

# Expanding Undergraduate Participation in Computational Biology: Resources and Lessons Learned from a Hands-on Workshop

Layla Oesper



Carleton

GLBIO – May 20, 2019

# Carleton College



Private Liberal Arts College in Northfield, MN

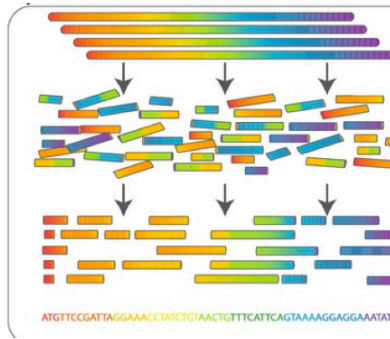
There are many small undergraduate focused institutions.

## The Students

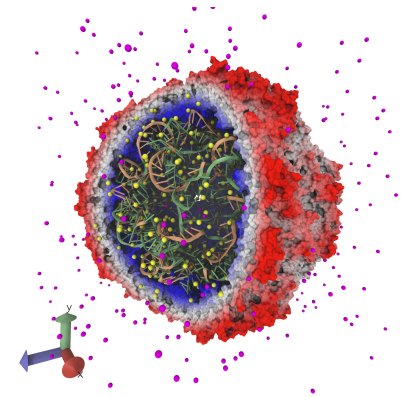
- 2,078 students
- All students are undergraduates
- ~80% students attend graduate school

## The Faculty

- 212 full-time faculty
- 9 Computer Science faculty
- Most faculty are the only expert in their field at the college
- Many areas of research aren't represented



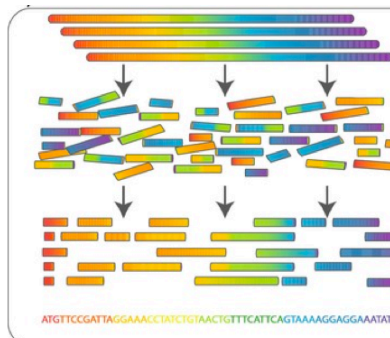
# Key Question



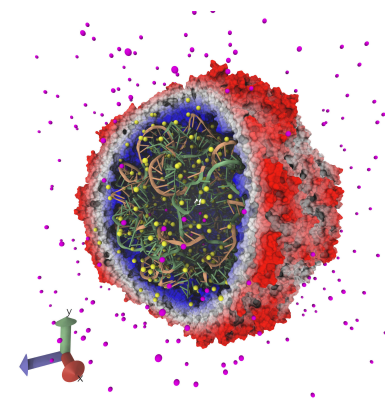
How do we expand participation in computational biology at small undergraduate focused institutions?



CRII-IIS-1657380



# Key Question



How do we expand participation in computational biology at small undergraduate focused institutions?



## Talk Outline

- Background and Inspiration
- Workshop Details, Outcomes and Resources
- Looking to the Future



# 2015 Bioinformatics and Genomics Workshop



Sarah Schaack

## The Basic Facts

- Two day workshop at Reed College in Portland, OR.
- Attended by students, faculty and staff.
- Targeted towards biologists.



Maia Benner

“The aim of the workshop is to introduce the world of publicly-available data and software for performing bioinformatic and genomic analyses.”

# 2015 Bioinformatics and Genomics Workshop



All materials are  
available on the  
website below!

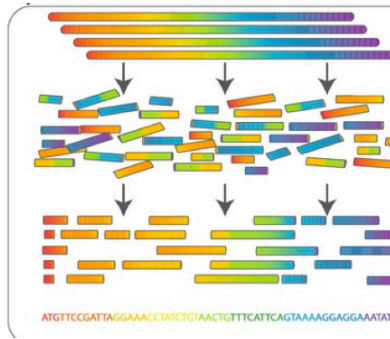
## Module Topics

1. Analyzing gene families
2. Genome Annotation
3. Protein Structure
4. Identifying Genomic Variants
5. Sequence analysis in R
6. Identifying Differentially Expressed Genes
7. Identifying Alternative Splicing
8. Pipeline development and Version Control

