

Cal Poly Bioinformatics:

Designing a Cross-Disciplinary Minor that Builds on the In-Concert Model for Bioinformatics Education

Jean Davison, Biology

Alex Dekhtyar, Computer Science

Anya Goodman, Chemistry/Biochemistry

Ed Himmelblau, Biology

Franz Kurfess, Computer Science

Theresa Migler-VonDollen, Computer Science

Andrew Schaffner, Statistics

Natalie Jo Schafer, Cal Poly Strategic Development Initiatives

Thank you!

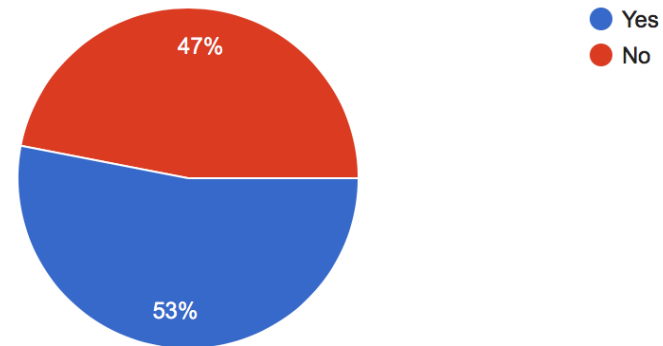
CAWIT Technology
Pathways Initiative

Prologue: Interest in and Challenges of Bioinformatics at Cal Poly

Interest in Bioinformatics: (among 164 Biology Freshman)

Would you consider taking additional coursework (beyond the credits required for your degree) to help you add coding experience?

164 responses



Limited Opportunities:

BIO 441 (Bioinformatics Applications): about 40 students per year

CSC 448 (Bioinformatics Algorithms): about 40 students per year

Difficult for BIO/BCHEM Students to enroll in Computer Science courses (e.g. very few complete existing Computer Science minor due to scheduling constraints)

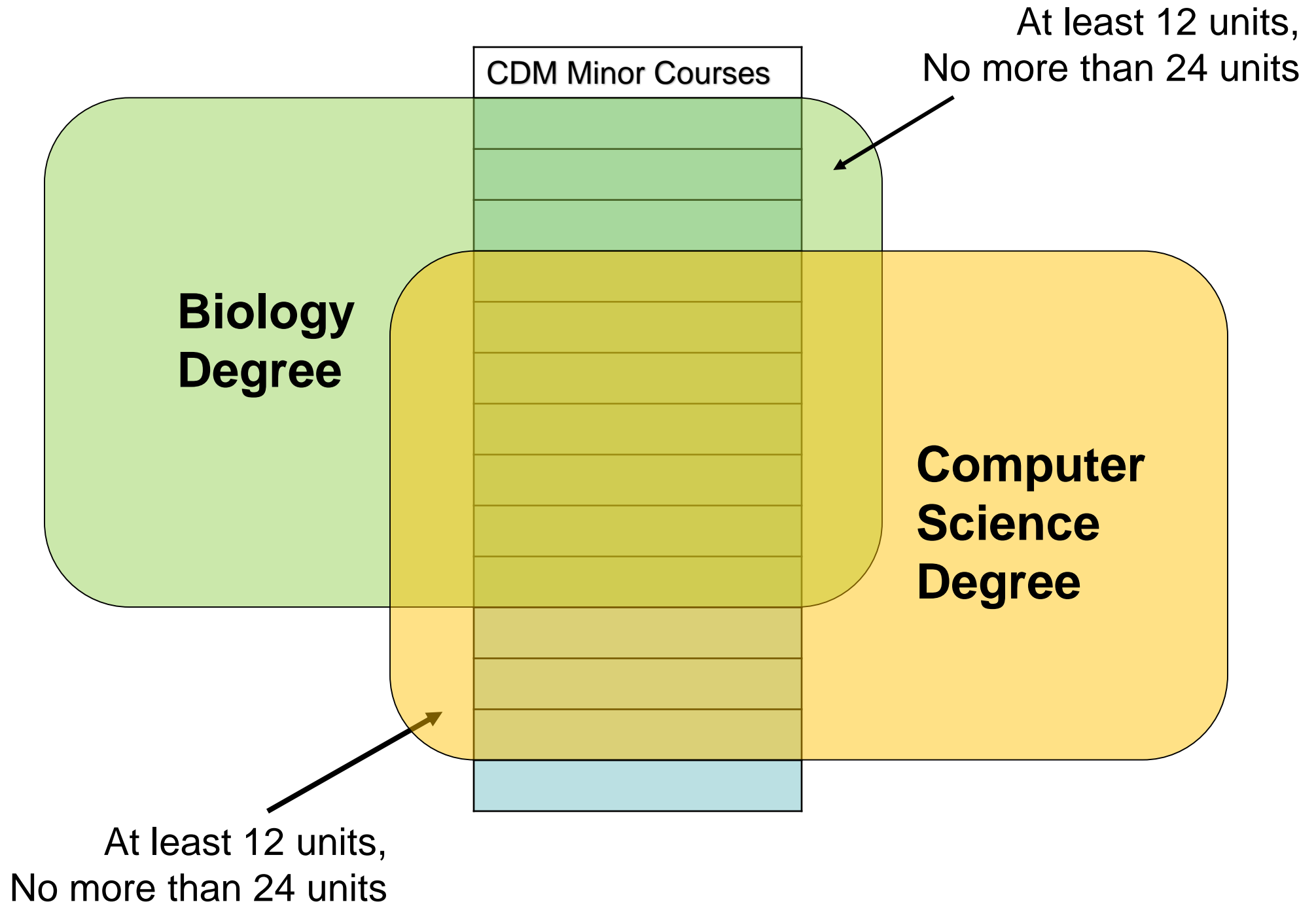
Create a Minor in Bioinformatics (Supported by CAWIT/TPI) that is informed by our experience teaching bioinformatics in a cross-disciplinary way (“in concert”).

Only open (or possible!) for particular majors.

