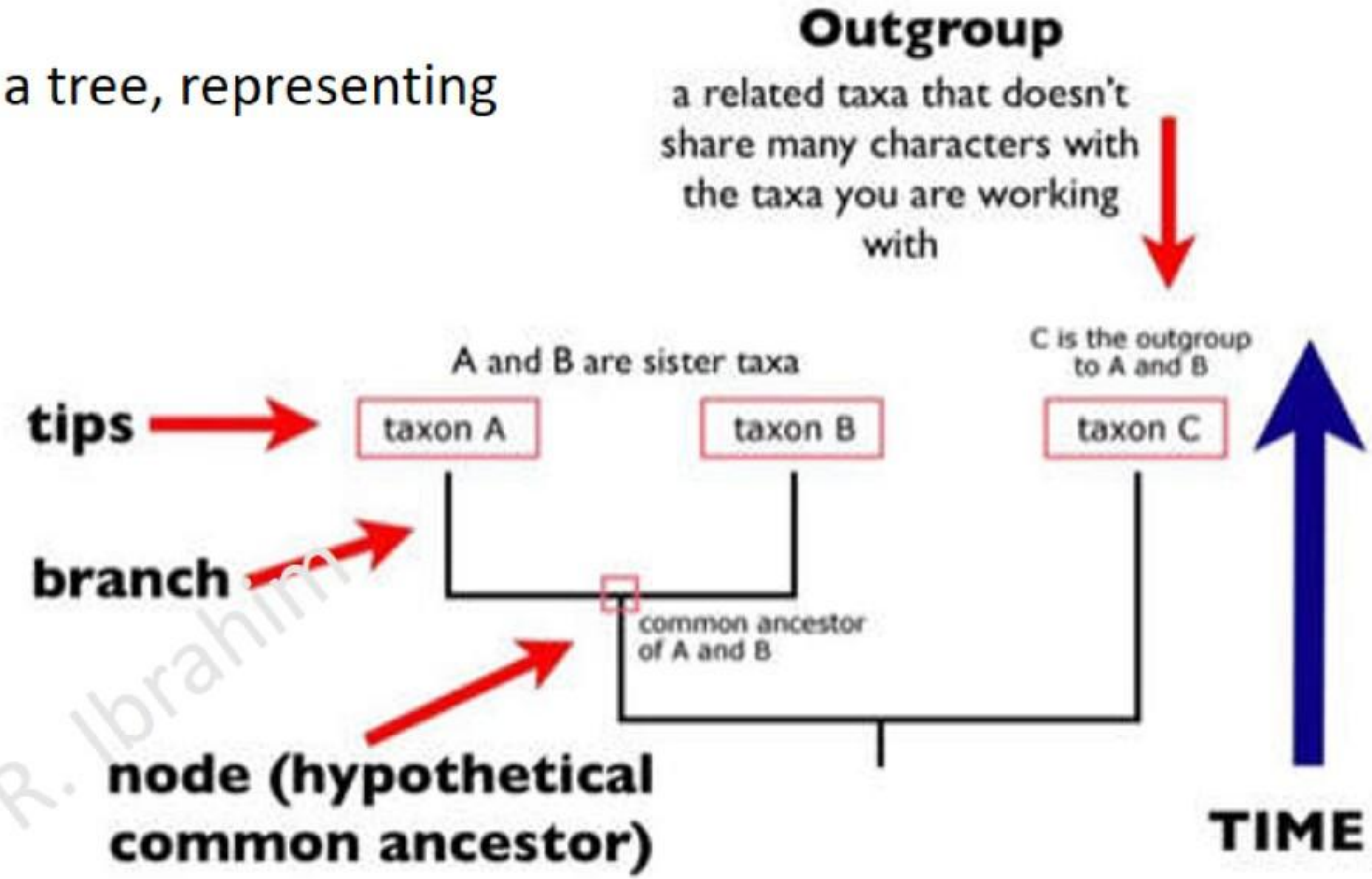
Root: the common ancestor of all taxa



Some important terminologies related to phylogenetic tree :

Branch: A line connecting nodes in a tree, representing evolutionary change

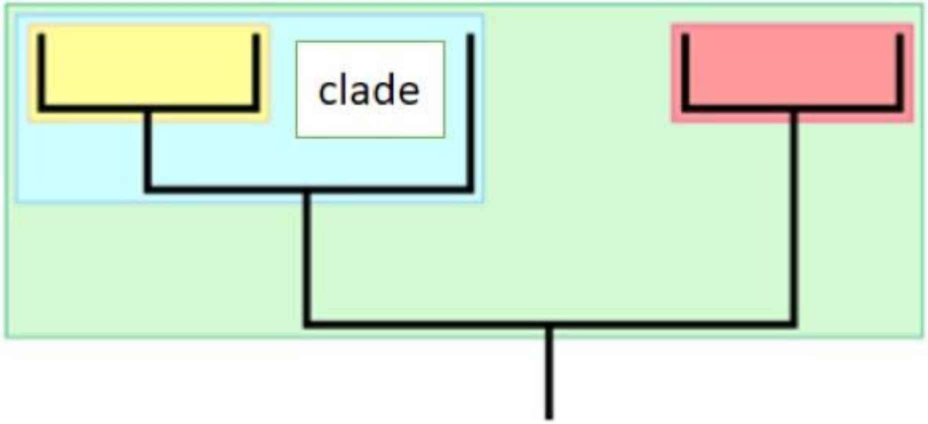
Sister taxa are two groups (species or other taxonomic units) that share an immediate common ancestor.

