City University of New York (CUNY)
CUNY Academic Works

Publications and Research | City College of New York

2014

A Course-Based Research Experience: How Benefits Change with Increased Investment in Instructional Time

Christopher D. Shaffer
Washington University in St. Louis

Consuelo J. Alvarez
Longwood University

April E. Bednarski
Lindenwood University

David Dunbar
Cabrini College

Anya L. Goodman
California Polytechnic State University

See next page for additional authors

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/cc_pubs/11
Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

Authors

Christopher D. Shaffer, Consuelo J. Alvarez, April E. Bednarski, David Dunbar, Anya L. Goodman, Catherine Reinke, Anne G. Rosenwald, Michael J. Wolyniak, Cheryl Bailey, Daron Barnard, Christopher Bazinet, Dale L. Beach, James E. J. Bedard, Satish Bhalla, John Braverman, Martin Burg, Vidya Chandrasekaran, Hui-Min Chung, Kari Clase, Randall J. DeJong, Shubha Govind, Gerard McNeil, and Karim Sharif

This article is available at CUNY Academic Works: https://academicworks.cuny.edu/cc_pubs/11

CBE—Life Sciences Education
Vol. 13, 111–130, Spring 2014

Article

A Course-Based Research Experience: How Benefits Change with Increased Investment in Instructional Time

Christopher D. Shaffer,[1] Consuelo J. Alvarez,[2] April E. Bednarski,[3,4] David Dunbar,[5] Anya L. Goodman,[6] Catherine Reinke,[7] Anne G. Rosenwald,[8] Michael J. Wolyniak,[9] Cheryl Bailey,[10,11] Daron Barnard,[12] Christopher Bazinet,[13] Dale L. Beach,[2] James E. J. Bedard,[14,15] Satish Bhalla,[16] John Braverman,[17] Martin Burg,[18] Vidya Chandrasekaran,[19] Hui-Min Chung,[20] Kari Clase,[21] Randall J. DeJong,[22] Justin R. DiAngelo,[23] Chunguang Du,[24] Todd T. Eckdahl,[25] Heather Eisler,[26] Julia A. Emerson,[27] Amy Frary,[28] Donald Frohlich,[29] Yuying Gosser,[30] Shubha Govind,[31] Adam Haberman,[32,33] Amy T. Hark,[34] Charles Hauser,[35] Arlene Hoogewerf,[22] Laura L. M. Hoopes,[36] Carina E. Howell,[37] Diana Johnson,[38] Christopher J. Jones,[39] Lisa Kadlec,[40] Marian Kaehler,[41] S. Catherine Silver Key,[42] Adam Kleinschmit,[14] Nighat P. Kokan,[43] Olga Kopp,[44] Gary Kuleck,[45,46] Judith Leatherman,[47] Jane Lopilato,[48] Christy MacKinnon,[49] Juan Carlos Martinez-Cruzado,[50] Gerard McNeil,[51] Stephanie Mel,[52] Hemlata Mistry,[53] Alexis Nagengast,[54] Paul Overvoorde,[55] Don W. Paetkau,[56] Susan Parrish,[57] Celeste N. Peterson,[58] Mary Preuss,[59] Laura K. Reed,[60] Dennis Revie,[61] Srebrenka Robic,[62] Jennifer Roecklein-Canfield,[63] Michael R. Rubin,[64] Kenneth Saville,[65] Stephanie Schroeder,[66] Karim Sharif,[67,68] Mary Shaw,[69] Gary Skuse,[70] Christopher D. Smith,[71] Mary A. Smith,[72] Sheryl T. Smith,[73] Eric Spana,[74] Mary Spratt,[75] Aparna Sreenivasan,[76] Joyce Stamm,[77] Paul Szauter,[78] Jeffrey S. Thompson,[79] Matthew Wawersik,[80] James Youngblom,[81] Leming Zhou,[82] Elaine R. Mardis,[83] Jeremy Buhler,[84] Wilson Leung,[1] David Lopatto,[85] and Sarah C. R. Elgin[1]

[1]Department of Biology and [84]Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130; [2]Department of Biological and Environmental Sciences, Longwood University, Farmville, VA 23909; [3]Chemistry Department, Lindenwood University, St. Charles, MO 63301; [5]Science Department, Cabrini College, Radnor, PA 19087; [6]Department of Chemistry and Biochemistry, California Polytechnic State University, San Luis Obispo, CA 93405; [7]Department of Biology, Linfield College, McMinnville, OR 97128; [8]Department of Biology, Georgetown University, Washington, DC 20057; [9]Biology Department,

DOI: 10.1187/cbe-13-08-0152
Address correspondence to: Christopher D. Shaffer (shaffer@biology.wustl.edu).

Present addresses: [4]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; [11]Howard Hughes Medical Institute, Chevy Chase, MD 20815; [15]Department of Biology, University of the Fraser Valley, Abbotsford, BC V2S 7M8, Canada; [33]Department of Biology, University of San Diego, San Diego, CA 92110; [46]College of Engineering & Science, University of Detroit Mercy, Detroit, MI 48221; [68]Biology Department, Massasoit Community College, Brockton, MA 02302.

© 2014 C. D. Shaffer *et al*. *CBE—Life Sciences Education* © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution—Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

Hampden-Sydney College, Hampden-Sydney, VA 23943; [10]Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, NE 68588; [12]Biology Department, Worcester State University, Worcester, MA 01602; [13]Department of Biological Sciences, St. John's University, Queens, NY 11439; [14]Department of Biology, Adams State University, Alamosa, CO 81101; [16]Department of Computer Science & Engineering, Johnson C. Smith University, Charlotte, NC 28216; [17]Department of Biology, Saint Joseph's University, Philadelphia, PA 19131; [18]Departments of Biomedical Sciences & Cell and Molecular Biology, Grand Valley State, Allendale, MI 49401; [19]Department of Biology, Saint Mary's College of California, Moraga, CA 94556; [20]Department of Biology, University of West Florida, Pensacola, FL 32514; [21]Technology Leadership & Innovation Department, Purdue University, West Lafayette, IN 47907; [22]Department of Biology, Calvin College, Grand Rapids, MI 49546; [23]Department of Biology, Hofstra University, Hempstead, NY 11549; [24]Department of Biology & Molecular Biology, Montclair State University, Montclair, NJ 07043; [25]Department of Biology, Missouri Western State University, St. Joseph, MO 64507; [26]Department of Biology, University of the Cumberlands, Williamsburg, KY 40769; [27]Department of Biology, Amherst College, Amherst, MA 01002; [28]Department of Biological Sciences, Mount Holyoke, South Hadley, MA 01075; [29]Biology Department, University of St. Thomas, Houston, TX 77006; [30]Grove School of Engineering and [31]Biology Department, City College of New York, New York, NY 10031; [32]Biology Department, Oberlin College, Oberlin, OH 44074; [34]Biology Department, Muhlenberg College, Allentown, PA 18104; [35]Department of Bioinformatics, St. Edwards University, Austin, TX 78704; [36]Department of Biology, Pomona College, Claremont, CA 91711; [37]Department of Biological Sciences, Lock Haven University of Pennsylvania, Lock Haven, PA 17745; [38]Department of Biological Sciences, George Washington University, Washington, DC 20052; [39]Department of Biological Sciences, Moravian College, Bethlehem, PA 18018; [40]Department of Biology, Wilkes University, Wilkes-Barre, PA 18701; [41]Biology Department, Luther College, Decorah, IA 52101; [42]Biology Department, North Carolina Central University, Durham, NC 27707; [43]Department of Natural Sciences, Cardinal Stritch University, Milwaukee, WI 53217; [44]Department of Biology, Utah Valley University, Orem, UT 84058; [45]Department of Biology, Loyola Marymount University, Los Angeles, CA 90045; [47]Department of Biological Sciences, University of Northern Colorado, Greeley, CO 80639; [48]Biology Department and [63]Department of Chemistry, Simmons College, Boston, MA 02115; [49]Biology Department, University of the Incarnate Word, San Antonio, TX 78209; [50]Department of Biology, University of Puerto Rico at Mayaguez, Mayaguez, PR 00680; [51]Department of Biology, York College–City University of New York, Jamaica, NY 11451; [52]Division of Biological Sciences, University of California–San Diego, La Jolla, CA 92093; [53]Department of Biology and [54]Departments of Chemistry and Biochemistry, Widener University, Chester, PA 19013; [55]Biology Department, Macalester College, St. Paul, MN 55105; [56]Department of Biology, Saint Mary's College, Notre Dame, IN 46556; [57]Biology Department, McDaniel College, Westminster, MD 21157; [58]Department of Biology, Suffolk University, Boston, MA 02108; [59]Department of Biological Sciences, Webster University, Webster Groves, MO 63119; [60]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35401; [61]Department of Biology, California Lutheran University, Thousand Oaks, CA 91360; [62]Department of Biology, Agnes Scott College, Decatur, GA 30030; [64]Department of Biology, University of Puerto Rico at Cayey, Cayey, PR 00736; [65]Department of Biology, Albion College, Albion, MI 49224; [66]Department of Biology, Webster University, St. Louis, MO 63119; [67]Department of Natural Sciences, LaGuardia Community College, Long Island City, NY 11101; [69]Department of Biology, New Mexico Highlands University, Las Vegas, NM 87701; [70]Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY 14623; [71]Department of Biology, San Francisco State University, San Francisco, CA 94132; [72]Department of Biology, North Carolina A&T State University, Greensboro, NC 27411; [73]Department of Biology, Arcadia University, Glenside, PA 19038; [74]Department of Biology, Duke University, Durham, NC 27708; [75]Department of Biology, William Woods University, Fulton, MO 65251; [76]Science and Environmental Policy, California State University–Monterey Bay, Seaside, CA 93955; [77]Department of Biology, University of Evansville, Evansville, IN 47722; [78]Biology Department, University of New Mexico, Albuquerque, NM 87106; [79]Department of Biology, Denison University, Granville, OH 43023; [80]Department of Biology, College of William & Mary, Williamsburg, VA 23187; [81]Department of Biology, California State University–Stanislaus, Turlock, CA 95382; [82]Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA 15213; [83]Genome Institute, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108; [85]Department of Psychology, Grinnell College, Grinnell, IA 50112