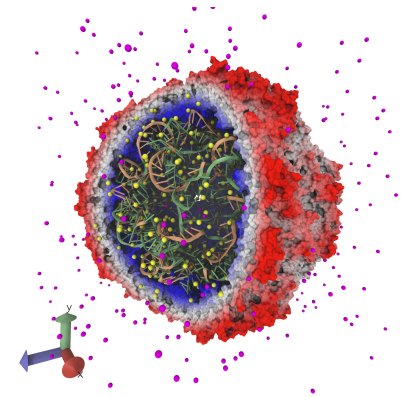
## Crash Courses

1. Using the terminal and command line
2. Using R: Data types, Structures and Subsets

<https://sites.google.com/site/bioinfo-genomics-wkshpv2/home>



# Key Question



How do we expand participation in computational biology at small undergraduate focused institutions?



## Talk Outline

- Background and Inspiration
- Workshop Details, Outcomes and Resources
- Looking to the Future

# 2018 Undergraduate Computational Biology Workshop

## The Basic Facts

- 1.5 day workshop held in Northfield, MN in Fall 2018
- Free to all student participants
- Open to students with a CS1 background. No biology required.

“This workshop will give students a hands on introduction to how computer science can be used to help answer important problems in Biology.”





# 2018 Undergraduate Computational Biology Workshop

What was the format of the workshop?

- Six different 90 minute modules, each created and led by a module leader
- Each module contains background section and a hands-on component
- Modules cover a VERY diverse set of topics



# 2018 Undergraduate Computational Biology Workshop



Rika Anderson  
Carleton College

Who where the module leaders?

Faculty at small undergraduate institutions  
with ties to computational biology



Getiria Onsongo  
Macalester College



Anna Ritz  
Reed College



Ameet Soni  
Swarthmore College



Anya Vostinar  
Grinnell College



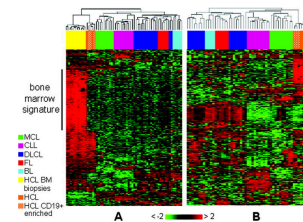
Catie Welsh  
Rhodes College

# Module Topics

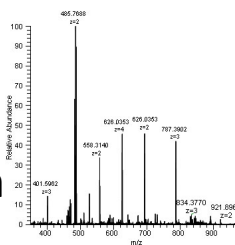
## I. Finding Friends in Molecular Interaction Networks



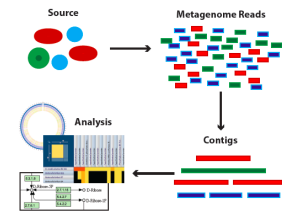
## IV. Machine Learning for Biological Data



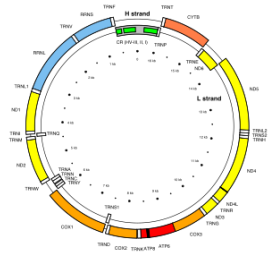
## II. Peptide and Protein Identification using MS/MS Data on the Galaxy-P platform



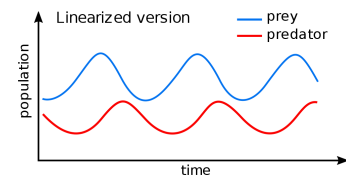
## V. Binning Genomes from Metagenomes



## III. Identifying Somatic Mutations in mtDNA



## VI. Artificial Life Agent-Based Simulations using the GPU



# Modules + Resources



CompBioW...  
2018

^ Home

About Us

Registration

^ Modules

Finding Friends in  
Molecular Interaction  
Networks

Peptide and Protein  
Identification using  
MS/MS Data on the  
Galaxy-P platfor

Identifying Somatic  
Mutations in mtDNA

Machine Learning for  
Biological Data

Binning Genomes  
from Metagenomes

Artificial Life Agent-  
Based Simulations  
using the GPU

Schedule

Assessment

## Computational Biology Workshop

September 29-30, 2018  
Carleton College, Northfield, MN

Welcome to the website for the 2018 Undergraduate Computational Biology Workshop at Carleton College.

