JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION, December 2012, p. 161-169
Copyright © 2012 American Society for Microbiology DOI: http://dx.doi.org/10.1128/jmbe.v13i2.451

Using Bioinformatics to Develop and Test Hypotheses: *E. coli*-Specific Virulence Determinants †

Joanna R. Klein[1]* and Theresa Gulsvig[2]
[1]*Northwestern College, St. Paul, MN 55113,* [2]*Central High School, St. Paul, MN 55104*

Bioinformatics, the use of computer resources to understand biological information, is an important tool in research, and can be easily integrated into the curriculum of undergraduate courses. Such an example is provided in this series of four activities that introduces students to the field of bioinformatics as they design PCR based tests for pathogenic *E. coli* strains. A variety of computer tools are used including BLAST searches at NCBI, bacterial genome searches at the Integrated Microbial Genomes (IMG) database, protein analysis at Pfam and literature research at PubMed. In the process, students also learn about virulence factors, enzyme function and horizontal gene transfer. Some or all of the four activities can be incorporated into microbiology or general biology courses taken by students at a variety of levels, ranging from high school through college. The activities build on one another as they teach and reinforce knowledge and skills, promote critical thinking, and provide for student collaboration and presentation. The computer-based activities can be done either in class or outside of class, thus are appropriate for inclusion in online or blended learning formats. Assessment data showed that students learned general microbiology concepts related to pathogenesis and enzyme function, gained skills in using tools of bioinformatics and molecular biology, and successfully developed and tested a scientific hypothesis.

INTRODUCTION

Bioinformatics is used extensively by researchers and is an area that students need to become competent in, especially considering rapid advances in genome sequencing projects (3). Just as in any inquiry-based lab, bioinformatics is most meaningful when students learn the tools while using them to test hypotheses. With this goal in mind, a classroom activity was designed for students to learn how to use some specific bioinformatics tools both in developing a hypothesis and then in testing whether the hypothesis is correct. In the process, students are also exposed to topics including bacterial enzyme function, virulence factors, and horizontal gene transfer.

This activity takes a case study approach in which students are asked to design a PCR-based diagnostic test for a pathogenic strain of *Escherichia coli.* The initial scenario has students testing for *E. coli* O157:H7. This strain of *E. coli* is a type of enterohemorrhagic *E. coli* (EHEC) that causes foodborne illness (8). It is often associated with contaminated ground beef, leafy greens, water, or unpasteurized milk or juice. Infection leads to bloody diarrhea and in children can lead to serious complications such as hemolytic uremic syndrome (HUS). Typically, diagnostic tests for *E. coli* O157:H7 utilize nonfermenting growth on sorbitol MacConkey agar or enzyme immunoassays (5). Increasingly, molecular tests for rapid detection and diagnosis are being sought, and for this purpose detecting the Shiga toxin gene (*Stx*) by PCR is utilized (5). Shiga toxin is an AB type toxin unique to O157:H7 strains of *E. coli* and was acquired by infection with a prophage (11). Additional pathogenic strains of *E. coli*, including Enterotoxigenic *E. coli* (ETEC), Enteropathogenic *E. coli* (EPEC) and Urophathogenic *E. coli* (UPEC) are also encountered by students in the final steps of the case study.

Through the case study, students use bioinformatics to identify genes that are unique to specific pathogens and therefore could be used in a PCR-based diagnostic test. The activity is broken into four parts, some or all of which can be used at the discretion of the instructor. In Part 1, students learn about PCR and gel electrophoresis using online virtual labs and textbook reading. In Part 2, students are asked to identify a gene that is unique to *E. coli* O157:H7 and therefore could be used in a diagnostic test. They are provided with a set of four unknown gene sequences, one of which (*stx*) is specific to O157:H7. They determine the identity of the sequences by performing BLAST (1) searches at NCBI. They review the function of the gene products and develop a hypothesis about which one might be unique to O157:H7. Then they test their hypothesis by using the integrated microbial genomes (IMG) database to search specific bacterial genomes for each gene. In Part 3, students delve

*Corresponding author. Mailing address: Northwestern College, 3003 Snelling Ave. N., St. Paul, MN, 55113. Phone: 651-286-7468. Fax: 651-286-7532. E-mail: jrklein@nwc.edu.
†Supplemental materials available at http://jmbe.asm.org.

into learning more about Shiga toxin using IMG (10), Pfam (12), and BLAST programs. In Part 4, students are asked to design a PCR-based assay for a different pathogenic strain of *E. coli*. While Part 2 walked them through the steps of how to do this for *E. coli* O157:H7, and provided them with the answer since one of the unknown gene sequences was indeed an O157:H7 specific gene, now they must apply these skills and additional critical thinking to a new situation. This part utilizes literature research, collaborative group work, and oral presentations to solidify concepts of bioinformatics, molecular diagnostics, and virulence factors in students' minds.

