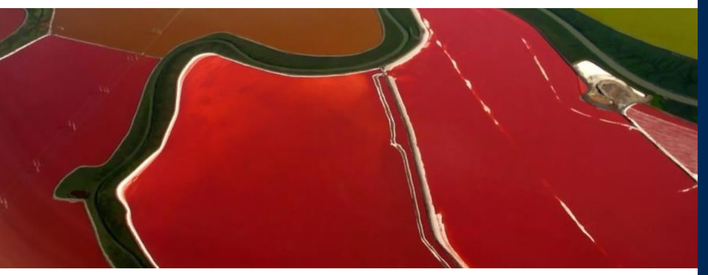
Genome Hunters:

A Quantitative Biology Course-Based Undergraduate Research Experience (CURE)

Robert Furrow, Joel Rodríguez-Medina, Alex Nord, and Mark Goldman

refurrow@ucdavis.edu



Salt ponds in the southern San Francisco Bay

Why a Q-Bio CURE?

Biology undergraduates vary in their attitudes towards and anxiety about math.[1,2] However, modern biological problems require strong foundational skills in math and data literacy.[3,4] An introductory Q-Bio CURE may **improve student learning,[5] build positive attitudes,[6,7] and prepare and retain a diverse cohort of students[8]** for careers in quantitative biology.

Course Learning Goals

> Cultivate **positive math-biology values**
> Improve ability to **collaborate in diverse teams**
> Develop **model thinking**
> Build **programming skills**

Assessing Attitudes (Pre-Post)

> Math-biology values[9], math anxiety[10], and programming anxiety[11]
> Science attitudes (CURE survey)[12]

Take a picture to see references and sample course materials

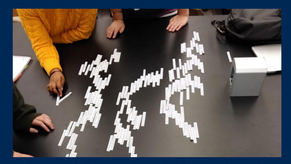
https://tinyurl.com/GenomeHunters

Topic: Halophile Genomics

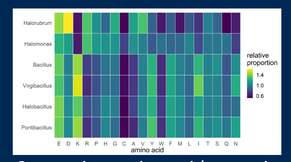
Students: 1st year undergraduates
Fall: culture microbes, quantitatively characterize halotolerance
Spring: bioinformatic analysis of the whole-genome sequences

Quantitative Labs (weeks 1-6)

1. Explore tangible example of concept
2. Design algorithm in small groups
3. Implement using code in R
4. Apply to novel microbial genomes



Assembling physical reads of text from Harry Potter



Comparing amino acid usage in proteomes of halophiles

bp_match_169	N50	L50	betaine_choline_r.total_genes_all_sol.GC
Halomonas citaheneis	219081	6	9	1013	62.12
Halorubrum terrestre	139206	10	NA	NA	65.29
Virgibacillus halodenitrificans	67417	155	20	1847	37.39
Halobacillus litoralis strain SL-4	1038528	2	14	1332	43.59
Halobacillus litoralis	396,879	3	16	1,266	43.83
Bacillus hwajinpoensis strain	797473	3	16	1374	40.22
Halomonas citaheneis	422380	4	9	1025	62.12
Halobacillus litoralis	237386	5	10	1254	47.22
Halomonas alkalicarsica	316952	4	7	1596	58.54
Pontibacillus yanchengensis	575184	3	16	1272	38.43
Halobacillus litoralis	690789	2	15	1471	47.61

Updating a shared google doc with the results of each week's analysis

Student-designed Projects (weeks 4-10)

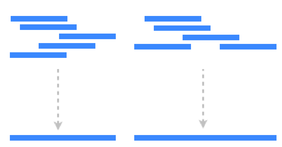
Peers use rubrics to review:

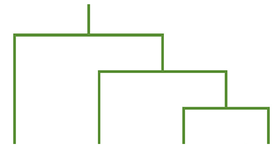
- initial idea pitches
- early results
- draft of full manuscript

Small-group workshops build skills:

- structuring an academic paper
- literature review[13]
- presentation best practices

Example Lessons





Genome assembly

1. Assemble physical "reads" of text
2. Formalize assembly algorithm
3. Use scaffolded R notebook to assemble example DNA reads and calculate coverage
4. Assess assembly quality of course genomes

Clustering

1. Visually cluster genomes by amino acid usage
2. Define a metric for distance btwn genomes
3. Use hierarchical clustering in R with their distance metric
4. Cluster the course genomes to infer halotolerance strategies

What's Next?

> Expand CURE to multiple sections
> Focus specifically on **math-biology self-efficacy**
> Bring in **grad student and postdoc mentors[7]**
> **Flip more coding material** using Swirl[14]

Acknowledgements

Sarah Andrews, Miranda Chen, and Nicole Chodkowski provided feedback on assessments. Marc Facciotti, John Albeck, Hyunsoo Kim, and Joel Rodriguez-Medina taught and provided input on assessment design and quantitative material. This project was funded by an HHMI grant to Mark Goldman.

Your Feedback

Thank you for any suggestions, questions, and ideas. Please swing by the July 27th community hour, 5-6pm ET, or email me (refurrow@ucdavis.edu) with any feedback.