

# Comparative and Regulatory Genomics-class assignment 2

## Predict the function of FG00500.1

Kristen Michelle Nader r0771801

November 10, 2020

Copy the file OrthoSearch.fsa

Fasta file contains fungal sequences. Names are taxonomy ID followed by a protein identifier

**Create a multiple sequence alignment and bootstrap a phylogenetic tree based on these fungal proteins.**

Listing 1: Multiple Sequence Alignment

```
clustalw2 -INFILE=OrthoSearch.txt -ALIGN -TREE -BOOTSTRAP -BOOTLABELS=branch  
-CLUSTERING=NJ -OUTPUTTREE=nj -TOSSGAPS -NEWTREE=OrthoSearch_OUTPUT.tree
```

**Visualize your tree in itol and root the tree. (Midpoint rooting or outgroup, encephalitozoa are an outgroup to the fungi taxids: 907965, 876142)**

**Locate [Search] the *Saccharomyces cerevisiae* (yeast) proteins in your tree (taxid:4932).**

What are the functions of these yeast proteins? Hint: go to SGD (<http://yeastgenome.org/>).

Table 1: Your first table.		
Systematic Name	Standard Name	Function
YCL040W	GLK1	Glucokinase
YDR516C	EMI2	Unknown Function
YFR053C	HXK1	Hexokinase isoenzyme 1
YGL253W	HXK2	Hexokinase isoenzyme 2

**Locate [Search] protein FGSG\_00500P0 (protein encoded by gene FG00500.1) in your tree (taxid:5518, *Fusarium graminearum*).**

**To which yeast protein is FGSG\_00500P0 orthologous?**

To determine which yeast protein is orthologous, we can check branch lengths. Instead of adding all branch lengths, we can check which branch length is smaller ; that of YFR053C or YGL253W.

**Which function do you predict for FGSG\_00500P0?**

Table 2: Your first table.		
Systematic Name	Standard Name	Distance
YFR053C	HXK1	0.1203
YGL253W	HXK2	0.1045

Then YGL253W is assumed to be an ortholog. We predict FGSG\_00500P0 to have a similar function- Hexokinase isoenzyme 2.

In addition, we can use STRING database to validate the results. Figure 1 and 2 display the protein network which also validates our prediction that the protein may be a Hexokinase.

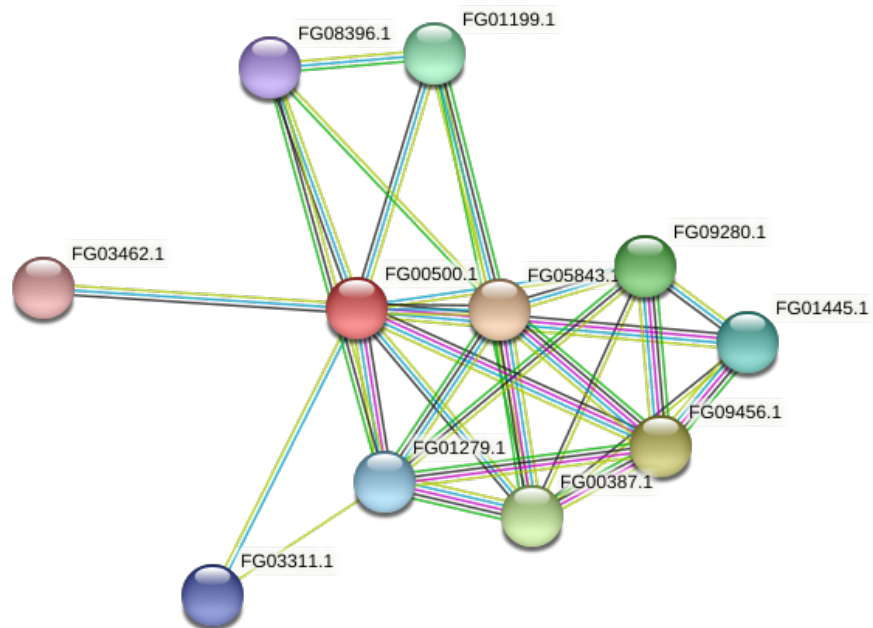


Figure 1: String DB result

## FG00500.1

### Information

Phosphotransferase; Hexokinase; Belongs to the hexokinase family

Identifier: FGSG\_00500P0, FG00500.1

Organism: *Fusarium graminearum*



- [show protein sequence](#)
- [homologs among STRING organisms](#)

