April 2012

Using Comparative Genomics for Inquiry-Based Learning to Dissect Virulence of Escherichia coli O157:H7 and Yersinia pestis

David J. Baumler
University of Wisconsin-Madison

Lois M. Banta
Williams College

Kai F. Hung
Eastern Illinois University, khung@eiu.edu

Jodi A. Schwarz
Vassar College

Eric L. Cabot
University of Wisconsin-Madison

See next page for additional authors

CBE—Life Sciences Education
Vol. 11, 81–93, Spring 2012

Article

Using Comparative Genomics for Inquiry-Based Learning to Dissect Virulence of *Escherichia coli* O157:H7 and *Yersinia pestis*

David J. Baumler,[*†] Lois M. Banta,[‡] Kai F. Hung,[§] Jodi A. Schwarz,[||] Eric L. Cabot,[†] Jeremy D. Glasner,[†] and Nicole T. Perna[†¶]

*BACTER Institute and [†]Genome Center of Wisconsin, University of Wisconsin, Madison, WI 53706; [‡]Department of Biology, Williams College, Williamstown, MA 01267; [§]Department of Biology, Eastern Illinois University, Charleston, IL 61920; [||]Department of Biology, Vassar College, Poughkeepsie, NY 12604; [¶]Department of Genetics, University of Wisconsin, Madison, WI 53706

Submitted April 12, 2010; Revised October 20, 2011; Accepted October 26, 2011
Monitoring Editor: Erin L. Dolan

Genomics and bioinformatics are topics of increasing interest in undergraduate biological science curricula. Many existing exercises focus on gene annotation and analysis of a single genome. In this paper, we present two educational modules designed to enable students to learn and apply fundamental concepts in comparative genomics using examples related to bacterial pathogenesis. Students first examine alignments of genomes of *Escherichia coli* O157:H7 strains isolated from three food-poisoning outbreaks using the multiple-genome alignment tool Mauve. Students investigate conservation of virulence factors using the Mauve viewer and by browsing annotations available at the A Systematic Annotation Package for Community Analysis of Genomes database. In the second module, students use an alignment of five *Yersinia pestis* genomes to analyze single-nucleotide polymorphisms of three genes to classify strains into biovar groups. Students are then given sequences of bacterial DNA amplified from the teeth of corpses from the first and second pandemics of the bubonic plague and asked to classify these new samples. Learning-assessment results reveal student improvement in self-efficacy and content knowledge, as well as students' ability to use BLAST to identify genomic islands and conduct analyses of virulence factors from *E. coli* O157:H7 or *Y. pestis*. Each of these educational modules offers educators new ready-to-implement resources for integrating comparative genomic topics into their curricula.

LEARNING OBJECTIVES

At the completion of these activities, students will:

1. have improved their ability to use BLAST;
2. be able to identify genomic islands from whole-genome alignments;
3. know one way to explore existing annotation for genes in a genomic island and determine whether any are involved in virulence;
4. be able to conduct analyses addressing conservation of virulence factors in *Escherichia coli* O157:H7 or *Yersinia pestis* strains; and
5. be able to apply these newly acquired skills to design a bioinformatic investigation of an *E. coli* outbreak.

DOI: 10.1187/cbe.10-04-0057
Address correspondence to: David J. Baumler (dbaumler@wisc.edu).

INTRODUCTION

The advent of genome sequencing and increased use of computational biology for analysis have dramatically changed the landscape for undergraduate student learning in the areas of genetics, molecular biology, and even ecology. A variety of curricular approaches attempt to help

students understand the impact that genomics will have in their lives and careers (Campbell, 2002; Dexter Dyer and LeBlanc, 2002; Honts, 2003; Petrill and Justice, 2007). A number of efforts are currently underway nationwide to develop new teaching approaches and resources for undergraduate curricula in bioinformatics and genomics (Forst and Goodner, 2006; Goodner and Wheeler, 2006; Baumler *et al*., 2008; Lopatto *et al*., 2008; Lopatto and Elgin, 2010; Shaffer *et al*., 2010; the Joint Genome Institute's "adopt a microbial genome" [www.jgi.doe.gov/education/genomeannotation.html]; the HHMI–Scientific Education Alliance's phage annotation initiative [www.hhmi.org/grants/sea]; and the Teaching Big Science at Small Colleges: A Genomics Collaboration initiative [http://serc.carleton.edu/genomics/index.html]). These pioneering educational initiatives reaffirm and build on the premise that learning objectives are met and exceeded when students find the topics engaging, exciting, and worthwhile (National Research Council, 2005). Most of these teaching resources address student learning about genomics and bioinformatics through active or inquiry-based learning (Gallagher *et al*., 1995; Checkley, 1997). Active learning of these topics requires the use of computers to develop and reinforce skills in bioinformatic gene analysis. Microbial genomes are especially suitable for teaching these subject areas, since they are relatively small, have modest computational requirements, and present a way to initiate student learning by investigation and problem solving of real-world, ill-defined problems (Wiggins and McTighe, 2005; Nehm, 2010). There are a variety of existing educational resources that focus on a single genome (e.g., Lopatto *et al*., 2008), and others designed to teach about genomic technologies (e.g., Shuster, 2011). However, despite the recent advances in DNA-sequencing technology that have resulted in an abundance of available genome sequences, relatively few curricular modules address genome-level questions. As of 2011, more than 1000 microbial genomes have been sequenced, representing a vast untapped resource for educators interested in using inquiry-based exercises that can compare multiple genomes from different strains of related microbes.

Comparative genomics is the study of the relationships of genome content, structure, and function across multiple organisms. This field is growing rapidly alongside advances in sequencing, particularly for studies of very closely related species and strains of bacteria. Comparative genomic techniques are used to discover genetic variation at scales ranging from large-scale chromosomal rearrangements to single-nucleotide polymorphisms or SNPs. The presence or absence of protein-coding genes (open reading frames or ORFs) often leads to new hypotheses regarding the distinguishing traits of each microorganism. While a number of comparative genomic tools exist, many are limited to the analysis of two genomes. One tool that permits the comparison of more than two genomes is the multiple-genome alignment tool Mauve (Darling *et al*., 2004). Mauve is particularly appealing as an educational tool, because it includes a user-friendly visualization of the results, and runs on widely available, modestly priced computer hardware. Mauve has been used extensively for comparative genomic analysis of numerous types of pathogenic microorganisms, such as *Pectobacterium* spp. (Glasner *et al*., 2008), *Salmonella* spp. (Vernikos *et al*., 2007), *E. coli* (Mau *et al*., 2006), and *Y. pestis* (Darling *et al*., 2008). These studies illustrate the applicability of comparative genomics for deciphering and identifying unique genomic regions that may play a role in the strain-to-strain variation of pathogenic microorganisms.

Key signatures of pathogenesis-promoting regions are genes encoding virulence factors or proteins that play a role in the ability of the pathogen to cause disease. Some types of microbial virulence factors are: adhesins (which help the microbe attach to host tissue), toxins (secreted proteins toxic to host cells or organs), secretion systems (which inject microbial proteins into the host cell or environment), and defenses against host barriers (which protect the microbe from the host's immune system or, e.g., allow passage through the acidic fluid in human stomachs to reach the intestine). For pathogenic members of the bacterial family *Enterobacteriaceae*, which includes the human pathogens *E. coli* O157:H7, *Salmonella*, and *Y. pestis*, a concerted effort to identify virulence factors based on experimental evidence and/or bioinformatic analysis has been conducted, and the resulting annotations that support the designation of genes as "putative or known virulence factors" are available in the A Systematic Annotation Package for Community Analysis of Genomes (ASAP) database (Glasner *et al*., 2006; http://asap.ahabs.wisc.edu/asap/home.php).

