

Gene Evolution Workshop 2023

June 13 and 14, 2023



The Carvunis Lab
■ Change and Innovation in Biological Systems ■ ■ ■

Workshop Team

Anne-Ruxandra Carvunis

Associate Professor
Department of Computational and Systems
Biology
University of Pittsburgh School of Medicine
Pittsburgh, PA
Email: carvunis@gmail.com

Jill Keeney

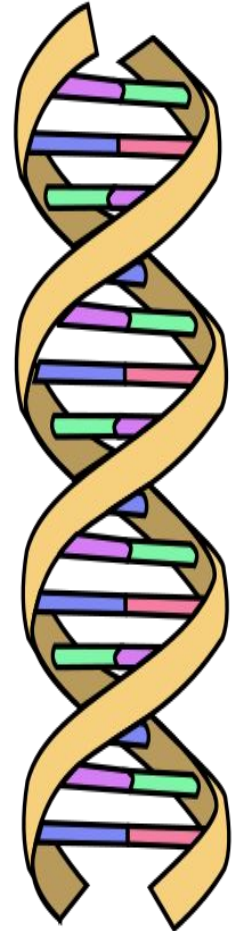
Charles A. Dana Professor of Biology
Juniata College
Huntingdon, PA
Email: keeney@juniata.edu

April Rich

PhD Candidate
Computational Biology
Carnegie Mellon - University of
Pittsburgh
Pittsburgh, PA
Email: aar75@pitt.edu

Nelson Castilho Coelho

Senior Research Specialist
Department of Computational and
Systems Biology
University of Pittsburgh School of
Medicine
Pittsburgh, PA
Email: castilho@pitt.edu



Schedule

June 13th:

- 12:00-1:00 (4-5 UTC) Introduction, background and goals
(Anne, Jill and April)
- 1:00-2:30 (5-6:30 UTC) Work through modules in groups
- 2:30-3:00 (6:30-7 UTC) Break (feel free to “visit” in the main room)
- 3:00-4:00 (7-8 UTC) Continue working through modules in groups
- 4:00-4:30 (8-8:30 UTC) Recap & assign homework

June 14th:

- 12:00-2:00 (4-6 UTC) Work through modules in groups cont.
- 2:00-3:00 (6-7 UTC) Presentations/discussions (in breakout rooms)
- 3:00-4:00 (7-8 UTC) Discussion on how to incorporate in the classroom (all)

RCN-UBE: Yeast ORFan Gene Project

Finding a place for ORFans to GO



Welcome

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

PAGES

- Assessment
- Lab Modules
- Yeast "adopt a proto-gene" project
- Member Affiliations
- Steering Committee

