

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

```
#decided to have input file entered in command line
```

```
#call program followed by genome name.
```

```
#the program assumes that a file with the extensions ptt and faa exist in the same directory.
```

```
#####INPUT Name of multiple seq file containing ORF of genome, open file and assign IN filehandle #####
```

```
unless(@ARGV==1) {die "please provide genome name in command line \n
```

```
file should contain multiple sequences in fasta format \n
```

```
a file with the ptt table should be in the same directory\n\n";}
```

```
#$num=0;
```

```
$filename=$ARGV[0];
```

```
@nameparts=split(/\./, $filename);
```

```
#print $parts[0];
```

```
$orfs="$nameparts[0]". ".faa";
```

```
$ptt="$nameparts[0]". ".ptt";
```

```
open(IN, "< $ptt") or die "cannot open $ptt:$!";
```

```
$line=<IN>; # read 1st line
```

```
if ($line=~ /complete genome/) { #look forheader
```

```
print "$line\n";}
```

```
$line=<IN>; # read 2nd line
```

```
print "$line\n";
```

```
$line=<IN>; # read 3rd line
```

```
if ($line=~ /Location Strand/) { #look for beginning of table
```

```
while (defined ($line=<IN>)){ # read through rest of table line by line
```

```
@parts=split /\t/, $line;
```

```
@fromto=split /\.\./, $parts[0];
```

```
$middle = (($fromto[1]+$fromto[0])/2);
```

```
print "$fromto[1]\t$fromto[0]\t$middle\t$parts[3]\t";
```

```
$gi_hash{$parts[3]}=$middle;
```

```
print "\n";
```

```
}
```

```
}
```

```
@gi_names = sort(keys(%gi_hash));
```

```
$total=scalar(@gi_names);
```

```
print "total number of GIs= $total\n";
```

```
foreach (@gi_names) {
```

```
print "gi number $_ is located at $gi_hash{$_}\n";
```

```
}
```

```
close(IN);
```

```
# read in and process faa file
```

```
open(IN, "< $orfs") or die "cannot open $filename:$!";
```

```
$outfilename = "$nameparts[0]". ".num.faa";
```

```
open(OUT, "> $outfilename")||die "cannot open: $!";
```

```
#####
```

```
while (defined ($line=<IN>)){ # read through file line by line
```

```
if ($line=~ />/) { #look for beginning of line starting with > (^ is an anchor for the beginning of the line)
```

```
$line =~ m/gi\(\d+\)\|/; #match gi number capture number in $1
```

```
$num=$gi_hash{$1};
```

```
$line =~ s/^>//;
```

```
# print "$1 $num \n";
```

```
$line= ">". "$num\t". " $line";
```

```
};
```

```
print "$line"; #print to screen
```

```
print OUT "$line"; #print to OUT
```

```
}
```

```
close(IN);
```

```
close(OUT);
```

```
#close(OUT);
```