Heidelberg Institute for Theoretical Studies

HITS

Phylogenetic Placement of Short Reads

Visualization, Species Delimitation and Taxonomic Annotation

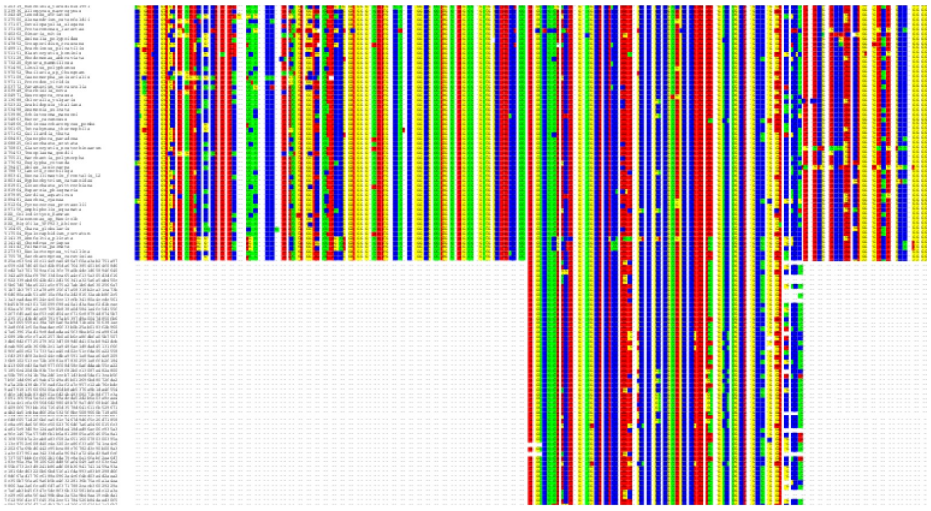
Lucas Czech and Alexey Kozlov, HITS Heidelberg

UniEuk 1st Implementation Meeting, Paris

2016-05-03

Input Data

Reference Tree

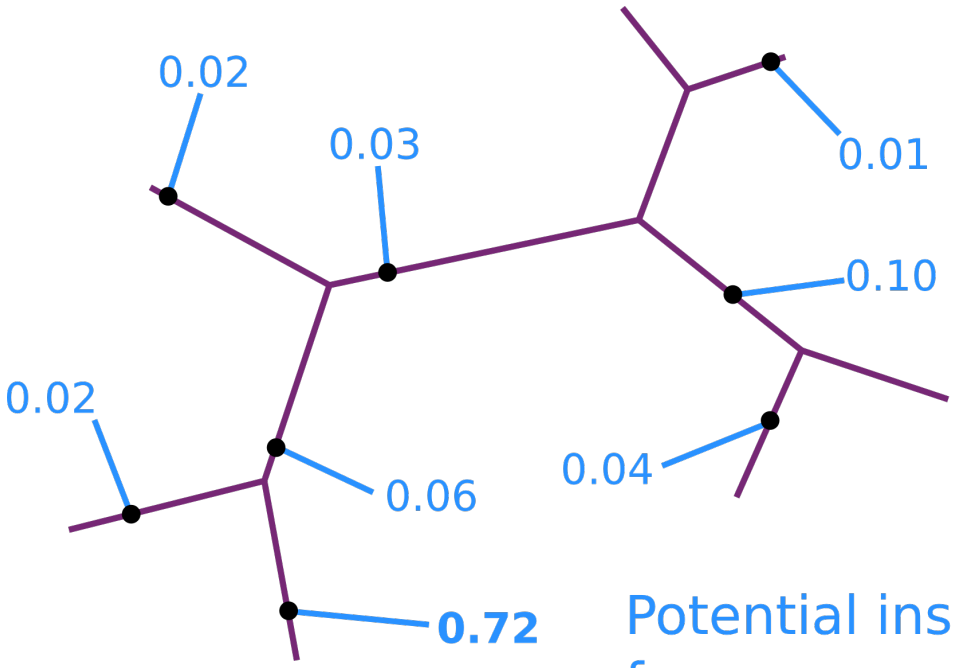


Reference Sequences

Query Sequences

HITS

Evolutionary Placement (EPA)

