

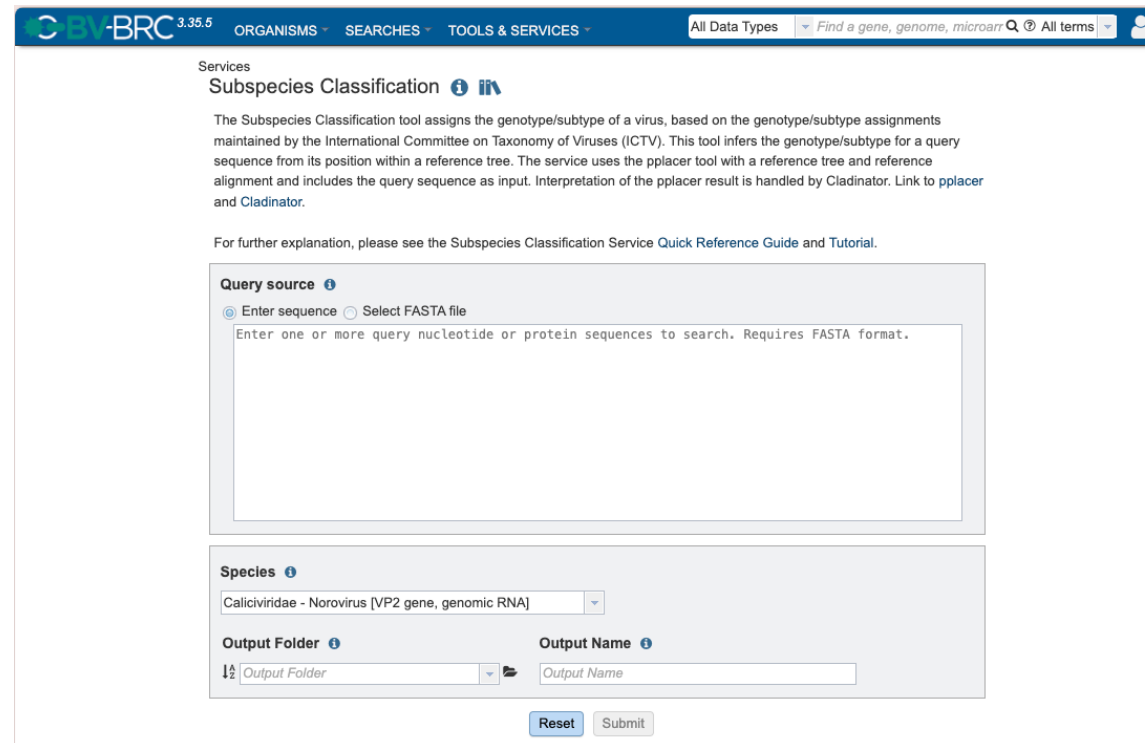
BV-BRC sub-species classification tools

Christian M. Zmasek, Ph.D.

J. Craig Venter Institute, BV-BRC

Sub-species classification tool in BV-BRC

<https://www.bv-brc.org/> > “TOOLS & SERVICES” > “Subspecies Classification (V)”



The screenshot shows the BV-BRC 3.35.5 website interface. The top navigation bar includes 'ORGANISMS', 'SEARCHES', and 'TOOLS & SERVICES'. A search bar contains the text 'Find a gene, genome, microarr' and a search icon. Below the navigation bar, the page title is 'Subspecies Classification'. A descriptive paragraph explains that the tool assigns the genotype/subtype of a virus based on ICTV assignments and uses the pplacer tool. Below the description, there are links to a 'Quick Reference Guide' and a 'Tutorial'. The main form area is divided into two sections: 'Query source' and 'Species'. The 'Query source' section has two radio buttons: 'Enter sequence' (selected) and 'Select FASTA file'. Below these is a text input field with the placeholder text 'Enter one or more query nucleotide or protein sequences to search. Requires FASTA format.' The 'Species' section has a dropdown menu currently showing 'Caliciviridae - Norovirus [VP2 gene, genomic RNA]'. Below the species dropdown are two input fields: 'Output Folder' and 'Output Name'. At the bottom of the form are 'Reset' and 'Submit' buttons.

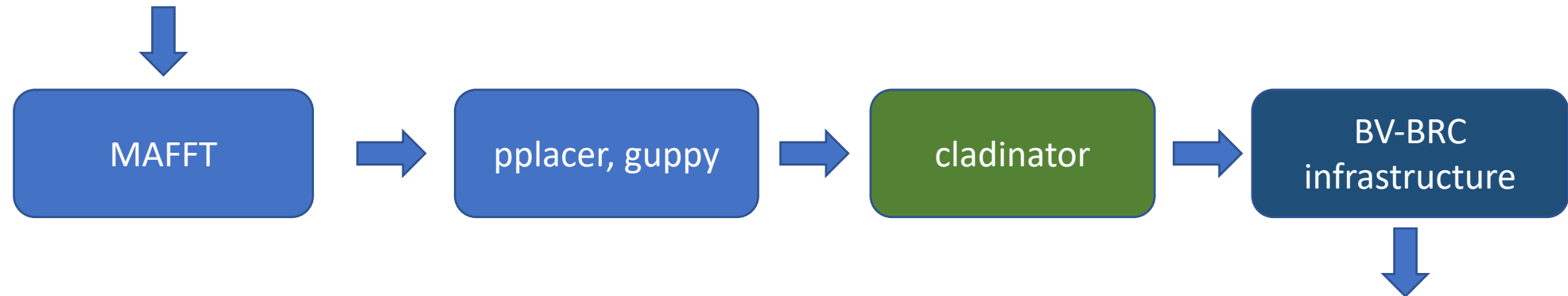
Sub-species classification tool in BV-BRC: Approach

Maximum likelihood placement of query sequence (gene, protein, or whole genome) on a curated reference phylogenetic tree