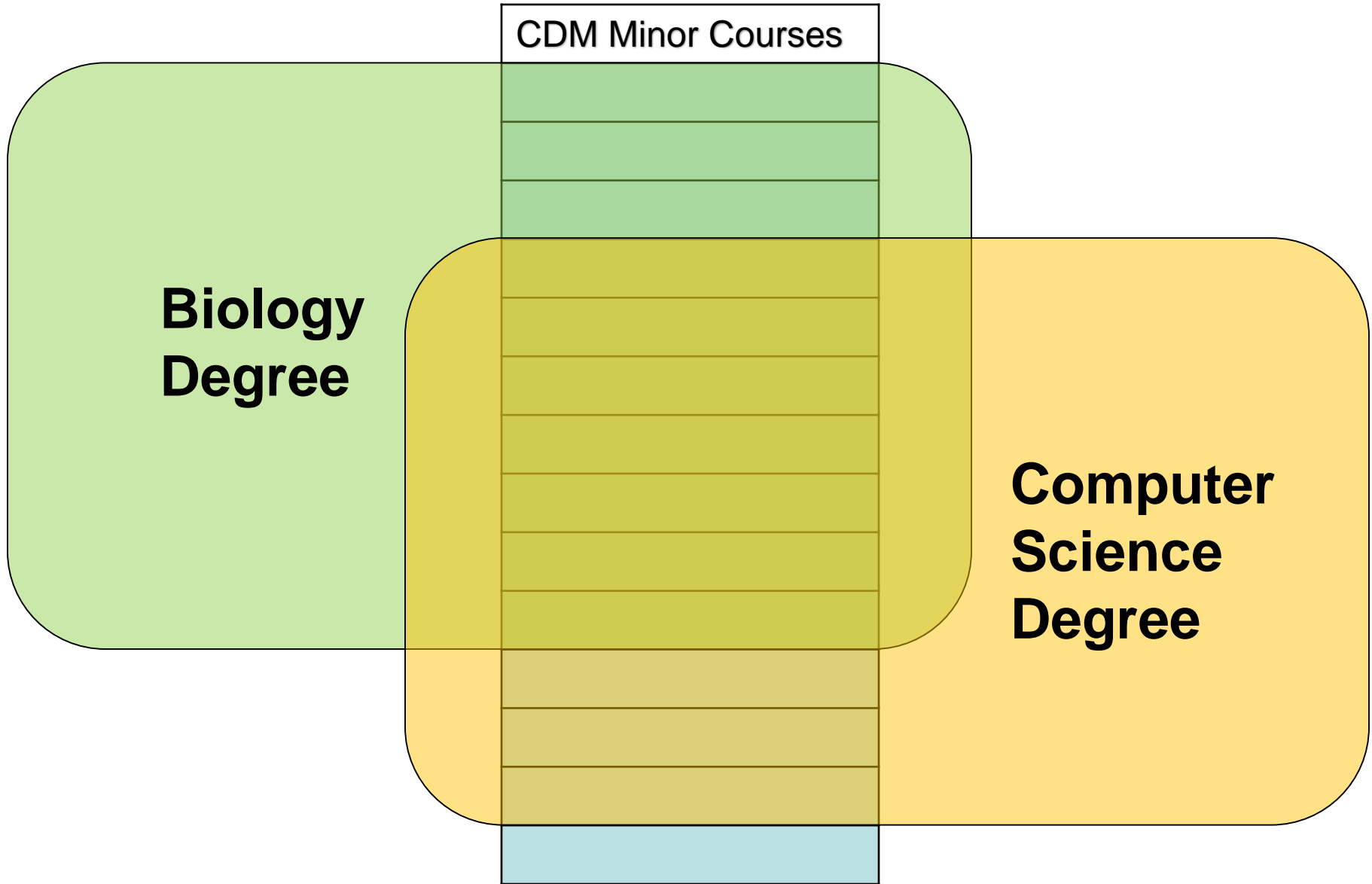
All student in the CDM take the same courses regardless of major.

At least 12 and no more than 24 units must be from outside the student's major degree.

Cross-Disciplinary Minors at Cal Poly



Cross-Disciplinary Minors at Cal Poly






 OPEN ACCESS

MESSAGE FROM ISCB

Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies

Lonnie Welch , Fran Lewitter, Russell Schwartz, Cath Brooksbank, Predrag Radivojac, Bruno Gaeta, Maria Victoria Schneider

Published: March 6, 2014 • <https://doi.org/10.1371/journal.pcbi.1003496>

265 Save	31 Citation
42,228 View	151 Share

Article

Authors

Metrics

Comments

Related Content



Download PDF 

Print

Share

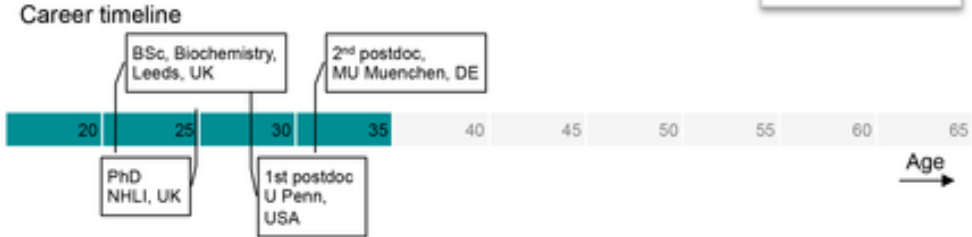
Bioinformatics User

Leon (bioinformatics user)

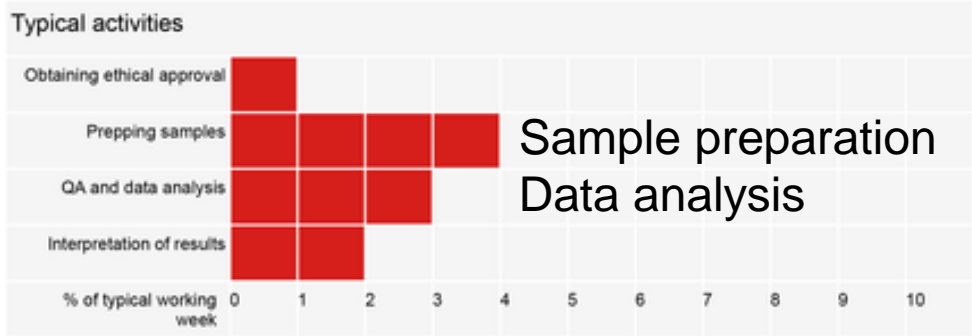
Leon is on his second postdoctoral fellowship, working on quorum sensing in bacteria. "I'm using a combination of transcriptomics, proteomics and metabolomics to understand these pathogenic changes better" he explains. "I end up with big spreadsheets of protein or gene IDs and I'm trying to piece together which signaling pathways are involved in flipping to the pathogenic state". He has been on an introductory Unix course but is much more comfortable with GUIs than with the command line. "I just have a visual brain", he says.



