It’s Impossible to Keep Up
Career-spanning Learning in the Life Sciences

Jason Williams
Cold Spring Harbor Laboratory, DNA Learning Center
@JasonWilliamsNY
Cold Spring Harbor Laboratory
CSHL DNA Learning Center

Hands on education in molecular biology/bioinformatics for secondary students (Grade 6-12); secondary and undergraduate faculty training; websites/multimedia
CSHL DNA Learning Center

588,416 Annual Exposures

- 31,064 DNALC Lab Field Trips
- 1,347 DNALC summer Campers
- 369 Extended Student Research
- 128,609 Labs at Licensed and Modeled Programs Worldwide
- 364,218 DNALC commercial Kits
- 62,809 Online DNALC Sequence Analysis Projects
Keeping up with the future...
The future is computational
Help researchers and educators learn how to use data and computation
The skillset is changing
Biology skills timeline

1980s:
- RFLP
- PCR

1990s - 2000s:
- Microarrays
- RNAi
- RNA—Seq
- CRISPR
- BLAST
- Databases

2010 and beyond?
- Big Data
- Cloud/HPC
- Machine Learning
- SKILL GAP
I went to school for...

I need to know this for my research...

```python
import urllib2
efetch = 'efetch.fcgi?'
s = eutils + efetch
targets = ['J04243', 'M60064']
idString = 'id=' + ','.join(targets)
s += idString + '&db=nucleotide&rettype=fasta'
fileObject = urllib2.urlopen(s)
data = fileObject.read().strip()
entries = data.split('\n\n')
title, sequence = entries[0].split('\n', 1)
print title.split(' ',1)[0]
```

# prints:
# >gi|154102|gb|J04243.1|STYHEMAPRF
A scientific digital divide?
digital divide affecting education even more
"... skill demands in STEM occupations have changed especially quickly. The faster rate of change in STEM is driven both by more rapid obsolescence of old skills and by faster adoption of new skills."
Training is the biggest need

- Training on integration of multiple data types
- Training on data management and metadata
- Training on scaling analysis to cloud/high performance computing
- Multi-step analysis workflows or pipelines
- Cloud computing
- Search for data & discover relevant datasets
- Support for bioinformatics and analysis
- Publish data to the community
- Updated analysis software
- Share data with colleagues
- Training on basic computing and scripting
- Sufficient data storage
- High-performance computing
95% of respondents indicate that bioinformatics should be integrated into the life science curriculum; 32% of faculty report achieving this.
## New faculty aren’t integrating

<table>
<thead>
<tr>
<th>Decade of Highest Degree</th>
<th>Formal Bioinformatics Training (%)</th>
<th>Faculty Integrating Bioinformatics (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980-1989</td>
<td>8.4</td>
<td>35.4</td>
</tr>
<tr>
<td>1990-1999</td>
<td>11.3</td>
<td>41.9</td>
</tr>
<tr>
<td>2000-2009</td>
<td>35.1</td>
<td>41.7</td>
</tr>
<tr>
<td>2010-2016</td>
<td>48.3</td>
<td>25.2</td>
</tr>
</tbody>
</table>
These gaps multiply and perpetuate

<table>
<thead>
<tr>
<th>REALITY</th>
<th>EQUALITY</th>
<th>EQUITY</th>
<th>JUSTICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>One gets <strong>more than</strong> is needed, while the other gets <strong>less than</strong> is needed. Thus, a huge disparity is created.</td>
<td>The assumption is that everyone benefits from the same supports. This is considered to be equal treatment.</td>
<td>Everyone gets the support they need, which produces equity.</td>
<td>All 3 can see the game without supports or accommodations because the cause(s) of the inequity was addressed. The systemic barrier has been removed.</td>
</tr>
</tbody>
</table>

Everyone is **INCLUDED** in the game. **No one** is left on the outside; we didn’t only remove the barriers keeping people out, we made sure they were valued & involved.
How to solve these problems?
Improve the quality and application of professional development
Null effects of boot camps and short-format training for PhD students in life sciences

David F. Feldon, Soojieong Jeong, James Peugh, Josipa Roksas, Cathy Maahs-Fladung, Alok Shenoy, and Michael Oliva

aDepartment of Instructional Technology & Learning Sciences, Utah State University, Logan, UT 84322-2830; bDepartment of Pediatrics, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH 45229-3026; cDepartment of Sociology, University of Virginia, Charlottesville, VA 22904; and dCurry School of Education, University of Virginia, Charlottesville, VA 22904

Edited by Dale Purves, Duke University, Durham, NC, and approved July 28, 2017 (received for review April 6, 2017)
A minimal set of principles for effective, inclusive, and career-spanning learning
“The unicycle” – fine for going alone
“The bicycle” – good for going far

