SLE712 Assessment 3- Part 2-Determine the limits of BLAST

Yasodha Jayawardena & Dhanusha Weerawardana

GitHub: <https://github.com/Dhanusha782/SLE712---Bioinformatics-And-Molecular-Biology-Tec.-Assesment-3-part-2ni.git>

# Introduction

BLAST is a program or algorithm which is important combinaion of biology and technology to compare biological sequences including, amino acid sequences, and DNA and RNA nucleotied sequences. During this assesment, students able to gain knowladge on BLAST program as wel as able to determine the limits of BLAST. Therefore, in order to compare the sequences, E.Coli gene sequence and Student56 gene sequence have used through out the process.

# Prerequisite libraries

There are three packages that we will need to get things working. seqinr package designed for sequence data procesing and analysis R.utils package designed for extract data from compressed files rBLAST package provide R interface for BLAST search Additionally, we source two functions which supplied for this practical to perform an analysis into the limits of BLAST.

#libraries that we need  
library("seqinr")  
library("R.utils")

## Loading required package: R.oo

## Loading required package: R.methodsS3

## R.methodsS3 v1.8.0 (2020-02-14 07:10:20 UTC) successfully loaded. See ?R.methodsS3 for help.

## R.oo v1.23.0 successfully loaded. See ?R.oo for help.

##   
## Attaching package: 'R.oo'

## The following object is masked from 'package:R.methodsS3':  
##   
## throw

## The following object is masked from 'package:seqinr':  
##   
## getName

## The following objects are masked from 'package:methods':  
##   
## getClasses, getMethods

## The following objects are masked from 'package:base':  
##   
## attach, detach, load, save

## R.utils v2.9.2 successfully loaded. See ?R.utils for help.

##   
## Attaching package: 'R.utils'

## The following object is masked from 'package:utils':  
##   
## timestamp

## The following objects are masked from 'package:base':  
##   
## cat, commandArgs, getOption, inherits, isOpen, nullfile, parse,  
## warnings

library("rBLAST")

## Loading required package: Biostrings

## Loading required package: BiocGenerics

## Loading required package: parallel

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':  
##   
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
## clusterExport, clusterMap, parApply, parCapply, parLapply,  
## parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, append, as.data.frame, basename, cbind, colnames,  
## dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,  
## grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,  
## order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,  
## rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,  
## union, unique, unsplit, which, which.max, which.min

## Loading required package: S4Vectors

## Loading required package: stats4

##   
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:base':  
##   
## expand.grid

## Loading required package: IRanges

##   
## Attaching package: 'IRanges'

## The following object is masked from 'package:R.oo':  
##   
## trim

## Loading required package: XVector

##   
## Attaching package: 'Biostrings'

## The following object is masked from 'package:seqinr':  
##   
## translate

## The following object is masked from 'package:base':  
##   
## strsplit

# Question 1-Download the whole set of E. coli gene DNA sequences and use gunzip to decompress to create a blast database.

Here we are used the download.file() command to obtain the E.coli dataset from ensemble FTP page and then we are decompress the file using gunzip command. makeblast() function use to creat a BLAST database. \*Number of sequences are present in the E.coli set= 4140

#Download the E.coli gene sequency from the Ensemble FTP page  
download.file("ftp://ftp.ensemblgenomes.org/pub/bacteria/release-42/fasta/bacteria\_0\_collection/escherichia\_coli\_str\_k\_12\_substr\_mg1655/cds/Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa.gz", destfile = "Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa.gz")  
#use gunzip to uncompress the file  
gunzip("Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa.gz",overwrite=TRUE)  
#create a blast database with makeblast function  
makeblastdb("Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa",dbtype = "nucl","-parse\_seqids")

# Question 2-Download the sample fasta sequences and read the sequence.

Here we are used the download.file() command to obtain the sample file and then read it into R. The str use to determine data structure. Next select allocated sequence 56 and read it as “myEcoli” and check its structure. Determine length and propotion of GC bases in allocated sequence by using getLength() command and GC() command respectively. *Length=276 bp* The propotion of GC bases= 0.5326087

#Download sample file  
download.file("https://raw.githubusercontent.com/markziemann/SLE712\_files/master/bioinfo\_asst3\_part2\_files/sample.fa", destfile = "sample.fa")  
#Read file into R as Ecoli  
Ecoli<-read.fasta("sample.fa")  
str(Ecoli) #check data structure

## List of 100  
## $ 1 : 'SeqFastadna' chr [1:615] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "1"  
## ..- attr(\*, "Annot")= chr ">1 "  
## $ 2 : 'SeqFastadna' chr [1:525] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "2"  
## ..- attr(\*, "Annot")= chr ">2 "  
## $ 3 : 'SeqFastadna' chr [1:702] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "3"  
## ..- attr(\*, "Annot")= chr ">3 "  
## $ 4 : 'SeqFastadna' chr [1:1176] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "4"  
## ..- attr(\*, "Annot")= chr ">4 "  
## $ 5 : 'SeqFastadna' chr [1:399] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "5"  
## ..- attr(\*, "Annot")= chr ">5 "  
## $ 6 : 'SeqFastadna' chr [1:1203] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "6"  
## ..- attr(\*, "Annot")= chr ">6 "  
## $ 7 : 'SeqFastadna' chr [1:867] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "7"  
## ..- attr(\*, "Annot")= chr ">7 "  
## $ 8 : 'SeqFastadna' chr [1:723] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "8"  
## ..- attr(\*, "Annot")= chr ">8 "  
## $ 9 : 'SeqFastadna' chr [1:663] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "9"  
## ..- attr(\*, "Annot")= chr ">9 "  
## $ 10 : 'SeqFastadna' chr [1:762] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "10"  
## ..- attr(\*, "Annot")= chr ">10 "  
## $ 11 : 'SeqFastadna' chr [1:1497] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "11"  
## ..- attr(\*, "Annot")= chr ">11 "  
## $ 12 : 'SeqFastadna' chr [1:333] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "12"  
## ..- attr(\*, "Annot")= chr ">12 "  
## $ 13 : 'SeqFastadna' chr [1:273] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "13"  
## ..- attr(\*, "Annot")= chr ">13 "  
## $ 14 : 'SeqFastadna' chr [1:837] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "14"  
## ..- attr(\*, "Annot")= chr ">14 "  
## $ 15 : 'SeqFastadna' chr [1:1419] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "15"  
## ..- attr(\*, "Annot")= chr ">15 "  
## $ 16 : 'SeqFastadna' chr [1:735] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "16"  
## ..- attr(\*, "Annot")= chr ">16 "  
## $ 17 : 'SeqFastadna' chr [1:309] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "17"  
## ..- attr(\*, "Annot")= chr ">17 "  
## $ 18 : 'SeqFastadna' chr [1:552] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "18"  
## ..- attr(\*, "Annot")= chr ">18 "  
## $ 19 : 'SeqFastadna' chr [1:1293] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "19"  
## ..- attr(\*, "Annot")= chr ">19 "  
## $ 20 : 'SeqFastadna' chr [1:1539] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "20"  
## ..- attr(\*, "Annot")= chr ">20 "  
## $ 21 : 'SeqFastadna' chr [1:909] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "21"  
## ..- attr(\*, "Annot")= chr ">21 "  
## $ 22 : 'SeqFastadna' chr [1:294] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "22"  
## ..- attr(\*, "Annot")= chr ">22 "  
## $ 23 : 'SeqFastadna' chr [1:627] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "23"  
## ..- attr(\*, "Annot")= chr ">23 "  
## $ 24 : 'SeqFastadna' chr [1:816] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "24"  
## ..- attr(\*, "Annot")= chr ">24 "  
## $ 25 : 'SeqFastadna' chr [1:1047] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "25"  
## ..- attr(\*, "Annot")= chr ">25 "  
## $ 26 : 'SeqFastadna' chr [1:600] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "26"  
## ..- attr(\*, "Annot")= chr ">26 "  
## $ 27 : 'SeqFastadna' chr [1:747] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "27"  
## ..- attr(\*, "Annot")= chr ">27 "  
## $ 28 : 'SeqFastadna' chr [1:1440] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "28"  
## ..- attr(\*, "Annot")= chr ">28 "  
## $ 29 : 'SeqFastadna' chr [1:1275] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "29"  
## ..- attr(\*, "Annot")= chr ">29 "  
## $ 30 : 'SeqFastadna' chr [1:777] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "30"  
## ..- attr(\*, "Annot")= chr ">30 "  
## $ 31 : 'SeqFastadna' chr [1:621] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "31"  
## ..- attr(\*, "Annot")= chr ">31 "  
## $ 32 : 'SeqFastadna' chr [1:393] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "32"  
## ..- attr(\*, "Annot")= chr ">32 "  
## $ 33 : 'SeqFastadna' chr [1:1083] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "33"  
## ..- attr(\*, "Annot")= chr ">33 "  
## $ 34 : 'SeqFastadna' chr [1:297] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "34"  
## ..- attr(\*, "Annot")= chr ">34 "  
## $ 35 : 'SeqFastadna' chr [1:1659] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "35"  
## ..- attr(\*, "Annot")= chr ">35 "  
## $ 36 : 'SeqFastadna' chr [1:249] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "36"  
## ..- attr(\*, "Annot")= chr ">36 "  
## $ 37 : 'SeqFastadna' chr [1:735] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "37"  
## ..- attr(\*, "Annot")= chr ">37 "  
## $ 38 : 'SeqFastadna' chr [1:321] "g" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "38"  
## ..- attr(\*, "Annot")= chr ">38 "  
## $ 39 : 'SeqFastadna' chr [1:162] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "39"  
## ..- attr(\*, "Annot")= chr ">39 "  
## $ 40 : 'SeqFastadna' chr [1:2160] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "40"  
## ..- attr(\*, "Annot")= chr ">40 "  
## $ 41 : 'SeqFastadna' chr [1:414] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "41"  
## ..- attr(\*, "Annot")= chr ">41 "  
## $ 42 : 'SeqFastadna' chr [1:2013] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "42"  
## ..- attr(\*, "Annot")= chr ">42 "  
## $ 43 : 'SeqFastadna' chr [1:411] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "43"  
## ..- attr(\*, "Annot")= chr ">43 "  
## $ 44 : 'SeqFastadna' chr [1:1629] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "44"  
## ..- attr(\*, "Annot")= chr ">44 "  
## $ 45 : 'SeqFastadna' chr [1:1179] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "45"  
## ..- attr(\*, "Annot")= chr ">45 "  
## $ 46 : 'SeqFastadna' chr [1:1428] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "46"  
## ..- attr(\*, "Annot")= chr ">46 "  
## $ 47 : 'SeqFastadna' chr [1:318] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "47"  
## ..- attr(\*, "Annot")= chr ">47 "  
## $ 48 : 'SeqFastadna' chr [1:1059] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "48"  
## ..- attr(\*, "Annot")= chr ">48 "  
## $ 49 : 'SeqFastadna' chr [1:996] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "49"  
## ..- attr(\*, "Annot")= chr ">49 "  
## $ 50 : 'SeqFastadna' chr [1:1221] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "50"  
## ..- attr(\*, "Annot")= chr ">50 "  
## $ 51 : 'SeqFastadna' chr [1:699] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "51"  
## ..- attr(\*, "Annot")= chr ">51 "  
## $ 52 : 'SeqFastadna' chr [1:999] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "52"  
## ..- attr(\*, "Annot")= chr ">52 "  
## $ 53 : 'SeqFastadna' chr [1:789] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "53"  
## ..- attr(\*, "Annot")= chr ">53 "  
## $ 54 : 'SeqFastadna' chr [1:444] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "54"  
## ..- attr(\*, "Annot")= chr ">54 "  
## $ 55 : 'SeqFastadna' chr [1:1260] "g" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "55"  
## ..- attr(\*, "Annot")= chr ">55 "  
## $ 56 : 'SeqFastadna' chr [1:276] "g" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "56"  
## ..- attr(\*, "Annot")= chr ">56 "  
## $ 57 : 'SeqFastadna' chr [1:765] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "57"  
## ..- attr(\*, "Annot")= chr ">57 "  
## $ 58 : 'SeqFastadna' chr [1:930] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "58"  
## ..- attr(\*, "Annot")= chr ">58 "  
## $ 59 : 'SeqFastadna' chr [1:240] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "59"  
## ..- attr(\*, "Annot")= chr ">59 "  
## $ 60 : 'SeqFastadna' chr [1:900] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "60"  
## ..- attr(\*, "Annot")= chr ">60 "  
## $ 61 : 'SeqFastadna' chr [1:573] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "61"  
## ..- attr(\*, "Annot")= chr ">61 "  
## $ 62 : 'SeqFastadna' chr [1:696] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "62"  
## ..- attr(\*, "Annot")= chr ">62 "  
## $ 63 : 'SeqFastadna' chr [1:1806] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "63"  
## ..- attr(\*, "Annot")= chr ">63 "  
## $ 64 : 'SeqFastadna' chr [1:258] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "64"  
## ..- attr(\*, "Annot")= chr ">64 "  
## $ 65 : 'SeqFastadna' chr [1:279] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "65"  
## ..- attr(\*, "Annot")= chr ">65 "  
## $ 66 : 'SeqFastadna' chr [1:1359] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "66"  
## ..- attr(\*, "Annot")= chr ">66 "  
## $ 67 : 'SeqFastadna' chr [1:255] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "67"  
## ..- attr(\*, "Annot")= chr ">67 "  
## $ 68 : 'SeqFastadna' chr [1:1023] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "68"  
## ..- attr(\*, "Annot")= chr ">68 "  
## $ 69 : 'SeqFastadna' chr [1:1704] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "69"  
## ..- attr(\*, "Annot")= chr ">69 "  
## $ 70 : 'SeqFastadna' chr [1:657] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "70"  
## ..- attr(\*, "Annot")= chr ">70 "  
## $ 71 : 'SeqFastadna' chr [1:489] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "71"  
## ..- attr(\*, "Annot")= chr ">71 "  
## $ 72 : 'SeqFastadna' chr [1:927] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "72"  
## ..- attr(\*, "Annot")= chr ">72 "  
## $ 73 : 'SeqFastadna' chr [1:423] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "73"  
## ..- attr(\*, "Annot")= chr ">73 "  
## $ 74 : 'SeqFastadna' chr [1:576] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "74"  
## ..- attr(\*, "Annot")= chr ">74 "  
## $ 75 : 'SeqFastadna' chr [1:975] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "75"  
## ..- attr(\*, "Annot")= chr ">75 "  
## $ 76 : 'SeqFastadna' chr [1:882] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "76"  
## ..- attr(\*, "Annot")= chr ">76 "  
## $ 77 : 'SeqFastadna' chr [1:417] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "77"  
## ..- attr(\*, "Annot")= chr ">77 "  
## $ 78 : 'SeqFastadna' chr [1:348] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "78"  
## ..- attr(\*, "Annot")= chr ">78 "  
## $ 79 : 'SeqFastadna' chr [1:1152] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "79"  
## ..- attr(\*, "Annot")= chr ">79 "  
## $ 80 : 'SeqFastadna' chr [1:1248] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "80"  
## ..- attr(\*, "Annot")= chr ">80 "  
## $ 81 : 'SeqFastadna' chr [1:927] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "81"  
## ..- attr(\*, "Annot")= chr ">81 "  
## $ 82 : 'SeqFastadna' chr [1:729] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "82"  
## ..- attr(\*, "Annot")= chr ">82 "  
## $ 83 : 'SeqFastadna' chr [1:2298] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "83"  
## ..- attr(\*, "Annot")= chr ">83 "  
## $ 84 : 'SeqFastadna' chr [1:477] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "84"  
## ..- attr(\*, "Annot")= chr ">84 "  
## $ 85 : 'SeqFastadna' chr [1:378] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "85"  
## ..- attr(\*, "Annot")= chr ">85 "  
## $ 86 : 'SeqFastadna' chr [1:1716] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "86"  
## ..- attr(\*, "Annot")= chr ">86 "  
## $ 87 : 'SeqFastadna' chr [1:915] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "87"  
## ..- attr(\*, "Annot")= chr ">87 "  
## $ 88 : 'SeqFastadna' chr [1:1137] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "88"  
## ..- attr(\*, "Annot")= chr ">88 "  
## $ 89 : 'SeqFastadna' chr [1:447] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "89"  
## ..- attr(\*, "Annot")= chr ">89 "  
## $ 90 : 'SeqFastadna' chr [1:1209] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "90"  
## ..- attr(\*, "Annot")= chr ">90 "  
## $ 91 : 'SeqFastadna' chr [1:1266] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "91"  
## ..- attr(\*, "Annot")= chr ">91 "  
## $ 92 : 'SeqFastadna' chr [1:1539] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "92"  
## ..- attr(\*, "Annot")= chr ">92 "  
## $ 93 : 'SeqFastadna' chr [1:363] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "93"  
## ..- attr(\*, "Annot")= chr ">93 "  
## $ 94 : 'SeqFastadna' chr [1:603] "g" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "94"  
## ..- attr(\*, "Annot")= chr ">94 "  
## $ 95 : 'SeqFastadna' chr [1:1743] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "95"  
## ..- attr(\*, "Annot")= chr ">95 "  
## $ 96 : 'SeqFastadna' chr [1:696] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "96"  
## ..- attr(\*, "Annot")= chr ">96 "  
## $ 97 : 'SeqFastadna' chr [1:1422] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "97"  
## ..- attr(\*, "Annot")= chr ">97 "  
## $ 98 : 'SeqFastadna' chr [1:468] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "98"  
## ..- attr(\*, "Annot")= chr ">98 "  
## $ 99 : 'SeqFastadna' chr [1:810] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "99"  
## ..- attr(\*, "Annot")= chr ">99 "  
## [list output truncated]