Submitted August 21, 2013; Revised October 21, 2013; Accepted October 30, 2013
Monitoring Editor: Diane K. O'Dowd

There is widespread agreement that science, technology, engineering, and mathematics programs should provide undergraduates with research experience. Practical issues and limited resources, however, make this a challenge. We have developed a bioinformatics project that provides a course-based research experience for students at a diverse group of schools and offers the opportunity to tailor this experience to local curriculum and institution-specific student needs. We assessed both attitude and knowledge gains, looking for insights into how students respond given this wide range of curricular and institutional variables. While different approaches all appear to result in learning gains, we find that a significant investment of course time is required to enable students to show gains commensurate to a summer research experience. An alumni survey revealed that time spent on a research project is also a significant factor in the value former students assign to the experience one or more years later. We conclude: 1) implementation of a bioinformatics project within the biology curriculum provides a mechanism for successfully engaging large numbers of students in undergraduate research; 2) benefits to students are achievable at a wide variety of academic institutions; and 3) successful implementation of course-based research experiences requires significant investment of instructional time for students to gain full benefit.

INTRODUCTION

A growing body of literature has established the benefits of research experiences for undergraduate students in the sciences (Seymour *et al.*, 2004; Lopatto, 2006, 2009; Laursen *et al.*, 2010). Indeed, integration of research experiences into the academic-year curriculum, along with the use of other active-learning strategies, is a central theme in calls for undergraduate biology education reform. (See *BIO2010: Transforming Undergraduate Education for Future Research Biologists* [National Research Council (NRC), 2003] and *Vision and Change in Undergraduate Biology Education: A Call to Action* [American Association for the Advancement of Science, 2011].) Such experiences can be particularly beneficial for first-generation students, underrepresented minorities, and at-risk students, significantly improving retention rates in the sciences for these groups (Nagda *et al.*, 1998; Hathaway *et al.*, 2002; Lopatto, 2006, 2007; Locks and Gregerman, 2008; Goins *et al.*, 2009). Perhaps as a consequence, the second recommendation in the recent President's Council of Advisors on Science and Technology (PCAST) report, *Engage to Excel* (2012, pp. 16, 25, and 38), is to "advocate and provide support for replacing standard laboratory courses with discovery-based research courses." Utilizing results from a diverse group of institutions, the growth in the Genomics Education Partnership (GEP) provides us with the opportunity to examine key features of undergraduate instruction that can contribute to student gains from a research experience embedded in an academic-year class (Shaffer *et al.*, 2010). Importantly, we find that these results are independent of institution characteristics. Rather, the key variable is the length of time spent on the project.

Genomics studies as a focus for undergraduate research provide many opportunities, and are especially useful for teaching institutions limited by minimal research infrastructure and budgetary support. First, there is a huge amount of raw genomics data, most archived and publicly available (National Center for Biotechnology Information [NCBI], www.ncbi.nlm.nih.gov). Most of this sequence data has been analyzed only by individual prediction programs at the time of posting, providing ample opportunities for undergraduates to improve this analysis and carry out their own investigations mining information from genomes. "Turning data into knowledge" (Brenner, 2002) is in fact the major bottleneck in genomics today, and a large-scale undergraduate project in sequence improvement and analysis results in new understandings and improved data sets that can benefit the research community as a whole (Boomer *et al.*, 2002; Elwess and Latourelle, 2004; Drew and Triplett, 2008). Further, with DNA sequencing continuously becoming less expensive, we can anticipate many more institutions initiating genome-sequencing or RNA-sequencing projects in the not-too-distant future, allowing students to explore local ecosystems and characterize genomes of local organisms (e.g., see Oleksyk *et al.*, 2012). In addition, the "tools of the trade" in genomics and bioinformatics are generally publicly available (e.g., at the NCBI) and lend themselves well to peer instruction. Today's students are often very adept at using computers, and students who are familiar with the computer-based tools of bioinformatics from previous experience prove to be effective coaches for newcomers, serving as peer instructors or teaching assistants (TAs). Peers can promote the development of a dynamic undergraduate research community in which students can serve as scientific mentors, with attendant benefits (Harrison *et al.*, 2010; Dunbar *et al.*, 2012). A faculty member with the requisite background in the fundamentals of genetics, molecular biology, and/or evolutionary biology can coach students to think critically about their observations and to question, probe, and analyze the data to address relevant scientific questions. Together, faculty and peer instructors can create the foundation for a lively genomics research team that is actively contributing to the scientific body of knowledge. Success in this project can provide students with the increased confidence that is a hallmark of research experiences for undergraduates. In the National Survey of Student Engagement (Bennett *et al.*, 2007), students deem research experience to be a high-impact practice.

To this end, several national programs have been established during the past few years that take advantage of genomics to engage undergraduates in research (Hanauer *et al.*, 2006; Hatfull *et al.*, 2006; Campbell *et al.*, 2007; Hingamp *et al.*, 2008; Shaffer *et al.*, 2010; Ditty *et al.*, 2010; Banta *et al.*, 2012; Singer *et al.*, 2013), in keeping with the recommendations of PCAST (2012), *Vision and Change* (AAAS, 2011), and *BIO2010* (NRC, 2003). Despite these reported successes and positive reports from long-range studies (Bauer and Bennett, 2003; Brodl, 2005), we have encountered some skepticism as to whether basing curriculum on a research project is an approach that is an efficient use of time and/or is broadly applicable to a diverse population of students, or whether it is best reserved for high-achieving students working in the summer, the current tradition.