Family	Genus	Species (Serotype)	Query	BV-BRC tool	Status	
<i>Caliciviridae</i>	<i>Norovirus</i>	Norwalk virus	VP1 gene, genomic RNA	maximum likelihood placement	complete	
			VP2 gene, genomic RNA	maximum likelihood placement	complete	
<i>Flaviviridae</i>	<i>Flavivirus</i>	Dengue virus virus	complete genome, genomic RNA	maximum likelihood placement	complete	
		Japanese encephalitis virus	complete genome, genomic RNA	maximum likelihood placement	complete	
		Murray Valley encephalitis virus	envelope protein E, genomic RNA	maximum likelihood placement	complete	
		St. Louis encephalitis virus	polyprotein gene, genomic RNA	maximum likelihood placement	complete	
		Tick-borne encephalitis virus	polyprotein gene, genomic RNA	maximum likelihood placement	complete	
		West Nile virus	complete genome, genomic RNA	maximum likelihood placement	complete	
		Yellow fever virus	polyprotein gene, genomic RNA	maximum likelihood placement	complete	
		Zika virus	complete genome, genomic RNA	maximum likelihood placement	complete	
		<i>Hepacivirus</i>	Hepatitis C virus	polyprotein gene, genomic RNA	maximum likelihood placement	complete
		<i>Pestivirus</i>	Bovine viral diarrhea virus	5' UTR region, genomic RNA	maximum likelihood placement	complete
<i>Orthomyxoviridae</i>	<i>Alphainfluenzavirus</i>	Influenza A virus H5	Hemagglutinin gene, genomic RNA	maximum likelihood placement	complete	
		Swine Influenza (H1) [global]	Hemagglutinin gene, genomic RNA	maximum likelihood placement	complete	
		Swine Influenza (H1) [US]	Hemagglutinin gene, genomic RNA	maximum likelihood placement	complete	
		Swine Influenza (H3) [global]	Hemagglutinin gene, genomic RNA	maximum likelihood placement	complete	
<i>Sedoreoviridae</i>	<i>Rotavirus</i>	Rotavirus A	complete genome, genomic RNA	specialized tool	complete	
<i>Adenoviridae</i>	<i>Mastadenovirus</i>	Human mastadenovirus A	complete genome, genomic DNA	maximum likelihood placement	tested, ready	
		Human mastadenovirus B	complete genome, genomic DNA	maximum likelihood placement	tested, ready	
		Human mastadenovirus C	complete genome, genomic DNA	maximum likelihood placement	tested, ready	
		Human mastadenovirus E	complete genome, genomic DNA	maximum likelihood placement	tested, ready	
		Human mastadenovirus F	complete genome, genomic DNA	maximum likelihood placement	tested, ready	
<i>Paramyxoviridae</i>	<i>Morbillivirus</i>	Measles morbillivirus	complete genome, genomic RNA	maximum likelihood placement	tested, ready	
	<i>Orthorubulavirus</i>	Mumps orthorubulavirus	complete genome, genomic RNA	maximum likelihood placement	tested, ready	
<i>Pneumoviridae</i>	<i>Orthopneumovirus</i>	RSV, Orthopneumovirus hominis		maximum likelihood placement	investigating	

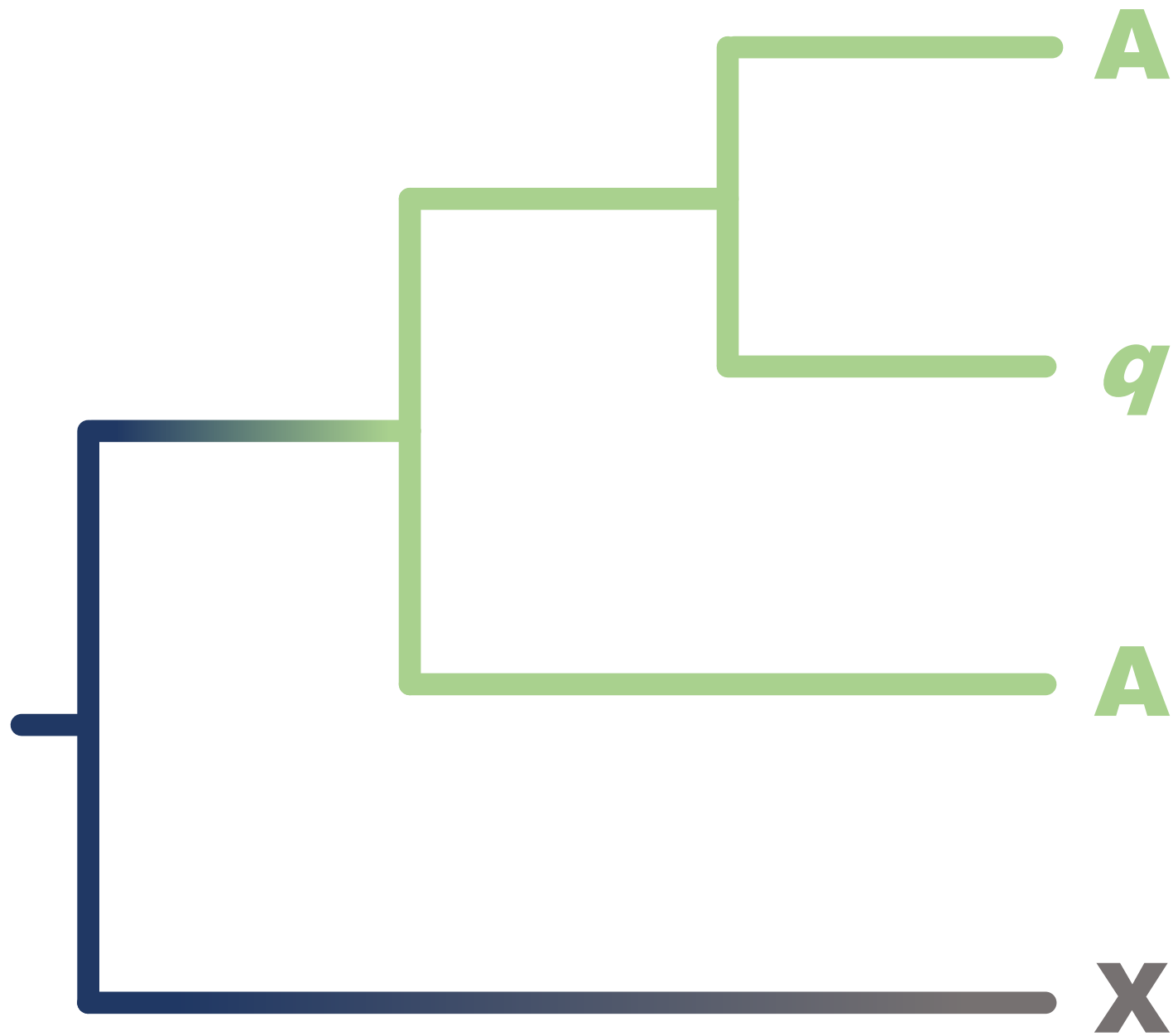
BV-BRC sub-species classification pipeline

- Expert-approved reference tree calculated by RAxML and resultant statistics output file
- Corresponding MSA
- User query sequence