#Read my allocated sequence (56) into R as myEcoli  
myEcoli<-Ecoli[[56]]  
#Lets check if the data has been imported properly  
myEcoli

## [1] "g" "t" "g" "g" "c" "t" "t" "c" "t" "g" "t" "t" "t" "c" "t" "a" "t" "c"  
## [19] "a" "g" "c" "t" "g" "t" "c" "c" "c" "t" "c" "c" "t" "g" "t" "t" "c" "a"  
## [37] "g" "c" "t" "a" "c" "t" "g" "a" "c" "g" "g" "g" "g" "t" "g" "g" "t" "g"  
## [55] "c" "g" "t" "a" "a" "c" "g" "g" "c" "a" "a" "a" "a" "g" "c" "a" "c" "t"  
## [73] "g" "c" "c" "g" "g" "a" "c" "a" "t" "c" "a" "g" "c" "g" "c" "t" "a" "t"  
## [91] "c" "t" "c" "t" "g" "c" "t" "c" "t" "c" "a" "c" "t" "g" "c" "c" "g" "t"  
## [109] "a" "a" "a" "a" "c" "a" "t" "g" "g" "c" "a" "a" "c" "t" "g" "c" "a" "g"  
## [127] "t" "t" "c" "a" "c" "t" "t" "a" "c" "a" "c" "c" "g" "c" "t" "t" "c" "t"  
## [145] "c" "a" "a" "c" "c" "c" "g" "g" "t" "a" "c" "g" "c" "a" "c" "c" "a" "g"  
## [163] "a" "a" "a" "a" "t" "c" "a" "t" "t" "g" "a" "t" "a" "t" "g" "g" "c" "c"  
## [181] "a" "t" "g" "a" "a" "t" "g" "g" "c" "g" "t" "t" "g" "g" "a" "t" "g" "c"  
## [199] "c" "g" "g" "g" "c" "a" "a" "c" "c" "g" "c" "c" "c" "g" "c" "a" "t" "t"  
## [217] "a" "t" "g" "g" "g" "c" "g" "t" "t" "g" "g" "c" "c" "t" "c" "a" "a" "c"  
## [235] "a" "c" "g" "a" "t" "t" "t" "t" "c" "c" "g" "c" "c" "a" "t" "t" "t" "a"  
## [253] "a" "a" "a" "a" "a" "c" "t" "c" "a" "g" "g" "c" "c" "g" "c" "a" "g" "t"  
## [271] "c" "g" "g" "t" "a" "a"  
## attr(,"name")  
## [1] "56"  
## attr(,"Annot")  
## [1] ">56 "  
## attr(,"class")  
## [1] "SeqFastadna"

str(myEcoli)

## 'SeqFastadna' chr [1:276] "g" "t" "g" "g" "c" "t" "t" "c" "t" "g" "t" "t" ...  
## - attr(\*, "name")= chr "56"  
## - attr(\*, "Annot")= chr ">56 "

#calculate length for my allocated gene  
seqinr::getLength(myEcoli)

## [1] 276

#calculate GC content for my allocated gene  
seqinr::GC(myEcoli)

## [1] 0.5326087

# Question 3-Use Blast to identify what E. coli gene your sequence matches best and show a table of percent identity, E-value and bit scores for top3 hits.

In order to creat a BLAST database, run the BLAST by the download.file(location of the file) command. Then res<- function can be applied to test the command and head() and str commands can be used to arrenge the table acoording to structure.

#sourced the function that ran blast searches in R  
download.file("https://raw.githubusercontent.com/markziemann/SLE712\_files/master/bioinfo\_asst3\_part2\_files/mutblast\_functions.R",destfile = "mutblast.R")  
source("mutblast.R")  
#test the function  
res<-myblastn\_tab(myseq = myEcoli,db= "Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa")  
#have a look at the blast results  
head(res)

## qseqid sseqid pident length mismatch gapopen qstart qend sstart send  
## 1 56 AAC73378 100.000 276 0 0 1 276 1 276  
## 2 56 AAC73368 100.000 276 0 0 1 276 1 276  
## 3 56 ABD18646 99.638 276 1 0 1 276 1 276  
## 4 56 AAC76469 98.188 276 5 0 1 276 1 276  
## 5 56 AAC74964 98.188 276 5 0 1 276 1 276  
## 6 56 AAC73133 98.188 276 5 0 1 276 1 276  
## evalue bitscore  
## 1 1.12e-151 531  
## 2 1.12e-151 531  
## 3 6.08e-150 525  
## 4 5.36e-143 502  
## 5 5.36e-143 502  
## 6 5.36e-143 502

str(res)

## 'data.frame': 13 obs. of 12 variables:  
## $ qseqid : int 56 56 56 56 56 56 56 56 56 56 ...  
## $ sseqid : Factor w/ 13 levels "AAC73132","AAC73133",..: 6 4 13 10 9 2 12 5 3 7 ...  
## $ pident : num 100 100 99.6 98.2 98.2 ...  
## $ length : int 276 276 276 276 276 276 276 82 82 82 ...  
## $ mismatch: int 0 0 1 5 5 5 26 0 0 1 ...  
## $ gapopen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ qstart : int 1 1 1 1 1 1 1 195 195 195 ...  
## $ qend : int 276 276 276 276 276 276 276 276 276 276 ...  
## $ sstart : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ send : int 276 276 276 276 276 276 276 82 82 82 ...  
## $ evalue : num 1.12e-151 1.12e-151 6.08e-150 5.36e-143 5.36e-143 ...  
## $ bitscore: int 531 531 525 502 502 502 381 158 158 152 ...

The first top 3 hits can be obtain by creating a character hits and applying running command for the hits.

###first three hits  
hits <-as.character(res$sseqid[1:3])  
hits

## [1] "AAC73378" "AAC73368" "ABD18646"

To identify the best matches between two gene sequences BLAST fuction can be used. Therefore, db<-read.fast command can be applied. The db is function for the database.

#identify what E.coil gene matches best  
db<-read.fasta("Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa")  
str(db)