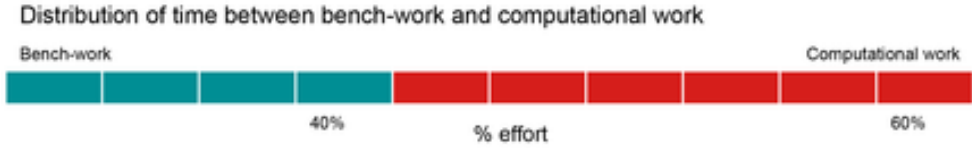
career path



typical activities

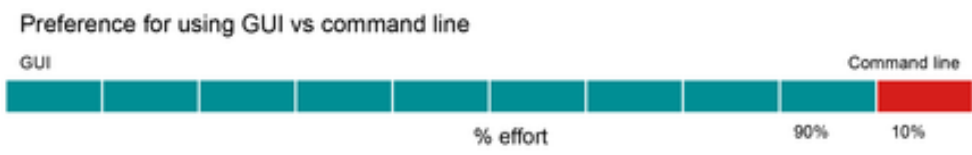


lab bench



workstation

GUI



command Line

<p>Drivers</p> <ul style="list-style-type: none"> Understanding what makes a usually harmless bacterium pathogenic in the lungs of people with cystic fibrosis 	<p>Goals</p> <ul style="list-style-type: none"> QA of -omics data Statistical analysis of data Data integration and pathway analysis 	<p>Pain points</p> <ul style="list-style-type: none"> Lack of access to departmental compute farm Sporadic to non-existent access to bioinformatics support
--	--	--

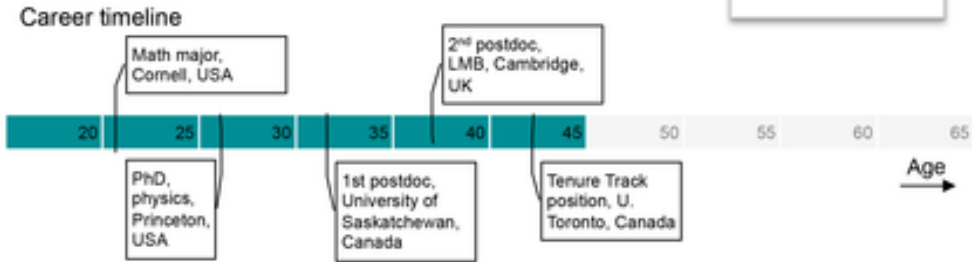
Bioinformatics Scientist

Martha (bioinformatics scientist)

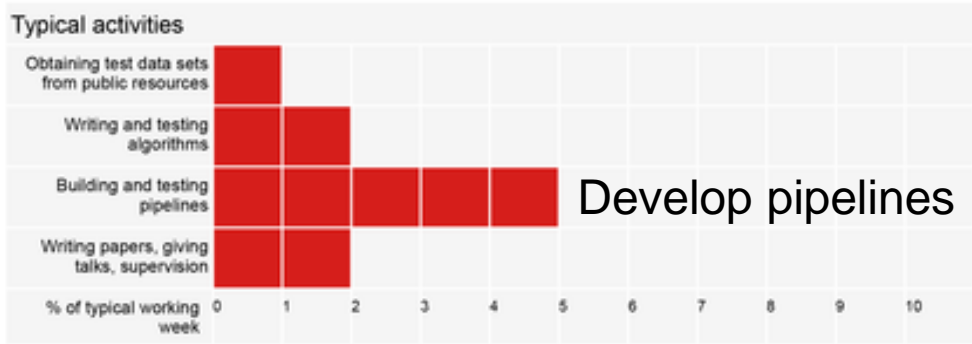
Martha is a senior bioinformatician in an international structural genomics consortium. Her biggest project is on predicting the functions of proteins whose structures have just been solved; she's building a structure-to-function prediction pipeline for the project. This is funded partly by the NIH and partly through industrial funding. She also has a fascination for predicting structure and usually has a student or two working on structural prediction projects.



career path



typical activities



Distribution of time between bench work and computational work



100% workstation

Preference using for GUI vs command line



GUI

command Line

Drivers

- Understanding the relationship between sequence, structure and function
- Application to target discovery and validation

Goals

- Create a structure-to-function pipeline for molecular biologists
- Predict structures de novo from models of similar, solved structures

Pain points

- Sometimes the guys in the lab expect her to fix their computers for them
- Finding students and more senior staff with adequate math

Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, et al. (2014) Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. PLOS Computational Biology 10(3): e1003496. <https://doi.org/10.1371/journal.pcbi.1003496>
<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003496>

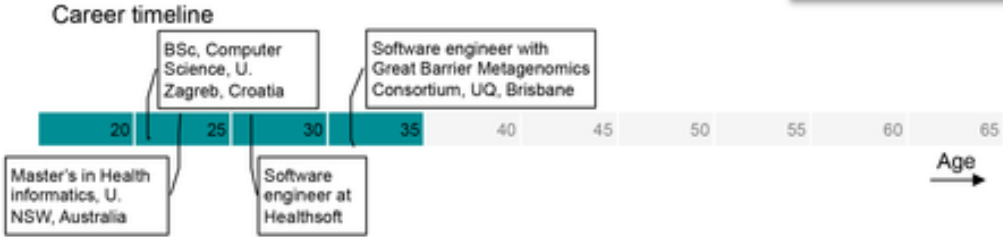
Bioinformatics Engineer

Ivan (bioinformatics engineer)

