

A multi-week metagenomic lab module for Introductory Biology Students

Karen Hogan^{1*} and Linda Robinson¹

¹Department of Biology, University of Pennsylvania, Philadelphia PA 19104.

*Correspondence to: Karen Hogan, 433 S University Ave, Philadelphia, PA 19104
hogank@sas.upenn.edu

Type of Manuscript: *CourseSource* Lesson Manuscript

Abstract

In order to engage our students in scientific research from the beginning of their undergraduate academic career, this laboratory module will introduce students to cutting edge molecular techniques, including high throughput sequencing, to answer student-generated questions about microbial diversity in their own environment. The development and adoption of culture-independent technologies, such as high throughput sequencing, has revolutionized the field of microbiology, and researchers are now able to characterize microbial diversity across a variety of previously unexplored environments. Introductory Biology students will work in pairs to: 1) generate a research question and formulate a hypothesis about the bacterial diversity in an environment of their choosing, 2) collect samples and relevant metadata from their environment, 3) extract DNA from their samples, 4) amplify the 16S rRNA gene from the bacterial DNA in their samples by PCR, 4) perform gel electrophoresis and prepare their samples for sequencing, 5) analyze the data using bioinformatics, 6) make conclusions about the bacterial diversity of their environment, and 7) give oral presentations summarizing their hypothesis and results. This module provides students with opportunities to apply the scientific method and use new technologies to explore their own environment.

Article Context: To make the submission process easier, you may want to examine the following form, which you will be asked to fill in during the submission process. Choose all applicable options that effectively **describe the conditions IN WHICH THE LESSON WAS TAUGHT**. Modifications to expand the usability of the Lesson will be addressed in the discussion.

****Not all categories will pertain to your article, in those cases, please select 'NA' when submitting on the website.**

Course

- Biochemistry
- Cell Biology
- Developmental Biology
- Genetics
- Microbiology
- Molecular Biology
- Introductory Biology
- Bioinformatics
- Evolution
- Ecology
- Anatomy-Physiology
- Neurobiology
- Plant Biology
- Science Process Skills

Course Level

- Introductory
- Upper Level
- Graduate
- High School
- Other

Class Type

- Lecture
- Lab
- Seminar
- Discussion Section
- On-line
- Other

Audience

- Life Sciences Major
- Non-Life Science Major
- Non-Traditional Student
- 2-year College
- 4-year College
- University
- Other

Class Size

- 1 – 50
- 51 – 100
- 101+

Assessment Type

- Assessment of individual student performance
- Assessment of student groups/teams
- Assignment
- Exam/quiz, in class
- Exam/quiz, take home
- Homework
- Answer clicker-type question(s)
- Answer essay question(s)
- Answer fill in the blank question(s)
- Answer multiple choice question(s)
- Answer short answer questions(s)
- Answer true/false question(s)
- Create a concept map
- Create a diagram, drawing, figure, etc.
- Create a website
- Create graph, table etc. to present data
- Design an experiment or research study
- Design/present a poster
- Give an oral presentation
- Informal in-class report
- Interpret data
- Order items (e.g. strip sequence)
- Participate in discussion
- Peer evaluation
- Post-test
- Pre-test
- Produce a video or video response
- Respond to metacognition/reflection prompt
- Self evaluation
- Solve problem(s)
- Written assignment: One minute paper
- Written assignment: Brochure
- Written assignment: Essay
- Written assignment: Figure and or figure legend
- Written assignment: Lab report
- Written assignment: Literature review
- Other

Lesson Length

- Portion of one class period
- One class period
- Multiple class periods
- One term (semester or quarter)
- One year
- Other

Key Scientific Process Skills

- Reading research papers
- Reviewing prior research
- Asking a question
- Formulating hypotheses
- Designing/conducting experiments
- Predicting outcomes
- Gathering data/making observations
- Analyzing data
- Interpreting results/data
- Displaying/modeling results/data
- Communicating results

Pedagogical Approaches

- Think-Pair-Share
- Brainstorming
- Case Study
- Clicker Question
- Collaborative Work
- One Minute Paper
- Reflective Writing
- Concept Maps
- Strip Sequence
- Computer Model
- Physical Model
- Interactive Lecture
- Pre/Post Questions
- Other

Bloom's Cognitive Level (based on learning objectives & assessments)

- Foundational: factual knowledge & comprehension
- Application & Analysis
- Synthesis/Evaluation/Creation
- Principles of how people learn
- Motivates student to learn material
- Focuses student on the material to be learned
- Develops supportive community of learners
- Leverages differences among learners
- Reveals prior knowledge
- Requires student to do the bulk of the work

Vision and Change Core Concepts

- Evolution
- Structure and Function
- Information flow, exchange and storage
- Pathways and transformations of energy and matter
- Systems

Vision and Change Core Competencies

- Ability to apply the process of science
- Ability to use quantitative reasoning
- Ability to use modeling and simulation
- Ability to tap into the interdisciplinary nature of science
- Ability to communicate and collaborate with other disciplines
- Ability to understand the relationship between science and society

Key Words: List 3 to 10 key words that are relevant for the Lesson (e.g. mitosis; meiosis; reproduction; egg; etc.)

