

Special Section: Innovative Laboratory Exercises—Focus on Genomic Annotation

Strategies for Using Peer-Assisted Learning Effectively in an Undergraduate Bioinformatics Course*[§]

Casey Shapiro[†]
Carlos Ayont
Jordan Moberg-Parker[‡]
Marc Levis-Fitzgerald[†]
Erin R. Sanders[§]

From the [†]Center for Educational Assessment, Office of Instructional Development, University of California, Los Angeles, Los Angeles, California 90095, [‡]Department of Molecular Biology, Interdepartmental Ph.D. Program, University of California, Los Angeles, Los Angeles, California 90095, [§]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, California 90095

Abstract

This study used a mixed methods approach to evaluate hybrid peer-assisted learning approaches incorporated into a bioinformatics tutorial for a genome annotation research project. Quantitative and qualitative data were collected from undergraduates who enrolled in a research-based laboratory course during two different academic terms at UCLA. Findings indicate that a critical feature of the peer-assisted learning approach is to have near-peer leaders with genome annotation experience, allowing them to communicate technical and conceptual aspects of the process in the context of a research project (a.k.a., the “big picture”). These characteristics are im-

portant for creating connections between the wet lab experiments and the computer lab activities, engendering excitement about the research project and fostering engagement in bioinformatics as a discipline. Likewise, it is essential to couple tutorial training in genome annotation with appropriate instructional materials, providing detailed, step-by-step instructions for database navigation. Finally, the assessment results support this hybrid peer-assisted learning approach as a model for undergraduates to successfully learn bioinformatics in a course setting. © 2012 by The International Union of Biochemistry and Molecular Biology, 41(1):24–33, 2013

Keywords: peer learning; bioinformatics; annotation; genomics; assessment

Modern research scientists have embraced diverse yet powerful bioinformatics tools as a means to retrieve and investigate the underlying biological information in genomic datasets. Such inquiries result in the discovery of novel genes or new functional properties previously not known or attributed to particular organisms [1]. Exposure to databases and various online computational resources can begin at the undergraduate level ([2–8], and references therein). The *in silico* nature of this undertaking may trigger unease among biology majors

with little to no experience in bioinformatics. The challenge for instructors is to find ways to acquaint these students with the online tools in a way that both engages student curiosity and confers the confidence and expertise needed to navigate the ever-evolving databases. IMG-ACT is a pipeline for integrating genome annotation projects into undergraduate courses ([3], and series introduction by Kerfeld). In this study, we describe how peer-assisted learning can be used to help faculty communicate the technical and conceptual information needed for students to develop reliable annotation results in IMG-ACT within a single quarter or semester time-frame.

Curriculum Overview

At UCLA, students in a research-based laboratory curriculum work in teams of three to four students to annotate genes in bacterial and bacteriophage genomes. The laboratory program consists of two consecutive 10-week quarters with students immersed in a research project, the goal of which is to

*Address for correspondence to: Erin R. Sanders; E-mail: erins@microbio.ucla.edu.

Received 16 October 2012

Additional Supporting Information may be found in the online version of this article.

DOI: 10.1002/bmb.20665

Published online in Wiley Online Library (wileyonlinelibrary.com)

discover novel bacteriophages. The first quarter (referred to as a *Research Immersion Laboratory*; course MIMG 103AL) offers students hands-on experience isolating bacterial viruses from unique environments such as the skin or soil microbiome, characterizing viral particles using electron microscopy, and sequencing viral genomes using next generation sequencing technology [9,10]. The second quarter (called *Advanced Research Analysis*; course MIMG 103BL) emphasizes rigorous quantitative and computational analysis of data collected during the first quarter. In the case of course 103BL, students utilize bioinformatics tools to investigate viral genome organization and predict functions of virus genes. With a limited amount of time to annotate an entire bacteriophage genome, students in course 103BL need to have already acquired the expertise to select the appropriate bioinformatics tools to reliably verify gene calls and predict gene functions. To achieve this goal, the IMG-ACT platform is employed during course 103AL to provide students bioinformatics training using a bacterial genome (analogous to the host organism for their phage discovery projects). An advantage of using a bacterial genome as the training dataset is that there are ample numbers of genes with a predicted function; these are much more straightforward to annotate than the genes of unknown function or hypothetical proteins comprising up to 80% of a bacteriophage genome. Since a goal in the first course is for students to learn how to use the bioinformatics tools, their training begins with a dataset that generates positive results that can be interpreted and discussed during class—a practical and satisfying outcome that students experience much less frequently with their phage genomes during the second quarter. Importantly, however, this experience gives them context for interpreting and troubleshooting the more ambiguous results associated with subsequent analysis of their bacteriophage genomes.

When Peer-assisted Instruction Meets the Bioinformatics Novice

To create a learner-centered classroom and manage a sizeable group of students participating in genome annotation with the training dataset, we developed a peer-assisted method of instruction [11]. Although the benefits of peer-assisted learning in undergraduate science courses have been documented previously [12–17], its impact on bioinformatics training in an undergraduate course, while not unprecedented [5], has not been systematically studied. Thus, we explored this pedagogical strategy and determined how it might be used effectively in the context of a bioinformatics research project.

