

How to use Bacteria Analysis System

G-language Project

February, 2003

1 Introduction

In this work, we introduce our Bacteria Analysis System (BAS) developed using the G-language Genome Analysis Environment. BAS is a collection of perl modules that facilitate the development of perl scripts for bioinformatics applications. The major characteristics of the system are:

1. Flexible environment able to be adjusted to needs, in order to perform high level analyses on complex subjects.
2. Natural user interface with which even the researchers inexperienced with computers can tirelessly manipulate.

By connecting the various analysis functions incorporated, including BioPerl programs, BAS realizes a high level analysis environment. Moreover, a graphical user interface built for BAS provides an easy-to-use interface for users.

2 Interface

2.1 Control Panel

Control Panel is the central interface of BAS for editing a file, operating Analysis Packages such as the BAS and other interfaces.

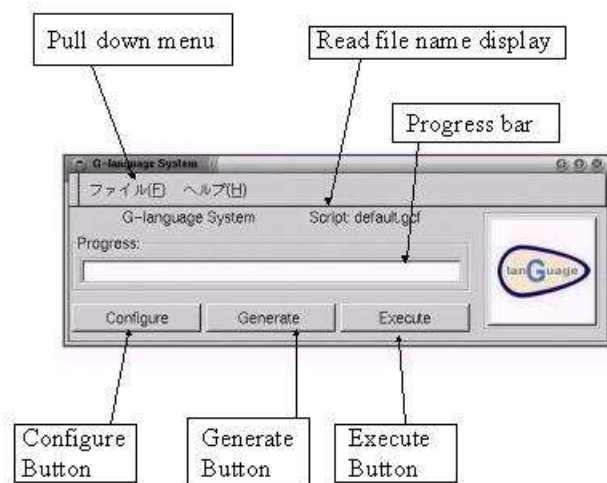


Figure 1: Control Panel

- Pull down menu list

File(F)	Load Script	read script file
	Exit	Exit the G-language Manager
Help	www.g-language.org	Jump to the web page of G-language Project
	About	author list

- Configure button
Click this button, you can open Configure Window and edit GCF.
- Generate button
Output a configured analysis script into a file.

- Excute button
Excute Bacteria Analysis System.
- Progress bar
Displays a progress of Bacteria Analysis System as bar graph.

2.2 System Console Window

This Window displays a progress of Bacteria Analysis System as script. For example, it displays the progress of each analysis.

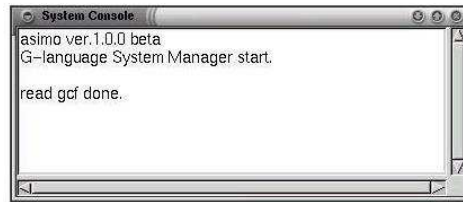


Figure 2: System Console Window

2.3 Configure Window

You can operate Bacteria Analysis Systems intuitively with this window. Now, we have 24 methods consisting of 10 classes. Basic usage is common to all method , but there are some particular options. Basic usage common to all methods is to specify the database at the Database configuration (default: Mycoplasma genitalium), to select methods to use at the Method configuration, to specify sequence data at the Sequence specification box (instance_G) and save the file with Save button or Save As button, and to excute analysis with the control panel.

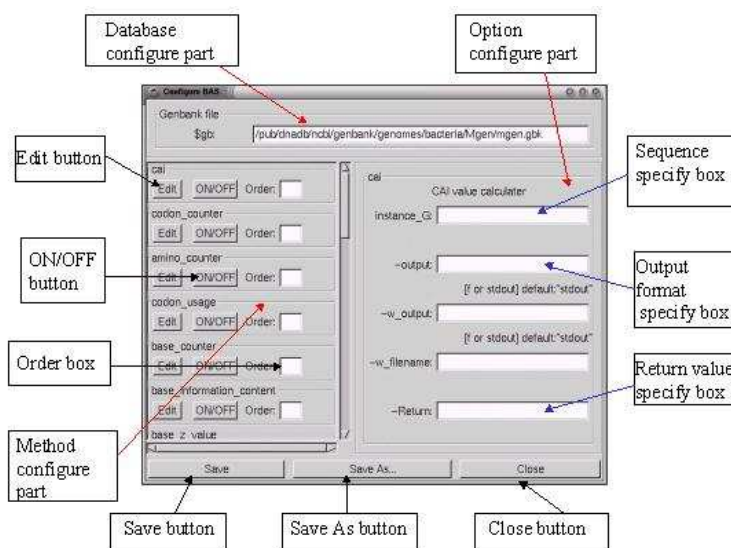


Figure 3: Configure Window

- Database configuration
Specify analysis data.
- Method configuration
Specify the methods of analyses.
- Option configuration
Specify the options for the analysis.
- Edit button
Click the button and display the option configuration window of each method on the right side.
- ON/OFF button
Click the button and operate ON/OFF switch of each method.
- Close button
Close the Configure Window.
- Sequence specification box(instance_G)
Specify the sequence for the analysis. (ex. \$gb, \$gb->{SEQ})
- Output format specification box(output)
Specify the output format of analysis result. There are some difference between each methods, but basically we have 3 options (f, g, show).

-output	f	save data into a file
	g	display analysis result as a graph
	show	display analysis result automatically

- Return value specification box(Return)
To analyze a data based on another analysis, you can specify the return value of each methods.

2.4 Text Output Window

After execution, most methods output standard text output result (shown below).

