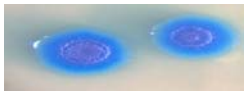
# GENOME DESIGNER FOR SYNTHETIC GENOME

2009 IGEM TEAM OF CHUNGBUK NATIONAL UNIVERSITY
CBNU-KOREA

## Abstract

Our team's final goal is developing a genome designer. We here propose a key tool to help the creation of a genome. The purpose of Essarker, which is a basic component of Genome Designer, is to help users design a genome comprising the essential genes of replication. Essarker is a standalone software to manage and retrieve required sequences of genomes, and explore the essential gene order and direction and the related orthologous genes. It also identifies and visualizes the positions and orientations of genes. In addition, it shows optimal ordering of essential genes and orthologs by statistical analysis. It will be useful to create a synthetic genome for fulfilling the needs of energy and food.

## Motivation



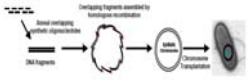
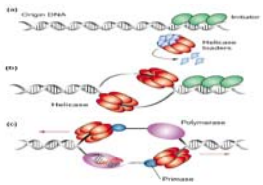


**Fig 1. What is Synthetic Genome?**

Synthetic genome is artificially compounded Genome and included mini genome and insert the desired genes. It can make materials to produce the desired. This picture shows general method of synthesizing synthetic cell



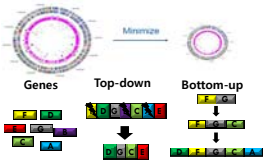



**Fig 2. Definition of minimal genome and scheme of minimal genome**

The minimal genome is smallest possible group of genes that would be sufficient to sustain a functioning cellular life form under the most favorable conditions imaginable. It is two way of the synthesizing minimal genome which are top-down and bottom up. Our group uses bottom-up way for designing minimal gen

**Fig 3. The principle of replication**

General overview of the initiation of DNA replication.

(a) Initiators bind origin DNA, and then remodel the DNA and recruit additional factors, such as helicases and helicase loaders.

(b) Helicases further unwind the DNA.

(c) This allows the addition of polymerases and primases to the origin.

Once sliding clamps, clamp loaders and single stranded DNA binding proteins are loaded onto the DNA, the replisomal machinery is complete and replication proceeds to the elongation Phase..

## Mini genome design

Mini genome is a self replicating DNA element comprising an origin and essential genes for replication.

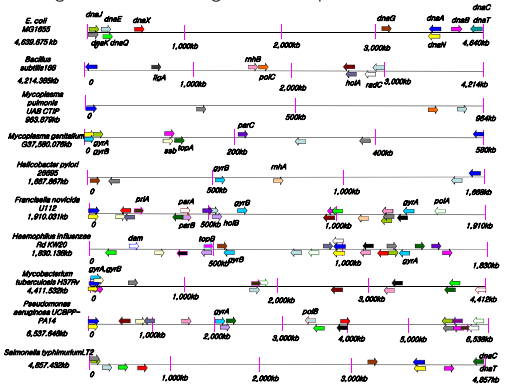


**Fig. 4 Arrangement of essential genes for DNA replication**

## Problems

1. Position
2. Direction
3. Operon structure
4. Intergenic space

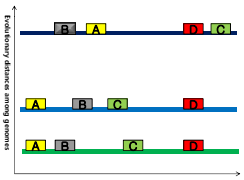



**Fig 5 Problem lists of genome design**

During genome designing needs various and much information. But many tools are separated in the web page and software for genome design. And also we consider various things like position, direction, operon structure.

## Analysis of essential genes for replication

**Fig 6. List of Database**

It is list of DB which from we get the essential genes information. Our team gets essential gene sequence from DEGG, KEGG and NCBI.

**Fig 7. Collect information**

| Bacteria name | Software | Coloring |
|---|---|---|
| Escherichia coli MG1655 | O | O |
| Bacillus subtilis 168 | O | O |
| Mycoplasma genitalium G37 | O | O |
| Staphylococcus aureus N315 | O | X |
| Francisella novicida U112 | O | X |
| Acidothiobacter ferrivorans ACF1 | O | X |
| Pseudomonas aeruginosa UCBPP-PA14 | O | X |
| Salmonella typhimurium LT2 | O | X |
| Mycobacterium tuberculosis H37Rv | O | X |
| Streptococcus pneumoniae CGSP14 | O | X |
| Haemophilus influenzae RdKW20 | O | X |
| Helicobacter pylori 26695 | X | X |
| Mycoplasma pulmonis UABCTP | X | X |