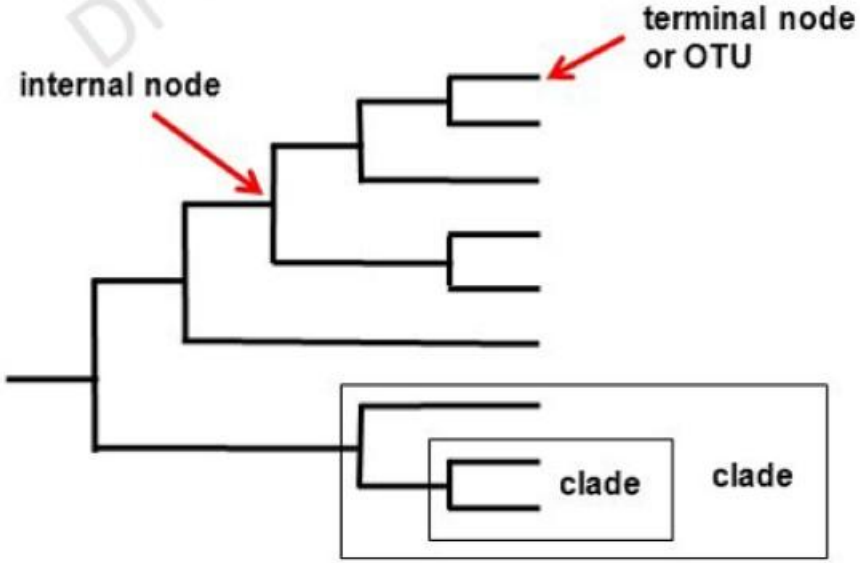


"**ROOTED** tree" - relationships in context of time

Some important terminologies related to phylogenetic tree

Clade: a group of two or more taxa or OTU that includes both their common ancestor and all their descendants.



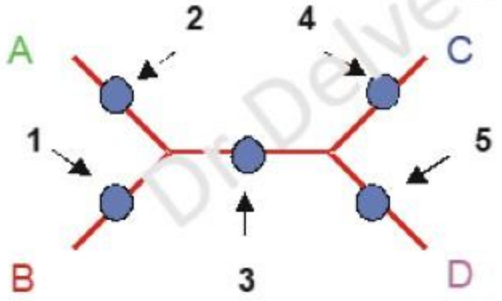


Phylogenetic trees

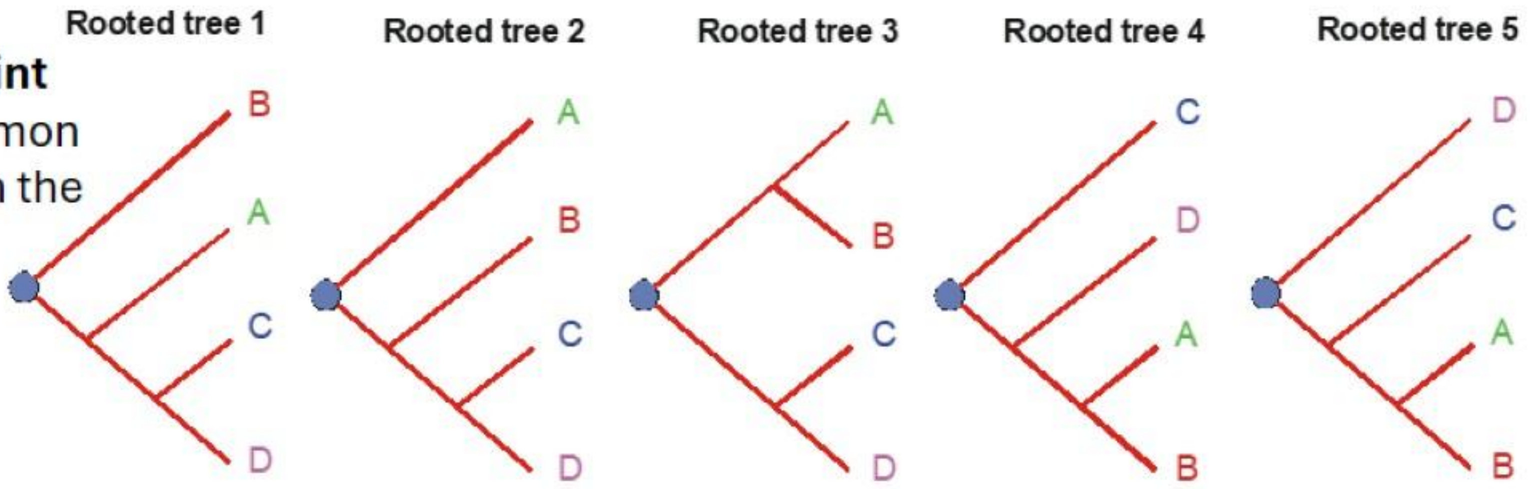
Trees can be unrooted or rooted:

- ## Unrooted trees
- A tree that shows **relationships** but **does not** indicate a common ancestor.