Historically, several types of peer-assisted instruction have been incorporated into higher education courses requiring the facilitation of group work [15,18,19]. Among them are near-peers who have completed the course (e.g. teaching assistants and tutors) as well as co-peers who are at an equivalent education level as the other students in the course. In recent decades, the peer learning pedagogy has received the support of national funding agencies like NSF, with the goal of develop-

ing, assessing, and disseminating best practices in undergraduate science education. Three variations of peer-assisted learning—Problem-Based Learning (PBL), Process-Oriented Guided Inquiry Learning (POGIL), and Peer-Led Team Learning (PLTL)—have emerged as means to enhance students' performance (grades) in science courses and persistence (retention, successful performance over time) in science curricula [16]. All three pedagogies are constructivist strategies that echo the national call for science educators to change the way we teach based on knowledge about the way people learn [20,21]. Although particular emphases such as classroom elements, roles of the peer leader, and group size differ for each of these pedagogies, they all share the same goals: (1) to promote higher-order cognitive skills, (2) to foster teamwork and collaborative problem-solving skills, and (3) to transition students from being passive, superficial learners to active, engaged and reflective learners [15,16].

Which peer-assisted learning strategy or combination of strategies an instructor uses typically depends on a number of factors [14–16,22]. Faculty involvement is critical to the successful implementation of peer-led instruction. Not only must faculty be content experts but also be supportive and closely involved in peer learning activities, including peer leader training and the development of instructional materials used in the course. The peer leaders themselves must have knowledge of the subject matter to be conveyed to the larger group. When integrated into undergraduate classes, PBL and PLTL peer leaders are individuals who have successfully completed the course—a fundamental source of their content knowledge (“expertise”)—and have received formal training in learning theory and small group dynamics [12,14,17,22]. With POGIL work groups, students have assigned roles, and the instructor facilitates interactions among group members [16]. To be most effective, peer learning activities need to be mandatory features of a course or at least required for those students who elect to participate at the start of the course (i.e., not offered as drop-in sessions). Instructional materials for peer learning need to be well-conceived and focused on the learner's level of experience [15]. The group assignment itself should be challenging but be comprised of a problem with an unambiguous and coherent solution. Finally, there are organizational considerations that must be taken into account such as the practicalities of the classroom characteristics and peer leader availability. For instance, ideal group size for PLTL should not exceed six to eight students because larger groups tend to change the dynamics of the peer learning environment. However, if the number of classrooms or peer leaders is a limiting factor, then modifications to the instructional strategy (e.g. hybridize different peer learning techniques) can be implemented to account for scale [16].

Implementing a peer-assisted learning pedagogy in bioinformatics at UCLA presented some challenges which ultimately resulted in the development of a hybrid version of this instructional model. First, bioinformatics is an activity that requires a computer and internet connection. So classroom (computer laboratory) availability was limited. Second, the research-

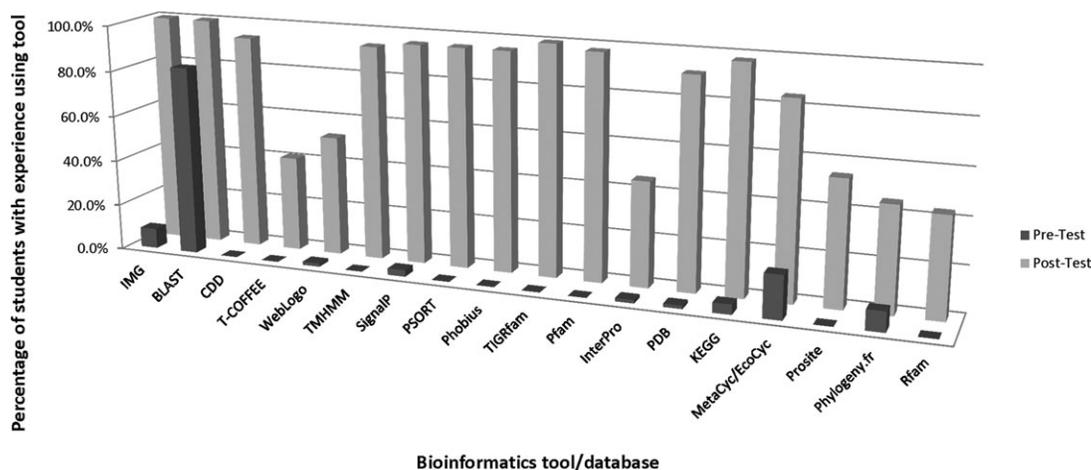


FIG 1

Survey results comparing student familiarity with gene annotation upon entering the course (pre-test) and then again upon completion of the two-quarter course sequence (post-test). The vertical bars in this figure represent the percent of students that claimed to have used each of the bioinformatics tools or databases listed on the horizontal access. For every tool save for BLAST, the percent increase of students who used it was very large.

based laboratory curriculum was introduced at UCLA in fall 2010. Being relatively new courses, there was not a large pool of MIMG 103AL and 103BL alumni who would qualify as suitable peer leaders for small group workshops. Third, the assigned task did not conform to what might be assigned in a strictly problem-solving session characteristic of PBL, POGIL, and PLTL. Specifically, the 103AL class is divided into teams of three to four students each assigned a different set of genes. The research goal is for each team to find and annotate orthologous genes in the bacterial genome using IMG-ACT. By verifying gene calls and reconstructing enzymatic pathways, students link genes to function within a relevant biological context for that organism. This assignment promotes collaborative learning among team members because the end product is the sum of its parts. Individual students annotate one or two genes in the pathway; its reconstruction depends on the cooperative efforts of the group (a sample Annotation Report, which is assembled by the team at the end of the quarter, is provided in the Supporting Information section). Group interdependence develops from this division of labor, and a shared end goal motivates them to interact [23]. The gene annotation assignment is challenging from a technical standpoint because the majority of students have had little to no experience with the databases they will be using throughout the quarter. Students enrolled in MIMG 103AL in fall quarter 2011 and spring quarter 2012 were administered a voluntary, anonymous survey at the beginning and end of the term (see “Methods” for detailed description of survey protocol). In these surveys, they were asked to report their level of knowledge of various tools and databases used in the bioinformatics project. As shown in Fig. 1 (see pre-test responses), with the exception of BLAST, less than 20% of the students had any experience using the tools or databases at the beginning of the term. These results make it clear that students need technical training (“how to

use the tools?”) as well as conceptual guidance (“why use a particular tool?” and “what do the results mean?”).