## List of 4140  
## $ AAC73112: 'SeqFastadna' chr [1:66] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73112"  
## ..- attr(\*, "Annot")= chr ">AAC73112 cds chromosome:ASM584v2:Chromosome:190:255:1 gene:b0001 gene\_biotype:protein\_coding transcript\_biotyp"| \_\_truncated\_\_  
## $ AAC73113: 'SeqFastadna' chr [1:2463] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73113"  
## ..- attr(\*, "Annot")= chr ">AAC73113 cds chromosome:ASM584v2:Chromosome:337:2799:1 gene:b0002 gene\_biotype:protein\_coding transcript\_bioty"| \_\_truncated\_\_  
## $ AAC73114: 'SeqFastadna' chr [1:933] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73114"  
## ..- attr(\*, "Annot")= chr ">AAC73114 cds chromosome:ASM584v2:Chromosome:2801:3733:1 gene:b0003 gene\_biotype:protein\_coding transcript\_biot"| \_\_truncated\_\_  
## $ AAC73115: 'SeqFastadna' chr [1:1287] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73115"  
## ..- attr(\*, "Annot")= chr ">AAC73115 cds chromosome:ASM584v2:Chromosome:3734:5020:1 gene:b0004 gene\_biotype:protein\_coding transcript\_biot"| \_\_truncated\_\_  
## $ AAC73116: 'SeqFastadna' chr [1:297] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73116"  
## ..- attr(\*, "Annot")= chr ">AAC73116 cds chromosome:ASM584v2:Chromosome:5234:5530:1 gene:b0005 gene\_biotype:protein\_coding transcript\_biot"| \_\_truncated\_\_  
## $ AAC73117: 'SeqFastadna' chr [1:777] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73117"  
## ..- attr(\*, "Annot")= chr ">AAC73117 cds chromosome:ASM584v2:Chromosome:5683:6459:-1 gene:b0006 gene\_biotype:protein\_coding transcript\_bio"| \_\_truncated\_\_  
## $ AAC73118: 'SeqFastadna' chr [1:1431] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73118"  
## ..- attr(\*, "Annot")= chr ">AAC73118 cds chromosome:ASM584v2:Chromosome:6529:7959:-1 gene:b0007 gene\_biotype:protein\_coding transcript\_bio"| \_\_truncated\_\_  
## $ AAC73119: 'SeqFastadna' chr [1:954] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73119"  
## ..- attr(\*, "Annot")= chr ">AAC73119 cds chromosome:ASM584v2:Chromosome:8238:9191:1 gene:b0008 gene\_biotype:protein\_coding transcript\_biot"| \_\_truncated\_\_  
## $ AAC73120: 'SeqFastadna' chr [1:588] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73120"  
## ..- attr(\*, "Annot")= chr ">AAC73120 cds chromosome:ASM584v2:Chromosome:9306:9893:1 gene:b0009 gene\_biotype:protein\_coding transcript\_biot"| \_\_truncated\_\_  
## $ AAC73121: 'SeqFastadna' chr [1:567] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73121"  
## ..- attr(\*, "Annot")= chr ">AAC73121 cds chromosome:ASM584v2:Chromosome:9928:10494:-1 gene:b0010 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73122: 'SeqFastadna' chr [1:714] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73122"  
## ..- attr(\*, "Annot")= chr ">AAC73122 cds chromosome:ASM584v2:Chromosome:10643:11356:-1 gene:b0011 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73124: 'SeqFastadna' chr [1:405] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73124"  
## ..- attr(\*, "Annot")= chr ">AAC73124 cds chromosome:ASM584v2:Chromosome:11382:11786:-1 gene:b0013 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73125: 'SeqFastadna' chr [1:1917] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73125"  
## ..- attr(\*, "Annot")= chr ">AAC73125 cds chromosome:ASM584v2:Chromosome:12163:14079:1 gene:b0014 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73126: 'SeqFastadna' chr [1:1131] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73126"  
## ..- attr(\*, "Annot")= chr ">AAC73126 cds chromosome:ASM584v2:Chromosome:14168:15298:1 gene:b0015 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73127: 'SeqFastadna' chr [1:1113] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73127"  
## ..- attr(\*, "Annot")= chr ">AAC73127 cds chromosome:ASM584v2:Chromosome:15445:16557:1 gene:b0016 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAT48122: 'SeqFastadna' chr [1:153] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAT48122"  
## ..- attr(\*, "Annot")= chr ">AAT48122 cds chromosome:ASM584v2:Chromosome:16751:16903:-1 gene:b4412 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73129: 'SeqFastadna' chr [1:210] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73129"  
## ..- attr(\*, "Annot")= chr ">AAC73129 cds chromosome:ASM584v2:Chromosome:16751:16960:-1 gene:b0018 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73130: 'SeqFastadna' chr [1:1167] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73130"  
## ..- attr(\*, "Annot")= chr ">AAC73130 cds chromosome:ASM584v2:Chromosome:17489:18655:1 gene:b0019 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73131: 'SeqFastadna' chr [1:906] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73131"  
## ..- attr(\*, "Annot")= chr ">AAC73131 cds chromosome:ASM584v2:Chromosome:18715:19620:1 gene:b0020 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73132: 'SeqFastadna' chr [1:504] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73132"  
## ..- attr(\*, "Annot")= chr ">AAC73132 cds chromosome:ASM584v2:Chromosome:19811:20314:-1 gene:b0021 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73133: 'SeqFastadna' chr [1:276] "g" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73133"  
## ..- attr(\*, "Annot")= chr ">AAC73133 cds chromosome:ASM584v2:Chromosome:20233:20508:-1 gene:b0022 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73134: 'SeqFastadna' chr [1:264] "t" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73134"  
## ..- attr(\*, "Annot")= chr ">AAC73134 cds chromosome:ASM584v2:Chromosome:20815:21078:-1 gene:b0023 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73135: 'SeqFastadna' chr [1:219] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73135"  
## ..- attr(\*, "Annot")= chr ">AAC73135 cds chromosome:ASM584v2:Chromosome:21181:21399:1 gene:b0024 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73136: 'SeqFastadna' chr [1:942] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73136"  
## ..- attr(\*, "Annot")= chr ">AAC73136 cds chromosome:ASM584v2:Chromosome:21407:22348:1 gene:b0025 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73137: 'SeqFastadna' chr [1:2817] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73137"  
## ..- attr(\*, "Annot")= chr ">AAC73137 cds chromosome:ASM584v2:Chromosome:22391:25207:1 gene:b0026 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73138: 'SeqFastadna' chr [1:495] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73138"  
## ..- attr(\*, "Annot")= chr ">AAC73138 cds chromosome:ASM584v2:Chromosome:25207:25701:1 gene:b0027 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73139: 'SeqFastadna' chr [1:450] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73139"  
## ..- attr(\*, "Annot")= chr ">AAC73139 cds chromosome:ASM584v2:Chromosome:25826:26275:1 gene:b0028 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73140: 'SeqFastadna' chr [1:951] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73140"  
## ..- attr(\*, "Annot")= chr ">AAC73140 cds chromosome:ASM584v2:Chromosome:26277:27227:1 gene:b0029 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73141: 'SeqFastadna' chr [1:915] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73141"  
## ..- attr(\*, "Annot")= chr ">AAC73141 cds chromosome:ASM584v2:Chromosome:27293:28207:1 gene:b0030 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73142: 'SeqFastadna' chr [1:822] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73142"  
## ..- attr(\*, "Annot")= chr ">AAC73142 cds chromosome:ASM584v2:Chromosome:28374:29195:1 gene:b0031 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73143: 'SeqFastadna' chr [1:1149] "t" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73143"  
## ..- attr(\*, "Annot")= chr ">AAC73143 cds chromosome:ASM584v2:Chromosome:29651:30799:1 gene:b0032 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73144: 'SeqFastadna' chr [1:3222] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73144"  
## ..- attr(\*, "Annot")= chr ">AAC73144 cds chromosome:ASM584v2:Chromosome:30817:34038:1 gene:b0033 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73145: 'SeqFastadna' chr [1:396] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73145"  
## ..- attr(\*, "Annot")= chr ">AAC73145 cds chromosome:ASM584v2:Chromosome:34300:34695:1 gene:b0034 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73146: 'SeqFastadna' chr [1:591] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73146"  
## ..- attr(\*, "Annot")= chr ">AAC73146 cds chromosome:ASM584v2:Chromosome:34781:35371:-1 gene:b0035 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73147: 'SeqFastadna' chr [1:786] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73147"  
## ..- attr(\*, "Annot")= chr ">AAC73147 cds chromosome:ASM584v2:Chromosome:35377:36162:-1 gene:b0036 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73148: 'SeqFastadna' chr [1:1554] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73148"  
## ..- attr(\*, "Annot")= chr ">AAC73148 cds chromosome:ASM584v2:Chromosome:36271:37824:-1 gene:b0037 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73149: 'SeqFastadna' chr [1:1218] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73149"  
## ..- attr(\*, "Annot")= chr ">AAC73149 cds chromosome:ASM584v2:Chromosome:37898:39115:-1 gene:b0038 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73150: 'SeqFastadna' chr [1:1143] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73150"  
## ..- attr(\*, "Annot")= chr ">AAC73150 cds chromosome:ASM584v2:Chromosome:39244:40386:-1 gene:b0039 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73151: 'SeqFastadna' chr [1:1515] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73151"  
## ..- attr(\*, "Annot")= chr ">AAC73151 cds chromosome:ASM584v2:Chromosome:40417:41931:-1 gene:b0040 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73152: 'SeqFastadna' chr [1:771] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73152"  
## ..- attr(\*, "Annot")= chr ">AAC73152 cds chromosome:ASM584v2:Chromosome:42403:43173:1 gene:b0041 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73153: 'SeqFastadna' chr [1:942] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73153"  
## ..- attr(\*, "Annot")= chr ">AAC73153 cds chromosome:ASM584v2:Chromosome:43188:44129:1 gene:b0042 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73154: 'SeqFastadna' chr [1:1287] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73154"  
## ..- attr(\*, "Annot")= chr ">AAC73154 cds chromosome:ASM584v2:Chromosome:44180:45466:1 gene:b0043 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73155: 'SeqFastadna' chr [1:288] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73155"  
## ..- attr(\*, "Annot")= chr ">AAC73155 cds chromosome:ASM584v2:Chromosome:45463:45750:1 gene:b0044 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73156: 'SeqFastadna' chr [1:1332] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73156"  
## ..- attr(\*, "Annot")= chr ">AAC73156 cds chromosome:ASM584v2:Chromosome:45807:47138:1 gene:b0045 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73157: 'SeqFastadna' chr [1:531] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73157"  
## ..- attr(\*, "Annot")= chr ">AAC73157 cds chromosome:ASM584v2:Chromosome:47246:47776:1 gene:b0046 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73158: 'SeqFastadna' chr [1:1863] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73158"  
## ..- attr(\*, "Annot")= chr ">AAC73158 cds chromosome:ASM584v2:Chromosome:47769:49631:1 gene:b0047 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73159: 'SeqFastadna' chr [1:480] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73159"  
## ..- attr(\*, "Annot")= chr ">AAC73159 cds chromosome:ASM584v2:Chromosome:49823:50302:1 gene:b0048 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73160: 'SeqFastadna' chr [1:843] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73160"  
## ..- attr(\*, "Annot")= chr ">AAC73160 cds chromosome:ASM584v2:Chromosome:50380:51222:-1 gene:b0049 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73161: 'SeqFastadna' chr [1:378] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73161"  
## ..- attr(\*, "Annot")= chr ">AAC73161 cds chromosome:ASM584v2:Chromosome:51229:51606:-1 gene:b0050 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73162: 'SeqFastadna' chr [1:822] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73162"  
## ..