We have developed two curricular modules to support student learning of bioinformatic skills via investigation of engaging questions in comparative microbial genomics. The extensive existing and recently updated ASAP annotation allows us to focus students' attention on downstream inferences, while the real-world clinical relevance appeals to the many premedical students in our courses. In the first module, students focus on three genomes of the human pathogen *E. coli* O157:H7 from separate foodborne outbreaks; a second module centers on student analysis of five genomes of the human pathogen *Y. pestis*. The skills addressed in both modules are the use of the basic local assignment search tool (BLAST; Altschul *et al*., 1990) and identification and analysis of genomic islands. Through these exercises, students are exposed to core concepts in microbial genomics, including functions and conservation of virulence factors; horizontal gene transfer; the evolution/structure/function of genomic islands; and SNPs (in genes unrelated to pathogenesis) as possible determinants of metabolic capabilities. In crafting the assignments for these modules, we focused on questions that allowed students to apply both their newly developed skills and these core concepts. For example, students use the literature resources linked to the ASAP database to investigate the function and potential relevance to pathogenicity of a gene on a genomic island; this task requires understanding of the concept of virulence factors and some knowledge of possible virulence factor functions, as well as the ability to use the Mauve alignment tools to identify a gene of interest. We also sought to foster higher-level thinking skills, by posing questions that require problem solving, analysis, and synthesis (Allen and Tanner, 2002).

The three *E. coli* O157:H7 strains in the first module were isolated from sickened individuals or contaminated food during outbreaks associated with ground beef in Michigan in 1982 (Perna *et al*., 2001); radish sprouts in Sakai City, Japan, in 1996 (Hayashi *et al*., 2001); and fresh bagged spinach in 17 states in the United States in 2006 (Manning *et al*., 2008). Strains of *E. coli* O157:H7 differ in the severity of disease

they cause in humans and can lead to bloody diarrhea, renal failure, hemolytic uremic syndrome (HUS), or death. The three strains of *E. coli* O157:H7 genomes used for this exercise, strain EDL933 (ground beef), Sakai (radish sprouts), or EC4042 (bagged spinach), each have different epidemiological statistics for those sickened from the outbreaks and differed in their hospitalization-to-death ratios of 23:0 (EDL933), 8938:3 (Sakai), and 205:3 (EC4042) (Hayashi *et al*., 2001; Rangel *et al*., 2005; Manning *et al*., 2008). On the basis of the epidemiological statistics, students learn that each of these strains differs in the severity of the disease that they cause in humans (Manning *et al*., 2008). In this module, students use Mauve to perform a comparative genomic analysis, looking for the presence or absence of virulence factor genes and eventually generating bioinformatics-based hypotheses to address strain-to-strain variation in pathogenicity and epidemiological outcomes.

The *Y. pestis* module reinforces the bioinformatic skills and core concepts learned in the first module, and provides an opportunity for students to delve further into analysis of genetic events, such as insertions, deletions, and SNPs responsible for interstrain variation. *Y. pestis* is the causative agent responsible for three historical global pandemics of the bubonic plague or Black Death (∼550 A.D., 1347 A.D., 1850 A.D.; Drancourt *et al*., 2004; Stenseth *et al*., 2008). This human pathogen is notorious for its ability to cause widespread death, as in the second pandemic, in which ∼30–60% of the entire European population succumbed. Through use of the instructional materials provided, students learn about two routes of transmission for *Y. pestis* infection of humans (Chamberlain, 2004). In the first, fleas transfer the bacterium by first biting infected small rodents, such as rats, and subsequently biting humans, resulting in bacterial infection in the human bloodstream (bacteremia). The second route is through human respiratory infection, in which infected individuals spread the bacteria to others via droplet infection. Students also learn that strains of *Y. pestis* are typically categorized into one of three biovars (Antiqua [east African origin], Mediaevalis [central Asian origin], or Orientalis [central Asian origin]) based on their experimentally determined ability to use the carbon sources arabinose or glycerol and the nitrogen source nitrate (Devignat, 1951), and discover that the inability of *Y. pestis* strains to utilize one of these two carbon sources or the nitrogen source can be traced to mutations in one of three genes, *araC*, *glpD*, or *napA*. This gene analysis provides an opportunity for instructors to introduce or reinforce students' content knowledge, including the connection between genetic traits and biochemical pathways; pseudogenes, missense, and nonsense codons; and the potential effects of single amino acid changes on metabolic enzyme activity. The five *Y. pestis* genomes in this module (CO92 [Parkhill *et al*., 2001], KIM [Deng *et al*., 2002], 91001 [Song *et al*., 2004], Antiqua and Nepal [Chain *et al*., 2006)]), include at least one representative of each biovar. Using BLAST, students classify each strain into biovars based on analysis of the three genes. BLAST comparisons with the genome of a sixth strain isolated from North America (YPE), allow students to deduce whether Pacific or Atlantic Ocean trade routes were most likely responsible for the arrival of *Y. pestis* in North America. Additionally, students use BLAST comparisons of modern *Y. pestis* DNA with sequences generated from the dental pulp of corpses of humans thought to have died during the first and second pandemics of the Black Death (Drancourt *et al*., 1998, 2004; Tran-Hung *et al*., 2007) to determine which biovar(s) are most similar to the strains that caused the two previous pandemics. In a final component of this module, students apply the comparative genomic techniques learned from the *E. coli* module to address variation in virulence factors among *Y. pestis* strains. One strain included in the genome alignment (91001) has lost the ability to cause disease in humans (Song *et al*., 2004). Students identify genomic islands conserved in four genomes (Antiqua, Nepal, KIM, CO92), but absent in the genome of strain 91001. Analysis of the genes contained within permits students to formulate new hypotheses about why the genes in these islands may be important for causing human disease.

To assess the effectiveness of these modules, we collected and analyzed data on students' self-efficacy, abilities, and content knowledge. Our pre- and posttest assessment data indicate that students reported gains in their ability to use BLAST and to analyze gene content and conservation in genomic islands. Analytical skills and content knowledge–based assessments support the claim that students achieved the learning objectives and were able to apply their newly acquired abilities to address open-ended questions requiring experimental design, deductive reasoning, and literature-based analyses of experimental evidence. Throughout this work, we refer to use of the BLAST and Mauve tools as "abilities" and the application of these abilities to formulating hypotheses, problem solving, analysis, and synthesis as "skills." These educational modules are the first to integrate epidemiological, phenotypic, paleomicrobiological, and bioinformatic analyses to help students learn about the consequences of variation in genome content among human pathogens. Moreover, they offer educators new, engaging resources to immediately integrate comparative genomic topics into their undergraduate curricula.

MATERIALS AND METHODS

General Study Design

These modules were used and learning-assessment data were collected from nine educational settings at three large public universities and two private, liberal arts colleges (Table 1). Consent was sought and granted by students in cohorts 1 and 6–9 in accordance with Institutional Review Board (IRB) guidelines for all student data presented from the use of these modules. Participation in the assessments by students in

Table 1. Description of student samples included in this study

Sample number	Course	Year	Number of students	*E. coli* module	*Y. pestis* module
1	Bioinformatics	2009	13	Y	Y
2	Microbiology	2008	15	Y	N
3	Microbiology	2008	12	Y	Y
4	Microbiology	2009	20	Y	N
5	Microbiology	2010	23	Y	Y
6	Genomics	2009	10	Y	N
7	Genomics	2010	17	Y	N
8	Microbiology	2011	14	Y	N
9	Microbiology	2008	14	Y	N

cohorts 2–5 was voluntary, in that students could choose not to complete the assessments. The IRB boards at all institutions either approved these studies or ruled these studies exempt. Analysis of statistical significance was performed using the Student's paired *t* test with a one-tailed distribution and two-sample equal variance.

