```
#!/usr/bin/perl
use LWP::Simple;

#It takes an array of EST AccNum, parse the EST pages, and produce a text
#file that is formatted so it can be directly loaded into MySQL database.

open (ACCFILE, "nrsortedacclist.txt");
@AccNum = <ACCFILE>;
close ACCFILE;

#sub dbestparser {

  $estindex = 0;

   foreach (@AccNum)
{
  $estpage =
get("http://www.ncbi.nlm.nih.gov:80/entrez/viewer.fcgi?save=0&cmd=&view=e
st&txt=0&val=$_");

      my $j = $estindex;

      @line = split(/\n/, $estpage);

      my $linenum = 0;

#read through each EST file line by line.
      foreach (@line)
      {
       my $keyword = substr($_, 0, 16);
       $keyword =~ s/\s{2,}//g;
         $linenum++;

       $content = substr($_, 16);

#parse dbEST id.

         if ($keyword eq "dbEST Id:")
      {
          $content =~ /(\d*)/;
        $estid[$j] = $1;
        next;
      }

#parse EST name.

       if ($keyword eq "EST name:")
       {
         $content =~ /([\d\w]*)/;
         $estname[$j] = $1;
         next;
       }

#parse GenBank Acc.
```

```
       if ($keyword eq "GenBank Acc:")
        {
          $content =~ /([\d\w]*)/;
          $gbAcc[$j] = $1;
          next;
        }

#parse GenBank ID.

        if ($keyword eq "GenBank gi:")
        {
          $content =~ /(\d*)/;
          $gbid[$j] = $1;
          next;
        }

#parse GDB ID.

          if ($keyword eq "GDB Id:")
          {
            $content =~ /(\d*)/;
            $gdbid[$j] = $1;
            next;
          }


#parse Clone ID.

        if ($keyword eq "Clone Id:")
        {
          $content =~ /(.*)/;
          $cloneid[$j] = $1;

          $content =~ /(\w*):/;
            $dborigin[$j] = $1;
          next;
        }

#parse Insert length.

          if ($keyword eq "Insert length:")
          {
            $content =~ /(\d*)/;
            $insert[$j] = $1;
            next;
          }

#parse EST sequence.

        if ($keyword eq "SEQUENCE")
        {
          my $i = $linenum;
```

```
          while ($line[$i])
          {
              $seqkeyword = substr($line[$i], 0, 16);
            $seqkeyword =~ s/\s+//g;
            $seqcontent = substr($line[$i], 16);

            if ($seqkeyword eq "")
             {
               $seq[$j] .= $seqcontent;
             }
            else
             {
               $seq[$j] =~ s/\s+//g;
             last;
             }
             $i++;
          }
        }

#parse dbEST putative ID.

          if ($keyword eq "PUTATIVE ID")
          {
            my $i = $linenum;

            while ($line[$i])
            {
            $putkeyword = substr($line[$i], 0, 16);
            $putkeyword =~ s/\s+//g;
            $putcontent = substr($line[$i], 16);

              if ($putkeyword eq "")
            {
              $putid[$j] .= $putcontent;
            }
              else
            {
              $putid[$j] =~ s/\n/\s/g;
              last;
            }
            $i++;
             }
          }

#parse dbEST library ID.

        if ($keyword eq "dbEST lib id:")
        {
           $content =~ /(\d*)/;
           $estlibid[$j] = $1;
           next;
        }

#parse dbEST library name.
```

```
          if ($keyword eq "Lib Name:")
          {
             $content =~ /(.*)/;
             $estlibname[$j] = $1;
             next;
          }

      }

$estindex++;
}

#write a text file to load data into database.

open (EST, ">estdata") || die "Can't create estdata!!";

my $totalest = @estid;
for (my $i = 0; $i < $totalest; $i++)
{
   print EST
"FED$i\t$estid[$i]\t$estname[$i]\t$gbAcc[$i]\t$gbid[$i]\t$gdbid[$i]\t$clo
neid[$i]\t$insert[$i]\t$seq[$i]\t$putid[$i]\t$estlibid[$i]\t$estlibname[$
i]\n";
}

close EST;

#}
```