Of practical significance is that this purely computer-based activity is a useful substitute for a lab on molecular methods for microbial identification when such a lab is not possible due to cost or time. Furthermore, it is an activity that could be incorporated into online learning.

Intended audience

The intended audience for this activity is microbiology and biology majors taking an introductory-level course in general microbiology or biology. It has also been used with high school students in a summer enrichment course on microbiology. Some or all of the activities described can be used, depending on the level and background of the students. Parts 1 and 2 are the core of the activity and are appropriate for all levels of students. Students can achieve significant learning gains after completing these first two parts. Parts 3 and 4 are most appropriate for students taking a microbiology course.

Learning time

The amount of learning time is dependent on which activities are used and to some extent on the background and level of the students. More time may be required to provide background information and discussion depending on prior student knowledge. Table 1 summarizes the ideal learning venues and approximate learning times.

Prerequisite student knowledge

Basic computer skills are required. Students should have an understanding of DNA and how it is replicated prior to doing the activity on PCR. They should also have a basic understanding of enzymes and metabolism and of the process of gene expression. Background knowledge about pathogenic and non-pathogenic strains of *E. coli* provides context for the case study and helps students generate a hypothesis about what genomes to test. For upper-level students in a microbiology course, this information may have been previously covered, but the requisite information can also be delivered in a brief mini-lecture format to introductory-level students. Previous exposure to laboratory tests for *E. coli* is helpful, but again, the requisite information can also be delivered in a brief mini-lecture to introductory level students.

Learning objectives

Parts 1 and 2 are the core of this activity. At completion of these two activities, students will be able to:

1. Explain the purpose, steps and materials of PCR and gel electrophoresis and interpret PCR amplified bands following gel electrophoresis.
2. Perform a BLAST search and interpret the results to determine the probable identity of unknown gene sequences.
3. Describe the function of several bacterial enzymes related to pathogenesis and general microbial physiology.
4. Develop a hypothesis about a gene that will differentiate *E. coli* O157:H7 from other bacterial strains.
5. Use the IMG database to search bacterial genomes for the presence of specific gene sequences.
6. Recognize the powers and limits of bioinformatics as a tool to study organisms.

Part 3 is an extension of the activity with these added learning objectives.

7. Use IMG and embedded resources including Pfam to learn about AB type toxins, including protein structure and function, gene structure and genomic position.
8. Perform a BLAST search and interpret the results to look for evidence of horizontal gene transfer.

Part 4 is an extension of the activity with these added learning objectives.

9. Describe virulence strategies and factors used by a specific strain of *E. coli.*
10. Apply bioinfomatic tools to generate and test hypotheses in new situations.
11. Design a PCR-based diagnostic test and predict expected results.
12. Effectively present scientific findings.

PROCEDURE

Materials

This is a computer-based activity and requires that students have access to a computer connected to the internet.

Student instructions

In this series of activities, students learn how to design a PCR-based diagnostic test for a pathogenic strain of *E. coli.*

TABLE 1.
Learning time and venue for activities.

Activity	Description	Learning Venue	Time
Part 1[a]	Online learning activity – PCR and gel electrophoresis	Outside of class	One half hour
Part 2	Hypothesis generation and testing	In class (ideal)[b]	One two-hour class period or two 60-minute class periods
Part 3	Online learning activity	Outside of class	< 1 hour
	Discussion	In class	10 minutes
Part 4	Initial group work on project	In class	30-60 minutes
	Group work	Outside of class	Several hours over the period of one week
	Student presentations	In class	10 minutes per group

[a] Students with previous experience with PCR and gel electrophoresis may not need to do this part although it does provide an introduction to the case-study scenario and would provide for a good review.
[b] Some work could be completed outside of class, as students use their own computers to access the databases, however explanation by the instructor is helpful initially and as student progress through the activity.

