Extracting and analyzing DNA from an unknown organism

Zach Ritter
James Madison University

Introduction

- **Biodiversity** is dwindling at an alarming rate (Herbert et al. 2003), yet we still don't know all the species that are present even in our back yards.

- **DNA barcoding** identifies organisms based on DNA sequences and sequence databases, enabling **novices** to ID local organisms and catalogue biodiversity (Herbert et al. 2003).

Goal

The goal of this study was to learn the techniques used in DNA barcoding while identifying a sample collected from the JMU Arboretum.
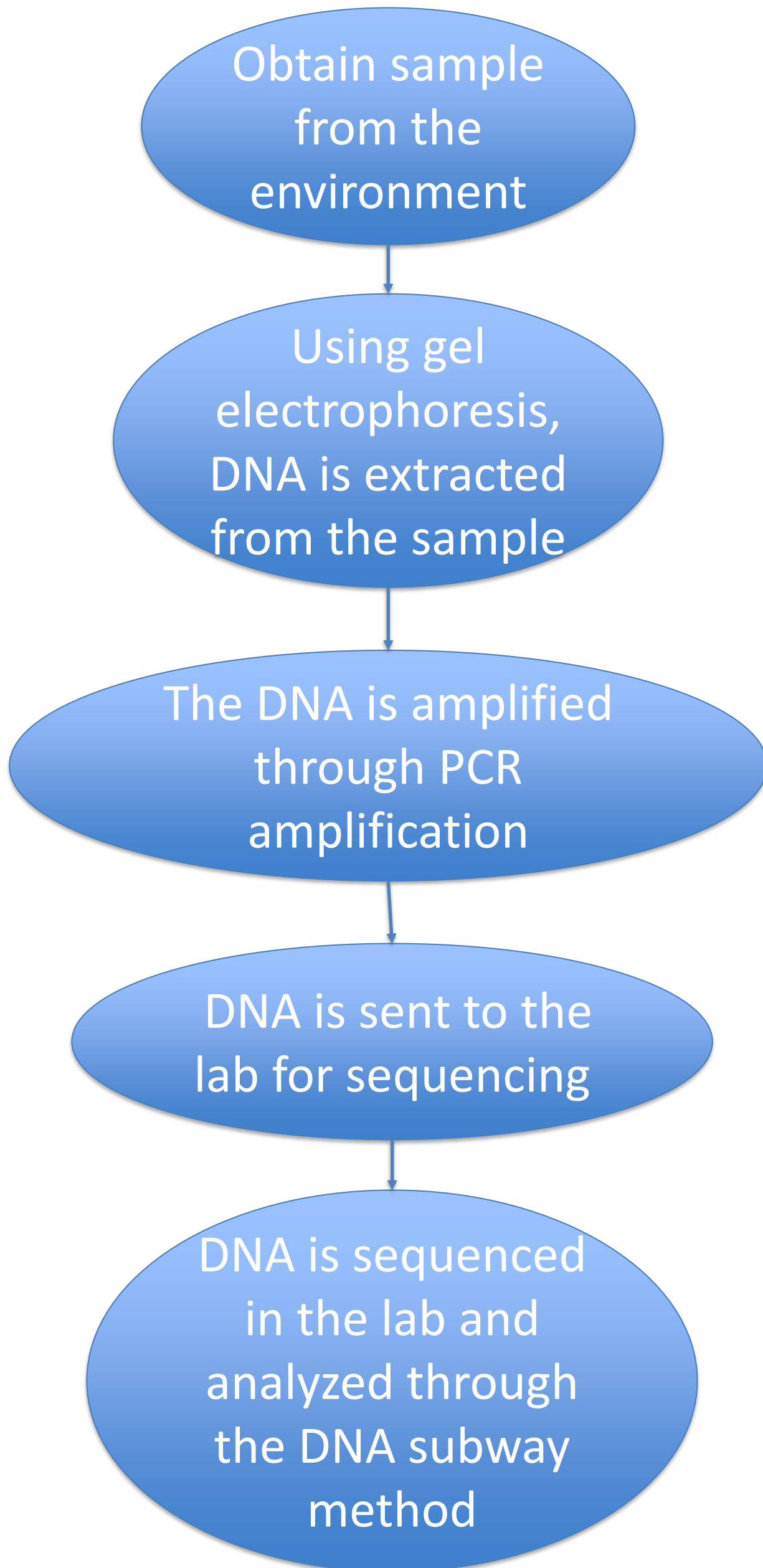
Methods