Participant Population

Student demographic information was collected for all samples. Of the 138 students who participated in these studies, 52 were male and 86 were female. Twenty-six were sophomores, 53 were juniors, 41 were seniors, and 18 were graduate students. The racial/ethnic composition of the participant pool was: 94 Caucasian, 21 Asian, 14 Latino/Hispanic, four African American, two African, one student from the Indian subcontinent, one Native American, and one Latino/Pacific Islander. The degree majors represented among the students included biology (81%), biochemistry (4%), chemistry (4%), economics (2%), and microbiology (9%).

Assessing Changes in Self-Efficacy

Student self-efficacy assessment was performed as previously described (Likert, 1932) using pre- and posttest self-report questions administered to students (Table 2). Pretest data represent a compilation of the data from samples 1–8 ($n = 124$). Posttest data were separated for classes that used only the *E. coli* module (samples 2, 4, 6, 7, and 8 [$n = 76$]) and for those that used both modules (samples 1, 3, and 5 [$n = 48$]). Self-efficacy assessment questions 1, 2, and 3 were used in all but one course (samples 1 through 8), while assessment question 4 was used in only five courses (samples 1, 5, 6, 7, and 8).

Assessing Changes in Abilities and Higher-Level Thinking Skills

To determine whether student learning gains occurred through use of these modules, formative assessments (Hutchings, 2000; Mettetal, 2001) of students' abilities to use their newly acquired BLAST and Mauve abilities were conducted by applying standardized rubrics to evaluate written responses to the course assignments (see course assignments 1 [for student samples 1, 4, 8, and 9] and 2 [samples 1 and 3] in the Supplemental Material) turned in by students or pairs of students (Tables 3 and 4). Written student products also provided the data to assess whether these resources can be used to promote higher-level analysis (Table 5) and synthesis (Supplemental Figure S7) skills through guided student inquiry (samples 2 and 3).

Assessing Changes in Content Knowledge

To further determine whether student learning gains occurred through use of these modules, assessments of students' content knowledge were conducted by evaluating written student responses to the course assignments (see course assignments 1 and 2 in the Supplemental Material), as described in the preceding section. Acquisition and retention of content knowledge were further assessed using qualitative evaluation of voluntary student written responses to pre- and posttest questions (sample 5; Table 6). Students' written responses to a take-home exam question about a hypothetical

Table 2. Pre- and posttest self-efficacy assessment ($n = 124$)

Student response option	Pretest (% ± SE)	Posttest *E. coli* only (% ± SE)	Posttest *E. coli* and *Y. pestis* (% ± SE)
Question 1: Circle the description that best describes you.			
I use Blast frequently and am confident in my ability with it.	6.5 ± 2.7	22.4 ± 4.3	41.0 ± 12.4
I am familiar with BLAST and could probably find my way around with it.	36.9 ± 18.6	77.6 ± 4.8	56.5 ± 11.4
I have heard of BLAST and have a vague idea of what it is.	33.7 ± 12.2	0 ± 0	2.5 ± 2.5
I have no idea what BLAST is.	22.8 ± 8.8	0 ± 0	0 ± 0
Question 2: I know how to use MAUVE to identify a unique genetic island that is not shared by closely related bacterial strains.			
Strongly disagree	67.6 ± 10.0	0 ± 0	0 ± 0
Disagree	30.4 ± 11.5	8.0 ± 5.2	16.0 ± 12.5
Agree	2.0 ± 1.3	62.6 ± 11.7	52.0 ± 8.1
Strongly agree	0 ± 0	29.4 ± 14.5	32.0 ± 20.7
Question 3: Once a genomic island is identified, I know how to analyze the genes contained within, and determine whether any may be involved in virulence.			
Strongly disagree	43.5 ± 16.9	0 ± 0	0 ± 0
Disagree	36.8 ± 12.7	2.5 ± 2.3	14.3 ± 12.5
Agree	18.4 ± 9.1	76.9 ± 2.9	70.4 ± 9.4
Strongly agree	1.3 ± 1.1	20.6 ± 0.5	15.3 ± 6.5
Question 4: Given a gene that confers a virulence trait in one pathogen, I know how to determine whether the gene is conserved in other related pathogens.			
Strongly disagree	31.7 ± 21.8	0 ± 0	0 ± 0
Disagree	39.0 ± 9.2	2.7 ± 1.8	5.2 ± 5.1
Agree	26.8 ± 14.5	64.8 ± 19.7	53.8 ± 2.4
Strongly agree	2.5 ± 2.3	32.5 ± 20.7	41.0 ± 5.6

Table 3. Scoring rubric for formative assessment of student learning based on written responses to assignment (course assignment 1 in the Supplemental Material) on *E. coli* module

Learning objective	Demonstrated competence with the tool[a]	Demonstrated skill in applying the concept/approach[a]
1. Improve student's ability to use BLAST	Did the student conduct a BLAST search of his/her assigned virulence gene?	Did the student determine whether the assigned gene was present in microorganisms other than *E. coli*?
2. Be able to identify genomic islands from whole genome alignments	Did the student identify a genomic island unique to a single genome?	Did the student identify a genomic island unique to a subset (two out of three) of the genomes?
3. Know one way to explore existing annotation for genes in a genomic island and determine whether any are involved in virulence	Did the student identify gene products located in a genomic island?	Did the student formulate a hypothesis as to how the proteins encoded on this genomic island may contribute to the microorganism's virulence?[b]
4. Be able to conduct analyses addressing conservation of genes in *E. coli* O157:H7 or *Y. pestis* strains	Did the student determine using BLAST and Mauve whether the assigned virulence gene was in other *E. coli* genomes?	Did the student analyze the BLAST and Mauve results correctly to determine whether multiple copies of the assigned virulence gene existed in any of the genomes?

[a]Items were scored as Yes/No unless otherwise noted.

[b]Students correctly identified a wide variety of phage-related genes (e.g., those encoding replication proteins, portal proteins, hydrolases, capsid components, and tape measure proteins), as well as transposase and integrase genes, as being unrelated to virulence. Genes identified by students as being highly relevant to differences in pathogenicity included shiga-like toxins and the host-cell adhesion gene *iha* (one student cited literature showing that Iha allowed the bacterium to adhere to kidney cells). Other students pointed to ureases and a short-chain dehydrogenase/reductase as potentially conferring the ability to survive in new environments and/or exploit unique nutritional resources. Finally, students hypothesized that complement resistance proteins and proteins involved in resistance to oxidative stress and phagocytic activity could contribute to a strain's virulence by enabling it to withstand host defenses.

Table 4. Scoring rubric for formative and skills assessment of student learning based on written responses to assignment (course assignment 2 in the Supplemental Material) on *Y. pestis* module

Learning objective	Demonstration of competence with the tools and skill in applying the concept/approach[a]	Demonstration of critical thinking and synthesis of information gained through the use of Mauve and BLAST[a]
1. Improve student's ability to use BLAST	Did the student conduct a BLAST search of the *glpD*, *napA*, and *araC* genes?	Did the student correctly interpret the SNP and BLAST data and successfully assign each strain and dental pulp sample to the correct biovar?
2. Be able to identify genomic islands from whole genome alignments	Did the student identify a genomic island absent from strain 91001 but present in the four pathogenic strains?	Not applicable
3. Know one way to explore existing annotation for genes in a genomic island and determine whether any are involved in virulence	Did the student explore the annotations for gene products located on a genomic island that is absent from strain 91001?	Did the student formulate a hypothesis as to how the proteins encoded on this genomic island may contribute to the microorganism's virulence?[b]
4. Be able to conduct analyses addressing conservation of genes in *E. coli* O157:H7 or *Y. pestis* strains	Did the student analyze the BLAST and Mauve results correctly to determine whether his/her assigned virulence gene is present in all five *Y. pestis* strains, and whether it is predicted to be functional in 91001?	Did the student successfully determine that the genetic evidence supports a Pacific trade route origin for the North American lineage?