There are four parts that are provided as four student handouts. Some or all of the parts may be assigned, depending on time, student background, and the learning goals of the instructor.

Part 1 is an activity for students to learn about PCR and gel electrophoresis. Students are directed to go through virtual labs on PCR and gel electrophoresis at the Genetic Science Learning Center, University of Utah, http://learn.genetics.utah.edu. They also read a section in their textbook about the use of PCR in clinical diagnosis. Based on what they learn, they answer a set of questions. This activity, Student Handout–Part 1 and the answer key, is provided as Appendix 1 and Appendix 2.

Part 2 of the activity is best done during class when the instructor is available to provide direction. In this activity, students determine the names of unknown gene sequences by performing BLAST searches. They then develop a hypothesis about the appropriateness of using one of these gene sequences to test for *E. coli* O157:H7. They test the hypothesis by searching various bacterial genomes for these genes using the IMG database. As a final step, students are asked to predict the results they would observe if they performed such a PCR based test. This activity, Student Handout–Part 2 and the answer key, can be found as Appendix 3 and Appendix 4. The unknown gene sequences to provide to students can be found in Appendix 5.

Part 3 of the activity is appropriate to do outside of class. Students use the IMG database and imbedded resources to learn more about the Shiga Toxin virulence factor found in *E. coli* O157:H7. They learn about protein structure and function of AB type toxins using the Pfam database. They also perform another BLAST search to consider where the Shiga-toxin gene originated and how *E. coli* O157:H7 acquired it. Students see evidence of horizontal gene transfer through transduction. This activity, Student Handout–Part 3 and the answer key, can be found as Appendix 7 and Appendix 8.

Part 4 of the activity asks the students to apply the skills and information they have learned in the previous activities to a new situation. In this activity, students are assigned a different pathogenic strain of *E. coli* (either ETEC, EPEC, or UPEC) and asked to develop a PCR-based diagnostic test for it. They begin by researching the bacterium to learn about its virulence factors and the genes that encode them. Based on this information, they make predictions about genes that may be unique to their assigned strain. They form a hypothesis to test their prediction and use bioinformatic tools to carry out the tests. They then give a presentation summarizing their findings to the class. This part incorporates critical thinking, collaborative work, and communication skills. This activity, Student Handout–Part 4, can be found as Appendix 9.

Faculty instructions

This series of activities is divided into four parts that are provided as four student handouts. Considering student level and learning goals, some or all of the activities can be used. While a basic description of each activity was provided in the student instructions, instructor guidelines for utilizing each part of the activity follow.

Part 1. Faculty should begin by assigning Student Handout 1.This is a pre-class activity that takes students about half an hour and provides basic information about PCR and gel electrophoresis in a virtual-lab format. Students may also be directed to read about PCR and the use of PCR in clinical diagnosis in a textbook to supplement the activity. If students are already familiar with these techniques, this activity can be omitted although it does provide an introduction to the case-study scenario and provides for a good review.

Part 2. This is the one essential activity in this series, where students learn how to use BLAST and IMG to develop and test a hypothesis. It can be done in one two-hour class or broken up into two shorter class sessions. Students begin by determining the identity of four unknown gene sequences. Prior to class, the instructor should

make available an electronic file containing four unknown sequences (Appendix 5). This can be done by emailing a file to the students, or posting to a class-management site. The four unknown genes (described below) were chosen because each is expected to be found differentially in *E. coli* strains. In addition, depending on timing and course content, they may represent genes that students have been previously exposed to, so prior material is recalled and reinforced. An instructor version of the unknown sequences lists the source of each sequence (Appendix 6).

Unknown Sequence 1 is glyceraldehyde-3-phosphate dehydrogenase (GAPDH). This protein catalyzes the sixth step of glycolysis and is a housekeeping gene present in most organisms. Unknown Sequence 2 is cytochrome c oxidase. This protein is part of the electron transport chain of certain organisms. It is absent from *E. coli* and students may have prior knowledge of this enzyme from doing the oxidase test in a previous lab. Unknown sequence 3 is tryptophanase. This enzyme catalyzes the formation of indole and pyruvate from tryptophan and is present in *E. coli.* Students may have prior knowledge of this enzyme from doing the indole test in a previous lab. Unknown sequence 4 is Shiga toxin subunit A. This is an AB type toxin that is present in the pathogenic strain *E. coli* O157:H7.