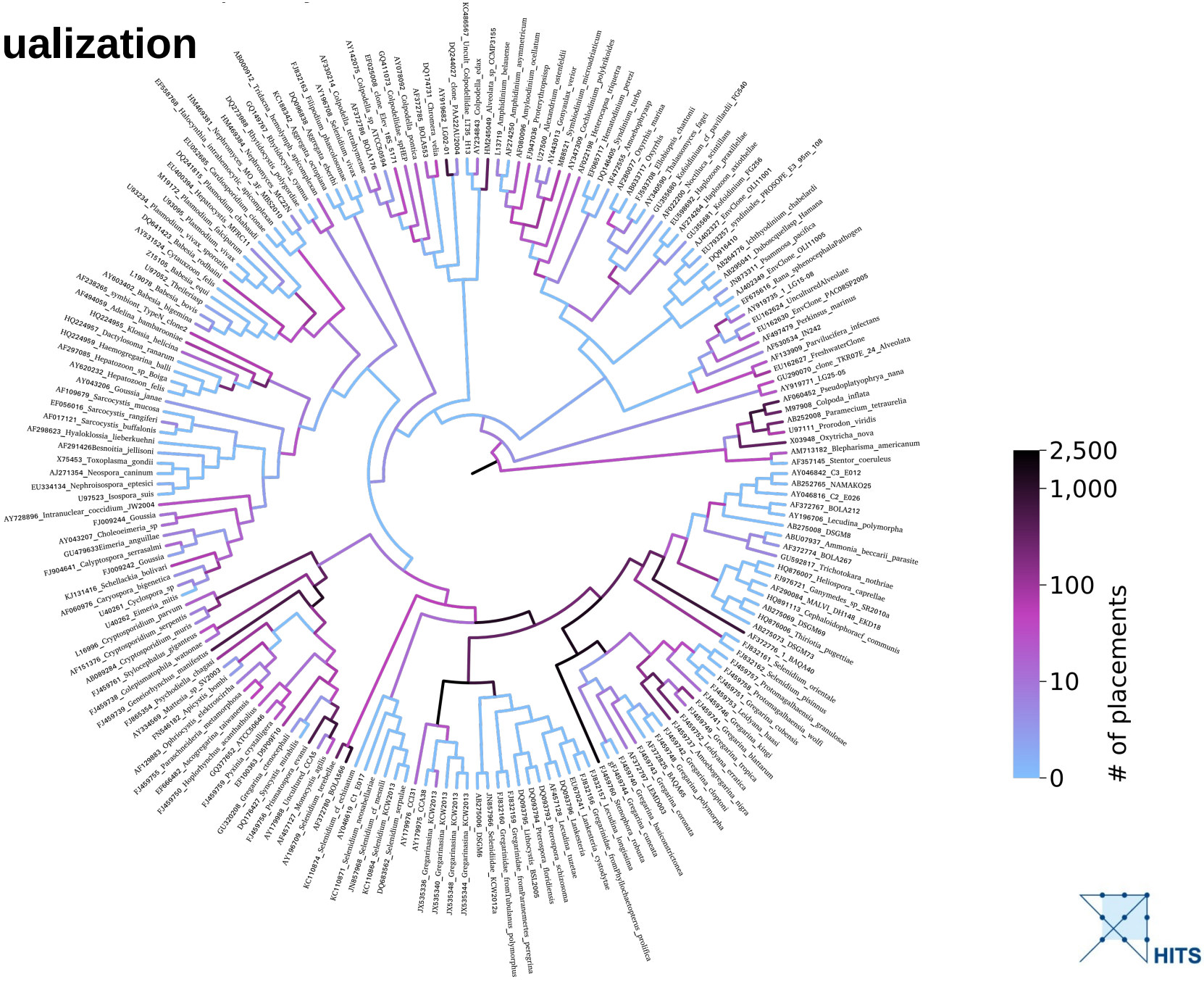


Reference tree

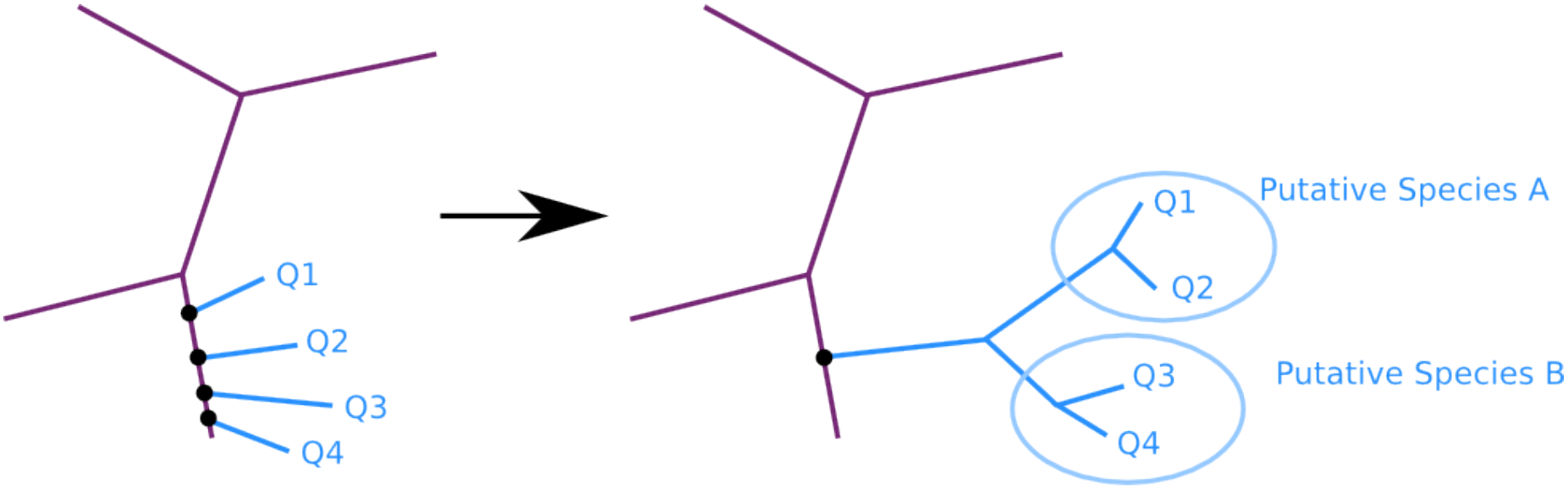
Potential insertion positions ("placements")
for a query sequence with likelihood ratios