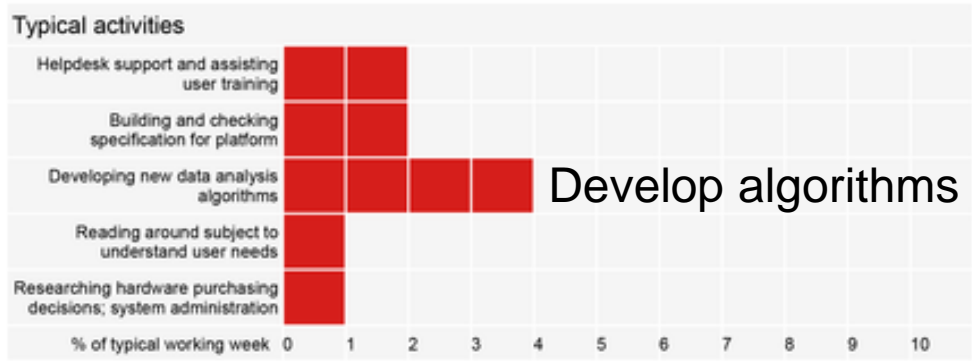
Ivan has just started a new support role in a bioinformatics core facility after working for an electronic health records company for four years. His main project is to develop a major new data integration platform for metagenomics data from coral reefs, but he also has to take his share of helpdesk queries on other projects. "I come from a computer science background, so talking the same language as the guys analysing the data is a bit of a challenge," he says. "I also didn't really figure that I'd be working on the GUI as well as the code – in my last job we had design folks to take care of that".



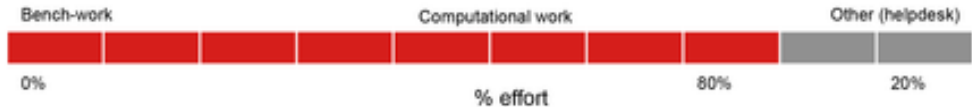
career path



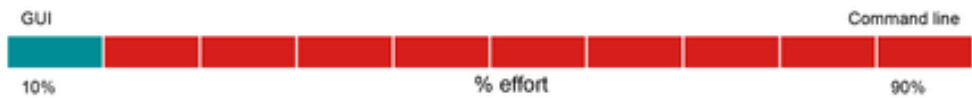
typical activities



Distribution of time between bench-work and computational work



Preference for using GUI vs command line



GUI

Drivers

- Writing algorithms and developing a platform to support novel research
- Supporting other research projects in a busy academic department

Goals

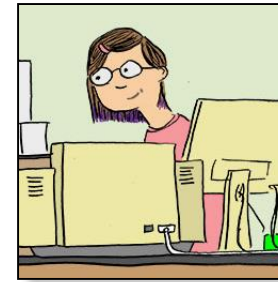
- Define a spec that meets the needs of his users
- Prototype and build part of the platform
- Make sure his part of the project complements others

Pain points

- Has to work with another software engineer who isn't a team player
- Sometimes struggles to interpret what his users want

Bioinformatics Minor Degree Holder

Well prepared to apply concepts and skills from outside her primary degree area. Proficient at cross-disciplinary collaboration in a bioinformatics setting.



Leon (bioinformatics user)

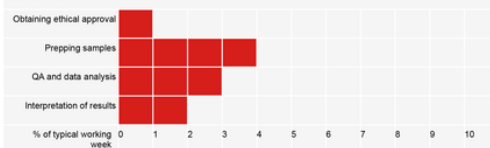
Leon is on his second postdoctoral fellowship, working on quorum sensing in bacteria. "I'm using a combination of transcriptomics, proteomics and metabolomics to understand these pathogenic changes better" he explains. "I end up with big spreadsheets of protein or gene IDs and I'm trying to piece together which signaling pathways are involved in flipping to the pathogenic state". He has been on an introductory Unix course but is much more comfortable with GUIs than with the command line. "I just have a visual brain", he says.



Career timeline



Typical activities



Distribution of time between bench-work and computational work



Preference for using GUI vs command line



Drivers

- Understanding what makes a usually harmless bacterium pathogenic in the lungs of people with cystic fibrosis

Goals

- QA of -omics data
- Statistical analysis of data
- Data integration and pathway analysis

Pain points

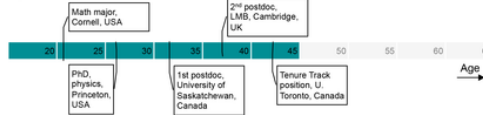
- Lack of access to departmental compute farm
- Sporadic to non-existent access to bioinformatics support

Martha (bioinformatics scientist)

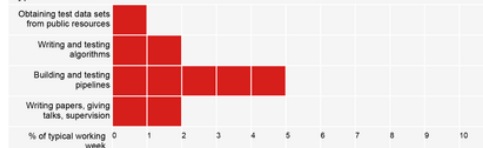
Martha is a senior bioinformatician in an international structural genomics consortium. Her biggest project is on predicting the functions of proteins whose structures have just been solved. She's building a structure-to-function prediction pipeline for the project. This is funded partly by the NIH and partly through industrial funding. She also has a fascination for predicting structure and usually has a student or two working on structural prediction projects.



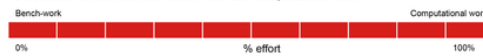
Career timeline



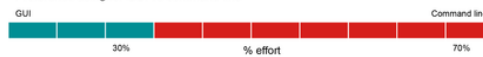
Typical activities



Distribution of time between bench work and computational work



Preference using for GUI vs command line



Drivers

- Understanding the relationship between sequence, structure and function
- Application to target discovery and validation

Goals

- Create a structure-to-function pipeline for molecular biologists
- Predict structures de novo from models of similar, solved structures

Pain points

- Sometimes the guys in the lab expect her to fix their computers for them
- Finding students and more senior staff with adequate math

Ivan (bioinformatics engineer)

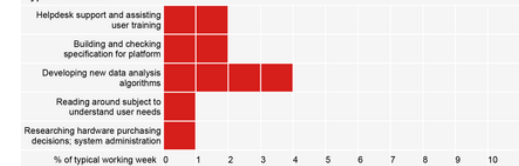
Ivan has just started a new support role in a bioinformatics core facility after working for an electronics health records company for four years. His main project is to develop a major new data integration platform for metagenomics data from coral reefs, but he also has to take his share of helpdesk queries on other projects. "I come from a computer science background, so talking the same language as the guys analysing the data is a bit of a challenge," he says. "I also didn't really figure that I'd be working on the GUI as well as the code - in my last job we had design folks to take care of that".