During class, students determine the name of the product encoded by each unknown sequence by doing a blastx search at NCBI. Although a written description of how to do this is provided in the student handout, it is often helpful to walk the class through one BLAST search and explain the three sections of the results page (Graphic summary, Descriptions and Alignments) and how to interpret this information. A tutorial on BLAST is provided by NCBI at http://www.ncbi.nlm.nih.gov/books/NBK21097/ (7) that may be helpful for instructors or students alike. Sample results from a BLAST search are found in Appendix 10.

Once students identify the name of each unknown gene sequence, they review or research the function of each protein. Based on what they learn, they make predictions about what bacteria would contain these genes. They are provided four categories for their predictions: i) found in all species of bacteria, ii) found in all *E. coli* strains, iii) found in just *E. coli* O157:H7, and iv) absent from *E. coli.* Then they write down a hypothesis about what gene is specific to *E. coli* O157:H7 and therefore would be an appropriate gene to test for the presence of using PCR.

The next step is to test the hypothesis by using the IMG database housed at the US Department of Energy Joint Genome Institute. This database contains the genomes of all sequenced microbial species and is an important tool for the microbiology student to be able to use. The students spend some time exploring this database and answer a few general questions. This is a point at which the instructor may like to walk the class through some features of IMG. Specific directions about how to use the site to answer these general questions are found in the student handout answer key (Appendix 4).

Students may need help in identifying bacterial genomes to search. We try to steer them towards selecting four or five genomes that fit in with the above predictions. We recommend students test a *Pseudomonas* strain, since in my course they have previously done biochemical tests on *Pseudomonas* including the oxidase and indole tests. They should recall that this bacterium tests positive for oxidase (and thus should have the cytochrome c oxidase gene) and negative for indole (and thus lacks the tryptophanase gene). They should also choose a nonpathogenic strain of *E. coli* and an *E. coli* O157:H7 strain.

When searching for the gene name by entering the keyword in the Find Genes tab, students may have to use some trial and error to determine the best keyword. We have found it useful to query with the gene names "glyceraldehyde," "cytochrome c oxidase," "tryptophanase," and "Shiga toxin." Sample results for a gene search at IMG can be found in Appendix 10.

Part 3. This is an extension activity that has students use online and textbook resources to learn more about Shiga toxin. If not previously covered, it is useful as an entry point to the topic of virulence factors and classroom discussion of the activity could lead to a more formal lecture on bacterial toxins and virulence factors. After being exposed to Shiga toxin in this activity, student interest in learning about AB toxins and other virulence factors is heightened and they are more invested in textbook reading assignments and follow-up mini-lectures on virulence factors.

In the same way, the activity ties in to topics related to horizontal gene transfer. The *stx* gene in O157:H7 was acquired from infection with a prophage, thus the top hits when a BLAST search is done that excludes *E. coli* are all phages. One notable exception is a homology in *Acinetobacter haemolyticus*, an environmental bacterium and an opportunistic pathogen resistant to multiple antibiotics. The finding of *stx* in an *A. haemolyticus* isolate from an infant with bloody diarrhea is an interesting example of the evolution of a new pathogenic strain and can provide for additional classroom discussion (6).

Part 4. In this exercise, students are assigned the task of developing a PCR-based assay for a different pathogenic strain of *E. coli* that is assigned by the instructor. Students work in groups to research the pathogen and its virulence factors. Then they make a hypothesis about a gene that will be unique to their strain and test their hypothesis *in silico* using the bioinformatic tools they have used in parts two and three. They diagram expected results and present their findings to the class.

There is a variety of pathogenic *E. coli* strains, but we suggest limiting the assignment to enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), or uropathogenic *E. coli* (UPEC). This is because students will ultimately be searching bacterial genomes at IMG and must know which genomes to search (e.g., which of the 106 *E. coli* genomes are actually UPEC). The genome name does not always include

a description of what type of *E. coli* it is, so determining this can be daunting to the novice. However, there are examples of genomes categorized as EPEC, ETEC and UPEC at IMG, making selection of genomes simpler.

Students begin by doing literature research to learn about their assigned strain and the virulence factors it contains. From there, they make a hypothesis about a gene or genes that may be unique to their strain. In some cases, a good candidate gene may be evident from what they read if its genome distribution has already been studied. Students then use IMG to test to see whether their gene is present in their strain but absent in other strains, as they did with Shiga toxin in Part 2.

