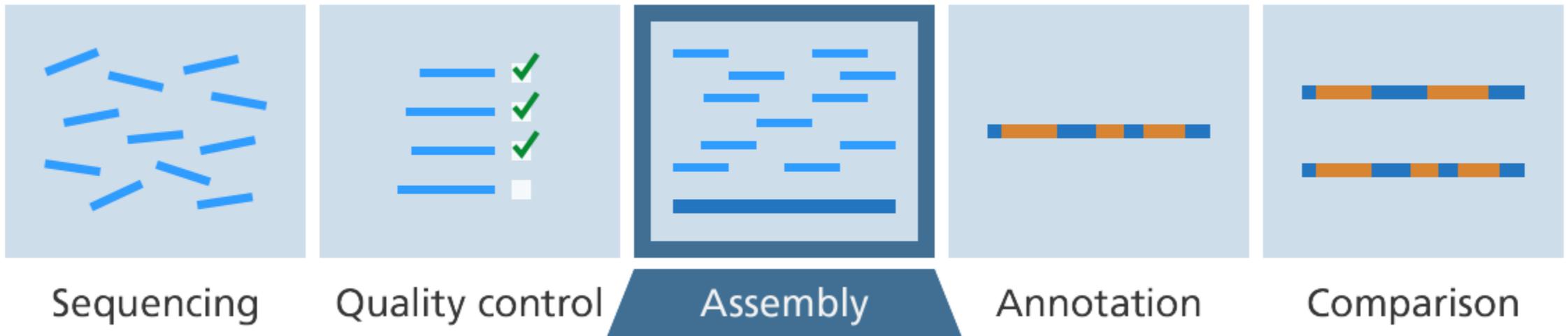


Bioinformatics as a high-impact remote lab activity - applying lessons learned from a remote T3 course

