Step 1. Phylogenetic inference.

(FASTA sequence file)   *(phylogenetic tree)*   *(lineage partition)*

FastTree2

expert clade annotation
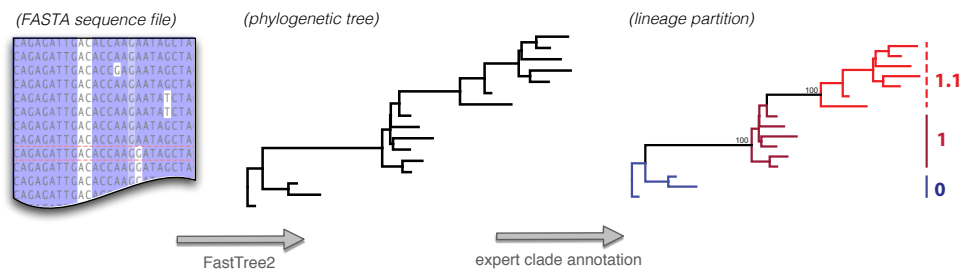
Step 2. Design LABEL module.

(annotation tree)   *(alignment hierarchy)*   *(profile HMM hierarchy)*

H5 — c-1 — 1
          — 1.1
   — 0

pHMM c-1   *pHMM 1*
pHMM 1.1
pHMM 0

pHMM training data

perl, muscle, Jalview

SAM's `modelfromalign`
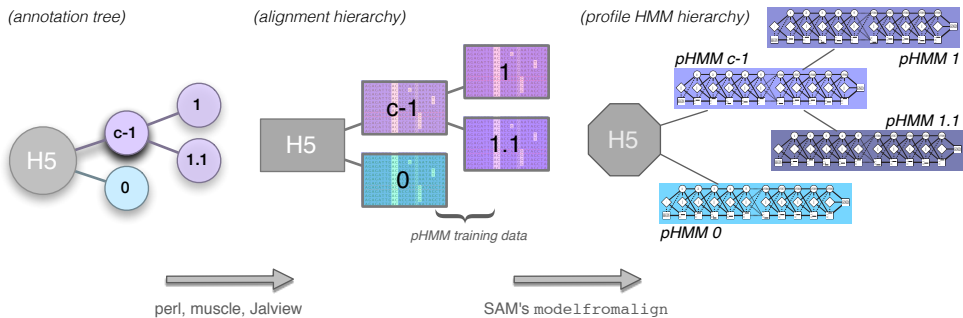
Step 3. Train support vector machine (SVM).

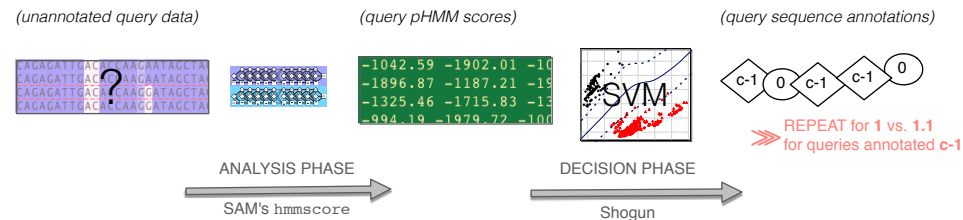*(SVM training data)*   *(scores & annotations)*   *(fully trained SVM)*

profile HMMs

```
-1042.59 -1902.01 -1052.40 -1052.
-1896.87 -1187.21 -1910.93 -1910.
-1325.46 -1715.83 -1331.56 -1331.
 -994.19 -1979.72 -1001.76 -1001.7
 -970.03 -1912.59  -980.01  -980.01
1 0 1 1 0 0 1 0 0 0 0 0 0 0 1 1 0
```

ANALYSIS PHASE

SAM's `hmmscore`

DECISION PHASE

Shogun

REPEAT for **1** vs. **1.1** according to Step 2 tree
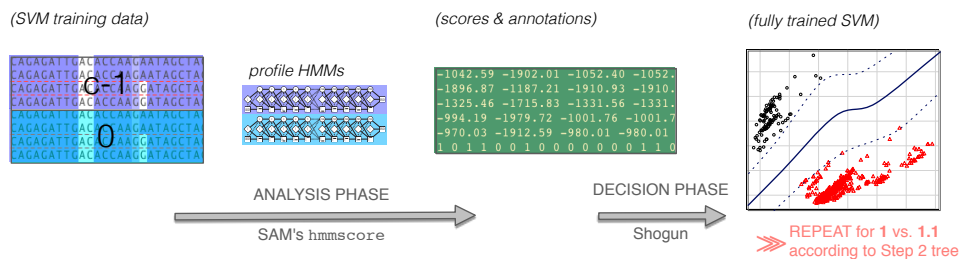
Step 4. Annotate query sequences.

(unannotated query data)   *(query pHMM scores)*   *(query sequence annotations)*

```
-1042.59 -1902.01 -10
-1896.87 -1187.21 -19
-1325.46 -1715.83 -13
 -994.19 -1979.72 -100
```

SVM

c-1 — 0 — c-1 — c-1 — 0

REPEAT for **1** vs. **1.1** for queries annotated **c-1**

ANALYSIS PHASE

SAM's `hmmscore`

DECISION PHASE

Shogun

Supp. Fig. 1: The lineage partition in step 1 is used to design the annotation tree in step 2, giving rise to a hierarchy of profile HMMs. Step 3 is repeated at each internal node of the annotation tree, where profile scores of training data are used to train SVMs. For annotating query sequences, step 4 uses the pHMMs and SVMs from steps 2 & 3 to assign query sequence annotations in a root-to-tips fashion corresponding with the annotation tree.