Source code of Original scripts “Alignment BLAST”, (BLAST database creation, Fasta slasher, OMA parameters)

Alignment BLAST

#!/bin/sh

#DB='DB\_list.txt'

g=`find /media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/Lost\_genes -name "\*fasta"`

mkdir /media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/Completed/

folder=/media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/Aligned/

mkdir -p $folder

for gen in $g

do

echo "Processing $gen"

gen\_c=`echo ${gen} | sed 's/.\*Lost\_genes\///' | sed -e 's/.fasta//'`

#folder=/media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/${gen\_c}

#mkdir -p $folder

#for genome in $f

#do

DB='/media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Cancer\_genes/Oncogenes/Genomes/BLAST\_DB/db.txt'

while read GENOME; do

echo "Aligning $gen\_c $GENOME"

#genome\_name=`echo $GENOME | sed -e 's/.db//' | sed "s/.\*Genomes\_hmmer\_np\///"`

#out=`echo $genome`

#out=${folder}/${GENOME}\_${gen\_c}\_tab.out

outn=${folder}/${GENOME}\_${gen\_c}.out

blastn -db $GENOME -query $gen -out $outn

done < $DB

echo "$gene\_c finished"

mv ${gen} /media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/Completed/

done

#DB='/media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/236\_shuf.txt'

#while read GENOME; do

# echo "Это строка: $GENOME"

#done < $DB

#echo "Hi!"

BLAST database creation

#dustmasker -in Ciona.faa -infmt fasta -parse\_seqids -outfmt maskinfo\_asn1\_bin -out Ciona\_dust.asnb

segmasker -in Carp.faa -infmt fasta -parse\_seqids -outfmt maskinfo\_asn1\_bin -out Carp.asnb

#makeblastdb -in Ciona.faa -input\_type fasta -dbtype prot -parse\_seqids -mask\_data Ciona\_dust.asnb -out Carp

makeblastdb -in Carp.faa -input\_type fasta -dbtype prot -parse\_seqids -mask\_data Carp.asnb -out Carp -title "Carp"

blastdbcmd -db Ciona -info

Fasta slasher

#!/bin/sh

#mkdir /media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/Lost\_genes/

LIST='/media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/names.txt'

while read NAME; do

echo "Extracting $NAME"

folder="/media/andrew/49a47684-4a07-4f13-b9bb-930c306d951f/Danio/Genes/Lost\_genes/"

out=${folder}/${NAME}.fasta

samtools faidx 149\_missed\_genes.fasta $NAME > $out

echo "$NAME finished"

done < $LIST

OMA parameters:

InputDataType := 'AA';

OutputFolder := 'Output';

ReuseCachedResults := true;

AlignBatchSize := 1e6;

MinScore := 181;

LengthTol := 0.61;

StablePairTol := 1.81;

InparalogTol := 3.00;

ParalogTol := -2.5\*StablePairTol;

VerifiedPairTol := 1.53;

MinSeqLen := 50;

UseOnlyOneSplicingVariant := true;

UseExperimentalHomologousClusters := false;

QuasiCliquesCutoff := 1.0:

StableIdsForGroups := false;

GuessIdType := false;

DoHierarchicalGroups := 'top-down';

MaxTimePerLevel := 1200;

SpeciesTree := 'estimate';

ReachabilityCutoff := 0.65;

MinEdgeCompletenessFraction := 0.8;