The GEP provides an opportunity to evaluate the efficacy of a research project specifically in genomics in both attaining student learning objectives and achieving informed student attitudes regarding research experiences, drawing from a diverse pool of students. The GEP is designed to engage undergraduates in a joint research project in genomics while introducing them to bioinformatics tools and resources with the goal of increasing their understanding of eukaryotic genes and genomes, as well as immersing them in the practice of science. This project, initiated in 2006, now has ~100 affiliated schools and has engaged more than 1000 students at ~60 different colleges and universities during the past year alone. Training materials to familiarize students with bioinformatics tools and relevant strategies have been developed at Washington University and at GEP member schools, with collaborative assessment and revision. Undergraduate students participating in the program can improve the quality of genomic sequence and annotate genes and other features, elucidating meaning from DNA sequence. Research questions in genomics are addressed using the results of these efforts, leading to student presentations both locally and nationally, and ultimately research publications. Since 2006, students involved in the program have worked with the Genome Institute at Washington University in St. Louis (WUSTL) to improve more than 7 million bases of draft genomic sequence from several species of *Drosophila.* Using a suite of bioinformatics tools selected and/or developed in collaboration between the Department of Biology and Department of Computer Science and Engineering at Washington University (often with additional input generated on their campus), students have produced hundreds of gene models using evidence-based manual annotation. Improved sequences are submitted to GenBank (2013) and used in studies exploring genome evolution (e.g., see Leung *et al.*, 2010; other manuscripts are currently in preparation).

Since our initial assessment of students enrolled in 2008–2009 (Shaffer *et al.*, 2010), the GEP has doubled in size and attracted a diverse group of schools in terms of size, educational mission, and public versus private support. Faculty members have collaboratively developed a variety of ways to use the GEP approach in their teaching, including short (~10 h) modules in a genetics course, longer modules within molecular biology laboratory courses, stand-alone genomics lab courses, and independent research studies. This diversity of schools and approaches has allowed us to look for critical variables for student success, measured both in terms of responses to an online SURE-style survey (SURE is the Survey of Undergraduate Research Experiences [Lopatto, 2004, 2007]), which we will refer to as the "learning survey," as well as an online knowledge-based quiz. In addition, to determine long-term impacts on students' subsequent actions and careers, we have surveyed occupations and attitudes of students one or more years after the completion of a GEP-affiliated course. We find that institutional characteristics have little correlation with student success, indicating that diverse students in diverse settings benefit from curriculum-based research experiences of this type. We see a similar impact on student attitudes from the GEP course compared with a traditional independent summer research experience as measured by the learning survey, but find that the impact correlates with the amount of time the instructor was able to devote to the project. The impact of time spent is seen not only with students at the end of their GEP course experience, but also in their reports on their experiences in subsequent years. The data make a strong argument for allotting more instructional time to research-based work in the undergraduate curriculum.

MATERIALS AND METHODS

The Student Research Project

The GEP is a collaborative effort between the Department of Biology, the Department of Computer Science and Engineering, and the Genome Institute, all located at WUSTL, and a growing number of colleges and universities. (See http://gep.wustl.edu/community/current_members for a list of currently affiliated schools, and Figure 6 later in this article for characteristics of the schools participating in this assessment.) The project is organized around a central database housed and maintained by WUSTL on a pair of SUSE Enterprise Linux servers that host a variety of services utilized by the GEP community, including curriculum modules, access to research projects and bioinformatics tools, and assessment and communications tools (see http://gep.wustl.edu and Shaffer *et al.*, 2010).

Our current research efforts focus on exploring the evolution of the small fourth chromosome (dot chromosome or Muller F element) using some of the 20 species of *Drosophila* for which genome sequence is currently available (Clark *et al.*, 2007; Baylor College of Medicine, 2012). This chromosome, previously studied primarily in *Drosophila melanogaster,* is unusual in that, while the chromosome as a whole exhibits heterochromatic properties (intense DAPI [4′,6-diamidino-2-phenylindole]) staining, late replication, no meiotic recombination, high repeat density, and high levels of chromatin marks associated with gene silencing, including HP1a and H3K9me3), the distal 1.2 Mb exhibit euchromatic properties such as replication during polyteny and a normal gene density (Riddle *et al.*, 2009, 2012). Leung *et al.* (2010) presented results from a comparison of the *D. virilis* and *D. melanogaster* dot chromosomes carried out by Washington University undergraduates in a pilot project for the current GEP initiative. Over the past several years, GEP undergraduates have analyzed this chromosomal region from *D. erecta*, *D. mojavensis*, and *D. grimshawi*, and a comparable euchromatic region from *D. erecta* and *D. mojavensis*, covering 40 million years of evolution in reference to *D. melanogaster*.

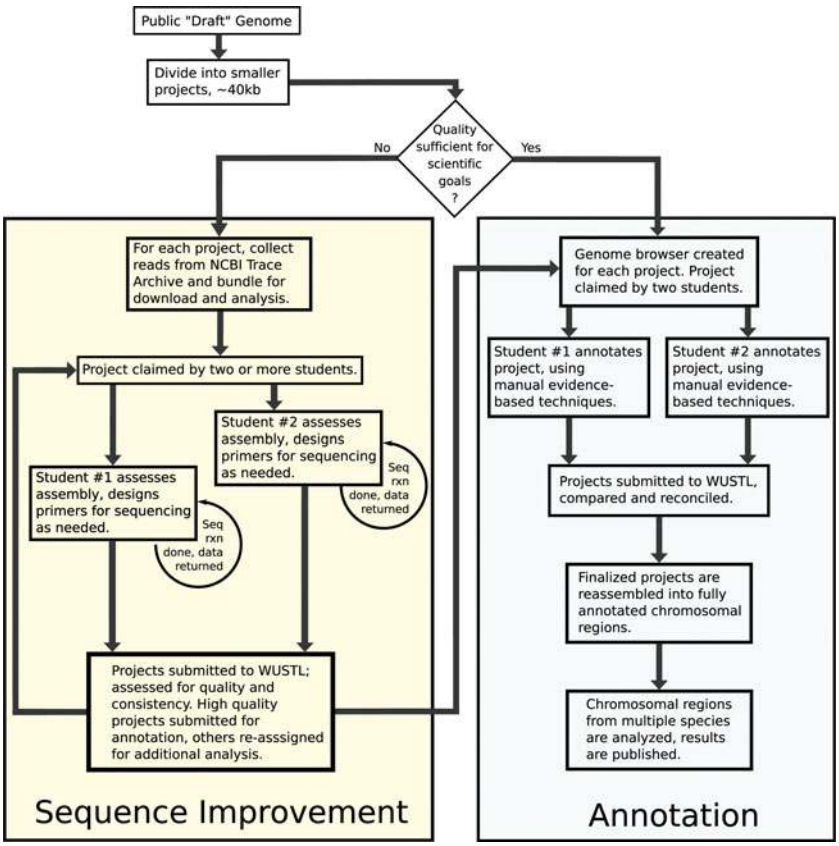

Figure 1. The GEP sequence improvement and annotation workflow. All projects are completed at least twice independently, and the results are reconciled before final assembly and analysis. While the quality of the genome sequence assembly available is sufficient to proceed directly to annotation for some species, in other cases it is necessary to improve the genome assembly ("finishing") by generating additional sequencing data and manually resolving misassemblies.

GEP students are challenged to verify the sequence assembly of an ~40-kb region of a *Drosophila* genome, and/or to annotate the region, identifying elements of interest and creating defendable gene models while working either individually or in teams (Figure 1). Projects are claimed and results submitted through the GEP website (Genomics Education Partnership, 2013a) using a standard, detailed reporting form along with appropriate sequence files. Each ~40-kb project is completed at least twice independently, and any discrepancies are resolved by experienced students working at WUSTL. GEP students and staff are able to draw on the expertise of the Genome Institute and the Department of Computer Science and Engineering for resolution of difficult issues. Reconciled projects are used in a reassembly of the chromosome region for characterization of the domain as a whole, and comparative analysis among species to chart evolutionary changes.

Most GEP students have been involved in annotation, with ~20% working on finishing as well; consequently, we will focus on the annotation experience in this paper. In the annotation research project, a predicted gene model (generated by an ab initio gene-finding algorithm such as GenScan, GeneID, etc.) serves as the starting hypothesis. While the overarching scientific question is predetermined by the group project (in this case, the properties and evolution of the *Drosophila* dot chromosome), we find that, with a little guidance, the student can be challenged to traverse the whole of the scientific research process, as shown in Table 1.

