Assignment 7 - Lesson Re-design

Helena Rasche

2022-02-07

The lesson to be re-designed is Computational Biology, Day 3: De Novo Genome Assembly, which unfortunately currently experiences poor coherence between assessment, learning objectives, and learning activities. This will be improved by taking the solid core of the lesson and expanding it to further test student's abilities and align with overall course and program goals.

The situation

Currently the lesson features the following components:

Aspect	Currently
Learning Outcomes	LO1) Compute and interpret a whole genome assembly, LO2) Judge the quality of a genome assembly.
Learning Activities	Conda (LO1), Spades (LO1), Qualimap (LO2), Comparison (LO2)
Assessment	k-mer exploration (no LOs.)

These are organised in the lecture as:

- An introductory lecture over genome assembly techniques
- Demonstration of assembly
- A hands-on portion where students run the assemblies (LO1)
  - Here students run "Conda", "Spades"
- Discussion of genome quality (LO2)
  - Here we do "Qualimap", "Comparison"
- Some formative assessments over their impressions of genome assembly qualities

Areas for Improvement

This results in very poor coherence currently; the described outcome "compute and interpret a whole genome assembly" was marginally achieved, though the 'interpret' step is somewhat lacking. However LO2 is missing very key components:

- There is minimal discussion of quality metrics, and what constitutes a good or bad result.
- There are no examples of Good, Mediocre, or Bad assemblies>
- While visualisations are available, they are not used.
- The data is assembled twice, with two different settings, which result in essentially indistinguishable results (to a trained eye!).

And this needs to be significantly improved, there are numerous low-hanging fruit which can be addressed. Additionally the assessment portion involves a take-home homework problem set is very theoretical, incredibly abstract, and related to genome assembly at a very low level. This hampers students applying knowledge obtained in class to effectively meet the learning outcomes.

Context of the Lesson

This lesson is part of the 3rd year course of BML where students need to obtain such skills as:

- Computer skills: [. . .] bio-informatics tools, [. . .]
- Research skills: problem analysis, research questions, [. . .]

And exists within the Minor Bioinformatica as a 2 EC class, where students are expected to:

> [. . .] you will focus on handling, storing, retrieving and visualizing massive amounts of biological data. While data generation is faster than data management, visualization gets more and more important. [. . .]

Improving Lesson-Context Coherence

The above context section gets to the core of the problem.

But here we analyse two relatively small genomes, this should be scaled up significantly across more samples to truly meet the goals of the course. This would additionally let us introduce variety of data, good to bad, and let the students separate those based on their quality scoring criteria like the N50, average read depth, or regions of low read depth as seen in Figure 1. On top of that, we fail to exercise good visualisation of the results that are readily available such as via Bandage (Figure 2, [Wick et al., 2015]), and need to introduce that so students have better comprehension of an incredibly abstract topic.

Students do not get the chance to explore the problem space themselves and optimise the results, something that would be improved with integration of more and larger datasets, allowing students to practice problem analysis and analysing questions applicable to later research.

Content Improvement Plan

The initial part of the lesson covering assembly background is useful theoretical knowledge for students and should be kept, but maybe augmented with student activities like "assembling" some sentence of text, to give students an idea of what machines do in the background. Especially if students are expected to understand k-mers. This activity is visualised in Figure 3.