This workshop will give students a hands on introduction to how computer science can be used to help answer important problems in Biology. The aim of the workshop is to help grow the computational biology community at liberal arts colleges in the region. Students from Carleton or other surrounding institutions who have taken an introductory computer science class are eligible to register to attend. No biology background is required. **Registration is now closed.**

**Workshop Organizer:** Layla Oesper, Carleton College

**Workshop Contributors:**

- Rika Anderson, Department of Biology, Carleton College
- Getiria Onsongo, Department of Computer Science, Macalester College
- Anna Ritz, Department of Biology, Reed College
- Ameet Soni, Department of Computer Science, Swarthmore College
- Anya Vostinar, Department of Computer Science, Grinnell College
- Catie Welsch, Department of Computer Science, Rhodes College

**Workshop Location:** The workshop will occur in the Weitz Center for Creativity, room 138. For a campus map see [here](#). For a map of all locations related to the workshop see [here](#).

<https://sites.google.com/carleton.edu/compbioworkshop2018>



# Finding Friends in Molecular Interaction Networks

## 1. Summary

Cells respond to external signals through protein-protein interactions. These interactions are often represented as a graph, and algorithms from graph theory can be used to generate hypotheses about protein regulation. In this module, I will introduce the computational problem of identifying candidate regulators of a specific protein of interest using molecular interaction networks. As a case study, we will aim to predict novel regulators of Fog signaling, which is involved in changing the shape of a cell. Students will try their hand at identifying candidate regulators in a newly-established Drosophila interactome, and will visualize their results using graph visualization software.

## 2. Presentation Materials:

Click [here](#).

## 3. Hands-on Exercise(s):

<https://repl.it/classroom/invite/W7alQQe> You will need to make a username and password.

## 4. Associated Materials/Files

NetworkX Cheat Sheet: click [here](#).

GraphSpace Cheat Sheet: click [here](#).

GraphSpace: <https://graphspace.org/>

CompBioWorkshop GraphSpace Group: <https://www.graphspace.org/groups/1067>

HTML Color Picker: <https://htmlcolorcodes.com/color-picker/>

## 5. Program/Software requirements

None. Hands-on activity is done through a web browser.

## 6. Advanced Material

### Further Reading

*Scale-Free Networks*. Barabasi and Bonabeau, Scientific American 2003. <http://barabasi.com/f/124.pdf>

*Data Science of the Facebook World*. Stephen Wolfram's blog, 2013. <http://blog.stephenwolfram.com/2013/04/data-science-of-the-facebook-world/>

*Interactome-based approaches to human disease*. Caldera et al., Current Opinion in Systems Biology 2017.

<https://www.sciencedirect.com/science/article/pii/S2452310016300154>

*A Cell-based Assay to Investigate Non-muscle Myosin II Contractility via the Folded-gastrulation Signaling Pathway in Drosophila S2R+ Cells*. Peters et al. Journal of Visual Experiments.

2018. <https://www.jove.com/video/58325/a-cell-based-assay-to-investigate-non-muscle-myosin-ii-contractility>

# Finding Friends in Molecular Interaction Networks

Introduces the computational problem of identifying candidate regulators of a specific protein of interest using molecular interaction networks.

## I. Working with Graphs

September 29-30, 2018 Computational Biology Workshop 6

Local Structure: Friend Groups

September 29-30, 2018 Computational Biology Workshop 10

The Computational Problem

**Given:** Undirected social network | **Given:** Set of my friends (positives)

September 29-30, 2018 Computational Biology Workshop 11

The Computational Problem with Adversaries

**Given:** Undirected social network | **Given:** Set of my friends (positives) and adversaries (negatives)

**Goal:** Identify the people that are most likely to be eventual friends

**Even Better:** Rank all non-friends by friendship confidence.

NetworkX

GraphSpace

## II. Working with Molecular Interaction Networks

September 29-30, 2018 Computational Biology Workshop 15

Protein-Protein Interactions

September 29-30, 2018 Computational Biology Workshop 20

Predicting Regulators of NMII

Absence of Fog ? Presence of Fog ?