Comparing Hybrid Pedagogical Strategies

What elements of peer-assisted learning work in our bioinformatics classroom? With the above challenges in mind, we devised two hybrid peer-assisted learning schemes, implementing one approach in fall 2011 and a second in spring 2012. The overarching goal was to determine how to use peer-assisted learning most effectively with a gene annotation project. We conducted focus groups to explore student experiences with the two approaches. We also examined scores on specific course assignments as a quantitative measure of student performance on the annotation project. Table I highlights the shared and unique components of each quarter’s instructional approach with a summary of each below.

Fall 2011 Cohort: Two-tiered Learning Utilizing Co-peers as Annotation Leaders

A class of 32 students was divided into two laboratory sections, each with 16 students and led by a Teaching Assistant (TA). The course instructor held a 75 minute annotation tutorial once per week in a computer lab. A different subset of students (two to three per laboratory section) met with the instructor to study one module in the annotation process each week. There were between two to three bioinformatics tools or databases covered in each module. During the annotation tutorial, the instructor reviewed the PowerPoint presentation containing detailed step-by-step instructions for navigating through each database (see URL provided in Table I). Each presentation contains extensive screenshots for online tools and sample results. During the annotation workshops, students were encouraged to practice annotating their own genes with the instructor, asking questions and clarifying uncertainties about using each database.

TABLE 1

Organizational components of peer-assisted learning strategies

	Fall 2011	Spring 2012
	Two-tiered learning utilizing co-peer leaders	Single-tiered learning using near-peer leaders
Tutorial schedule:	Once per quarter for 75 min in computer classroom.	Once per week for 75 min in computer classroom.
Tutorial leader:	Course instructor.	Course instructor or Teaching Assistant (TA). ^a
Tutorial enrollment:	Four to six students per week with two three representatives from each lab section; student group was different every week.	20–24 students per week; two sections.
Instructional materials:	PowerPoint slides ^b with step-by-step instructions with screenshots for navigating each tool/database, including sample results; organized into weekly modules with two to three tools/databases covered per module.	
Tutorial assignment:	(1) Complete the module for tutorial session attended prior to giving presentation to lab section. (2) Practice group presentation.	n/a
Role of peer leaders:	(1) Give group presentation during lab using same slides as tutorial instructor. (2) Serve as content expert for the module presented; mentor peers in use of tools and interpretation of results.	Serve as content expert for all modules; mentor peers in use of tools and interpretation of results.
Team work:	Complete modules each week either during designated class times ^c or outside of class.	
Compliance check:	Teaching Assistants for laboratory sections check student IMG-ACT notebooks three times per quarter to confirm students are keeping up with the weekly assignments and that the work meets quality standards.	

^a TAs for the tutorials (near-peers) were the same as those in the laboratory sections for the course. TAs were former undergraduate students who had successfully completed the course and had experience with all tools and databases used in the annotation project. TAs also completed a formal pedagogical training course (MIMG 495) and discussed instructional materials with course instructor prior to delivering content or providing project guidance to students.

^b The PowerPoint slides are available on the UCLA website at: <http://www.mimg.ucla.edu/faculty/sanders/education.html>.

^c There was some time for hands-on annotation work immediately following the peer-led tutorial. There also was time allotted during laboratory sections for students to go to computer lab to work with teams on their annotation projects. Students were encouraged to meet with their teams outside of class as well.

After the tutorial, students are expected to complete the module then work with the other tutorial attendee(s) from their laboratory section to prepare a short 15 minute oral presentation to be given in their class the following week using slides provided by the instructor. The goal of this exercise was for the tutorial attendees to become experts on particular modules, teach their fellow MIMG 103AL students the given steps each week, and serve as co-peer mentors during the annotation process throughout the 10-week quarter. Using this two-tiered method of peer-assisted instruction, with tier 1 knowledge transfer from

instructor to co-peer leaders and tier 2 from co-peer leaders to other students in the class, students were expected to learn how to navigate genomic databases, extract and interpret meaningful results, and reconstruct biochemical pathways in different bacteria.

Spring 2012 Cohort: Single-tiered Learning Using Near-peer Annotation Leaders

A class of 43 students was divided into three laboratory sections, each with 13 to 16 students and led by a Teaching



TABLE II

Demographic information for study sample by percent

Term	Gender		URM ^a	Grade level	
	Men	Women		Junior	Senior
Fall 2011 (<i>n</i> = 32)	43.75	56.25	12.5	0.0	100.0
Spring 2012 (<i>n</i> = 43)	37.2	62.8	16.3	69.7	27.9

^a Under-represented Racial Minority (URM) students included Students who self-identified as Latina/o, African American, or American Indian.

Assistant (TA). A 75 minute annotation tutorial was offered once per week in two available computer labs, each with approximately half the class (20–24 students). One tutorial section was led by the course instructor and the other by TAs, former undergraduates who had successfully completed the course and acquired formal pedagogical training through a departmental TA training course (near-peers). During the tutorials, the instructor or near-peers would review the PowerPoint presentations, with step-by-step instructions for the modules. Every student had their own computer workstation and was able to practice annotating their genes with the instructor or near-peers throughout the tutorial. Student teams were expected to work on their annotation assignments during allotted class time or outside of class all through the quarter. This single-tiered method of peer-assisted instruction facilitated transfer of knowledge from the instructor or near-peer leaders to students in the class.