Faculty Implementation Strategies

GEP faculty members have utilized the GEP project in a variety of types of courses, including a first course in genetics, a molecular biology course, a lab in bioinformatics, independent study, and many courses that fall in between these descriptors. To capture the rich experience of the partnership, faculty responses to the two questions "How do you use GEP materials in your curriculum?" and "What advice would you offer to instructors interested in adapting this approach to fit the needs of their students?" are given in the Supplemental Material. One of the advantages of a consortium is the ability to share training materials and other curriculum items, which are posted on the GEP website. Current members have attended a 3- to 5-d workshop at Washington University to gain familiarity with the material. Much of the information provided at these workshops is readily accessible online, and we invite all educators to use the curriculum resources posted on the GEP website, under a Creative Commons license. (For an example of a ready-to-use script, see "An Introduction to NCBI BLAST" at http://gep.wustl.edu/curriculum/course_materials_WU/annotation/tutorials_and_walkthroughs.) While most GEP members participate in the broad research effort, claiming projects from the dot chromosome or comparison domain under study, an annotation problem can also be used simply to teach about eukaryotic genes and genomes. An example of a 3-wk lab module that introduces students to the structure of eukaryotic genes is posted at http://gep.wustl.edu/curriculum/course_materials_GEP_partners/able_workshop (Emerson *et al.*,

Table 1. Traversing the scientific process through student investigations within the GEP

The general process	The GEP process
1. Define the question	What genes or other features are present in this segment of a *Drosophila* species genome? What is the most likely gene structure? How has this region evolved?
2. Gather background information	*D. melanogaster* has been very well annotated; its evolutionary relationship with species under study allows for comparative analysis.
3. Experimental design	Students must decide which computational tools to use directly (e.g., BLAST to look for evidence of homology) and learn to seek other evidence from results displayed in a genome browser (e.g., ab initio gene predictors to look for computational evidence for the presence of a gene).
4. Collect experimental observations	Students generate BLAST results and collect other results (e.g., RNA-seq data) from a genome browser for their region of the genome.
5. Analyze collected data	Students create a gene model and test it using the collected observations; what gene model is best supported by the evidence?
6. Disseminate results	Students write papers and/or prepare talks or posters on their results, describing their results and defending their conclusions; pooled results are submitted to GenBank and linked to FlyBase.

2013), while a quick start for a longer investigation is posted at http://gep.wustl.edu/curriculum/quickstart.

Assessment Instruments

Learning Surveys. For the purpose of comparing the GEP experience with a summer research experience, we constructed a student survey that combines verbatim the 20 items previously used by a published survey (SURE survey [Lopatto, 2004, 2007]) with new items specific to the use of GEP materials. The SURE survey is a postexperience survey that asks students to respond to a list of 20 different knowledge or attitudinal benefits with respect to their research experience. Students are asked to rate their gains from 1 (none or very small gain) to 5 (very large gain). Using the SURE questions allows comparison of GEP student gains with those reported by students spending a summer working in a research laboratory. To compare attitudes of GEP alumni (students who completed a course with GEP materials one or more years prior) with those of students who had just completed the GEP curriculum, we prepared an alumni survey that contained questions identical to a portion of the GEP survey, as well as demographic questions about the alumni themselves. (All survey items are provided in the Supplemental Material.)

Knowledge Quizzes. We assessed knowledge gains of GEP students involved in annotation using a quiz composed of 20 multiple-choice questions. Quiz questions, written by the GEP faculty, were designed to test both conceptual knowledge about genes and genomes and specific skills related to the annotation process. The quiz questions assess a range of skills, from mastery of basic terminology and concepts to more complex cognitive skills, including data analysis and evaluation. [See Supplemental Table S1 for a tally of quiz questions as they relate to the revised Bloom's taxonomy (Anderson *et al.*, 2001).] Two versions of the quiz have been created to avoid the effect of repeated testing. This quiz is distinct from that used in a prior publication (Shaffer *et al.*, 2010) in both the attention to Bloom's taxonomy in question design and in the pre/postcourse administration protocol. The online quizzes are available by request.

Data Collection and Analysis. Both the learning survey and the knowledge quizzes were accessed by the students through the GEP website. Confidentiality was maintained by applying a cryptographic hash function to an identifier provided by the student. Participation was entirely voluntary, and students were able to opt out of the entire process or any single question. Approval to conduct assessment for scholarly purposes was obtained from the local Institutional Review Board (IRB) at each participating institution. For a comparison group for the quiz, we recruited students at participating schools who had completed the prerequisites to the GEP-affiliated course but were not engaged in the GEP research-based curriculum.

During the academic years 2010–2011 and 2011–2012, of ~2000 students eligible to participate, 751 postcourse learning surveys were collected; 1026 students took the precourse knowledge quizzes, and 748 took the postcourse knowledge quizzes. Students from 57 schools contributed to this data set. Data loss due to students' missing the pretest, the posttest, or using faulty or no identification yielded a matched data set of 394 sets of quizzes. To avoid the possibility that any improvement from pre- to postcourse scores was due to student exposure to the precourse quiz, we used two similar quizzes covering the same material but using different questions, as noted above. Students taking the precourse quiz were randomly assigned to one version and then given the other version when they returned for the postcourse quiz. Comparison of these data with the posttest-only data revealed no significant differences based on either the experience of the pretest or the version of the quizzes that the students encountered (unpublished data). Student affiliation with a participating partner school is maintained, allowing cross-correlation with institutional characteristics (as reported by U.S. News & World Report Staff [2011] and verified by the GEP faculty) and with course characteristics reported by the GEP faculty members.

For the alumni survey, GEP faculty members obtained IRB approval at their institutions and disseminated a link to the online survey to all of their GEP alumni. Of 1645 students eligible, 473 students (29%) from 41 institutions participated. As above, participation was entirely voluntary, and students were able to opt out of the entire process or any single question. The survey requested identification of the student's GEP school, type of GEP course, and other course details; asked about career status/plans and the value of the GEP experience; and invited recommendations and comments.

Except for items that collected demographic or categorical data or that provided text boxes asking for comments in which participants were allowed to enter any arbitrary input, all items across all surveys asked participants to indicate their feelings/responses on a 1–5 scale (e.g., see section on *Learning Surveys*). We treated these responses as numerical data. Unless otherwise stated, all averages reported are means, and errors are reported as ± 2 SEM; significance was determined at $p < 0.05$. To compare classroom-based GEP responses with those reported by summer (SURE) research students, we used an independent-groups t test. To test for correlation between the institutional characteristics and student learning outcomes (both learning survey responses and knowledge gains shown on quizzes), we applied multiple linear regression using IBM SPSS Statistics Version 20. This package was also used for all subsequent analysis. To look for any difference between groups of students with respect to both knowledge and learning gains, we used one-factor analysis of variance (between groups). The quartiles used in the analysis of outcomes relative to time spent in class on GEP material were defined by rank ordering all classes by time spent (as reported by GEP faculty members) and separating them into four groups with equal numbers of participating schools.

For the numerical analyses of comments, we classified each sentence as "positive" if the sentence contained a conventionally positive emotion ("liked," "loved," "appreciated," "enjoyed", etc.) and as "negative" if the sentence contained a conventionally negative emotion ("disliked," "confused," etc.). Other groupings were built on the key words indicated, including similar phrases, variants, and synonyms thereof; for example, "independence," "independent," "independently," were grouped together along with synonyms such as "freedom" or "on my own."

To describe quiz questions with respect to categories in Bloom's revised taxonomy, 10 GEP faculty members evaluated each question independently. The results indicated that many of the questions required the use of more than one skill category. Therefore, we created four intervals of overlapping adjacent categories (shown in Table S1) and calculated the modal score for each question with respect to those four categories. In most cases, there was agreement on this binning (10 out of 10 responses falling into these two categories), but responses were quite wide-ranging in a few cases, with only six of 10 responses falling in these two categories.