September 29-30, 2018 Computational Biology Workshop 22

The Computational Problem

**Given:** Undirected interactome | **Given:** Set of Fog proteins (positives) and unrelated proteins (negatives)

**Goal:** Identify the proteins that are most likely to be Fog/NMII regulators

**Even Better:** Rank all unlabeled nodes by their regulator confidence.

Hands-on Components: <https://repl.it/classroom/invite/W7aIQQe>

# Finding Friends in Molecular Interaction Networks

Introduces the computational problem of identifying candidate regulators of a specific protein of interest using molecular interaction networks.

repl.it

Part 1: Working with Graphs

my repls student notifications talk learn/teach loesper (0)

share your program enhance your work after submitting export to repl

back to classroom run

main.py example\_edge.. example\_node..

```
1 # import the NetworkX module
2 # learn more: https://networkx.github.io/documentation/stable/
3 import networkx as nx
4
5 # import the GraphSpace and GSGraph classes from graphspace_python
6 # learn more: https://python.org/pypi/graphspace_python
7 from graphspace_python.api.client import GraphSpace
8 from graphspace_python.graphs.classes.gsgraph import GSGraph
9
10 def main():
11     """
12     Main function creates the graph and calculates some
13     basic statistics.
14     Inputs: None
```

Python 3.6.1 (default, Dec 2015, 13:05:11)  
[GCC 4.8.2] on linux

Due: --

submit

Instructions from your teacher:

Graph object and how to get information about a graph using the NetworkX API.

### 3 GraphSpace

GraphSpace (<http://graphspace.org/>) is a webservice for uploading, sharing, and interacting with graphs. Make another tab in your browser and log into GraphSpace (<http://graphspace.org/>) using the username **compbio@reed.edu** and password **compbio**. You should be a member of the "CompBioWorkshop" group under Groups (<http://graphspace.org/groups/1067>). You are going to visualize your graph and add it to the list of graphs in this group.

The diagram illustrates the GraphSpace workflow. It features a central circular flow involving a 'Scientific Community' and a 'Researcher'. The cycle consists of three main stages: 'Network Generation' (receiving 'Cytoscape.js JSON' input), 'Network Exploration' (displaying a network visualization), and 'Network Dissemination' (outputting 'JSON', 'URL', and 'PNG' files). Below the main cycle, there are two panels for 'Filter nodes and edges' with 'Current rank' indicators (10 and 50). To the right, there are panels for 'Node/Edge Popups' and 'Sharing' options. A 'Search' bar is located at the top right of the interface.

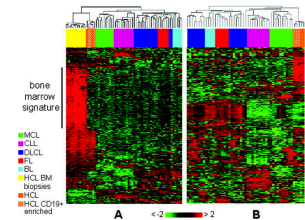
Privacy - Terms

# Module Topics

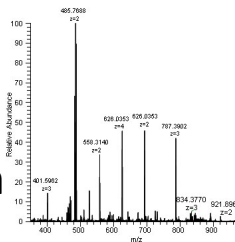
## I. Finding Friends in Molecular Interaction Networks



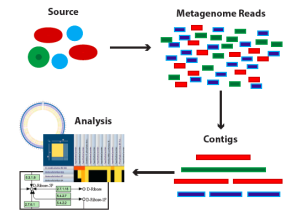
## IV. Machine Learning for Biological Data



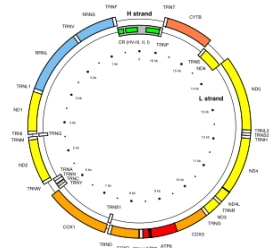
## II. Peptide and Protein Identification using MS/MS Data on the Galaxy-P platform