[a]Tasks were scored as Yes/No unless otherwise indicated.

[b]Students typically pinpointed genes implicated in iron uptake and metabolism as being potentially relevant to virulence, because bacteria require iron and exhibit tight regulation of these functions. Other positive findings included fimbrial proteins, which the students suggested were involved in adhesion to host cells. Students correctly inferred that putative phage tail proteins, antirepressors, host-specificity proteins, and transposases are unlikely to be involved in virulence. One student found a putative sulfatase and sulfatase modifier and concluded that these genes were insufficient to cause virulence because they were found in the nonpathogenic *E. coli* strain K12 as well.

Table 5. Learning objectives and observed student outcomes upon completion of *E. coli* O157:H7 or *Y. pestis* modules[a]

	Demonstrated student achievement					
	E. coli O157:H7 module			*Y. pestis* module		
Learning objective	Student outcome	Competency with tool	Skill in applying concept/approach	Student outcome	Ability and application competency	Synthesis/ critical-thinking skills
1. Improve student's ability to use BLAST	Students were able to correctly identify the gene and protein sequence of the genes possibly involved in virulence.	98% (47/48)	90% (43/48)	Students were able to correctly deduce the biovars of the five strains and the dental pulp samples using BLAST results, as well as obtaining protein and ORF information following BLAST searches. In their answers, they were able to explain what the e values mean and draw the correct conclusions.	100% (17/17)	94% (16/17)
2. Be able to identify genomic islands from whole-genome alignments	Students were able to correctly identify islands of conservation or dissimilarity from whole-genome alignments.	100% (48/48)	98% (47/48)	Students were able to correctly identify islands of conservation or dissimilarity from whole-genome alignments.	100% (17/17)	Not applicable
3. Know one way to explore existing annotation for genes in a genomic island and determine whether any are involved in virulence	While the students were able to identify the hypothetical functions of suspected virulence genes, their conclusions about whether the genes are actually involved in virulence were tentative, because they did not feel that they understood the virulence pathways in *E. coli* very well.	98% (47/48)	81% (39/48)	Students were able to identify annotated functions of predicted proteins encoded on a chosen island, as well as obtain information on related proteins using InterPro Scan data. Students were able to deduce that the presence or absence of a few virulence factors may or may not be sufficient evidence to conclude whether a strain is virulent or not.	100% (17/17)	88% (15/17)
4. Be able to conduct analyses addressing conservation of genes in *E. coli* O157:H7 or *Y. pestis* strains	Students were able to identify whether the genes exist in all three genomes or just one, and whether identified islands are present in different strains.	96% (46/48)	83% (40/48)	Students were able to determine whether their assigned virulence gene is present in all five strains, but fewer than half analyzed whether the gene was functional. Students utilized BLAST similarity comparisons to draw conclusions on whether the strains originated from the Pacific or Atlantic region.	100% (11/11) for gene presence; 45% (5/11) for gene function	71% (12/17)

[a]Data are reported as percent of students who successfully accomplished the task outlined in the scoring rubrics (Tables 3 and 4).

Table 6. Assessment of student acquisition and retention of content knowledge

Question and evaluation of answer	Pretest	Posttest
What is a genomic island?[a]	$n = 23$	$n = 14$
No idea/no answer	57%	0%
Wrong	39%	14%
Partial credit	4%	50%
Correct	0%	36%
How does genetic information get into an island? How do you know how it got into an island?[b]		
No idea/no answer	65%	14%
Wrong	13%	0%
Partial credit	22%	43%
Correct	0%	43%
Suppose you had a group of bacterial strains from different regions of the world that are all the same species, yet some are more virulent than others. If you sequenced the genomes of all these strains, what feature(s) would you look for in those genome sequences that might confer strain-to-strain variability in virulence among bacterial strains of the same species?[c]		
No idea/no answer	48%	7%
Wrong	17%	0%
Partial credit	13%	14%
Acceptable	22%	36%
Good	0%	43%

[a]Full credit was given for responses that included the concept of sequences or sets of genes unique to one strain within a bacterial species that may confer virulence. Partial credit responses lacked the possible link to virulence.

[b]Full credit responses included both some mention of modes of horizontal gene transfer (e.g., phage transduction) and evidence, such as phage gene remnants and/or transposition-related sequences (transposases, insertion sequences). Partial credit responses typically failed to answer the second question.

[c]Good answers described possible virulence factor functions (e.g., toxins, iron uptake, adhesins, etc.). Acceptable answers invoked differences in gene content without mentioning specific potential functions. Partial credit was given for "genomic islands."

E. coli outbreak served as an additional mode of assessing student content knowledge (sample sets 7 and 8).

Instructional Resources

Information about the computational requirements for using Mauve and materials, methods, and figures pertaining to the *E. coli* and *Y. pestis* genome alignments are provided for instructors (see the Supplemental Material). Introductory materials for instructors to use with their courses are available as PowerPoint slides for *E. coli* O157:H7 (Supplemental Slide S1) and *Y. pestis* (Supplemental Slide S2). These slides introduce the tools used in these modules and provide background genomic, epidemiological, and historical information. For instructors and/or students, information about the use and interpretation of BLAST results can be found in the BLAST information guide available at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html) and additional simple explanations about using BLAST can be found in chapter 7 of the book *Bioinformatics for Dummies* (Claverie and Notredame, 2003).

Student Instructional Resources

Instructions for students are provided to assist with using the Mauve alignments and the ASAP database (see instructional material 1 and 2 in the Supplemental Material). Additional independent exercises consisting of multiple inquiry-based questions that can be used with these modules are provided (see course assignments 1–3 in the Supplemental Material).

RESULTS

To determine the extent to which our student learning objectives were achieved, we used a variety of assessment instruments, including pre- and postmodule knowledge-based questionnaires, scoring of student responses to guided inquiry-based individual assignments, and open-ended investigations/exam questions that required students to apply their newly acquired abilities and their understanding of core concepts, such as the potential functions of virulence factors or the effects of SNPs on metabolic capabilities. In addition to these tools, which were used to assess acquisition of abilities and content knowledge, self-efficacy assessment rating scales were used to investigate whether students, through use of these problem-based learning modules, gained confidence in their abilities to use BLAST, Mauve, and the ASAP database to identify and analyze the contents of genomic islands in two microbial pathogens, and to answer interesting questions about conservation of virulence factors. In our analyses of these self-efficacy data, student responses for the posttest assessment were separated into two groups; the first student group was composed of those who solely used the *E. coli* O157:H7 module, and the second set represented those who completed both the *E. coli* O157:H7 and the *Y. pestis* modules.

Learning Objective 1: Did Students Report Gains in Their Abilities to Use BLAST through Use of One or Both Modules?

Self-efficacy assessment revealed that the majority of the students surveyed (56.5%) had little familiarity or confidence using BLAST prior to this exercise, and use of these modules resulted in significant ($P < 0.05$) increases in student responses that indicated familiarity (76.4% [± 4.8]) or even confidence (22.4% [± 4.4]) using BLAST upon completion of the *E. coli* module alone (Table 2). Furthermore, students who used both modules felt more confident using BLAST (41.0% [± 12.3]), in comparison with those who completed only the *E. coli* module (22.4% [± 4.4]) (Table 2). Formative assessment of the students' abilities to conduct a BLAST search was determined using the rubrics shown in Tables 3 and 4 to score written responses handed in by students upon completion of a module. Essentially all students were able to use the Mauve interface successfully to conduct a BLAST search in *E. coli* for an assigned virulence factor (Table 5). Ninety percent of students properly conducted BLAST searches against other microbial genomes (Table 5). Finally, the *Y. pestis* module asks students to apply their skills with BLAST to determine the biovar most likely to have caused the first and second pandemics, and almost all students accomplished this task correctly. In their written responses to the assignments (provided in the Supplemental Material), students typically cited a correct interpretation of e values to support their

conclusions. Overall, these results indicate that the module instructions are clear and provide evidence for success in helping students improve their self-efficacy and ability in using BLAST for inquiry-based research.