2017-2022



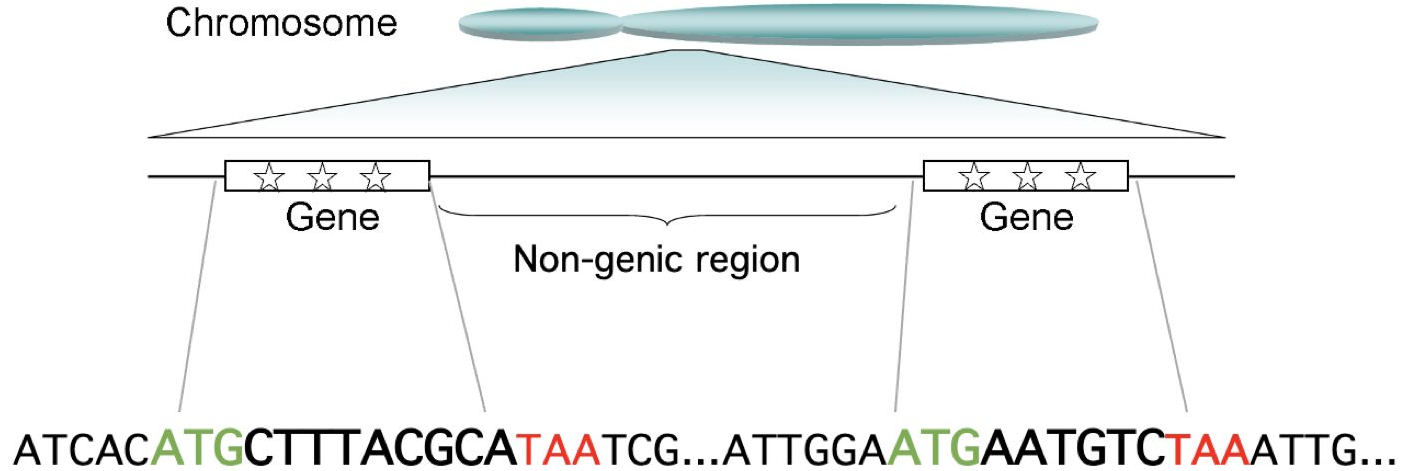
National
Science
Foundation

Adopt a Proto-gene Initiative

2022-2026

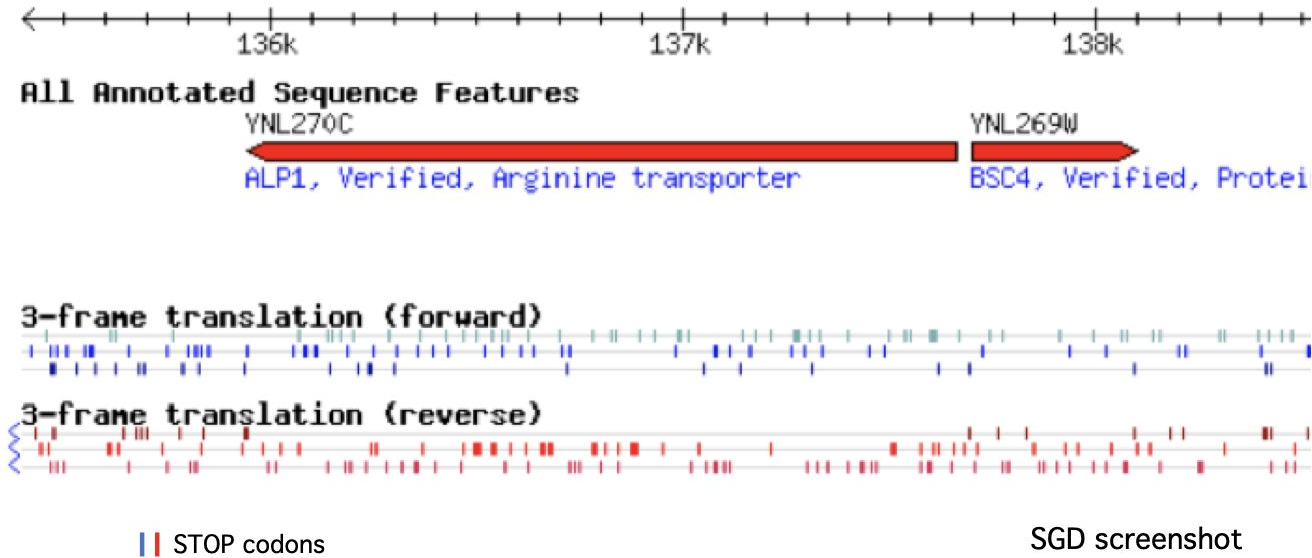
Genome = genes + non-genic sequences

Gene = DNA sequence coding for a functional protein



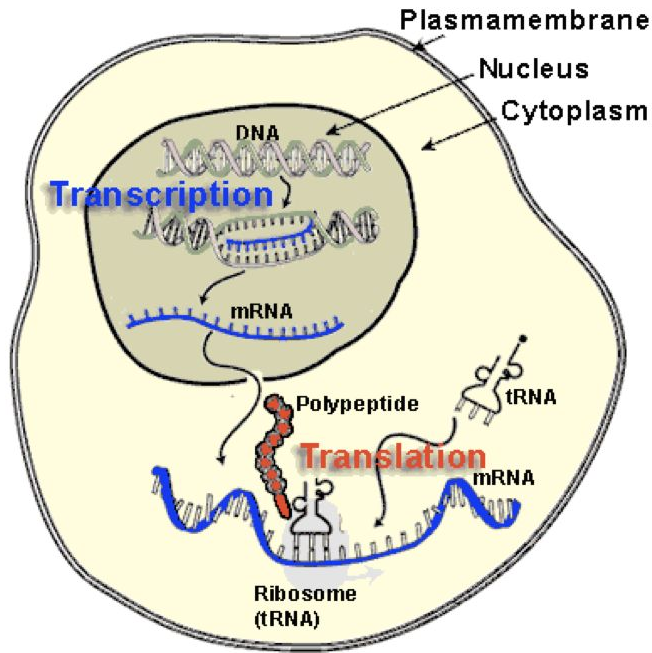