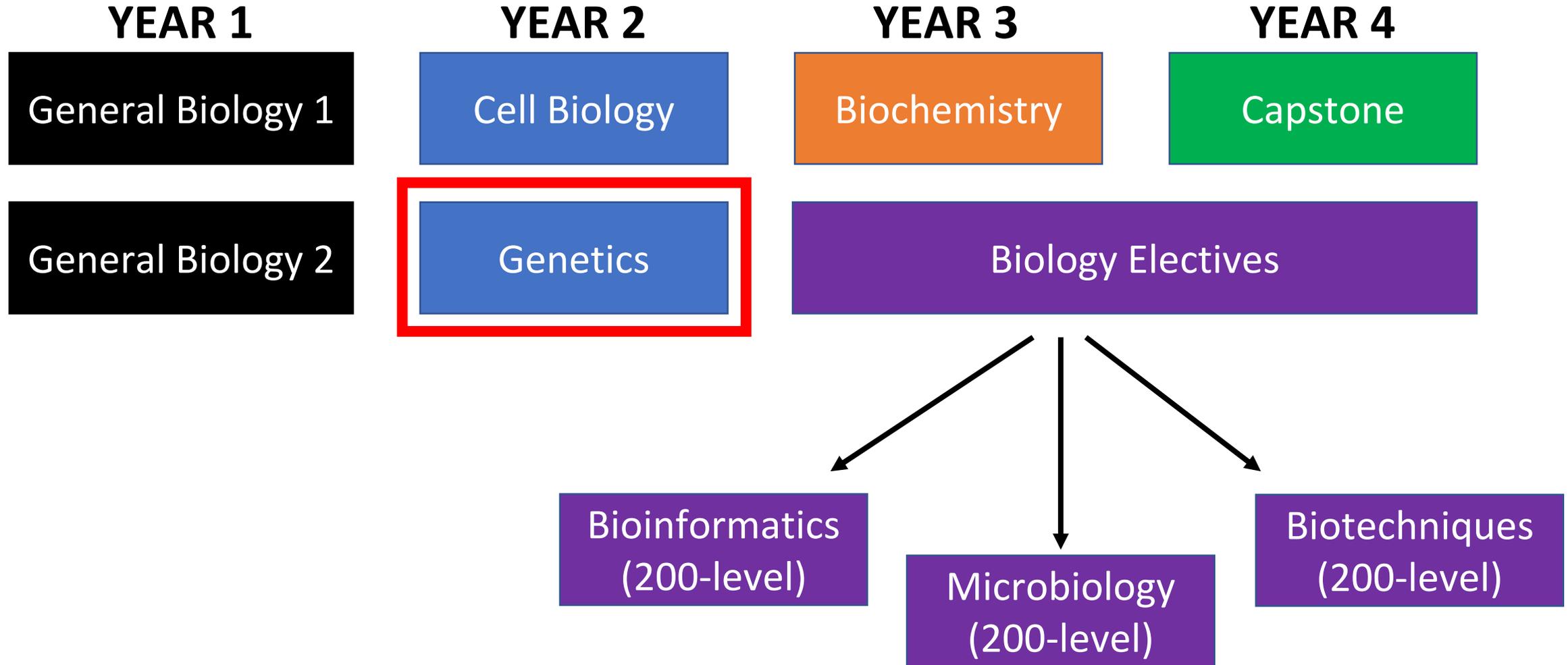
Anne N. Reid, PhD

Assistant Professor, Biology and Biomedical Sciences

Salve Regina University



Bioinformatics in the Biology Curriculum at SRU



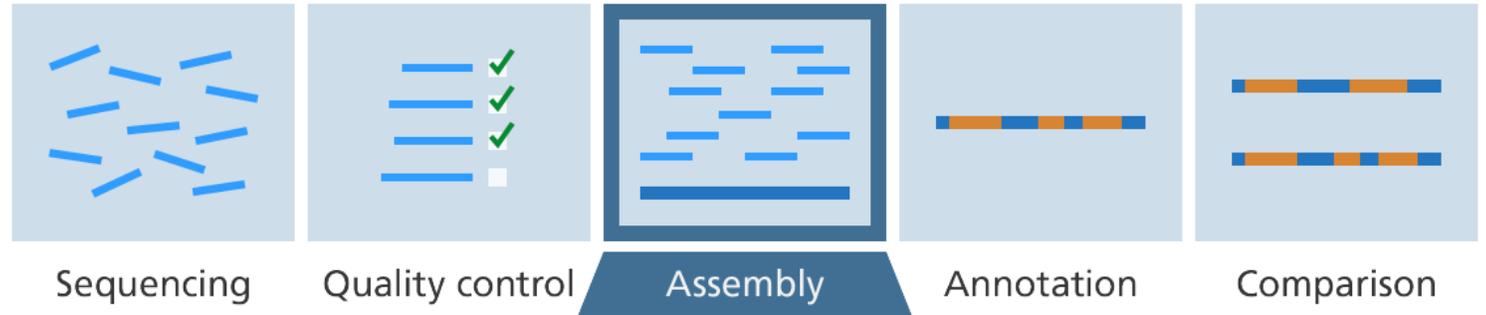
Bioinformatics T3: Train-the-trainer (July 2020)

The Integration of Bioinformatics into an Undergraduate Biology Curriculum

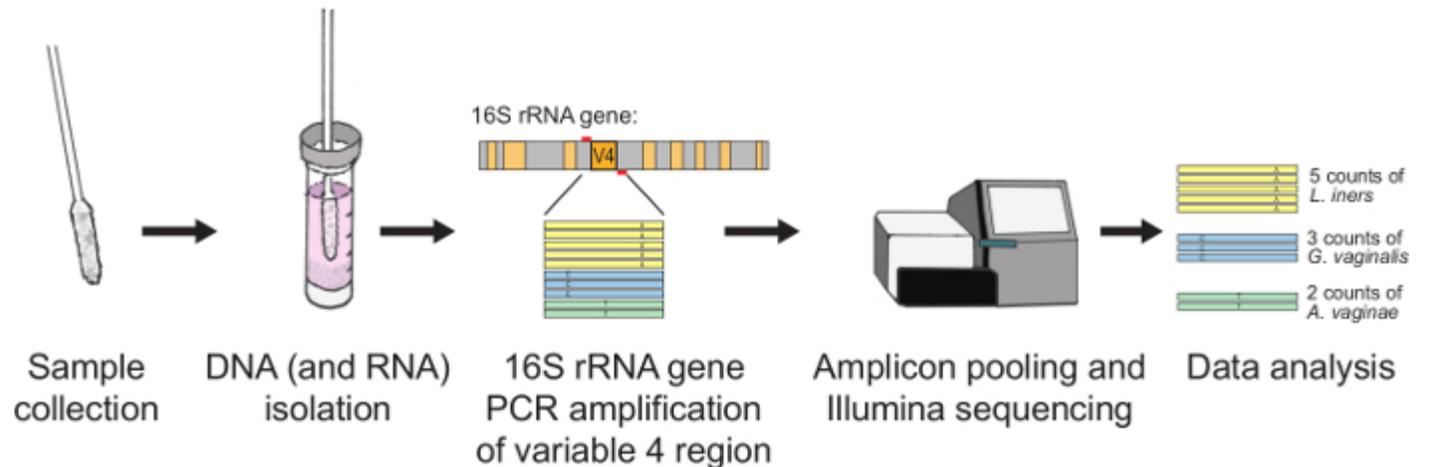
MDI Biological Laboratory, Bar Harbor, ME

