

MGAW : a Microbial Genome Annotation Workbench under Web-based Analysis Interface

Hwajung Seo¹, Hyeweon Nam¹, Daesang Lee^{1,2}, Hongseok Tae¹, Kiejung Park¹

Keywords: annotation, genome, ontology, COG, viewer, browser

1 Introduction

As the genome projects have produced tremendous bioinformatic data, annotation is considered as an essential part of genome sequencing projects to elucidate the value of the sequence[1]. Through annotation systems, molecular biologists could deal with genome data easily and all related information could be accessed, edited, or updated without additional efforts. To satisfy these needs, we have developed MGAW(Microbial Genome Annotation Workbench), a web-based microbial genome annotation system, which provides web-based analysis interface for gene prediction, homology search, promoter analysis, motif analysis and gene ontology analysis. The annotated information can be retrieved with database searching and browsed with a genome map browser and a gene classification viewer. Public microbial genome databases are imported and can be searched and browsed through the same interface.

2 Methods

MGAW database contains from contig data to functional analysis data of the final gene set. The interface of each annotation tool was implemented not only for running each tool and viewing the result but also for monitoring the progress. Analysis results are saved in the database with primary keys indicating the relationship between data.

Gene prediction in a genome project is the first step of annotation. General methods for gene prediction are applied in this system and a few analysis options are provided. For promoter analysis, we implemented a general promoter pattern search and a two-component analysis search against all the predicted genes. For motif analysis, the Prosite DB patterns are searched against all protein sequences which are translated from the gene prediction. Fast algorithms were developed to accomplish fast searching for motif patterns of regular expression. The progress/status of promoter and motif analysis can be monitored through web interface. For homology analysis of all the predicted genes, we implemented the interface for NCBI BLAST and both COG(Clusters of Orthologous Groups) and GO(Gene Ontology) databases were used to classify the homology search result[4].

Database searching module was implemented to query for the annotation results of an in-progress or finished genome project and a linear map browser was implemented to visualize the whole genome map and detailed annotation information for each selected gene by further clicking. A gene classification viewer was implemented to show gene ontology analysis result with COG for a whole genome. A circular map is generated after retrieving gene ontology information of all the

¹ Information and Technology Institute, SmallSoft Co., Ltd. Junmin-Dong 461-71, Yusung-Gu, Daejeon, 305-811, South Korea. E-mail: {hjseo, hwnam, dslee, hstae, kjpark}@smallsoft.co.kr

² Dept. of Biological Science, KAIST, Kusung-dong, Yusung-gu, Daejeon, 305-701, South Korea. E-mail: dslee@bioneer.kaist.ac.kr

genes of a genome and calculating a few features for the whole genome area. A few options were implemented to select a specified category, a region and a drawing mode

For public microbial genome data, input programs were implemented to parse the genome data of GenBank format and import into MGAW databases. Each imported genome can be searched and browsed as an annotated genome can be.

3 Results and Discussions

MGAW has stepwise and intuitive interface as shown in Figure 1. We have tested and improved through a few microbial genome projects.

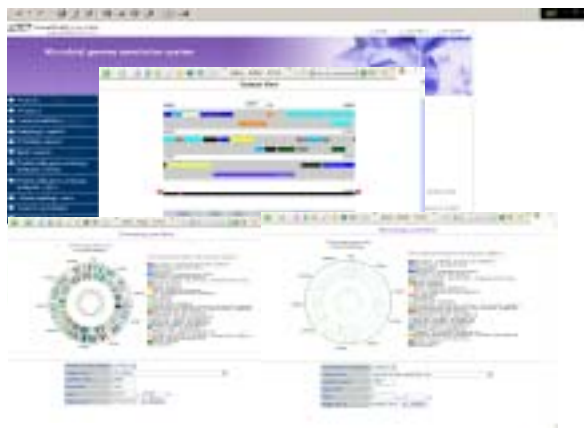


Figure 1: Overview of MGAW features : annotation interface, genome browsing, gene ontology viewing

More features will be added in the near future, including gene ontology viewing for GO, genome alignment, and a lot of comparative genome analysis modules. MGAW will be very helpful not only for analysis of public microbial genome annotation but also as a practical information system of genomics/comparative genomics for real genome projects.

4 References

- [1] Lee, D., Seo, H., Tae, H., Kim, Nam, H. and Park, K. 2003. Development of a web-based genome annotation system. *Research in Computational Molecular Biology (RECOMB 2003)*..
- [2] Seo, H., Kim, K.-B., Tae, H., Park, W., and Park, K. 2002, Development of gene ontology analysis and classification tools for microbial genome annotation. *Research in Computational Molecular Biology (RECOMB 2002)*.
- [3] Tae, H., Seo, H., Nam, H., Lee, D. and Park, K. 2003. Development of a web-based genome annotation system and two analysis tools *Intelligence system for Molecular Biology (ISMB 2003)*.
- [4] Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V. 2001. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*. 29:22-28.