Learning Objective 2: Were Students Able to Identify Genomic Islands Using Mauve?

While multiple comparative genomic software tools exist, Mauve is unique, in that it allows more than two genomes to be aligned, has a student-friendly visual interface to interpret the genome alignment, and when using genomes obtained from the ASAP database, contains direct Web links from the genes to annotation pages, thus permitting students to survey additional information regarding the genes' roles in pathogenesis. We sought to determine whether this comparative genomic tool could be used effectively by students to identify polymorphisms that differentiate strains within a given species. Not surprisingly, our pretest self-efficacy assessment data indicated that almost none of our students felt knowledgeable about comparing multiple genomes to identify genomic islands (regions unique to a single genome; Table 2). The posttest data revealed that 92% or 84% of students surveyed gained confidence in using a Mauve multiple-genome alignment to identify a genomic island upon completion of one or both exercises, respectively ($P < 0.0001$; Table 2). Analysis of written student responses to the assignments corroborated these data, indicating that the vast majority of students successfully identified a genomic island unique to a single genome or to a subset of the genomes (Table 5). On the basis of a pretest knowledge assessment, 96% of students in one class (sample 5) had no prior knowledge of what a genomic island is; by the end of the course, 5 wk after using the second module, 86% of the students had acquired and retained at least partial (50%) or accurate (36%) understanding of this core concept (Table 6). Thus, these educational modules based on Mauve appear to be an effective tool to enable students to learn important comparative genomic approaches and to acquire content knowledge.

Learning Objective 3: Did Students Learn How to Determine Whether Any Genes in a Genomic Island May Be Involved in Virulence?

Genomic comparison of *E. coli* K-12 (strain MG1655) and O157:H7 (strain EDL933) revealed that many of the differences that distinguish these *E. coli* strains may be attributed to the contents of genomic islands unique to the pathogenic strain (Perna *et al.*, 2001). Genomic island analysis can also be applied to the genomes of multiple pathogenic strains to identify variations and develop new hypotheses that address epidemiological data associated with historical outbreaks. This learning objective addressed the insight that variations in the gene content of genomic islands may correlate with variations in pathogenesis from one strain to another. Therefore, we sought to determine whether changes occurred in students' self-efficacy and in their ability to analyze the genes contained in genomic islands and to develop hypotheses to address strain-to-strain variability among microbial pathogens. Based on student self-reports (Table 2), it is clear that most students (80%) did not possess the knowledge to approach such a scientific problem prior to use of these modules. Posttest data demonstrated a significant ($P < 0.0005$) increase in self-reported ability to discern whether any of the genes/ORFs contained within a genomic island may be involved in virulence, upon completion of one (98%) or both (88%) modules (Table 2). It is interesting to note that 14% of the students who completed both modules still did not feel that they had learned enough to address this type of problem.

We used student responses to the course assignments (see the Supplemental Material) to independently assess students' abilities to identify the genes contained within genomic islands and to relate their predicted functions to potential roles in pathogenicity. These data demonstrate that essentially all of the students successfully analyzed the gene contents of a genomic island by using the annotation links to ASAP, and that 80–90% were able to formulate a hypothesis as to why the presence or absence of the genomic island contents may contribute to the microorganism's virulence (Table 5). Not unexpectedly, students who completed the *Y. pestis* module were more likely to have successfully addressed possible roles for the island-borne genes than students exposed to the roles of virulence factors for the first time during the *E. coli* module (Table 5). Pre- and posttest assessment revealed substantial gains in students' content knowledge about the types of functions (e.g., adhesins, iron-carrying proteins, toxins, and secretion systems) characteristic of virulence factors (Table 6), consistent with the notion that exploration of possible virulence factors encoded within islands may have helped the students recall these functions.

Many of the genetic islands present in the *E. coli* and *Y. pestis* strains contain a large number of phage-related and insertion element–related genes, suggesting they originated through a phage-transduction event. Depending on how familiar they are with phage biology, students may or may not recognize these as phage genes or remnants from the annotations provided in ASAP. As students explore the genomic island content, instructors can use the presence of these genes as a starting point for a discussion of horizontal gene transfer. In one class, written lab responses to the open-ended question "From your analyses, what is one mechanism that has caused divergence of these genomes?" showed that 10 out of 14 pairs of students were able to correctly deduce that phage transduction was responsible (data not shown). Interestingly, a pretest knowledge question posed to another class of students at the same institution revealed that 22% of the students had some inkling (almost certainly from an earlier genetics prerequisite course) of phage transduction as a possible explanation for the presence of genetic information within a genomic island, even though only 4% of those students could articulate a cogent definition of a genomic island before using the module (Table 6). After experience with both modules, 86% of the students gave a response that was at least partially correct.

Learning Objective 4: Through Use of One or Both of These Modules, Did Students Improve Their Ability to Conduct Analyses of Virulence Factors in Multiple Genomes of Microbial Pathogens?

To determine whether students could use BLAST as a tool to address the conservation of genes thought to play a role in pathogenesis, we provided students with genes identified as virulence factors by scientific experts at the ASAP

genome database and asked students to determine whether the gene/ORF was conserved in the other strains included in the comparative genomic analysis. Student pretest data indicated that most students surveyed (70%) did not feel confident in their ability to address this type of scientific question prior to these exercises (Table 2). Therefore, these modules provided opportunities for students to learn how to apply BLAST to search for virulence genes from one genome to another. The posttest data revealed that use of the modules combining BLAST, the ASAP database, and the multiple-genome alignment tool Mauve resulted in significant ($P < 0.05$) gains in student self-efficacy. Specifically, 97% or 95% of students expressed confidence ("agree" or "strongly agree") in approaching this task upon completion of one or both modules, respectively (Table 2). Additional practice appeared to enhance student learning, with 41.1% (±5.5) of the students strongly agreeing that they could address this type of question upon completion of both modules, as compared with 32.4.% (±10.4) of those who used only the *E. coli* module. Our formative assessment data confirmed that almost all students were able to determine conservation of an assigned virulence gene across numerous *E. coli* O157:H7 genomes, but students were somewhat less competent at using BLAST to address the question "Is this gene or a homolog found in other Enterobacteria?" (Tables 3 and 5). Tellingly, less than half of the students who used the *Y. pestis* module ascertained whether their assigned virulence gene was actually predicted to be functional in all the genomes, although all of the students reported whether it was present, and the assignment explicitly asked about gene product function (Table 5). Some students using this module also exhibited difficulty in fully synthesizing the data needed to determine whether the North American lineage is more likely attributable to Atlantic or Pacific trade routes (Tables 4 and 5). In this case, four of 11 pairs of students in one sample drew the wrong conclusion, because they relied exclusively on the SNP analysis and failed to notice a large deletion in the *glpD* BLAST analysis.

Questions of conservation lend themselves well to a "teachable moment" regarding the choice of nucleotide versus protein BLAST. In one group of 28 students, students were asked to provide a written response justifying their choice of using BLASTP or BLASTN. Twelve of the 14 pairs of students provided answers that were complete and exhibited clear comprehension of relevant concepts, including third position wobble. One pair gave an answer that was adequate, although not thorough, while the last pair's response invoked introns, an informative answer, in that it revealed a misconception grounded in a basic understanding of the Central Dogma, concerning the absence of splicing in bacteria.

