

Workflow

Commands

gentrez

NCBI Search Results

seqret

NC_000964 Genome FlatFile

gphx

Genome Map PNG Image

sort & cut

Gene List Sorted by PHX

head & tail

Top/Worst 100 PHX Gene List

genret

Sequence Before Start Codons

emma

Aligned Sequences

extractalign

Aligned Regions

kweblogo

Sequence Logo

1. Search NCBI Entrez using *gentrez*

```
% gentrez genome "bacillus subtilis" -outfile stdout
```

2. Retrieve genome flatfile of NC_000964 using *seqret*

```
% seqret refseqn:NC_000964 -feature -osformat:GenBank -outseq bsub.gbk
```

3. Identify predicted highly expressed (PHX) genes using *gphx*

```
% gphx bsub.gbk -outfile phx.csv
```

4. Sort gene names by gene expression measure values

```
% grep ^\0 phx.csv | sort -t , -k 1 -nr | cut -d , -f 6 > sorted_genes.list
```

5. Get the top/worst 100 genes from sorted file

```
% [head|tail] -n 100 sorted_phx.list > phx_[top|bottom]100.list
```

6. Extract sequence around the start codon of the genes using *genret*

```
% genret bsub.gbk -gene @phx_[top|bottom]100.list -access before_startcodon  
-outfile [top|bottom]100_before.fasta
```

7. Align the sequences using *emma*

```
% emma [top|bottom]100_before.fasta -outseq [top|bottom]100.aln  
-dendoutfile [top|bottom]100.dnd
```

8. Extract 20bp sequence before the start codon using *extractalign*

```
% extractalign [top|bottom]100.fasta -region 80-100 -outseq [top|bottom]100_extract.aln
```

9. Draw a sequence logo using *kweblogo*

```
% kweblogo [top|bottom]100_extract.aln -goutfile [top|bottom]100.png
```

GEMBASSY Tools

Other EMBOSS & EMBASSY Tools

UNIX Command Tools