Assessment Methodology

To understand the role of peer interactions on learning gene annotation and evaluate different approaches to delivering the information required to use the bioinformatics tools and databases, we utilized a mixed method design. This strategy consisted of collecting and analyzing both quantitative and qualitative data. As noted by Creswell and Plano Clark [24], quantitative and qualitative methods can work together to provide a more complete picture of the research problem. Three sources of data were utilized in this study: pre/post student surveys, course assignment scores, and two semistructured focus group interviews. Data were collected separately for each of the student groups, fall 2011 and spring 2012, respectively.

Quantitative Design

The quantitative data for this study came from two sources, the IMG-ACT pre/post survey and students' course assignments. The IMG-ACT surveys provide quantitative data illustrating students' familiarity with gene annotation upon entering and completing MIMG 103AL. The surveys were taken anonymously. The pre-survey was an assignment given to all

75 students enrolled in the courses. The post-survey was a not a course assignment in fall 2011 and as such the response rates for this survey differ from the pre-survey. The final analytic sample for this data source included 69 pre-survey responses (30 fall 2011, 39 spring 2012) and 46 post-survey responses (5 fall 2011, 41 spring 2012) yielding a response rate of 92% on the pre-survey and 61% on the post-survey.

Scores from specific assignments related to the gene annotation portion of the course provided another source of quantitative data. In MIMG 103AL, students were asked to complete an annotation notebook as well as write a team annotation report by the end of the term. These assignment scores, along with the combined total of all their annotation assignments, comprise the quantitative data used to understand students' technical and conceptual understanding of the gene annotation project. Scores were compiled for all students who completed the course, which yielded a total sample of 32 students in fall 2011 and 43 students in spring 2012.

Demographic information for the overall sample population for these data sources can be found in Table II.

For the IMG-ACT surveys and course assignment information, descriptive analyses were conducted, providing a basic understanding of students' experiences with gene annotation as well as their overall performance in the class. To find potential differences in their levels of understanding of gene annotation process, three independent sample *t*-tests were conducted using the gene annotation notebook grade, final gene annotation report grade, and the overall combined gene annotation grade.

Qualitative Design

The second phase of data collection consisted of two focus group interviews. These involved students who previously enrolled in MIMG 103AL in fall 2011 or who were currently enrolled in course 103AL in spring 2012. Students in the fall 2011 cohort who were still involved in a gene annotation project were invited to participate in the fall 2011 focus group. All students enrolled in course 103AL in spring 2012 were invited to participate in the spring 2012 focus group. This yielded a student sample for the fall 2011 that was comprised of five women (20% URM) and the spring 2012 sample included three women and four men (28.6% URM). Participants from the fall

TABLE III

Comparison of mean scores for gene annotation assignments

	Fall 2011	Spring 2012	Difference ^a	Significance ^b
Annotation notebook ^c (total possible score = 60 pts)	55.06	57.33	2.27	0.008
Annotation report ^d (total possible score = 50 pts)	44.50	46.70	2.20	0.000
Annotation total score ^e (proportion correct)	0.95	0.96	0.01	0.463

^a Difference between fall 2011 and spring 2012 mean scores for each assignment category.

^b Difference between mean scores for each assignment analyzed using an independent samples *t*-test. The *p*-value cut-off for significance is 0.05. *p*-Values less than 0.05 indicate there is a significant difference between scores of students in the two cohorts whereas *p*-values greater than 0.05 show that there is no significant difference between mean scores.

^c The notebook was evaluated for each individual on a project team. There were three notebook checks conducted by the TAs during the quarter, each worth 20 points each for a total possible score of 60 pts.

^d The report was evaluated per team with contributions by individuals counting towards the total number of points possible for this assignment.

^e In addition to the notebook and report, there were other points students could earn related to the annotation project. In fall 2011, there were up to 80 points awarded for attendance and participation in the annotation tutorial one week per quarter, presenting the PowerPoint slides during lab sections, and evaluating the presentations as feedback for the presenters. In spring 2012, there were up to 50 points awarded for weekly attendance and participation in the annotation tutorial. Because the maximum number of possible points for the annotation assignments was different for each quarter, the proportion of possible points (rather than total number of points) was reported and compared.

2011 group were all graduating seniors, while the seven students from spring 2012 were juniors. These groups were fairly representative of the overall student population in the two courses (see Table II).

The focus group interviews were semistructured and ranged from 40 to 45 minutes. The semistructured strategy was used as it can allow researchers to “delve deeply into a topic and to understand thoroughly the answers provided” while also keeping the style “somewhat conversational” [25]. The focus group interview questions were developed by the authors (see Supporting Information) and consisted of seven general questions focused on students’ experiences with gene annotation and aspects of the course that proved helpful learning the process. The interview protocol also probed more specifics from the IMG-ACT survey on students’ familiarity with gene annotation. Four additional questions were posed to the fall 2011 cohort to understand their experiences with the two-tiered peer instruction strategy used that term. All of the interviews were digitally recorded, transcribed, and then checked for accuracy.

To interpret the focus group interview data, a coding scheme was developed by examining the transcriptions and organizing the material into meaningful segments [26] based on the initial focus group protocol and preliminary data analyses from the IMG-ACT survey. These segments consisted of students’ direct quotes, which allowed for the students’ voices to emerge and for direct interpretation of the data. Themes were identified upon inspection of these segments in conjunction with the quantitative data. These themes, which will be discussed in more detail in the results section, are the perspectives of multiple individuals and are supported by the text.