HITS

Visualization



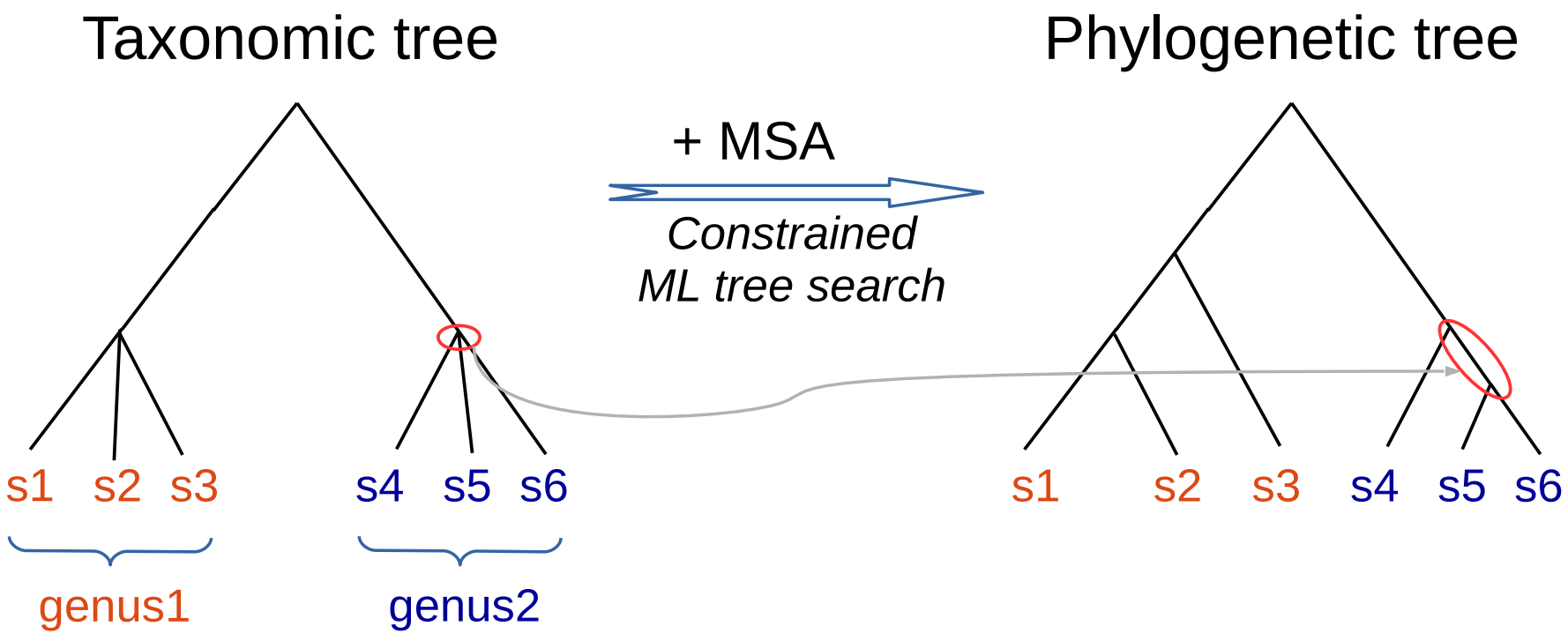
Species Delimitation (PTP)