- Item 1: Text assembly activity, students split into 3 groups, and get to assemble a sentence split with different K-mer values to
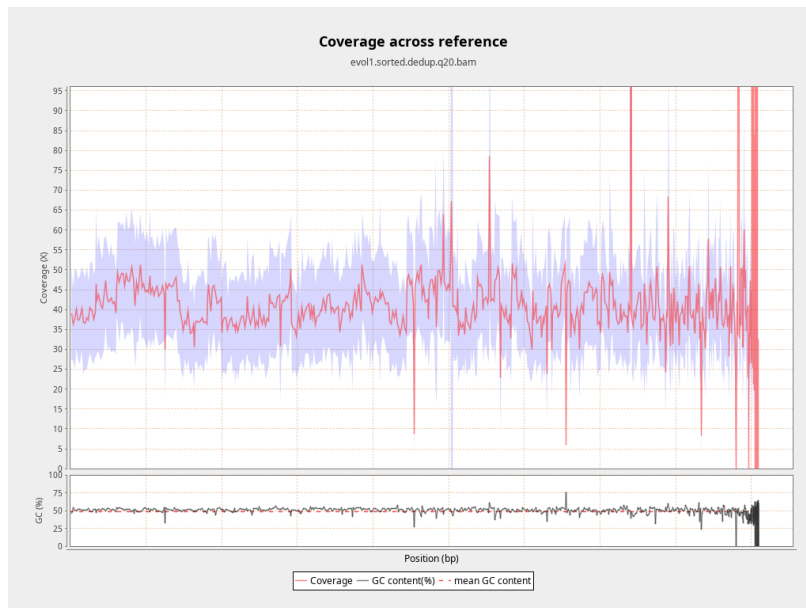
Figure 1: A graph from the tool qualimap, showing average coverage across the reference, areas of low coverage may be suspicious and require investigation, or reassembly, or resequencing.
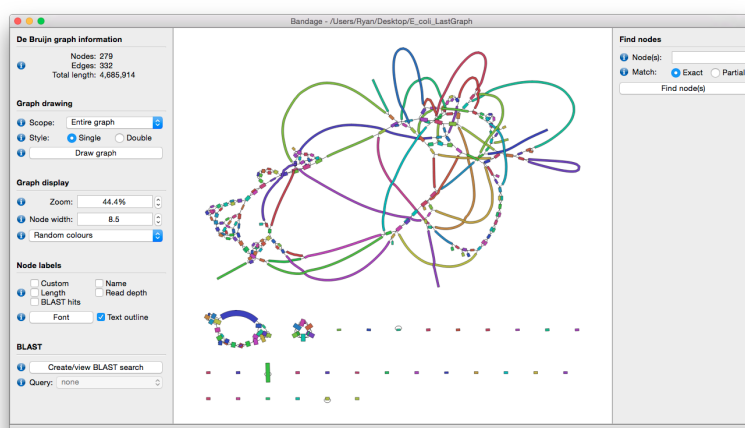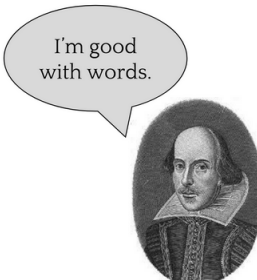

Figure 2: A genome graph extracted from the assembly process and visualised with Bandage, showing relatively poor assembly and lots of repeats and ambiguous sequences.

understand consequences.

- **_Reads_**
  ```
  ds, Romans, count
  ns, countrymen, le
  Friends, Rom
  send me your ears;
  crymen, lend me
  ```

- **_Overlaps_**
  ```
  Friends, Rom
      ds, Romans, count
          ns, countrymen, le
              crymen, lend me
                  send me your ears;
  ```

> I'm good with words.

Figure 3: A screenshot from the slide deck covering assembly, this would be turned into an activity. Students would be split into three groups, each group with a different read length (i.e. k-mer), and each group would try and re-assemble the original text based on a pile of papers handed to them. This could be conducted with text in any language, even intentionally in a language students don't speak, to help them remove the effects of previous knowledge on solving the puzzle. One group would have 5-mers and would struggle, the other 7-mers and do better but still struggle with some repeats, and the last would have something longer like 15-mers to experience how that improves the process. The activity as written is equivalent to short and long read single-end sequencing, but we could use it as well to introduce paired-end sequencing, via another round of short-read sequencing but with paired-end results (i.e. 5 letters, a gap of 10, another 5 letters as one piece of paper.)

Next, the lesson needs to expand to include a section on evaluation of genomes by quality metrics as well as a discussion of those major quality metrics that are used in the field.

- Item 2: (Short) presentation of quality metrics (e.g. N50, Bandage), explanations of their meaning, and then good vs bad results.

Finally during the lesson, we will replace the dataset we're using such that we can have multiple assemblies of various resulting qualities. Potentially this can be done through a small workflow further exercising bioinformatics skills they need in the workplaces of their future employers.

- Item 3: More, smaller genomes of varying quality, tasking students to return with a ranking of best to worst assemblies.

Lastly the assessments must be augmented, discussing k-mer contributions is good but not sufficient. This must be expanded to include a further assembly and quality evaluation, asking them to explore the parameter space on their own: which parameters resulted in the best assembly? Why do they think that is the best possible assembly given the data?