NH INBRE

1. Assembly of a novel bacterial genome.



2. Bacterial community analysis via 16S rRNA metagenomics.

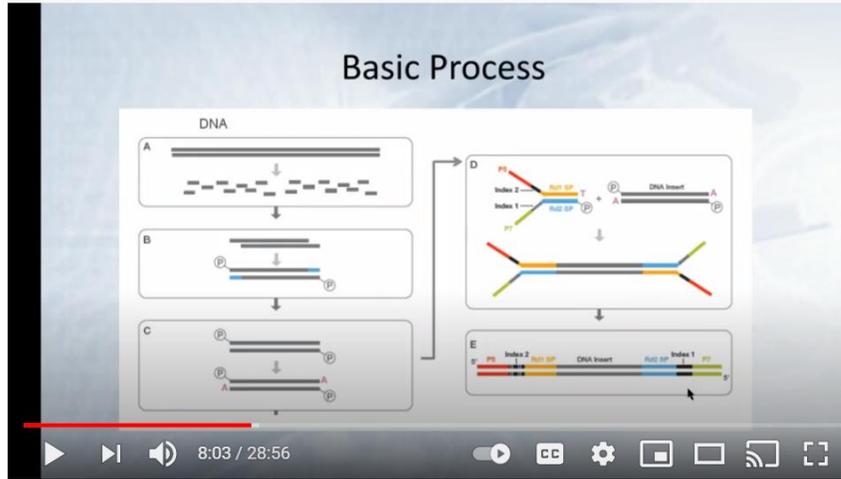


Bioinformatics T3: Train-the-trainer (July 2020)

Theory videos

Access to Ron (data, programs)

Step-by-step instructions



How-to videos

```
hcgsh@hcgsl-Lenovo-G50-80: ~$ ls
12:38:59 airjordan> ls
160329_MKT_Water      example_reads  old                Silval19_release_aligned_rep_files  uniprot_cluster
160601_KerieDavies   Genomic_16s   paladin_data      smithsonian                          uniprot_sprot.dat
160617_HelenAntiBioTest  hassell       Project_ISURF     special_chars.txt                    uniprot_sprot.fasta
anguina_genome        Helena_150424  projects          tagmentation_data                    vibrio_coverage
augustus              maker         Project_Thomas_CinderYeast  test_directory                        transcription_error
bin                   mapped        random
example_assembly     notebooks     Silval19_release

example_assembly> cd ex
example_assembly/ example_reads/
12:39:03 airjordan> cd example_assembly/
12:39:03 example_assembly/
12:39:14 example_assembly> ls
contigs.fasta  prokka_report  quast_report  spades.log  warnings.log
12:39:17 example_assembly>
```

Read Quality Check w/ FASTQC

manual: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

FASTQC explained

- Run Fastqc

FastQC is a program to summarize read qualities and base composition. Since we have millions of reads there is no practical way to do this by hand. We call the program to parse through the fastq files and do the hard work for us. The input to the program is one or more fastq file(s) and the output is an html file with several figures. The link above describes what each of the output figures are describing. I mainly focus on the first graph which visualizes our average read qualities and the last figure which shows the adapter content. Note that this program does not do anything to your data, as with the majority of the assessment tools, it merely reads it.

```
# make a directory to store the output
mkdir fastqc_raw-reads
# run the program
fastqc Sample_*/*_R1_* Sample_*/*_R2_* -o fastqc_raw-reads
ls fastqc_raw-reads
# the resulting folder should contain a zipped archive and an html file, we can ignore the zipped archive wh:
```