Results

Pre/Post Surveys: Students Report Increased Levels of Experience Using Bioinformatics Tools

As discussed above, students in MIMG 103AL started the term with little to no experience with most bioinformatics tools (see Fig. 1). In comparison, by the end of the term, students reported large increases in their level of experience with each database. The annotation project visibly succeeded in exposing students to a wide range of bioinformatics tools that they otherwise may never have used.

Analysis of Scores on Annotation Assignments Suggests There is No Difference in Quality of Student Work Between Quarters

There were two assignments given in MIMG 103AL each quarter related specifically to the gene annotation project. Every student was assigned either two or three genes to annotate and was asked to maintain a digital notebook throughout the quarter, recording the results from their database searches each week for all of their assigned genes. Student teams were required to submit a final report comprised of individual gene annotation summaries (IMG-ACT online reports) and a collective evaluation of the team results for each gene in a biochemical pathway, justifying (or contradicting) the accuracy of the automated gene call (see sample report in Supporting Information). Reported in Table III are the mean scores by quarter for both assignments, the difference between mean scores on these two assignments for the fall and spring quarters, and the results of an independent sample *t*-test for each grade category.



The *t*-test results reveal significant differences between the scores of students in the fall 2011 and spring 2012 cohorts on the annotation notebook ($p = 0.008$; with a 2.27 point mean difference between the two groups) and the annotation report ($p = 0.000$; with a 2.20 point mean difference between the two groups). Given that the mean scores for these assignments differ by less than 2.5 points, it is important to interpret these differences with caution. In particular, while the means are statistically different, practically speaking these differences could be attributed to noise in the dataset. For example, as shown in Table II, the two cohorts are comprised of students at different grade levels, with fall 2011 being made up entirely of seniors and spring 2012 reflecting a mixture of juniors and seniors. The differences in scores could be impacted by demographic dissimilarities with seniors potentially less engaged or enthusiastic about the project than juniors [27]. Additionally, the two cohorts were taught by different instructors; thus, performance evaluation criteria may have been somewhat different between the two cohorts. The significant differences observed with the notebooks and the reports disappear when other annotation related activities for which points were awarded are taken into account. As shown in the last row of Table III, there is no significant difference between mean total scores ($p = 0.463$). Taken together, these results indicate that, despite changing the peer-assisted instructional approach from one quarter to the next, there was little to no difference in overall levels of technical and conceptual understanding of gene annotation by students in fall 2011 compared with those in spring 2012 at least with respect to their performance on graded, medium to low stake assignments.

Deconstructing Focus Group Responses Provides Insights into Student Experience

The resulting five themes that emerged from the analysis of the focus group discussions are shown in Table IV. Interview responses revealed substantial insights about students' overall experience with genome annotation tools (databases), identified critical instructional materials for bioinformatics work, showed the extent to which students appreciated a training exercise and understood its application to their bacteriophage research project, and unveiled essential characteristics of an effective peer leader.

Consistent with the IMG-ACT pre-survey findings, very few students in the focus groups reported they had any familiarity with gene annotation at the start of the term. For instance, some had heard of BLAST, but had never used it. Additionally, as reported in the IMG-ACT post-survey results, the majority of students felt they became more comfortable with the databases over time. Notably, students in both focus groups spoke about the importance of the PowerPoint slides in helping them learn and become more comfortable with the material. All students felt that the slides were crucial to their understanding about how the databases worked (e.g. bioinformatics technique—the “how”). Students in the spring cohort also stated that having access to individual computer workstations during

the tutorial provided an opportunity for them to have hands-on experience with the databases, thereby creating an active (versus passive) learning environment.

Perhaps surprisingly, students in the fall cohort felt they did not acquire enough information from their co-peer leaders, TAs, or course instructor about why they were using such a diverse array of bioinformatics tools. While they understood that each database produced different types of results related to protein function, students did not appreciate how to connect the database function with the scientific question they were trying to answer (conceptual knowledge). For instance, if a student finds a gene with amino acid sequence homology to a transporter by BLAST search, then s/he should predict that the protein contains transmembrane helices and localizes to a membrane, testing their hypothesis using tools such as TMHMM and PSORT, respectively [28–30]. Furthermore, understanding solely how to do the data analysis (technical knowledge) was not a constructive learning experience for students; they also demanded context and relevance for their computational tasks (a.k.a., the “big picture”). For example, students in fall described a disconnect between the bacteriophage discovery project in the wet laboratory and the genome annotation project in the computer lab. Students did not appear to understand how the tools they were learning in MIMG 103AL could be applied to bioinformatics investigations with their own phage genomes in MIMG 103BL. These two shortcomings—lack of conceptual knowledge and ‘big picture’ understanding—were not echoed by students in the spring focus group. Instead, they felt the near-peer leaders (TAs who had completed the course previously) and the instructor (not the same faculty member who trained the fall cohort) imparted both the “how” and the “why” during weekly tutorial discussions. This finding suggests the expertise near-peer leaders bring to the classroom is an important aspect to creating a positive learning experience with genome annotation projects.

Interestingly, students in both cohorts reported feeling that a conceptual understanding of genome annotation is important for student engagement in the bioinformatics component of their research projects—that is, getting students excited and motivated to do bacteriophage genome annotation in MIMG 103BL. As a student in the fall cohort noted, “...I felt really comfortable with the program and I had a full grasp of it. And it was kind of exciting finding functions of genes.” Conceptual and technical knowledge of the bioinformatics project, whether acquired in MIMG 103AL or 103BL, enabled students to make connections that also promoted project ownership, or pride, in their work. One student in the spring cohort indicated that he became more excited about the project over time, noting the instructor showed the class “pictures of how these things would look like, [making] it... interesting.” Other students in the spring cohort spoke about how much they appreciated the chance to connect what they learned in other classes to something “you can do... it was pretty awesome” (student in spring cohort).