CORE PRINCIPLES

Inclusive
Catalytic
Best Evidence
Effective

COMMUNITY PRINCIPLES

Reach
Scale
Sustain
The Bicycle Principles for Effective, Inclusive, and Career-spanning Short-format Training

Improving Professional Development in the Life Sciences and Beyond

Announcement
November 2022

We’re collecting feedback: Let us know what you think about the recommendations to improve short-format training. We will be conducting surveys and focus groups from now through February 2023. Participants will be compensated for their time.
bioRxiv posts many COVID-19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

Optimizing Short-format Training: an International Consensus on Effective, Inclusive, and Career-spanning Professional Development in the Life Sciences and Beyond


doi: https://doi.org/10.1101/2023.03.10.531570

This material is based upon work supported by the National Science Foundation under DRL/EHR:2027025.
Example recommendations
A. Professionalize the training of short-format training instructors and instructional designers
D. Operationalize equitable and inclusive practice in short-format training as an ethical obligation

Large crowd of people having fun
E. Deploy short-format training to counter inequity

Judge in robe fighting for justice
K. Communicate standards of instruction through badging
Support infrastructure that works for research and education
foss.cyverse.org

Spring 2023 FOSS workshop will be virtual on Thursdays 11:00AM - 1:00PM US Arizona Mountain Standard Time starting January 19th, 2023

CyVerse's 10-week virtual workshop teaches you the principles, practices, and how-tos for doing collaborative open science using cutting-edge, open source cyberinfrastructure.

To see how our FOSS workshop can support your work, check out the workshop curriculum over the years:

<table>
<thead>
<tr>
<th>Name</th>
<th>Dates</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring FOSS 2023</td>
<td>Jan 19 - Mar 30, 2023</td>
<td>Fifth virtual workshop series</td>
</tr>
<tr>
<td>Fall FOSS 2022</td>
<td>Sept 15 - Nov 18, 2022</td>
<td>Fourth virtual workshop series</td>
</tr>
<tr>
<td>Fall FOSS 2021</td>
<td>Sept 7 - Nov 18, 2021</td>
<td>Third virtual workshop series</td>
</tr>
<tr>
<td>Spring FOSS 2021</td>
<td>Feb 9 - Apr 21, 2021</td>
<td>Second virtual workshop series</td>
</tr>
<tr>
<td>Summer FOSS 2020</td>
<td>July 28 - Nov 3, 2020</td>
<td>First virtual workshop series</td>
</tr>
<tr>
<td>Spring FOSS 2020</td>
<td>Feb 17 - 21, 2020</td>
<td>Second in-person workshop at UArizona</td>
</tr>
<tr>
<td>Spring FOSS 2019</td>
<td>Jun 3-7, 2019</td>
<td>First in-person workshop at UArizona</td>
</tr>
</tbody>
</table>
Change requires more than sharing ideas. We must build communities and support them.
Building community

Why short-format training?

In many areas of the life sciences new technologies and approaches (especially, but not only computational ones) are changing rapidly. It’s not possible for formal training (undergraduate/graduate) to keep pace, but short-format training can fill these gaps. Short-format training comes with its own set of challenges, and this community works together to address them.
I don’t know what the future looks like
I don’t know what the future looks like

Community Discussion – What will bioinformatics training and learning look like in the age of ChatGPT and AI?
April 7, 2023 // No Comments

What will it mean to teach and learn about bioinformatics with the assistance of AI?

Continue reading »

April post/Meeting info
“The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn”

– A. Toffler
Developing Foundations for Nanopore DNA Sequencing Course-based Undergraduate Research Experiences at Minority-Serving Institutions

- Pilot (2-years)
- Simplify lab and bioinformatics protocols
- Support faculty needs and understand barriers to use

DUE# 2216349
DNA Barcoding Programs
Using DNA Barcodes to Identify and Classify Living Things

DNA barcodes allow non-experts to objectively identify species—even from small, damaged, or industrially processed material. A "DNA barcode" is a unique DNA sequence that identifies each living thing. Short DNA barcodes, about 700 nucleotides in length, can be quickly processed from thousands of specimens and unambiguously analyzed by computer programs. With DNA barcodes, we can discover and catalog biodiversity on our planet using tools developed at the DNA Learning Center.

DNA Barcoding Program Outcomes

- **1,640** Total GenBank Publications
- **146** First GenBank Barcodes
- **279** New Sequence Variants
- **1,331** Unique GenBank Authors
- **544** Total Species Identified
Envisioning the Next Bioscience Workforce: A Summit on Industry Trends and Needs

Monday, June 26, 2023
8:30 a.m. – 6:30 p.m. EDT
NATIONAL ACADEMY OF SCIENCES, WASHINGTON, D.C.