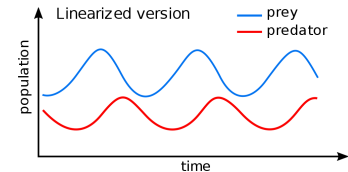
## V. Binning Genomes from Metagenomes



### III. Identifying Somatic Mutations in mtDNA



## VI. Artificial Life Agent-Based Simulations using the GPU



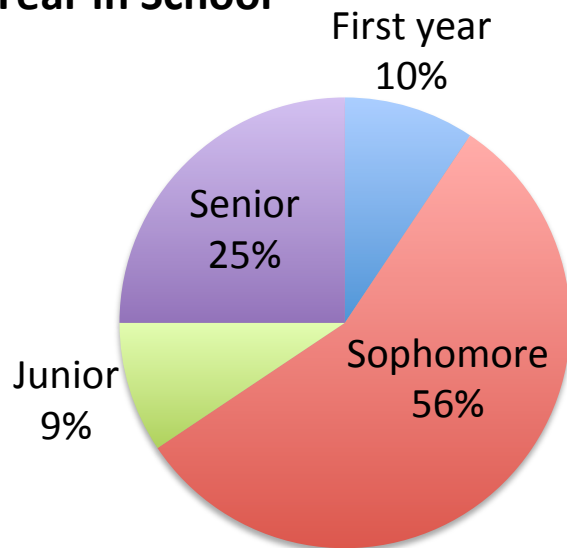


# 2018 Undergraduate Computational Biology Workshop

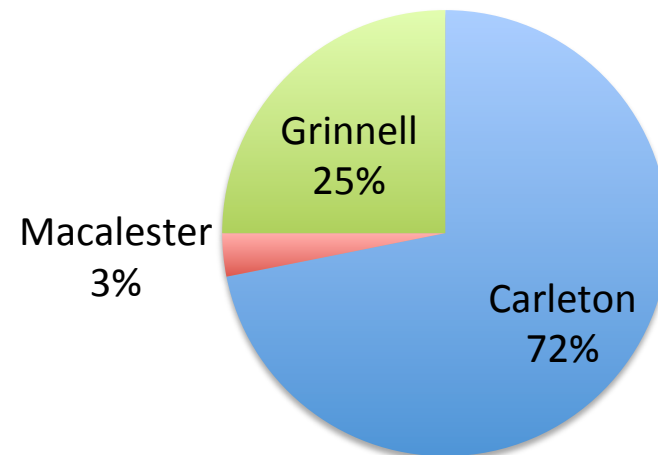
Who where the attendees?

32 undergraduate student participants registered (plus one parent).

