lab notebook assignment 1
directory
/projects/bgmp/ewi/bioinfo/Bi623/Assignments/Reciprocal-Best-Hits

Goal of reciprocal best hit:
- species 1 gene x to species 2 gene y has the lowest e-value out of the group species 1 gene x to any gene in species 2
- species 2 gene y to species 1 gene x has the lowest e-value out of the group species 2 gene y to any gene in species 1
- there are no e-value repeats in either combination
    - if species 1 gene x to any gene in species 2 (or the other way around) contains multiple of the lowest e-value, then discard the pair

Plan
initialize dictionary
- iterate through
    - if the gene of species 1 doesnt exist yet, add the species 2 gene and e-value pair to dictionary
    - if the gene does exist:
        - if the evalue is lower than the e-value there already, replace the pair with the current
        - if it is equal, add the pair to the list
        - if it is greater, dont do anything

bash command to sort files:
Human Zebrafish:
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Dre_query_Hsa_db.txt > sorted_zfish_against_human.txt
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Hsa_query_Dre_db.txt > sorted_human_against_zfish.txt

Human Eel:
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Hsa_query_Eel_db.txt > sorted_human_against_eel.txt
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Eel_query_Hsa_db.txt > sorted_eel_against_human.txt

Human Babywhale:
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Hsa_query_Pka_db.txt > sorted_human_against_babywhale.txt

sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Pka_query_Hsa_db.txt > sorted_babywhale_against_human.txt

Zebrafish Eel:
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Dre_query_Eel_db.txt > sorted_zfish_against_eel.txt
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Eel_query_Dre_db.txt > sorted_eel_against_zfish.txt

Zebrafish Babywhale:
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Dre_query_Pka_db.txt > sorted_zfish_against_babywhale.txt
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Pka_query_Dre_db.txt > sorted_babywhale_against_zfish.txt

Eel Babywhale:
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Eel_query_Pka_db.txt > sorted_eel_against_babywhale.txt
sort -k1,1 -k11,11g /projects/bgmp/shared/Bi623/PS1/blasthits/
Pka_query_Eel_db.txt > sorted_babywhale_against_eel.txt

Results
Human_Zebrafish_RBH.tsv
Number of Reciprocal Best Hits:7895
Human_Electric_Eel_RBH.tsv
Number of Reciprocal Best Hits:8949
Human_Eletric_Baby_Whale_RBH.tsv
Number of Reciprocal Best Hits:8944
Zebrafish_Electric_Eel_RBH.tsv
Number of Reciprocal Best Hits:10185
Zebrafish_Baby_Eletric_Whale_RBH.tsv
Number of Reciprocal Best Hits:9341
Electric_Eel_Electric_Baby_Whale_RBH.tsv
Number of Reciprocal Best Hits:10662

Runtime:
User time (seconds): 1.63
        System time (seconds): 0.14
        Percent of CPU this job got: 99%
        Elapsed (wall clock) time (h:mm:ss or m:ss): 0:01.79
        Average shared text size (kbytes): 0
        Average unshared data size (kbytes): 0
        Average stack size (kbytes): 0

Average total size (kbytes): 0
Maximum resident set size (kbytes): 82812
Average resident set size (kbytes): 0
Major (requiring I/O) page faults: 0
Minor (reclaiming a frame) page faults: 19454
Voluntary context switches: 59
Involuntary context switches: 10
Swaps: 0
File system inputs: 0
File system outputs: 0
Socket messages sent: 0
Socket messages received: 0
Signals delivered: 0
Page size (bytes): 4096
Exit status: 0