

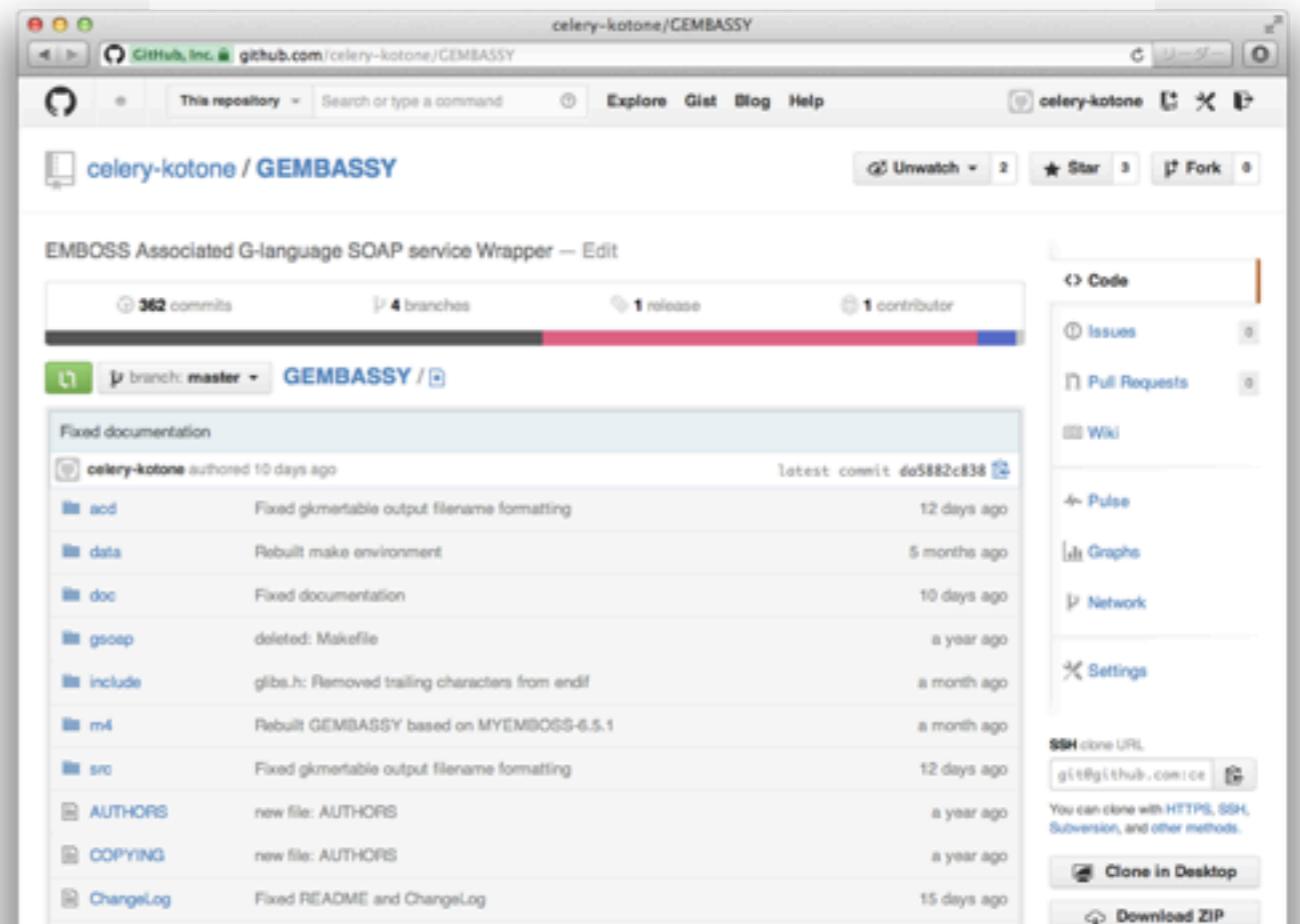
GEMBASSY:

an EMBOSS associated software package for genome analysis using G-language SOAP/REST web services

