```
#! /usr/bin/perl -w

use Bio::SeqIO;

##############################################################################
#####################
# Additional_file_1.pl
# author: Leandro Costa do Nascimento
# E-mails: leandro@lge.ibi.unicamp.br or l.costa.nascimento@gmail.com

# Article: A web-based bioinformatics interface applied to Genosoja Project:
databases and pipelines
# Nascimento et al., 2011
# Bioinformatics - Genomics and Expression Laboratory (LGE)
http://www.lge.ibi.unicamp.br
# GENOSOJA database: http://www.lge.ibi.unicamp.br/soja

# Usage: perl Additional_file_1.pl <folder_with_the_files> <gbk_file>
<output_fasta>
# Bugs: Probably many! =D
##############################################################################
#####################

### Global variables - Don't edit
#######################################################################
my $accession = "";
my $cultivar = "";
my $tissue = "";
my $cult = 0;
my $tis = 0;
my $conf = 0;
my $confirma = 0;
my $sequencia = "";
my $cria = 0;

my $dir = "";
my $gbk = "";
my $new = "";
my %cont;
##############################################################################
##########################

### Parameters section
##############################################################################
####
sub show_parameters{
   print "Usage: perl Additional_file_1.pl <folder_with_the_files> <gbk_file>
<output_fasta>\n\n";
   exit(0);
}

($dir, $gbk, $new) = @ARGV;

if(!(defined($dir))){
   show_parameters();
}

if(!(defined($gbk))){
   show_parameters();
}

if(!(defined($new))){
   show_parameters();
}
##############################################################################
##########################

open SEQ, "<$dir/$gbk";
```

```
while(<SEQ>){
   chomp;
   my $linha = $_;

   if(/ACCESSION\s+(.*)/){
      $accession = $1;
      $cultivar = "";
      $tissue = "";
      $sequencia = "";
      $cria = 0;
      $confirma = 0;
      $cult = 0;
      $tis = 0;
   }

   if(/^\s+\/cultivar\=\"([^\"]+)\"/){
      $cultivar = $1;
      $cultivar =~ s/\r//g;
      $cultivar =~ s/\n//g;

      if($cultivar ne ""){
         $cult = 1;
      }
   }

   if(/^\s+\/tissue\_type\=\"([^\"]+)\"/){
      $tissue = $1;
      $tissue =~ s/\r//g;
      $tissue =~ s/\n//g;

      if($tissue ne ""){
         $tis = 1;
      }
   }

   if(/^\/\//){
      $conf = 0;
      $cria = 1;
   }

   if($conf == 1){
      $linha =~ s/\s//g;
      $linha =~ s/\d//g;
      $linha =~ s/\n//g;
      $linha =~ s/\r//g;
      $sequencia = $sequencia . $linha;
   }

   if(/^ORIGIN\s+$/){
      $conf = 1;
   }

   if(($cria == 1) && ($cult == 1) && ($tis == 1)){
      open NEW, ">>$new";
         print NEW ">$accession cultivar:$cultivar tissue:$tissue\n";
         print NEW "$sequencia\n";
      close NEW;

      $accession = "";
      $cultivar = "";
      $tissue = "";
      $sequencia = "";
      $cria = 0;
      $confirma = 0;
      $cult = 0;
      $tis = 0;
   }
```

```
    }
close SEQ;
```