- attr(\*, "Annot")= chr ">AAC73162 cds chromosome:ASM584v2:Chromosome:51609:52430:-1 gene:b0051 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73163: 'SeqFastadna' chr [1:990] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73163"  
## ..- attr(\*, "Annot")= chr ">AAC73163 cds chromosome:ASM584v2:Chromosome:52427:53416:-1 gene:b0052 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73164: 'SeqFastadna' chr [1:1287] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73164"  
## ..- attr(\*, "Annot")= chr ">AAC73164 cds chromosome:ASM584v2:Chromosome:53416:54702:-1 gene:b0053 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73165: 'SeqFastadna' chr [1:2355] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73165"  
## ..- attr(\*, "Annot")= chr ">AAC73165 cds chromosome:ASM584v2:Chromosome:54755:57109:-1 gene:b0054 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73166: 'SeqFastadna' chr [1:816] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73166"  
## ..- attr(\*, "Annot")= chr ">AAC73166 cds chromosome:ASM584v2:Chromosome:57364:58179:1 gene:b0055 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73169: 'SeqFastadna' chr [1:660] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73169"  
## ..- attr(\*, "Annot")= chr ">AAC73169 cds chromosome:ASM584v2:Chromosome:59687:60346:-1 gene:b0058 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73170: 'SeqFastadna' chr [1:2907] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73170"  
## ..- attr(\*, "Annot")= chr ">AAC73170 cds chromosome:ASM584v2:Chromosome:60358:63264:-1 gene:b0059 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73171: 'SeqFastadna' chr [1:2352] "g" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73171"  
## ..- attr(\*, "Annot")= chr ">AAC73171 cds chromosome:ASM584v2:Chromosome:63429:65780:-1 gene:b0060 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73172: 'SeqFastadna' chr [1:696] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73172"  
## ..- attr(\*, "Annot")= chr ">AAC73172 cds chromosome:ASM584v2:Chromosome:65855:66550:-1 gene:b0061 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73173: 'SeqFastadna' chr [1:1503] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73173"  
## ..- attr(\*, "Annot")= chr ">AAC73173 cds chromosome:ASM584v2:Chromosome:66835:68337:-1 gene:b0062 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73174: 'SeqFastadna' chr [1:1701] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73174"  
## ..- attr(\*, "Annot")= chr ">AAC73174 cds chromosome:ASM584v2:Chromosome:68348:70048:-1 gene:b0063 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73175: 'SeqFastadna' chr [1:879] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73175"  
## ..- attr(\*, "Annot")= chr ">AAC73175 cds chromosome:ASM584v2:Chromosome:70387:71265:1 gene:b0064 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73176: 'SeqFastadna' chr [1:765] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73176"  
## ..- attr(\*, "Annot")= chr ">AAC73176 cds chromosome:ASM584v2:Chromosome:71351:72115:1 gene:b0065 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73177: 'SeqFastadna' chr [1:699] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73177"  
## ..- attr(\*, "Annot")= chr ">AAC73177 cds chromosome:ASM584v2:Chromosome:72229:72927:-1 gene:b0066 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73178: 'SeqFastadna' chr [1:1611] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73178"  
## ..- attr(\*, "Annot")= chr ">AAC73178 cds chromosome:ASM584v2:Chromosome:72911:74521:-1 gene:b0067 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73179: 'SeqFastadna' chr [1:984] "g" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73179"  
## ..- attr(\*, "Annot")= chr ">AAC73179 cds chromosome:ASM584v2:Chromosome:74497:75480:-1 gene:b0068 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73180: 'SeqFastadna' chr [1:1656] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73180"  
## ..- attr(\*, "Annot")= chr ">AAC73180 cds chromosome:ASM584v2:Chromosome:75644:77299:-1 gene:b0069 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ ACO59989: 'SeqFastadna' chr [1:132] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "ACO59989"  
## ..- attr(\*, "Annot")= chr ">ACO59989 cds chromosome:ASM584v2:Chromosome:77388:77519:1 gene:b4662 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAT48123: 'SeqFastadna' chr [1:1179] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAT48123"  
## ..- attr(\*, "Annot")= chr ">AAT48123 cds chromosome:ASM584v2:Chromosome:77621:78799:1 gene:b0070 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73182: 'SeqFastadna' chr [1:606] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73182"  
## ..- attr(\*, "Annot")= chr ">AAC73182 cds chromosome:ASM584v2:Chromosome:78848:79453:-1 gene:b0071 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73183: 'SeqFastadna' chr [1:1401] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73183"  
## ..- attr(\*, "Annot")= chr ">AAC73183 cds chromosome:ASM584v2:Chromosome:79464:80864:-1 gene:b0072 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73184: 'SeqFastadna' chr [1:1092] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73184"  
## ..- attr(\*, "Annot")= chr ">AAC73184 cds chromosome:ASM584v2:Chromosome:80867:81958:-1 gene:b0073 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73185: 'SeqFastadna' chr [1:1572] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73185"  
## ..- attr(\*, "Annot")= chr ">AAC73185 cds chromosome:ASM584v2:Chromosome:81958:83529:-1 gene:b0074 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73186: 'SeqFastadna' chr [1:87] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73186"  
## ..- attr(\*, "Annot")= chr ">AAC73186 cds chromosome:ASM584v2:Chromosome:83622:83708:-1 gene:b0075 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73187: 'SeqFastadna' chr [1:945] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73187"  
## ..- attr(\*, "Annot")= chr ">AAC73187 cds chromosome:ASM584v2:Chromosome:84368:85312:1 gene:b0076 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73188: 'SeqFastadna' chr [1:1725] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73188"  
## ..- attr(\*, "Annot")= chr ">AAC73188 cds chromosome:ASM584v2:Chromosome:85630:87354:1 gene:b0077 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73189: 'SeqFastadna' chr [1:492] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73189"  
## ..- attr(\*, "Annot")= chr ">AAC73189 cds chromosome:ASM584v2:Chromosome:87357:87848:1 gene:b0078 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73191: 'SeqFastadna' chr [1:1005] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73191"  
## ..- attr(\*, "Annot")= chr ">AAC73191 cds chromosome:ASM584v2:Chromosome:88028:89032:1 gene:b0080 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73192: 'SeqFastadna' chr [1:459] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73192"  
## ..- attr(\*, "Annot")= chr ">AAC73192 cds chromosome:ASM584v2:Chromosome:89634:90092:1 gene:b0081 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73193: 'SeqFastadna' chr [1:942] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73193"  
## ..- attr(\*, "Annot")= chr ">AAC73193 cds chromosome:ASM584v2:Chromosome:90094:91035:1 gene:b0082 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73194: 'SeqFastadna' chr [1:366] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73194"  
## ..- attr(\*, "Annot")= chr ">AAC73194 cds chromosome:ASM584v2:Chromosome:91032:91397:1 gene:b0083 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73195: 'SeqFastadna' chr [1:1767] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73195"  
## ..- attr(\*, "Annot")= chr ">AAC73195 cds chromosome:ASM584v2:Chromosome:91413:93179:1 gene:b0084 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73196: 'SeqFastadna' chr [1:1488] "g" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73196"  
## ..- attr(\*, "Annot")= chr ">AAC73196 cds chromosome:ASM584v2:Chromosome:93166:94653:1 gene:b0085 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73197: 'SeqFastadna' chr [1:1359] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73197"  
## ..- attr(\*, "Annot")= chr ">AAC73197 cds chromosome:ASM584v2:Chromosome:94650:96008:1 gene:b0086 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73198: 'SeqFastadna' chr [1:1083] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73198"  
## ..- attr(\*, "Annot")= chr ">AAC73198 cds chromosome:ASM584v2:Chromosome:96002:97084:1 gene:b0087 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73199: 'SeqFastadna' chr [1:1317] "a" "t" "g" "g" ...  
## ..- attr(\*, "name")= chr "AAC73199"  
## ..- attr(\*, "Annot")= chr ">AAC73199 cds chromosome:ASM584v2:Chromosome:97087:98403:1 gene:b0088 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73200: 'SeqFastadna' chr [1:1245] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73200"  
## ..- attr(\*, "Annot")= chr ">AAC73200 cds chromosome:ASM584v2:Chromosome:98403:99647:1 gene:b0089 gene\_biotype:protein\_coding transcript\_bi"| \_\_truncated\_\_  
## $ AAC73201: 'SeqFastadna' chr [1:1068] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73201"  
## ..- attr(\*, "Annot")= chr ">AAC73201 cds chromosome:ASM584v2:Chromosome:99644:100711:1 gene:b0090 gene\_biotype:protein\_coding transcript\_b"| \_\_truncated\_\_  
## $ AAC73202: 'SeqFastadna' chr [1:1476] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73202"  
## ..- attr(\*, "Annot")= chr ">AAC73202 cds chromosome:ASM584v2:Chromosome:100765:102240:1 gene:b0091 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73203: 'SeqFastadna' chr [1:921] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73203"  
## ..- attr(\*, "Annot")= chr ">AAC73203 cds chromosome:ASM584v2:Chromosome:102233:103153:1 gene:b0092 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73204: 'SeqFastadna' chr [1:831] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73204"  
## ..- attr(\*, "Annot")= chr ">AAC73204 cds chromosome:ASM584v2:Chromosome:103155:103985:1 gene:b0093 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73205: 'SeqFastadna' chr [1:1263] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73205"  
## ..- attr(\*, "Annot")= chr ">AAC73205 cds chromosome:ASM584v2:Chromosome:103982:105244:1 gene:b0094 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73206: 'SeqFastadna' chr [1:1152] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73206"  
## ..- attr(\*, "Annot")= chr ">AAC73206 cds chromosome:ASM584v2:Chromosome:105305:106456:1 gene:b0095 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73207: 'SeqFastadna' chr [1:918] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73207"  
## ..- attr(\*, "Annot")= chr ">AAC73207 cds chromosome:ASM584v2:Chromosome:106557:107474:1 gene:b0096 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73208: 'SeqFastadna' chr [1:513] "g" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73208"  
## ..- attr(\*, "Annot")= chr ">AAC73208 cds chromosome:ASM584v2:Chromosome:107705:108217:1 gene:b0097 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73209: 'SeqFastadna' chr [1:2706] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73209"  
## ..- attr(\*, "Annot")= chr ">AAC73209 cds chromosome:ASM584v2:Chromosome:108279:110984:1 gene:b0098 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73210: 'SeqFastadna' chr [1:390] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73210"  
## ..- attr(\*, "Annot")= chr ">AAC73210 cds chromosome:ASM584v2:Chromosome:111044:111433:1 gene:b0099 gene\_biotype:protein\_coding transcript\_"| \_\_truncated\_\_  
## $ AAC73212: 'SeqFastadna' chr [1:198] "a" "t" "g" "t" ...  
## ..- attr(\*, "name")= chr "AAC73212"  
## ..- attr(\*, "Annot")= chr ">AAC73212 cds chromosome:ASM584v2:Chromosome:111649:111846:-1 gene:b0101 gene\_biotype:protein\_coding transcript"| \_\_truncated\_\_  
## $ AAC73213: 'SeqFastadna' chr [1:744] "a" "t" "g" "c" ...  
## ..- attr(\*, "name")= chr "AAC73213"  
## ..- attr(\*, "Annot")= chr ">AAC73213 cds chromosome:ASM584v2:Chromosome:111856:112599:-1 gene:b0102 gene\_biotype:protein\_coding transcript"| \_\_truncated\_\_  
## $ AAC73214: 'SeqFastadna' chr [1:621] "a" "t" "g" "a" ...  
## ..- attr(\*, "name")= chr "AAC73214"  
## ..- attr(\*, "Annot")= chr ">AAC73214 cds chromosome:ASM584v2:Chromosome:112599:113219:-1 gene:b0103 gene\_biotype:protein\_coding transcript"| \_\_truncated\_\_  
## [list output truncated]