**Year in School**



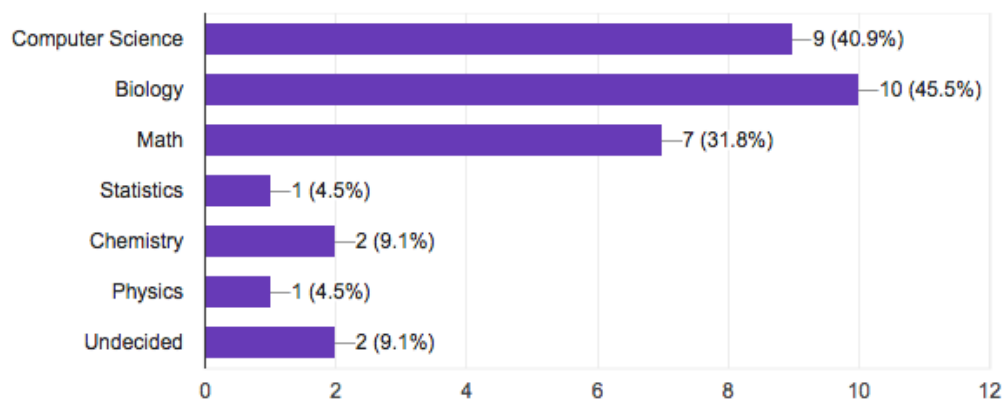
**School Attended**



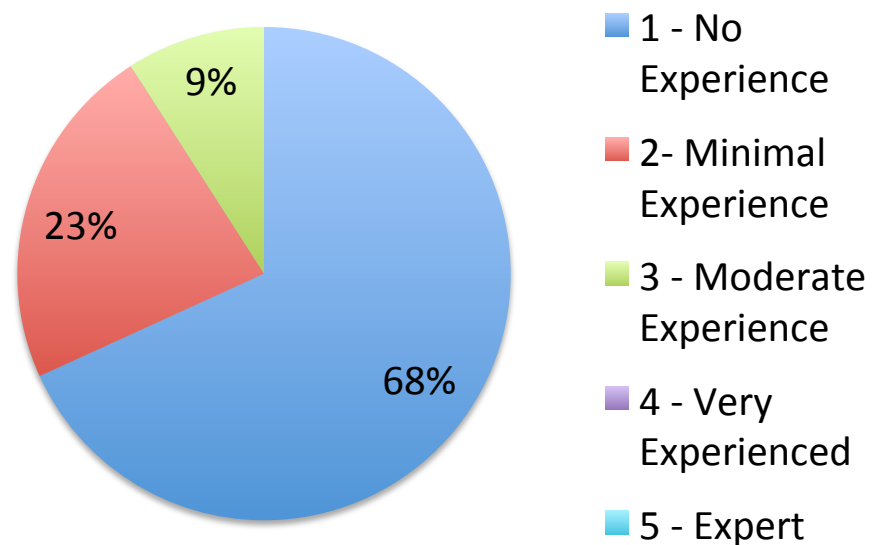
# 2018 Undergraduate Computational Biology Workshop

Who where the attendees?

Major or intended Major



Previous Computational Biology Experience



# Workshop Outcomes

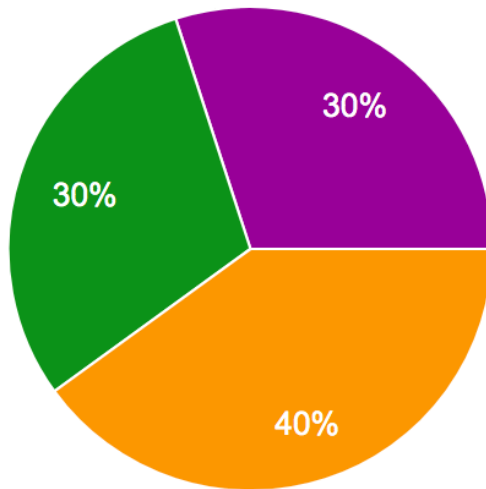
“This workshop will give students a hands on introduction to how computer science can be used to help answer important problems in Biology.”



Did we achieve this goal?

