

Development of the Bacteria Analysis System on G-Language Genome Analysis Environment

Koya Mori^{1,2} Kazuharu Arakawa¹ Masaru Tomita^{1,3}
 s98982km@sfc.keio.ac.jp t98901ka@sfc.keio.ac.jp mt@sfc.keio.ac.jp

¹ Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0035, Japan

² Department of Policy Management, Keio University, Fujisawa 252-8520, Japan

³ Department of Environmental Information, Keio University, Fujisawa 252-8520, Japan

Keywords: analysis software, development environment, computational analysis, bioinformatics

1 Introduction

Although various genome analysis software are being developed around the world, most lack the following two requirements to become a standard among researchers:

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.

In this work, we introduce and discuss our Bacteria Analysis System (BAS) developed to fulfill the above requirements on the G-language Genome Analysis Environment (<http://www.g-language.org>). By connecting the various analysis functions incorporated, BAS realizes a high level analysis environment. Moreover, a graphical user interface built for BAS provides an easy-to-use interface for users.

2 Methods

Analysis functions incorporated in BAS as shown in Table 1 perform typical bioinformatics analyses, and the results are displayed visually for the users. Each analysis function has variety of options to adjust the analysis according to the needs of users. These analysis functions can also be connected and ordered, in order to perform systematic analysis for thorough understanding of the complex features. This flexibility of BAS allows connection of different analysis methodologies, which is often required in the analyses of bioinformatics. Experienced users can also utilize their own scripts with the functions provided by G-language Genome Analysis Environment and customize the analysis system. In this case, the users can also

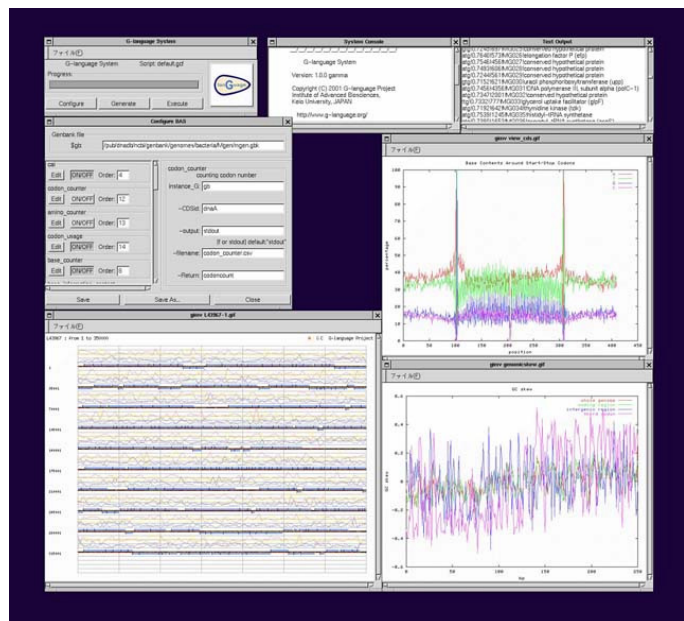


Figure 1: BAS GUI.

Table 1: Methods implemented in BAS.

Method	Function
Cai	Calculates Codon Adaptation Index for each genes
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 calculating for the whole genome, prints out the codon table.
base_counter	Count 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 individual information around start/stop codons.
foreach_RNAfold	Calculates the free energy of a given sequence using RNAfold.
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 GC skew
Genomicskew	Graphs the GC skew for the whole genome, coding regions, intergenic regions, and the third codon
cum_gcskew	Calculates the cumulative GC skew
Gcwin	Calculates the GC content
Markov	Calculates the Markov analysis for all oligomers.
over_lapping_finder	Finds overlapping genes
Palindrome	Searches for palindrome sequences
foreach_tandem	Finds each length and pattern of repeats from a sequence
graphical_LTR_search	Converts a sequence to a gif image in order to find LTR
seq2gif	Converts a sequence to a gif image
genome_map	Creates a graphical map of the genome

directly edit the Perl program file generated from the graphical user interface of BAS for further modification. The above system of analysis may be saved in a form of text script named GCF, and by using this text file, users can perform a systemized analysis on different databases without making changes in the software. Because GCF is a text file, users can directly edit the configuration with text editors.

BAS is built with a graphical user interface for the users to easily access the analyses of bioinformatics. (Figure 1) By modifying the GCF configuration file of BAS, users can control the analysis system, instead of writing software programs to perform original analysis. This configuration of GCF is also possible on the graphical user interface. All modifications including the adjustment of the options of analysis functions can be performed via the graphical user interface.

References

- [1] Stein, L., How Perl saved the human genome project, *TPJ*, 1:5–9, 1996.
- [2] Wall, L., Christiansen, T., and Schwartz, R.L., *Programming Perl* (Second Ed.), O'Reilly & Associates, 1996.
- [3] <http://www.bioperl.org/>