head(names(db))

## [1] "AAC73112" "AAC73113" "AAC73114" "AAC73115" "AAC73116" "AAC73117"

head(db)

## $AAC73112  
## [1] "a" "t" "g" "a" "a" "a" "c" "g" "c" "a" "t" "t" "a" "g" "c" "a" "c" "c" "a"  
## [20] "c" "c" "a" "t" "t" "a" "c" "c" "a" "c" "c" "a" "c" "c" "a" "t" "c" "a" "c"  
## [39] "c" "a" "t" "t" "a" "c" "c" "a" "c" "a" "g" "g" "t" "a" "a" "c" "g" "g" "t"  
## [58] "g" "c" "g" "g" "g" "c" "t" "g" "a"  
## attr(,"name")  
## [1] "AAC73112"  
## attr(,"Annot")  
## [1] ">AAC73112 cds chromosome:ASM584v2:Chromosome:190:255:1 gene:b0001 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding gene\_symbol:thrL description:thr operon leader peptide"  
## attr(,"class")  
## [1] "SeqFastadna"  
##   
## $AAC73113  
## [1] "a" "t" "g" "c" "g" "a" "g" "t" "g" "t" "t" "g" "a" "a" "g" "t" "t" "c"  
## [19] "g" "g" "c" "g" "g" "t" "a" "c" "a" "t" "c" "a" "g" "t" "g" "g" "c" "a"  
## [37] "a" "a" "t" "g" "c" "a" "g" "a" "a" "c" "g" "t" "t" "t" "t" "c" "t" "g"  
## [55] "c" "g" "t" "g" "t" "t" "g" "c" "c" "g" "a" "t" "a" "t" "t" "c" "t" "g"  
## [73] "g" "a" "a" "a" "g" "c" "a" "a" "t" "g" "c" "c" "a" "g" "g" "c" "a" "g"  
## [91] "g" "g" "g" "c" "a" "g" "g" "t" "g" "g" "c" "c" "a" "c" "c" "g" "t" "c"  
## [109] "c" "t" "c" "t" "c" "t" "g" "c" "c" "c" "c" "c" "g" "c" "c" "a" "a" "a"  
## [127] "a" "t" "c" "a" "c" "c" "a" "a" "c" "c" "a" "c" "c" "t" "g" "g" "t" "g"  
## [145] "g" "c" "g" "a" "t" "g" "a" "t" "t" "g" "a" "a" "a" "a" "a" "a" "c" "c"  
## [163] "a" "t" "t" "a" "g" "c" "g" "g" "c" "c" "a" "g" "g" "a" "t" "g" "c" "t"  
## [181] "t" "t" "a" "c" "c" "c" "a" "a" "t" "a" "t" "c" "a" "g" "c" "g" "a" "t"  
## [199] "g" "c" "c" "g" "a" "a" "c" "g" "t" "a" "t" "t" "t" "t" "t" "g" "c" "c"  
## [217] "g" "a" "a" "c" "t" "t" "t" "t" "g" "a" "c" "g" "g" "g" "a" "c" "t" "c"  
## [235] "g" "c" "c" "g" "c" "c" "g" "c" "c" "c" "a" "g" "c" "c" "g" "g" "g" "g"  
## [253] "t" "t" "c" "c" "c" "g" "c" "t" "g" "g" "c" "g" "c" "a" "a" "t" "t" "g"  
## [271] "a" "a" "a" "a" "c" "t" "t" "t" "c" "g" "t" "c" "g" "a" "t" "c" "a" "g"  
## [289] "g" "a" "a" "t" "t" "t" "g" "c" "c" "c" "a" "a" "a" "t" "a" "a" "a" "a"  
## [307] "c" "a" "t" "g" "t" "c" "c" "t" "g" "c" "a" "t" "g" "g" "c" "a" "t" "t"  
## [325] "a" "g" "t" "t" "t" "g" "t" "t" "g" "g" "g" "g" "c" "a" "g" "t" "g" "c"  
## [343] "c" "c" "g" "g" "a" "t" "a" "g" "c" "a" "t" "c" "a" "a" "c" "g" "c" "t"  
## [361] "g" "c" "g" "c" "t" "g" "a" "t" "t" "t" "g" "c" "c" "g" "t" "g" "g" "c"  
## [379] "g" "a" "g" "a" "a" "a" "a" "t" "g" "t" "c" "g" "a" "t" "c" "g" "c" "c"  
## [397] "a" "t" "t" "a" "t" "g" "g" "c" "c" "g" "g" "c" "g" "t" "a" "t" "t" "a"  
## [415] "g" "a" "a" "g" "c" "g" "c" "g" "c" "g" "g" "t" "c" "a" "c" "a" "a" "c"  
## [433] "g" "t" "t" "a" "c" "t" "g" "t" "t" "a" "t" "c" "g" "a" "t" "c" "c" "g"  
## [451] "g" "t" "c" "g" "a" "a" "a" "a" "a" "c" "t" "g" "c" "t" "g" "g" "c" "a"  
## [469] "g" "t" "g" "g" "g" "g" "c" "a" "t" "t" "a" "c" "c" "t" "c" "g" "a" "a"  
## [487] "t" "c" "t" "a" "c" "c" "g" "t" "c" "g" "a" "t" "a" "t" "t" "g" "c" "t"  
## [505] "g" "a" "g" "t" "c" "c" "a" "c" "c" "c" "g" "c" "c" "g" "t" "a" "t" "t"  
## [523] "g" "c" "g" "g" "c" "a" "a" "g" "c" "c" "g" "c" "a" "t" "t" "c" "c" "g"  
## [541] "g" "c" "t" "g" "a" "t" "c" "a" "c" "a" "t" "g" "g" "t" "g" "c" "t" "g"  
## [559] "a" "t" "g" "g" "c" "a" "g" "g" "t" "t" "t" "c" "a" "c" "c" "g" "c" "c"  
## [577] "g" "g" "t" "a" "a" "t" "g" "a" "a" "a" "a" "a" "g" "g" "c" "g" "a" "a"  
## [595] "c" "t" "g" "g" "t" "g" "g" "t" "g" "c" "t" "t" "g" "g" "a" "c" "g" "c"  
## [613] "a" "a" "c" "g" "g" "t" "t" "c" "c" "g" "a" "c" "t" "a" "c" "t" "c" "t"  
## [631] "g" "c" "t" "g" "c" "g" "g" "t" "g" "c" "t" "g" "g" "c" "t" "g" "c" "c"  
## [649] "t" "g" "t" "t" "t" "a" "c" "g" "c" "g" "c" "c" "g" "a" "t" "t" "g" "t"  
## [667] "t" "g" "c" "g" "a" "g" "a" "t" "t" "t" "g" "g" "a" "c" "g" "g" "a" "c"  
## [685] "g" "t" "t" "g" "a" "c" "g" "g" "g" "g" "t" "c" "t" "a" "t" "a" "c" "c"  
## [703] "t" "g" "c" "g" "a" "c" "c" "c" "g" "c" "g" "t" "c" "a" "g" "g" "t" "g"  
## [721] "c" "c" "c" "g" "a" "t" "g" "c" "g" "a" "g" "g" "t" "t" "g" "t" "t" "g"  
## [739] "a" "a" "g" "t" "c" "g" "a" "t" "g" "t" "c" "c" "t" "a" "c" "c" "a" "g"  
## [757] "g" "a" "a" "g" "c" "g" "a" "t" "g" "g" "a" "g" "c" "t" "t" "t" "c" "c"  
## [775] "t" "a" "c" "t" "t" "c" "g" "g" "c" "g" "c" "t" "a" "a" "a" "g" "t" "t"  
## [793] "c" "t" "t" "c" "a" "c" "c" "c" "c" "c" "g" "c" "a" "c" "c" "a" "t" "t"  
## [811] "a" "c" "c" "c" "c" "c" "a" "t" "c" "g" "c" "c" "c" "a" "g" "t" "t" "c"  
## [829] "c" "a" "g" "a" "t" "c" "c" "c" "t" "t" "g" "c" "c" "t" "g" "a" "t" "t"  
## [847] "a" "a" "a" "a" "a" "t" "a" "c" "c" "g" "g" "a" "a" "a" "t" "c" "c" "t"  
## [865] "c" "a" "a" "g" "c" "a" "c" "c" "a" "g" "g" "t" "a" "c" "g" "c" "t" "c"  
## [883] "a" "t" "t" "g" "g" "t" "g" "c" "c" "a" "g" "c" "c" "g" "t" "g" "a" "t"  
## [901] "g" "a" "a" "g" "a" "c" "g" "a" "a" "t" "t" "a" "c" "c" "g" "g" "t" "c"  
## [919] "a" "a" "g" "g" "g" "c" "a" "t" "t" "t" "c" "c" "a" "a" "t" "c" "t" "g"  
## [937] "a" "a" "t" "a" "a" "c" "a" "t" "g" "g" "c" "a" "a" "t" "g" "t" "t" "c"  
## [955] "a" "g" "c" "g" "t" "t" "t" "c" "t" "g" "g" "t" "c" "c" "g" "g" "g" "g"  
## [973] "a" "t" "g" "a" "a" "a" "g" "g" "g" "a" "t" "g" "g" "t" "c" "g" "g" "c"  
## [991] "a" "t" "g" "g" "c" "g" "g" "c" "g" "c" "g" "c" "g" "t" "c" "t" "t" "t"  
## [1009] "g" "c" "a" "g" "c" "g" "a" "t" "g" "t" "c" "a" "c" "g" "c" "g" "c" "c"  
## [1027] "c" "g" "t" "a" "t" "t" "t" "c" "c" "g" "t" "g" "g" "t" "g" "c" "t" "g"  
## [1045] "a" "t" "t" "a" "c" "g" "c" "a" "a" "t" "c" "a" "t" "c" "t" "t" "c" "c"  
## [1063] "g" "a" "a" "t" "a" "c" "a" "g" "c" "a" "t" "c" "a" "g" "t" "t" "t" "c"  
## [1081] "t" "g" "c" "g" "t" "t" "c" "c" "a" "c" "a" "a" "a" "g" "c" "g" "a" "c"  
## [1099] "t" "g" "t" "g" "t" "g" "c" "g" "a" "g" "c" "t" "g" "a" "a" "c" "g" "g"  
## [1117] "g" "c" "a" "a" "t" "g" "c" "a" "g" "g" "a" "a" "g" "a" "g" "t" "t" "c"  
## [1135] "t" "a" "c" "c" "t" "g" "g" "a" "a" "c" "t" "g" "a" "a" "a" "g" "a" "a"  
## [1153] "g" "g" "c" "t" "t" "a" "c" "t" "g" "g" "a" "g" "c" "c" "g" "c" "t" "g"  
## [1171] "g" "c" "a" "g" "t" "g" "a" "c" "g" "g" "a" "a" "c" "g" "g" "c" "t" "g"  
## [1189] "g" "c" "c" "a" "t" "t" "a" "t" "c" "t" "c" "g" "g" "t" "g" "g" "t" "a"  
## [1207] "g" "g" "t" "g" "a" "t" "g" "g" "t" "a" "t" "g" "c" "g" "c" "a" "c" "c"  
## [1225] "t" "t" "g" "c" "g" "t" "g" "g" "g" "a" "t" "c" "t" "c" "g" "g" "c" "g"  
## [1243] "a" "a" "a" "t" "t" "c" "t" "t" "t" "g" "c" "c" "g" "c" "a" "c" "t" "g"  
## [1261] "g" "c" "c" "c" "g" "c" "g" "c" "c" "a" "a" "t" "a" "t" "c" "a" "a" "c"  
## [1279] "a" "t" "t" "g" "t" "c" "g" "c" "c" "a" "t" "t" "g" "c" "t" "c" "a" "g"  
## [1297] "g" "g" "a" "t" "c" "t" "t" "c" "t" "g" "a" "a" "c" "g" "c" "t" "c" "a"  
## [1315] "a" "t" "c" "t" "c" "t" "g" "t" "c" "g" "t" "g" "g" "t" "a" "a" "a" "t"  
## [1333] "a" "a" "c" "g" "a" "t" "g" "a" "t" "g" "c" "g" "a" "c" "c" "a" "c" "t"  
## [1351] "g" "g" "c" "g" "t" "g" "c" "g" "c" "g" "t" "t" "a" "c" "t" "c" "a" "t"  
## [1369] "c" "a" "g" "a" "t" "g" "c" "t" "g" "t" "t" "c" "a" "a" "t" "a" "c" "c"  
## [1387] "g" "a" "t" "c" "a" "g" "g" "t" "t" "a" "t" "c" "g" "a" "a" "g" "t" "g"  
## [1405] "t" "t" "t" "g" "t" "g" "a" "t" "t" "g" "g" "c" "g" "t" "c" "g" "g" "t"  
## [1423] "g" "g" "c" "g" "t" "t" "g" "g" "c" "g" "g" "t" "g" "c" "g" "c" "t" "g"  
## [1441] "c" "t" "g" "g" "a" "g" "c" "a" "a" "c" "t" "g" "a" "a" "g" "c" "g" "t"  
## [1459] "c" "a" "g" "c" "a" "a" "a" "g" "c" "t" "g" "g" "c" "t" "g" "a" "a" "g"  
## [1477] "a" "a" "t" "a" "a" "a" "c" "a" "t" "a" "t" "c" "g" "a" "c" "t" "t" "a"  
## [1495] "c" "g" "t" "g" "t" "c" "t" "g" "c" "g" "g" "t" "g" "t" "t" "g" "c" "c"  
## [1513] "a" "a" "c" "t" "c" "g" "a" "a" "g" "g" "c" "t" "c" "t" "g" "c" "t" "c"  
## [1531] "a" "c" "c" "a" "a" "t" "g" "t" "a" "c" "a" "t" "g" "g" "c" "c" "t" "t"  
## [1549] "a" "a" "t" "c" "t" "g" "g" "a" "a" "a" "a" "c" "t" "g" "g" "c" "a" "g"  
## [1567] "g" "a" "a" "g" "a" "a" "c" "t" "g" "g" "c" "g" "c" "a" "a" "g" "c" "c"  
## [1585] "a" "a" "a" "g" "a" "g" "c" "c" "g" "t" "t" "t" "a" "a" "t" "c" "t" "c"  
## [1603] "g" "g" "g" "c" "g" "c" "t" "t" "a" "a" "t" "t" "c" "g" "c" "c" "t" "c"  
## [1621] "g" "t" "g" "a" "a" "a" "g" "a" "a" "t" "a" "t" "c" "a" "t" "c" "t" "g"  
## [1639] "c" "t" "g" "a" "a" "c" "c" "c" "g" "g" "t" "c" "a" "t" "t" "g" "t" "t"  
## [1657] "g" "a" "c" "t" "g" "c" "a" "c" "t" "t" "c" "c" "a" "g" "c" "c" "a" "g"  
## [1675] "g" "c" "a" "g" "t" "g" "g" "c" "g" "g" "a" "t" "c" "a" "a" "t" "a" "t"  
## [1693] "g" "c" "c" "g" "a" "c" "t" "t" "c" "c" "t" "g" "c" "g" "c" "g" "a" "a"  
## [1711] "g" "g" "t" "t" "t" "c" "c" "a" "c" "g" "t" "t" "g" "t" "c" "a" "c" "g"  
## [1729] "c" "c" "g" "a" "a" "c" "a" "a" "a" "a" "a" "g" "g" "c" "c" "a" "a" "c"  
## [1747] "a" "c" "c" "t" "c" "g" "t" "c" "g" "a" "t" "g" "g" "a" "t" "t" "a" "c"  
## [1765] "t" "a" "c" "c" "a" "t" "c" "a" "g" "t" "t" "g" "c" "g" "t" "t" "a" "t"  
## [1783] "g" "c" "g" "g" "c" "g" "g" "a" "a" "a" "a" "a" "t" "c" "g" "c" "g" "g"  
## [1801] "c" "g" "t" "a" "a" "a" "t" "t" "c" "c" "t" "c" "t" "a" "t" "g" "a" "c"  
## [1819] "a" "c" "c" "a" "a" "c" "g" "t" "t" "g" "g" "g" "g" "c" "t" "g" "g" "a"  
## [1837] "t" "t" "a" "c" "c" "g" "g" "t" "t" "a" "t" "t" "g" "a" "g" "a" "a" "c"  
## [1855] "c" "t" "g" "c" "a" "a" "a" "a" "t" "c" "t" "g" "c" "t" "c" "a" "a" "t"  
## [1873] "g" "c" "a" "g" "g" "t" "g" "a" "t" "g" "a" "a" "t" "t" "g" "a" "t" "g"  
## [1891] "a" "a" "g" "t" "t" "c" "t" "c" "c" "g" "g" "c" "a" "t" "t" "c" "t" "t"  
## [1909] "t" "c" "t" "g" "g" "t" "t" "c" "g" "c" "t" "t" "t" "c" "t" "t" "a" "t"  