RESULTS

Gene Annotation Requires Students to "Think Like a Scientist"

The GEP facilitates the process by which a student confronts the challenges of annotating a 40- to 60-kb stretch of *Drosophila* DNA. To generate a defendable annotation, students must analyze and evaluate multiple available lines of evidence to generate gene models within their claimed sequence. The GEP project is set up using the genome browser software developed by the Genome Bioinformatics group at the University of California–Santa Cruz (Kent *et al.*, 2002). The browser is hosted at WUSTL and populated with in-house-generated evidence tracks for each student project (GEP, 2013b). By selecting appropriate tracks, the student can see the results of a BLASTX search against *D. melanogaster* (identifying conserved protein-coding regions), the predictions obtained with several different ab initio and evidence-based gene finders, the results from RepeatMasker, any RNA-seq data available, TopHat analysis of the RNA-sequencing data (suggesting exon/intron splice sites), predicted splice site donors/acceptors, various conservation tracks, and so on (Figure 2). Inevitably, some of the lines of evidence supporting the presence of a gene will be contradictory, particularly for the details of exon/intron structure. The student must decide which collection of evidence should be given most weight in deriving his or her final gene model and be prepared to defend that conclusion.

The execution of a GEP research project is designed to require a student to traverse all six categories of cognitive skills found in Bloom's taxonomy (Bloom and Krathwohl, 1956; Anderson *et al.*, 2001; Table 2). Students start with practice annotation problems (posted on the GEP website at http://gep.wustl.edu/curriculum/course_materials_WU/annotation/annotation_exercises) and are guided by a general GEP protocol. Research outcomes will include the expected and the unexpected—all "rules" are broken at least some of the time! Among the latter, students have identified new genes not present in *D. melanogaster*, instances of stop-codon read-through, changes in exon number, shifts in splice sites, use of noncanonical splice sites, insertions of additional amino acids, and the loss and gain of isoforms. Another annotation challenge can be found with pseudogenes, which are rare in *Drosophila* but do occur. Changes in gene order and orientation within the element are observed, and ~10% of all genes have moved from one chromosome to another during 40 million years of evolution (for examples, see Leung *et al.*, 2010). Thus, while students are guided by their knowledge of the well-annotated *D. melanogaster* genome, each new species presents numerous challenges and reveals a new perspective on genome structure. At the end of their analysis, students submit a standard report to the GEP. In addition, most faculty members (96%) require students to communicate their findings to their colleagues, mentors, and/or the broader community as part of the overall process. As reported by GEP faculty, a variety of mechanisms have been implemented, including oral (83%), written (94%), and poster (40%) formats.

Project results are compared and reconciled by undergraduates working at WUSTL during the summer. We find complete congruence in 50–65% of submitted gene models, varying with the level of difficulty as judged by the degree of sequence divergence from our reference species, *D. melanogaster*. (Work on a given species usually stretches over more than one semester; as a consequence, some recorded discrepancies occur simply because the *D. melanogaster* annotation has changed between the first and second round of GEP student annotation.) Common errors, most of them readily rectified by experienced students, include missing an annotation for a gene present in that project, missing a possible isoform, or choosing a nonoptimal intron/exon splice site. Generally, students make poor choices when they rely exclusively on one source of data (BLAST alignments, gene predictors, or other), rather than assessing all of the data available to identify genes, isoforms, and splice sites.

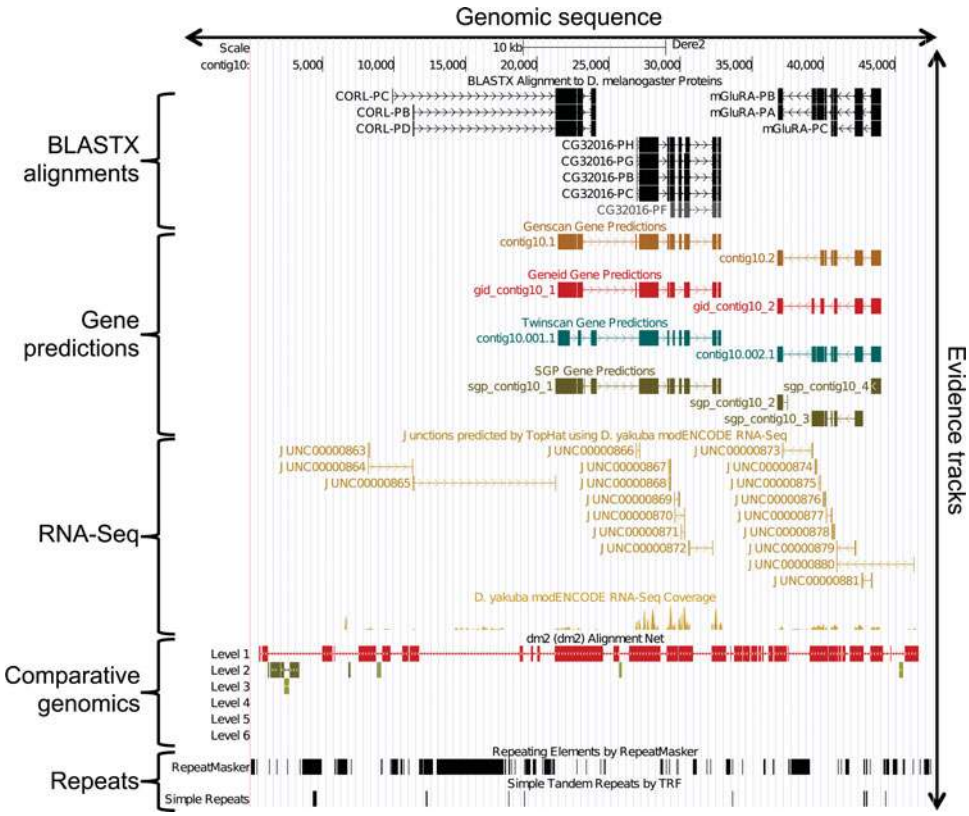

Figure 2. The GEP UCSC Genome Browser. Students are challenged to analyze and evaluate the available evidence assembled on the genome browser to create optimal gene models and explore other genomic features. A sample of the available tracks is shown here. Often the available evidence is contradictory; e.g., see the discrepancy between some of the gene predictions and the organization of genes suggested by homology with *D. melanogaster* (BLASTX alignments).

Collectively, GEP students in collaboration with the Genome Institute at WUSTL have examined and improved the quality of more than 7 million base pairs of sequence from these species and have generated more than 1000 gene models, providing a detailed picture of the differing characteristics of the genes present on the fourth chromosome and the pattern of evolution of this domain. We are currently preparing a manuscript describing student-generated results from this study that will have ~500 student and ~50 faculty coauthors and will acknowledge contributions from many classes.

Participation in GEP Projects Promotes Knowledge Gains

GEP projects are designed to easily integrate student research into laboratory curriculum suitable for use during the academic year while remaining fundamentally grounded in research. Success toward this goal is measured by student gains in both cognitive and affective domains. The desired outcome is that GEP students are learning the concepts and skills being taught, becoming engaged and confident in their skills, and perceiving their work as a valuable research contribution.

Table 2. GEP students completing a research experience utilize a range of cognitive skills from Bloom's taxonomy[a]

Taxanomic skill	Utilization by GEP students
Remembering	Correctly use and define terms
Understanding	Explain steps required for gene annotation
Applying	Use BLAST to identify sequences similar to the sequence of interest
	Use FlyBase, UCSC Genome Browser, Gene Record Finder to retrieve information about a gene
Analyzing	Diagram possible gene structure
	Identify features (CDS, exons, repeats) in a genomic DNA sequence
Evaluating	Evaluate alternative gene models
	Select most likely gene model and support choice using multiple lines of evidence
Creating	Assemble a well-documented annotation for a region of the genome under study

[a]Bloom and Krathwohl, 1956; Anderson *et al.*, 2001.

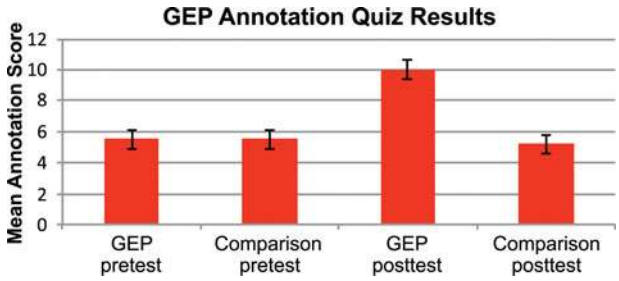

Figure 3. GEP students show gains in their understanding of genes and genomes. The bars depict the mean scores on a 20-point quiz on genes and genomes attained by GEP students who participated in annotation and by a comparison group over 2010–2011 and 2011–2012. The GEP pretest data include 1026 observations; the GEP posttest data include 748 observations. The increase in scores was evaluated in two ways: first, with an independent-groups t test ($t = 26.1$, $df = 1836$, $p < 0.05$), and second, with a paired t test for matched data ($t = 21.9$, $df = 393$, $p < 0.05$). The comparison pretest data include 133 observations, while the comparison posttest represents 87 results from non-GEP students. The error bars represent two SEM. Cronbach's alpha statistic averaged 0.84 for these quizzes in 2011–2012.