<http://www.g-language.org/gembassy>



<http://github.com/celery-kotone/GEMBASSY/>



°Hidetoshi Itaya, Kazuki Oshita, Kazuharu Arakawa, Masaru Tomita
Institute for Advanced Biosciences, Keio University

GEMBASSY:



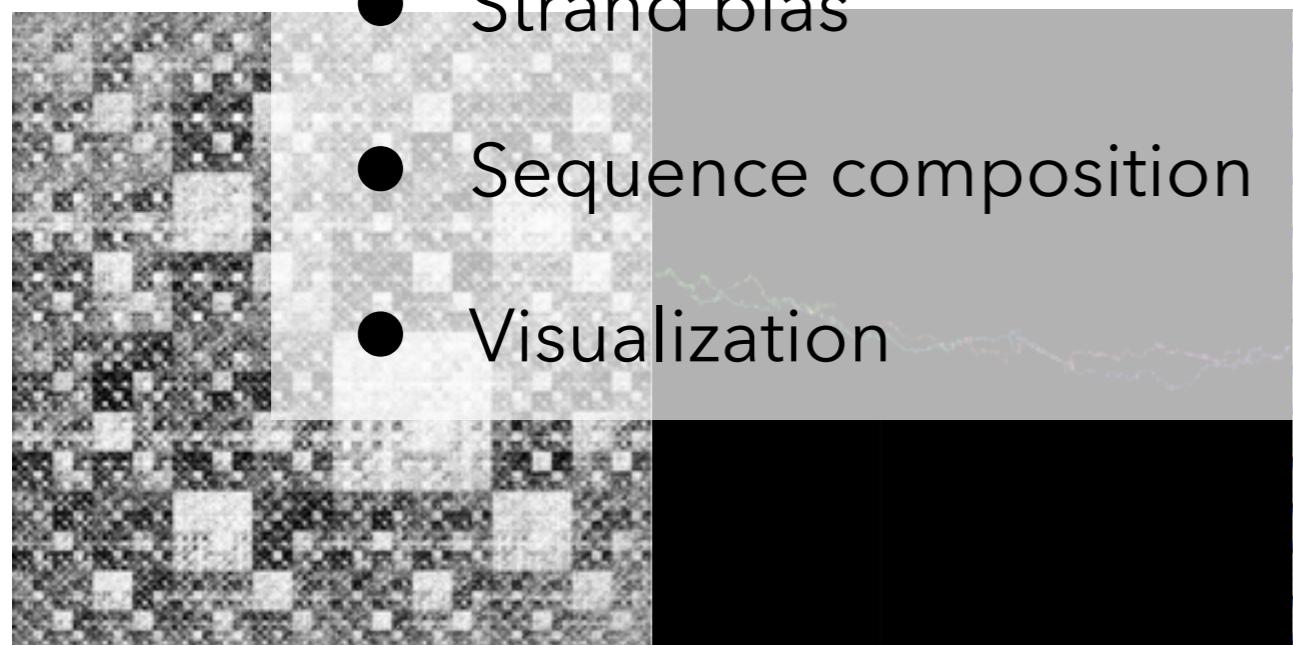