Finally, the focus group discussions helped uncover which aspects of peer-assisted instruction were critical for successful

TABLE IV

Focus group interview results by theme and term

Outcomes	Fall 2011	Spring 2012
1. Initial familiarity with genome annotation ("what")	Majority not familiar with gene annotation.	Majority not familiar with gene annotation.
2. Gaining a technical understanding of genome annotation ("how")	PowerPoint slides crucial for technical (step-by-step) understanding.	PowerPoint slides crucial for technical (step-by-step) understanding. Access to individual workstation facilitated technical understanding.
3. Gaining a conceptual understanding of genome annotation ("why")	Students wanted more "big picture" information from co-peer mentors and the instructor.	Students gained more "big picture" understanding of when and why to use specific databases from tutorials given by the instructor or near-peer leaders.
4. Engagement in material and pride in work ("when")	Half of the students wanted to continue to learn about annotation once they felt they had a good understanding of the material. Half wanted to continue to learn about gene annotation because they were excited about finding out more about their phages.	The majority of students discussed their interest in the project and how this class was very informative. The majority of students also said they felt excitement and pride in their work, which encouraged them to speak to peers about the course or show their work to their peers.
5. Who is the "expert"? ("who")	Peer mentor instruction acquired during tutorials did not make those students an "expert" on the material. Instead, the student(s) who had already successfully completed the assignment was considered the "expert." In general, co-peers were helpful, but primarily for technical help navigating the databases.	Near-peer leaders who had taken the course and were trained on teaching pedagogies were viewed as peer "experts." Students felt comfortable seeking the near-peer leader's assistance mostly for technical questions, but also for a conceptual understanding of material.

dissemination of the genome annotation process. Recall that the fall cohort experienced a two-tiered learning approach (Table I) in which co-peers became the leaders responsible for teaching fellow students the "how" and "why" of genome annotation. Scores on annotation assignments indicate students achieved these goals with a mean of over 90% on the annotation notebook and 89% on the final report (Table III). However, after discussing their experience with this instructional strategy with the focus group participants from the fall cohort, we learned that it was not necessarily the presenters who provided the expertise needed to help students navigate the databases and interpret the results. Instead, those who had already successfully completed a module (had the experience) emerged as co-peer leaders in the course. As one student said, expertise came from "...whether or not you did it, not whether you taught it." These experienced students, now

considered the "experts" in their laboratory section, were most useful for technical (the "how"), rather than conceptual (the "why"), support. Consequently, conceptual knowledge was not obviously attained until students were well into the second course, MIMG 103BL, working on their own bacteriophage genomes. As one student in the fall cohort described, "by the end of the quarter [MIMG 103BL] I figured it out. And then we were able to fix our phage... and seeing it through to the end... was great." This finding underscores two major points. First, the annotation assignments in MIMG 103AL may only be measuring technical achievements and scores do not reflect student conceptual understanding of gene annotation. One student from fall 2011 substantiated this sentiment about the notebooks during the focus group, stating "I would just get through it basically and have no real understanding". Second, co-peer leaders can successfully disseminate the technical



information needed to help students complete the annotation tasks and were viewed as “lifelines for understanding” (student in fall cohort), but their level of expertise (and thus value to other students in the class) relied on experience using the databases, not simply attending the tutorials with the instructor or presenting the PowerPoint slides. Of note, one of the tutorial assignments in fall 2011 was for students to complete the module prior to giving their presentations to the laboratory sections. However, we learned that this requirement was not always met and so diminished the designated co-peer leader’s role in instruction.

Our experiences with the fall cohort informed changes to the peer learning approach for the spring cohort. Specifically, in spring we used a single-tiered learning approach in which near-peer leaders met with the same group of students each week for the tutorial session (Table I). These meetings took place in a computer lab so individual students could practice using the online tools as the near-peer leader was going through the material with the class. One disadvantage to this strategy was that the group size approached that of a discussion section (recitation), so there was concern that the dynamics of the experience would change, potentially discouraging team work or promoting passive learning. However, student responses in the focus group from spring quarter suggest otherwise, confirming that they viewed their near-peer leaders as “experts” and felt comfortable seeking their help with both technical and conceptual questions related to genome annotation. Interestingly, while students admitted that they still felt a bit like novices at the end of MIMG 103AL, they also indicated that they thought the annotation project prepared them for their MIMG 103BL projects to be undertaken the next quarter, that they had developed a commitment and desire to understand more.

Conclusions and Future Directions

Taken together, the findings of this study have shaped how we move forward using peer learning effectively to teach undergraduates how to do genome annotation. Critical ingredients for successfully carrying out this instructional methodology in a bioinformatics classroom include providing students technical understanding of the databases (the “how”) as well as conceptual knowledge about the overall annotation process (the “why”). Near-peer leaders seem to be more effective than co-peers at communicating the conceptual knowledge and the “big picture” to students, the latter of which is critical for fostering excitement about their research projects and getting them engaged in the bioinformatics components in particular. The students valued having detailed instructions for using the databases provided in the PowerPoint slides. In fact, they were viewed as the most important tools in the course. As one student puts it, “They take you step-by-step so there is no question in what you’re supposed to do.” Technical understanding was also enhanced by having the tutorial sessions in a computer lab where students had access to individual work stations. Furthermore, this environ-

ment was conducive to active learning since students could have a hands-on experience with near-peer leaders.