HITS

Taxonomic Annotation with EPA

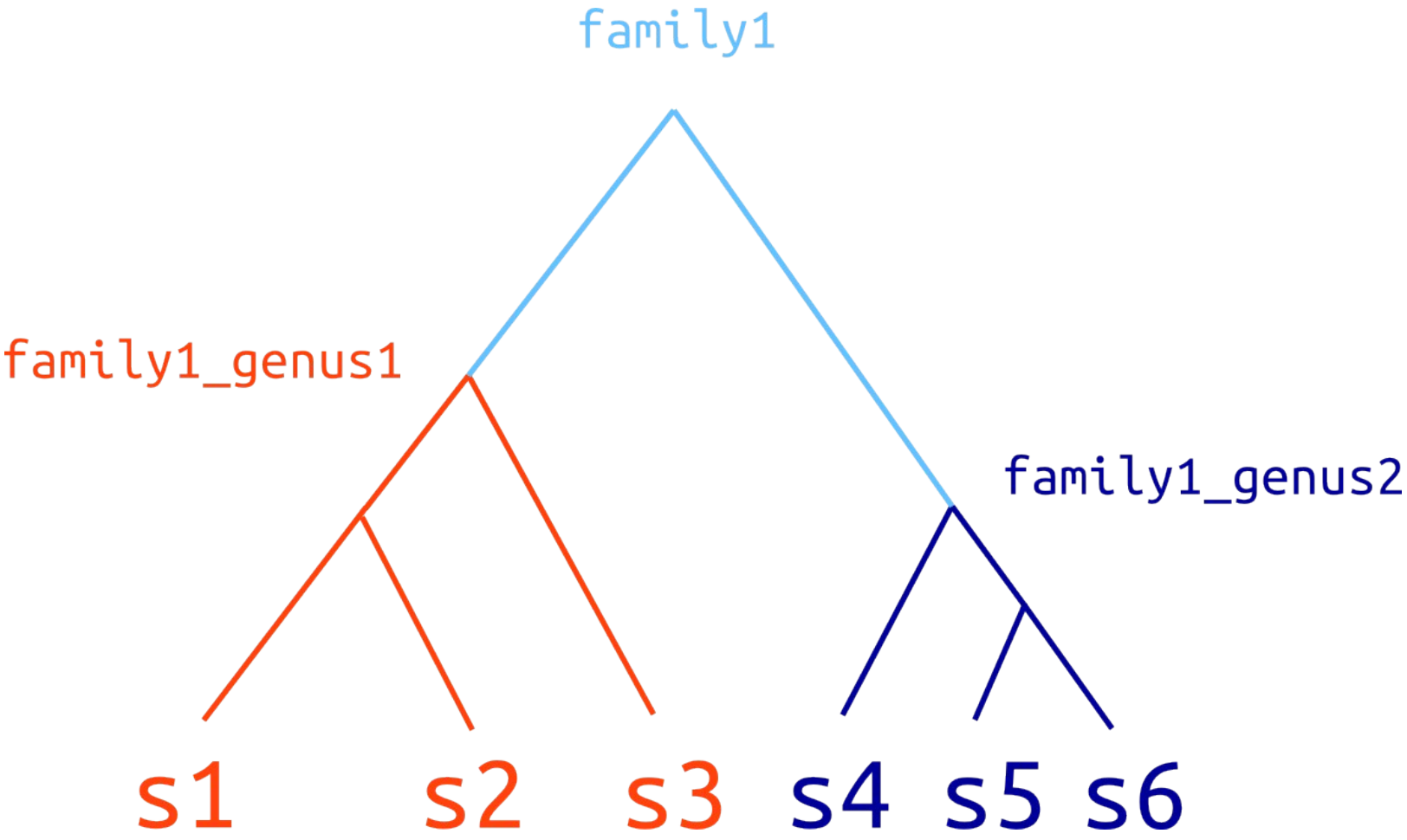
Step 1: Build Reference Tree



HITS

Taxonomic Annotation with EPA

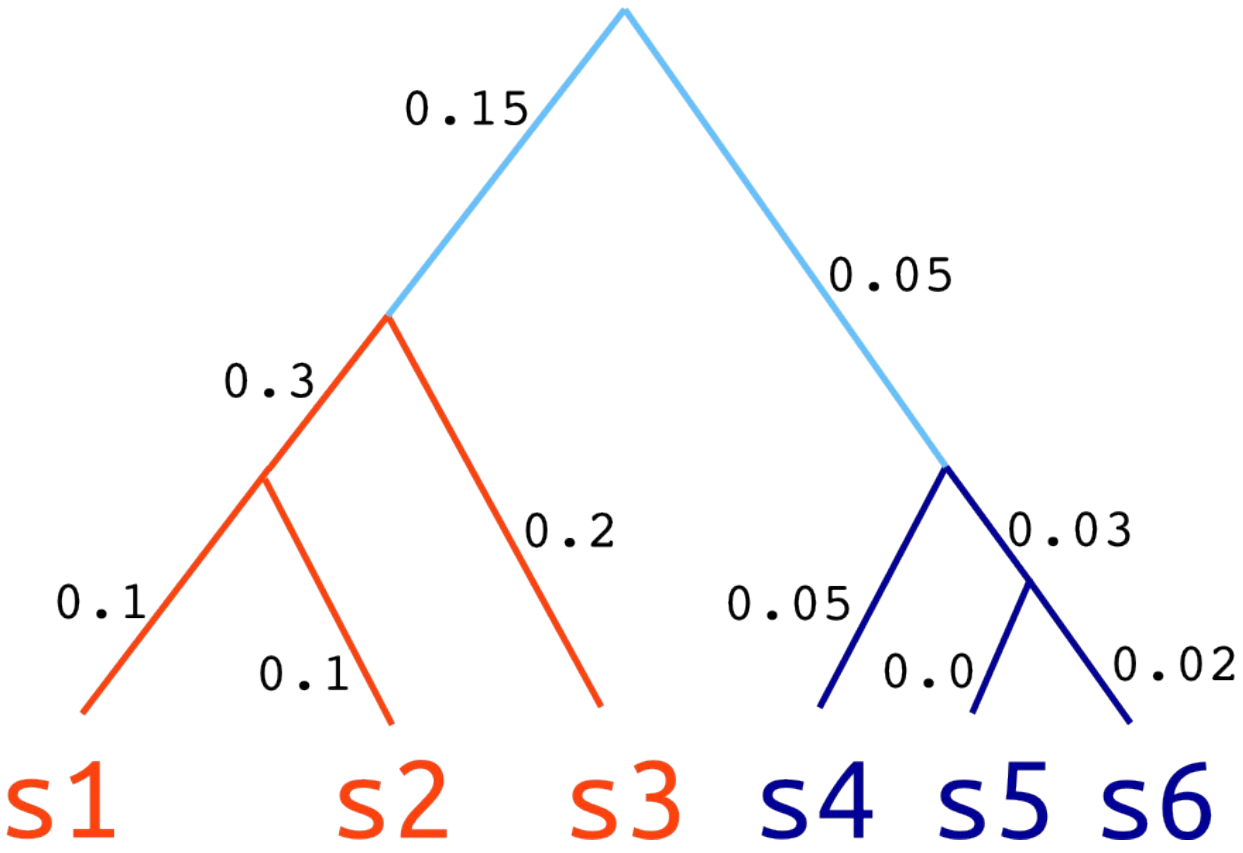
Step 2: Label the Branches



HITS

Taxonomic Annotation with EPA

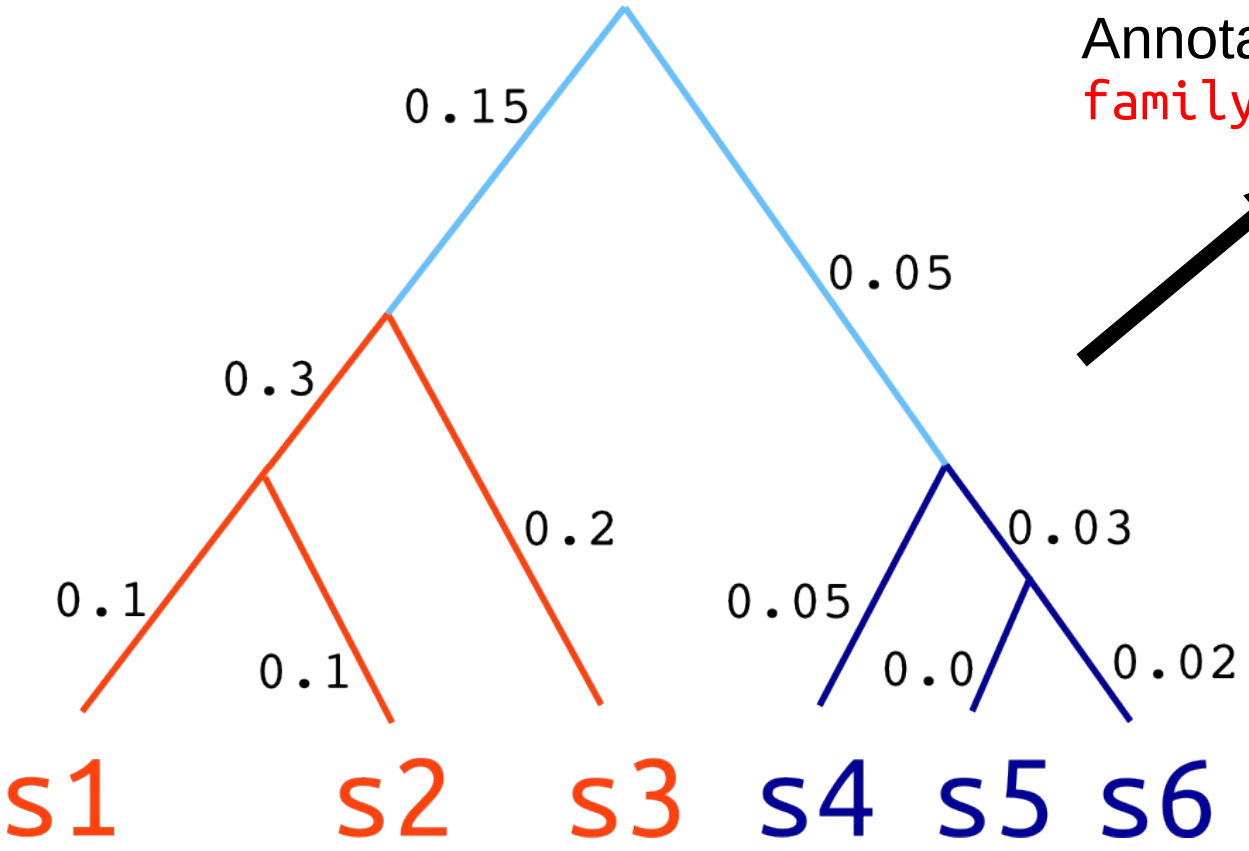
Step 3: Run EPA



HITS

Taxonomic Annotation with EPA