Updated Learning Objectives

- Compute multiple whole genome assemblies in such a way to develop big data processing skills (Apply+Procedural)
- Learn to evaluate quality metrics so that they can separate good and bad assemblies (Analyse+Conceptual, Evaluate+Procedural)
- Visualise assemblies so that they understand presentation of various failure modes (Apply+Procedural, Evaluate+Conceptual)

Lesson Re-design

Students will attend this lesson with previous theoretical experience doing assemblies, this portion is a review for them. They have gone

avans _hogeschool_

through the motions of assembly but not done it by hand, nor understood the intricacies of parameter selection. Students at this stage in their career are planning to go on to a company or research institute where they will need to apply these skills to analyse genomic sequences and help coordinate and design sequencing projects. When those sequencing experiments occasionally fail for various reasons, they will need to understand why they failed and how to resolve those issues, be it parameter exploration or resequencing. This lesson should serve students well as a very practical lesson delving into comparative analysis which provides key information for them.

The above described updated learning objectives will allow students to efficiently approach targeted skills and knowledge as described in the minor bioinformatic and BML programmes above. This lesson falls near the end of their curriculum and as such can spend more time focused on an in-depth understanding of assemblies and their associated failure modes. The module is titled "Computational Biology", which is furthered through their development of genome assembly skills and computational scaling of analyses across multiple samples, and visualisation thereof. This is a task of which they need to become Intermediate practitioners, in both knowledge and application, within the 2ECs available. By this point in their career they have already become Novices with assembly and genomics as topics, letting them now deepen their knowledge and meet overall Minor Bioinformatica Learning Objectives such as massive data handling and visualisation.

This lesson's theme is assembly which is supported through several phases and materials: 1. an assembly activity 2. an initial presentation on assembly 3. a demonstration portion 4. a hands-on portion where students work in duos to assemble multiple genomes 5. a subsequent presentation on quality metrics and visualisation 6. a further hands on where students apply metrics, visualise, and evaluate all genomes for quality 7. a group discussion of the results

These will be matched formative assessments in the form of homework problems which they need to resolve on their own given knowledge from the lesson. Here they will be tasked with assembling a genome and exploring the parameter options to optimise the resulting assembly. This will be complemented with a discussion section where they should explain why they've chosen those specific parameters to modify. This will aim to assess their ability to independently answer research questions.

The work forms used in this redesigned lesson will involve lectures, problem based learning (genome assembly with real world data), group work in duos (assembly, quality evaluation), and finally group discussion with peer teaching (as students will explain which one is better and why to each other.) The in class teaching time is 2.5 hours which should allow plenty of time for varied work forms to help students achieve learning goals.

References

Ryan R. Wick, Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, 31(20):3350–3352, 06 2015. ISSN 1367-4803. DOI: 10.1093/bioinformatics/btv383. URL https://doi.org/10.1093/bioinformatics/btv383.

avans
hogeschool

LESSON PREPARATION FORM

Lecturer: Helena Rasche   Date:

 Group: ATGM/BML   Number of students: 10   Classroom:

Subject/lesson: Computational Biology, Lesson 3: Genome Assembly

Starting situation:
What do the students already know about the subject and what can they already do? How do they feel about it? Have they already gained work experience?
Describe the composition of the group. When and where does the lesson take place? And similar.

Students will attend this lesson with previous theoretical experience doing assemblies, this portion is a review for them. They have gone through the motions of assembly but not done it by hand, nor understood the intricacies of parameter selection. Students at this stage in their career are planning to go on to a company or research institute wh ere they will need to apply these skills to analyse genomic sequences and help coordinate and design sequencing projects. When those sequencing experiments occasionally fail for various reasons, they will need to understand why they failed and how to resolve those issues, be it parameter exploration or resequencing. This lesson should serve stude nts well as a very practical lesson delving into comparative analysis which provides key information for them.

Objective/lesson objective:
Describe the objective(s) of the lesson according to the 3C model, taking account of the taxonomy level according to Bloom.

LO1: Compute multiple whole genome assemblies in such a way to develop big data processing skills (Apply+Procedural)
LO2: Learn to evaluate quality metrics so that they can separate good and bad assemblies (Analyse+Conceptual, Evaluate+Procedural)
LO3: Visualise assemblies so that they understand presentation of various failure modes (Apply+Procedural, Evaluate+Conceptual)

Educational resources:
Which learning materials do you use during your lesson? (book,
smartboard, whiteboard, paper, etc.)