Be aware that this step may not always be as straightforward for other genes as it was presented in Part 2. In Part 2, a keyword search was utilized at IMG, and sometimes it is difficult to find the right keyword to search for, or the same keyword might be used to describe different genes. In this case, it may be necessary to determine the gene sequence of the candidate gene and do a BLAST search of selected genomes to test the hypothesis. This adds complexity to the assignment, but uses skills and resources students have had previous exposure to, so is still doable. One other observation is that some students may do literature searches for molecular tests that have already been developed and published for their particular strain and use that to guide them in this assignment.

Among the three strains assigned, students had the easiest time with ETEC since it contains two well defined and unique virulence factors, heat-labile enterotoxin and heat-stable enterotoxin (13). EPEC and UPEC contain multiple virulence factors, many of which are found in multiple types of *E. coli* (for reviews see Reference 4 and 9). However, students who are assigned these strains have exposure to a broad variety of virulence strategies used by pathogens as they wade through candidate genes.

Suggestions for determining student learning

Several tools for measuring student learning are available. First, the student handouts can be collected to assess all learning objectives. Answer keys for parts 1 to 3 are provided as appendices. Second, the presentations students give in Part 4 can be assessed. A sample scoring rubric is provided as Appendix 11. Third, a pre–posttest on 12 key terms and concepts from the activity is provided (Appendix 12 and Appendix 13). This primarily measures objectives 1, 3, and 6. Fourth, a post-activity assessment measures whether students met learning objectives 1 to 6 (Appendix 14). The post-activity assessment is a useful tool for assigning grades, and an answer key and scoring rubric is provided for this purpose (Appendix 15 and Appendix 16).

Sample data

Sample results for a BLAST search and a gene search at IMG can be found in Appendix 10. Sample student answers for the 12-question pre–post tests are provided in Appendix 13 and a sample student presentation is provided as Appendix 17.

Safety issues

There are no safety concerns with this computer-based activity. However, it might be useful to discuss with students safety concerns that would be encountered if they were actually doing tests for pathogenic *E. coli* strains proposed in the case study.

DISCUSSION

Field testing and evidence of student learning

This activity was developed and used at Northwestern College over a period of three years in multiple settings. Parts 1 and 2 have been used most extensively, as described below. Based on faculty feedback, parts 3 and 4 were developed more recently and have been tested once.

Parts 1 and 2 have been used together in the following contexts:

i) twice in a microbiology course for biology majors (ranging from freshman through senior) with enrollment of 12–24 students;
ii) once in an organismal biology course for biology majors in their first year (Principles of Biology 2), student enrollment of 27;
iii) once in a summer enrichment course on microbiology for 14 high school students in their junior and senior years and 4 college biology majors who served as mentors for the high school students; and
iv) piloted by instructors of microbiology courses at two other institutions.

All four parts of the activity were used recently in a microbiology course for biology majors (ranging from freshman through senior) with an enrollment of 12 students. Assessment results, student comments and faculty feedback from early trials contributed to the refinement of the activities as presented herein.

Assessment data presented in this paper is from the most recent course offerings of Microbiology and Principles of Biology 2, spring semester 2012. Microbiology is an upper-level course for biology majors, while Principles of Biology 2 is an introductory-level course, so the usefulness of this activity with students of various levels is compared.

Principles of Biology 2 is the second in a series of three introductory courses biology majors are required to take. It introduces the topics of comparative animal anatomy and physiology and organismal diversity of prokaryotes, protists, invertebrates, and vertebrates. Twenty-seven students were enrolled in the course, primarily in the second semester of their first year of college. They had taken an introductory cell and molecular biology course and were concurrently

enrolled in an introductory ecology and evolution course. In the introductory cell and molecular biology course, students had studied basic prokaryotic cell structure, learned about PCR, and carried out gel electrophoresis, but otherwise had little prior knowledge of microbiology topics. Parts 1 and 2 of the activity were done during a two-week unit on bacteriology at the end of the semester. Part 1 was an out-of-class assignment and Part 2 was done during two consecutive 65-minute class sessions. Mini-lectures and other in-class activities introducing and elaborating on topics related to this activity were also done during the two-week period.