Step 4: Derive Taxonomic Annotation



Annotation for Query:
family1_genus1

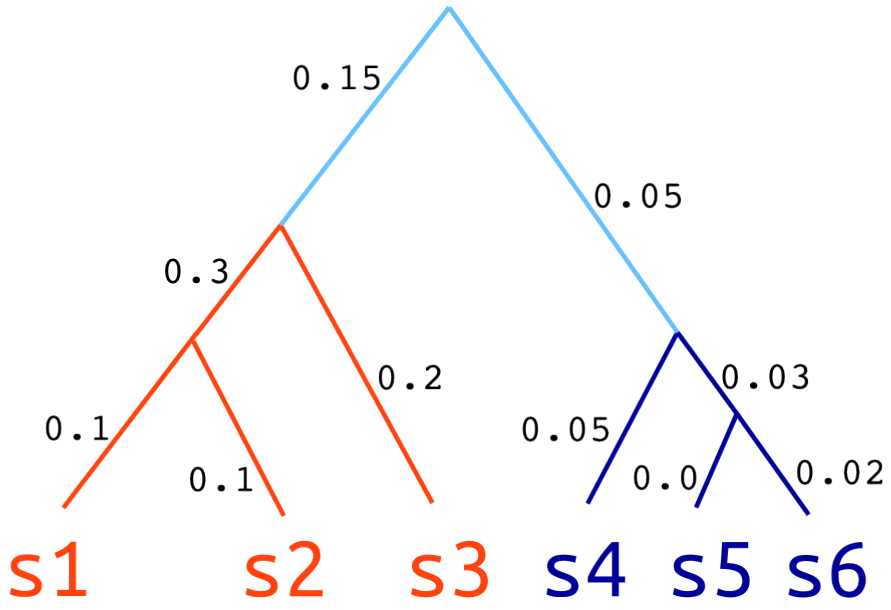
HITS

Software

- PaPaRa: Align Short Reads to Reference MSA
  http://sco.h-its.org/exelixis/software
- RaxML-EPA: Phylogenetic Placement
  https://github.com/stamatak/standard-RAxML
- Genesis: Library for e.g. Clade Annotation and Visualization
  http://genesis-lib.org/
- SATIVA: Detect Taxonomically Mislabeled Sequences
  https://github.com/amkozlov/sativa
- PTP: Species Delimitation
  http://sco.h-its.org/exelixis/software
- PUmPER: Extend Trees with new Sequences from GenBank
  https://github.com/fizquierdo/perpetually-updated-trees

