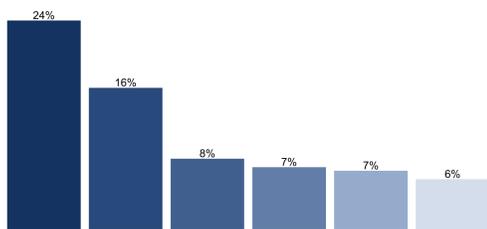


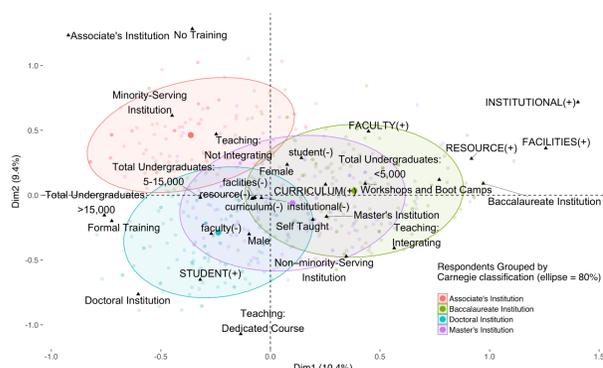
Challenges

Faculty do not have the training to bring bioinformatics into the classroom

In 2016, NIBLSE (Network for Integrating Bioinformatics into Life Sciences Education) conducted the largest-ever U.S. survey on bioinformatics in undergraduate education, collecting more than 1200 responses. Although 95% of faculty believe that bioinformatics should be included in teaching, only 40% did so. The survey identified key barriers to including bioinformatics in the curricula, many of which were reported disproportionately for underrepresented minority faculty and faculty at Minority-serving Institutions (MSIs).



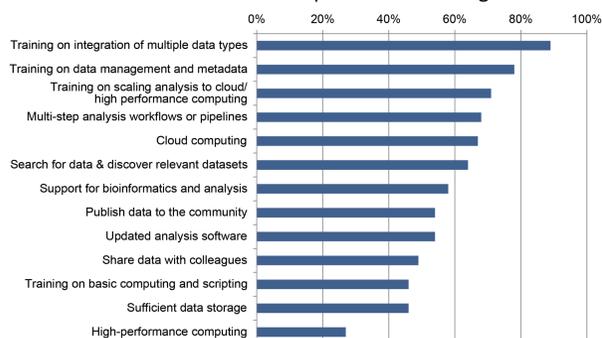
NIBLSE Survey: Faculty-identified barriers to integrating bioinformatics – training is the barrier



NIBLSE Survey: Multiple correspondence analysis of barriers by institution type – MSIs get left behind

Training in computational skills is the most unmet need for NSF-funded researchers

In 2016, the CSHL DNA Learning Center conducted a needs analysis as part of outreach for CyVerse and obtained 704 responses from NSF-funded PIs across all BIO directorates. Out of more than a dozen areas where investigators reported inadequacies, the top three unmet needs were for computational training.



CyVerse Survey: Physical infrastructure is not the problem, training on how to use it effectively is

Opportunities

Genomics presents unique opportunities to increase participation of underrepresented minorities in science. Advancing technologies enable students and teachers to work with the same data, the same tools, at the same time as researchers. Exploiting these opportunities in the context of course-based research experiences (CUREs) increases student retention in STEM and will usher in the next generation of diverse researchers.

CyVerse: Egalitarian data-driven discovery



Initially funded by NSF as the iPlant Collaborative in 2008, CyVerse is now a community of over 40,000 users in the US and abroad. CyVerse makes computational tools and resources broadly available to researchers that would not otherwise have access. The result of this investment, is that less resourced institutions have “a seat at the table” – the technology and training they need to participate in the biological big data revolution.

(NSF DBI 0735191 and DBI 1265383)



CyVerse Tools and Services: A “full-stack” cyberinfrastructure allows easy data sharing, access to cloud, and high-performance computing through easy-to-use web interfaces

93% of biology researchers work with, or will be working with large genomics datasets.

