

Required Parameters – What Does it Take to Bring Bioinformatics into the Classroom at the National Level?

J. Williams*,†, M. Spector*, C. Ghiban*,‡, and D. Micklos*,†
 *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, †CyVerse, T.W. Keating Bioresearch Building

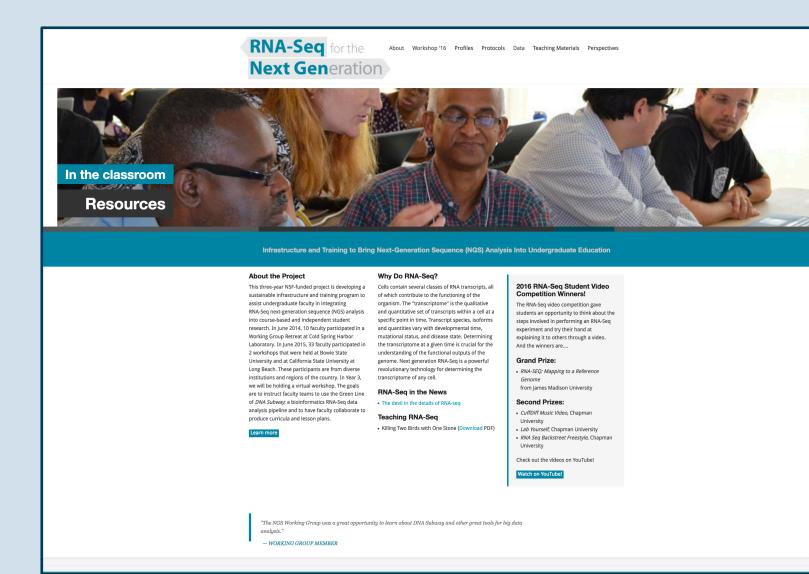
Bioinformatics is as essential to modern biology as the microscope, but most biologists (and biology students) face a training gap that separates them from the skills needed to analyze data, or to be informed collaborators. How can we scale training and resources to effect national impact?

RNA-Seq for the Next Generation

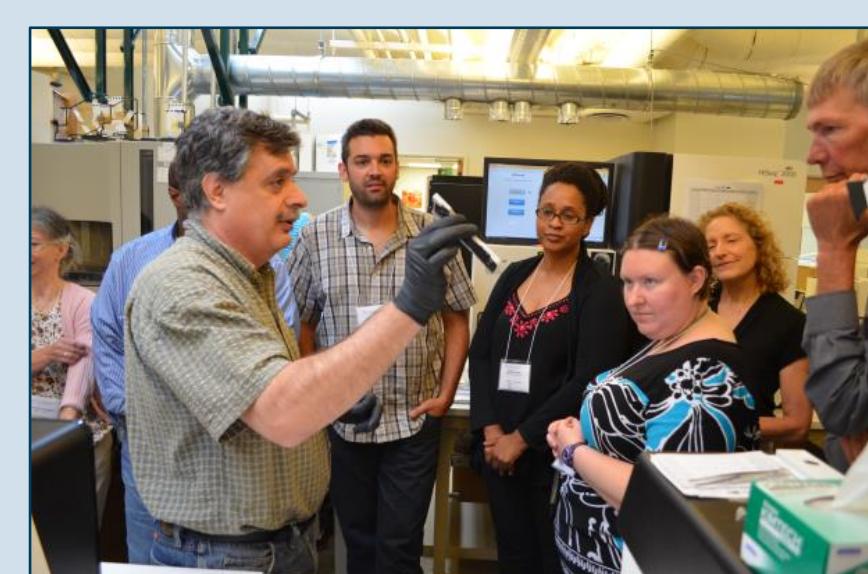
We developed *RNA-Seq for the Next Generation* (www.rnaseqforthenextgeneration.org) to introduce faculty and students to high-throughput (next generation) sequencing – exploring several important biological concepts and techniques, by making use of sophisticated bioinformatics and computation. The project's three years engaged faculty across the US in three cohorts (initial working group, in-person, and virtual dissemination). Learning materials and RNA-Seq pipeline are freely available online.