(intergenic, non-coding)

Yeast genome : 6,000 genes and a multitude of random open reading frames (ORFs)

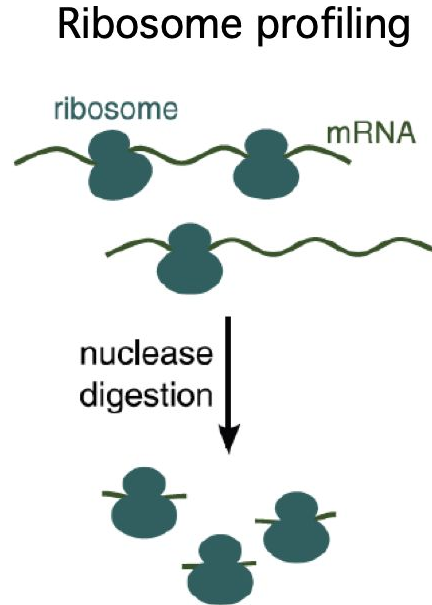


~6,000 protein-coding genes including 700 ORFans (Genes of Unknown Function)
>150K non-genic ORFs

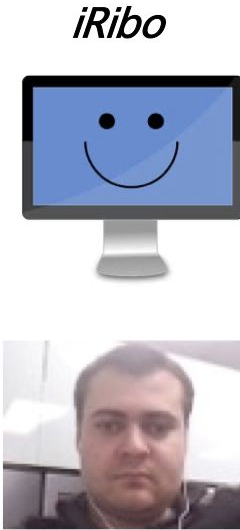
What is translated?