Unrooted tree

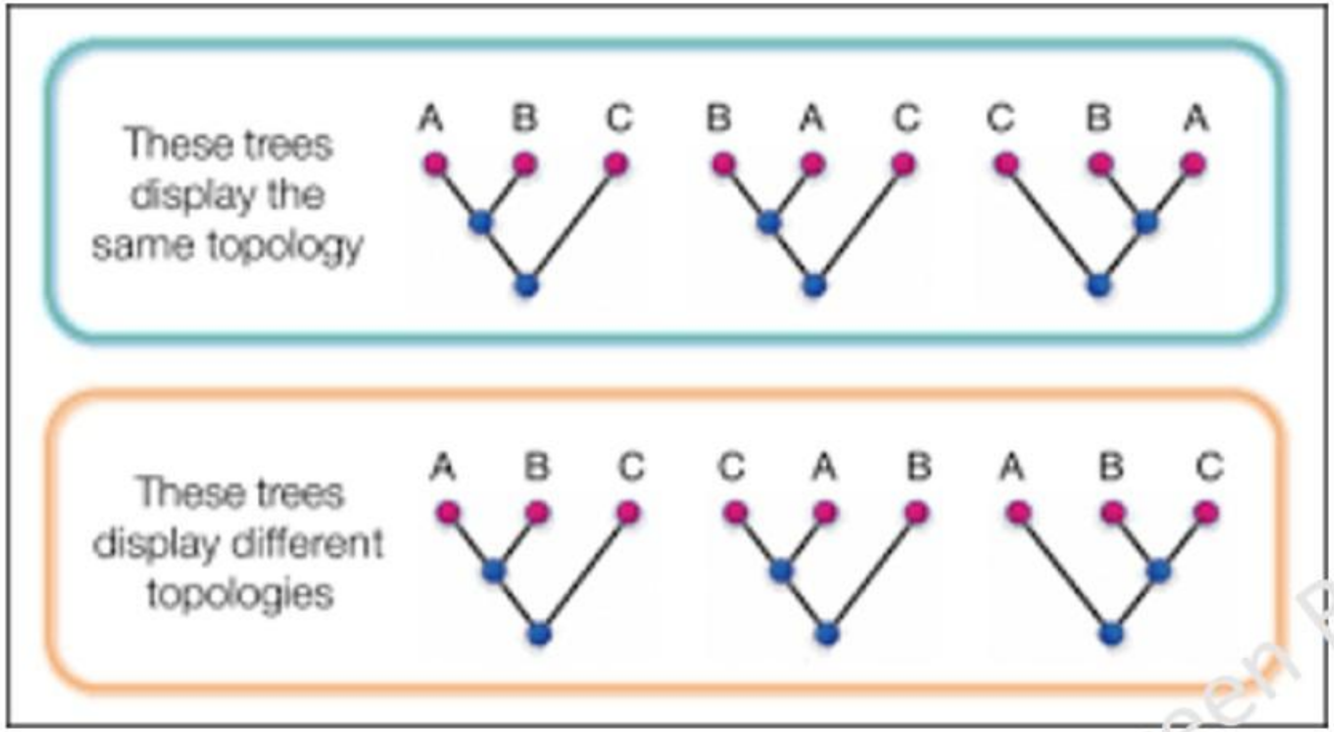


- ## Rooted trees
- A tree with a **starting point (root)**, showing the common ancestor of all species in the tree.



These trees show five different evolutionary relationships among the taxa!

Topology: represents the branching pattern of the tree. Branches can rotate on internal nodes.



Construction of a phylogenetic tree involves the following steps:

(1) Selection of the appropriate molecular marker (genes/proteins/mitochondrial DNA/ rRNA)

(2) Multiple sequence alignment

(3) Selection of a model of evolution (ex: **Nucleotide Substitution Models** (For DNA Trees): These models describe how one nucleotide (A, T, C, G) **mutates into another** over time.

(4) Construction of the phylogenetic tree methods.

(5) Assessment of the reliability of the tree (ex: bootstrapping)

4. Construction of the phylogenetic tree Methods

- This is a highly specialized area of computation and statistics. Therefore, only some overall principles are discussed here. The methods to construct phylogenetic trees can be classified into two major types:
- **(1) Distance-based methods**: These methods calculate how different two sequences are and then cluster them based on their similarity. For example, the unweighted pair group method with arithmetic mean (UPGMA) and Neighbor Joining (NJ).

(2) Character-based method

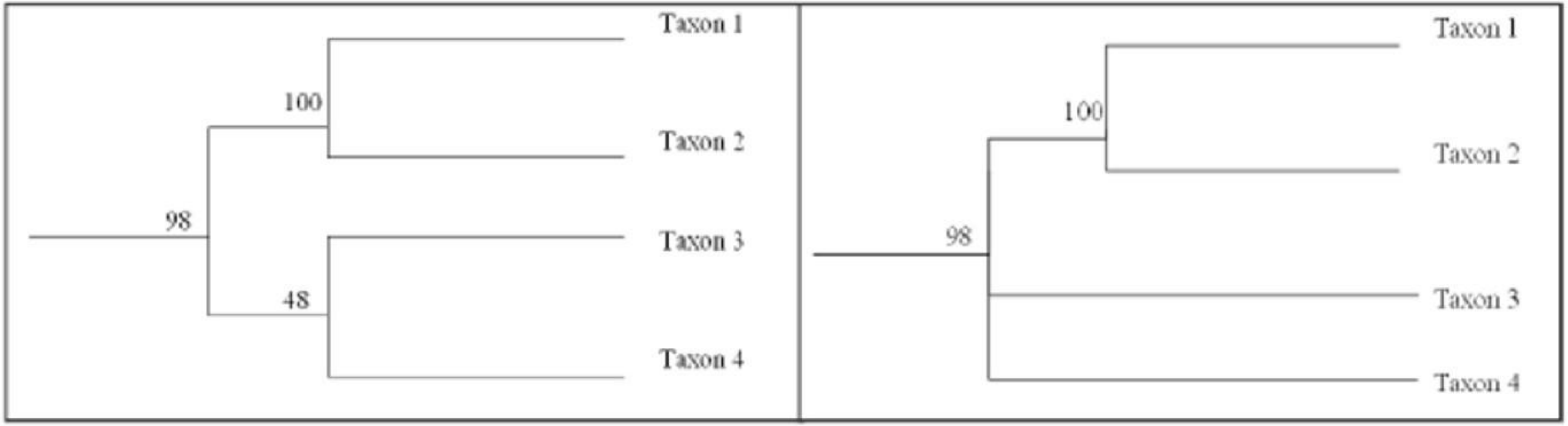
- Also called the discrete method: utilize the sequence itself rather than the pairwise distance obtained from the sequence features. There are three popular character-based methods:
- Maximum parsimony (MP)
- Maximum likelihood (ML)
- **Bayesian Inference** (BI)

5. Assessment of the reliability of the tree

- **Bootstrapping** in phylogenetics is a method used to check how **reliable** a tree is. It works by:

1. **Resampling** the original data multiple times (usually 1000 times is recommended).
2. **Building new trees** from these samples.
3. **Checking how often a specific branch appears in these trees.**