## [1927] "a" "t" "c" "t" "t" "c" "g" "g" "c" "a" "a" "g" "t" "t" "a" "g" "a" "c"  
## [1945] "g" "a" "a" "g" "g" "c" "a" "t" "g" "a" "g" "t" "t" "t" "c" "t" "c" "c"  
## [1963] "g" "a" "g" "g" "c" "g" "a" "c" "c" "a" "c" "g" "c" "t" "g" "g" "c" "g"  
## [1981] "c" "g" "g" "g" "a" "a" "a" "t" "g" "g" "g" "t" "t" "a" "t" "a" "c" "c"  
## [1999] "g" "a" "a" "c" "c" "g" "g" "a" "c" "c" "c" "g" "c" "g" "a" "g" "a" "t"  
## [2017] "g" "a" "t" "c" "t" "t" "t" "c" "t" "g" "g" "t" "a" "t" "g" "g" "a" "t"  
## [2035] "g" "t" "g" "g" "c" "g" "c" "g" "t" "a" "a" "a" "c" "t" "a" "t" "t" "g"  
## [2053] "a" "t" "t" "c" "t" "c" "g" "c" "t" "c" "g" "t" "g" "a" "a" "a" "c" "g"  
## [2071] "g" "g" "a" "c" "g" "t" "g" "a" "a" "c" "t" "g" "g" "a" "g" "c" "t" "g"  
## [2089] "g" "c" "g" "g" "a" "t" "a" "t" "t" "g" "a" "a" "a" "t" "t" "g" "a" "a"  
## [2107] "c" "c" "t" "g" "t" "g" "c" "t" "g" "c" "c" "c" "g" "c" "a" "g" "a" "g"  
## [2125] "t" "t" "t" "a" "a" "c" "g" "c" "c" "g" "a" "g" "g" "g" "t" "g" "a" "t"  
## [2143] "g" "t" "t" "g" "c" "c" "g" "c" "t" "t" "t" "t" "a" "t" "g" "g" "c" "g"  
## [2161] "a" "a" "t" "c" "t" "g" "t" "c" "a" "c" "a" "a" "c" "t" "c" "g" "a" "c"  
## [2179] "g" "a" "t" "c" "t" "c" "t" "t" "t" "g" "c" "c" "g" "c" "g" "c" "g" "c"  
## [2197] "g" "t" "g" "g" "c" "g" "a" "a" "g" "g" "c" "c" "c" "g" "t" "g" "a" "t"  
## [2215] "g" "a" "a" "g" "g" "a" "a" "a" "a" "g" "t" "t" "t" "t" "g" "c" "g" "c"  
## [2233] "t" "a" "t" "g" "t" "t" "g" "g" "c" "a" "a" "t" "a" "t" "t" "g" "a" "t"  
## [2251] "g" "a" "a" "g" "a" "t" "g" "g" "c" "g" "t" "c" "t" "g" "c" "c" "g" "c"  
## [2269] "g" "t" "g" "a" "a" "g" "a" "t" "t" "g" "c" "c" "g" "a" "a" "g" "t" "g"  
## [2287] "g" "a" "t" "g" "g" "t" "a" "a" "t" "g" "a" "t" "c" "c" "g" "c" "t" "g"  
## [2305] "t" "t" "c" "a" "a" "a" "g" "t" "g" "a" "a" "a" "a" "a" "t" "g" "g" "c"  
## [2323] "g" "a" "a" "a" "a" "c" "g" "c" "c" "c" "t" "g" "g" "c" "c" "t" "t" "c"  
## [2341] "t" "a" "t" "a" "g" "c" "c" "a" "c" "t" "a" "t" "t" "a" "t" "c" "a" "g"  
## [2359] "c" "c" "g" "c" "t" "g" "c" "c" "g" "t" "t" "g" "g" "t" "a" "c" "t" "g"  
## [2377] "c" "g" "c" "g" "g" "a" "t" "a" "t" "g" "g" "t" "g" "c" "g" "g" "g" "c"  
## [2395] "a" "a" "t" "g" "a" "c" "g" "t" "t" "a" "c" "a" "g" "c" "t" "g" "c" "c"  
## [2413] "g" "g" "t" "g" "t" "c" "t" "t" "t" "g" "c" "t" "g" "a" "t" "c" "t" "g"  
## [2431] "c" "t" "a" "c" "g" "t" "a" "c" "c" "c" "t" "c" "t" "c" "a" "t" "g" "g"  
## [2449] "a" "a" "g" "t" "t" "a" "g" "g" "a" "g" "t" "c" "t" "g" "a"  
## attr(,"name")  
## [1] "AAC73113"  
## attr(,"Annot")  
## [1] ">AAC73113 cds chromosome:ASM584v2:Chromosome:337:2799:1 gene:b0002 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding gene\_symbol:thrA description:Bifunctional aspartokinase/homoserine dehydrogenase 1"  
## attr(,"class")  
## [1] "SeqFastadna"  
##   
## $AAC73114  
## [1] "a" "t" "g" "g" "t" "t" "a" "a" "a" "g" "t" "t" "t" "a" "t" "g" "c" "c"  
## [19] "c" "c" "g" "g" "c" "t" "t" "c" "c" "a" "g" "t" "g" "c" "c" "a" "a" "t"  
## [37] "a" "t" "g" "a" "g" "c" "g" "t" "c" "g" "g" "g" "t" "t" "t" "g" "a" "t"  
## [55] "g" "t" "g" "c" "t" "c" "g" "g" "g" "g" "c" "g" "g" "c" "g" "g" "t" "g"  
## [73] "a" "c" "a" "c" "c" "t" "g" "t" "t" "g" "a" "t" "g" "g" "t" "g" "c" "a"  
## [91] "t" "t" "g" "c" "t" "c" "g" "g" "a" "g" "a" "t" "g" "t" "a" "g" "t" "c"  
## [109] "a" "c" "g" "g" "t" "t" "g" "a" "g" "g" "c" "g" "g" "c" "a" "g" "a" "g"  
## [127] "a" "c" "a" "t" "t" "c" "a" "g" "t" "c" "t" "c" "a" "a" "c" "a" "a" "c"  
## [145] "c" "t" "c" "g" "g" "a" "c" "g" "c" "t" "t" "t" "g" "c" "c" "g" "a" "t"  
## [163] "a" "a" "g" "c" "t" "g" "c" "c" "g" "t" "c" "a" "g" "a" "a" "c" "c" "a"  
## [181] "c" "g" "g" "g" "a" "a" "a" "a" "t" "a" "t" "c" "g" "t" "t" "t" "a" "t"  
## [199] "c" "a" "g" "t" "g" "c" "t" "g" "g" "g" "a" "g" "c" "g" "t" "t" "t" "t"  
## [217] "t" "g" "c" "c" "a" "g" "g" "a" "a" "c" "t" "g" "g" "g" "t" "a" "a" "g"  
## [235] "c" "a" "a" "a" "t" "t" "c" "c" "a" "g" "t" "g" "g" "c" "g" "a" "t" "g"  
## [253] "a" "c" "c" "c" "t" "g" "g" "a" "a" "a" "a" "g" "a" "a" "t" "a" "t" "g"  
## [271] "c" "c" "g" "a" "t" "c" "g" "g" "t" "t" "c" "g" "g" "g" "c" "t" "t" "a"  
## [289] "g" "g" "c" "t" "c" "c" "a" "g" "t" "g" "c" "c" "t" "g" "t" "t" "c" "g"  
## [307] "g" "t" "g" "g" "t" "c" "g" "c" "g" "g" "c" "g" "c" "t" "g" "a" "t" "g"  
## [325] "g" "c" "g" "a" "t" "g" "a" "a" "t" "g" "a" "a" "c" "a" "c" "t" "g" "c"  
## [343] "g" "g" "c" "a" "a" "g" "c" "c" "g" "c" "t" "t" "a" "a" "t" "g" "a" "c"  
## [361] "a" "c" "t" "c" "g" "t" "t" "t" "g" "c" "t" "g" "g" "c" "t" "t" "t" "g"  
## [379] "a" "t" "g" "g" "g" "c" "g" "a" "g" "c" "t" "g" "g" "a" "a" "g" "g" "c"  
## [397] "c" "g" "t" "a" "t" "c" "t" "c" "c" "g" "g" "c" "a" "g" "c" "a" "t" "t"  
## [415] "c" "a" "t" "t" "a" "c" "g" "a" "c" "a" "a" "c" "g" "t" "g" "g" "c" "a"  
## [433] "c" "c" "g" "t" "g" "t" "t" "t" "t" "c" "t" "c" "g" "g" "t" "g" "g" "t"  
## [451] "a" "t" "g" "c" "a" "g" "t" "t" "g" "a" "t" "g" "a" "t" "c" "g" "a" "a"  
## [469] "g" "a" "a" "a" "a" "c" "g" "a" "c" "a" "t" "c" "a" "t" "c" "a" "g" "c"  
## [487] "c" "a" "g" "c" "a" "a" "g" "t" "g" "c" "c" "a" "g" "g" "g" "t" "t" "t"  
## [505] "g" "a" "t" "g" "a" "g" "t" "g" "g" "c" "t" "g" "t" "g" "g" "g" "t" "g"  
## [523] "c" "t" "g" "g" "c" "g" "t" "a" "t" "c" "c" "g" "g" "g" "g" "a" "t" "t"  
## [541] "a" "a" "a" "g" "t" "c" "t" "c" "g" "a" "c" "g" "g" "c" "a" "g" "a" "a"  
## [559] "g" "c" "c" "a" "g" "g" "g" "c" "t" "a" "t" "t" "t" "t" "a" "c" "c" "g"  
## [577] "g" "c" "g" "c" "a" "g" "t" "a" "t" "c" "g" "c" "c" "g" "c" "c" "a" "g"  
## [595] "g" "a" "t" "t" "g" "c" "a" "t" "t" "g" "c" "g" "c" "a" "c" "g" "g" "g"  
## [613] "c" "g" "a" "c" "a" "t" "c" "t" "g" "g" "c" "a" "g" "g" "c" "t" "t" "c"  
## [631] "a" "t" "t" "c" "a" "c" "g" "c" "c" "t" "g" "c" "t" "a" "t" "t" "c" "c"  
## [649] "c" "g" "t" "c" "a" "g" "c" "c" "t" "g" "a" "g" "c" "t" "t" "g" "c" "c"  
## [667] "g" "c" "g" "a" "a" "g" "c" "t" "g" "a" "t" "g" "a" "a" "a" "g" "a" "t"  
## [685] "g" "t" "t" "a" "t" "c" "g" "c" "t" "g" "a" "a" "c" "c" "c" "t" "a" "c"  
## [703] "c" "g" "t" "g" "a" "a" "c" "g" "g" "t" "t" "a" "c" "t" "g" "c" "c" "a"  
## [721] "g" "g" "c" "t" "t" "c" "c" "g" "g" "c" "a" "g" "g" "c" "g" "c" "g" "g"  
## [739] "c" "a" "g" "g" "c" "g" "g" "t" "c" "g" "c" "g" "g" "a" "a" "a" "t" "c"  
## [757] "g" "g" "c" "g" "c" "g" "g" "t" "a" "g" "c" "g" "a" "g" "c" "g" "g" "t"  
## [775] "a" "t" "c" "t" "c" "c" "g" "g" "c" "t" "c" "c" "g" "g" "c" "c" "c" "g"  
## [793] "a" "c" "c" "t" "t" "g" "t" "t" "c" "g" "c" "t" "c" "t" "g" "t" "g" "t"  
## [811] "g" "a" "c" "a" "a" "g" "c" "c" "g" "g" "a" "a" "a" "c" "c" "g" "c" "c"  
## [829] "c" "a" "g" "c" "g" "c" "g" "t" "t" "g" "c" "c" "g" "a" "c" "t" "g" "g"  
## [847] "t" "t" "g" "g" "g" "t" "a" "a" "g" "a" "a" "c" "t" "a" "c" "c" "t" "g"  
## [865] "c" "a" "a" "a" "a" "t" "c" "a" "g" "g" "a" "a" "g" "g" "t" "t" "t" "t"  
## [883] "g" "t" "t" "c" "a" "t" "a" "t" "t" "t" "g" "c" "c" "g" "g" "c" "t" "g"  
## [901] "g" "a" "t" "a" "c" "g" "g" "c" "g" "g" "g" "c" "g" "c" "a" "c" "g" "a"  
## [919] "g" "t" "a" "c" "t" "g" "g" "a" "a" "a" "a" "c" "t" "a" "a"  
## attr(,"name")  
## [1] "AAC73114"  
## attr(,"Annot")  
## [1] ">AAC73114 cds chromosome:ASM584v2:Chromosome:2801:3733:1 gene:b0003 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding gene\_symbol:thrB description:homoserine kinase"  
## attr(,"class")  
## [1] "SeqFastadna"  
##   
## $AAC73115  
## [1] "a" "t" "g" "a" "a" "a" "c" "t" "c" "t" "a" "c" "a" "a" "t" "c" "t" "g"  
## [19] "a" "a" "a" "g" "a" "t" "c" "a" "c" "a" "a" "c" "g" "a" "g" "c" "a" "g"  
## [37] "g" "t" "c" "a" "g" "c" "t" "t" "t" "g" "c" "g" "c" "a" "a" "g" "c" "c"  
## [55] "g" "t" "a" "a" "c" "c" "c" "a" "g" "g" "g" "g" "t" "t" "g" "g" "g" "c"  
## [73] "a" "a" "a" "a" "a" "t" "c" "a" "g" "g" "g" "g" "c" "t" "g" "t" "t" "t"  
## [91] "t" "t" "t" "c" "c" "g" "c" "a" "c" "g" "a" "c" "c" "t" "g" "c" "c" "g"  
## [109] "g" "a" "a" "t" "t" "c" "a" "g" "c" "c" "t" "g" "a" "c" "t" "g" "a" "a"  
## [127] "a" "t" "t" "g" "a" "t" "g" "a" "g" "a" "t" "g" "c" "t" "g" "a" "a" "g"  
## [145] "c" "t" "g" "g" "a" "t" "t" "t" "t" "g" "t" "c" "a" "c" "c" "c" "g" "c"  
## [163] "a" "g" "t" "g" "c" "g" "a" "a" "g" "a" "t" "c" "c" "t" "c" "t" "c" "g"  
## [181] "g" "c" "g" "t" "t" "t" "a" "t" "t" "g" "g" "t" "g" "a" "t" "g" "a" "a"  
## [199] "a" "t" "c" "c" "c" "a" "c" "a" "g" "g" "a" "a" "a" "t" "c" "c" "t" "g"  
## [217] "g" "a" "a" "g" "a" "g" "c" "g" "c" "g" "t" "g" "c" "g" "c" "g" "c" "g"  
## [235] "g" "c" "g" "t" "t" "t" "g" "c" "c" "t" "t" "c" "c" "c" "g" "g" "c" "t"  
## [253] "c" "c" "g" "g" "t" "c" "g" "c" "c" "a" "a" "t" "g" "t" "t" "g" "a" "a"  
## [271] "a" "g" "c" "g" "a" "t" "g" "t" "c" "g" "g" "t" "t" "g" "t" "c" "t" "g"  
## [289] "g" "a" "a" "t" "t" "g" "t" "t" "c" "c" "a" "c" "g" "g" "g" "c" "c" "a"  
## [307] "a" "c" "g" "c" "t" "g" "g" "c" "a" "t" "t" "t" "a" "a" "a" "g" "a" "t"  
## [325] "t" "t" "c" "g" "g" "c" "g" "g" "t" "c" "g" "c" "t" "t" "t" "a" "t" "g"  
## [343] "g" "c" "a" "c" "a" "a" "a" "t" "g" "c" "t" "g" "a" "c" "c" "c" "a" "t"  
## [361] "a" "t" "t" "g" "c" "g" "g" "g" "t" "g" "a" "t" "a" "a" "g" "c" "c" "a"  
## [379] "g" "t" "g" "a" "c" "c" "a" "t" "t" "c" "t" "g" "a" "c" "c" "g" "c" "g"  
## [397] "a" "c" "c" "t" "c" "c" "g" "g" "t" "g" "a" "t" "a" "c" "c" "g" "g" "a"  
## [415] "g" "c" "g" "g" "c" "a" "g" "t" "g" "g" "c" "t" "c" "a" "t" "g" "c" "t"  
## [433] "t" "t" "c" "t" "a" "c" "g" "g" "t" "t" "t" "a" "c" "c" "g" "a" "a" "t"  
## [451] "g" "t" "g" "a" "a" "a" "g" "t" "g" "g" "t" "t" "a" "t" "c" "c" "t" "c"  
## [469] "t" "a" "t" "c" "c" "a" "c" "g" "a" "g" "g" "c" "a" "a" "a" "a" "t" "c"  