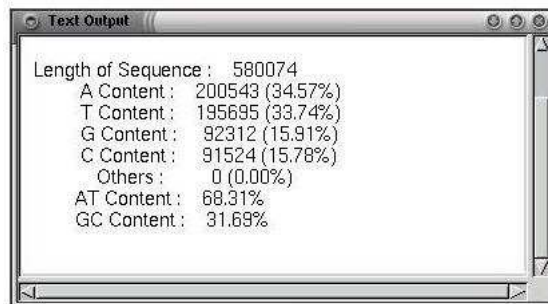


Figure 4: Text Output Window

In addition to this standard text output result, some methods output graph or table shown below.

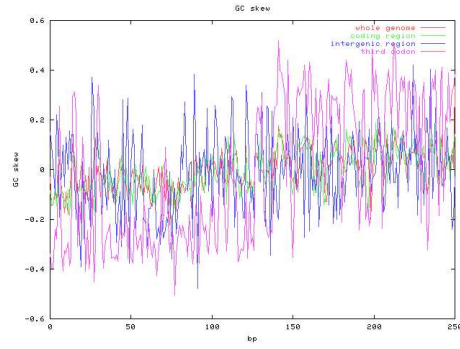


Figure 5: Genomic GCskew

First	second				Third
	T	C	A	G	
T	TTT F 0.865	TCT S 0.188	TAT Y 0.742	TGT C 0.297	T
	TTC F 0.135	TCC S 0.061	TAC Y 0.258	TGC C 0.203	C
	TTA L 0.473	TCA S 0.244	TAA / 0.725	TGA M 0.642	A
	TTG L 0.133	TCG S 0.018	TAG / 0.275	TGG M 0.358	G
C	CTT L 0.187	CCT P 0.490	CAT H 0.649	CGT R 0.225	T
	CTC L 0.047	CCC P 0.121	CAC H 0.351	CGC R 0.099	C
	CTA L 0.119	CCA P 0.257	CAA Q 0.809	CGA R 0.044	A
	CTG L 0.041	CCG P 0.032	CHG Q 0.191	CGC R 0.034	G
A	ATT I 0.620	ACT T 0.472	AAT N 0.613	AAT S 0.389	T
	ATC I 0.218	ACC T 0.191	AAC N 0.287	ACC S 0.100	C
	ATA I 0.154	ACA T 0.207	AAR K 0.742	AGA R 0.443	A
	ATG H 0.380	ACG T 0.030	ARG K 0.258	AGG R 0.149	G
G	GTT Y 0.620	GCT A 0.492	GAT D 0.851	GAT G 0.498	T
	GTC Y 0.095	GCC A 0.075	GAC D 0.139	GAC G 0.109	C
	GTA Y 0.213	GCA A 0.386	GAH E 0.803	GAG G 0.248	A
	GTG Y 0.112	GCG A 0.047	GHG E 0.197	GGG G 0.146	G

yellow minus charge red plus charge blue noncharge green nonpolar
 exception GTG H 0.013
 TTG H 0.007

Figure 6: Codon Table

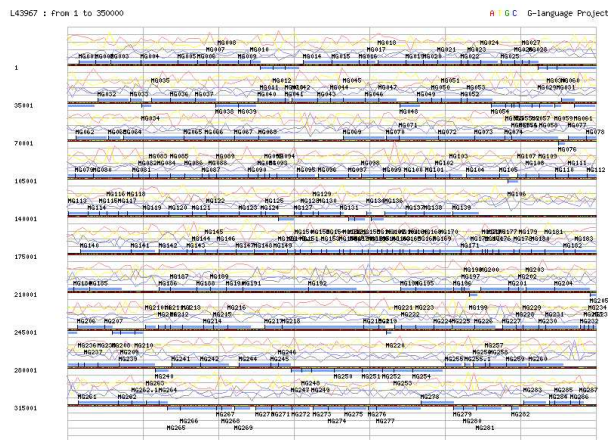


Figure 7: Genome Map

3 Methods list

Class	Method	Function
CAI	CAI	Calculates Codon Adaptation Index for each genes.
Codon	codon_counter	Counts the number of codons in CDS.
	amino_counter	Counts the number of amino acids in CDS.
	codon_usage	Calculates the codon usages for each CDS. When using the whole genome, prints out the codon table.
Consensus	base_counter	Counts consensus patterns around start/stop codons.
	base_information_content	Calculates information content around start/stop codons and graph the results.
	base_z_value	Calculates Z values around start/stop codons.
	base_entropy	Calculates entropy around start/stop codons.
	base_relative_entropy	Calculates relative entropy around start/stop codons and graph the results.
	base_individual_information_matrix	Calculates the individual information around start/stop codons.
GCskew	view_cds	Creates a graph showing the average A, T, G, C contents around the start/stop codons.
	find_ori_ter	Predicts the replicational origin and terminus.
	gcskew	Calculates the GCskew.
	genomicskew	Graphs the GCskew for the whole genome, coding regions, intergenic regions, and the third codon.
	cum_gcskew	Calculates the cumulative GCskew.
	gcwin	Calculates the GC content.
FreeEnergy	foreach_RNAfold	Calculates the free energy of a given sequence using RNAfold.
Markov	markov	Calculates the Markov analysis for all oligomers.
OverLapping	over_lapping_finder	Finds over lapping genes.
PatSearch	palindrome	Searches for palindrome sequences.
TandemRepeat	foreach_tandem	Finds each length and pattern of repeats from a sequence.
	graphical_LTR_search	Converts a sequence to a png image in order to find LTR.
Util	seq2png	Converts a sequence to a png image.
	genome_map	Creates a graphical map of the genome.