Figure 2: Result from FGSG\_00500P0

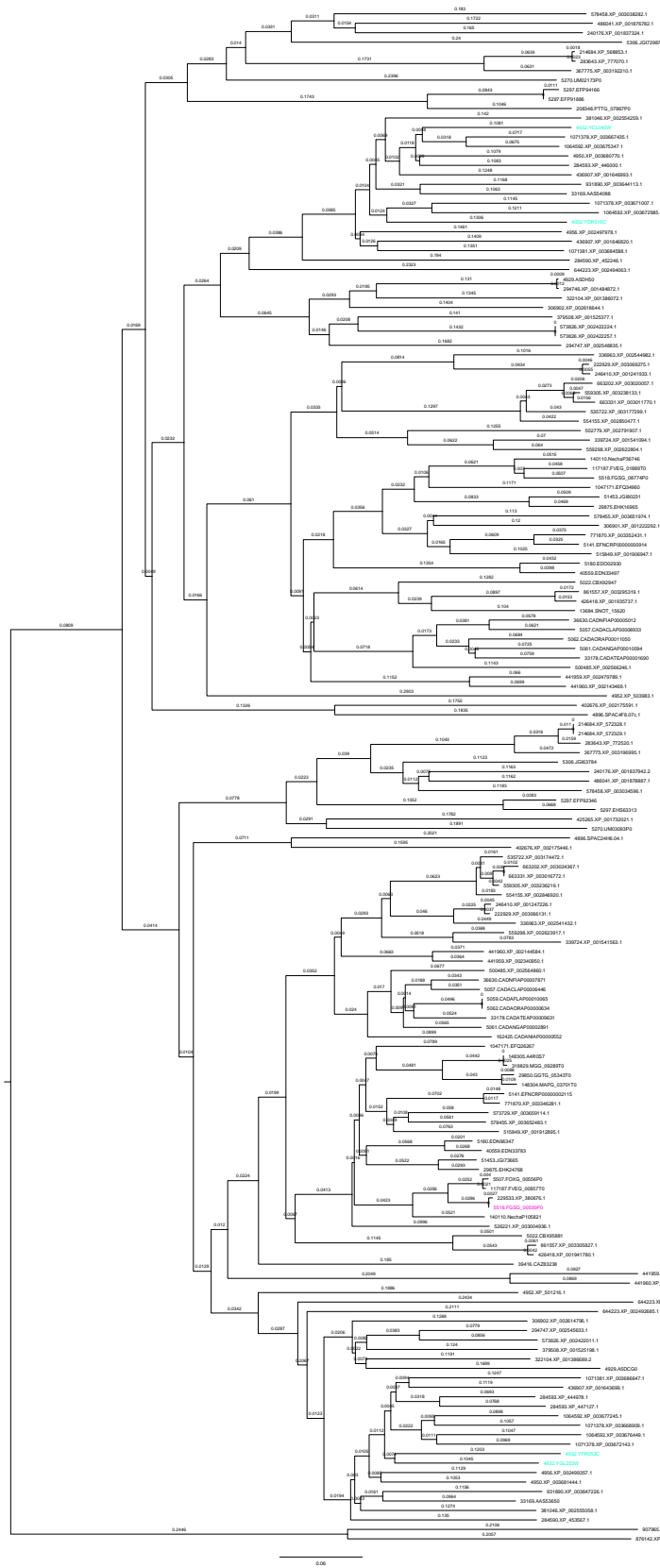


Figure 3: Midpoint rooting including branch lengths. Multiple sequence alignment with clustalw2. Teal represents yeast proteins, Pink represents protein of unknown function in *Fusarium graminearum*