Steps for obtaining the candidate homologs in *Hypsibius exemplaris*

Step 1: Obtaining the first set of candidate genes using HMMER

a. building HMMs

hmmbuild <hmm output file> <input file>
→ used the *Drosophila melanogaster* protein isoforms.

b. HMMER search

esl-sfetch --index <database>
→ indexing the *H. exemplaris* predicted protein sequences as a database

hmmsearch -E 10e-6 --tblout <output file#1> <query> <database>
→ the output file will be in a table format (.tblout) with names of the genes from the database

grep -v "^#" <output file#1> | awk '{print $1}' | esl-sfetch -f <database> -> <output file#2>
→ using the names of the genes from output file#1, the sequences will be obtained ("fetced") from the *H. exemplaris* predicted protein sequences database

Step 2: Obtaining the second set of candidate genes using BLASTp

a. building database for BLASTp search

makeblastdb -dbtype prot -in <input file> -out <output file#2>
→ making the 1st set of candidate gene as a database for the BLASTp search

esl-sfetch –index <output file #2>
→ indexing the 1st set of candidate gene so it can be used for the esl-sfetch command later

b. BLASTp search

blastp -db <output file#2> -query <input file> -evalue 10e-6 -outfmt 6 | awk '{print $2}' | uniq | esl-sfetch -f <output file#2> -> <output file #3>
→ only hits with an e-value less than 10^{-6} were retained. The query (input file) used here is the same input file used for building HMMs in Step 1a (i.e., the *D. melanogaster* protein isoform).

Step 3: Obtaining the third set of candidate genes using BLASTp (reciprocal BLAST)

This step used the online BLASTp search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) wherein the 2nd set of candidate genes is used as query against the *D. melanogaster* protein database (taxid: 7227). To be retained, hits must have a max score greater than 80, a percent identity greater than 20%, and e-value less than 10^{-6}.