Career timeline



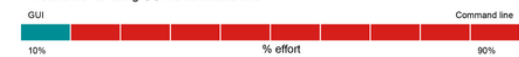
Typical activities



Distribution of time between bench-work and computational work



Preference for using GUI vs command line



Drivers

- Writing algorithms and developing a platform to support novel research
- Supporting other research projects in a busy academic department

Goals

- Define a spec that meets the needs of his users
- Prototype and build part of the platform
- Make sure his part of the project complements others

Pain points

- Has to work with another software engineer who isn't a team player
- Sometimes struggles to interpret what his users want

Core Competencies

Computation

Programming

Software Engineering

Algorithm Design and
Analysis

Machine learning/data
mining

Databases

Scripting Languages

Ability to use stats
software packages

Open source
repositories

Web UI

Version Control

Core Competencies

Computation

Programming

Software Engineering

Algorithm Design and
Analysis

Machine learning/data
mining

Databases

Scripting Languages

Ability to use stats
software packages

Open source
repositories

Web UI

Version Control

Biology

Molecular biology

Genetics

Genomics (-omics)

Evolution

Core Competencies

Computation

Programming

Software Engineering

Algorithm Design and Analysis

Machine learning/data mining

Databases

Scripting Languages

Ability to use stats software packages

Open source repositories

Web UI

Version Control

Biology

Molecular biology

Genetics

Genomics (-omics)

Evolution

Bioinformatics

Retrieve and manipulate public data

Use (manage, interpret, analyze) large data sets

Explain common bioinformatics algorithms.

Access bioinformatics resources

Communicate genomic data

Core Competencies

Computation

Programming

Software Engineering

Algorithm Design and Analysis

Machine learning/data mining

Databases

Scripting Languages

Ability to use stats software packages

Open source repositories

Web UI

Version Control

Biology

Molecular biology

Genetics

Genomics (-omics)

Evolution

Bioinformatics

Retrieve and manipulate public data

Use (manage, interpret, analyze) large data sets

Explain common bioinformatics algorithms.

Access bioinformatics resources

Communicate genomic data

Statistics

Descriptive Statistics

Data Visualization

Experimental design

Statistical Inference

Statistical Models (intro to regression / ANOVA)

R and R libraries

Statistical Software

Probability

Core Competencies

Computation

Programming

Software Engineering

Algorithm Design and Analysis

Machine learning/data mining

Databases

Scripting Languages

Ability to use stats software packages

Open source repositories

Web UI

Version Control

Biology

Molecular biology

Genetics

Genomics (-omics)

Evolution

Bioinformatics

Retrieve and manipulate public data

Use (manage, interpret, analyze) large data sets

Explain common bioinformatics algorithms.

Access bioinformatics resources

Communicate genomic data

Statistics

Descriptive Statistics

Data Visualization

Experimental design

Statistical Inference

Statistical Models (intro to regression / ANOVA)

R and R libraries


Statistical Software

Probability

Teaching *in concert*

a novel approach to interdisciplinary collaborative project-based instruction

Comparative Genomics



Team 5 and 6
(Kristin LaForge and Anne Welker)

Dot contigs 20-38
3L fosmids 22-42



BIOINFORMATICS ALGORITHMS

REDUNDANCY TEAM

COLLABORATION

REFLECTION

CC MEMBERS:
Kristin LaForge
Anne Welker

SOFTWARE

DNA Analysis Library:
These different statistical analyses can be used to partition knowledge on how relationships are formed by DNA sequences. For the purpose of this analysis, the computer generated the designed library of functions to compare the binary GC content. Kristin LaForge, Anne Welker, and Anne Welker.

Sequence Classification:
Information about the GC content provides a way to separate unrelated DNA sequences from the data. The assessment that can be applied to the data is all of the sequences were related. While the GC content is a useful metric for identifying related sequences, it is not sufficient to identify related sequences in all cases.

Repeat Polymorphisms Sequence Location:
Certain repetitive elements in the genome are found in the sequences of DNA. The location of these elements is important for understanding the structure and function of the genome.

Leading the field in collaboration and synergy

Ed Himelblau
Biological Sciences

Anya Goodman
Chemistry and Biochemistry



Alex Dekhtyar
Computer Science

Theresa Migler
Computer Science

In concert approach to the upper level bioinformatics course:

Lectures – separate, discipline-specific

BIO Applications course

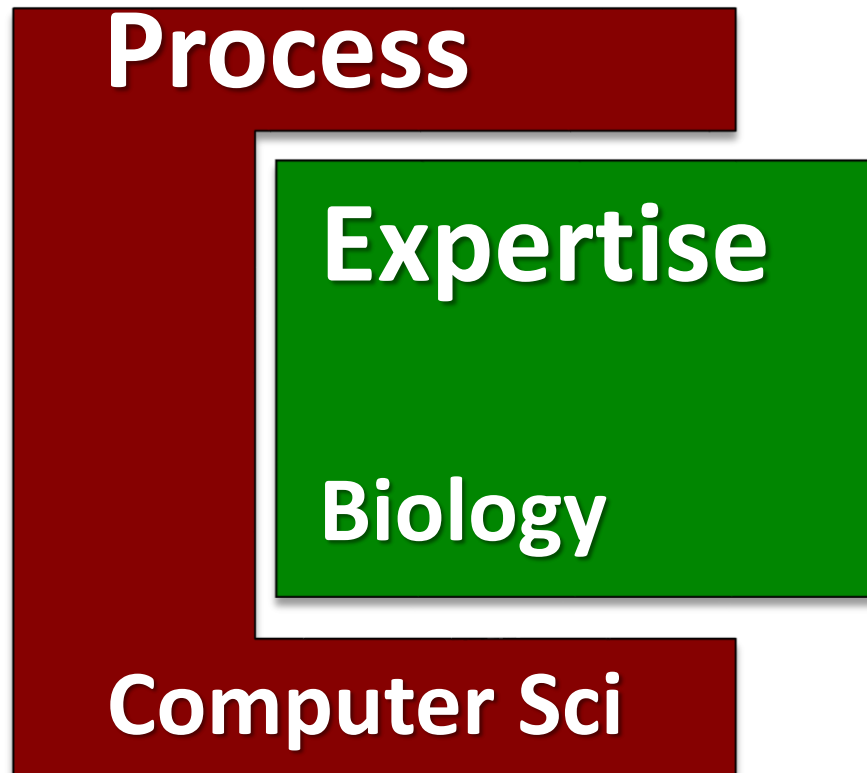
- Database and software use
- Genome research
- New technologies

