

```

#!/usr/bin/perl
use strict;
use warnings;
use File::Basename;

#####
# SCRIPT TO QUICKLY BLAST S5_genome_XXX sequence
#####
# e.g RUN ON VITAL-IT (need a database...)
# BEFORE RUNNING THE SCRIPT, TYPE IN TERMINAL:
# module add Blast/ncbi-blast/latest;
# USAGE:
# ./blastSeq.pl cdsfile n sequence_idX
#   -> cdsfile => the file with all CDS to retrieve the sequences of the corresponding id
#   -> n => the number of query locus
#   -> sequence_idX => n sequence ids
# EXAMPLE:
# ./blastSeq.pl Pseud_S5_CDS.fasta 2 S5_genome_4542 S5_genome_900
# OUTPUT
# -> txt file created in current directory for all blast query
# -> print on terminal screen the 5 top results for each query

my $cds = $ARGV[0];
my @query_id = @ARGV[2..($ARGV[1]+1)];
my %all_seq;
my $seq = "";
my $id;
my $query_file = "S5_query.fasta";

# for vital-it:
# do not know why but it doesn't work !!!
# should be type before launching the script
system("module add Blast/ncbi-blast/latest;");

open(CDS, $cds) || die;
while(my $line = <CDS>){
    chomp($line);
    if($line =~ /^>/){

        if(length $seq){           # add to the dict, but do not do that for the 1st iteration
            $all_seq{$id} = $seq;
            $seq = "";
        }

        print ("hello");
        last;
    }
    ($id = $line) =~ s/^>//; # store the id without the > that starts the line

    } else{
        $seq .= $line; # concatenate the sequence parts
    }
}
# do not forget the last sequence
$all_seq{$id} = $seq;
close(CDS);

system("rm -f $query_file"); # because we append, make sure to start from scratch
system("touch $query_file");

foreach my $pos(@query_id){
    # print("$pos\n");
    open(my $queryF, '>>', $query_file) or die("open: $$");
    print($queryF ">$pos\n$all_seq{$pos}\n" );
    # print $x;
}

##### BLAST QUERY
#####
my $command;

```

```

my $blast_output = join("_", @query_id)."_blast.txt";

#$command = "formatdb -i swissprot -p T -o T";
#system($command);

#$command = "blastx -db swiss -query $query_file -out $blast_output";
# -p programm name, -d database, -i query file name
#-max_target_seqs 1

$command = "blastx -db EXPASY/UPKB/UniProtKB -query $query_file -out $blast_output";

system($command);

open(BLAST, $blast_output) || die;
my $k = 0;
my $next;
while(my $line = <BLAST>){
    chomp($line);
    if($line =~ /Query=/){
        print "$line\n";
        $k = 0;          # I want to print the 5 best alignments
    }
    if ($line =~ /Sequences producing/){    # this line will be printed, if not wanted, put at the end
of the loop
        $next = 1;
    }
    if(length $line and $next and $k<6 ){    #line not empty # 5 seq + 1 line Sequences producing...
        print "$line\n";
        $k ++;
    }
    if($k==6){
        $next = 0;
    }
}
close(BLAST);

```