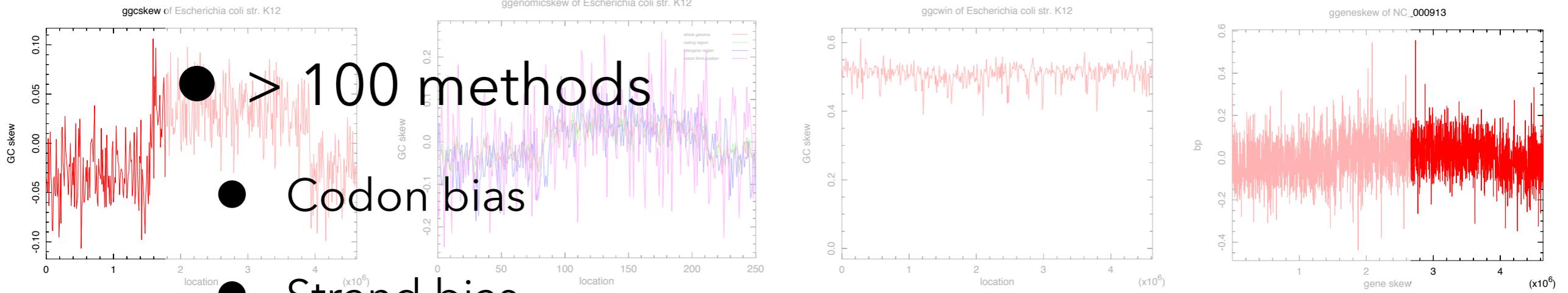
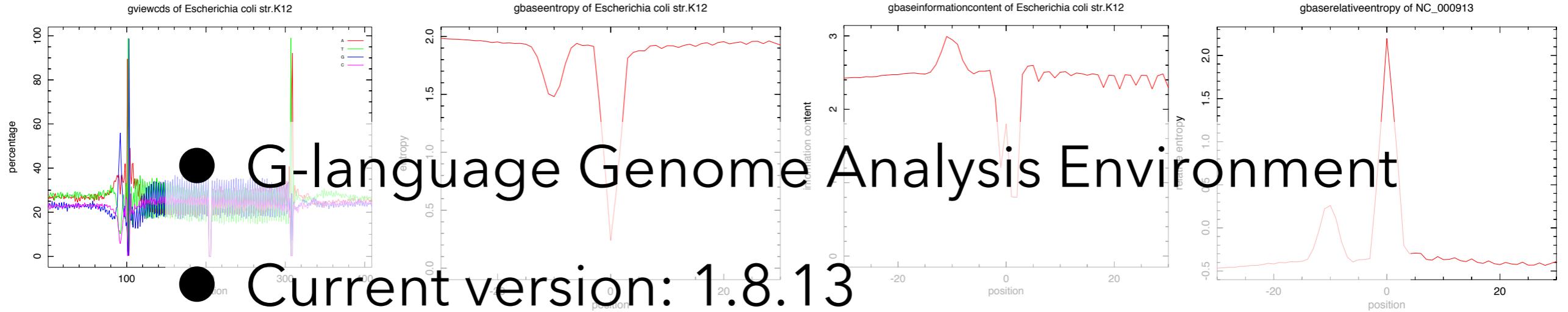
- EMBOSS Associated Software (GEMBASSY) Package
- 53 tools wrapping G-language REST/SOAP Web Services
- Available from website, GitHub, and EMBOSS Explorer
- GNU General Public License version 2