Figure 1. The step by step method for taking a sample and extracting DNA to analyze

- Obtain sample from the environment
- Using gel electrophoresis, DNA is extracted from the sample
- The DNA is amplified through PCR amplification
- DNA is sent to the lab for sequencing
- DNA is sequenced in the lab and analyzed through the DNA subway method

Results

Figure 2. Organism photo(s) collected from Harrisonburg, VA, Feb. 2020

Figure 3. Photo of best BLAST match – Western Red Cedar (*Thuja plicita*).
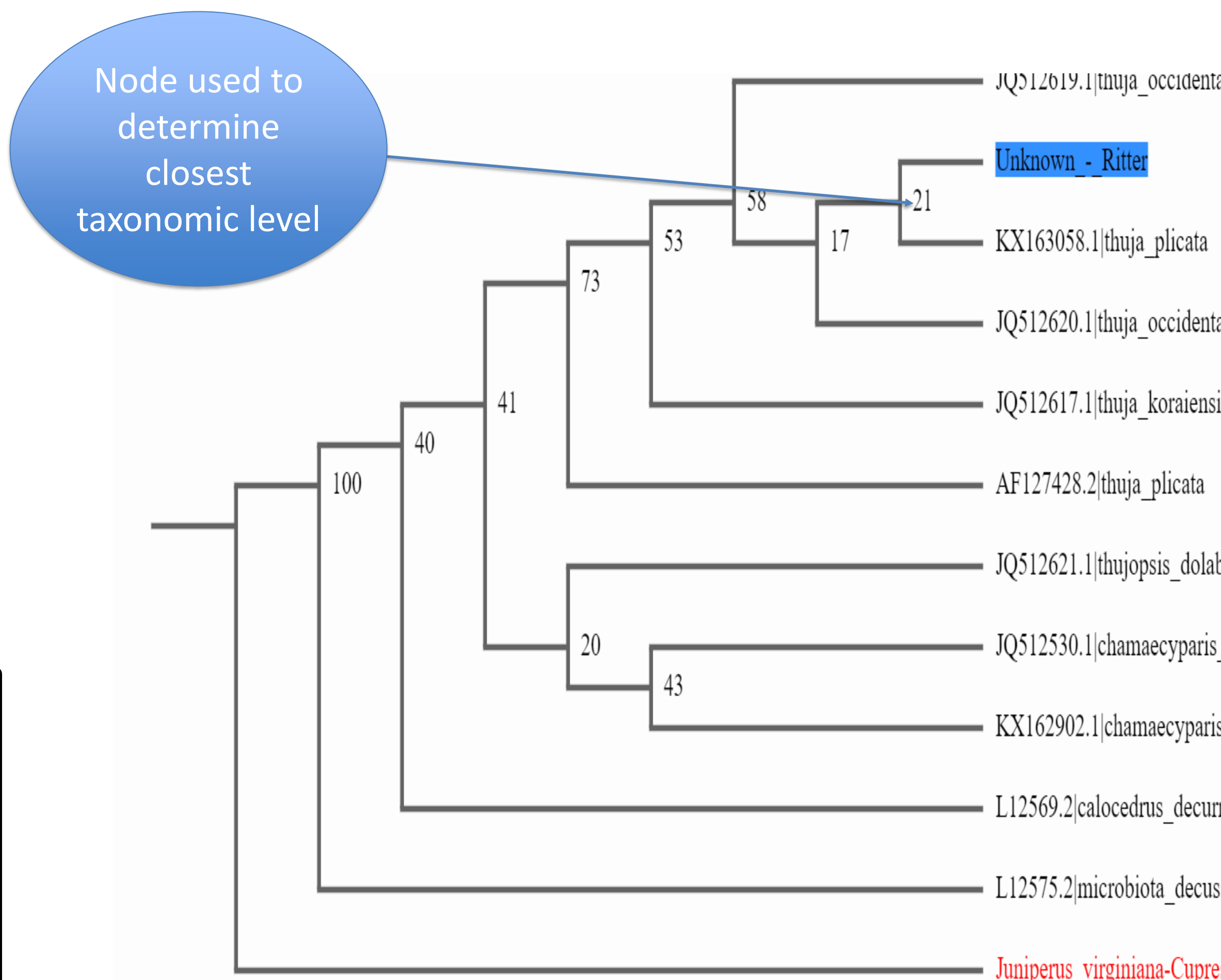Source: https://www.conifers.org/cu/Thuja_plicata.php