Faculty Locations



Project Website

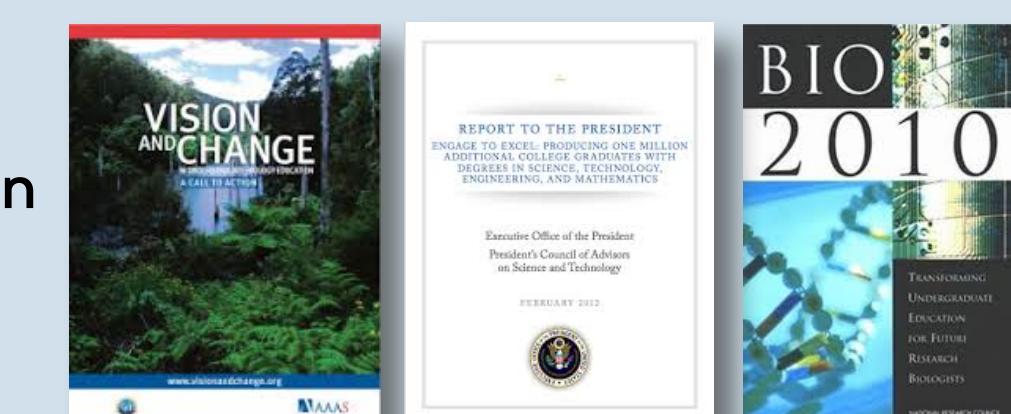


In-person Workshop

CUREs – the “Gold Standard” of STEM Education

The emerging consensus of science education over the last decade is that undergraduate education in Science, Technology, Engineering, and Mathematics (STEM) should work towards replacing traditional laboratory education with discovery-based research courses¹. Four key elements of this reform include:

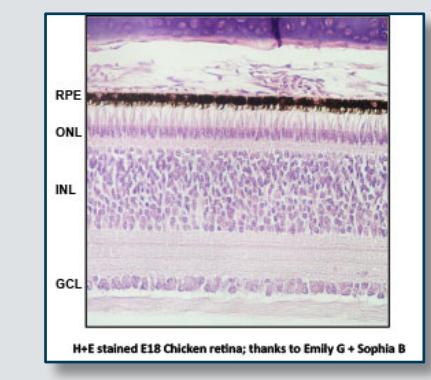
1. Focus on the first two years of college
2. Foster conceptual understanding and practice, rather than memorization
3. Inquiry-centered approaches that begin with students' own questions
4. Increase opportunities for interdisciplinary and collaborative work



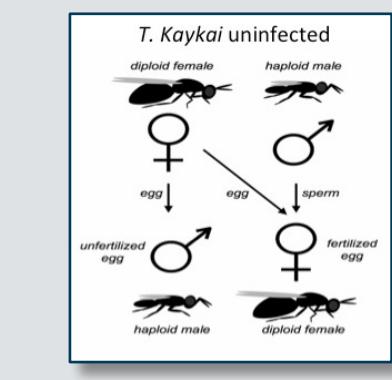
A CURE (course-based research experience) scales open-ended research questions into collaborative projects that involve large numbers of students. CURE benefits include increased participation from groups that may not have previously considered a STEM career, and increased graduation and retention rates in STEM degrees².

Lessons Learned from the *RNA-Seq for the Next Generation* Project

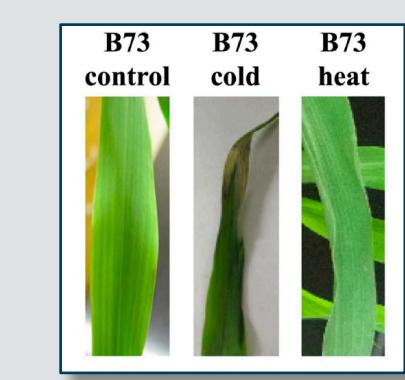
--CURE-based-model



Ray Enke
James Madison University, VA
retina development in chicken



Latanya Hammonds-Odie
Georgia Gwinnett College, GA
parthenogenesis induction in wasps



