**Yeast ORFan/GUF project**

<http://www.yeastorfanproject.com/>

Funded by the National Science Foundation.

Contact: Jill Keeney, Juniata College Huntingdon, PA keeney@juniata.edu

**The Yeast ORFan Gene Project is a consortium of undergraduate researchers and faculty teaching undergraduates to coordinate resources and design strategies to assign molecular functions to genes of unknown function in the model organism *S. cerevisiae* (Baker’s yeast).**

Background: When the genomic sequence of the model eukaryote Saccharomyces cerevisiae (Baker’s or Brewer’s yeast) was completed in 1996, the expectation was that an understanding of the integrated functioning of the collection of genes in this single-celled eukaryote would shortly follow.  Despite almost 20 years of intense collaborative effort among yeast researchers, nearly 10% of open reading frames (ORFs) are considered uncharacterized.  Determining the function of these orphan genes (ORFans) will require mining the current yeast genomic data, compiled in the [Saccharomyces genome database](http://yeastgenome.org/) (SGD), to most effectively design ORF-specific experiments in cell and molecular biology, and comparative genomics. The “Yeast ORFan Gene Project” is a network of yeast researchers and educators with a focus on teaching experimental design to students and to overcome the challenges and absorb the risks of researching individual ORFans. The major activity of the network is week-long summer workshops for faculty to collaborate on adapting bioinformatic and wet-lab modules for use in classes at their home institutions. Members of the network have access to modules and protocols developed at the workshops, and a pre- and post-course assessment.

**Example ORFan/GUF for an introductory level course**

YLR064W, PER33, ORF, Verified

Protein that localizes to the endoplasmic reticulum; also associates with the nuclear pore complex; deletion extends chronological lifespan; highly conserved across species, orthologous to human TMEM33 and paralogous to Pom33p; protein abundance increases in response to DNA replication stress.

**Example ORFan/GUF for an upper level course**

YHR033W ORF, Uncharacterized

Putative protein of unknown function; epitope-tagged protein localizes to the cytoplasm; YHR033W has a paralog, PRO1, that arose from the whole genome duplication.

Select one of these and work through the tools on the back of this sheet.

For convenience of accessing links, this handout is at <http://www.yeastorfanproject.com/summer-workshops/>

Try these tools with one of the ORFan/GUFs:

1. Look up the GO terms on SGD (<https://www.yeastgenome.org/>)

Molecular Function

Biological Process

Cellular Component

2. Run the following selected tools:

a. BLAST-P (Module 1). This can be run directly from the YLR064W protein tab on SGD

b. Philius transmembrane predictor (Module 6) [**http://www.yeastrc.org/philius/pages/philius/runPhilius.jsp**](http://www.yeastrc.org/philius/pages/philius/runPhilius.jsp)

[For more comprehensive data, you can search the protein database: <http://www.yeastrc.org/pdr/pages/front.jsp>]

c. Yeast Protein Localization database (YPL) (Module 6)<http://yeastgfp.yeastgenome.org/>

d. Gene mania (Module 8) <http://www.genemania.org/> On the search bar in the upper left corner click on the picture of the *Homo sapiens*, a drop down menu should appear, change the selection to “**Saccharomyces cerevisiae**” and enter your gene name into the search box.

3. From the (very limited) information gathered, what might be the function of this ORF? Is there an experiment you could do to test this hypothesis?

Questions for discussion/thought:

1. What course(s) do you teach? [classroom/lab; what level course; enrollment/section]
2. Do you include a CURE (Course based Undergraduate Research Experience) component?
3. How could you incorporate yeast gene discovery into your course(s)? What knowledge/resources would you need to do so?

**Consider applying to the June 2020 workshop!**