Together, these data demonstrate that these modules can be used both to teach about gene conservation and to help students develop analytical and synthesis skills. To further illustrate this point, one instructor used the assignment of virulence factors as a springboard for a larger unit on type 3 secretion systems (T3SSs). The "locus of enterocyte effacement" (LEE) genomic island is a virulence determinant encoding a number of proteins, including the components of a T3SS, required for attachment of enteropathogenic *E. coli* strains to host intestinal cells (Jerse *et al*., 1990). Found in many plant and animal pathogens, T3SSs function to deliver bacterial virulence factors that typically subvert host defenses and/or manipulate host processes, such as cytoskeletal rearrangements (Galan and Collmer, 1999). The LEE T3SS translocates substrates that cause the formation of a raised pedestal or "docking platform" on the surface of host cells and the Tir protein that serves as a receptor for the bacterial membrane adhesion intimin (Kenny *et al*., 1997). Prior to any class discussion on T3SSs, students explored the conservation of the LEE genes as part of the *E. coli* module. As one aspect of this analysis, the students were asked to investigate the literature links for their assigned LEE gene in the ASAP database, and to report in writing "What do we know?" (evidence that this gene contributes to virulence); "What do we think we know?" (predicted or known function of the encoded gene product); and "What do we need to know more about?" (open/unanswered questions about the protein, but also background concepts or elements with which the students were unfamiliar). Even in the absence of any prior knowledge about T3SSs, the morphological changes incited by the pathogen, or the key delivered substrate Tir, pairs of students were able to identify each of these pieces of information for their assigned gene (Table 7). To gain practice in synthesizing disparate pieces of data, the students were then organized into groups around related gene products and asked to assemble a concept map, with each student "expert" contributing information on his/her individual virulence factor (see course assignment 3 in the Supplemental Material). The exercise culminated with the entire class assembling one large, and highly accurate, concept map (Supplemental Figure S7), demonstrating that students can use the information in the ASAP database, together with literature searches, as a tool for independent analysis and synthesis of information about a complex regulatory system.

Learning Objective 5: Could Students Apply These Newly Acquired Skills to Design a Bioinformatic Investigation of an **E. coli** *Outbreak?*

To determine whether students could apply the combination of newly acquired skills and content knowledge, we asked them to design a bioinformatic study to approach the challenge of determining which genes might be contributing to the unique virulence of a new outbreak strain. Student responses were evaluated on five criteria that encompassed biology, bioinformatics, and experimental process, as outlined in Table 8. Sixty-six percent of students provided responses that demonstrated understanding of virulence factors/genes, and 61% of students could define the evolution, structure, and function of pathogenicity islands (Table 8). Sixty-nine percent of students recognized the need to examine genomic islands, 72% of students were able to describe a bioinformatic approach to identify genomic islands, and 50% of students were able to describe how to determine the function of genes contained in these islands (Table 8). Although students did not conduct wet-lab experiments as part of these modules, 39% of students understood that experimental verification is needed to further demonstrate that one or more of the genes contained within an island might function as virulence factors (Table 8).

Overall, these data indicate that the student gains in abilities and content knowledge aligned with student self-reports of increased self-efficacy relevant to the five learning objectives for these comparative genomic educational modules. Both learning assessments and observed student outcomes

Table 7. Summary of student investigations of genes contained within the LEE island

	Students were able to identify:[a]			
Assigned virulence gene on LEE island	Evidence for role in virulence	Known or predicted function	Open/unanswered questions[b]	Unfamiliar concepts/ elements[c]
escN	X	X		
escJ	X	X	X	
rOrf1	X	X	X	
espZ	X	X	X	
cesT	X	X	X	X
cesD	X	X	X	X
espG	X	X		X
espH[d]	X	X	X	X
espH[d]	X	X		X
espF	X	X	X	
sepL	X	X	X	
grlA	X	X	X	
grlR	X	X	X	
	X	X	X	X

[a] An X indicates students successfully identified the information indicated for their assigned virulence gene.

[b] Answers included the need for additional experiments to confirm reported protein–protein interactions and how those interactions affect virulence, functions of gene products reported to be regulated by the assigned LEE-encoded protein, mechanisms of action of the LEE-encoded protein, and how the protein in question can act as a chaperone if it does not directly interact with its target.

[c] Answers included definitions of terms, including brush border remodeling, membrane ruffling, T3SS, and attachment/effacing lesions.

[d] Two pairs of students investigated this gene.

Table 8. Student responses to the following question: "Imagine you are a genomicist working for CDC and there has been an outbreak of enterohemorrhagic disease that has resulted in illness of 100,000 people and deaths of 1000 of those patients, making this the most deadly outbreak of *E. coli* that has ever been reported. Bacteria cultured from the fecal material of some of the patients all revealed the same strain of *E. coli* that had never before been described. You have funds to sequence the genome of this strain. Design a bioinformatic study in which you approach the challenge of determining which genes might be contributing to the extreme virulence of this strain."

Criteria	Student response	% ($n = 36$)
Biology		
1. Student can define and describe the concept of a virulence factor, and briefly describe at least two classes of genes that function as virulence factors.	Provided a complete, well-reasoned, answer	66.7
	Provided an incomplete/superficial answer	27.8
	Did not address this topic	5.5
2. Student can define the evolution, structure, and function of pathogenicity islands.	Provided a complete, well-reasoned, answer	61.1
	Provided an incomplete/superficial answer	30.6
	Did not address this topic	8.3
Bioinformatics		
3. Student can explain that a comparison of genome sequences between closely related strains of *E. coli* will reveal genomic islands unique to one or more strains, and these may represent pathogenicity islands.	Provided a complete, well-reasoned, answer	69.4
	Provided an incomplete/superficial answer	27.8
	Did not address this topic	2.8
4. Student can describe a bioinformatics approach to identify genomic islands.	Provided a complete, well-reasoned, answer	72.2
	Provided an incomplete/superficial answer	27.8
	Did not address this topic	0.0
5. Student can describe a bioinformatics approach to learn about the potential functions of genes located within the genomic island.	Provided a complete, well-reasoned, answer	50.0
	Provided an incomplete/superficial answer	41.7
	Did not address this topic	8.3
Experimental process		
6. Student understands that experimental verification is needed to demonstrate that one or more of the genes contained within an island might function as virulence factors	Provided a complete, well-reasoned, answer	38.9
	Provided an incomplete/superficial answer	25.0
	Did not address this topic	36.1

illustrate that topics of bioinformatics, comparative genomics, and virulence factor analysis are appropriate subjects for students to study to learn about these new computational inquiry-based approaches to address scientific questions.

DISCUSSION

Comparative genomics offers a unique scientific approach that enables student exploration of the relationship between evolutionary variations in genomes and epidemiological outcomes. In an effort to help update undergraduate education in the fields of genomics and bioinformatics, we designed two modules that reinforce the use of BLAST and introduce comparative genomic alignment techniques to identify strain-to-strain differences in bacterial genomes of microorganisms that cause human disease. These tools enable students to perform these analyses in much the same way as real-world scientific researchers do. Learning assessment was conducted using pre- and posttest student questionnaires to determine whether a set of five learning objectives were achieved and to examine gains in students' self-efficacy, skills, and content knowledge in approaching these challenging tasks. In this paper, we discuss the learning objectives, results, and some of the changes we made to improve these modules based on classroom use and student feedback.

