## Microteaching Lesson Plan Template

*In this lesson plan, please include detailed annotations that articulate the rationale of your instructional decisions, based on the scholarship of teaching and learning, learner-centered teaching literature, and your own experiences as a teacher and/or learner.*

### Session Information:

|  |  |
| --- | --- |
| **Instructor Name: Ken Youens-Clark**  **Instructor Status:** GTA  **Course Title: BE487/587 Introduction to Metagenomics**  **Lesson Topic: Using k-mers to Determine Sequence Similarity**  **Date & Time: 23 October 2018**  **Place/Site:**  **Number of students: 12** | **Materials Needed:**   1. Room equipment / Integrated tools: HDMI connector to projector 2. Other: |
| **Context:**  At this point in the semester, where are your students? What process can you use? At what level of challenge?  Students would have been introduced to the basic concepts of metagenomics which is the study of communities of microbes (bacteria, fungi, viruses, archaea, etc.) of unknown composition. One goal of many studies in metagenomics is to identify the organisms present and/or their function which requires comparing the DNA (or RNA) sequences either to databases of known sequences or other samples (perhaps of unknown composition, too). K-mers allow for computationally rapid comparison as compared to alignment which can be more accurate. The challenge is to make students understand the trade-offs (speed/accuracy) when choosing between k-mers vs alignment. | **Instructional Objectives:**  (Formula: “Students will be able to”+ learning behavior + content + means/strategies)  1) Discuss what sequence similarity means (e.g., written languages, DNA)  2) Identify meaningless vs meaningful information (e.g., articles in English/repetitive DNA vs unusual words/important genes)  3) Identify k-mers (AKA n-grams)  4) Contrast k-mer analysis to alignment (speed/accuracy)  5) Explain varying statistical significance of longer k-mers in DNA |

### 

### 

### Session Schedule:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Time** | **Agenda Point** | **Teach from Objective(s)** | **Instructor - What I Do** | **Students - What They Do** |
| i2-2:05 | Similarities of texts | #1 | Engage students in discussion of (dis)similar texts e.g., newspapers (WSJ vs NYT), authors (Hemingway vs Austen), styles (poems vs scientific journals) | Students will ponder genres and writing styles that influence composition which is communicated via WORDS which is the central idea (words/meaning == DNA) |
| 2:05-2:08 | Consider meaning of words | #2 | Some words are common and convey nothing novel (a, an, the). Some words are repeated and set a text apart. Some words are spelled differently in different contexts (Am vs Br English) but convey the same information. | Students will be asked to identify the words in presented texts which do/don’t convey important information; consider minor respellings as sequence variations in DNA that do not alter genetic content. |
| 2:08-2:10 | Identify k-mers | #3 | Explain n-gram/k-mer | Students will write 3-mers of presented sequence on whiteboard |
| 2:10-2:12 | Show alignment | #4 | Show pictures of good/messy multiple sequence alignment (MSA) | Ponder the difficultly of implementing code to do MSA |
| 2:12-2:13 | Explain statistics | #5 | Explain the statistics of finding a given k-mer in a given sequence, show how this lead to discovery of a promoter |  |
| 2:13-2:14 | Show word vs kmer analysis of En/Fr/Ge/It texts | #4 | Show clustering of texts from 4 languages when counting words vs k-mers | Discuss the different clusters WRT words vs k-mers esp considering the size of k |
| 2:15 | Summary / Closure | all | Recap k-mers and alignment | Profit! |

### Comments/Notes:

This is a shortened version of a lecture I developed last year for the Metagenomics class I helped Dr. Bonnie Hurwitz teach. I’ve given it 3 times now, varying the focus depending on the audience (once was to a high school class). It’s usually an hour long, and so I’m struggling which parts to sacrifice for time. I may have to cut the stats part which is technically more important but less entertaining than the pictures comparing texts in varying languages.