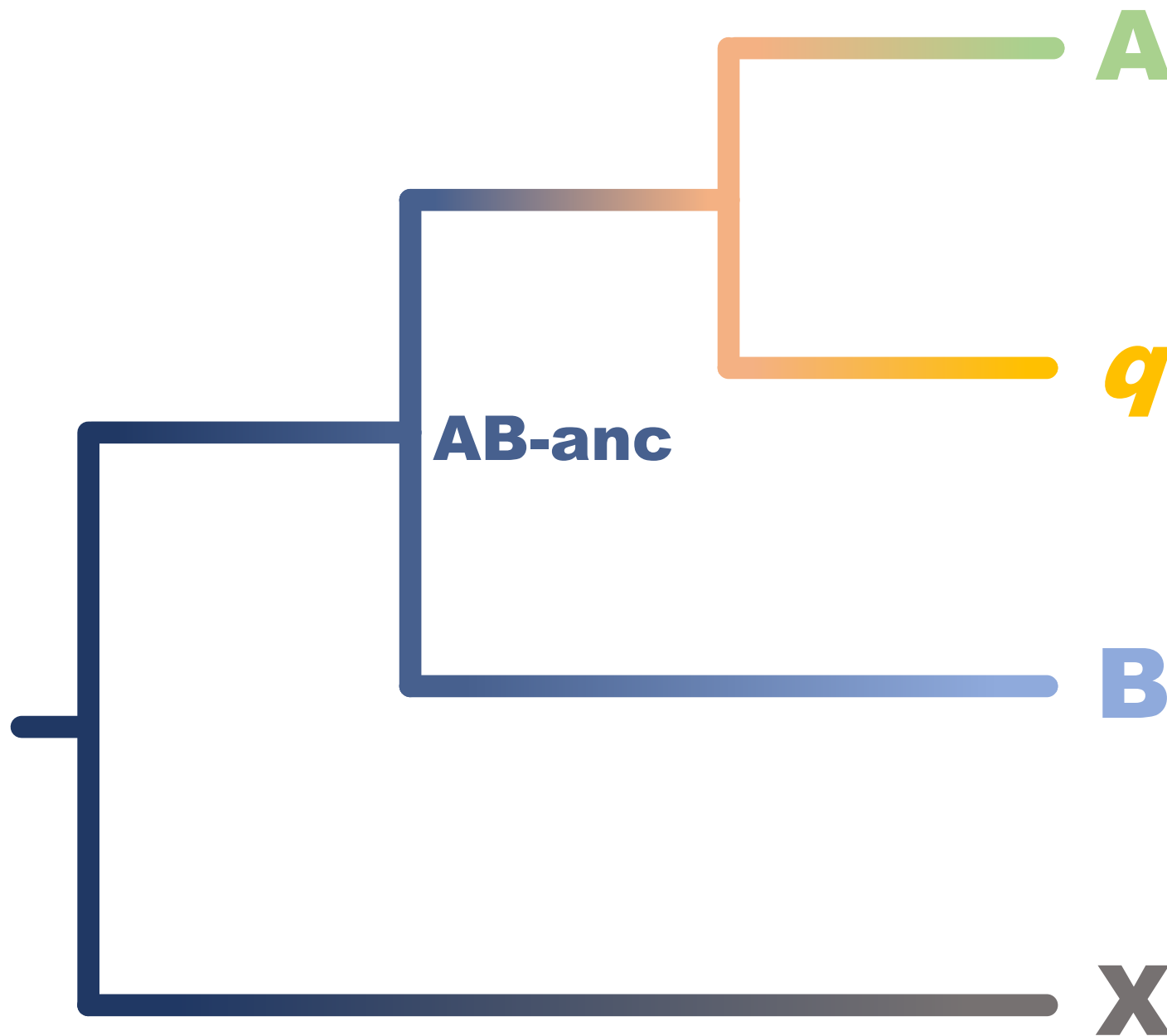


- User readable output
- Machine readable output
- Interactive tree with placement(s) of query (Archaeopteryx.js)

Cladinator: Analysis of query
placements in hierarchically
annotated phylogenetic trees

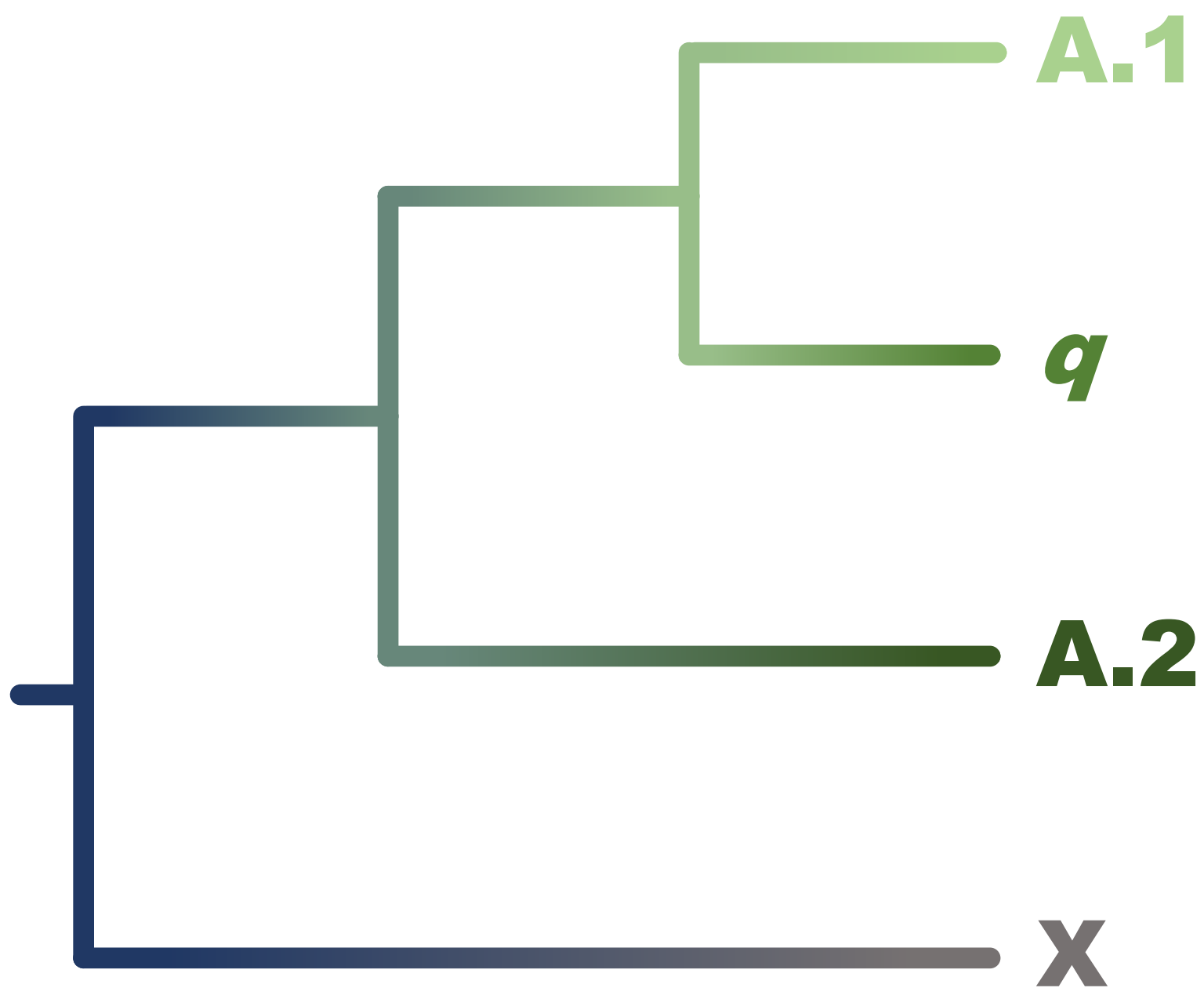


q is A-type:
bracketed by A and
A.



q is of unknown type: bracketed by A and B (it could be "C", "A", or "A.x").