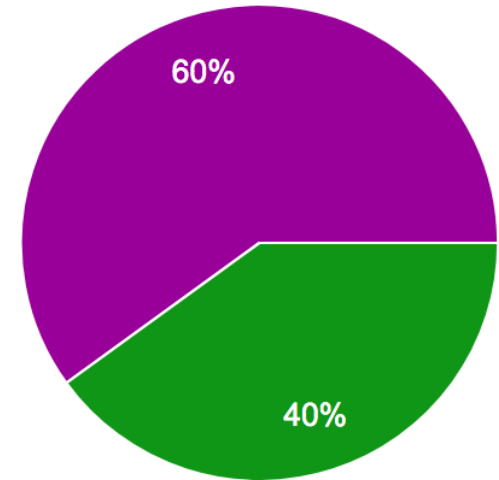
# Outcomes of the Workshop for Students

**I am likely to pursue future computational biology opportunities.**

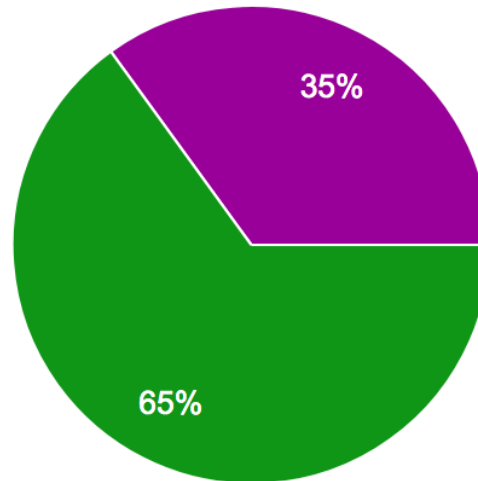


- 1 - Strongly Disagree
- 2 - Disagree
- 3 - Undecided
- 4 - Agree
- 5 - Strongly Agree

**The workshop exposed me to novel and exciting applications involving computation and biology**



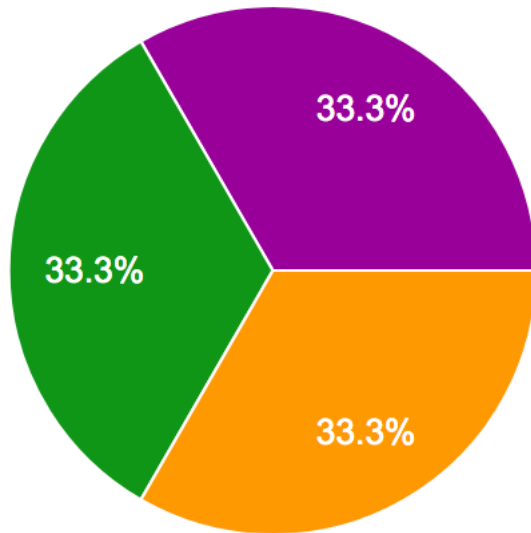
**I am aware of a variety of ways that computation can be combined with biology.**



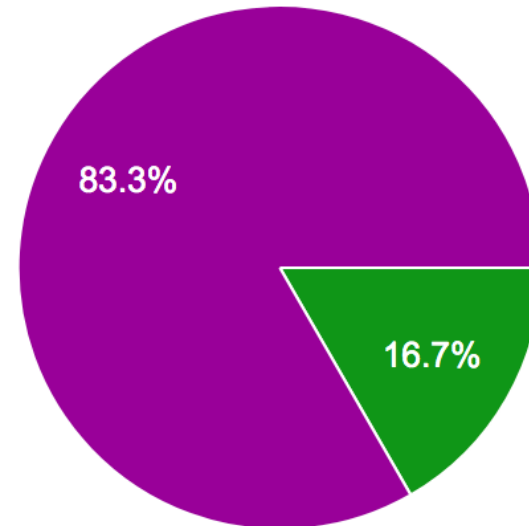


# Outcomes of the Workshop for the Module Leaders

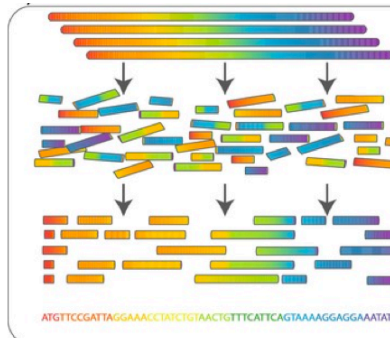
**I will use modules created by others for the workshop again in the future in my own classes or otherwise.**



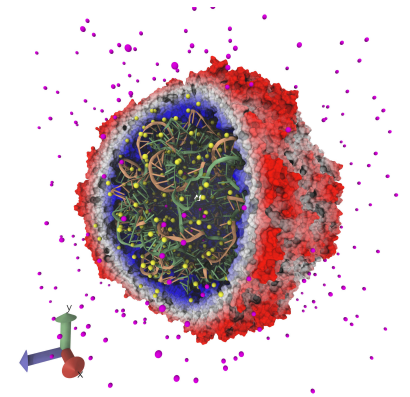
**This workshop helped to expand my computational biology professional network**



- 1 - Strongly Disagree
- 2 - Disagree
- 3 - Undecided
- 4 - Agree
- 5 - Strongly Agree



# Key Question



How do we expand participation in computational biology at small undergraduate focused institutions?