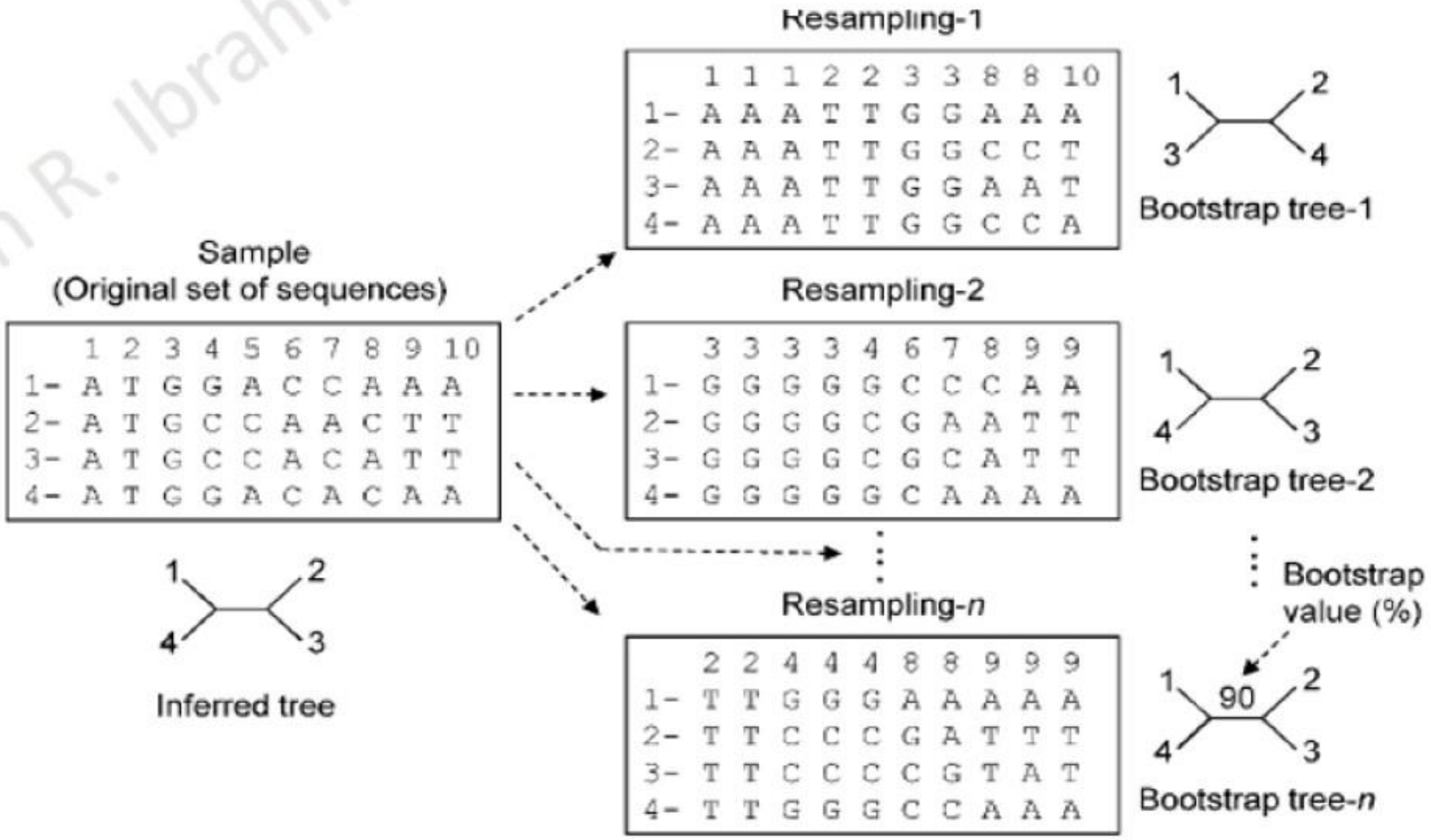
- **How to interpret bootstrap values?**
- **Higher values (e.g., 90-100%)** → Strong support for that branch.
- **Lower values (e.g., below 70%)** → Weak support, meaning the relationship may not be reliable.



Standard

Collapsed

Principles of bootstrapping the phylogenetic tree



Sample
(Original set of sequences)

	1	2	3	4	5	6	7	8	9	10
1-	A	T	G	G	A	C	C	A	A	A
2-	A	T	G	C	C	A	A	C	T	T
3-	A	T	G	C	C	A	C	A	T	T
4-	A	T	G	G	A	C	A	C	A	A

Inferred tree

Resampling-1

	1	1	1	2	2	3	3	8	8	10
1-	A	A	A	T	T	G	G	A	A	A
2-	A	A	A	T	T	G	G	C	C	T
3-	A	A	A	T	T	G	G	A	A	T
4-	A	A	A	T	T	G	G	C	C	A

Bootstrap tree-1

Resampling-2

	3	3	3	3	4	6	7	8	9	9
1-	G	G	G	G	G	C	C	C	A	A
2-	G	G	G	G	C	G	A	A	T	T
3-	G	G	G	G	C	G	C	A	T	T
4-	G	G	G	G	G	C	A	A	A	A

Bootstrap tree-2

Resampling-*n*

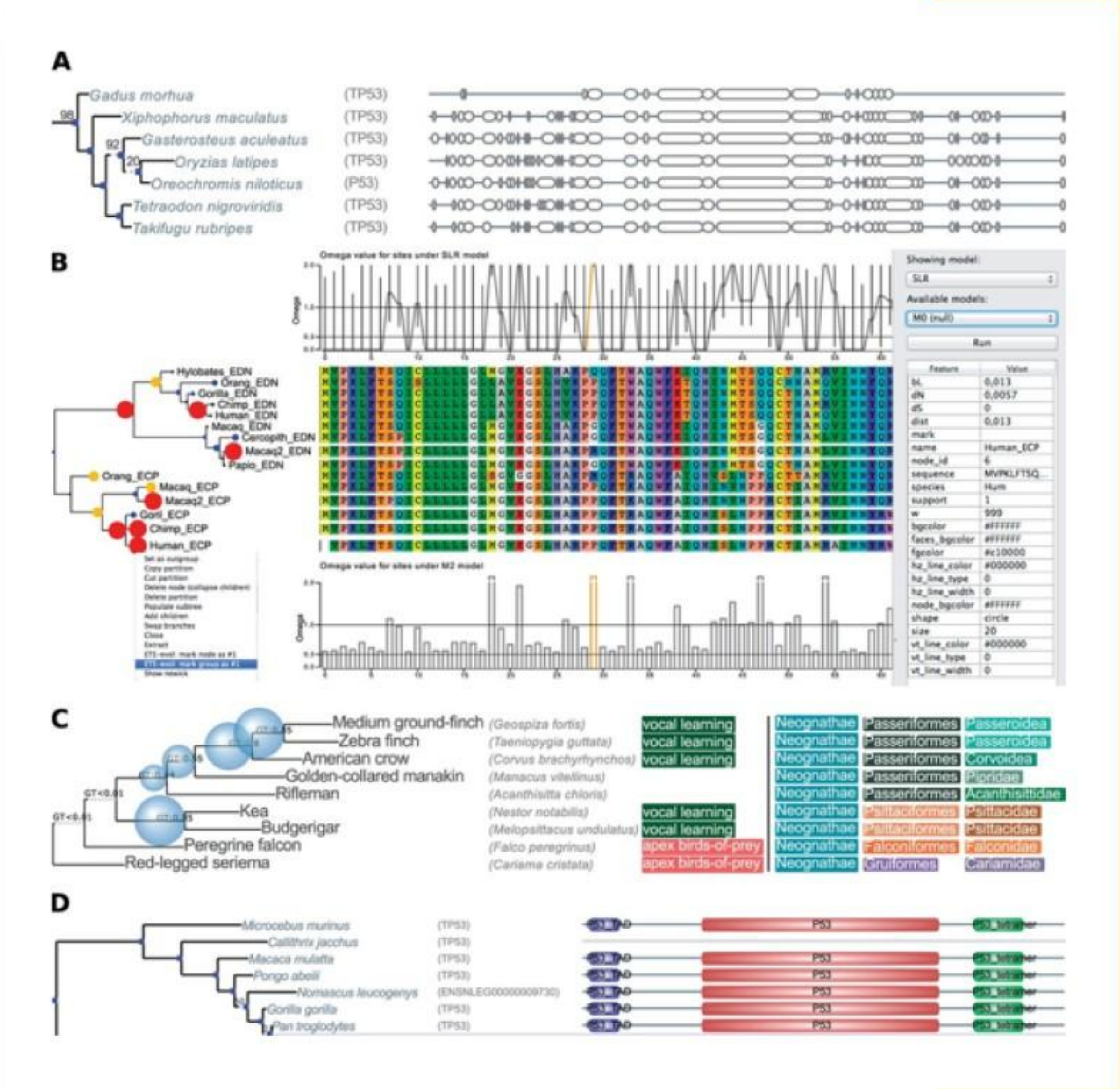
	2	2	4	4	4	8	8	9	9	9
1-	T	T	G	G	G	A	A	A	A	A
2-	T	T	C	C	C	G	A	T	T	T
3-	T	T	C	C	C	C	G	T	A	T
4-	T	T	G	G	G	C	C	A	A	A