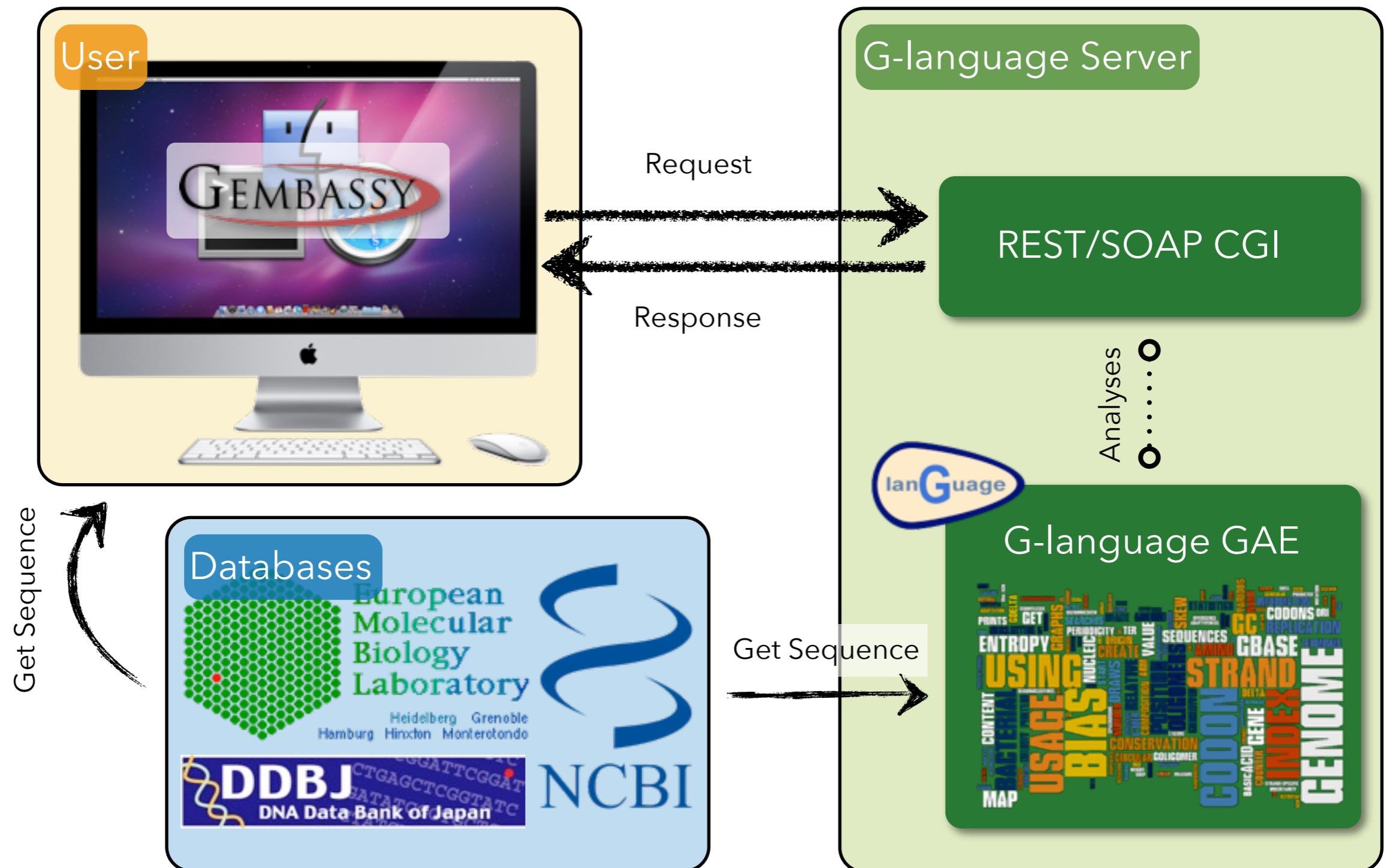
http://soap.g-language.org/gembassy/emboss_explorer

```
% gphx nc_000964.gbk
Identify predicted highly expressed gene
Codon usage output file [nc_000964.gphx]: stdout
Sequence: NC_000964
BgC,BgH,E_g,phx,pa,gene
0.2049,0.5814,0.3524,0,0,dnaA
0.3277,0.5478,0.5982,0,0,dnaN
0.5482,0.8992,0.6096,0,1,yaaA
0.2243,0.6320,0.3548,0,0,recF
0.5314,0.7526,0.7060,0,1,yaaB
0.2482,0.4458,0.5567,0,0,gyrB
0.2859,0.3850,0.7426,0,0,gyrA
0.3098,0.6364,0.4868,0,0,yaaC
0.3468,0.3939,0.8805,0,0,guAB
0.2461,0.4059,0.6063,0,0,dacA
0.4680,0.2929,1.5978,1,0,yaaD
0.3604,0.5515,0.6535,0,0,yaaE
0.2810,0.5049,0.5565,0,0,serS
0.4448,0.6043,0.7361,0,1,dck
0.3819,0.7195,0.5308,0,0,dgk
0.2730,0.6159,0.4433,0,0,yaaH
0.3329,0.5067,0.6570,0,0,yaaI
0.4420,0.6178,0.7154,0,1,tadA
0.2040,0.5886,0.3465,0,0,dnaX
0.4475,0.3348,1.3365,1,0,yaaK
```