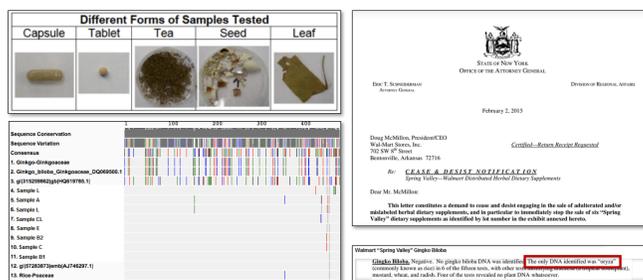
- CyVerse surveys 2008-2017

Case studies: Cyberinfrastructure as the enabler of CUREs



This “citizen science” program is an approachable way to introduce students to molecular biology and bioinformatics. Using the CyVerse’s DNA Subway, students can analyze and identify unknown organisms, foods, and medicines. Novel barcodes are submitted to GenBank with students credited as submitters.

- Since 2012, UBP and a companion research program has involved **1051** students in **384** teams in students across New York City
- Using CyVerse DNA Subway, UBP students have generated **94** novel sequence submissions to Genbank and generated more than **10,900** sequencing reads
- **32%** average minority participation rate since 2012



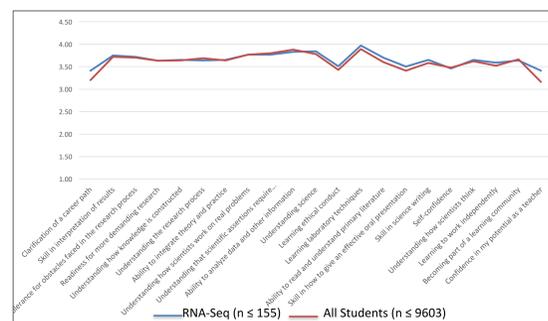
The first UBP competition winners from Hostos-Lincoln Academy of Science, Bronx, NY (pictured above) discovered medicinal “gingko” products contained only rice. Similar findings by the NY State Attorney General confirmed these results two years later.

RNA-Seq for the Next Generation



This three-year NSF-funded project developed sustainable infrastructure and training to assist undergraduate faculty in integrating RNA-Seq and next-generation sequencing (NGS) analysis into CUREs. Using CyVerse’s DNA Subway, participants analyzed data using XSEDE high-performance supercomputing. (NSF DUE 1323522)

- Trained **82** faculty across the US – mostly at small colleges and/or primarily undergraduate institutions
- Produced **53** novel datasets, and analyzed more than **990** gigabases of sequence data across **36** eukaryotic organisms
- Reached more than **1,600** students through incorporation of datasets and bioinformatics analyses into **28** courses

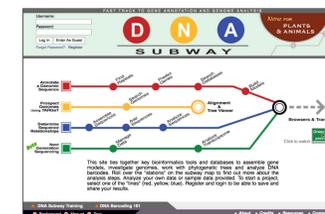


CURE survey results (attitudinal and learning gains) as a direct result of student participation in the RNA-Seq for the Next Generation program

Projects and Resources



www.dnasubway.org



CyVerse’s DNA Subway bundles research-grade bioinformatics tools, high-performance computing, and databases into web-based, classroom-friendly workflows. “Riding” DNA Subway lines, students can predict and annotate genes in up to 150kb of DNA (Red Line), identify homologs in sequenced genomes (Yellow Line), identify species using DNA barcodes and phylogenetic trees (Blue Line), and examine RNA-Seq datasets for differential transcript abundance (Green Line).



qubeshub.org/groups/niblse

The Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE; pronounced “nibbles”) is an NSF Research Coordination Network for Undergraduate Biology Education (RCN-UBE). The long-term goal of NIBLSE is to establish bioinformatics as an essential component of undergraduate life sciences education.

(NSF RCN UBE 1539900)



DATA CARPENTRY
BUILDING COMMUNITIES TEACHING UNIVERSAL DATA LITERACY

www.datacarpentry.org

Originally supported as a collaboration of NSF Bio Centers, Data Carpentry targets learners who have little to no prior computational experience and develops and teaches workshops on the fundamental data skills needed to conduct research.

Connect!



Find me at the summit or contact me:
Williams@cshl.edu



http://tiny.cc/includes_2017

This poster includes the work of several people and projects. Download a copy of the poster and some additional papers/resources