Bootstrap value (%)

Bootstrap tree-*n*

Phylogenetic tree

Practical part



Note:

Generally, **sequence alignment** is used for **detecting mutation and variation**, **structure and function of the gene**, **evolutionary relationship between sequences (gene/protein).**

Some gene related terminologies:

1. **Homologous** genes: **Genes that share a common ancestor**. While homologous genes can be similar in sequence, similar sequences are not necessarily homologous. There are two types: **Orthologs and Paralogs**

2. **Analogs:** Genes (or gene products) that **do similar functions** but **do NOT come from a common ancestor**. They evolved **independently**

Term	Meaning	Example
Homologous genes	Share a common ancestor	β-globin gene in mice and rats
Orthologs	Homologs in different species	β-globin gene in mice and rats
Paralogs	Homologs within the same species	Human hemoglobin and myoglobin genes
Analogous genes	Same function, different origin	Human hemoglobin and insect hemocyanin

Types of Distance Trees Produced by BLAST

- When using **BLAST (Basic Local Alignment Search Tool)**, you can generate **distance-based phylogenetic trees** based on sequence similarity. These trees are built using **pairwise distances** between sequences and fall into two main types:
- Neighbor-Joining (NJ) Tree
- Fast Evolution Method

Phylogenetic tree based on rRNA

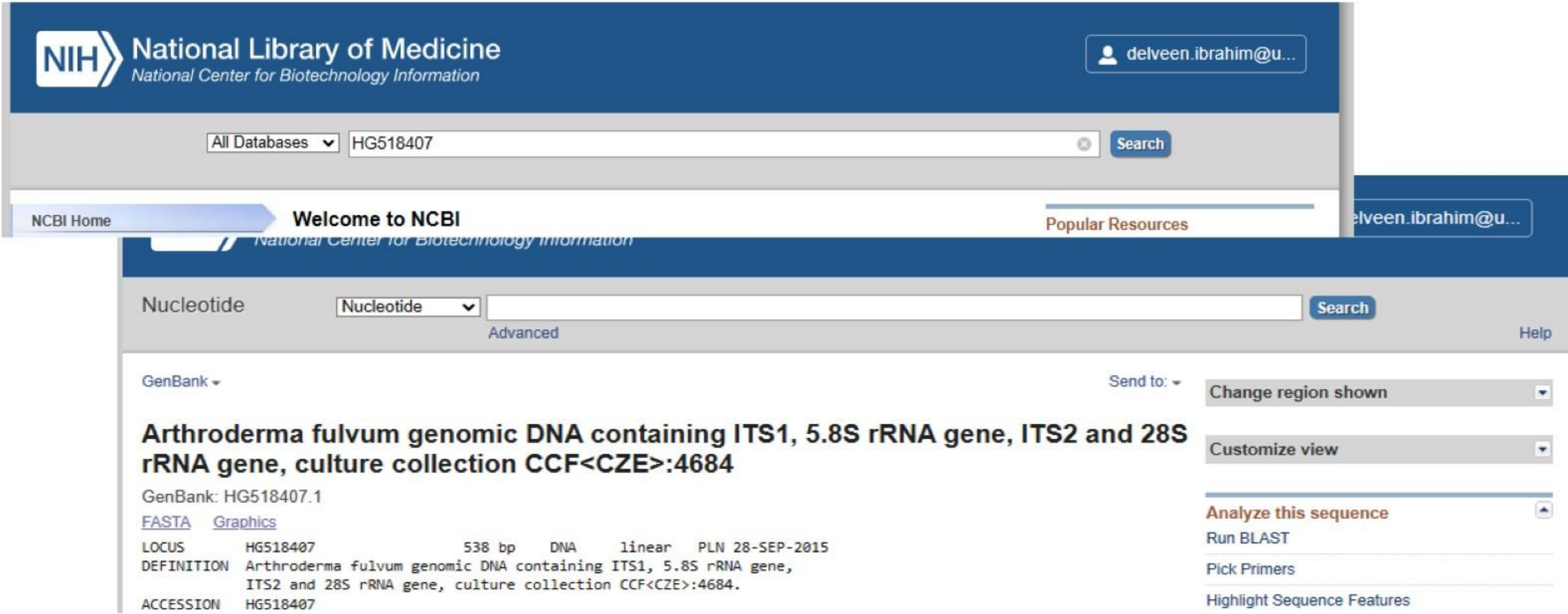
- You will be given a sequence for a specific region within rRNA, which is ITS (Internal transcribed spacer). Answer each of the following using BLAST:
- Your query sequence is (HG518407)

1. To which fungi does this accession number belongs
2. Find the relation between the identified fungi and the following sequences, with the following accession numbers.

