**Advanced Size Reduction algorithm in Pseudocode**

**INPUT:**

NKFile: a Newick file containing the tree that needs to be reduced

Abundance Threshold (default value is 1%)

Number of nodes wanted in the final Tree (default value is 30)

**START**

**Read** NKFile and **store** the result as a tree T

**Create** a List named Kept\_Sequences

#List of clonotypes we want to keep in the reduced tree

**Add** all the clonotypes of T to Kept\_Sequences

#We are keeping all the clonotypes at first, and then taking some off one at the time

**WHILE** length of Kept\_Sequences > Number of nodes wanted in the final tree :

**FOR** each leaf L of T :

**IF** L is not an observed sequence : #would mean that an internal node inferred by clonalTree would now be a leaf, which is not relevant

**Delete** L from Kept\_Sequences

**ENDIF**

**ENDFOR**

**IF** no leaf has been deleted yet :

**IF** all leaves are of abundance > abundance Threshold or among the ten most abundant clonotypes : # would mean that all the leaves are relevant, we can’t prune the tree anymore

**Break** from the While loop

**ENDIF**

**GET** the leaves L1 and L2 that are separated by the minimum hamming distance

**IF** L1’s abundance>L2’s abundance :

**IF** L2’s abundance<abundance threshold and L2 is not among the 10 most abundant clonotypes :

**Delete** L2 from Kept\_Sequences

**ENDIF**

**ELSE :**

**IF** L1’s abundance<abundance Threshold and L1 is not among the 10 most abundant clonotypes :

**Delete** L1 from Kept\_Sequences

**ENDIF**

**ENDIF**

**Prune** Tto keep only Kept\_Sequences

**OUTPUT :**

**The pruned tree T**