The screenshot shows the 'genret' tool configuration page within the EMBOSS Explorer interface. The URL in the address bar is http://soap.g-language.org/gembassy/emboss_explorer/. The left sidebar lists various tools under categories like DATA RETRIEVAL, DISPLAY, DOCUMENTATION, FEATURE TABLES, and NUCLEIC CODON USAGE. The main panel for 'genret' includes sections for 'Input section', 'Advanced section', and 'Output section'. In the 'Input section', there are three input fields: 1. A database accession number input field with placeholder 'USA'. 2. A local sequence file input field with placeholder 'ファイルを選択'. 3. A manual sequence entry input field. Below these are fields for 'List of gene name(s) to report' and 'Name of gene feature to access'. The 'Advanced section' contains an 'Extra arguments to pass to method' input field and a 'Include to use sequence accession ID as query?' checkbox. The 'Output section' is currently empty.

G-language GAE





EMBOSS (> 6.5.7)

gSOAP Toolkit (version 2.8)

libcurl-devel (> 7.29.0)

G-language SOAP Service WSDL File:

Mac OS X/UNIX

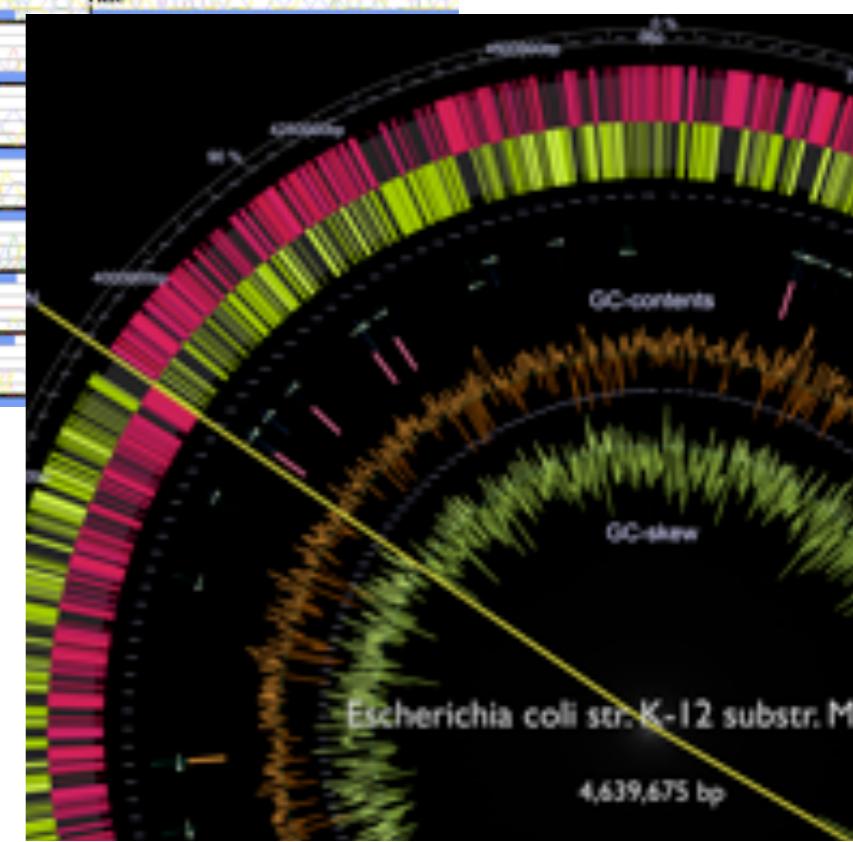
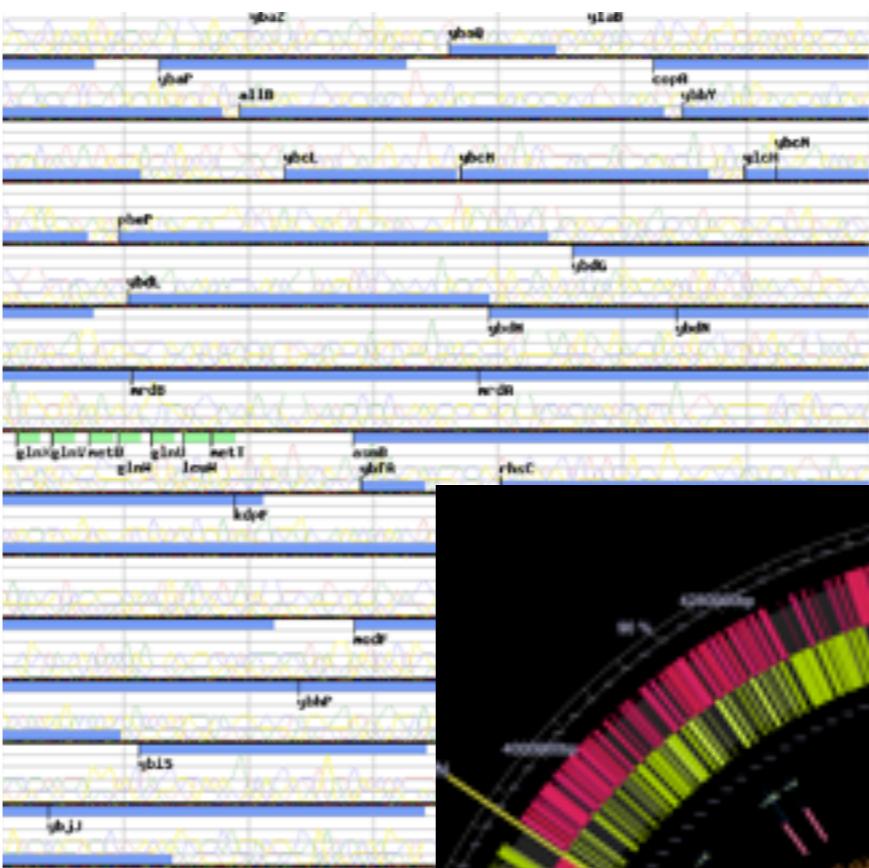
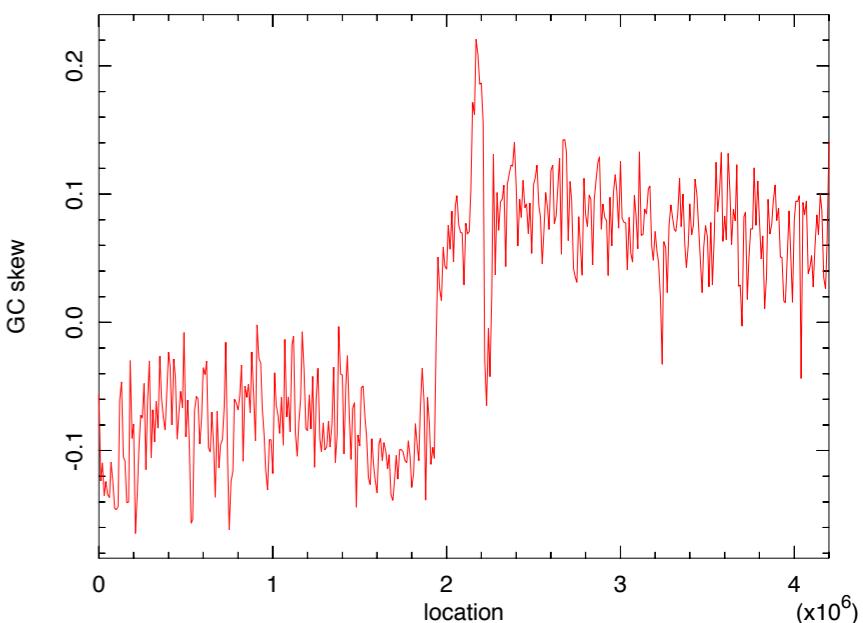
Working with GEMBASSY

```
% seqret -feature refseqn:NC_000964 -osformat genbank  
Read and write (return) sequences  
output sequence(s) [nc_000964.genbank]:nc_000964.gbk
```

```
% ggcskew nc_000964.gbk
```