- NR_155417.1
- NR_119467.1
- NR_168788.1
- NR_121226.1
- NR_144887.1
- NR_077142.1

3. Construct a phylogenetic tree using BLAST, record your results.

To find out to which fungi the provided accession number belong

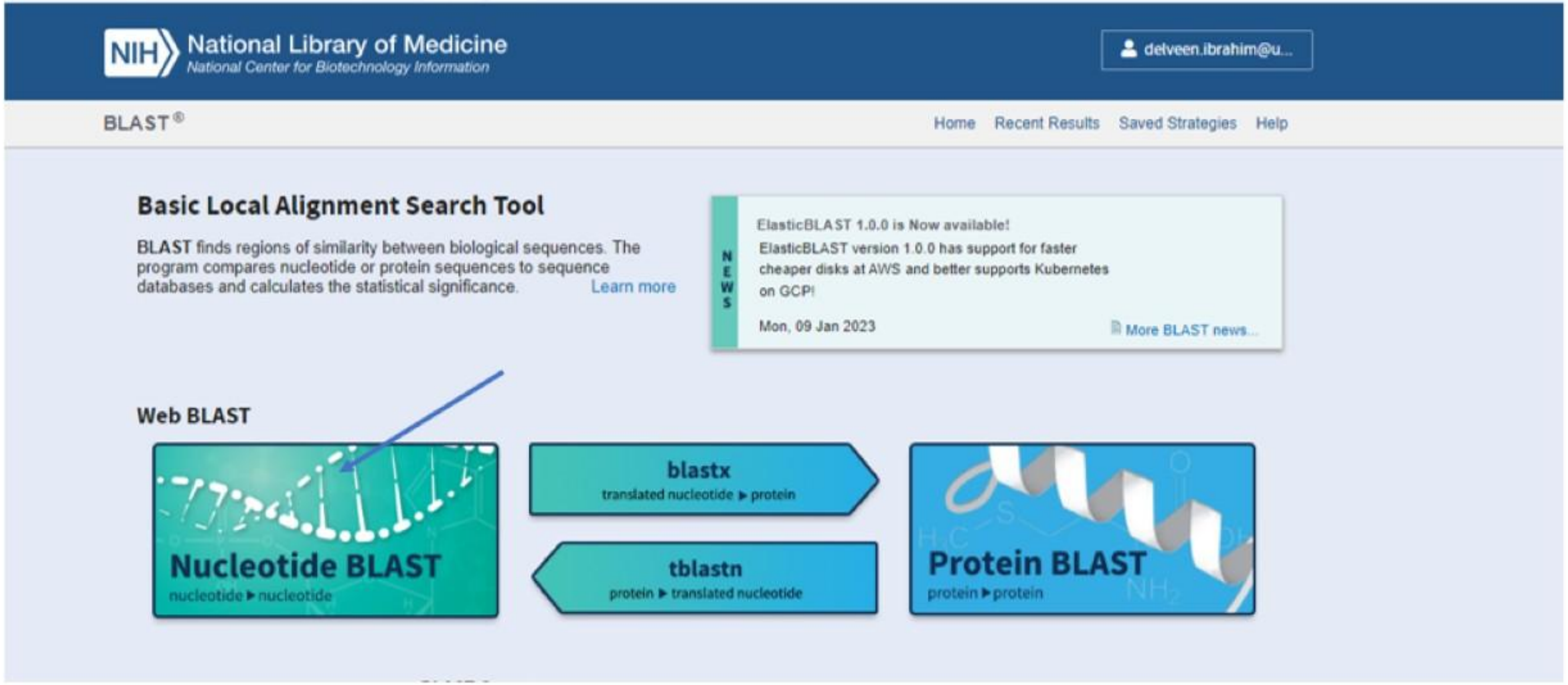


Arthroderma fulvum genomic DNA containing ITS1, 5.8S rRNA gene, ITS2 and 28S rRNA gene, culture collection CCF<CZE>:4684

GenBank: HG518407.1

```
LOCUS       HG518407                 538 bp    DNA     linear   PLN 28-SEP-2015
DEFINITION  Arthroderma fulvum genomic DNA containing ITS1, 5.8S rRNA gene,
            ITS2 and 28S rRNA gene, culture collection CCF<CZE>:4684.
ACCESSION   HG518407
```

For creating a distance tree, Go to BLAST and then choose (blastn)



