**R Scripts, input and output data summary**

**C# Code Output files:**

1. Distance matrix MEGA: *Distance\_Matrix\_MEGA.meg* – Standard processing (C#)
2. Distance matrix R / APE: *Distance\_Matrix\_R\_APE.csv* – Standard processing (C#)
3. Excel Summary light: *PairwiseComparisonsLight.csv* – Standard processing (C#)
4. Excel Summary details: *PairwiseComparisons.csv* – Line change in processing (C#)

**R scripts & outputs:**

**Word performance (Table 1 in paper):**

Input: PairwiseComparisons.csv

Script: WordPerformance.R

Output: WordPerformance.csv

**Distribution IE vs. Random (Table 7 in paper):**

Input: Distances\_within\_IE.csv and Distances\_within\_Dummies.csv

Script: HistogramOverlapIE\_Dummies.R

Output: Histogram in R Studio console

**Resampling: generate random languages**

Input: AllWords.csv

Script: GenerateRandomLanguages.R

Output: newRandomLanguages.csv

**Dating nodes (1) – table with distance between families/subfamilies:**

Input: PairwiseComparisons.csv

Script: DatingNodes.R

Output: MonitorClades.csv

**Dating nodes (2) – Train calibrated data and predict:**

Input: Calibrating data in vector

Distances.csv for query

Script: Dating\_Regression.R

Output: Dating.csv

**Long Range Analysis:**

Input: PairwiseComparisons.csv

Script: MacroFamilies.R

Output: Output.csv

**Phylogeny in R:**

Input: Distance\_Matrix\_R\_APE.csv (output from C# code)

Script: LoadMatricesInR.R

Output: SVG Files from R Studio Console