Slide sets

The screenshot shows the MDI Biological Laboratory website. The main content area is titled 'Bioinformatics T3' and lists 'Materials' and 'Lecture Slides'. Under 'Materials', there is a list of five items: 1. BASH notes, 2. Genome Assembly and Assessment, 3. Genbank Download and Comparative Genomics, 4. Metabarcoding with Qiime 2, and 5. Downloading Experiments from NCBI. Under 'Lecture Slides', there is a list of five items: 1. Kelley Thomas - Genome Sequencing, 2. Kelley Thomas - Genome Assembly and Annotation, 3. Kelley Thomas - Intro to Metabarcoding and QIIME, 4. Ben King - Genomic Resources (Zoom Lecture Recording), and 5. Karen Bieluch - Draft survey (as of 7-2-20) and Course Evaluation Study slides.

Bioinformatics T3 goes remote

- + Screen share (dual monitors)
- + No one left behind
- Engagement (zoom fatigue, pace, distraction)

```
hcgsh@hcgsh-Lenovo-G50-80: -
12:38:59 airjordan> ls
160329_WKT_Water          example_reads  old           Silv
160601_KerieDavies       Genomic_16s   paladin_data  smit
160617_HelenAntiBioTest  hassell       Project_ISURF  spec
anguina_genome           Helena_150424 projects      tagm
augustus                 maker         Project_Thomas_CinderYeast  test
bin                      mapped        random        [redacted]
example_assembly         notebooks     Silval19_release  tran
12:39:03 airjordan> cd ex
example_assembly/ example_reads/
12:39:03 airjordan> cd example_assembly/
12:39:14 example_assembly> ls
contigs.fasta  prokka_report  quast_report  spades.log  warnings.log
12:39:17 example_assembly> █
```

Bringing bioinformatics to the SRU Genetics lab

Modality

- face-to-face, 12 students in lab
- few students needing to attend remotely (computer access, quarantine, illness)

Experiment

- genome assembly & annotation

Data Set

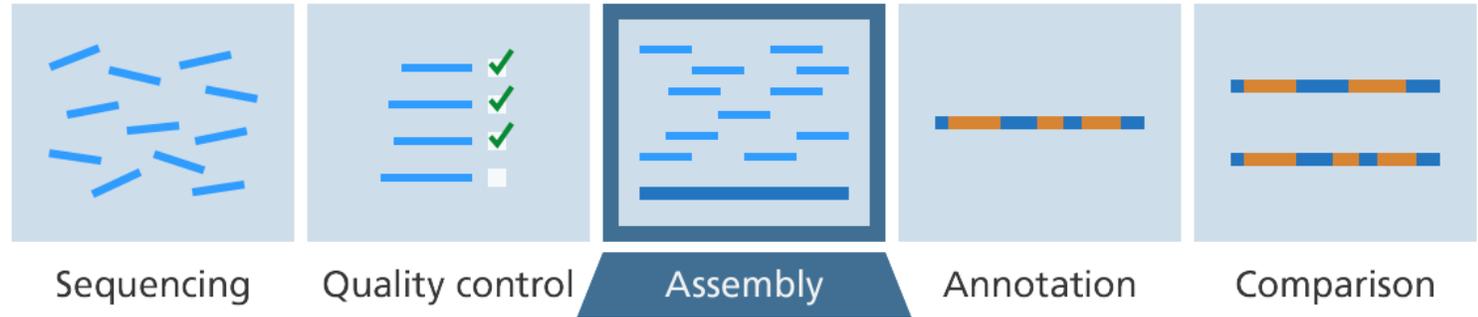
- sequence data for *Staphylococcus* isolates recovered from animal sources
- each student assigned one isolate

