Improving the Bioinformatics Curriculum

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

2018 Galaxy Community Conference / Bioinformatics Open Source Conference – Portland OR, USA
Preparing students and families to thrive in the gene age
The view from Neptune
Bioinformatics skills

- Stats
- Comp. Science
- Biology
- Data Science
- Bioinformatics
- Comput. biology

Image credit: https://genomesjigsaw.wordpress.com/2015/06/27/seq/
Conspiracy Theory:

The bioinformatics curriculum is dominated by unseen forces
“The biggest problem with life science education is the hegemonic control of our curriculum by the medical school overlords”
### What skills are in the biology toolbox?

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Molecular Skills</th>
<th>Computational Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980s</td>
<td>RFLP, PCR</td>
<td>Word processing</td>
</tr>
<tr>
<td>1990s - 2000s</td>
<td>Microarrays, RNAi</td>
<td>BLAST Databases</td>
</tr>
<tr>
<td>2010 and beyond</td>
<td>RNA-Seq, CRISPR</td>
<td>SKILL GAP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Big Data, Cloud/HPC, Machine Learning</td>
</tr>
</tbody>
</table>

- **Molecular**
  - RFLP
  - PCR
  - Microarrays
  - RNAi
  - RNA-Seq
  - CRISPR

- **Computational**
  - Word processing
  - BLAST
  - Databases
  - SKILL GAP
  - Big Data
  - Cloud/HPC
  - Machine Learning
People should be the target to scale

1980s:

2018:
How can we...

Accelerate science?
Solve important problems?
Make people happier?

Answer: Teach computational skills
93% of survey respondents indicate they are or will be working with large datasets
“How do you rate your level of bioinformatics skills?”

- Beginner (or don't use bioinformatics): 53%
- Intermediate: 35%
- Advanced: 12%
Revealing unseen forces

(ἀποκάλυψις)
For some faculty, the computational skills they are looking for might as well be on Neptune.
Network for Integrating Bioinformatics into Life Science Education (NSF-RCN-UBE)

- ~30 Undergraduate educators, experts in assessment, industry, and other fields
- Interested in:
  - Faculty and student preparation
  - Methods and resources for integration
  - Assessment
Largest-ever survey on bioinformatics in undergraduate education (U.S.)
(n ≈ 1260 responses)

• Understand faculty perceptions and behaviors in addressing bioinformatics
  o Is bioinformatics important? Are you including? If not why?

• Gather information on bioinformatics related syllabi
  o What do you teach, when, how?
95% of respondents indicate that bioinformatics should be integrated into the life science curriculum; 40% of faculty report achieving this.
Composite survey respondent:
• white male/female
• self-taught in bioinformatics
• PhD earned in 2000–2009

S/he works at:
• non−minority-serving
• doctoral-granting institution
• Ugrad enrollment < 5,000
Barriers to Integration of Bioinformatics into Undergraduate Life Sciences Education


doi: https://doi.org/10.1101/204420
<table>
<thead>
<tr>
<th>Question</th>
<th>Number of Keyword-coded Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In your opinion, what do you think are the most important challenges currently facing those educating undergraduate life scientists in bioinformatics?</td>
<td>734 (59.6%)</td>
</tr>
<tr>
<td>2A. What is the level of the courses you teach in which you would like to include bioinformatics content? (multiple choice)</td>
<td></td>
</tr>
<tr>
<td>2B. Please describe briefly; include any barriers to development and/or implementation</td>
<td>364 (29.6%)</td>
</tr>
<tr>
<td>3*. What is preventing you from including bioinformatics content in these courses?</td>
<td>313 (25.4%)</td>
</tr>
<tr>
<td>4. At your current institution, do you face any technical barriers in teaching bioinformatics, e.g., availability of a computer lab, different operating systems, access to high performance computing for teaching, IT support? Please describe</td>
<td>511 (41.5%)</td>
</tr>
</tbody>
</table>
Training is the biggest need
41% of respondents who are underrepresented minorities in STEM report training barriers vs. 28% of non-URM faculty
Barriers differ by institution type

- Faculty Issues: Lack of expertise/training
- Student Issues: Lack of background skills/knowledge
- Student Issues: Lack of interest
- Curriculum Issues: Quickly changing technologies

Institution Types:
- Doctoral (n=358)
- Baccalaureate (n=228)
- Master's (n=204)
- Associate's (n=176)
MCA: Institution type
23% of respondents at minority-serving institutions report integrating bioinformatics vs.
43% at non-minority-serving institutions.
Students preparation?