## [487] "a" "g" "t" "c" "c" "a" "c" "t" "g" "c" "a" "a" "g" "a" "a" "a" "a" "a"  
## [505] "c" "t" "g" "t" "t" "c" "t" "g" "t" "a" "c" "a" "t" "t" "g" "g" "g" "c"  
## [523] "g" "g" "c" "a" "a" "t" "a" "t" "c" "g" "a" "a" "a" "c" "t" "g" "t" "t"  
## [541] "g" "c" "c" "a" "t" "c" "g" "a" "c" "g" "g" "c" "g" "a" "t" "t" "t" "c"  
## [559] "g" "a" "t" "g" "c" "c" "t" "g" "t" "c" "a" "g" "g" "c" "g" "c" "t" "g"  
## [577] "g" "t" "g" "a" "a" "g" "c" "a" "g" "g" "c" "g" "t" "t" "t" "g" "a" "t"  
## [595] "g" "a" "t" "g" "a" "a" "g" "a" "a" "c" "t" "g" "a" "a" "a" "g" "t" "g"  
## [613] "g" "c" "g" "c" "t" "a" "g" "g" "g" "t" "t" "a" "a" "a" "c" "t" "c" "g"  
## [631] "g" "c" "t" "a" "a" "c" "t" "c" "g" "a" "t" "t" "a" "a" "c" "a" "t" "c"  
## [649] "a" "g" "c" "c" "g" "t" "t" "t" "g" "c" "t" "g" "g" "c" "g" "c" "a" "g"  
## [667] "a" "t" "t" "t" "g" "c" "t" "a" "c" "t" "a" "c" "t" "t" "t" "g" "a" "a"  
## [685] "g" "c" "t" "g" "t" "t" "g" "c" "g" "c" "a" "g" "c" "t" "g" "c" "c" "g"  
## [703] "c" "a" "g" "g" "a" "g" "a" "c" "g" "c" "g" "c" "a" "a" "c" "c" "a" "g"  
## [721] "c" "t" "g" "g" "t" "t" "g" "t" "c" "t" "c" "g" "g" "t" "g" "c" "c" "a"  
## [739] "a" "g" "c" "g" "g" "a" "a" "a" "c" "t" "t" "c" "g" "g" "c" "g" "a" "t"  
## [757] "t" "t" "g" "a" "c" "g" "g" "c" "g" "g" "g" "t" "c" "t" "g" "c" "t" "g"  
## [775] "g" "c" "g" "a" "a" "g" "t" "c" "a" "c" "t" "c" "g" "g" "t" "c" "t" "g"  
## [793] "c" "c" "g" "g" "t" "g" "a" "a" "a" "c" "g" "t" "t" "t" "t" "a" "t" "t"  
## [811] "g" "c" "t" "g" "c" "g" "a" "c" "c" "a" "a" "c" "g" "t" "g" "a" "a" "c"  
## [829] "g" "a" "t" "a" "c" "c" "g" "t" "g" "c" "c" "a" "c" "g" "t" "t" "t" "c"  
## [847] "c" "t" "g" "c" "a" "c" "g" "a" "c" "g" "g" "t" "c" "a" "g" "t" "g" "g"  
## [865] "t" "c" "a" "c" "c" "c" "a" "a" "a" "g" "c" "g" "a" "c" "t" "c" "a" "g"  
## [883] "g" "c" "g" "a" "c" "g" "t" "t" "a" "t" "c" "c" "a" "a" "c" "g" "c" "g"  
## [901] "a" "t" "g" "g" "a" "c" "g" "t" "g" "a" "g" "t" "c" "a" "g" "c" "c" "g"  
## [919] "a" "a" "c" "a" "a" "c" "t" "g" "g" "c" "c" "g" "c" "g" "t" "g" "t" "g"  
## [937] "g" "a" "a" "g" "a" "g" "t" "t" "g" "t" "t" "c" "c" "g" "c" "c" "g" "c"  
## [955] "a" "a" "a" "a" "t" "c" "t" "g" "g" "c" "a" "a" "c" "t" "g" "a" "a" "a"  
## [973] "g" "a" "g" "c" "t" "g" "g" "g" "t" "t" "a" "t" "g" "c" "a" "g" "c" "c"  
## [991] "g" "t" "g" "g" "a" "t" "g" "a" "t" "g" "a" "a" "a" "c" "c" "a" "c" "g"  
## [1009] "c" "a" "a" "c" "a" "g" "a" "c" "a" "a" "t" "g" "c" "g" "t" "g" "a" "g"  
## [1027] "t" "t" "a" "a" "a" "a" "g" "a" "a" "c" "t" "g" "g" "g" "c" "t" "a" "c"  
## [1045] "a" "c" "t" "t" "c" "g" "g" "a" "g" "c" "c" "g" "c" "a" "c" "g" "c" "t"  
## [1063] "g" "c" "c" "g" "t" "a" "g" "c" "t" "t" "a" "t" "c" "g" "t" "g" "c" "g"  
## [1081] "c" "t" "g" "c" "g" "t" "g" "a" "t" "c" "a" "g" "t" "t" "g" "a" "a" "t"  
## [1099] "c" "c" "a" "g" "g" "c" "g" "a" "a" "t" "a" "t" "g" "g" "c" "t" "t" "g"  
## [1117] "t" "t" "c" "c" "t" "c" "g" "g" "c" "a" "c" "c" "g" "c" "g" "c" "a" "t"  
## [1135] "c" "c" "g" "g" "c" "g" "a" "a" "a" "t" "t" "t" "a" "a" "a" "g" "a" "g"  
## [1153] "a" "g" "c" "g" "t" "g" "g" "a" "a" "g" "c" "g" "a" "t" "t" "c" "t" "c"  
## [1171] "g" "g" "t" "g" "a" "a" "a" "c" "g" "t" "t" "g" "g" "a" "t" "c" "t" "g"  
## [1189] "c" "c" "a" "a" "a" "a" "g" "a" "g" "c" "t" "g" "g" "c" "a" "g" "a" "a"  
## [1207] "c" "g" "t" "g" "c" "t" "g" "a" "t" "t" "t" "a" "c" "c" "c" "t" "t" "g"  
## [1225] "c" "t" "t" "t" "c" "a" "c" "a" "t" "a" "a" "t" "c" "t" "g" "c" "c" "c"  
## [1243] "g" "c" "c" "g" "a" "t" "t" "t" "t" "g" "c" "t" "g" "c" "g" "t" "t" "g"  
## [1261] "c" "g" "t" "a" "a" "a" "t" "t" "g" "a" "t" "g" "a" "t" "g" "a" "a" "t"  
## [1279] "c" "a" "t" "c" "a" "g" "t" "a" "a"  
## attr(,"name")  
## [1] "AAC73115"  
## attr(,"Annot")  
## [1] ">AAC73115 cds chromosome:ASM584v2:Chromosome:3734:5020:1 gene:b0004 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding gene\_symbol:thrC description:L-threonine synthase"  
## attr(,"class")  
## [1] "SeqFastadna"  
##   
## $AAC73116  
## [1] "g" "t" "g" "a" "a" "a" "a" "a" "g" "a" "t" "g" "c" "a" "a" "t" "c" "t"  
## [19] "a" "t" "c" "g" "t" "a" "c" "t" "c" "g" "c" "a" "c" "t" "t" "t" "c" "c"  
## [37] "c" "t" "g" "g" "t" "t" "c" "t" "g" "g" "t" "c" "g" "c" "t" "c" "c" "c"  
## [55] "a" "t" "g" "g" "c" "a" "g" "c" "a" "c" "a" "g" "g" "c" "t" "g" "c" "g"  
## [73] "g" "a" "a" "a" "t" "t" "a" "c" "g" "t" "t" "a" "g" "t" "c" "c" "c" "g"  
## [91] "t" "c" "a" "g" "t" "a" "a" "a" "a" "t" "t" "a" "c" "a" "g" "a" "t" "a"  
## [109] "g" "g" "c" "g" "a" "t" "c" "g" "t" "g" "a" "t" "a" "a" "t" "c" "g" "t"  
## [127] "g" "g" "c" "t" "a" "t" "t" "a" "c" "t" "g" "g" "g" "a" "t" "g" "g" "a"  
## [145] "g" "g" "t" "c" "a" "c" "t" "g" "g" "c" "g" "c" "g" "a" "c" "c" "a" "c"  
## [163] "g" "g" "c" "t" "g" "g" "t" "g" "g" "a" "a" "a" "c" "a" "a" "c" "a" "t"  
## [181] "t" "a" "t" "g" "a" "a" "t" "g" "g" "c" "g" "a" "g" "g" "c" "a" "a" "t"  
## [199] "c" "g" "c" "t" "g" "g" "c" "a" "c" "c" "t" "a" "c" "a" "c" "g" "g" "a"  
## [217] "c" "c" "g" "c" "c" "g" "c" "c" "a" "c" "c" "g" "c" "c" "g" "c" "g" "c"  
## [235] "c" "a" "c" "c" "a" "t" "a" "a" "g" "a" "a" "a" "g" "c" "t" "c" "c" "t"  
## [253] "c" "a" "t" "g" "a" "t" "c" "a" "t" "c" "a" "c" "g" "g" "c" "g" "g" "t"  
## [271] "c" "a" "t" "g" "g" "t" "c" "c" "a" "g" "g" "c" "a" "a" "a" "c" "a" "t"  
## [289] "c" "a" "c" "c" "g" "c" "t" "a" "a"  
## attr(,"name")  
## [1] "AAC73116"  
## attr(,"Annot")  
## [1] ">AAC73116 cds chromosome:ASM584v2:Chromosome:5234:5530:1 gene:b0005 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding gene\_symbol:yaaX description:DUF2502 family putative periplasmic protein"  
## attr(,"class")  
## [1] "SeqFastadna"  
##   
## $AAC73117  
## [1] "a" "t" "g" "c" "t" "g" "a" "t" "t" "c" "t" "t" "a" "t" "t" "t" "c" "a"  
## [19] "c" "c" "t" "g" "c" "g" "a" "a" "a" "a" "c" "g" "c" "t" "t" "g" "a" "t"  
## [37] "t" "a" "c" "c" "a" "a" "a" "g" "c" "c" "c" "g" "t" "t" "g" "a" "c" "c"  
## [55] "a" "c" "c" "a" "c" "g" "c" "g" "c" "t" "a" "t" "a" "c" "g" "c" "t" "g"  
## [73] "c" "c" "g" "g" "a" "g" "c" "t" "g" "t" "t" "a" "g" "a" "c" "a" "a" "t"  
## [91] "t" "c" "c" "c" "a" "g" "c" "a" "g" "t" "t" "g" "a" "t" "c" "c" "a" "t"  
## [109] "g" "a" "g" "g" "c" "g" "c" "g" "g" "a" "a" "a" "c" "t" "g" "a" "c" "g"  
## [127] "c" "c" "t" "c" "c" "g" "c" "a" "g" "a" "t" "t" "a" "g" "c" "a" "c" "g"  
## [145] "c" "t" "g" "a" "t" "g" "c" "g" "c" "a" "t" "c" "a" "g" "c" "g" "a" "c"  
## [163] "a" "a" "a" "c" "t" "g" "g" "c" "g" "g" "g" "t" "a" "t" "c" "a" "a" "c"  
## [181] "g" "c" "c" "g" "c" "t" "c" "g" "c" "t" "t" "t" "c" "a" "t" "g" "a" "c"  
## [199] "t" "g" "g" "c" "a" "g" "c" "c" "a" "g" "a" "t" "t" "t" "c" "a" "c" "g"  
## [217] "c" "c" "g" "g" "c" "g" "a" "a" "t" "g" "c" "c" "c" "g" "c" "c" "a" "g"  
## [235] "g" "c" "g" "a" "t" "t" "c" "t" "g" "g" "c" "g" "t" "t" "t" "a" "a" "a"  
## [253] "g" "g" "t" "g" "a" "t" "g" "t" "c" "t" "a" "c" "a" "c" "c" "g" "g" "c"  
## [271] "t" "t" "g" "c" "a" "g" "g" "c" "c" "g" "a" "a" "a" "c" "c" "t" "t" "c"  
## [289] "a" "g" "c" "g" "a" "a" "g" "a" "c" "g" "a" "t" "t" "t" "c" "g" "a" "t"  
## [307] "t" "t" "t" "g" "c" "c" "c" "a" "a" "c" "a" "g" "c" "a" "t" "t" "t" "g"  
## [325] "c" "g" "a" "a" "t" "g" "c" "t" "t" "t" "c" "c" "g" "g" "c" "t" "t" "g"  
## [343] "t" "a" "t" "g" "g" "c" "g" "t" "a" "c" "t" "c" "c" "g" "c" "c" "c" "g"  
## [361] "c" "t" "c" "g" "a" "t" "t" "t" "a" "a" "t" "g" "c" "a" "g" "c" "c" "t"  
## [379] "t" "a" "t" "c" "g" "t" "c" "t" "g" "g" "a" "a" "a" "t" "g" "g" "g" "g"  
## [397] "a" "t" "c" "c" "g" "t" "c" "t" "t" "g" "a" "g" "a" "a" "t" "g" "c" "c"  
## [415] "c" "g" "a" "g" "g" "g" "a" "a" "a" "g" "a" "t" "c" "t" "g" "t" "a" "t"  
## [433] "c" "a" "a" "t" "t" "c" "t" "g" "g" "g" "g" "a" "g" "a" "t" "a" "t" "c"  
## [451] "a" "t" "c" "a" "c" "c" "a" "a" "c" "a" "a" "g" "c" "t" "g" "a" "a" "c"  
## [469] "g" "a" "g" "g" "c" "g" "c" "t" "c" "g" "c" "a" "g" "c" "a" "c" "a" "a"  
## [487] "g" "g" "c" "g" "a" "t" "a" "a" "t" "g" "t" "g" "g" "t" "g" "a" "t" "t"  
## [505] "a" "a" "t" "c" "t" "g" "g" "c" "g" "t" "c" "a" "g" "a" "t" "g" "a" "g"  
## [523] "t" "a" "t" "t" "t" "t" "a" "a" "a" "t" "c" "g" "g" "t" "g" "a" "a" "g"  
## [541] "c" "c" "g" "a" "a" "g" "a" "a" "a" "t" "t" "g" "a" "a" "t" "g" "c" "c"  
## [559] "g" "a" "g" "a" "t" "t" "a" "t" "c" "a" "a" "g" "c" "c" "g" "g" "t" "g"  
## [577] "t" "t" "c" "c" "t" "c" "g" "a" "t" "g" "a" "g" "a" "a" "g" "a" "a" "c"  
## [595] "g" "g" "c" "a" "a" "a" "t" "t" "t" "a" "a" "g" "a" "t" "c" "a" "t" "c"  
## [613] "a" "g" "c" "t" "t" "c" "t" "a" "c" "g" "c" "t" "a" "a" "g" "a" "a" "a"  
## [631] "g" "c" "a" "c" "g" "c" "g" "g" "t" "c" "t" "g" "a" "t" "g" "a" "g" "t"  
## [649] "c" "g" "t" "t" "t" "c" "a" "t" "t" "a" "t" "t" "g" "a" "a" "a" "a" "t"  
## [667] "c" "g" "g" "c" "t" "g" "a" "c" "c" "a" "a" "a" "c" "c" "a" "g" "a" "g"  
## [685] "c" "a" "a" "c" "t" "g" "a" "c" "t" "g" "g" "t" "t" "t" "t" "a" "a" "t"  
## [703] "a" "g" "c" "g" "a" "a" "g" "g" "t" "t" "a" "c" "t" "t" "c" "t" "t" "t"  
## [721] "g" "a" "t" "g" "a" "a" "g" "a" "t" "t" "c" "c" "t" "c" "c" "a" "g" "c"  
## [739] "a" "a" "t" "g" "g" "c" "g" "a" "a" "c" "t" "g" "g" "t" "g" "t" "t" "t"  
## [757] "a" "a" "a" "c" "g" "c" "t" "a" "c" "g" "a" "g" "c" "a" "g" "c" "g" "t"  
## [775] "t" "a" "a"  
## attr(,"name")  
## [1] "AAC73117"  
## attr(,"Annot")  
## [1] ">AAC73117 cds chromosome:ASM584v2:Chromosome:5683:6459:-1 gene:b0006 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding gene\_symbol:yaaA description:peroxide resistance protein, lowers intracellular iron"  
## attr(,"class")  
## [1] "SeqFastadna"