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

NEWS

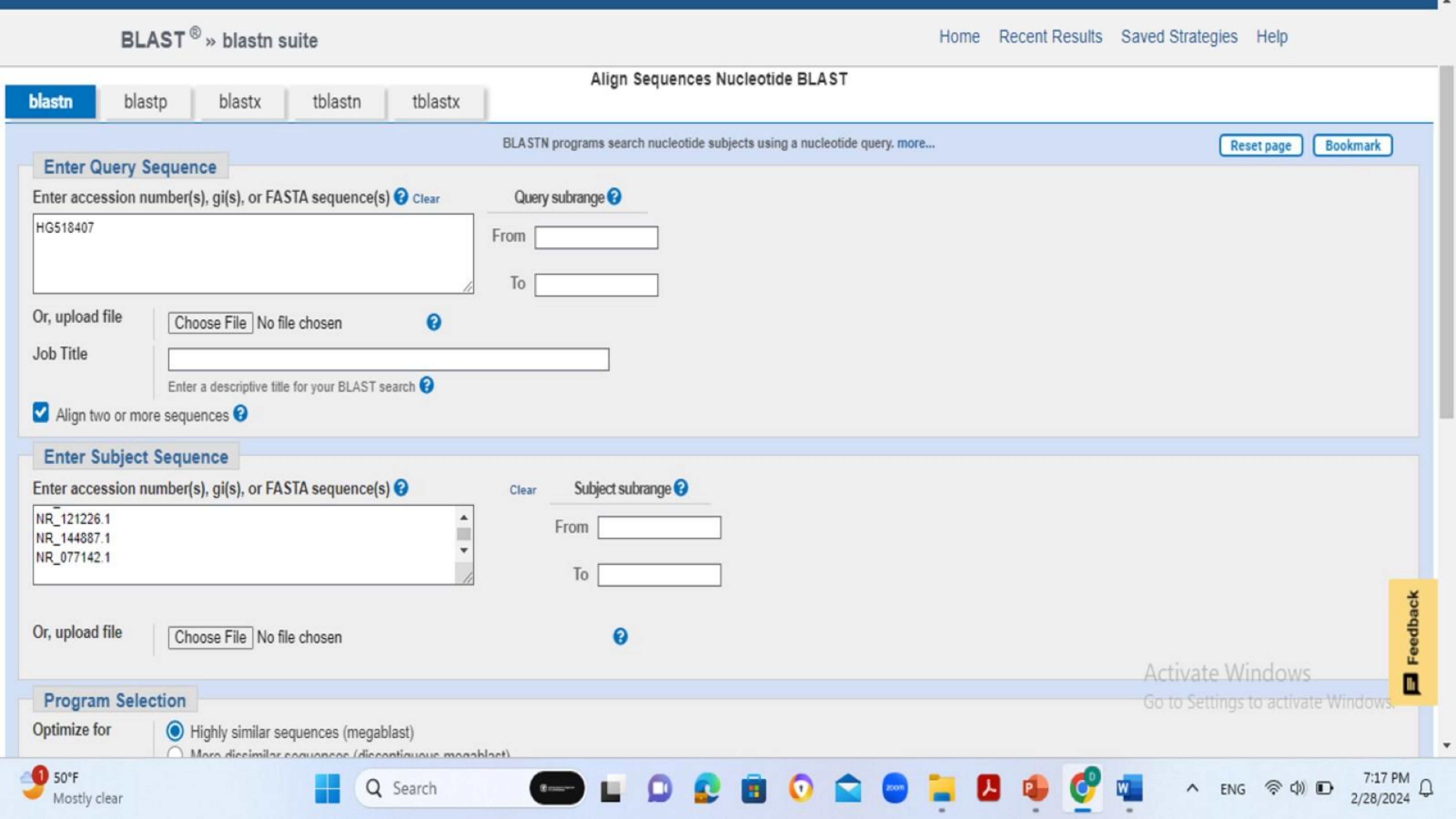
ElasticBLAST 1.0.0 is Now available!

ElasticBLAST version 1.0.0 has support for faster cheaper disks at AWS and better supports Kubernetes on GCP!

Mon, 09 Jan 2023

More BLAST news...

Web BLAST



Align Sequences Nucleotide BLAST

BLASTN programs search nucleotide subjects using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

HG518407

Query subrange
From
To

Or, upload file: No file chosen

Job Title:

Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

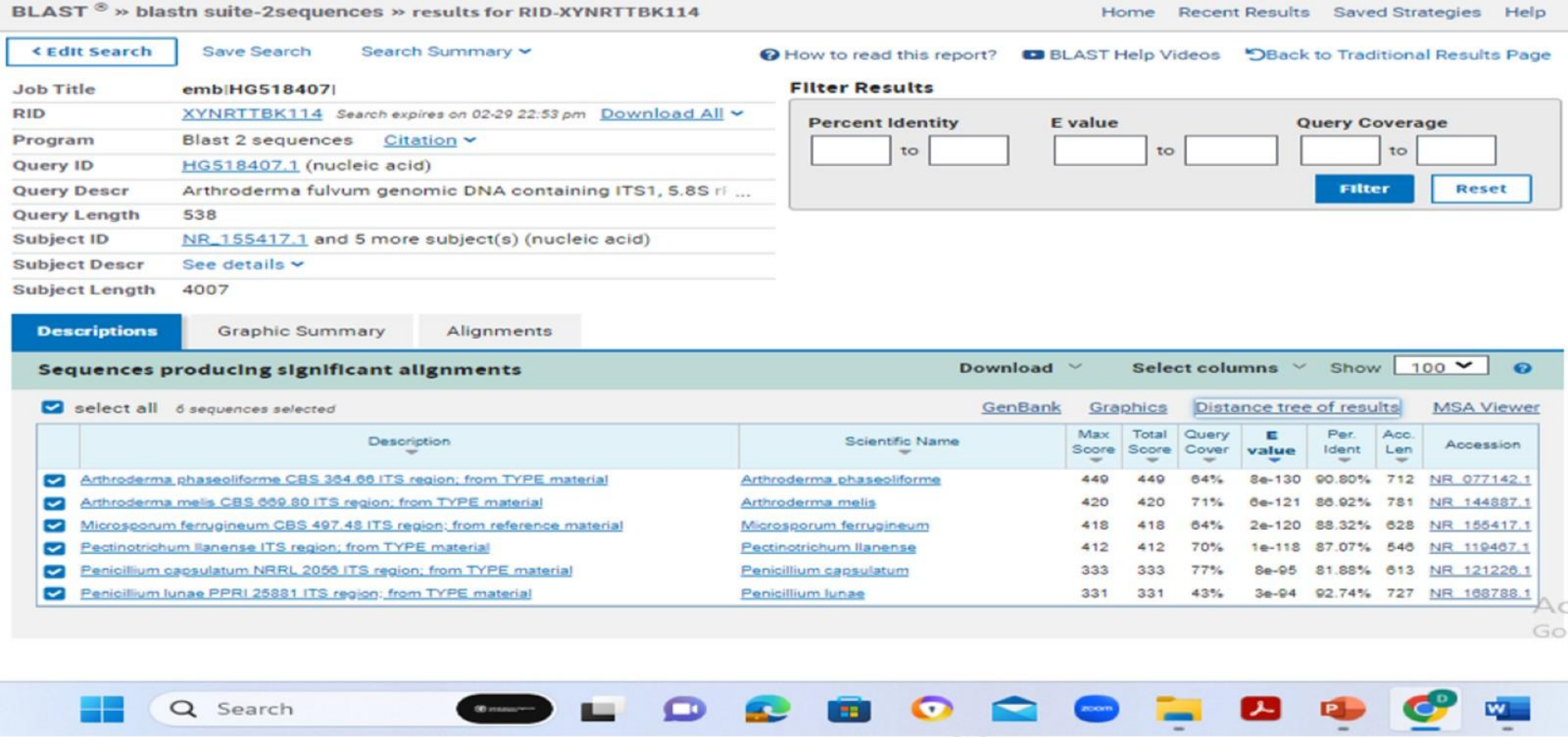
NR_121226.1
NR_144887.1
NR_077142.1

Subject subrange
From
To

Or, upload file: No file chosen

Program Selection

Optimize for
● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)



Job Title emb|HG518407|

RID XYNRTTBK114 *Search expires on 02-29 22:53 pm* Download All

Program Blast 2 sequences Citation

Query ID HG518407.1 (nucleic acid)

Query Descr Arthroderma fulvum genomic DNA containing ITS1, 5.8S ri ...