The impact of this activity on achieving learning objectives 1–6 was assessed in several ways. First, all student handouts were collected and read by the instructor for formative assessment, monitoring student progress towards achieving the objectives. Additional explanations about specific questions were made in class as necessary. Second, students were given a pre–post test in which they were asked to define 12 key terms and concepts covered in the activity (see Appendix 12 and Appendix 13). Some of the terms tested their understanding of the biology of the situation (pathogenesis and specific protein function) while others tested their understanding of PCR and the bioinformatics tools employed. This same test was given prior to the start of the unit on microbiology and again at the end of the unit. Each item was scored and received two points (mastery of term), one point (partial mastery of term) or zero points (no attempt or incorrect answer). Thus a score of 24 indicates student mastery of each term, while a score of 12 indicates partial mastery. Results are provided in Table 2. Students began the activity with very little prior knowledge, the pretest average score for the class being only 2.6 points. After completing the activity, the average score increased to 12.2. These final scores may seem low; however, they do show an appropriate gain (partial mastery) for this level of student. Furthermore, this summative assessment was done for the purposes of this research, but did not factor into student grades, so it is possible that students did not use their best effort, reducing overall scores.

Third, students were given a 12-question post-activity assessment in which they were asked to answer a variety of questions by applying skills and knowledge we hoped they gained in the activity (provided as Appendix 14 and Appendix 15). A rubric was developed to rate student responses at four levels: mastery, proficient, basic, and developing (2) (Appendix 16). The results are provided in Table 3. We found that the students scored below proficient (score of 3) on a few questions. One of these was question 1, which tested objective 6 and asked students to decide if the following statement was true or false and explain their rationale. "Q: If a bacterium cannot be easily grown in the lab, very little information can be learned about how it functions." Most students did answer false, but failed to explain that bioinformatics is a useful tool in such a situation. Students fared better on question 4, which also tested objective 6. When asked how to compare two species of bacteria, students were able to explain that computers or bioinformatic tools were possibilities. Students also had difficulty with question 2, which related to objectives 3 and 6. It asked them to describe two ways to determine if a bacterium has the enzyme tryptophanase. Their answers tended to be incomplete, probably due to their lack of previous exposure to biochemical testing procedures for bacteria. Question 3, which tested objective 2, asked students to provide a written description of the BLAST search tool, which was

TABLE 3.
Post-activity assessment results.

Question	Objective Measured	Biology 2 Average Score[a,b]	Microbiology Average Score[a,c]
1	6	2.2	3.0
2	3, 6	2.8	3.5
3	2	2.4	2.3
4	5,6	3.2	3.8
5	3, 4	3.6	4.0
6	1	3.9	3.9
7	1	3.0	3.8
8	1	3.0	4.0
9	1	3.4	3.3
10	2	3.7	4.0
11	2	3.1	4.0
12	2	3.4	4.0

[a] The average score out of 4 points possible.
[b] 23 Freshman biology majors taking Principles of Biology 2.
[c] 12 Freshman-Senior biology majors taking a Microbiology course.

TABLE 2.
Pre–post test results.

Course	Number of Students	Pretest[a]	Posttest[a]	Gain
Principles of Biology 2	22	2.6	12.2	9.6
Microbiology	12	10.4	17.2	6.8

[a]The average score out of 24 points.

difficult for them. However, students were able to interpret BLAST results in question 10. To improve performance on questions for which they were less than proficient, we suggest the addition of instructor-led discussions. These would allow students to practice articulating the concepts and techniques they have learned and to connect the activity to the overall picture of bioinformatics.

Lastly, students were asked to self-evaluate their learning related to objectives 1–5 (Table 4). Students had the least confidence in objective 3 regarding their ability to describe the function of several bacterial enzymes related to pathogenesis and general microbial physiology. This is not too surprising since a relatively short amount of time was devoted to these topics. In the self-evaluation, students were also asked to rate the following statement: "I found the PCR and bioinformatics activity to provide the appropriate level of challenge." Fifty-eight percent of students answered strongly agree and 42% answered agree. Student comments related to this question included: "It was challenging for me, but after putting in some extra effort, I got it," "It was very interesting and was slightly challenging, more so because I had never completed anything similar to this activity," and "It was not too difficult, but it definitely stretched my abilities."