1. Metagenomics
2. Bioinformatics
3. Microbial ecology
4. Microbiology
5. DNA extraction
6. PCR
7. Gel electrophoresis
8. Scientific method

Scientific Teaching Context

Learning Goal(s)

- Students will know how genomics, and specifically metagenomics, is broadly used to answer questions and test hypotheses in contemporary research.
- Students will apply the scientific method to ask and answer a research question rooted in metagenomics.
- Students will develop skills commonly used in metagenomic studies including molecular techniques and high throughput sequencing.

Learning Objective(s)

- Students can describe high throughput sequencing and how it's being used to answer research questions in variety of fields including, but not limited to, human health, ecology, biotechnology, etc.
- Student will be able to propose a research question with a testable hypothesis that can be answered using high throughput sequencing.
- Students will understand why the 16S ribosomal RNA gene is commonly used for microbial identification.
- Students will be able to define, explain, and perform the techniques used to generate libraries for high throughput sequencing, including DNA extraction, 16S rRNA gene PCR, and agarose gel electrophoresis
- Students will know how high throughput sequencing technologies work, specifically Illumina sequencing.
- Students will use appropriate informatics to analyze sequence data they generate.
- Students will be able to communicate and discuss the implications of their findings to a larger audience.

1 Introduction

2 Current trends in undergraduate science education include an increasing shift away from
3 “cookbook” lab experiments, which typically include pre-defined questions and scoring students on how well
4 they achieve the expected results, to open-ended authentic research questions where students propose the
5 research question, design the experiments themselves, analyze the data, and communicate their findings to
6 their peers and faculty. Engaging our students in scientific research from the very beginning of their
7 undergraduate academic career using cutting-edge technologies is a central goal of our proposal.

8 High throughput sequencing is relatively new and being rapidly incorporated in life science
9 research programs ranging from molecular biology to ecosystem ecology. Our understanding of biological
10 processes, particularly “black box” processes, is expanding at an unprecedented rate due to the tremendous
11 amount of data that can be produced using high throughput sequencing. For example, even though it is well
12 known that microbes play essential and critical roles in ecosystem functioning, researchers were previously
13 unable to determine the microbial diversity in an environment as most microbes could not be cultivated in
14 the laboratory. The development and adoption of culture-independent technologies, such as direct
15 sequencing of DNA by high throughput sequencing, has revolutionized the field of microbiology. As a result,
16 researchers are now able to characterize the microbial biodiversity in a habitat without cultivation, and
17 studies consistently find that microbial diversity across a variety of previously unexplored or underexplored
18 environments is much greater than earlier thought. Understanding the immense metabolic diversity present
19 in these environments can lead to novel gene discoveries, which will have far reaching effects in a variety of
20 fields.

21 A central feature of this module is that students will perform the majority of the steps associated
22 with their experiment themselves, in order to analyze the biodiversity of their microbial communities of
23 interest. First and second-year intended and declared STEM majors will perform this module as part of an
24 Introductory Biology lab curriculum. Therefore, no pre-requisite student knowledge is required. The module
25 will encompass five three-hour lab sessions, or approximately half of one semester of the lab course.

26 *Intended Audience*

27 Intended student population is first- and second-year intended and declared STEM majors.

28 *Required Learning Time*

29 Five three-hour lab sessions with additional outside class time for preparing for class and completing assignments.

30 *Pre-requisite student knowledge*

31 No pre-requisite student knowledge required.

32 Scientific Teaching Themes

33 Explain how the Lesson relates to the Scientific Teaching Themes of:

34 *Active Learning*

35 Students will propose research questions with hypotheses, collect samples and relevant
36 metadata, complete wet lab activities including DNA extraction, PCR, agarose gel electrophoresis,
37 perform bioinformatics data analyses and present their findings to instructors and peers during
38 class-wide presentations.

39 *Assessment*

40 Student learning and mastery will be assessed by using pre- and post-lab written assignments,
41 exams, and oral presentations. Individual scores on these assessments will be used as part of
42 determining each student's final lab grade. In addition, student learning for the entire
43 metagenomics module will be evaluated by a pre- and post-learning assessment that will be
44 completed by students anonymously.

45 *Inclusive Teaching*

46 Students will be selecting samples from their environment, where they live, work, play and study.
47 Students enrolled in these introductory biology courses often declare majors in Biochemistry,
48 Bioengineering, the Biological Basis of Behavior, or Biology, commonly with a concentration in
49 Computational Biology, Ecology, Molecular and Cell Biology, or Neurobiology. Given that the
50 development of high throughput sequencing is a direct result of interdisciplinary collaborations
51 between the fields of chemistry, biology, ecology, computational sciences, and engineering we feel
52 that this is the most appropriate audience of students to expose to this lab curriculum. The mixture
53 of reading, writing, and oral assignments along with lab activities is designed to include many
54 different learning styles in the classroom.