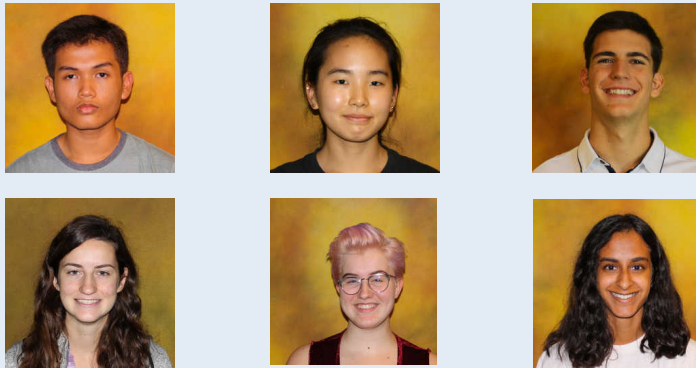
## Talk Outline

- Background and Inspiration
- Workshop Details, Outcomes and Resources
- Looking to the Future

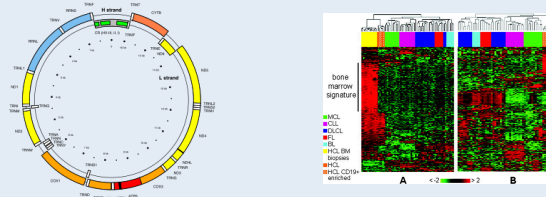
# Future Workshops?

## Things that Worked Well

- Hire Student Beta Testers



- Wide breadth of modules



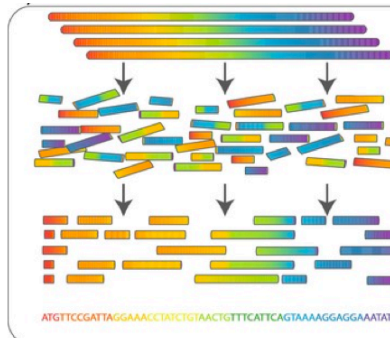
- One and a half day length

## Things that were Challenging

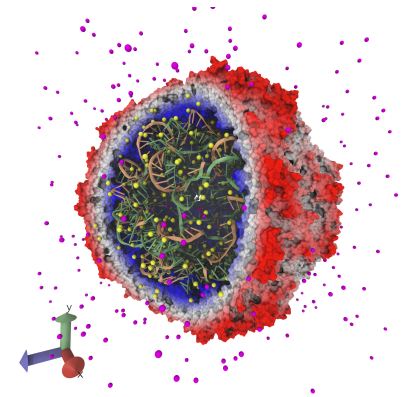
- Getting students from other colleges to attend.
- Technical difficulties: Online vs. installed software tradeoff.



- Length/depth of modules matching student background.



# Key Question



How do we expand participation in computational biology at small undergraduate focused institutions?



## Talk Outline

- Background and Inspiration
- Workshop Details, Outcomes and Resources
- Looking to the Future

# Acknowledgements

## Module Leaders

Anna Ritz  
Rika Anderson  
Getiria Onsongo  
Ameet Soni  
Catie Welsh  
Anya Vostinar

## Administrative Assistance



Sue Jandro

## Technical Assistance



Mike Tie



Rebecca Barkmeier



Troy Barkmeier

## Workshop Helpers/ Beta Testers

Alief Moulana  
Allison Kim  
David Galambos  
Kate Engel  
Rosemary Wonnell  
Saahithi Rao

## Inspiration and Advising



Sarah Schaack

## Funding Sources



CRII-IIS-1657380



Carleton

**Resources:** <https://sites.google.com/carleton.edu/compbioworkshop2018>