Query Length 538

Subject ID NR_155417.1 and 5 more subject(s) (nucleic acid)

Subject Descr See details

Subject Length 4007

Filter Results

Percent Identity ☐ to ☐

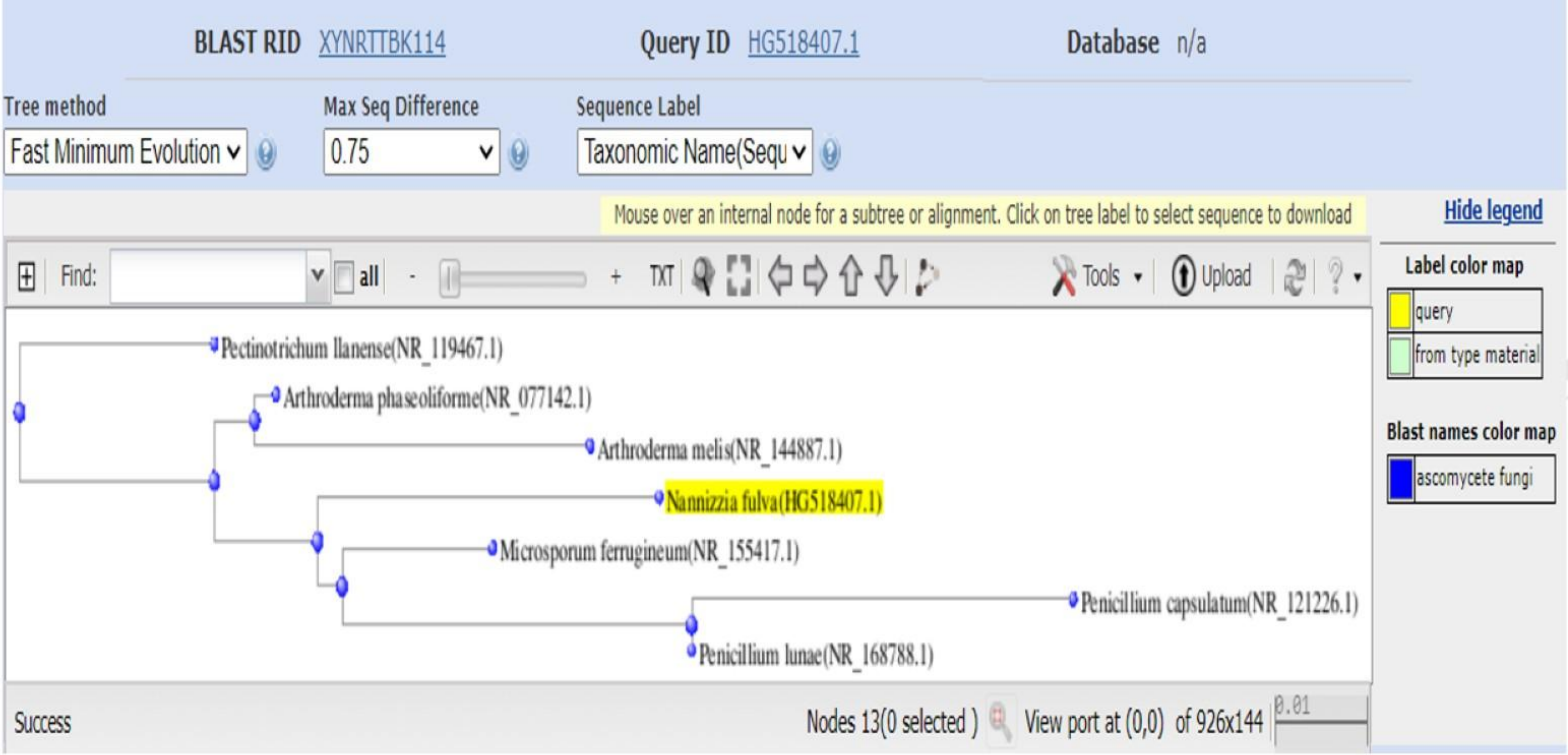
E value ☐ to ☐

Query Coverage ☐ to ☐

Sequences producing significant alignments

☑ select all *6 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Arthroderma phaseoliforme CBS 364.66 ITS region; from TYPE material	Arthroderma phaseoliforme	449	449	64%	8e-130	90.80%	712	NR_077142.1
☑	Arthroderma melis CBS 669.80 ITS region; from TYPE material	Arthroderma melis	420	420	71%	6e-121	86.92%	781	NR_144887.1
☑	Microsporum ferrugineum CBS 497.48 ITS region; from reference material	Microsporum ferrugineum	418	418	64%	2e-120	88.32%	628	NR_155417.1
☑	Pectinotrichum llanense ITS region; from TYPE material	Pectinotrichum llanense	412	412	70%	1e-118	87.07%	546	NR_119467.1
☑	Penicillium capsulatum NRRL 2056 ITS region; from TYPE material	Penicillium capsulatum	333	333	77%	8e-95	81.88%	613	NR_121226.1
☑	Penicillium lunae PPRI 25881 ITS region; from TYPE material	Penicillium lunae	331	331	43%	3e-94	92.74%	727	NR_168788.1



BLAST RID XYNRTTBK114

Query ID HG518407.1

Database n/a

Tree method: Fast Minimum Evolution

Max Seq Difference: 0.75

Sequence Label: Taxonomic Name(Sequ

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Success

Nodes 13(0 selected)

View port at (0,0) of 926x144