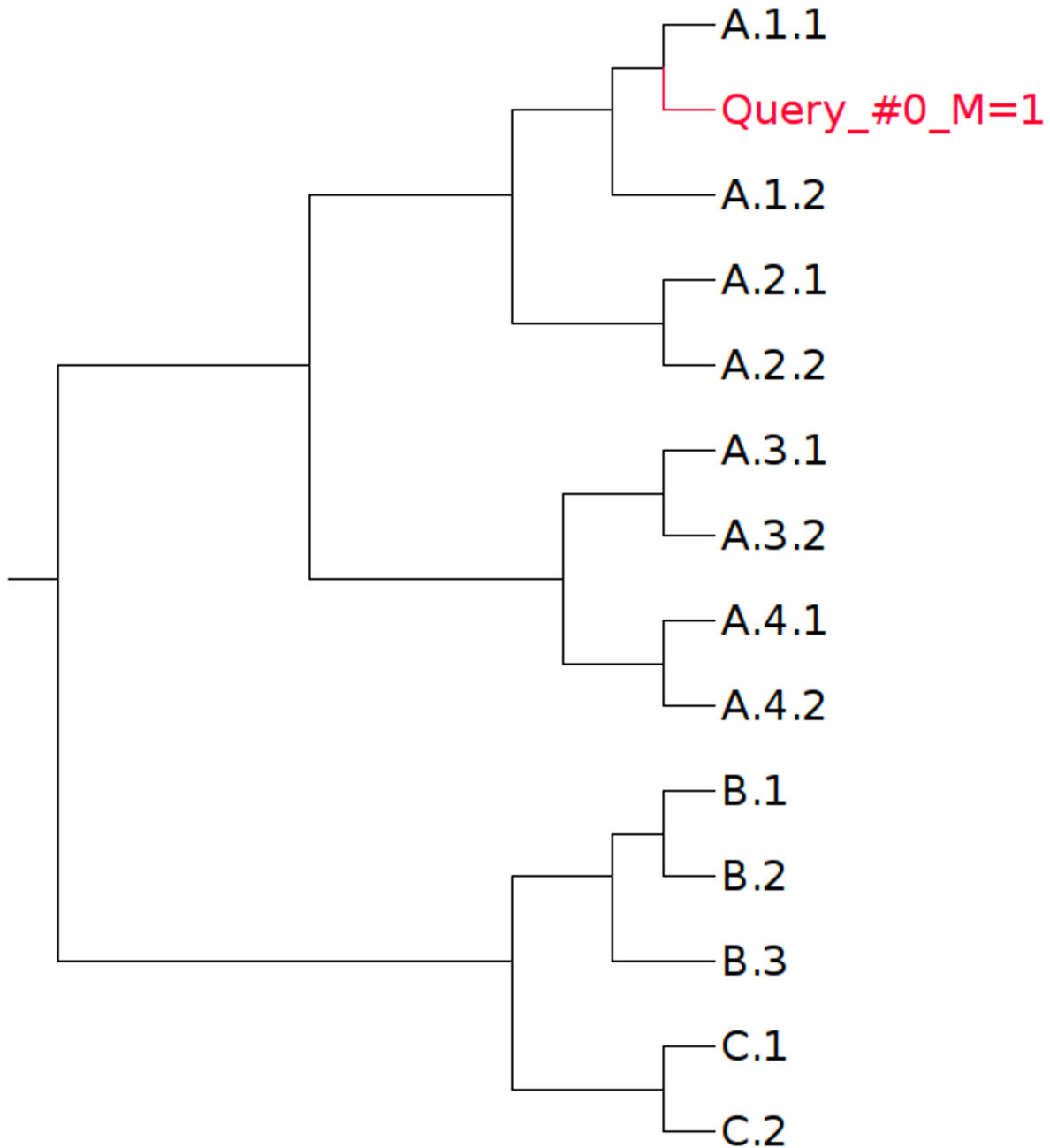
Naïvely, it looks like q must be of A-type, but we do **not** know at which point along the branch going from AB-ancestor to A, the type changes from AB-ancestor-type to A-type.



q is of A-type:
bracketed by A.1
and A.2 (it could
be "A.3", "A.1", or
"A.1.x").

cladinator

- Command-line program to analyze outputs produced by pplacer (or similar programs) when used to analyze reference trees with hierarchical annotations (such as virus type and subtype)
- Produces two output types:
 - One is human readable
 - One is easily parse-able