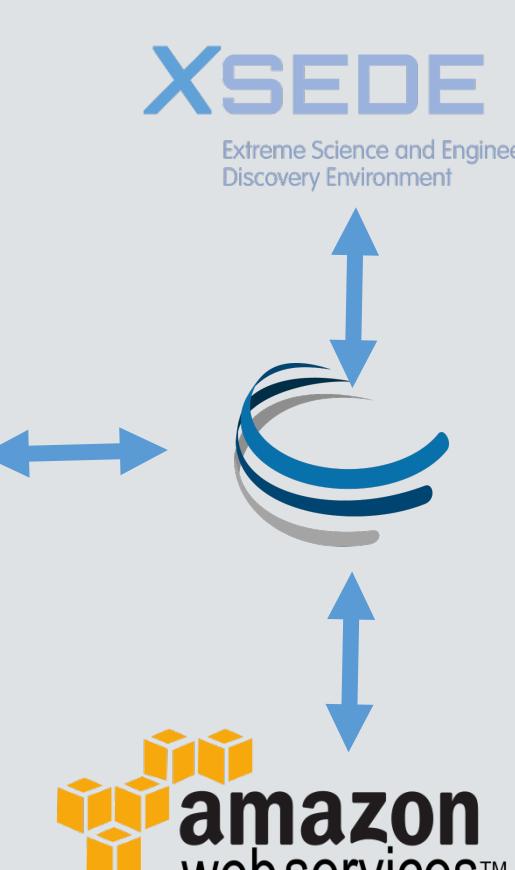
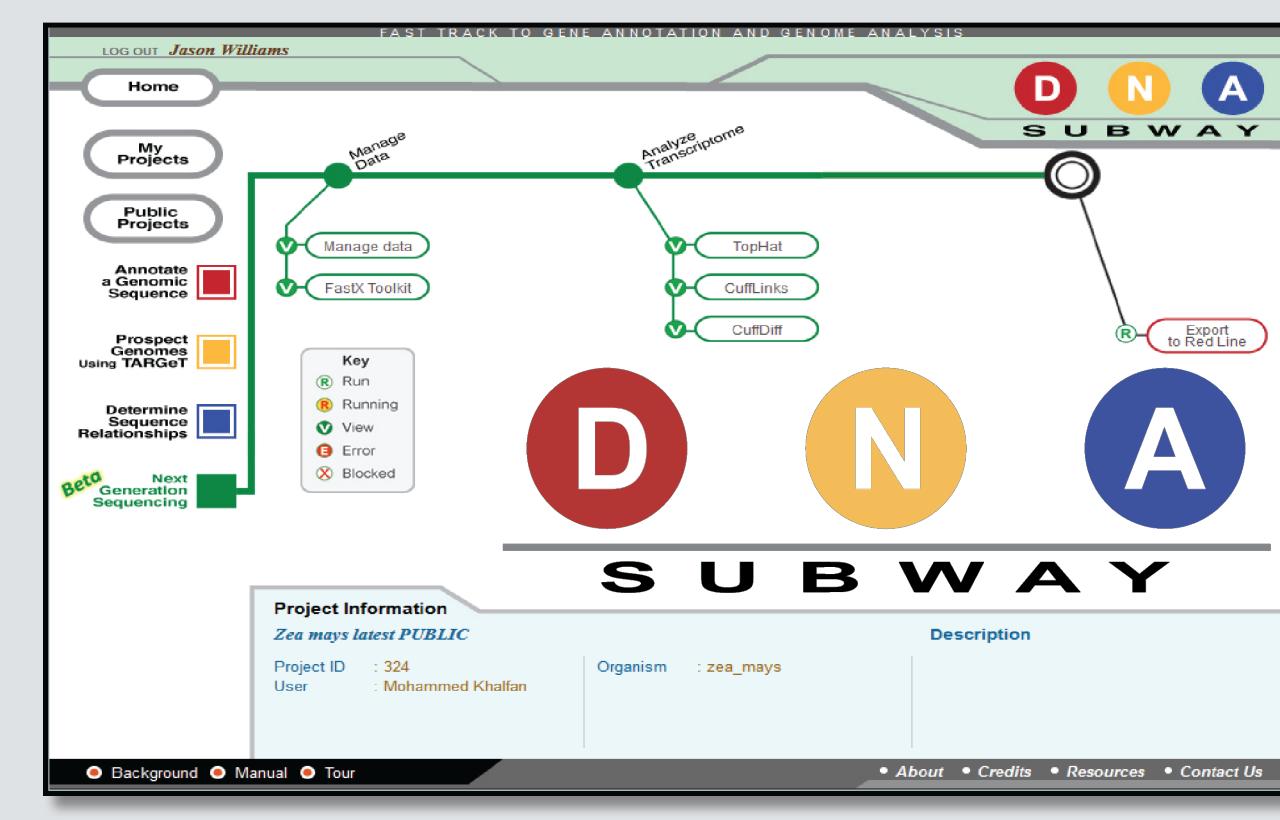
Irina Makarevitch
Hamline University, MO
abiotic stress responses in Maize



