*First get summary (?) of 1000 hits*

blastn -db nr -query week1\_query.fa -out blastresult -remote -max\_target\_seqs 1000

*After checking E value is less than 0.0001, get alignments from all 1000 hits.*

blastn -db nr -query week1\_query.fa -out alignblastresult -remote -num\_alignments 1000

*Change the format to FASTA file*

Using Python TextMate:

#!/bin/bash/env python

file = open("alignblastresult", 'r')

string\_list= file.read().split(">")

result = ""

f = open("output2", 'w')

for i in range (1, len(string\_list)):

gene\_name= string\_list[i].split("\n")[0]

result += ">" + gene\_name + '\n'

sequences = string\_list[i].split("Query")

for j in range (1, len(sequences)):

result += sequences[j].split("\n")[2].split()[-2] + '\n'

f.write(result)

result = ""

f.close()

file.close()

*Change FASTA to amino acid seq*

transeq text -outseq fakeoutput

*align the AA sequences*

mafft fakeoutput > align