



# De Novo Assembly of the *Chryseobacterium haifense* Genome

Thomas Sontag<sup>1</sup>, Jordan Krebs<sup>1</sup>, Jeffrey Newman<sup>1</sup>

<sup>1</sup>Department of Biology, Lycoming College, 700 College Place, Williamsport, Pa 17701



## RAST Sequence Based Comparison

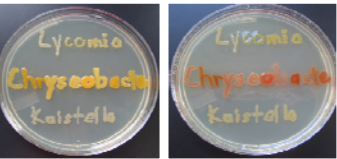
## Abstract

To clarify the phylogenetic position of *Chryseobacterium haifense*, we sought to sequence its genome. The DNA was sequenced with 40x coverage using Ion Torrent technology as a part of the GCAT-SEEK consortium. The Ion Torrent produced about 1,000,000 reads which were then assembled into 2,761 contigs using NextGene software.

The NextGene assembly was imported into Geneious software to construct supercontigs. A series of *de novo* assemblies were created to determine the optimal parameters for Geneious assembly. The *de novo* assemblies were created either without trimming, or with trimming of 15, 25, 35, 45, 50, or 60 base pairs from the 5' and 3' ends of the contigs.

A correlation was observed between the number of base pairs trimmed and the number of supercontigs created by Geneious from the NextGene contigs. The number of contigs used increased while the number of supercontigs decreased. This trend was not observed with the trimming of 45 and 60 base pairs. However, the assembly with 50 base pairs did follow this trend. The assembly with 50 base pairs trimmed was the optimal *de novo* assembly for *C. haifense*.

The supercontigs were edited for ambiguous bases in order to be analyzed using Rapid Annotation using Subsystem Technology (RAST). A sequence based comparison matched orthologs from *C. haifense* to two reference organisms, *C. gleum* and *Flavobacteriaceae* sp. 3519-10. RAST was used to determine the order and orientation of the supercontigs for assembly into scaffolds in preparation for gap closure and finishing. The completed genome assembly may complement other data suggesting that *C. haifense* belongs to a novel genus.



## Background

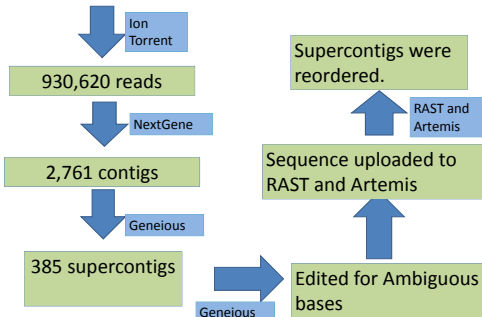
*Chryseobacterium haifense* is not the typical color of a *Chryseobacterium* and other phenotypic differences also suggest that it may not belong to that genus.

## Hypothesis

The genome of *C. haifense* will be significantly different than the genome of *C. gleum* indicating a need for a novel genus.

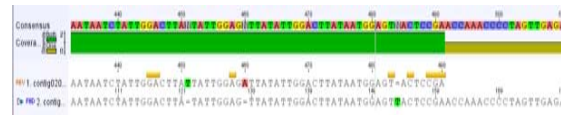
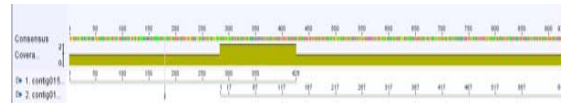
## Methods

gDNA was isolated from *C. haifense*



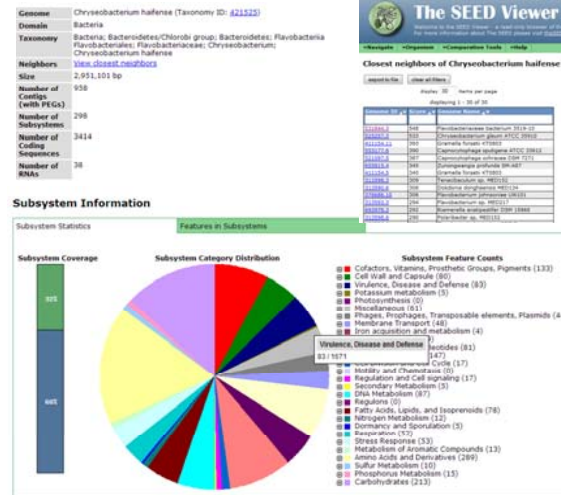
## Results

### Geneious Supercontigs



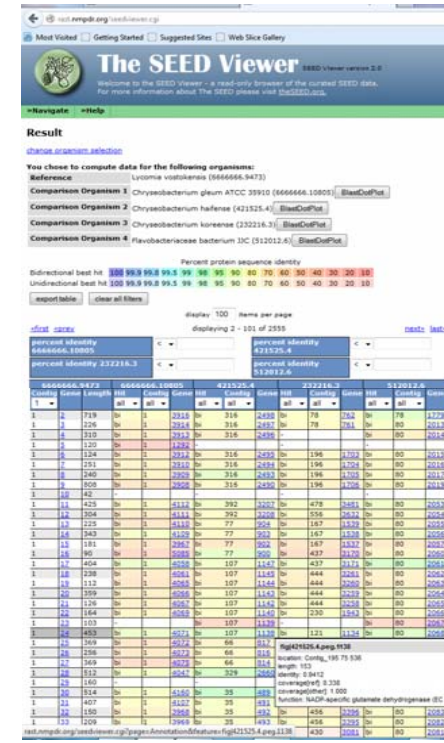
- User-Friendly Geneious program used to identify overlaps between NextGene Contigs, assemble into "Supercontigs"
- Mismatches create ambiguities that must be repaired before upload into RAST for Annotation

### Organism Overview of *C. haifense* using RAST

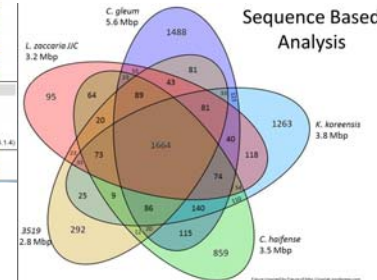


## Acknowledgments

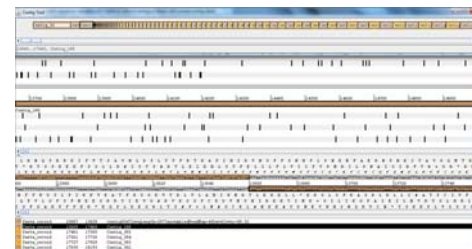
The author would like to thank the GCAT seek consortium for funding of the genome sequencing and the Lycoming College Biology Department for funding of this project.



- Sequence based comparison with related genomes suggests adjacent contigs/ supercontigs.
- Contigs/supercontigs can be reordered & reverse complemented in Artemis
- Detailed sequence/ORF analysis can determine if
  - contigs/supercontigs overlap , or
  - if gap exists to be closed by PCR and sequencing during finishing.
- Comparison of genes allows identification of species specific genes, genes shared by different groups



## Supercontig Reordering using Artemis



## Conclusions

- Upon completing the reordering and assembly of the supercontigs, the gene content will be re-analyzed
- The analyzed genome will provide insight to the proper taxonomic position for *C. haifense*.