55 Lesson Plan

56 **Overview.** Below is an outline of our proposed 5-week laboratory module. The module has been
57 designed for a large multi-section introductory lab course however it would suitable to run this unit
58 with a single section that has small student enrollment. During this module students will work in
59 pairs to maximize student engagement in the laboratory activities, with the exception of the
60 bioinformatics tutorial in week 3, at which point students will work individually. Students will have
61 freedom to sample a public space of their choosing including, but not limited to, soil, water, or swab
62 samples. Students are restricted from sampling humans, human artifacts, private facilities or
63 companies, or dangerous environments. At the beginning of the unit, students will be trained on the

64 appropriate University of Pennsylvania's Environmental Health and Radiation safety protocols and
65 we will discuss the risks associated with sampling microorganisms in the wild.

66 **Week 1: Introduction, Generating a Research Question & Sample Collection.** During the first
67 week of this module, students will learn the conceptual background necessary to design a bacterial
68 biodiversity experiment that uses high throughput sequencing. Working in pairs and guided by the
69 lab coordinators and teaching assistants, students will decide on a research question and the
70 environment they want to sample, formulate a hypothesis about the bacterial biodiversity in their
71 environment, and design their experiment. Student pairs will then go out into the environment and
72 aseptically collect their samples and relevant metadata following the sample collection protocol
73 specified in the course lab manual. Student pairs will transport samples back to the lab and samples
74 will be stored at -20°C until week 2.

75 **Week 2: DNA Extraction & PCR.** In the second week, student pairs will perform a DNA extraction
76 on their samples using the MoBio PowerSoil DNA extraction kit. Student pairs will use the PCR to
77 amplify the v4 region of the 16S rRNA genes from the bacterial DNA present in their sample. Each
78 student pair will use their own uniquely barcoded reverse primer in their PCRs. The specific PCR
79 protocol will be specified in the course lab manual. Student pairs will perform the PCR in duplicate.

80 **Week 3: Agarose Gel Electrophoresis, Troubleshooting, and Bioinformatics Tutorial.** In the
81 third week, student pairs will pool the duplicate PCRs and check the quality of their PCR libraries
82 using gel electrophoresis to make sure they generated the desired product. If their reaction is
83 successful, they will prepare their samples to be sent for sequencing. If they are not successful they
84 will have the opportunity to troubleshoot their DNA extraction and/or PCR reaction and prepare the
85 samples for sequencing before the lab coordinators ship the samples to the outside facility for
86 sequencing. During this lab session, students will also be introduced to some of the bioinformatics
87 tools that they will use to analyze their results and complete a bioinformatics tutorial using a
88 provided dataset. This tutorial will be completed individually to maximize student engagement in
89 this activity.

90 **Weeks 4-7: Sample submission & Sequencing.** The students' library samples will be sent for high
91 throughput sequencing to an outside facility. Currently, the EMP and related staff at the Argonne
92 National Laboratory in Lemont, IL has agreed to process the students' samples and have guaranteed
93 a turn-around time of 2-3 weeks. We have also discussed the possibility of sending students' samples
94 to UPenn's Microbiome Core for sequencing when that facility is up and running. A critical feature of
95 the relationship between our courses and whichever sequencing facility we select will be the ability
96 of the sequencing facility to process the students' samples in a timely manner at a fair price so that
97 students will be able to process their sequencing data before the end of the semester. During these
98 weeks, students complete other lab activities associated with the courses.

99 **Week 8: Data Analysis.** Once the sequencing data is available, student pairs will download their
100 data and spend a lab session analyzing the data using advanced bioinformatics. This analysis will
101 provide students with the taxonomic classification of the bacteria present in their sample. Students
102 can then make conclusions about the bacterial biodiversity of their environment and determine if
103 their findings support their research hypothesis. They can compare their results to the established
104 literature to see how their findings relate to current microbial ecology research.

105 **Week 9: Communication of Research Findings.** Student pairs will give oral presentations that
106 summarize the background behind their group's original hypothesis, whether or not the data
107 supports their hypothesis, and other interesting findings (more in-depth description of one or more
108 taxa in their sample and their societal importance—economic value, interesting phenotypes,
109 pathogenesis) from their results.

110 Acknowledgments

111 Begin the Acknowledgements on a new page. The acknowledgements can include multiple paragraphs.

112 GCAT SEEK faculty and staff

113 Svetlana Kozik, Amanda Cottone, Lori Spindler, Staver Bezhani, Chris Jang & Graduate TAs

114 Introductory Biology Faculty (Brenda Casper, Fevzi Daldal, Greg Guild, Mark Goulian, Mike Lampson, Dustin
115 Lampson)

116 Aurora McCreary, Staff at the Earth Microbiome Project, including Sarah Owens and Jarrad Marcell

117 Adam Linder and Paul Doherty

118