CS Algorithms course

- Algorithms
- Data management
- Software Development

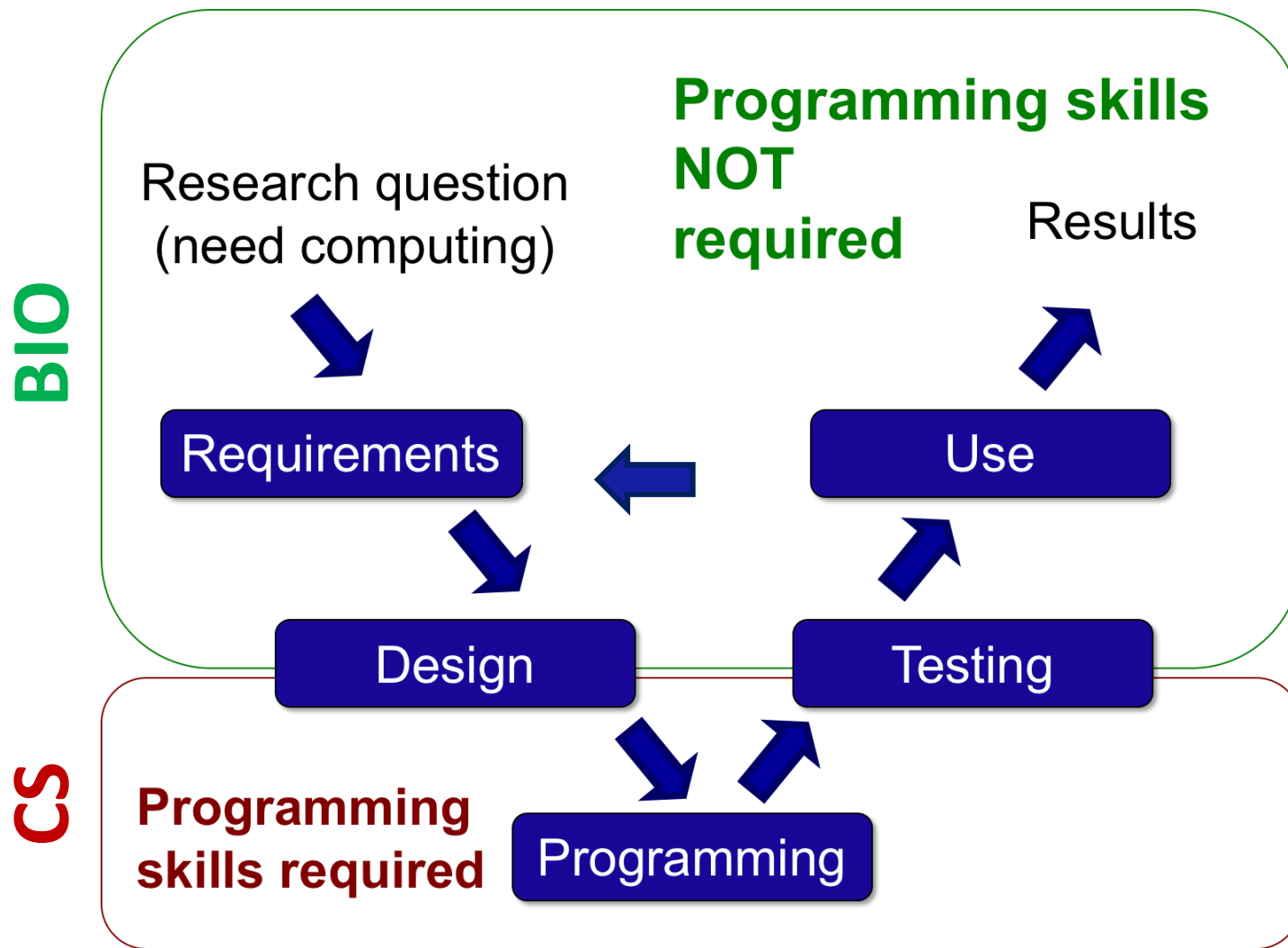
In concert approach to the upper level bioinformatics course:

Lectures – separate, discipline-specific



Laboratories –
collaborative
assignments
structured around
clearly defined
interdependent roles

Interdisciplinary collaboration structured around **clearly defined interdependent roles**



Software development cycle

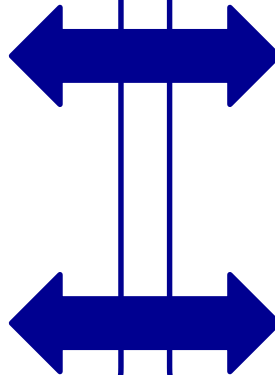
Interdisciplinary collaboration structured around **clearly defined interdependent** roles

BIO Applications course

- Discuss the problem
- Write program requirements
- Test software
- Use software

CS Algorithms course

- Elicit program requirements
- Design/build software
- Test software
- Deliver/maintain



In concert approach to the upper level bioinformatics course:

BIO Applications course

- Database and software use
- Genome research
- New technologies

Add:

- Comp. thinking
- Effective collaboration across disciplines

CS Algorithms course

- Algorithms
- Data management
- Software Development

Add:

- Effective collaboration across disciplines

In concert approach: implementation

9-12 teams

30-36 CS students

2-3 CS

20-25 BIO students

2-3 BIO

1-3 teaching assistants, two instructors

10-week quarter

60 hours

meet twice a week:

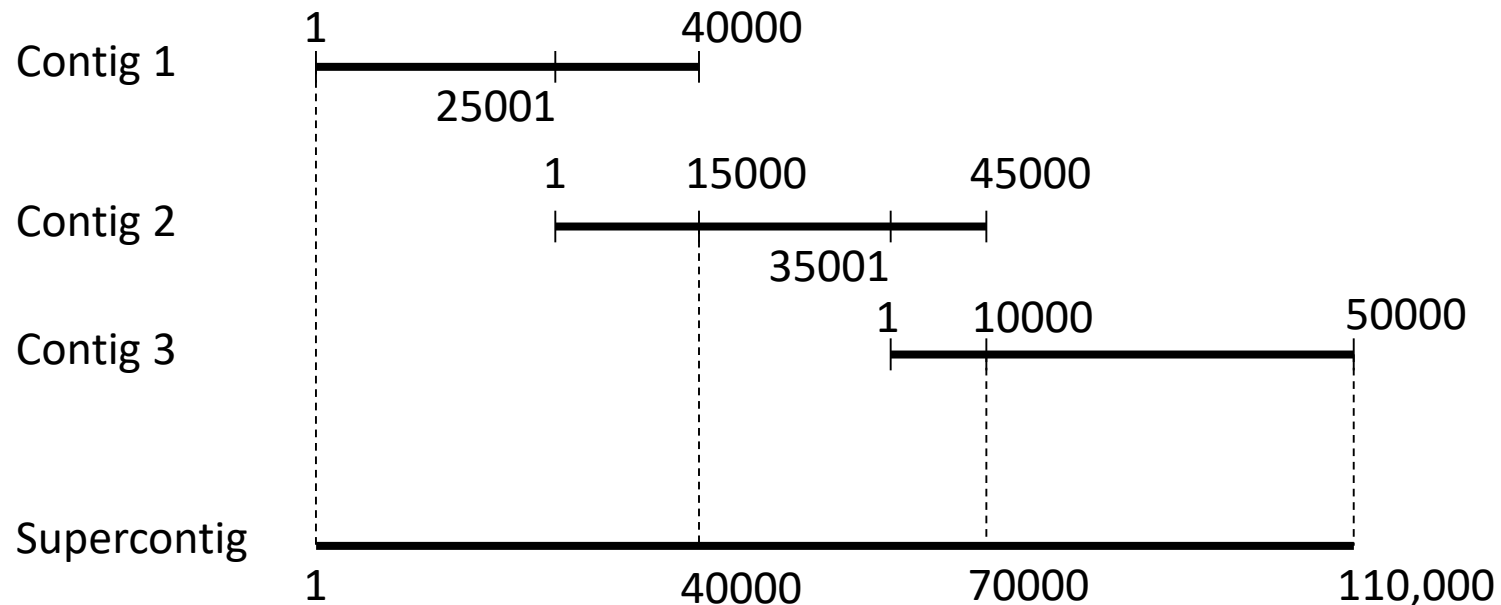
1.5 hr lec (separate)

1.5 hr lab (joint)



Example lab assignments:

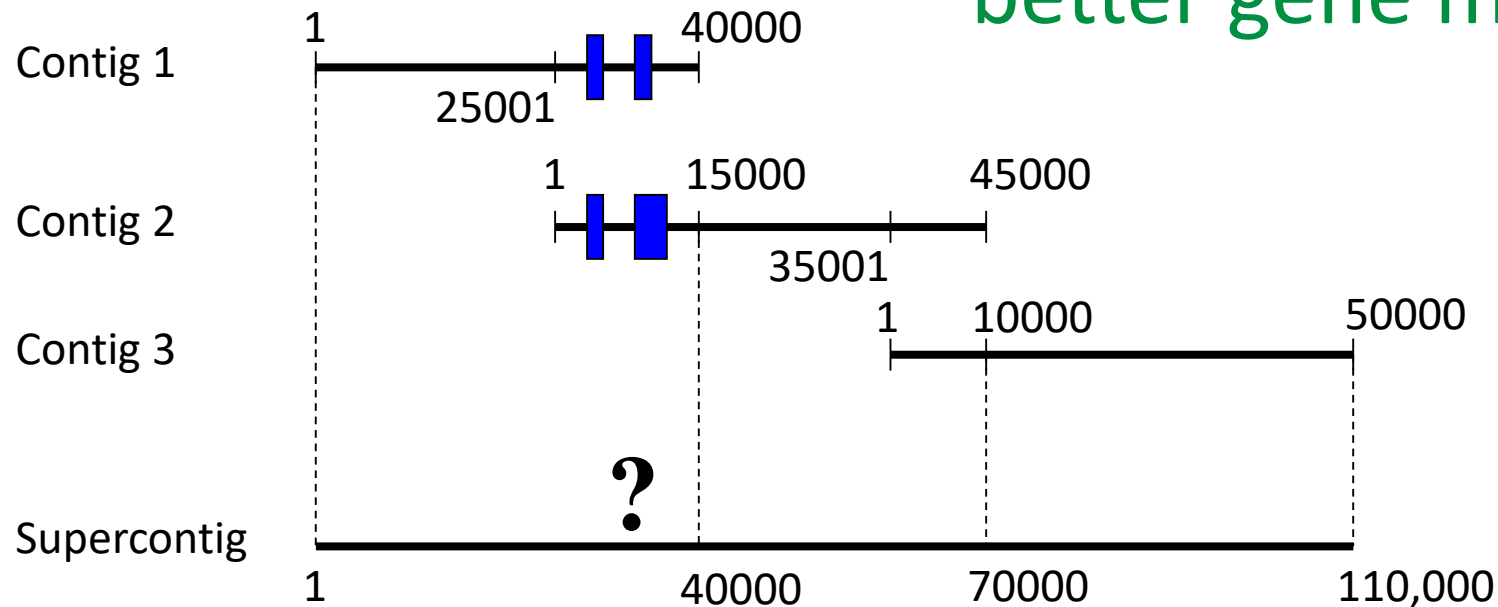
Question/Problem for BIO: How large is the genome region spanned by all annotated contigs?