Figure 4. PHYLIP NJ tree of unknown sample (blue box), 10 selected BLAST hits, and 1 sequence from the common reference data, including the outgroup (red text).
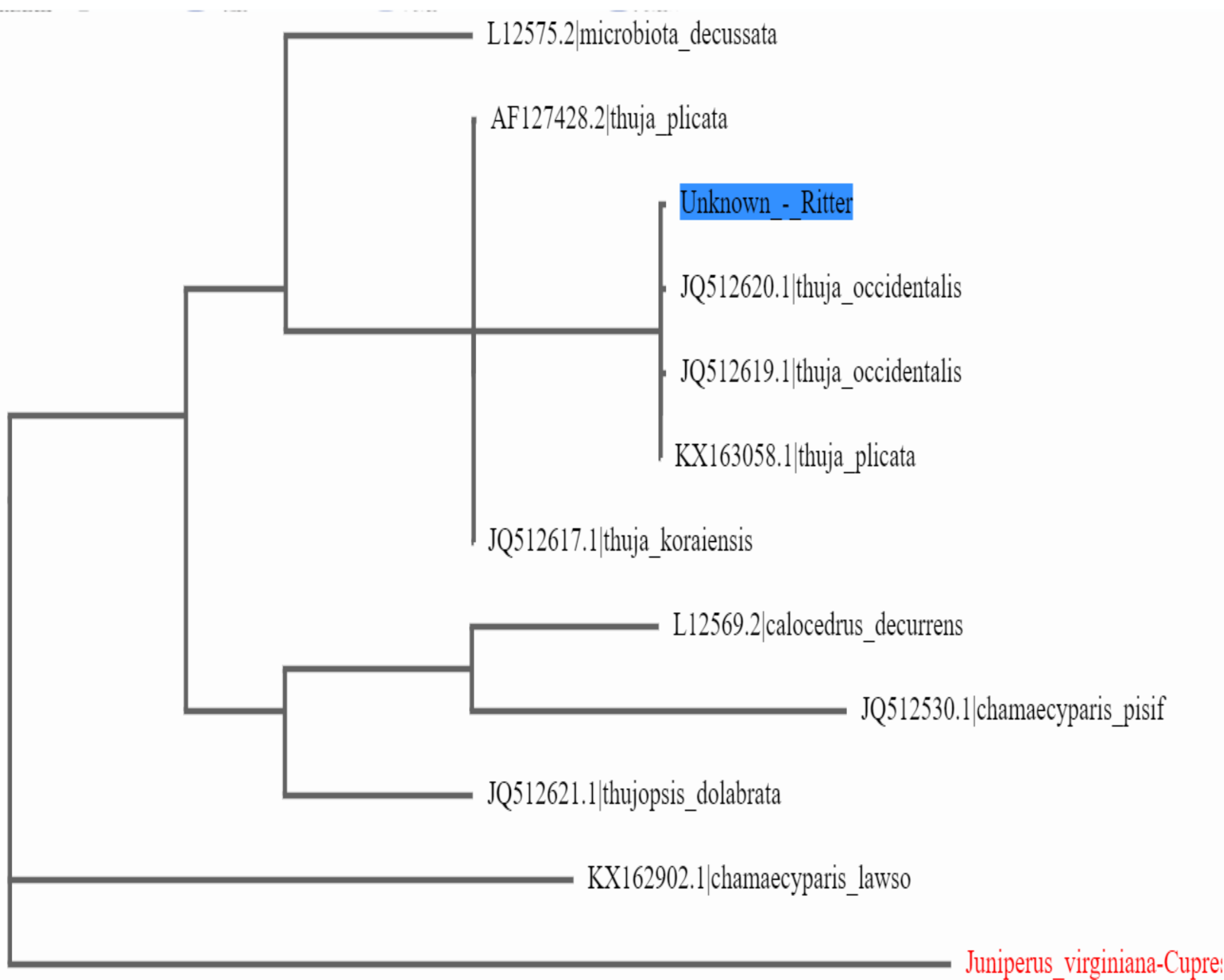