To address the first and fourth learning objectives, students used BLAST to analyze virulence factors and their conservation among genomes. Assessment data showed that using one or both of these modules improved students' confidence in using and competence with this type of analysis. Since the initial publication by Altschul *et al*., (1990) describing the BLAST algorithm, the tool has been referred to as "one of the most widely used bioinformatics programs," and this immense use is evidenced in the sheer number (>30,000) of citations of this publication in the scientific literature. This widespread use of BLAST in the scientific community argues for the importance of students gaining familiarity and confidence in using the tool. Familiarity with BLAST is also crucial for an entry-level understanding about gene annotations and deriving information about new genes. Such information is commonly based on prior results from BLAST comparisons with known genes/proteins (Baumler *et al*., 2008; the Joint Genomic Institute's "adopt a microbial genome" program; and the HHMI–Scientific Education Alliance's phage annotation initiative). Our student learning–assessment data revealed that, upon completion of one or both of these educational modules, 100% of students are familiar with and feel they can use BLAST to work/solve future problems and have gained confidence in their abilities to apply this technique and the associated content knowledge to future scientific inquiries. While this self-reported competence is certainly an overstatement of skill level, it is a good indicator of improvement through use of the exercise. Additional assessment tools corroborated these gains by documenting student skill levels (Table 5) and increases in content knowledge (Tables 6 and 8). Unfortunately, our ability to measure increases in acquisition and retention of content knowledge (Table 6) was partially compromised by the low posttest response rate, reflecting the fact that students in this cohort were under no obligation to fill out the assessment surveys. A tendency among students to provide short and less than complete answers in the rush of the end of a semester has previously been noted (Harris *et al*., 2009). In our case, it is entirely possible that students who were unsure of the answers may have chosen not to take the time to respond, and/or that students who had a clearer grasp of the material rushed to finish and wrote incomplete answers. This is not the case for the data reported in Table 8, in which students responded to an exam question, rather than to a voluntary survey.

In our pilot implementations in three educational contexts, the use of both educational modules, as compared with just the *E. coli* module, generally resulted in increased student confidence for all of our learning objectives. While these data are encouraging, it appears that there were some students who, upon completion of the second (*Y. pestis*) module, became less confident in determining whether genomic island contents contribute to virulence. This may be due a larger number of genes with candidate virulence factor annotations in *E. coli* O157:H7 ($n = 394$) than *Y. pestis* ($n = 148$). Thus, students have a higher probability of locating genes thought to encode virulence factors (ORFs encoding putative or known virulence factors/total ORFs) when surveying genomes of *E. coli* O157:H7 (7.7%) in comparison with *Y. pestis* (3.6%). Additionally, the second module focuses on genomic island analysis by asking students to analyze those regions apparently lost in the genome of the 91001 strain, and to speculate on which genes may account for this strain's inability to cause disease in humans. The genes responsible for this phenomenon are currently unknown, and the solution is not straightforward; therefore, students may have felt that they did not examine the contents of the genomic islands correctly. This point notwithstanding, when learning-assessment results were compared for students who used both modules versus the single module, it appeared that the use of the second module generally reinforced gains in self-efficacy and ability (Tables 2 and 5). This increase was likely achieved by posing a second intriguing set of scientific problems, in which the same tool can be used to further analyze similarities and differences of gene content among microbial pathogens.

Once students gained confidence in comparing single genes using BLAST, we sought to scaffold skills by providing students with hands-on experience in utilizing genome-scale BLAST queries, stored and readily available in Web-accessible databases, such as ASAP. One distinguishing feature of the ASAP database in comparison with other microbial genomic resources is that ASAP contains a copious amount of information added as "annotations." These annotations provide standard information, such as what the gene product/protein is (product annotation) and what the protein does (function), along with many other annotation types not commonly found in other genomic resources, including mutant phenotypes, curator comments, molecular interactions, overexpression phenotypes, and virulence factor classifications. Another distinct feature of the ASAP database is a clickable link alongside each annotation that directs the user to the source of evidence from which the annotation was derived, such as a scientific publication indexed in PubMed or other bioinformatic resource(s). In the case of pathogenic microorganisms, such as *E. coli* O157:H7 or *Y. pestis*, some genes contain numerous annotations supporting the notion that the gene product may contribute to the organism's ability to cause human disease.

This extensive collection of annotations for virulence factor genes serves as a powerful resource for students as they analyze the gene contents of genomic islands. In the *E. coli* O157:H7 module, students are asked to determine whether any of the genes in an island are similar to those for which evidence exists supporting their role in pathogenesis. In the *Yersinia* module, using a combination of BLAST, ASAP, and Mauve enables students to analyze genotypic differences across multiple microbial genomes and derive bioinformatics-driven hypotheses that may address known epidemiological, historical, experimental, and phenotypic information. The extended LEE unit we have described allowed students to discover for themselves, in a cooperative learning situation, what is known about one well-characterized virulence determinant. Among the outcomes demonstrated by the students who worked on this unit were successful evaluation of gaps in their own knowledge, ability to formulate questions (Table 7), and the ability to synthesize a complex set of information (Supplemental Figure S7); these are abilities associated with critical thinking and intellectual maturity (Allen and Tanner, 2002; Beck *et al*., 2010). We offer this unit as a model for instructors who wish to leverage the well-documented advantages of student-driven inquiry over conventional lectures (e.g., Knight and Wood, 2005; Armbruster *et al*., 2009) with the rich resource provided by the ASAP annotations.

In our experience, students tend to learn more when intrigued by problems or when questions with no "correct" answer are provided, and we used this pedagogical strategy to motivate student learning. Our approach, in which students use BLAST to address unresolved research questions and develop hypotheses about strain-to-strain differences in virulence, pathogenicity, epidemiology, and paleomicrobiology is integrative and encourages students to achieve the learning objectives of this exercise by engaging them in real-world problems. In addition to teaching about cross-genome comparisons, these modules expose students to information about a variety of virulence factors and the mechanisms by which they are thought to play a role in pathogenicity. The additional supplemental information provided as a component of these modules represents a comprehensive up-to-date survey of experimental literature related to genes in *E. coli* O157:H7 or *Y. pestis*. The different types of virulence factors are subcategorized and briefly described to provide students and instructors requisite knowledge about microbial virulence factors. This resource offers may possible avenues for expanded curricula based on these starting exercises.

Finally, one more unique and powerful feature of this set of instructional modules is that it can be used in a broad range of curricular settings. In this study, these modules were used in upper-level microbiology courses to help students learn how to study virulence, but they were also used in a genomics course to allow students to see how genomes evolve and confer new traits on organisms, and in a bioinformatics course. This set of instructional materials would likewise be entirely suitable for courses in other diverse areas of biology, including evolution or epidemiology, and could be combined with a set of wet-lab activities in which students transform *E. coli*, for example, to teach students about biology, not just about bioinformatics. In sum, these tools can serve as a vehicle to foster student understanding of bacterial biology, concepts of virulence and disease, proteins and genes, evolutionary mechanisms, how to work with genomic data, and how to use bioinformatics tools, potentially all in 1 wk.

CONCLUSION

Overall, these two modules represent a novel pedagogical approach to teaching molecular microbiology and comparative genomics; they focus on real-life, human disease–related questions that motivate students to learn and use bioinformatics by employing the same tools and resources used by researchers in the field. Learning-assessment results demonstrate that significant student gains in self-efficacy, ability, and content knowledge were achieved in 1) using BLAST; 2) understanding how to identify genomic regions of interest from a multiple genomic alignment; 3) analyzing the contents of genomic islands to derive bioinformatics-driven hypotheses relating to strain differences; and 4) learning about gene conservation. Our data further indicate that the depth of learning for most students is greater after using both modules. These educational modules are the first to integrate epidemiological, phenotypic, paleomicrobiological, and bioinformatics questions to help students learn about the consequences of variation in genome content among human pathogens. They offer instructors in a range of disciplines engaging new pedagogical resources for integrating comparative genomic topics into undergraduate curricula and have been implemented, tested, and refined in a variety of course contexts.

