Title: A Lecture on High-Throughput Sequencing Using a Specific Scenario and Clicker Questions

Jenifer Cruickshank

Affiliations:

State University of New York at Oswego, Oswego, New York

Scientific Teaching Context Page

Learning Goal(s):
- Students will understand the basic processes of next-generation sequencing (NGS) and sequence analysis.
- Students will appreciate what kinds of research questions can be answered with information garnered from NGS data.

Learning Objective(s):
- Choose an appropriate sequencing technology for a particular research question.
- Define the types of differences one would look for when comparing two genome sequences.

1. Introduction:
This lesson is intended to introduce students to next-generation sequencing (NGS) technologies, the types of data NGS generates, and why one would be interested in large-scale sequence information. The set-up of the lesson is a clicker-based lecture that will present a scenario at the start of class that progresses (virtually) through (re)sequencing, assembling, and analyzing a eukaryotic genome, then comparing the newly-generated genome to the pre-existing genome sequence of a closely-related species.

This lesson is intended for one lecture period (80 minutes) in a Genetics course. Enrolled students are primarily second and third year undergraduate biology and zoology majors with a smaller proportion of biochemistry majors. This class period would be late in the semester when students should have a solid understanding of DNA structure, what a gene is, eukaryotic gene structure, and basic principles of evolution. Students will prepare specifically for this lecture by reading several relevant sections on genomics in the course textbook and by watching several online videos on different sequencing techniques (Sanger, Illumina, NanoPore?)

2. Scientific Teaching Themes:

Active Learning: Students will actively engage with the material with clicker questions, which will be of the "What should be our next step?" variety. A clicker question will be posed, students will answer without discussion. Assuming a small minority answer correctly, a mini-lecture on the subtopic will be presented followed by student discussion with a neighbor then a reposing of the question. There will be a homework question asking for another specific application of high-throughput sequencing.

Assessment: The clicker questions will provide an assessment of students' own learning in addition to assessment information for the instructor. A homework assignment after the class will ask student to expand beyond the specific example from the lecture. Subsequent exams will include questions over the material covered in this lecture.

Inclusive Teaching: Answers to clicker questions are anonymous and allow full participation by students regardless of their level of extroversion. The follow-up homework question allows students to identify an application of high-throughput sequencing that is of interest to them.

3. Lesson Plan:

Pre-class preparation:

Because of the lengthy run times for most eukaryotic genome analysis software programs, all necessary analyses will have already been done and be accessible during class. In this case, it is whole genome sequence from a yak. Access to the internet during class is also necessary. Visual lecture information and clicker questions will be presented on PowerPoint slides.

In-class:

The set-up: A rich hobby farmer wants to know if the yak (*Bos grunniens*) bull she just purchased is pure yak or whether he has any cattle (*Bos taurus*) in his pedigree. She is willing to pay for any needed "DNA tests" to answer this question.

Clicker question 1: What would be good to know at this point? (Granted, this one is a bit of a gimme.)

  A. Is yak genome information available?
  B. Is cattle genome information available?
  C. Can we see a picture of the bull?
  D. Both A and B would be good to know.

Presumably, D is chosen by a large majority which prompts an internet source by the instructor for "yak genome" and "cattle genome". The yak genome search should turn up the Yak Genome Database, and there are multiple options for the cattle genome, I would go with Ensembl. Information from the front page of the Yak Genome Database notes the size of the genome (2657 Mb) and that coverage is 65X, prompting

Clicker question 2: What does 65X coverage mean?

  A. Every base is present 65 times in the yak sequence data.
  B. On average, a given base is present 65 times in the yak sequence data.
  C. 65 yak genomes were sequenced.
  D. 65% of the yak genome is included in this yak sequence data.

Presumably, B is not chosen by a large majority which prompts a minilecture on the process of high-throughput sequencing with examples of Illumina and NanoPores(?).

Repose clicker question 2.

Continue the narrative, we get a blood sample from the yak and extract the DNA (it's really good quality). Now what?

Clicker question 3: What do we do with the DNA?

      A. Sequence the mRNAs from this blood sample.
      B. Sequence the DNA on an Illumina platform.
      C. PCR a few genes and run the PCR products on a gel.
      D. Sequence the mtDNA.

Depending on how students answer, briefly review high-throughput sequencing and discuss incorrect answers.

Continue the narrative: we prepare and send the DNA away for sequencing. Show what the data files look like when they come back.

Clicker question 4: What should we do with this data first?

      A. Assemble the sequences into contigs.
      B. Blast the sequences against the database.
      C. Assemble the sequences into one long DNA sequence.
      D. Assess the quality of the sequences.

Presuming a majority do not correctly choose D, give a mini-lecture on how sequencing errors can arise and how they can be identified (phred scores and kmer graphs).

Repose clicker question 4.

(Yet to be completed.) Continue along through genome assembly, gene annotation, and comparative genomics. The final clicker question may be what parts of the genome should we compare to find out if the yak bull has cattle DNA. Certain genes? coding sequence? introns? SNPs? Followed by discussion on that topic.