Presentations (powerpoint), documentation (webpages), assembly activity (web-based group activity with paper assembly fallback during in-person lessons), compute cluster which students access remotely to run computations and get hands-on experience.

Assessment procedure:
Provide a description of the final assessment of the unit of study and at which of Bloom's taxonomy levels the assessment will be carried out.

At the end of the lesson they will go home with an assignment including an unknown genome sequence which they will need to assemble and check for quality. (Apply+Procedural) This task will require re-assembling the genome with several different parameters to optimise the assembly and get the best genome out (Evaluate+Procedural). They will then need to write a short statement on why they choose those parameters and what has led them to believe that that is the optimal genome assembly (Evaluate+Concetual)

Schedule (how long?)	Content (what?)	Teaching and learning activities/work forms (how?)		Justify: how will this be used to reach the objective?
		Teacher	Student	
"Cold Open" 20 minutes	We'll begin with the assembly exercise, students given either paper or digital pieces of paper that they need to re-assemble into their original sentence. Some will have mistakes or low coverage portions so they can make some guesses	Activity introduction, Observation	Students will apply existing knowledge from e.g. legpuzzel solving to assemble the sentences	This goes towards **LO1**, learning about the procedure of whole genome assembly by giving them a fun introductory activity where they can transition from excitement of being in the class to a critical thinking state and begin to Apply an algorithm and begin to Analyze the algorithm they're making intuitively.
Lecture ~20 min	Now we'll take the applied knowledge and turn theoretical, focusing on how algorithms work in practice	Presentation	Listen / note taking / knowledge integration	Here we focus on the Understand+Procedural components of **LO1**, allowing students to know how assembly algorithms work at a theoretical level so they can begin understanding what parameters they can adjust, and what effect they might have.
Assembly: Demo ~10 minutes	Here students will see a demo of assembly and listen to the teacher's explanation of parameters	Demonstration	Listen / note taking / memorisation	Students now obtain the precise knowledge they'll need to Remember+Procedural steps required to accomplish assembly (**LO1**)
Assembly: Hands-on ~20 minutes	Students move into breakout rooms where they do the assembly for multiple genomes, and report on their results. (most/fewest contigs (grep), biggest/smallest (wc))	Observation	Hands-on activity where students work in duos to accomplish the activity and report back results	Students Apply+Procedural knowledge from the previous step to do the assembly and then Analyse+Factual results of their process. (Complete **LO1**)
~15 min	Break			
Quality Contrtol: Demo ~20 minutes	Instructor demonstrates quality control and discusses several metrics that are used. Demonstrates QUAST. Demonstrates Bandage.	Demonstration	Listen / note taking / memorisation	Students now obtain the precise knowledge they'll need to Remember+Procedural steps required to accomplish assembly (**LO2** & **LO3**)
Quality Control: Hands on ~30 minutes	Students move into breakout rooms where they do the quality metrics across their genomes, and investigate bandage where necessary.	Observation	Hands-on **activity** where students work in duos to accomplish the activity and report back results	Students Apply+Procedural knowledge from the previous step to do the assembly and then Analyse+Factual results of their process. Additionally they need to Evaluate+Procedural to conclude the best/worst assemblies. (Complete **LO2** & **LO3**.)
Group Discussion 10 minutes	Students return from breakout rooms and discuss their results. Which were best, which were worse, did anyone have differing observations	Lead discussion	Students report on their results, and discuss when there are conflicting answers	Here they Evaluate+Factual results of their assemblies against other groups. Potentially the ordering can be different due to factors teachers control, giving room for discussion. They need to Evaluate+Conceptual their results and check for inconsitencies.
Conclusion: ~5 Min	Recap k-mers, assembly, tools used (conda, spades, quast, bandage), discuss homework assignment (assemble and change parameters to optimise! Read the documentation!)	Presentation	Listening	Students will now need to use what they learned from **LO1**, **LO2**, & **LO3** to evaluate an unknown genome's assembly and optimise the results, as well as explainin why they choose those parameters based on available information. They will need to read documentation to accomplish this.