ACKNOWLEDGMENTS

This work was funded by the United States Department of Energy Genomics:GTL and SciDAC Programs (DE-FG02-04ER25627) for the BACTER Institute Post-Doctoral Fellowship (D.J.B); the National Library of Medicine, National Institutes of Health (NIH), grant number 5T15LM007359 to the Computation and Informatics in Biology and Medicine Training Program for a Post-Doctoral Traineeship (D.J.B); and in part with federal funds from the National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services, under contract number HHSN266200400040C. We thank Dr. Madeline Fisher for writing assistance and Drs. Bradley Anderson, Bob Mau, and Aaron Darling for technical assistance with this work.

REFERENCES

Allen D, Tanner K (2002). Approaches to cell biology teaching: questions about questions. Cell Biol Educ *1*, 63–67.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410.

Armbruster P, Patel M, Johnson E, Weiss M (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. CBE Life Sci Educ *8*, 203–213.

Baumler DJ, Hung KF, Cabot EL, Glasner JD, Greene JM, Perna NT (2008). Annotation of the Bacteriophage 933W Genome: An In-class Interactive Web-based Exercise. American Society for Microbiology MicrobeLibrary. Curriculum Collection. http://archive.microbelibrary.org/advsearchres.asp?rcat=4 (accessed 12 November 2011).

Beck RJ, Skinner WF, Schwabrow LA (2010). Researching Assessment Methods in Tutorial Education. Final Report to Teagle Foundation. www.teaglefoundation.org/learning/pdf/2010_Lawrence.pdf (accessed 15 October 2011).

Campbell AM (2002). Genomics in the undergraduate curriculum—rocket science or basic science. Cell Biol Educ *1*, 70–72.

Chain PS, Hu P, Malfatti SA, Radnedge L, Larimer F, Vergez LM, Worsham P, Chu MC, Andersen GL (2006). Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. J Bacteriol *188*, 4453–4463.

Chamberlain NR 2004. Transmission Cycles of Plague. American Society for Microbiology MicrobeLibrary. www.microbelibrary.org/library/resources/2823-transmission-cycles-of-plague (accessed 12 November 2011).

Checkley K 1997. Problem-based learning. ASCD Curriculum Update *Summer*, 3.

Claverie J-M, Notredame C (2003). Bioinformatics for Dummies, Hoboken, NJ: Wiley Publishing Inc.

Darling AC, Mau B, Blattner FR, Perna NT (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res *14*, 1394–1403.

Darling AE, Miklós I, Ragan MA (2008). Dynamics of genome rearrangement in bacterial populations. PloS Genet *4*, e1000128.

Deng W *et al.* (2002). Genome sequence of *Yersinia pestis* KIM. J Bacteriol *184*, 4601–4611.

Devignat R (1951). Varieties of *Pasteurella pestis*; new hypothesis. Bull World Health Organ *4*, 247–263.

Dexter Dyer B, LeBlanc MD (2002). Incorporating genomics research into undergraduate curricula. Cell Biol Educ *1*, 101–104.

Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D (1998). Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. Proc Natl Acad Sci USA *95*, 12637–12640.

Drancourt M, Roux V, Dang LV, Tran-Hung L, Castex D, Chenal-Francisque V, Ogata H, Fournier PE, Crubézy E, Raoult D (2004). Genotyping, Orientalis-like *Yersinia pestis*, and plague pandemics. Emerg Infect Dis *10*, 1585–1592.

Forst S, Goodner B (2006). Comparative bacterial genomics and its use in undergraduate education. Biol Control *38*, 47–53.

Galan JE, Collmer A (1999). Type III secretion machines: bacterial devices for protein delivery into host cells. Science *284*, 1322–1328.

Gallagher S, Stephien WJ, Sher BT, Workman D (1995). Implementing problem-based learning in science classrooms. School Sci Math *95*, 136–146.

Glasner JD, Rusch M, Liss P, Plunkett G III, Cabot EL, Darling A, Anderson BD, Infield-Harm P, Gilson MC, Perna NT (2006). ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Res *34*, D41–D45.

Glasner JD *et al.* (2008). Niche-specificity and the variable fraction of the *Pectobacterium* pan-genome. Mol Plant Microbe Interact *21*, 1549–1560.

Goodner B, Wheeler C (2006). Functional genomics: using reverse genetics to test bioinformatics predictions. American Society for Microbiology MicrobeLibrary. Curriculum Collection. http://archive.microbelibrary.org/advsearchres.asp?rcat=4 (accessed 12 November 2011).

Harris MA, Peck RF, Colton S, Morris J, Neto EC, Kallio J (2009). A combination of hand-held modules and computer imaging programs helps students answer oral questions about molecular structure and function: A controlled investigation of student learning. CBE Life Sci Educ *8*, 29–43.

Hayashi T *et al.* (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res *8*, 11–22.

Honts JE (2003). Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. Cell Biol Educ *2*, 233–247.

Hutchings P 2000. Introduction to Opening Lines: Approaches to the Scholarship of Teaching and Learning. Carnegie Foundation for the Advancement of Learning. www.carnegiefoundation.org/publications/opening-lines-approaches-scholarship-teaching-and-learning (accessed 12 November 2011).

Jerse AE, Yu J, Tall BD, Kaper JB (1990). A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc Natl Acad Sci USA *87*, 7839–7843.

Kenny B, DeVinney R, Stein M, Reinscheid DJ, Frey EA, Finlay BB (1997). Enteropathogenic E. coli (EPEC) transfers its receptor for intimate adherence into mammalian cells. Cell *91*, 511–520.

Knight JK, Wood WB (2005). Teaching more by lecturing less. Cell Biol Educ *4*, 298–310.

Likert R (1932). A technique for the measurement of attitudes. Arch Psychol *140*, 1–55.

Lopatto D, Elgin SC (2010). The Genomics Education Partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. CBE Life Sci Educ *9*, 55–69.

Lopatto D *et al.* (2008). Genomics Education Partnership. Science *322*, 684–685.

Manning SD *et al.* (2008). Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc Natl Acad Sci USA *105*, 4868–4873.

Mau B, Glasner JD, Darling AE, Perna NT (2006). Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*. Genome Biol *7*, R44.

Mettetal G (2001). The what, why, and how of classroom action research. JoSoTL *2*, 6–13.

National Research Council (2005). Washington, DC: National Academies Press.

Nehm RH (2010). Understanding undergraduates' problem solving processes. J Biol Microbiol Educ *11*, 119–122.

Parkhill J *et al.* (2001). Genome sequence of *Yersinia pestis*, the causative agent of the plague. Nature *413*, 523–527.

Perna NT *et al.* (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature *409*, 529–533.

Petrill SA, Justice LM (2007). Bridging the gap between genomics and education. Mind Brain Educ *1*, 153–161.

Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL (2005). Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. Emerg Infect Dis *11*, 603–609.

Shaffer CD *et al.* (2010). The Genomics Education Partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. CBE Life Sci Educ *9*, 55–69.

Shuster (2011). Can genetics and genomics nursing competencies be successfully taught in a prenursing microbiology course? CBE Life Sci Educ *10*, 216–221.

Song Y *et al.* (2004). Complete genome of *Yersinia pestis* strain 91001, an isolate avirulent to humans. DNA Res *11*, 179–197.

Stenseth NC, Atshabar BB, Begon M, Belmain SR, Bertherat E, Carniel E, Gage KL, Leirs H, Rahalison L (2008). Plague: past, present, and future. Plos Med *5*, e3.

Tran-Hung L, Tran-Thi N, Aboudharam G, Raoult D, Drancourt M (2007). A new method to extract dental pulp DNA: application to universal detection of bacteria. Plos One *2*, e1062.

Vernikos GS, Thomson NR, Parkhill J (2007). Genetic flux over time in the *Salmonella* lineage. Genome Biol *8*, R100.

Wiggins G, McTighe J (2005). Understanding by Design, expanded 2nd ed., Alexandria, VA: Association for Supervision and Curriculum Development.