Richard Barker
University of Wisconsin-Madison, WI
Arabidopsis responses to micro-gravity

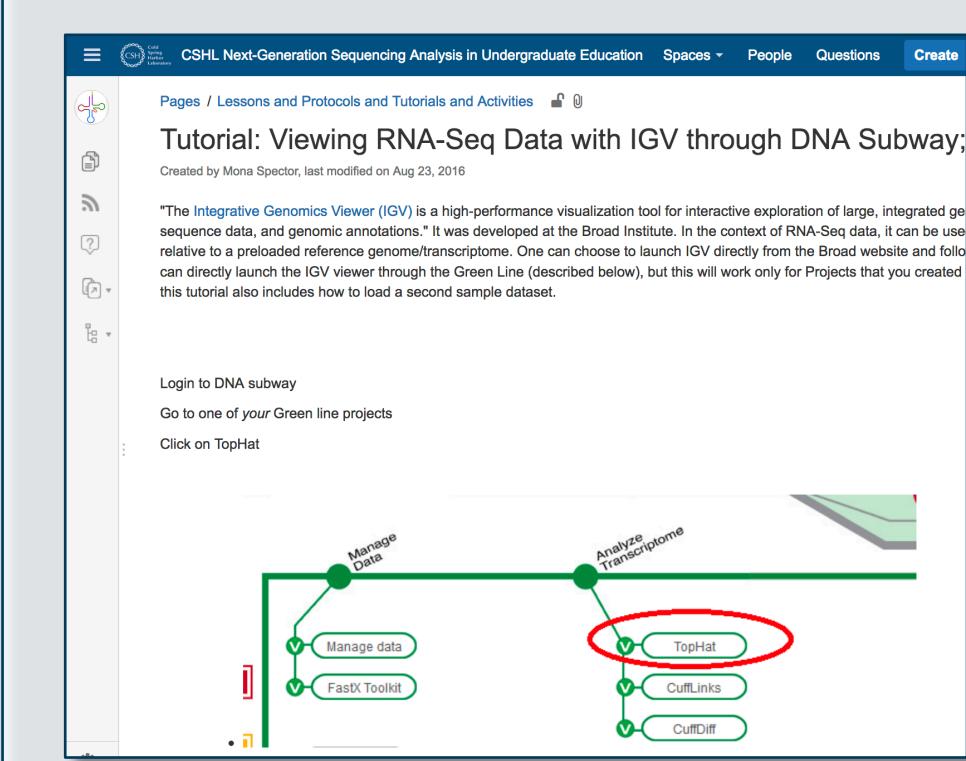
- Discovery-based research questions
- Collaborative effort and team participation
- [wet-lab] hands-on; students and faculty care about data generated
- [multiple entry levels] student participation at multiple levels of preparation (freshman-senior)