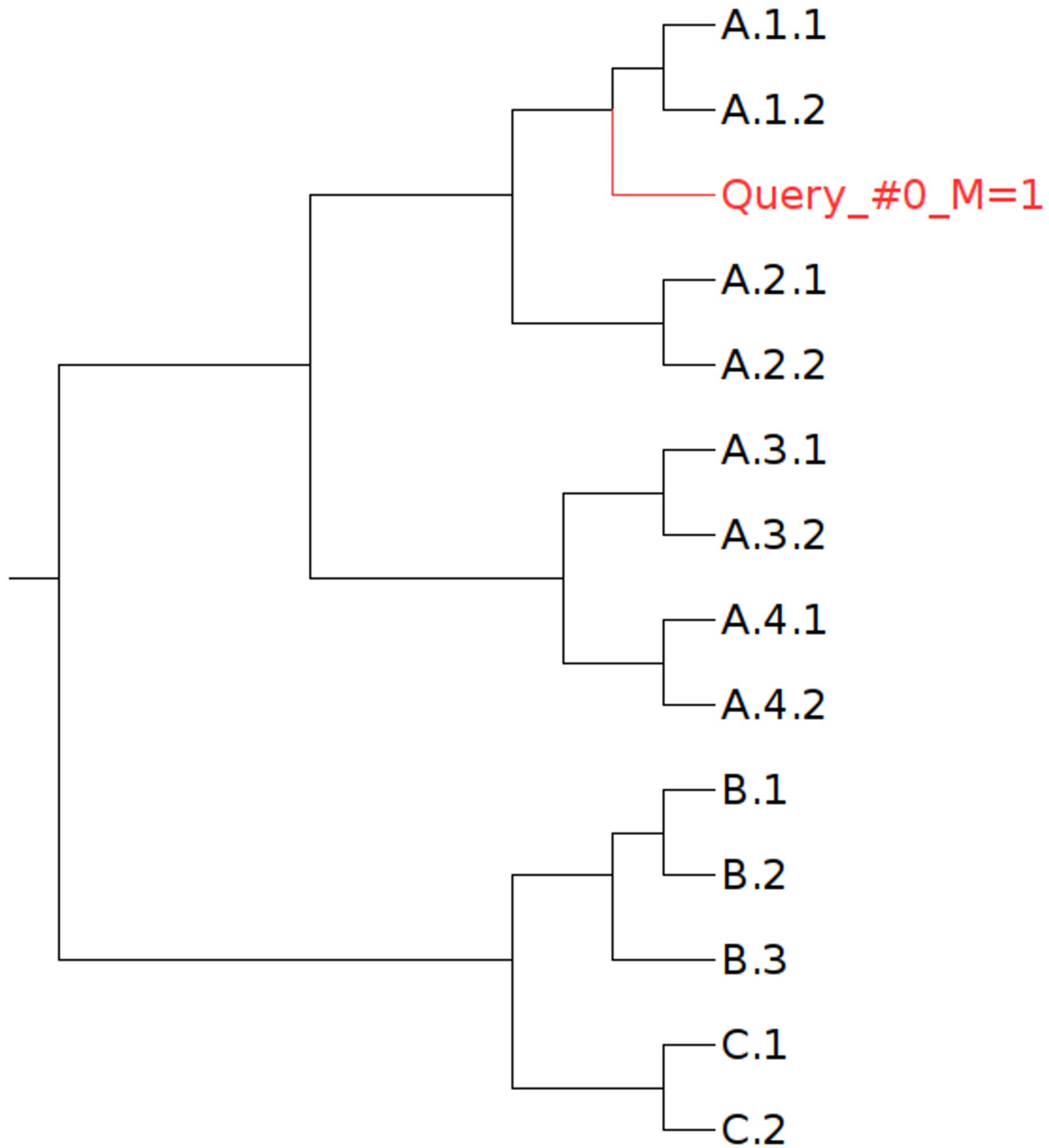
Results

Consensus: A.1

Bracketed by: A.1.1, A.1.2

Probability: 1.0

Potential for novel A.1 sub-species



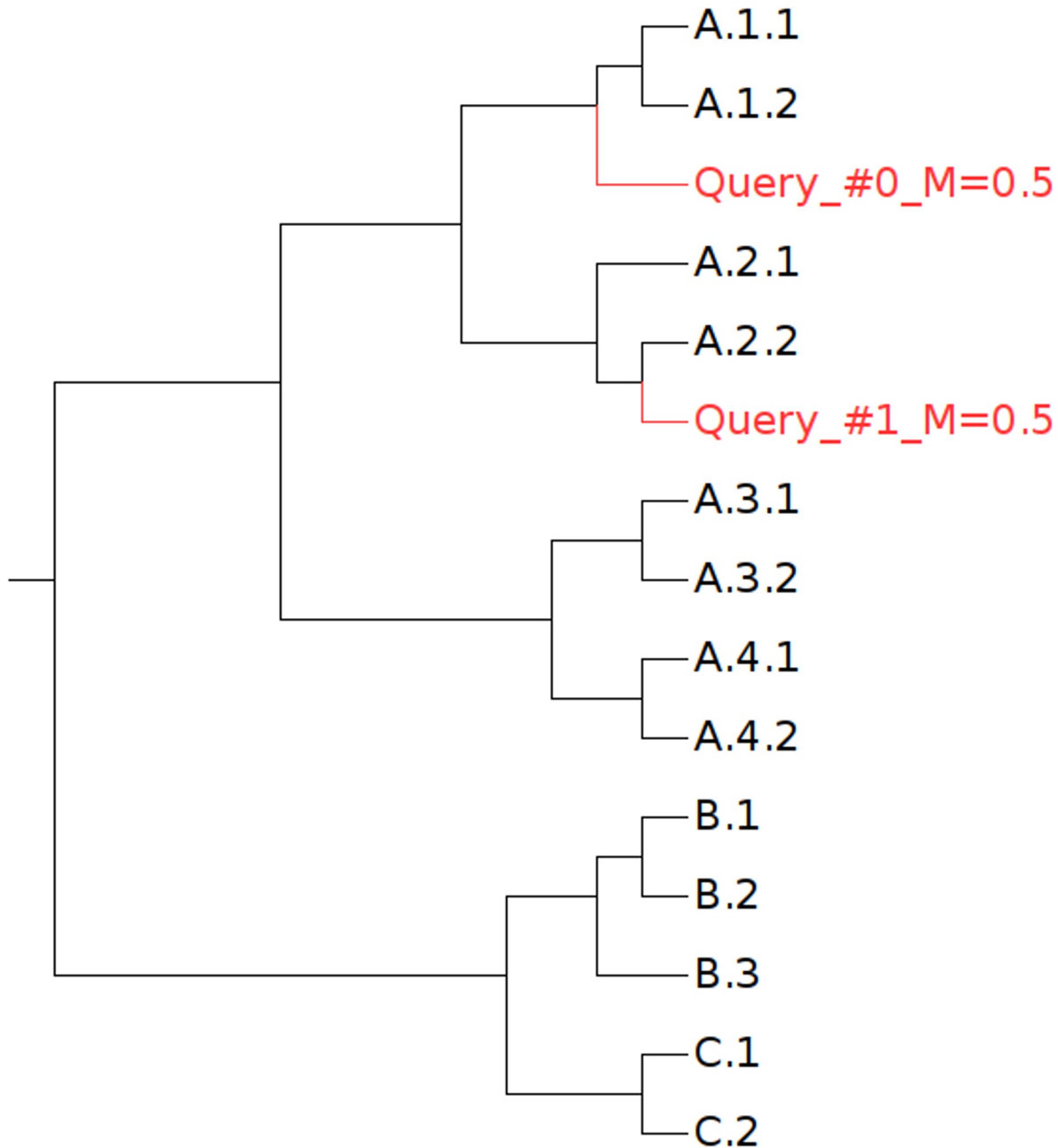
Results

Consensus: A

Bracketed by: A.1._, A.2._

Probability: 1.0

Potential for novel A sub-species



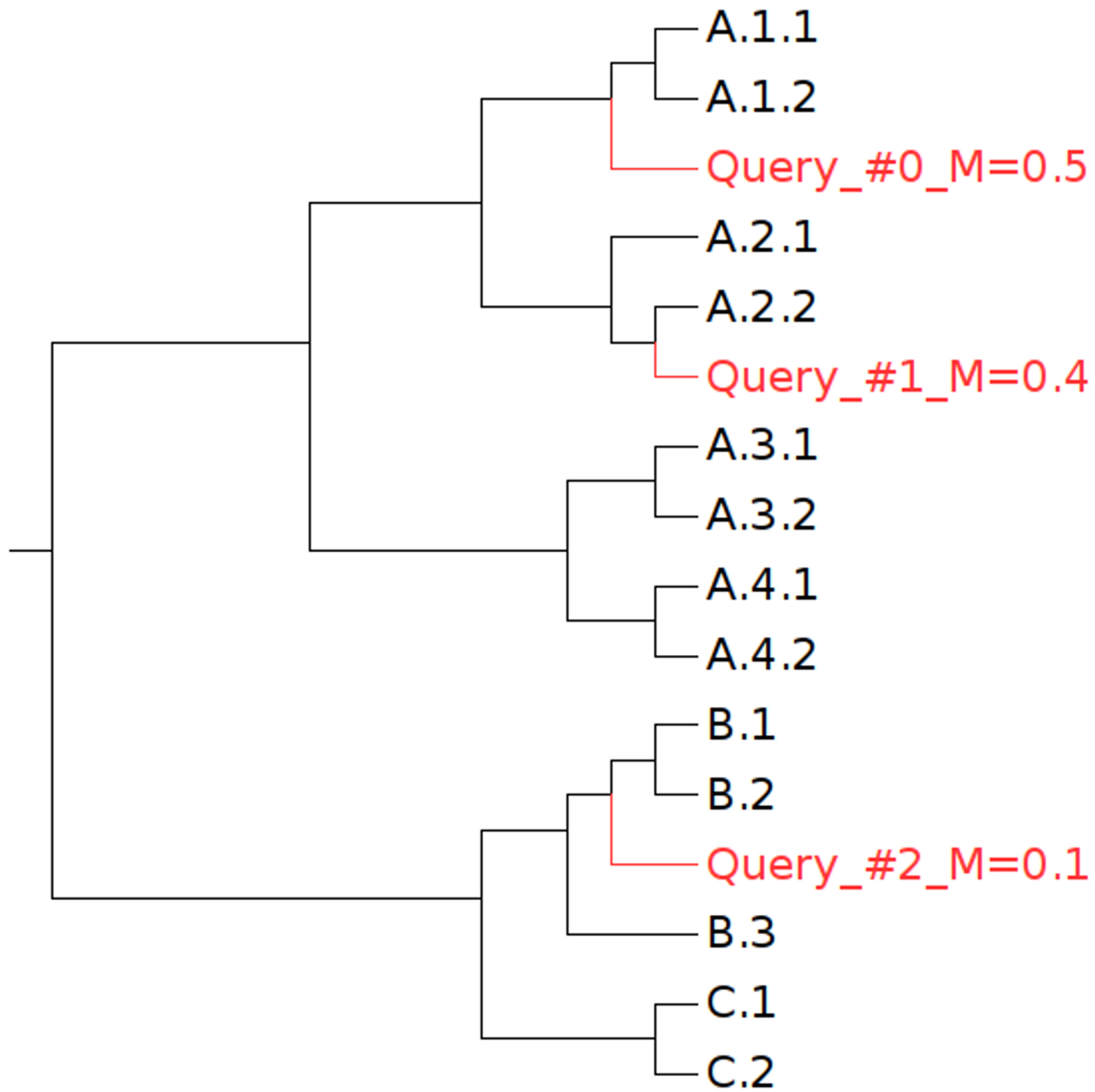