Talking Glossary of Genetics



Ingolia et al., Science 2009

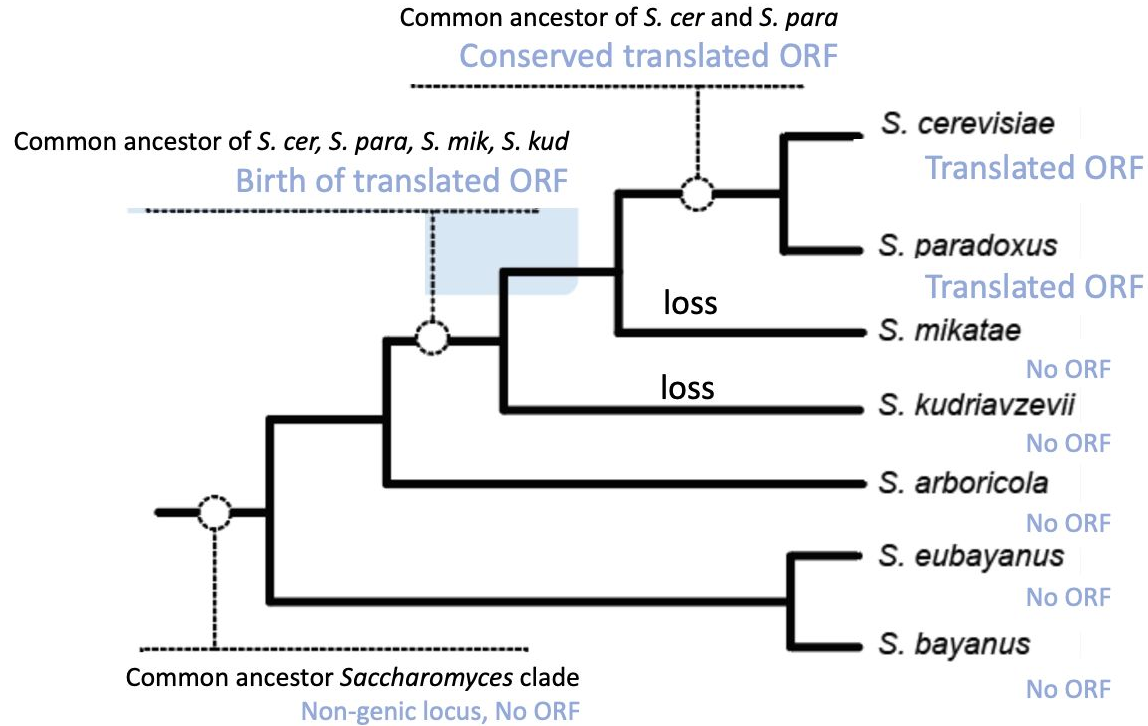


Dr Aaron Wacholder

Wacholder et al, Cell Systems, 2023

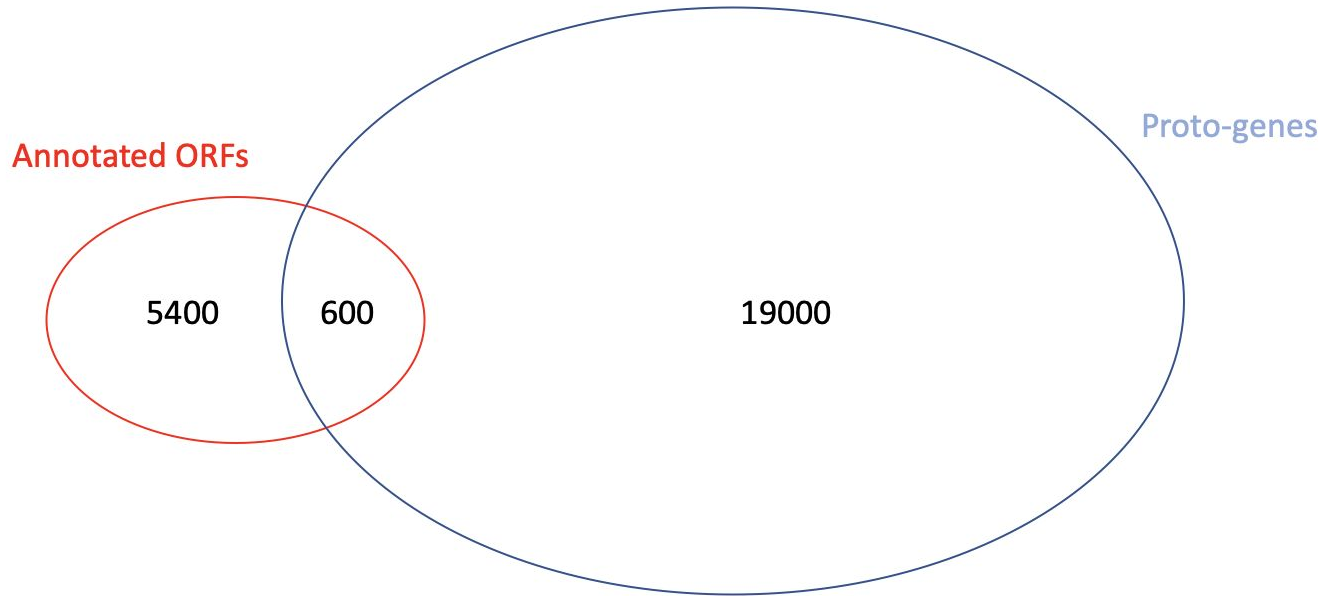
6,000 genes + 19,000 proto-genes

Proto-genes: translated ORFs of recent 'de novo' origins



An immense number of ORFans in need of scientific care!