--interfaces-and-computation



- [Command-line] may reach better-prepared faculty and students, but GUI required to reach a large audience
- Use open-source tools like CyVerse, Galaxy, Jupyter Notebooks;
- Organize methods to serve the needs of effective teaching
- Computational capacity for large numbers of simultaneous users; [cloud-based] solutions drive adoption

--documentation-and-training



- Bioinformatics training is a weak point in faculty preparation⁴; diverse support mechanisms required
- Extensive documentation drives use and adoption
- Feedback and makes materials classroom-ready
- [in-person] training develops an enthusiastic and supportive community

With guidance, students can participate in the generation of RNA for sequencing, and faculty are able to generate data related to their own research investigations. RNA-Seq datasets can be explored by large numbers of students who do not need to complete every aspect of the investigation to benefit. Ultimately students can be co-investigators on published datasets.

DNA Subway, a ‘classroom-friendly’ web-based GUI was the interface to centralized computing through CyVerse-developed Agave API. DNA Subway connected users with resources available at the Texas Advanced Computing Center (TACC) and is extensible to AWS. We did not integrate a CLI into DNA Subway, but provided training in R and RStudio to enable advanced analyses.

Approaches

Data generation is complex and may be expensive to scale; still, students can explore novel aspects of the data. The many steps of RNA-Seq are conceptually challenging, so many aspects of the analysis may need extensive preparation to be covered successfully.

Developing a totally novel user interface is resource intensive, but rewarding for projects that attempt to cultivate wide-spread adoption. Interface implementation limits our ability to adapt as new software emerges. Development of a robust pipeline requires major effort.

Extensive documentation was developed through an easy-to-update project wiki. Polished teaching materials were iteratively developed and published to a public site. Monthly virtual meetings of the working group kept faculty engaged and participating. Assessment component included feedback that addressed both technical and pedagogical challenges.