We assessed knowledge gains concerning genes and genomes using the quizzes described above. Comparing the performance of all GEP students on a pre- and postquiz, we find that the average score increased by 4.4 points (out of 20) at the end of the course (Figure 3). This increase was not seen in the comparison students who had not been exposed to the material through a research-based curriculum. We note that students were provided no external incentive to make an effort to perform well on this quiz, suggesting that the gains shown reflect secure knowledge. Those students who participated in sequence improvement (finishing) showed similar gains on a finishing quiz (Supplemental Figure S1). These results are consistent with the previously reported knowledge gains for a smaller group of students measured using an earlier version of the quizzes (Shaffer *et al*., 2010).

To investigate these gains in detail, we categorized quiz questions based on the cognitive skills tested using Bloom's taxonomy and found that questions correctly answered by GEP students on the postquiz spanned all skill categories. To look specifically for gains in higher-order cognitive skills, we compared results for quiz questions that tested lower-level skills (Bloom's levels 1–2; see Table S1) with results for the other quiz questions, which were designed to test higher-level skills. Comparing the performance of all GEP students on the pre- and postquiz, we find that the average score for lower-level quiz questions increased from 4.20–5.75 points, while the average score for higher-level quiz questions increased from 1.56–4.4 points (total $n = 382$ for this matched set). These data suggest that students' engagement in the annotation project provided practice and subsequent learning gains across a wide range of cognitive skills (see Table 2), and point to gains in higher-order skills. Although the absolute scores on the current quiz were lower than those previously reported (Shaffer *et al*., 2010), the degree of change is similar. We attribute the shift in absolute scores to the relative difficulty of the current versions of the quiz. Nearly half of the questions on the new quizzes used here test higher-order cognitive skills, while the earlier version focused on lower-level skills.

Students Engaged in GEP Show Learning Gains Similar to Students in an Independent Summer Research Project

We assessed student attitudes and perceptions of their GEP-related experience using a modified SURE learning survey (Lopatto *et al*., 2008). For comparison, we used data collected from students who were involved in a traditional summer undergraduate research experience in research laboratories. In a section identical to the SURE survey, we asked students to report their learning gains on 20 items. The average responses from the classroom-based GEP students show greater gains than those reported by summer (SURE) students on 19 of 20 items, although for many of the items, the difference was not statistically significant (Figure 4). The results from this larger and more diverse pool of students confirm our prior finding that the GEP research project is as effective as a summer research experience by this measure (Shaffer *et al*., 2010), but the larger data set allows a more detailed analysis as well.

The survey comparison reveals several noteworthy differences between the GEP and the SURE experiences. The two single items with the largest difference involved statements wherein students indicated how much they felt they gained in understanding science. For both the statement "Understanding how knowledge is constructed in this field" and the more general "Understanding science," GEP student responses averaged above 3.85, while the averages of the SURE student responses were below 3.53. (For both of these categories the SE for the averages of the SURE and GEP responses is <0.04.) We suggest that it is the process of grappling with contradictory evidence and the need to generate a defendable resolution in a rather short period of time (during one semester) that elicits these gains. While desirable, this sort of challenge does not always occur during the summer research experience. Similar differences were also seen for "Understanding that scientific assertions require supporting evidence" and "Ability to analyze data and other information." Another noteworthy item is "Learning laboratory techniques," for which the SURE responses averaged significantly above those of the GEP students (3.82 vs. 3.46). It is clear that not all students view their acquisition of new computer-based skills as "learning laboratory techniques," presumably because they associate laboratory techniques with traditional bench or field experimentation.

Because they are designed to be group projects centered on genomics, the GEP courses provide the opportunity for students to gain additional skills not reflected in the SURE design. Indeed, GEP students also reported significant and reproducible gains in learning computer skills, skill in reasoning from data, self-confidence in discussing science with peers and mentors, skill in scientific writing, and learning to work as part of a team (average ~3.8 on a scale of 1–5, as above). Despite the reported gains in learning, students reported smaller gains in their willingness to take additional courses in math and computer science. The results were consistent year to year (Figure 5).

As part of the GEP-specific attitudinal survey, students were asked to assess how much they gained from the various teaching materials and course activities. As reported previously (Shaffer *et al*., 2010), students gave the highest ratings to working on their own projects, in agreement with previous findings stressing the importance of student "ownership" (Hanauer *et al*., 2006). Student comments were

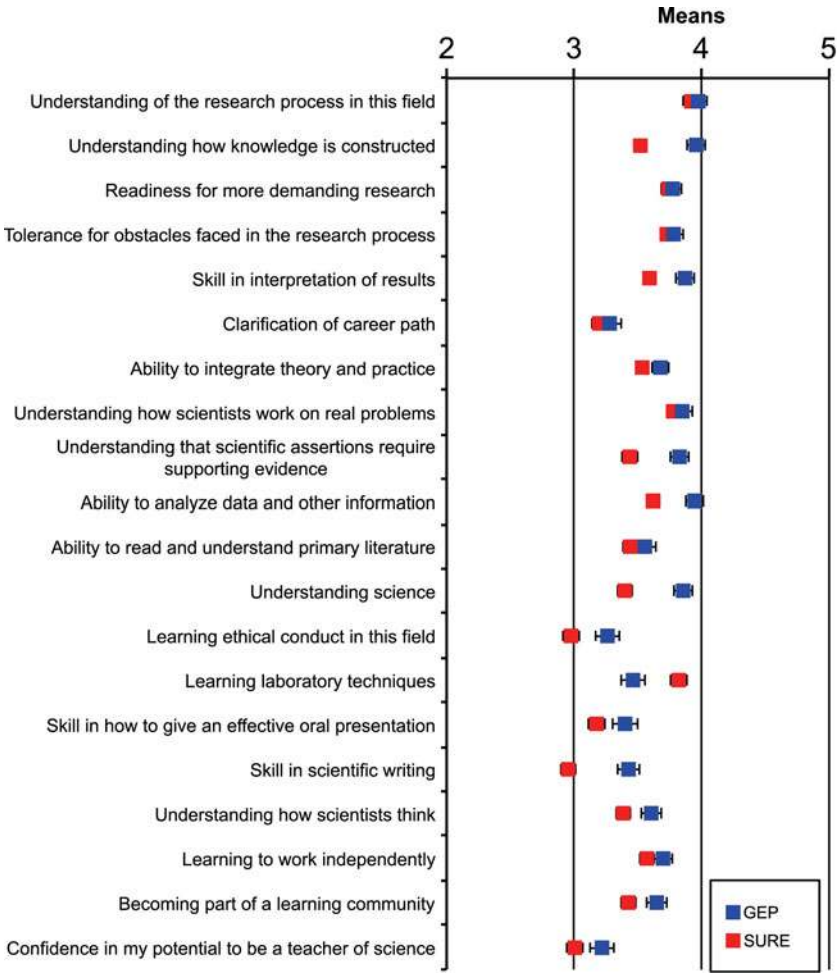

Figure 4. Self-reported student learning gains using the SURE survey. Blue squares indicate the mean for GEP students, while red squares indicate the mean for SURE summer research students, 2009. Error bars represent two SEs below and above the means. The SE for the averages of the GEP and SURE responses was <0.04. Data shown combine results from surveys given in academic years 2010–11 and 2011–12; the data include between 652 and 751 responses on each of the 20 items from GEP students. The comparison group is the 2009 SURE survey of 1653 students who had just completed a summer in the lab. The large number of students allows for smaller error estimates than in our previous study (Lopatto *et al.*, 2008).

invited at the end of the survey; these comments also stressed the importance of being responsible for their own projects, while participating as a team member.

