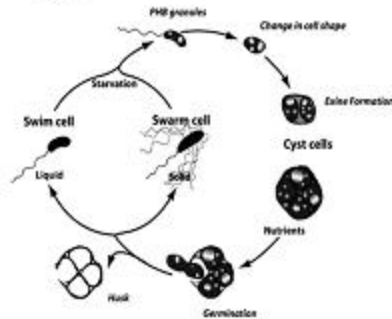# Genome Annotation of Rhodospirillum Centenum
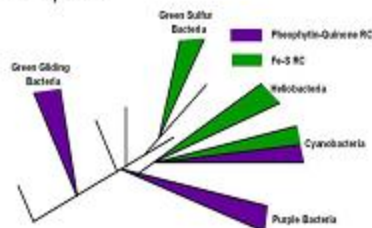
Problem?

Annotate the newly sequenced genome of R. Centenum. Use Computational methods to accurately and efficiently find genes and assign functions.

Rhodospirillum Centenum?

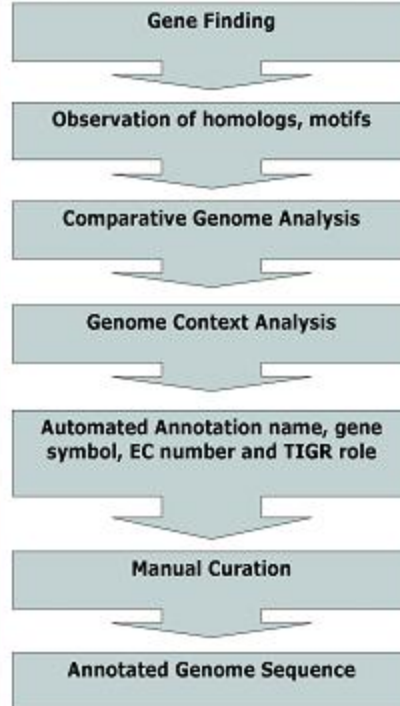- R. Centenum is a purple bacteria with a complex life cycle

- R. Centenum is a model organism for photosynthesis, phototaxis and cellular development

- Association of R. Centenum with plants exhibit beneficial effects on plant growth and yields through their aerobic capability of nitrogen fixation

Methods

- **Gene Finding**
- **Observation of homologs, motifs**
- **Comparative Genome Analysis**
- **Genome Context Analysis**
- **Automated Annotation name, gene symbol, EC number and TIGR role**
- **Manual Curation**
- **Annotated Genome Sequence**

- How to find gene regions with accuracy
- How to observe significant homologs or signal patterns
- How to improve function assignment by comparing other genomes of related species
- How to assign annotations automatically by scaling multiple observations

Comparative Genome Analysis

Finding clusters of orthologous groups can provide insight into gene family and function.

BBH -> Triangle -> BAG clustering

- BBH: reciprocal best-hit relationship between proteins in two genomes.
- Triangle: 3 BBH that have more than 2 nodes (proteins) in common will form a Triangle.
- BAG clustering: clustering of sequences using graph theory.
    - *biconnected component:* a maximal biconnected subgraph of G where there are at least two paths for any pair of vertcies.
    - *articulation point :*a vertex whose removal disconnects the graph.
    - a set of vertices in each biconnected component forms a family of sequences
- COGs : Clusters of Orthologous Groups database at NCBI

Genomic Context Analysis

- co-occurrence of genes across different species
- gene fusion events
- gene neighborhood

Results

The sequencing is in the process of assembly and almost finished.

Once the sequence data are ready we will be able to analyze the genome and get meaningful results.

Future Work

The genome annotation methods are switching from the basic homologue search to the context search approach.

As well as using existing methods to find information on the genes we could work on designing novel computational methods to find relationships among genes within and across the genomes.

Reference

- Kwangmin Choi, et al. PLATCOM: A Platform for Computational Comparative Genomics, Bioinformatics 2005
- S. Kim. Graph theoretic sequence clustering algorithms and their applications to genome comparison, *Computational Biology and Genome Informatics* 2003
- Berleman, J. & Bauer, C. Characterization of cyst cell formation in the purple photosynthetic bacterium, *Rhodospirillum centenum.* Microbiology 2003

Mira Han   Indiana University School of Informatics   mirhan@indiana.edu