HITS

Taxonomic Annotation with EPA, Revisited

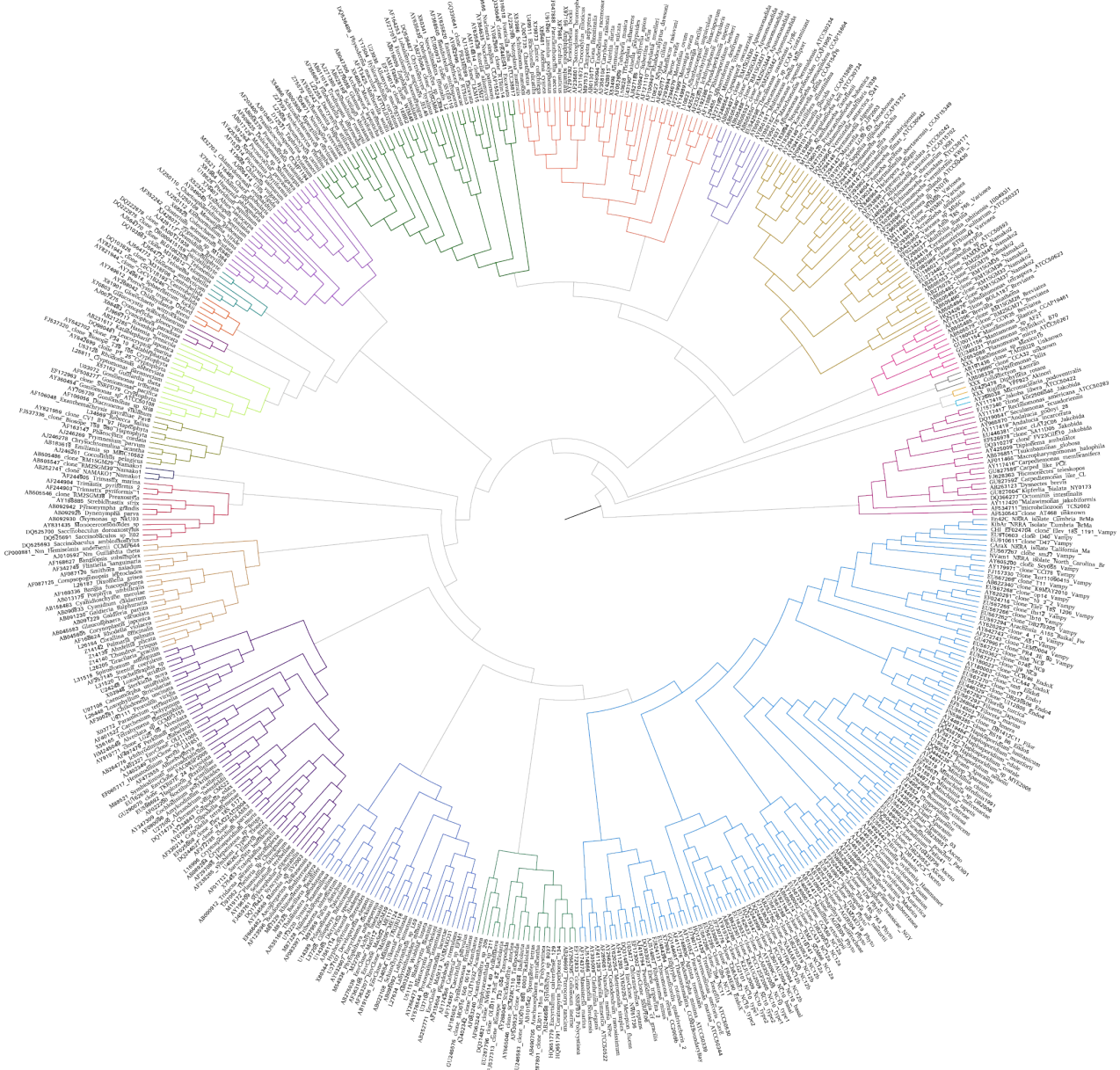


Per-rank LW score evaluation:

```
LW(family1_genus1) = 0.1  + 0.1 + 0.3 + 0.2 = 0.7
LW(family1)        = 0.15 + 0.05            = 0.2
LW(family1_genus2) = 0.05 + 0.03 + 0.02     = 0.1
```

Final assignment: `family1_genus1`

Confidence family / genus: `1.0 / 0.7`

HITS

Clade Annotated Tree