Results

Consensus: A

Bracketed by: n/a

Probability: 1.0

Uncertain placement within A



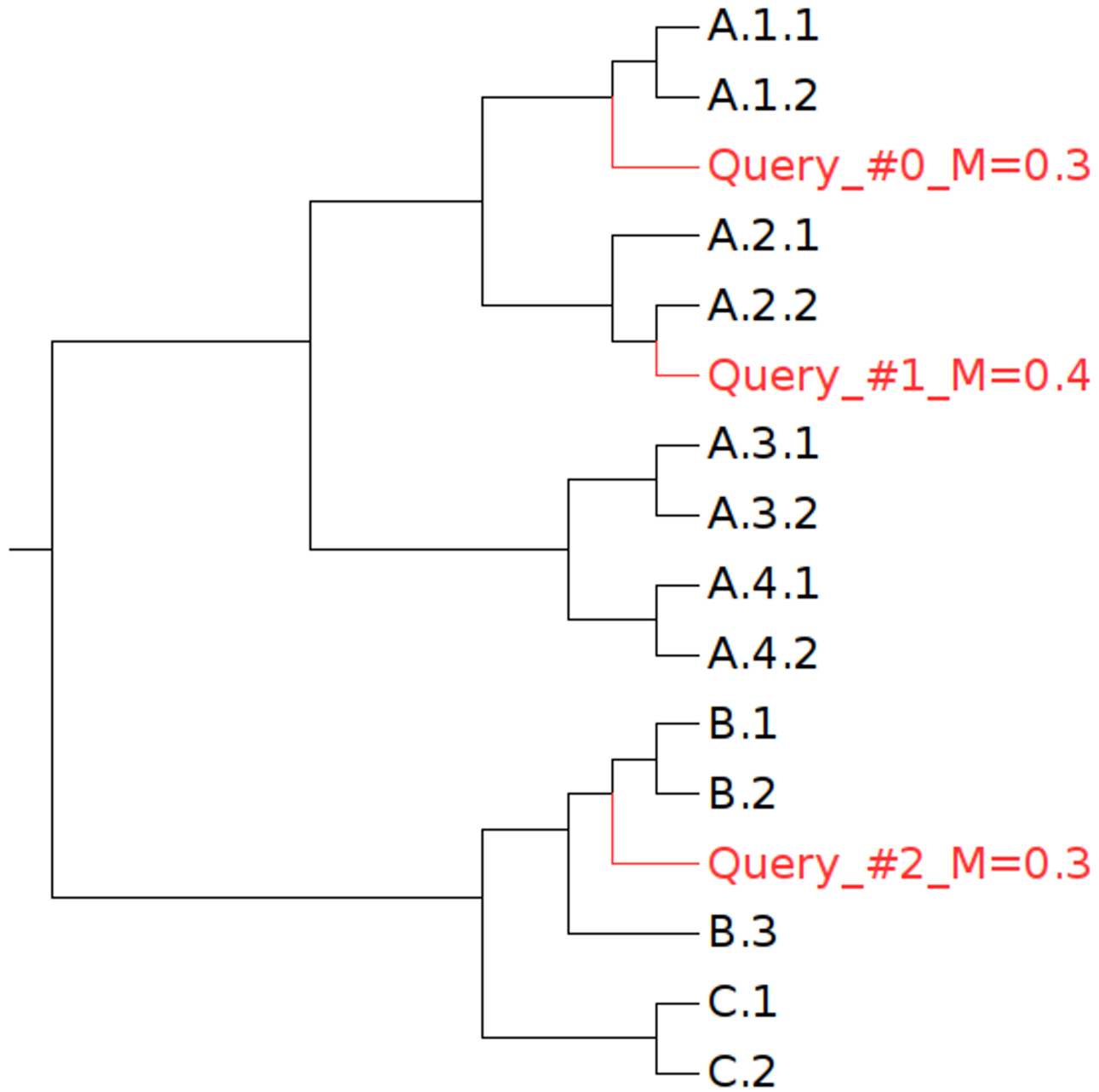
Results

Consensus: A

Bracketed by: n/a

Probability: 0.9

Uncertain placement within A



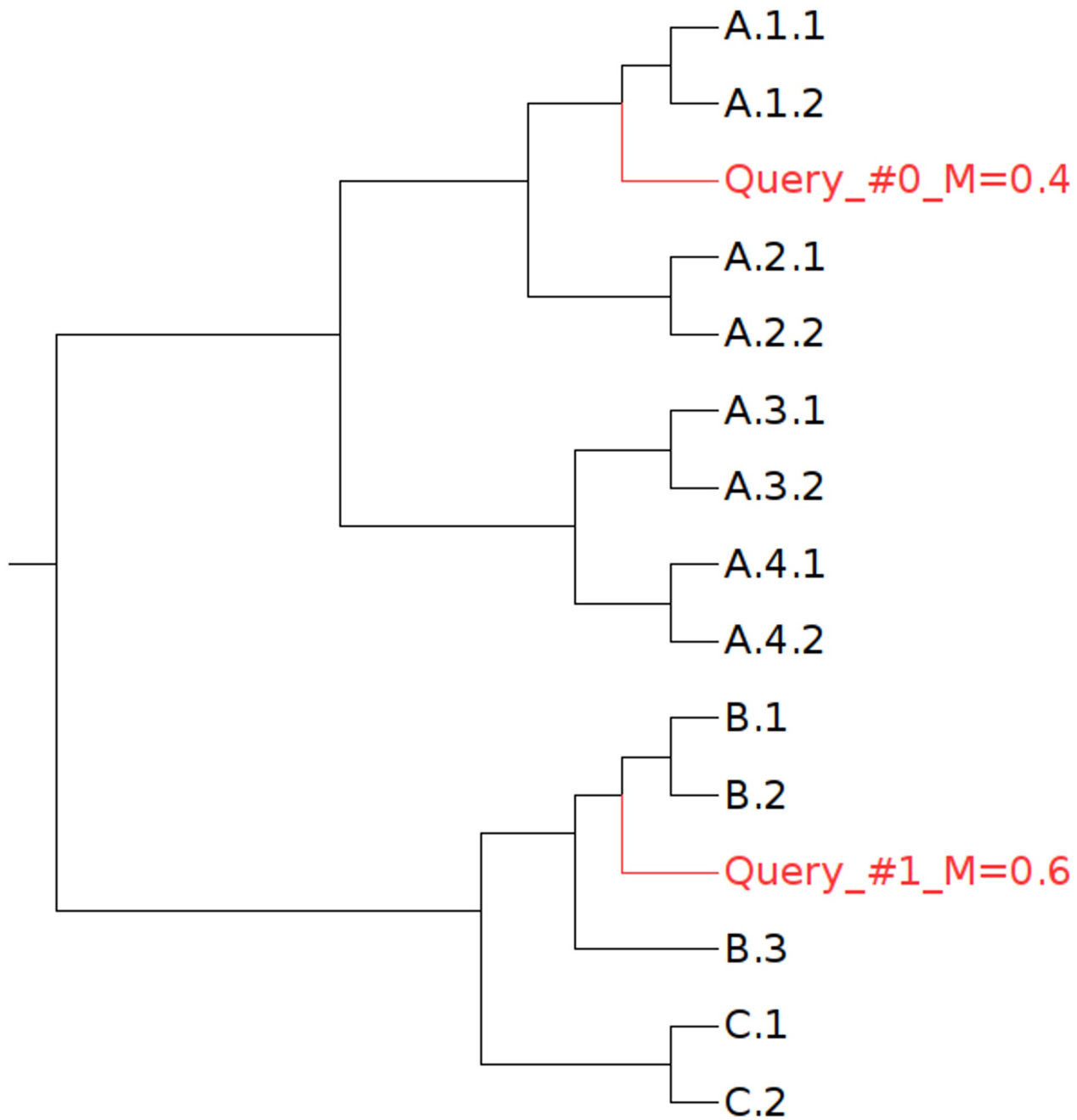
Results

Consensus: A

Bracketed by: n/a

Probability: 0.7