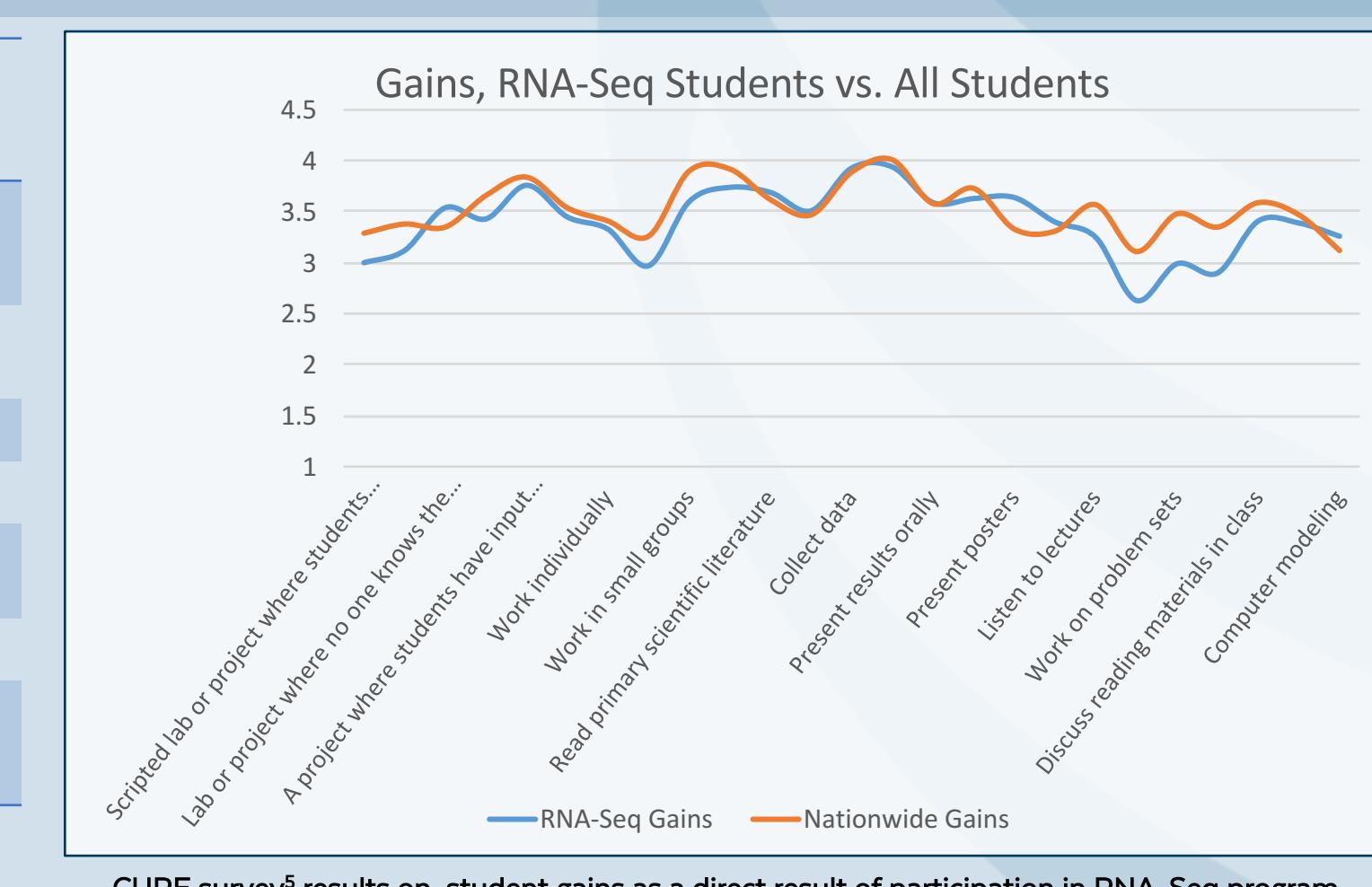
Challenges

Documentation is almost always out-of-date, so constant updating is required. In-person workshops are (in our opinion) necessary, but may be expensive and time-consuming to coordinate. Faculty enthusiasm for learning about bioinformatics is high, but hand-holding will be needed at the beginning!

Results and Impact

Task/Tool	Z	P	Pre-workshop μ	Post-workshop μ
Uploading data to DNA Subway	-4.186	.000	2.22	4.04
Using FastX tool kit	-4.323	.000	1.63	4.15
Using TopHat	-4.239	.000	1.74	4.08
Using Cuffdiff	-4.236	.000	1.70	4.04
Visualizing reads using IGV	-3.685	.000	1.73	3.35
Building graphs in R	-3.747	.000	1.81	3.35
Conducting a pathway analysis	-4.161	.000	1.78	3.46

Changes pre/post-workshop in bioinformatics teaching confidence using Wilcoxon signed-rank test (n = 23, Year-1/2 faculty, 87% respondents)



Teaching materials and faculty profiles online at:
www.rnaseqforthenextgeneration.org

- [1] President's Council of Advisors on Science and Technology, Engage to Excel: Producing One Million Additional College Graduates With Degrees in Science, Technology, Engineering, and Mathematics. Washington, D.C.; 2010.
- [2] Rodriguez A, et al. 2013. Simmons SL, Dolan EL. Early Engagement in Course-Based Research Increases Graduation Rates and Completion of Science, Engineering, and Mathematics Degrees. *CBET Life Sci Educ*. 2013;15(2):20.
- [3] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562-578.
- [4] Cummings MP, Temple GG. Broader incorporation of bioinformatics in education: opportunities and challenges. *Brief in Bioinfo* 2010; 11(6): 537-543.
- [5] CURE Survey: <https://www.grinnell.edu/academics/areas/psychology/assessments/cure-survey> (Accessed, October, 2016).

- Trained 84 faculty across the US – mostly at small colleges and/or primarily undergraduate institutions.
- Analyzed ~990 Gigabases of sequence data across 36 eukaryotic organisms.
- 6 publications involving faculty-student research
- Reached more than 1,370 students through faculty incorporation of datasets and bioinformatics analyses.