Figure 5. PHYLIP ML tree of unknown sample (blue box), 10 selected BLAST hits, and 1 sequence from the common reference data, including the outgroup (red text).
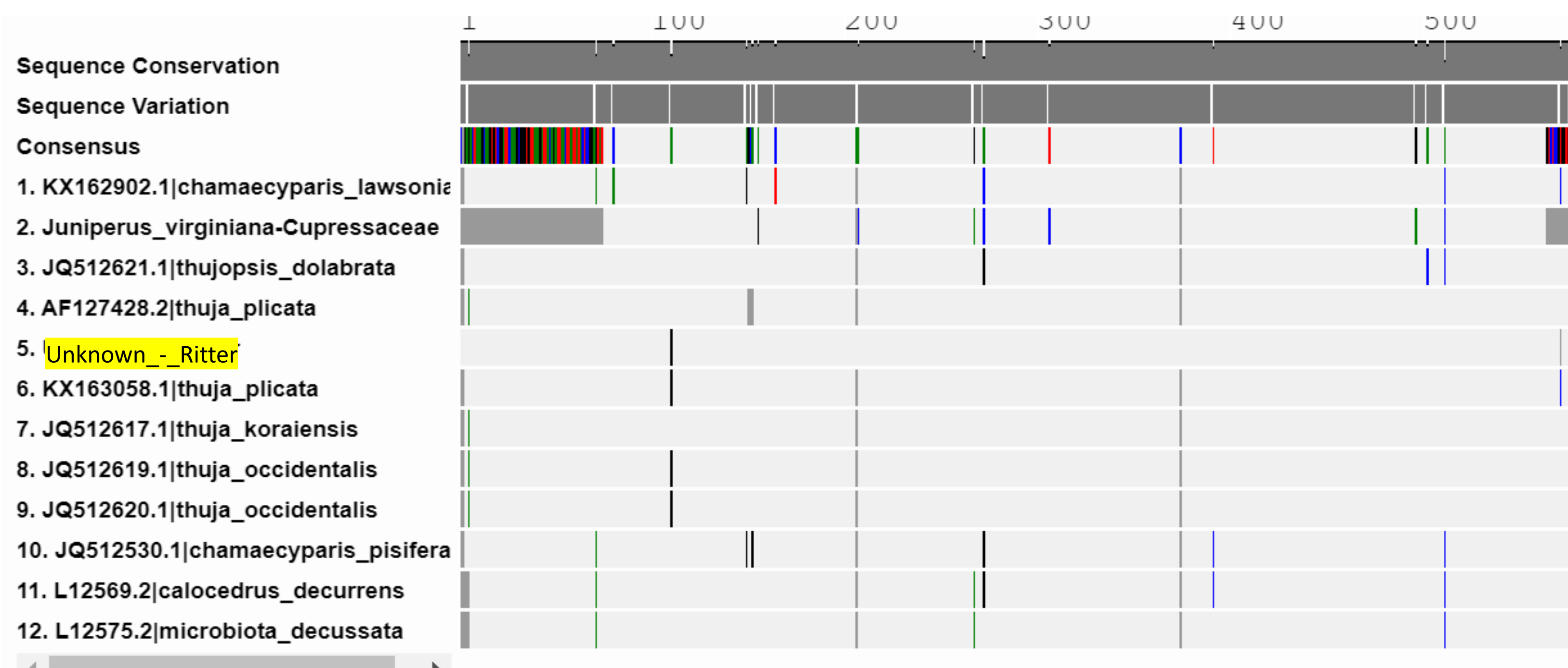

Figure 6. Trimmed MUSCLE alignment of the unknown sequence (highlighted), 10 selected BLAST hits, and 1 sequence(s) from common reference data. Colored vertical lines indicate sequence polymorphism. Details listed under public DNA subway project – Ritter_plant_April142020_BIO140, project ID 208060

Results

Table 1. Top BLAST matches arranged by bit score (query date: 14APR2020).

