

Network analysis of pseudogene-gene relationships: from pseudogene evolution to their functional potentials

Authors: Travis S Johnson, Sihong Li, Jonathan R Kho, Kun Huang, Yan Zhang

Supplementary Materials

The Supplementary Materials contain two zip files that are downloadable from https://github.com/yanzhanglab/PGG_DB.

Supplementary File 1: alignment_matrix.csv.zip

This file is an alignment matrix between the pseudogenes and gene families. Each row (except the last) is a gene family consensus sequence and each column is a pseudogene. The individual cells are the alignment score from ClustalW between the pseudogene sequence (column) and the consensus sequence (row). The last row is not a consensus sequence but the index (row number) of the highest scoring alignment in that column. This row number is the family to which the pseudogene was assigned. These new PGG families that are inclusive of pseudogenes and genes were then aligned within themselves (see below).

Supplementary File 2: pgAmats.zip

This file contains the complete adjacency matrix of each PGG family. Each family has its own csv file containing the sequence alignment score (of every pair of members in the family) from the CUDA-align algorithm where each row is a pseudogene/gene and each column is a pseudogene/gene. These were used to generate the MSTs used in our manuscript. There is also much more information contained within these files that was not fully reported in the manuscript due to page limit.