Help characterize them with your students



Our research: What do they do? How do they evolve? Do they participate in making each species unique?
This workshop: proto-genes as a teaching tool combining computational and evolutionary approaches

Gene evolution: how to access modules and resources

This workshop is an expansion of the Yeast ORFan/GUF project

Modules are accessed at the following web site:

<https://www.yeastorfanproject.com/lab-modules/yeast-adopt-a-proto-gene-project/>

- Five modules
- Available as both PDF and a read-only link to a google doc for download.
- YouTube video walkthroughs of modules will be available later this year.

Workshop goals:

1. Explore an assigned unannotated protogene through completion of the five modules
2. Make a short (1-3 minute) presentation to deliver to your breakout room on Day 2.
 - a. Summary of what was discovered about assigned protogene
 - b. Time and goals for use in a selected course. [Template](#) is provided
 - i. [Alternate template for student participants](#) who are not currently planning for a specific course.
 - c. Leave workshop with a specific plan for incorporating gene exploration in one of your courses and/or research projects
3. Provide us with feedback on the modules

Tuesday schedule

now - 2:30 PM ET

now - 6:30 PM UTC

work through modules in groups
start with Module 1: Genome browser

2:30 - 3:00 PM ET

6:30 - 7:00 PM UTC

break

3:00 - 4:00 PM ET

7:00 - 8:00 PM UTC

continue working in groups but start on the ancestral reconstruction module (module 5)

4:00 - 4:30 PM ET

8:00 - 8:30 PM UTC

Recap & introduce tomorrow's agenda

You do not need to finish all the modules today,
there will be time to work on them tomorrow

The modules

1. Genome Browser
2. Cellular Localization
3. Structure Prediction
4. Coexpression
5. Ancestral Reconstruction

Each module has an associated worksheet with questions for you to answer

Groups are small to encourage discussion and working together

If you have questions



Nelson



John



Carly

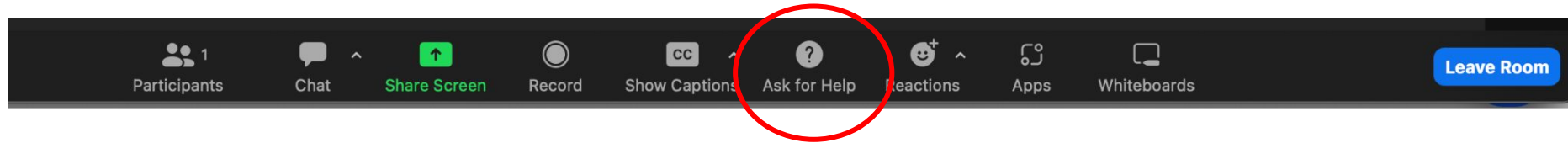


Alistair



April

If you have questions



Slack: for feedback and announcements

- There is a slack channel for each module
 - If you have comments or suggestions, please type any feedback into the corresponding module's slack channel.

`#module1-genomebrowser`

`#module2-localization`

`#module3-structure`

`#module4-coexpression`

`#module5-ancestralreconstruction`

- Any updates or announcements will be posted in the `#announcements` slack channel and also posted in the zoom chat
- You can find your assigned proto-gene in the pinned comment in the `#announcements` slack

Any questions?

type any questions in the chat

Wrap-up Day 1

- Hope everyone made some progress on the modules with their assigned protogene!
- Homework! Download the template for the presentation to your computer, and name the file as orf(your protogene)_first name_last name.
 - For example: orf12345_Jill_Keeney
- Be thinking about what you want to put in it (and start if you want!)
- Day 2, Wednesday:
 - Noon-2 (4-6 UTC): work on modules and your presentation.
 - 2-3 (6-7 UTC): in small groups, present to each other and discuss plans
 - 3-4 (7-8 UTC): all together, report back and discuss