Students at Diverse Institutions Show Similar Gains in Project Outcomes Assessment

Having a large data set from diverse institutions allows us to correlate various aspects of the GEP experience with desirable outcomes. The GEP is made up of a very diverse group of schools (see current members at GEP website), allowing us to look for possible moderators influencing student performance. Institutional parameters examined included public or private status, size (total enrollment), degree types granted in biology, selectivity of admission, and other publicly available data. We collected additional data on the character of the student body, including the percentage of the student body that is residential versus commuter, minority, first generation to college, or nontraditional (more than 25 yr old). Thus, we selected 10 characteristics of interest and

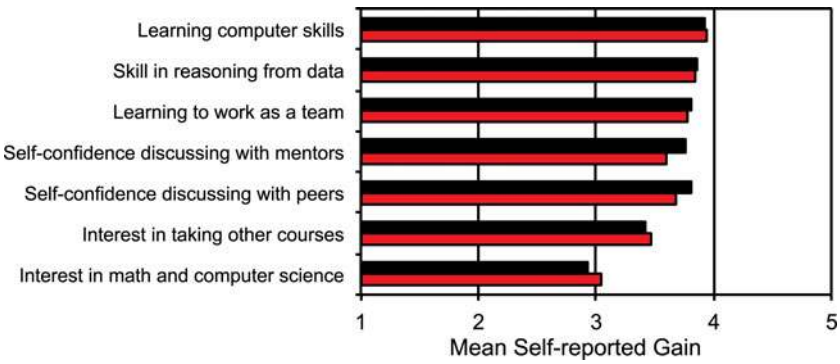

Figure 5. Additional gains from a GEP experience not queried in the SURE survey. Student self-reported gains, assessed on a scale of 1 (no gain) to 5 (very large gain). GEP 2011 (337–344 cases): black; GEP 2012 (391–394 cases): red.

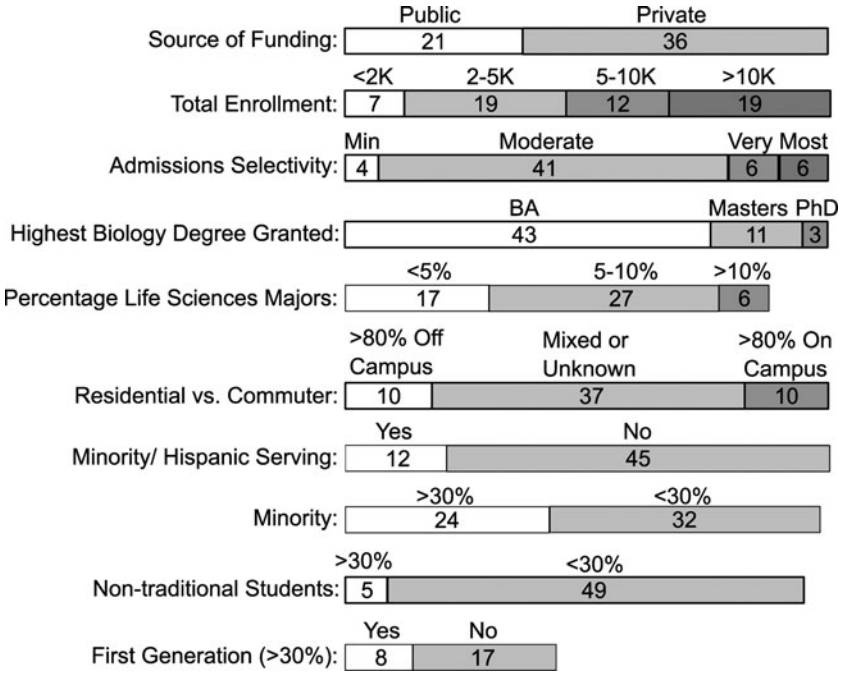

Figure 6. Diversity of GEP institutions with students participating in the above assessment during 2010–2011 and 2011–2012. For the purposes of this survey, nontraditional students are defined as those over age 25. Total number of schools represented is 57. Some schools do not collect some of the above data, resulting in some incomplete data sets. Data from U.S. News & World Report Staff (2011) or supplied by the institution.

binned the schools into various categories based on these data (Figure 6).

We then tested these categorical data to see whether the characteristics correlated with student outcomes as measured above. Using multiple linear regression to test for correlation, we found that neither student improvement on the annotation quizzes nor positive responses on the GEP SURE-matched learning survey correlated significantly with most characteristics of the home institution. Correlations were not found to be statistically significant for public versus private school, student body size, the presence or absence of advanced degree programs in biology, or service to any particular type of student (e.g., first generation, nontraditional, commuter, and/or minority). All institutional characteristics (shown in Figure 6) taken together accounted for only ~6% of the total variance within the annotation quizzes and self-reported SURE learning benefits (averaged over 20 benefits). Looking at student subpopulations, we find that students across all ethnic groups benefited; there is no statistically significant difference in pre/postquiz gains and no difference in GEP SURE learning benefits among ethnic groups. While we understand that these measures may not have captured all the ways in which institutional differences can affect student outcomes, we have found from individual experiences that the GEP offers a pedagogical approach that has been successful in many settings. To expand on this theme, we as faculty have compiled our personal experiences in response to the question "What has been the impact on your students?" These responses are provided in the Supplemental Material (Text S2). We are encouraged to believe that students from diverse institutions and backgrounds can greatly benefit from a research-oriented laboratory experience in genomics such as the one outlined here.

Time Devoted to the Project Is a Critical Factor in Gaining a Research Experience.

As noted above, we have utilized GEP project materials in a variety of types of courses. See the table of faculty members on the GEP website (http://gep.wustl.edu/wiki/index.php/Table_of_Faculty) for a listing of syllabi, indicating the participating school and type of course. We used the table of faculty reports to gain insights into how the GEP curriculum was being implemented. When the GEP research project is a central focus of the course, 48 faculty members report that they organize the class such that students do the majority of the work in class and typically (40 of 48) devote a total of 25–45 instructional hours to the project (lecture, discussion, demonstration, lab work). Some faculty members (six) use a course design in which students are expected to do a substantial fraction of their analysis outside class time; for these courses, 10–25 h of class time is devoted to the project. An independent study course may focus on the GEP project all semester or be combined with another topic. In contrast to these, a 3-wk lab module designed primarily to introduce students to the structure of eukaryotic genes may require as little as 10 h (see example posted at http://gep.wustl.edu/curriculum/course_materials_GEP_partners/able_workshop; Emerson *et al.*, 2013). While the latter utilization of GEP material serves its immediate purposes of teaching the structure of eukaryotic genes and exposing students to bioinformatics tools and databases, it raises questions as to whether the students can gain what can be classified as "research experience" in such a short time interval. To address this question, we compared student gains as assessed through the postcourse survey with the amount of time spent using GEP materials. Faculty members reported participation levels that ranged from 3 to 64 h per term. From these data, we created four quartiles (Q1–Q4) of

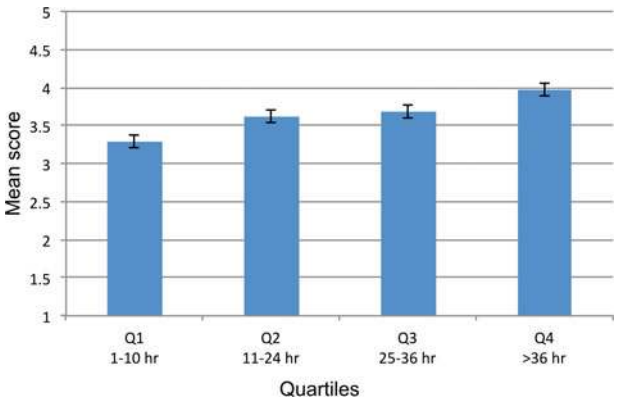

Figure 7. The 20 learning gains (SURE) reported by GEP students were averaged and plotted against the four quartiles of annotation project hours. There was a significant difference in average learning gains across quartile groups ($F = 11.9$, $df = 3, 374$, $p < 0.05$). Pairwise contrasts indicate a significant difference between Q2 and Q4, but not between Q3 and Q4.The number of class hours utilized by each quartile group are shown above. The number of student respondents in each quartile is: Q1: 58; Q2: 139; Q3: 65; Q4: 116. (Respondents tallied here are only those who answered all 20 questions on the SURE survey.) Error bars represent two SEM.

instruction time. A bar graph showing the relationship between annotation instruction time and student learning outcomes (SURE-type survey) is shown in Figure 7. The results show a strong relationship, with higher learning gains (the average of all 20 SURE questions) resulting from more time devoted to the project.