Calculates the GC skew of the input sequence

Created ggcskew.ps



```
% ggenomemap3 nc_000964.gbk
```

```
% gcircularmap nc_000964.gbk
```

```
% gphx nc_000964.gbk
```

Identify predicted highly expressed gene
Codon usage output file [nc_000964.gphx]:

```
% grep ^0 nc_000964.gphx |  
sort -t , -k 3 -nr |  
cut -d , -f 6 > genes_sorted.txt
```

```
% head -n 100 genes_sorted.txt > top_genes.txt
```

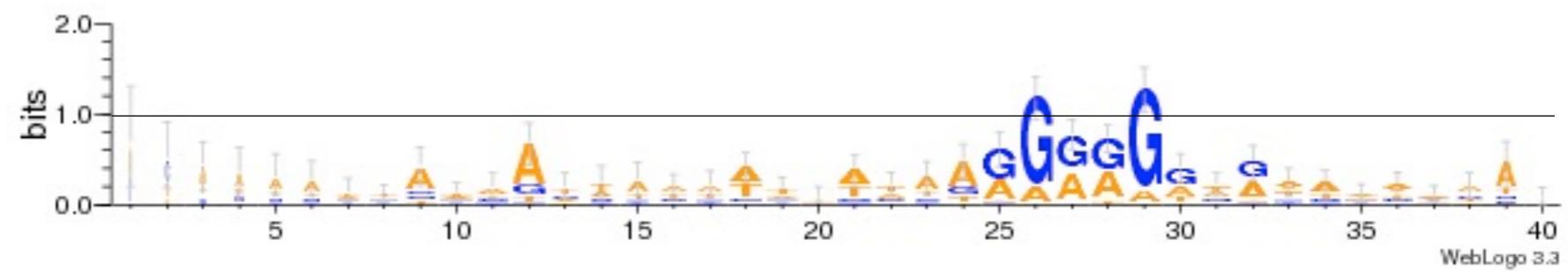
```
% tail -n 100 genes_sorted.txt > bottom_genes.txt
```

Single gene All genes
(rbsC or *)

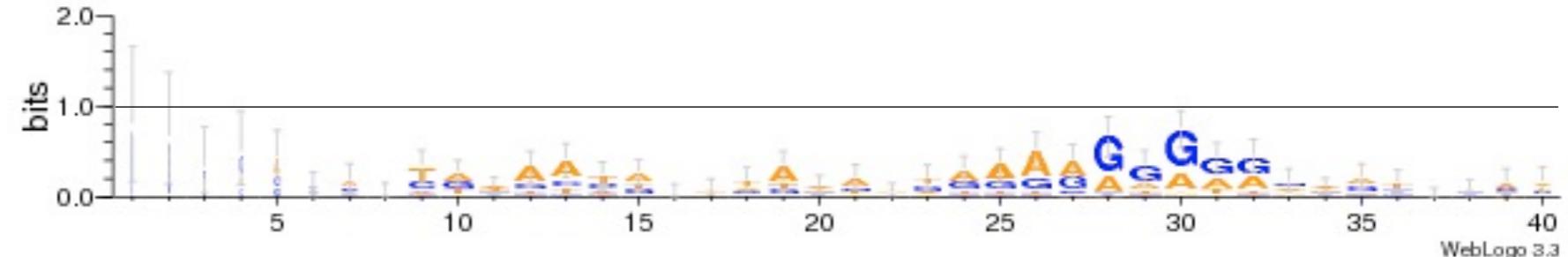
% genret nc_000964.gbk @[top|bottom]_genes.txt
before_startcodon [top|bottom]_before_startcodon.fasta

emma -> extractalign -> kweblogo

top.png



bottom.png



Acknowledgments

This research is supported by funds from the Yamagata Prefectural Government and Tsuruoka City

Many thanks to organizers of BOSC 2013

ISMB Posters

ISMB Poster - B08

GEMBASSY: an EMBOSS associated package for genome analysis using G-language SOAP/REST web services

Hidetoshi Itaya, Kazuharu Arakawa, Masaru Tomita