Programs/algorithms for CS: Contig and GFF merger (string matching, data manipulation)

Example lab assignments:

Questions for BIO: For independently annotated genes, do annotations match? If not, which is a better gene model?



Programs/algorithms for CS: Annotation QC
(global alignment)

In concert learning objective: cooperate effectively with colleagues within and across disciplines

Assessment: student products and exit survey

Questions:

1. What were the **benefits** of working with partners from a different discipline?
2. What were the **challenges** of working with partners from a different discipline?
3. If you overcame the challenges, **how**?

1. What were the benefits of working with partners from a different discipline?

CS: "I liked that most of the pressure from the course came from our biology partners and having things ready for them, rather than pressure just to get a good grade."



1. What were the benefits of working with partners from a different discipline?

BIO: "This was a new experience ... It required all to be respectful of each other's lack of understanding with the other's background but be able to clearly discuss each for precision + accuracy"

BIO: "Learning how different we think from each other and how to communicate more effectively."



2. What were the challenges of working with partners from a different discipline?

“Really the same as the benefits...
differences in background knowledge

and
communication.”



2. What were the challenges of working with partners from a different discipline?

BIO: “They had little/no prior knowledge of the concepts, so everything had to be very clear and precise while also very detailed.”

CS: “Transferring expert knowledge between two disciplines is really hard.”



3. If you overcame the challenges, how?

“Dogged determination, sacrifice, and communication efforts/skills got us through.

When these failed, we failed.”



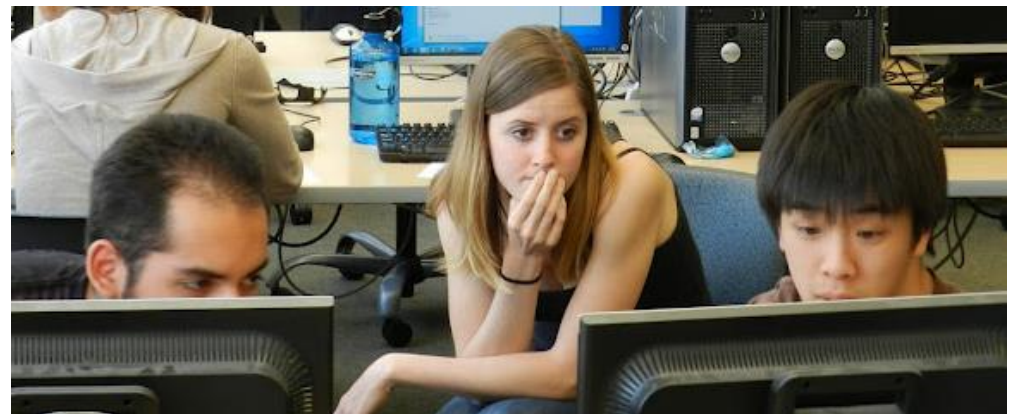
“Learning to speak up when something is not correct on either side of the team and trying to teach each other bits of background info ...”

3. If you overcame the challenges, how?

“by trying to fully understand the information before communicating it to CS partners”

“we had to discuss frequently and rewind our explanations until they made sense”

"We overcame the challenges by talking about the problem and teaching each other things that the other majors did not know"



Student challenges: bridging a cultural divide

Count: 1,2,3,...

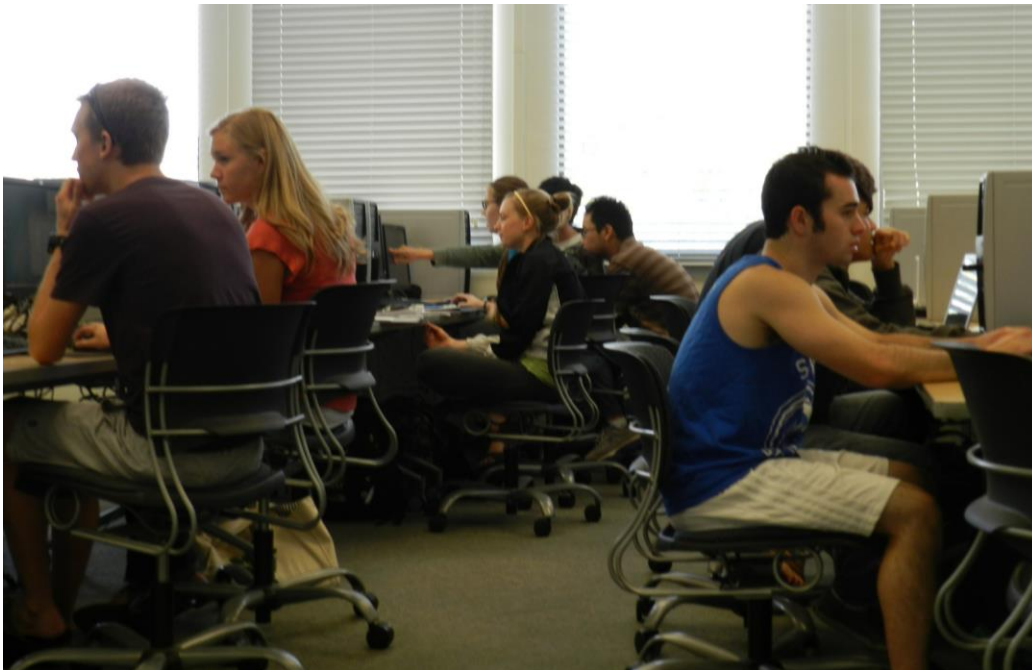
DNA Sequence

Do it over and over

Count 0,1,2,...

String

Loop



Instructor challenges:

- Managing student researchers (keeping track of student progress, providing timely feedback).
- Institutional support (scheduling, rooms, release time for development assignments).
- Timing/organization of assignments.
- Assessment (computational thinking communication skills)

Acknowledgment:

Genomics Education Partnership

Sponsored by Washington University in St. Louis, HHMI and NSF

Cal Poly
Learn by Doing