Overall comments about the activity were positive, for example, "I enjoyed this activity since it helped me understand PCR and bioinformatics more," "It was very interesting and was also exciting to learn about the many resources we are exposed to," and "This activity was interesting and something I've never done before. The directions were clear and helpful. I can't believe it all actually worked out — it was cool to see."

All four parts of the activity were used in the upper-level microbiology course. This course enrolled 12 students ranging from well-prepared freshman to senior. The activity was done towards the end of the semester before we began discussion of human–microbe interactions. Students had experience performing and interpreting a variety of biochemical identification tests in the course, but any prior experience with PCR or gel electrophoresis came from other courses. A similar set of assessments to those described above were given to the microbiology students. It is clear from the 12 question pre–post test (Table 2), that students began the activity with more background understanding than students in the introductory course, the average score being 10.2 points. However, there were some terms they had not previously encountered. After completing the activity, their performance on these questions improved and the average score was 17.2 points. This score may be lower than one might expect for advanced students, but, as in the introductory course, since these scores did not factor into student grades, the students may not have performed at their best level.

Microbiology students also performed better on the post-activity assessment (Table 3). This is likely due to two factors. First, these students had more general microbiology experience. Second, these students completed all four parts of the activity. Parts 3 and 4 likely contributed significantly to their overall understanding of the concepts. They did struggle with providing a written description of the BLAST search tool (Question 3). The question was intentionally vague to see if they would come up with key terms such as e-value, score, and query coverage. Rewording of the

TABLE 4.
Student evaluation of PCR and bioinformatics activity.

After completing the PCR and Bioinformatics activity, I am able to:	**Strongly Agree (5)**	**Agree (4)**	**Neutral (3)**	**Disagree (2)**	**Strongly Disagree (1)**	**Average**
Explain the purpose, steps and materials of PCR and gel electrophoresis and interpret PCR amplified bands following gel electrophoresis (Objective 1)	50%	46%	4%			4.5
Perform a BLAST search and interpret the results to determine the probable identity of unknown gene sequences (Objective 2)	38%	58%	4%			4.3
Describe the function of several bacterial enzymes related to pathogenesis and general microbial physiology. (Objective 3)	17%	62%	17%	4%		3.9
Develop a hypothesis about a gene that will differentiate *E. coli* O157:H7 from other bacterial strains. (Objective 4)	42%	50%	8%			4.3
Use the IMG database to search bacterial genomes for the presence of specific gene. (Objective 5)	50%	33%	13%	4%		4.3

Note: 24 students enrolled in Principles of Biology 2 took the survey and the percentage of student responses to each choice is provided.

question to provide more direction for what answer is expected would likely lead to improved scores. Alternatively, additional instructor-led discussion embedded within the activity would help the students move towards articulating their understanding more completely.

Objectives 7 and 8 were assessed by collecting student handouts for Part 3. All students met the objectives by being able to correctly answer questions related to AB toxins and used BLAST results to conclude that the toxin was likely acquired from a phage (data not shown).

Part 4 of the activity was assessed by evaluating oral presentations. Relatively little direction was given in this first iteration of the activity, yet each of the four groups successfully identified a unique virulence determinant, used bioinformatic tools to test its genomic distribution, and outlined a feasible PCR diagnostic test (data not shown). While the students successfully performed the task, the quality of the presentations was not exceptional. A sample student presentation is provided in Appendix 17. Based on the sample data, handouts were refined to provide improved direction and expectations.

Student comments about the four parts of the activity were collected and reveal that each part had value.

Part 1
- "Necessary in order to establish a background for subsequent parts."
- "A good introduction."
- "Good interactive way to learn general information."

Part 2
- "Really helped me become more familiar with NCBI and IMG"
- "It was challenging to use IMG, but very exciting."

Part 3
- "This helped because we had to think for ourselves and apply information."
- "This part introduced me to bacterial toxins and got me asking questions about how a toxin might undergo horizontal gene transfer."

Part 4
- "Most helpful! Hard, but necessary for understanding."
- "Challenging, thought provoking and interesting."
- "This activity really solidified all we learned."
- "I really enjoyed this learning activity. It was challenging and made me think very hard. It also helped us work together in order to complete part 4. So we helped each other learn."

It is noteworthy that while Part 4 was challenging for students, they report that it was a critical activity and contributed greatly to their learning. We highly recommend that it be included in upper-level courses to solidify all learning objectives.