Schedule

Week 1: Logging onto Ron and Introduction to BASH

Week 2: Read trimming and quality assessment, submit genome for assembly

Week 3: Genome quality assessment and annotation



Pre-lab activities



Bringing bioinformatics to the SRU Genetics lab

	#Raw data				QUAST						#BUSCO					
First	Num_reads	Total bp	Estimated_Coverage	Calculated Coverage	Genome_Length bps, can refine after	Real Sequence Length	Largest Sequence	# Contigs	N50	GC%	Single-Copy	Duplicated	Fragmented	Missing	Total # of BUSCOs	Ortho database (Firmicutes or Bacteria)
	# reads x 2 x 250	bps / genome length														
	1,812,230	906,115,000	324	315	2800000	2724833	834768	60	288611	32.68	223	1	1	7	224	Firmicutes
	1,723,602	861,801,000	308	285	2800000	2867391	641856	161	254725	32.79	227	0	0	5	232	Firmicutes
	2,096,221	1,048,110,000										0	0	9	148	bacteria
	1,970,611	985,305,500	351.89									0	1	9	148	Bacteria
	1,432,882	716,441,000														
	1,869,588	934,794,000	311.56									0	0	5	232	Firmicutes
	1,808,029	904,014,500										0	0	5	97.8	firmicutes
	1,434,940	717,470,000										0	0	5	232	firmicutes
	1,514,414	757,207,000	270.43									1	0	0	97.8	Firmicutes
					2800000	2724833	834768	60	288611	32.68						
					2800000	2867391	641856	161	254725	32.79						
					2800000	2875490	881177	71	379725	32.68						
					2800000		703722	29	391028	32.78						
					2800000											
					2800000	2818328	625293	133	268492	32.7						
					2800000	2799994	465340	47	238241	32.79						
					2800000	2755989	701612	24	325980	32.73						
					2800000	2840053	881178	44	379725	32.74						
					2800000	2442451	434917	109	232218	35.87						
					2800000	7566564	228283	3975	3693	41.91						
					2800000	7566564	228283	3975	3693	41.9						

Success Stories

Hearing students 'speak bioinformatics'

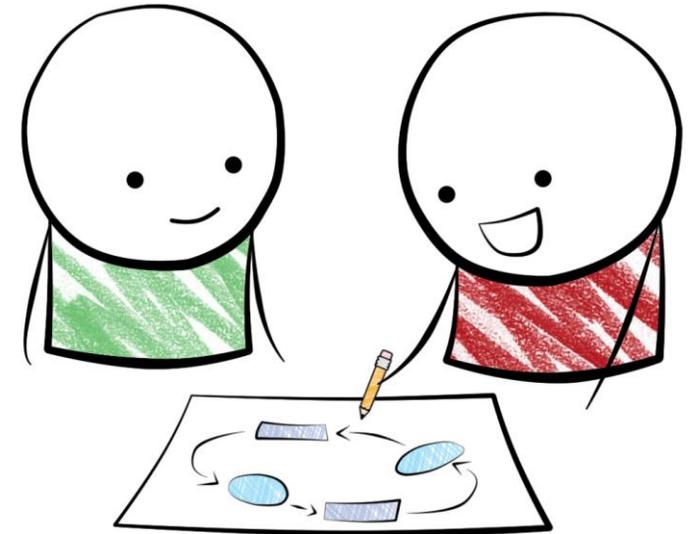
...improved quality score after trimming...

...when I grepped the file...

Students finding a home



Peer teaching



Key Realizations

- leave ample time for logging onto Ron (not just week 1)
 - username/password: Slack / email from Devin
 - typing passwords 'blindly' (particularly during password reset)
 - assign this before first lab
- no one left behind only works if students speak up
 - especially difficult in remote / hybrid environment
 - a few students fell entire class behind
 - expand spreadsheet to include more of the data collected
- preferred modality – face-to-face with students logged in to WebEx/Zoom
 - face-to-face improves focus, attention
 - share screen when stuck, whether remote or in person
 - keep remote students feeling like part of the class
 - maintain social distancing in face-to-face setting

Key Realizations

- students need help understanding what constitutes data in this world
 - what goes into Notebook
 - construct tables / figures of data, results / discussion statements
 - NCBI to collect data from other published genomes, compare to their own
- instructor notes helpful for students seeking to catch up
 - posted to Canvas AFTER that day's lab
 - prevent students from jumping ahead
 - useful for students seeking to repeat analysis

Thank You!

RI-INBRE - support to attend T3
NH-INBRE – resources, support

*happy to discuss, share materials,
hear suggestions, etc...
anne.reid@salve.edu