- Student Issues: Lack of background skills/knowledge
- Student Issues: Lack of interest
- Resource Issues: Access to software
- Student Issues: Intimidated by topic
- Institutional Issues: Lack of inter-departmental cooperation

Percentage of respondents

Bioinformatics Integration in Teaching

- Not Integrating Bioinformatics (n=591)
- Integrating Bioinformatics (n=263)
- Dedicated Course (n=132)
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>
MCA: Degree decade
How can we address some challenges?
Meet faculty where they are
Bioinformatics core competencies for undergraduate life sciences education


Published: June 5, 2018 • https://doi.org/10.1371/journal.pone.0196878
NIBLSE Incubator: RNA-seq Analysis

Overview

RNA-seq Analysis

In this computer lab module, students learn how to process an RNA-seq data set to identify differentially expressed genes (DEGs). The samples for this data set were collected from yeast cells expressing either a control gene or a pathogen effector gene. The goal is to identify potential cellular targets of the pathogen.
Reach those who need to be reached
Opportunities to improve careers and outcomes

- **Community colleges**: more than 50% of students start in STEM fields start at community colleges (NAE & NRC 2012)

- **Freshman and sophomore students**: students who completed three semesters of course-based research (CUREs) had a 23% higher probability of graduating with a STEM degree and 17% higher six-year graduation rate (Rodenbusch et al. 2016)

- **Underrepresented students**: Since the late 1980’s, URMs have aspired to baccalaureate STEM majors at the same rate as their peers (NEST 2014).
Fight the war not the battle
Unmet needs for Bio Big Data

- 2016, Collected 704 / 3987 respondents across the 4 NSF Bio Directorates
  - Environmental Bio/ Molecular and Cellular / Bio Infrastructure/
    Organismal systems
- Asked about current and future needs in training and infrastructure
- Satisfied by research group size/directorate
“Does your institution meet this need?”
(‘no’ responses)

- 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
Not more training, better training

Null effects of boot camps and short-format training for PhD students in life sciences

David F. Feldon\textsuperscript{a,1}, Soojong Jeong\textsuperscript{a}, James Peugh\textsuperscript{b}, Josipa Roksa\textsuperscript{c,d}, Cathy Maahs-Fladung\textsuperscript{e}, Alok Shenoy\textsuperscript{a}, and Michael Oliva\textsuperscript{a}

\textsuperscript{a}Department of Instructional Technology & Learning Sciences, Utah State University, Logan, UT 84322-2830; \textsuperscript{b}Department of Pediatrics, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH 45229-3026; \textsuperscript{c}Department of Sociology, University of Virginia, Charlottesville, VA 22904; and \textsuperscript{d}Curry 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)

- Longitudinal study of 294 PhD students from 53 US institutions
- No statistically significant outcomes over 115 variables (e.g. publication/abstracts, writing samples, scholarly engagement, etc.)
- Consistent with published pedagogical research that boot camps don’t lead to sustainable learning*
Barriers Differing Significantly by Level of Bioinformatics Training

Shown as percentage of respondents within each level of bioinformatics training n=1052
Question effect size at 80% power is 0.102, sufficient for detecting medium effects [3]
Sibling Organizations

“teach foundational computational, coding, and data science skills to researchers worldwide”

Software Carpentry

• Original Organization (1998)
• Basic computing and software development skills (Linux, programming, version control, databases)

Data Carpentry

• Original funding by NSF-Bio (2013)
• Data Management skills (working with tabular/tidy data)
• Domain specific (Ecology, Genomics, Geospatial, Social)
Communicate with people
The way we programmers explain what we've written
Biological Data Science

November 7-10\textsuperscript{th} 2018

Cold Spring Harbor Laboratory

Download Slides

@JasonWilliamsNY