#	Accession #	Details	Aln. Length	Bit Score	e	Mis-matches
1(1).	KX163058.1	**Thuja plicata** - Thuja plicata voucher OMH-A60 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, partial cds	573	1014	0.0	1
2(2).	JQ512620.1	**Thuja occidentalis** - Thuja occidentalis isolate PDBK2009-0719 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, complete cds	573	1009	0.0	2
3(3).	JQ512619.1	**Thuja occidentalis** - Thuja occidentalis isolate PDBK2009-0406 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, complete cds	573	1009	0.0	2
4(4).	JQ512618.1	**Thuja occidentalis** - Thuja occidentalis isolate PDBK2009-0055 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, complete cds	573	1009	0.0	2
5(5).	AY237154.1	**Thuja plicata** -	573	1009	0.0	2
6(6).	JQ512617.1	**Thuja koraiensis** - Thuja koraiensis isolate PDBK2010-1013 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, complete cds	573	1005	0.0	3
7(7).	JQ512621.1	**Thujopsis dolabrata** - Thujopsis dolabrata isolate PDBK2009-0142 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, complete cds	573	996	0.0	5
8(8).	AF127428.2	**Thuja plicata** - Thuja ribulose-1,5-bisphosphate carboxylase large subunit (rbcL) gene,partial cds	573	994	0.0	6
9(9).	HQ590300.1	**Thuja occidentalis** - Thuja occidentalis voucher JAG 0016-b ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene,partial cds	562	994	0.0	1
10(10).	HQ590299.1	**Thuja occidentalis** - Thuja occidentalis voucher AP264 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene,partial cds	562	994	0.0	1

Table 2. Summary of taxonomic identification data

Technique	Best supported taxonomic ID(s)
BLAST	*Thuja plicata*
MUSCLE (% similarity)	*Thuja plicata (99.82%)*
NJ Tree	Genus: Thuja
ML Tree	Genus: Thuja
Morphology (photos)	Matches *Thuja plicata* features
FINAL TAXONOMIC ID	*Thuja plicata*

Discussion

Interpretation of DNA subway results

- SEQUENCE: The length of the sequence is longer than 400, making the sequence reliable and trustworthy. Also, there was no red triangle in DNA subway signaling weak sequences.
- BLAST. I am confident in the top blast hit because the bit score (1014) is extremely high for comparing organisms. Also, there is only 1 mismatch and the error is 0, meaning there is high evidence that the blast hit is a match.
- MUSCLE: I am confident in the muscle data for my organism because it shows a 99.82 percent match, which is extremely high. This means only a few of the sequences did not line up, giving strong evidence of similarity.
- NJ TREE: The NJ tree shows a lot about the evolution of my organism. I can see that the thuja genus seems to be the prominent genus throughout the tree, as thuja is the genus on the node closet to the unknown organism.
- ML TREE: The ML tree gives the maximum likelihood of the evolution of the unknown organism. This shows that the organism evolved once again from the genus thuja as the two organisms before the unknown has the genus thuja.
- **Taxonomy of specimen**
- The unknown organism is *Thuja pliciata*, or western red cedar.
- The organism shows strong similarity in morphologic design of the western red cedar shown in the images online, as both are evergreen conifers with rough and folded leaves
- **Biodiversity**
- The western red cedar is mainly found in the Pacific Northwest in cool, moist environments.

References

Herbert, P., Cywinsky, A., Ball, S., and deWaard, J. (2003). Biological Identifications through DNA Barcodes. Proc. Roy. Soc. B. 270:313-321.