

```

1 //create protein distance graph
2 graph G = (v,e)
3
4 For each gene  $i \in \{\text{all genes}\}$ 
5     Vertex  $v_1 := (i)$ 
6     G.ADD_VERTEX( $v_1$ ) if  $v_1 \notin \{\text{VERTICES}(G)\}$ 
7     For each gene  $j \in \{\text{all genes}\}$ 
8         Vertex  $v_2 := (j)$ 
9         G.ADD_VERTEX if  $v_2 \notin \{\text{VERTICES}(G)\}$ 
10        Alignment := BLASTP.BestHit( $i, j$ )
11        Next if |Alignment| == 0
12        Edge  $e := (v_1, v_2)$ 
13         $e.score = \text{Alignment.score}$ 
14         $e.alignmentLength = \text{ClustalW gap-free alignment length}$ 
15         $e.distance = \text{protdist distance}$ 
16        G.ADD_EDGE( $e$ )
17
18 tree T := known evolutionary tree of taxa
19 For each node  $\in \{T\}$ 
20     Next if |node.children|  $\neq 2$ 
21     Outgroup := {taxa  $\notin$  node.childrenTaxa}
22
23     //create seeds
24     Seeds := {}
25     For each gene  $i \in \{\text{node}[0]\}$ 
26         Seeds[i] := MAX { G[i][x].score } where  $x \in \{\text{node}[1].\text{genes}\}$ 
27     For each gene  $i \in \{\text{node}[1]\}$ 
28         Seeds[i] := MAX { G[i][x].score } where  $x \in \{\text{node}[0].\text{genes}\}$ 
29
30     SORT Seeds descending
31
32     //create clusters
33     geneUsed := {}
34     For each seed  $\in \{\text{Seeds}\}$ 
35         minDistance := seed.distance
36         minAlignmentLength := seed.alignmentLength
37
38         //depth first search
39         Cluster := {genes  $\in \{\text{seed}\}$ }
40         Seen := {}
41         For each gene  $i \in \{\text{Cluster}\}$ 
42             Next If Seen[i] == TRUE
43             For each gene  $j \in \{G[i]\}$ 
44                 If G[i][j].distance < maxDistance
45                     AND G[i][j].alignmentLenth > minAlignmentLength
46                     AND Seen[j] == FALSE
47                     ADD j to Cluster
48                     Seen[j] := TRUE
49
50         For each gene  $i \in \{\text{Cluster}\}$ 
51             If geneUsed.i == TRUE OR i.taxa  $\in \{\text{Outgroup}\}$ 
52                 deleteCluster := TRUE
53             geneUsed.i := TRUE
54
55         If deleteCluster == TRUE
56             Delete Cluster
57         Else
58             Return Cluster

```