# Question 4-Make a set number of point mutations to your sequence of interest and check the number of mismatchesbetween the original and mutated sequence.

In this question, first seqinr package has to be run and it is a Exploratory data analysis and data visualization for biological sequence and then created a blast database by the command makeblast. Thereafter the cat command can be applied to produce the out by converting argument in to a character vectors. Then we have given the command mutator to extract mutation with 30 substitutions.

seqinr::write.fasta(myEcoli,names="myEcoli",file.out = "myEcoli.fa")  
makeblastdb("myEcoli.fa", dbtype = "nucl","-parse\_seqids")  
res <- myblastn(myseq = myEcoli, db= "myEcoli.fa")  
res

## [1] "BLASTN 2.6.0+"   
## [2] ""   
## [3] ""   
## [4] "Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb"   
## [5] "Miller (2000), \"A greedy algorithm for aligning DNA sequences\", J"   
## [6] "Comput Biol 2000; 7(1-2):203-14."   
## [7] ""   
## [8] ""   
## [9] ""   
## [10] "Database: myEcoli.fa"   
## [11] " 1 sequences; 276 total letters"   
## [12] ""   
## [13] ""   
## [14] ""   
## [15] "Query= 56"   
## [16] ""   
## [17] "Length=276"   
## [18] " Score E"   
## [19] "Sequences producing significant alignments: (Bits) Value"   
## [20] ""   
## [21] "myEcoli 531 8e-156"  
## [22] ""   
## [23] ""   
## [24] ""   
## [25] "Query\_1 1 GTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAAC 60"   
## [26] "myEcoli 1 ............................................................ 60"   
## [27] ""   
## [28] "Query\_1 61 GGCAAAAGCACTGCCGGACATCAGCGCTATCTCTGCTCTCACTGCCGTAAAACATGGCAA 120"   
## [29] "myEcoli 61 ............................................................ 120"   
## [30] ""   
## [31] "Query\_1 121 CTGCAGTTCACTTACACCGCTTCTCAACCCGGTACGCACCAGAAAATCATTGATATGGCC 180"   
## [32] "myEcoli 121 ............................................................ 180"   
## [33] ""   
## [34] "Query\_1 181 ATGAATGGCGTTGGATGCCGGGCAACCGCCCGCATTATGGGCGTTGGCCTCAACACGATT 240"   
## [35] "myEcoli 181 ............................................................ 240"   
## [36] ""   
## [37] "Query\_1 241 TTCCGCCATTTAAAAAACTCAGGCCGCAGTCGGTAA 276"   
## [38] "myEcoli 241 .................................... 276"   
## [39] ""   
## [40] ""   
## [41] "Lambda K H"   
## [42] " 1.33 0.621 1.12 "   
## [43] ""   
## [44] ""   
## [45] "Effective search space used: 71289"   
## [46] ""   
## [47] ""   
## [48] " Database: myEcoli.fa"   
## [49] " Posted date: Jun 12, 2020 12:27 PM"   
## [50] " Number of letters in database: 276"   
## [51] " Number of sequences in database: 1"   
## [52] ""   
## [53] ""   
## [54] ""   
## [55] "Matrix: blastn matrix 1 -2"

cat(res,fill=TRUE)

## BLASTN 2.6.0+ Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb   
## Miller (2000), "A greedy algorithm for aligning DNA sequences", J   
## Comput Biol 2000; 7(1-2):203-14. Database: myEcoli.fa   
## 1 sequences; 276 total letters Query= 56 Length=276   
## Score E   
## Sequences producing significant alignments: (Bits) Value   
##   
## myEcoli 531 8e-156   
##   
## Query\_1 1 GTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAAC 60   
## myEcoli 1 ............................................................ 60   
## Query\_1 61 GGCAAAAGCACTGCCGGACATCAGCGCTATCTCTGCTCTCACTGCCGTAAAACATGGCAA 120   
## myEcoli 61 ............................................................ 120   
##   
## Query\_1 121 CTGCAGTTCACTTACACCGCTTCTCAACCCGGTACGCACCAGAAAATCATTGATATGGCC 180   
## myEcoli 121 ............................................................ 180   
##   
## Query\_1 181 ATGAATGGCGTTGGATGCCGGGCAACCGCCCGCATTATGGGCGTTGGCCTCAACACGATT 240   
## myEcoli 181 ............................................................ 240   
## Query\_1 241 TTCCGCCATTTAAAAAACTCAGGCCGCAGTCGGTAA 276   
## myEcoli 241 .................................... 276   
## Lambda K H 1.33 0.621 1.12   
## Effective search space used: 71289 Database: myEcoli.fa   
## Posted date: Jun 12, 2020 12:27 PM Number of letters in database: 276   
## Number of sequences in database: 1 Matrix: blastn matrix 1 -2

###create a mutated copy with 30 substitutions  
myEcolimutator<- mutator(myEcoli,30)  
res <- myblastn\_tab(myseq = myEcolimutator, db= "myEcoli.fa")  
res

## qseqid sseqid pident length mismatch gapopen qstart qend sstart send  
## 1 56 myEcoli 93.841 276 17 0 1 276 1 276  
## evalue bitscore  
## 1 2.62e-126 433

# Question 5-Determine the number and proportion of sites that need to be altered to prevent the BLAST search from matching the gene of origin.

mutator #randomize with mutator

## function (myseq, nmut)   
## {  
## myseq\_mod <- myseq  
## mypos <- sample(seq\_along(myseq), nmut)  
## myseq\_mod[mypos] <- sample(c("a", "c", "g", "t"), length(mypos),   
## replace = TRUE)  
## return(myseq\_mod)  
## }

myblastn\_tab

## function (myseq, db)   
## {  
## mytmpfile1 <- tempfile()  
## mytmpfile2 <- tempfile()  
## write.fasta(myseq, names = attr(myseq, "name"), file.out = mytmpfile1)  
## system2(command = "/usr/bin/blastn", args = paste("-db ",   
## db, " -query", mytmpfile1, "-outfmt 6 -evalue 0.05 -ungapped >",   
## mytmpfile2))  
## res <- NULL  
## if (file.info(mytmpfile2)$size > 0) {  
## res <- read.csv(mytmpfile2, sep = "\t", header = FALSE)  
## colnames(res) <- c("qseqid", "sseqid", "pident", "length",   
## "mismatch", "gapopen", "qstart", "qend", "sstart",   
## "send", "evalue", "bitscore")  
## }  
## unlink(c(mytmpfile1, mytmpfile2))  
## if (!is.null(res)) {  
## res <- res[order(-res$bitscore), ]  
## }  
## res  
## }  
## <bytecode: 0x56283dd9b270>

In this section we have created a function, which is to sumarise and reports risults as 0 or 1 and tested the function by the command myfunc. The commanded number or mutation is 50.

myfunc <- function (myseq,nmut){  
 mutseq <- mutator(myseq=myseq,nmut=nmut)  
 res<-myblastn\_tab(myseq=mutseq, db= "Escherichia\_coli\_str\_k\_12\_substr\_mg1655.ASM584v2.cds.all.fa")  
 if(is.null(res)){myres=0} else {myres=1}  
 return(myres)}  
  
myfunc(myseq=myEcoli,nmut=50)

## [1] 0

In order to acive a reliable answer test must run multiple times, thereby replicate command applied to run the test repeatedly and number of repeated test given as 100. Thereafter the results summarized into decimal numbers in between 0 and 1 to get a proportion the number of sucessful BLASTs. Then the interest number of mutant to be evaluated comanded by n<-c code and gave function to run multiple tests for all the interst values. Finally, by the command finalres tested the BLASTs results.

replicate(n=100,expr = myfunc(myseq=myEcoli,nmut=50))

## [1] 1 1 1 0 1 1 0 0 0 1 0 1 0 1 1 0 0 1 0 0 0 1 0 0 0 0 0 1 1 0 0 1 0 0 0 1 1  
## [38] 0 1 1 0 0 1 0 0 0 0 1 0 1 0 0 1 0 1 0 1 0 0 0 0 1 0 1 0 1 0 1 1 0 1 1 0 0  
## [75] 1 1 0 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1 1 1 0 0 1 1 0

mean (replicate(n=100,expr = myfunc(myseq=myEcoli,nmut=50)))

## [1] 0.41

n<-c(0,10,20,30,40,50,60,70,80,90,100)  
myfunc\_rep <-function (nmut){mean (replicate(n=100,expr = myfunc(myseq=myEcoli,nmut=50)))}  
finalres <-sapply(n,myfunc\_rep)  
finalres

## [1] 0.43 0.45 0.41 0.50 0.41 0.50 0.47 0.45 0.46 0.50 0.48

In this section we have commanded same fuction for the full sequence.

myfunc(myseq=myEcoli,nmut=276)

## [1] 0

replicate(n=100,expr = myfunc(myseq=myEcoli,nmut=276))

## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## [38] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## [75] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

mean (replicate(n=100,expr = myfunc(myseq=myEcoli,nmut=276)))

## [1] 0

n<-c(0,10,20,30,40,50,60,70,80,90,100)  
myfunc\_rep <-function (nmut){mean (replicate(n=100,expr = myfunc(myseq=myEcoli,nmut=276)))}  
finalres <-sapply(n,myfunc\_rep)  
finalres

## [1] 0 0 0 0 0 0 0 0 0 0 0

# Question 6-Chart or table that shows how the increasing proportion of mutated bases reduces the ability for BLAST to match the gene of origin

Bdata <- c(0,10,20,30,40,50,60,70,80,90,100)  
Bdata <- as.data.frame(Bdata)  
Bdata

## Bdata  
## 1 0  
## 2 10  
## 3 20  
## 4 30  
## 5 40  
## 6 50  
## 7 60  
## 8 70  
## 9 80  
## 10 90  
## 11 100

Bdata$repsite<-c(1,1,1,0.95,0.75,0.49,0.25,0.12,0.05,0.02,0.01)  
Bdata

## Bdata repsite  
## 1 0 1.00  
## 2 10 1.00  
## 3 20 1.00  
## 4 30 0.95  
## 5 40 0.75  
## 6 50 0.49  
## 7 60 0.25  
## 8 70 0.12  
## 9 80 0.05  
## 10 90 0.02  
## 11 100 0.01

plot(Bdata, type="b", col="red",col.main="blue", col.sub="green",  
 main="How increasing no.random bases affects BLAST performance",sub="100 repeates,using sequence number 56",  
 xlab="Number of site randomised", ylab="Propotion of successful BLASTs"  
 )