**Table 1. list of essential gene species**

Our team uses data for testing genome designer which is 13 species and making ortholog clustering use 3species.

| No. | Name of the gene | Function |
|---|---|---|
| 1 | dnaA | Chromosomal replication initiator protein |
| 2 | dnaN | DNA polymerase III beta subunit |
| 3 | gyrB | DNA gyrase subunit B |
| 4 | gyrA | DNA gyrase subunit A |
| 5 | dnaE | DNA polymerase III alpha subunit |
| 6 | dnaX | DNA polymerase III tau subunit |
| 7 | recF | Recombination protein F |
| 8 | recR | Recombination protein R |
| 9 | recO | Recombination protein O |
| 10 | ligA | DNA ligase |
| 11 | rnhA | Ribonuclease H |
| 12 | rnhB | Ribonuclease HII |
| 13 | dnaG | DNA primase |
| 14 | rpoD | RNA polymerase sigma factor |
| 15 | rpoA | RNA polymerase alpha subunit |
| 16 | rpoB | RNA polymerase beta subunit |
| 17 | rpoC | RNA polymerase beta' subunit |
| 18 | rplK | 50S ribosomal protein L11 |
| 19 | rplA | 50S ribosomal protein L1 |
| 20 | rplJ | 50S ribosomal protein L10 |
| 21 | rplL | 50S ribosomal protein L7/L12 |
| 22 | rpsL | 30S ribosomal protein S12 |
| 23 | rpsG | 30S ribosomal protein S7 |
| 24 | fusA | Elongation factor G |
| 25 | tufA | Elongation factor Tu |

**Table 2. Essential genes for DNA replication**

## Software :Essarker

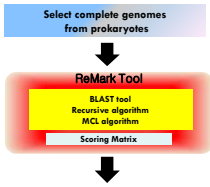
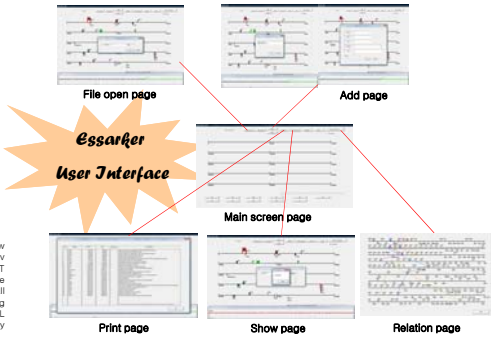



**Fig. 8 Overview of clustering orthologs**

Clustering orthologs of essential gene We here used a new hybrid method combining the Recursive and the Markov CLuster (MCL) algorithms without using the BLAST thresholding. In the first step, we use InParanoid to produce n(n-1)/2 ortholog tables from n genomes. After combining all the tables into one, our clustering algorithm clusters ortholog pairs recursively in the table. Then, our method employs MCL algorithm to compute the clusters and refines the clusters by adjusting the inflation factor.




## Future

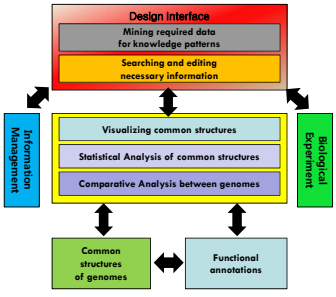



**Fig 9. Final architecture**

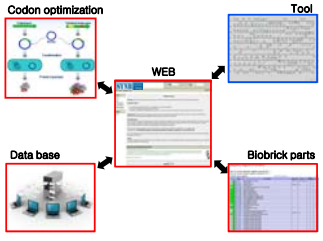



**Fig 10. Scheme of future**

## References

1. Dasgupta, S., Lobner-Olesen, A., 2004, Host controlled plasmid replication: Escherichia coli minichromosomes, Elsevier Inc. plasmid 52, 151-168
2. Edwards, J. S., Ibarra, R. U., Palsson, B.O., Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth, 2002, nature. 420, 186-189
3. Forster, A. C., Church, G. M., Synthetic biology projects in vitro, 2007, Genome Res. 17, 1-6
4. Ren Zhang, Yan Lin, DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes, 2009, Nucleic Acids Research. 37, 455-458
5. Baba, T., Takeshi, A., Hasegawa, M., Yuki, T., Yoshiko, O., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., Mori, H., 2006, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, Molecular Systems Biology. 21, 1-11
6. kim, s. s., Kang, J., Chang, Y. J., Li, J., Ryu, K. H., 2008, Clustering orthologous proteins across phylogenetically distant species, Proteins. 71, 1113-1122.

## Sponsors