Evaluating the hybrid peer-assisted learning approach in the two courses offered us a unique lens by which to decide how best to continue bioinformatics training as an important feature of the new laboratory curriculum. We will retain TAs as near-peer leaders for the tutorials. And while computer lab availability may limit our ability to decrease the class size, we will recruit more undergraduate alumni to assist near-peers with tutorial instruction during MIMG 103AL. These undergraduate assistants can facilitate mentoring of smaller groups within the larger class, helping to transform the tutorials into problem-solving sessions for teams. These pedagogical interventions should enable us to achieve the three primary goals of peer-assisted learning in bioinformatics: Promoting higher-order cognitive skills (conceptual, not just technical, understanding of genome annotation), fostering collaborative problem-solving (team-based, hypothesis-driven genome annotation projects), and creating an active learning environment (hands-on experience with computation).

Acknowledgements

The authors thank Lori Scott at Augustana College for her assistance collecting IMG-ACT survey data. They also thank Cheryl Kerfeld at the Joint Genome Institute for facilitating the student genome annotation project using IMG-ACT and for encouraging their participation in the BAMBED laboratory series. Special thanks to the undergraduate members of the honors course who assisted with development of the PowerPoint tutorials: Catherine Lee, Vivian Lau, Sarah Dodge, Anita Viswanath, Natalya Watson, Joan Khuu, Caroline Keegan, Eric Dugum, Johnny Wu, and Diana Lee. Data for only UCLA students was presented in this study. Approval by UCLA Institutional Review Board was obtained for focus group work and all existing data, including IMG-ACT survey results for UCLA (IRB #12-000713).

References

- [1] S. Leininger, T. Urich, M. Schloter, L. Schwark, J. Qi, G. W. Nicol, J. I. Prosser, S. C. Schuster, C. Schleper. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809.
- [2] D. J. Baumler, L. M. Banta, K. F. Hung, J. A. Schwarz, E. L. Cabot, J. D. Glasner, N. T. Perna (2012) Using comparative genomics for inquiry-based learning to dissect virulence of *Escherichia coli* O157:H7 and *Yersinia pestis*. *CBE Life Sci. Educ.* 11, 81–93.
- [3] J. L. Ditty, C. A. Kvaal, B. Goodner, S. K. Freyermuth, C. Bailey, R. A. Britton, S. G. Gordon, S. Heinhorst, K. Reed, Z. Xu, E. R. Sanders-Lorenz, S. Axen, E. Kim, M. Johns, K. Scott, C. A. Kerfeld. (2010) Incorporating genomics and bioinformatics across the life sciences curriculum. *PLoS Biol.* 8, e1000448.
- [4] C. D. Shaffer, C. Alvarez, C. Bailey, D. Barnard, S. Bhalla, C. Chandrasekaran, V. Chandrasekaran, H. Chung, D. R. Dorer, C. Du, T. T. Eckdahl, J. L. Poet, D. Frohlich, A. L. Goodman, Y. Gosser, C. Hauser, L.L.M. Hoopes, D. Johnson, C. J. Jones, M. Kaehler, N. Kokan, O. R. Kopp, G. A. Kuleck, G. McNeil, R. Moss, J. L. Myka, A. Nagengast, R. Morris, P. J. Overvoorde, E. Shoop, S. Parrish, K. Reed, E. G. Regisford, D. Revie, A. G. Rosenwald, K. Saville, S. Schroeder, M. Shaw, G. Skuse, C. Smith, M. Smith, E. P. Spana, M. Spratt, J.