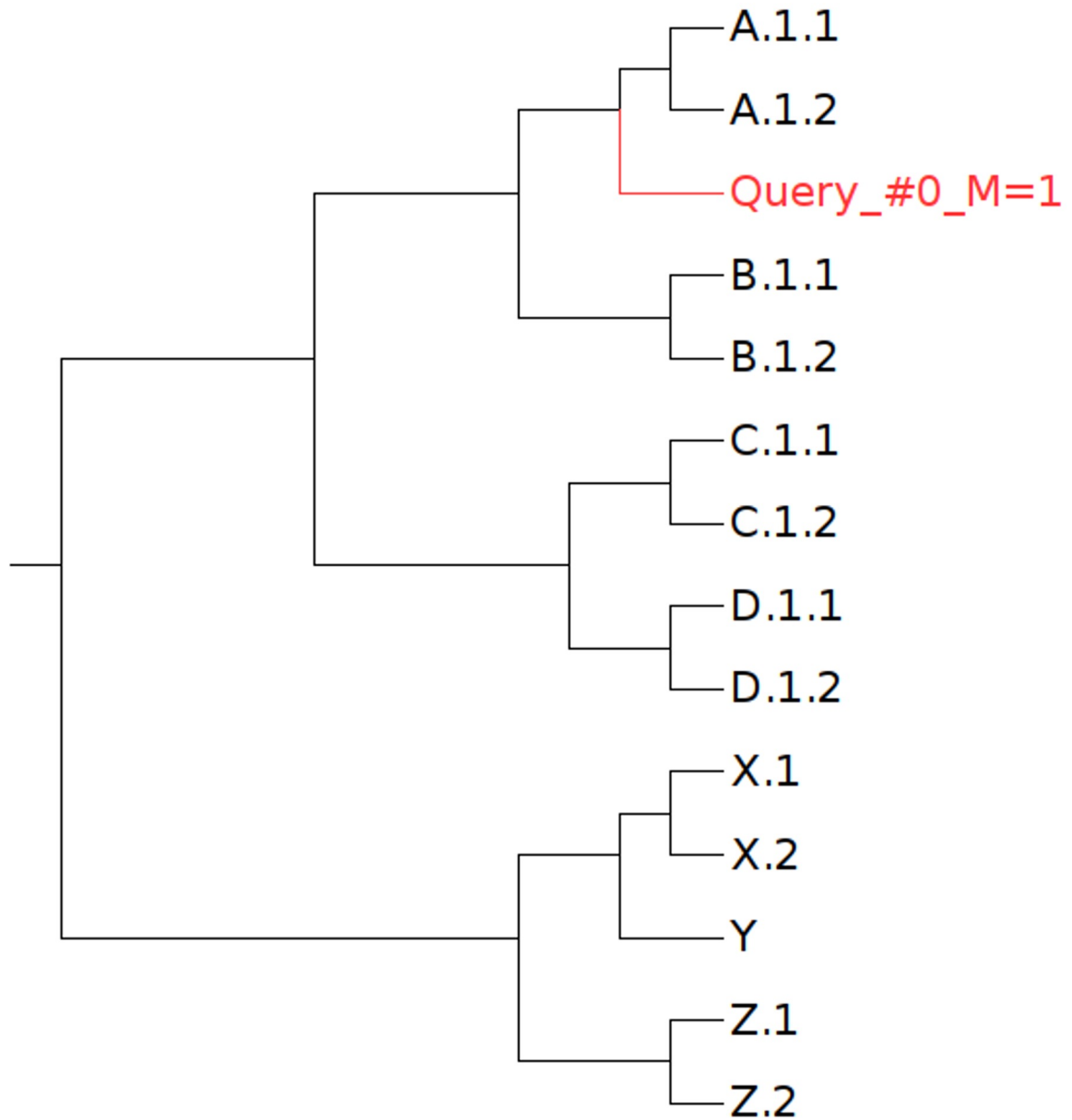
Uncertain placement within A



Results

No consensus

Uncertain placement



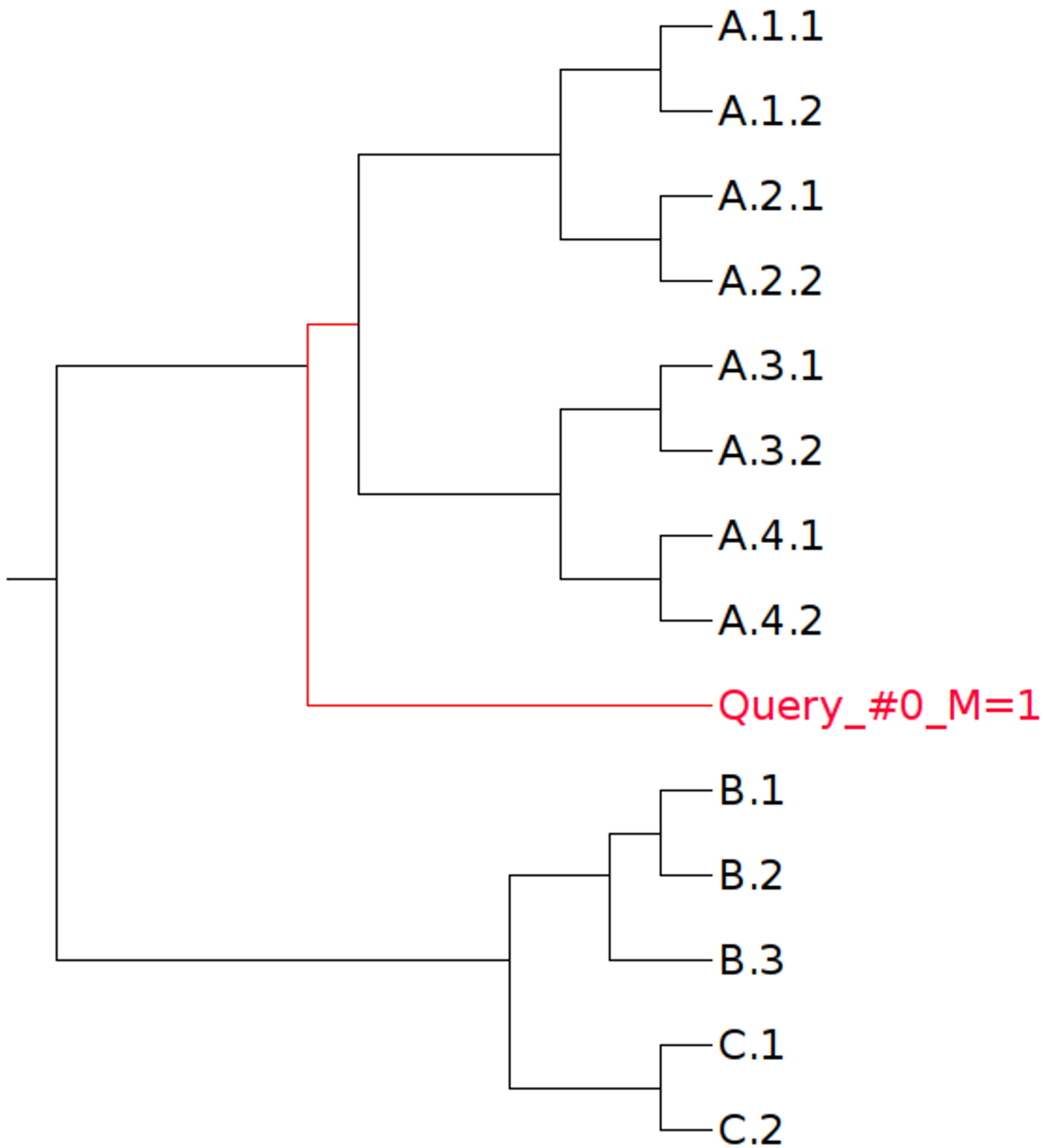
Results

No Consensus

Bracketed by: A.1._, B.1._

Probability: 1.0

Potential for novel sub-species



Results

No Consensus

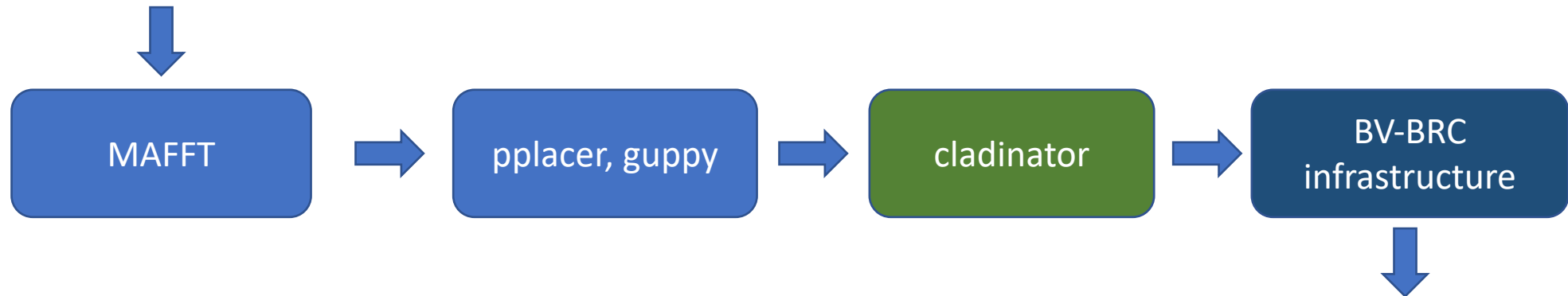
Bracketed by: n/a

Probability: 1.0

Potential for novel sub-species

Usage of cladinator in BV-BRC

- Expert-approved reference tree calculated by specific version of RAxML and resultant statistics output file
- Corresponding MSA
- User query sequence



- User readable output
- Machine readable output
- Interactive tree with placement(s) of query (Archaeopteryx.js)