Possible modifications

While the majority of this activity was done during a face-to-face session, it is amenable to use in an online-only setting. Extension of the activity could involve discussion of actual techniques used in the public health or clinical lab to identify *E. coli* O157:H7. More advanced students could design primers to amplify the desired gene. A wet-lab component could also be done in which students perform PCR and gel electrophoresis. If non-pathogenic strains that contain the virulence gene are available, student- or instructor-designed primers could be used in a mock case study scenario where the diagnostic PCR based assay is performed. Alternatively, completely unrelated strains and primers could be used in such an instructor-designed mock case study. We would expect that the addition of a wet-lab component would lead to an increase in student performance on assessment questions related to PCR and gel electrophoresis.

SUPPLEMENTAL MATERIALS

Appendix 1: Student Handout – Part 1
Appendix 2: Student Handout – Part 1 — Answer Key
Appendix 3: Student Handout – Part 2
Appendix 4: Student Handout – Part 2 — Answer Key
Appendix 5: Unknown Sequences – Student Version
Appendix 6: Unknown Sequences – Instructor Version
Appendix 7: Student Handout – Part 3
Appendix 8: Student Handout – Part 3 — Answer Key
Appendix 9: Student Handout – Part 4
Appendix 10: Sample BLAST and IMG results
Appendix 11: Presentation grading criteria
Appendix 12: Pre–post assessment
Appendix 13: Pre–post assessment — Answer Key
Appendix 14: Post-activity assessment
Appendix 15: Post-activity assessment — Answer Key
Appendix 16: Post-activity assessment — Scoring Rubric
Appendix 17: Sample student presentation

ACKNOWLEDGMENTS

We would like to thank Shellie Kieke, Concordia University–St. Paul, and Ruth Gyure, Western Connecticut State University, for piloting this lab in their microbiology courses and providing feedback and assessment data. The authors declare that there are no conflicts of interest.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman**. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res.**25**:3389–3402.

2. **Andrade, H. G.** 1999. Understanding rubrics. *In* ALPS: The thinking classroom. http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm. Accessed 2011 July 27.
3. **Ditty, J. L., C. A. Kvaal, B. Goodner, S. K. Freyermuth, C. Bailey, et al.** 2010. Incorporating genomics and bioinformatics across the life sciences curriculum. PLoS Biol. **8**(8): e1000448.
4. **Emody, L., M. Kerényi, and G. Nagy.** 2003. Virulence factors of uropathogenic *Escherichia coli*. Int. J. Antimicrob. Agents **22**(2): 29–33.
5. **Grys, T. E., L. M. Sloan, J. E. Rosenblatt and R. Patel.** 2009. Rapid and sensitive detection of Shiga toxin-producing *Escherichia coli* from nonenriched stool specimens by real-time PCR in comparison to enzyme immunoassay and culture. J. Clin. Microbiol. **47**: 2008–2012.
6. **Grotiuz, G., A. Sirok, P. Gadea, G. Varela, and F. Schelotto.** 2006. Shiga toxin 2-producing *Acinetobacter haemolyticus* associated with a case of bloody diarrhea. J. Clin. Microbiol. **44**(10): 3838–3841.
7. **Madden, T**. The BLAST sequence analysis tool. 2002 Oct 9 [Updated 2003 Aug 13]. *In* J. McEntyre and J. Ostell (ed.), The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US). http://www.ncbi.nlm.nih.gov/books/NBK21097/.
8. **Madigan, M. T., J. M. Martinko, P. V. Dunlap and D. P. Clark**. 2009. Brock Biology of Microorganisms, 12th ed. Benjamin Cummings, San Francisco.
9. **Mellies, J. L., A. M. S. Barron, and A. M. Carmona.** 2007. Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. Infect. Immun. **75**:4199–4210.
10. **Markowitz, V. M, A. C. I-Min, K. Palaniappan, K. Chu, E. Szeto,Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides.** 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucl. Acids Res. **40** (D1): D115–D122.
11. **O'Brien, A. D., J. W. Newland, S. F. Miller, R. K. Holms, and H. W. Smith**. 1984. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. Science **226**:694–696.
12. **Punta, M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn** (2012) The Pfam protein families database. Nucl. Acids Res. **40**:D290–D301.
13. **Qadri, F, A-M. Svennerholm, A. S. G. Faruque, and R.B. Sack.** 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin. Microbiol. Rev. **18**(3): 465–483.