- Stamm, J. S. Thompson, M. Wawersik, B. A. Wilson, J. Youngblom, W. Leung, J. Buhler, E. R. Mardis, D. Lopatto, S.C.R. Elgin. (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.
- [5] C. A. Kerfeld, R. W. Simons (2007) The undergraduate genomics research initiative. *PLoS Biol.* 5, e141.
- [6] K. M. Scott, S. M. Sievert, F. i N. Abril, L. A. Ball, C. J. Barrett, R. A. Blake, A. J. Boller, P. S. G. Chain, J. A. Clark, C. R. Davis, C. Detter, K. F. Do, K. P. Dobrinski, B. I. Faza, K. A. Fitzpatrick, S. K. Freyermuth, T. L. Harmer, L. J. Hauser, M. Hügler, C. A. Kerfeld, M. G. Klotz, W. W. Kong, M. Land, A. Lapidus, F. W. Larimer, D. L. Longo, S. Lucas, S. A. Malfatti, S. E. Massey, D. D. Martin, Z. McCuddin, F. Meyer, J. L. Moore, L. H. Ocampo Jr., J. H. Paul, I. T. Paulsen, D. K. Reep, Q. Ren, R. L. Ross, P. Y. Sato, P. Thomas, L. E. Tinkham, G. T. Zeruth. (2006) The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* Xcl-2. *PLoS Biol.* 4, e383.
- [7] J. Chen, G.B. Call, E. Beyer, C. Bui, A. Cespedes, A. Chan, J. Chan, S. Chan, A. Chhabra, P. Dang, A. Deravanesian, B. Hermogeno, J. Jen, E. Kim, E. Lee, G. Lewis, J. Marshall, K. Regalia, F. Shadpour, A. Shemmassian, K. Spivey, M. Wells, J. Wu, Y. Yamauchi, A. Yavari, A. Abrams, A. Abramson, L. Amado, J. Anderson, K. Bashour, E. Bibikova, A. Bookatz, S. Brewer, N. Buu, S. Calvillo, J. Cao, A. Chang, D. Chang, Y. Chang, Y. Chen, J. Choi, J. Chou, S. Datta, A. Davarifar, P. Desai, J. Fabrikant, S. Farnad, K. Fu, E. Garcia, N. Garrone, S. Gasparyan, P. Gayda, C. Goffstein, C. Gonzalez, M. Guirguis, R. Hassid, A. Hong, J. Hong, L. Hovestreydt, C. Hu, F. Jamshidian, K. Kahen, L. Kao, M. Kelley, T. Kho, S. Kim, Y. Kim, B. Kirkpatrick, E. Kohan, R. Kwak, A. Langenbacher, S. Laxamana, C. Lee, J. Lee, S.Y. Lee, T.H. Lee, T. Lee, S. Lezcano, H. Lin, P. Lin, J. Luu, T. Luu, W. Marrs, E. Marsh, S. Min, T. Minasian, A. Misra, M. Morimoto, Y. Moshfegh, J. Murray, C. Nguyen, K. Nguyen, E. Nodado 2nd, A. O'Donahue, N. Onugha, N. Orjiakor, B. Padhiar, M. Pavel-Dinu, A. Pavlenko, E. Paz, S. Phaklides, L. Pham, P. Poulouse, R. Powell, A. Pusic, D. Ramola, M. Ribbens, B. Rifai, D. Rosselli, M. Saakyan, P. Saarikoski, M. Segura, R. Singh, V. Singh, E. Skinner, D. Solomin, K. Soneji, E. Stageberg, M. Stavchanskiy, L. Tekchandani, L. Thai, J. Thiyanaratnam, M. Tong, A. Toor, S. Tovar, K. Trangsrud, W.Y. Tsang, M. Uemura, M. Unkovic, E. Vollmer, E. Weiss, D. Wood, S. Wu, W. Wu, Q. Xu, K. Yackle, W. Yarosh, L. Yee, G. Yen, G. Alkin, S. Go, D. M. Huff, H. Minye, E. Paul, N. Villarasa, A. Milchanowski, U. Banerjee. (2005) Discovery-based science education: Functional genomic dissection in drosophila by undergraduate researchers. *PLoS Biol.* 3, e59.
- [8] S. Forst, B. Goodner (2005) Comparative bacterial genomics and its use in undergraduate education. *Biol. Control* 38, 47–53.
- [9] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Goodwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L.I. Alenquer, T. P. Jarvie, K. B. Jirage, J. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. Fredrik Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, J. M. Rothberg. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.
- [10] S. C. Schuster (2008) Next-generation sequencing transforms today's biology. *Nat. Methods* 5, 16–18.
- [11] J. J. Velez, K. J. Wolf, J. Cano (2007) Student perceived benefits of peer-assisted instruction. *Proc. AAAE Res. Conf.* 34, 731–733.
- [12] C. C. Wamser (2006) Peer-Led team learning in organic chemistry: Effects on student performance, success, and persistence in the course. *J. Chem. Educ.* 83, 1562–1566.
- [13] D. K. Grosser, V. Roth (1998) The workshop chemistry project: Peer-led team learning. *J. Chem. Educ.* 75, 185–187.
- [14] I. J. Quitadamo, C. J. Brahler, G. J. Crouch (2009) Peer-led team learning: A prospective method for increasing critical thinking in undergraduate science courses. *Sci. Educ.* 18, 29–39.
- [15] M. Micari, G. Light (2009) Reliance to independence: Approaches to learning in peer-led undergraduate science, technology, engineering, and mathematics workshops. *Int. J. Sci. Educ.* 31, 1713–1741.
- [16] T. Eberlein, J. Kampmeier, V. Minderhout, R. S. Moog, T. Platt, V.-P. Nelson, H. B. White (2008) Pedagogies of engagement in science: A comparison of PBL, POGIL, and PLTL. *Biochem. Mol. Biol. Educ.* 36, 262–273.
- [17] T. Platt, E. Barber, A. Yoshinaka, V. Roth (2003) An innovative selection and training program for problem-based learning (PBL) workshop leaders in biochemistry. *Biochem. Mol. Biol. Educ.* 31, 132–136.
- [18] B. Goldschmid, M. L. Goldschmid (1976) Peer teaching in higher education: A review. *Higher Educ.* 5, 9–33.
- [19] N. A. Whitman (1988) *Peer Teaching: To Teach is to Learn Twice*, Education, Ashe-Eric Higher, Washington, DC.
- [20] National Research Council (2000) *How People Learn: Brain, Mind, Experience, and School*, The National Academies Press, Washington, DC.
- [21] National Research Council (2003) *Bio 2010: Transforming Undergraduate Education for Future Research Biologists*, The National Academies Press, Washington, DC.
- [22] L. Gafney, Varma-Nelson P (2008) Development and Implementation of Peer-Led Team Learning, in Karen C. Cohen. (ed.) *Peer-Led Team Learning: Evaluation, Dissemination, and Institutionalization of a College Level Initiative*, Springer Science Plus Business Media, BV.
- [23] E. G. Cohen (1994) Restructuring the classroom: Conditions for productive small groups. *Rev. Educ. Res.* 64, 1–35.
- [24] J. W. Creswell, P. V. L. Clark (2011) *Designing and Conducting Mixed Methods Research*, Sage Publications, Thousand Oaks, CA.
- [25] M. C. Harrell, M. A. Bradley (2009) *Data-Collection Methods: Semi-Structured Interviews and Focus Groups*, RAND Corp, Santa Monica, CA.
- [26] J. W. Creswell (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd ed., Sage Publications, Thousand Oaks, CA.
- [27] B. Valentine (2001) The legitimate effort in research papers: Student commitment versus faculty expectations. *J. Acad. Librarianship* 27, 107–115.
- [28] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- [29] K. Nakai, P. Horton (1999) PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24, 34–36.
- [30] E. L. Sonnhammer, V. G. Heijne, A. Krogh (1998) A hidden markov model for predicting transmembrane helices in protein sequences. *Int. Conf. Intelligent Syst. Mol. Biol.* 6, 175–182.