These results argue that it is necessary to invest significant course time (combined lecture, discussion, lab work time, etc.) for students to gain a research experience, as defined by the questions in the SURE survey, when a genome annotation project is used. Faculty observations suggest that a time investment is needed to gain familiarity with using the bioinformatics tools before students can begin to ask their own questions, feel comfortable in interrogating the data, and gain confidence in their own analytical abilities. Nonetheless, the instructional time needed is quite short compared with a summer in the lab, in which multiple weeks of preparation and work are necessary to provide mentorship for a relatively small number of students. The average time spent by Q2 faculty was 20.4 h, while that spent by Q4 faculty was 45.5 h.

To investigate in detail the influence of time spent on student learning gains, we mapped the average gain for students in each quartile onto the individual potential learning gains from the student surveys. The results on comparing the average gains of the Q1 and Q4 students, as seen in Figure 8, reemphasize the advantage of devoting extended class time to the development of a successful project. The seven additional learning benefits that GEP students evaluated show similar differences (Figure 9). Note that reporting the data in this way allows a tally of all responses to a given item, resulting in a significantly larger pool of respondents. The results support our prior conclusion, specifically, that a significant commitment of class time (>36 h) is required to obtain the full benefits of a research experience.

Alumni Attitudes Also Show Increased Value with More Time Invested

To determine the effect of this curriculum-based research experience on student career trajectories, we obtained demographic and attitudinal data on students who had formerly taken a GEP course or served as a TA in a GEP course. Data were collected during the summer and fall of 2012 by an online survey that included demographic and attitudinal questions. The earliest cohort surveyed took GEP classes in 2005, the last cohort in Fall 2011. We had 473 valid responses (29% of the students contacted); however, because all questions were voluntary, the total number of responses to any one question rarely totaled 473. The students reported that they came from 41 different institutions. The diversity of institutions represented is similar to that found in the student survey (Figure 6; see data in Figure S2). Of the respondents, 200 described themselves as male, 268 as female; the pool was 59% Caucasian, 20% Asian, 17% underrepresented minorities (African American, African, Hispanic), and 5% mixed plus other. All others declined to respond.

To investigate the current occupations of the alumni respondents, we gave them a series of 16 career categories and allowed students to select one or more as appropriate. Figure 10 shows the total number of alumni that selected each category. The three largest areas of occupation (pooling some categories) reported by the alumni were being a postsecondary student in science (pursuing an MA, PhD, or other professional degree: 30%), medical school student (pursuing an MD or MD/PhD: 22%), and employment in science (19%). Only 9% indicated they were no longer in science (pooled PhD nonscience, other professional nonscience, and employed nonscience). These results are very encouraging; however, as in any survey of this type, the findings may be impacted by response bias; that is, those students continuing in science may have been more inclined to respond to the request to participate in the survey.

To examine the attitudes these alumni had toward science in general and their experience with the GEP in particular, we asked them to complete a survey much the same as our original postcourse learning survey. Alumni were asked to reflect on and evaluate their experiences and learning gains, given that some time had passed since they had taken the GEP course. Figure 11 shows the average response to these questions, with 1 being "strongly disagree" and 5 being "strongly agree," that their GEP experience fulfilled each particular goal. In addition to the overall average for these topics, we also cross-correlated the results with the extent of the GEP experience to which each student was exposed. Students were asked to pick one of four responses as describing the type of GEP course that they took. These were: 1) a course devoted primarily to the GEP project, 2) a course spending half or more of the time on the GEP project, 3) a course that spent a quarter of the time on the GEP project, and 4) a course in which just one to three labs were devoted to the GEP project. Figure 11 shows the average responses for each of these groups (numerical data in Table S2). In general, there was a clear and consistent increase in attitudinal gains as students spent more course time working on their GEP project. The average of all items for each group was 1) 4.02, 2) 3.62, 3) 3.30, and 4) 3.03.

Items common to both the alumni and current student surveys allow us to directly compare responses from current

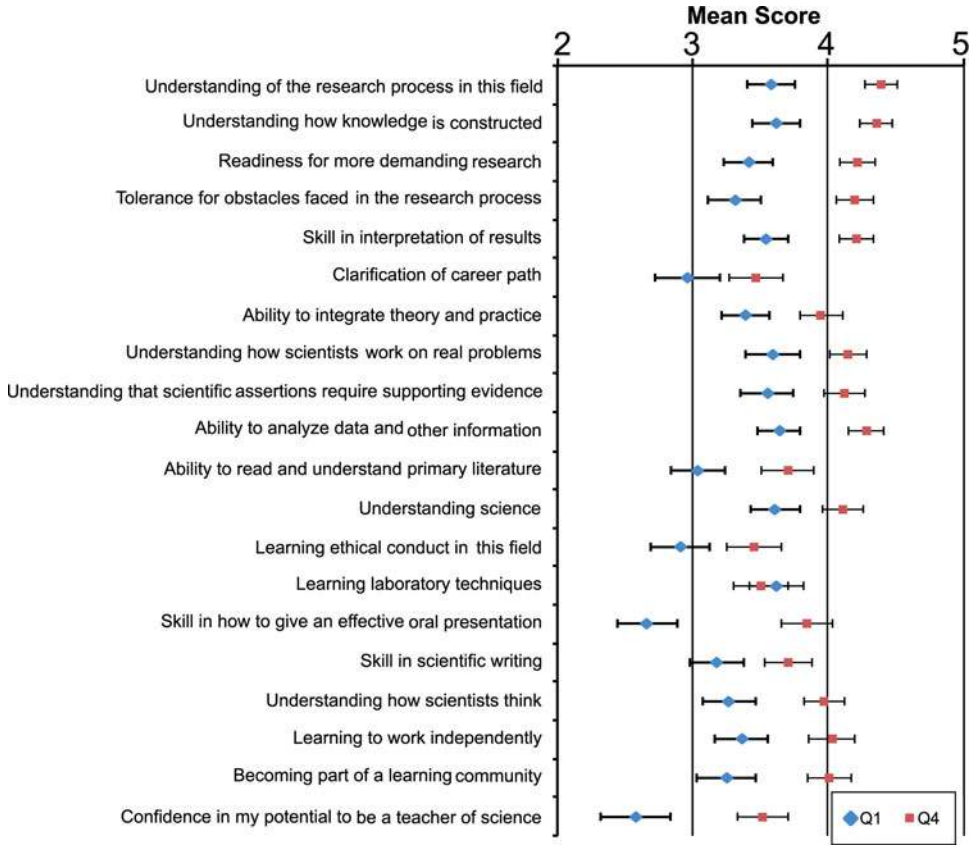

Figure 8. Comparison of student responses on the 20 learning gain items (mean and SEM) from the SURE survey. The data are classified by instructor reports of the number of hours devoted to the annotation project. These were divided into four quartiles as shown in Figure 7; the responses from the Q1 (1–10 h) and Q4 (>36 h) students are shown here. The Q1 group includes 86–112 observations; the Q4 group includes 149–175 observations. (Respondents tallied here were those who answered the specific question on the SURE survey.) Error bars represent 2 SEM.

2012–2013 GEP students with those of the alumni. Figure 12 shows these comparison data. In most cases, the alumni had a very positive response to their GEP experience, showing an average response to each item that was ∼0.25–0.4 units higher than the average from the current GEP 2012 cohort. One item, "Genomics is awesome," showed a difference larger than the typical range, with a difference of 0.63. These results suggest that the students value the experience more as they find that what they learned in their annotation project has applicability to their further pursuits.

Again, one might be concerned about response bias. In particular, we asked whether the students who had spent more time working on the GEP research project were disproportionately represented in the pool of students who responded to the alumni survey. We found that 30% of the alumni respondents fell into the group having a brief GEP experience (one to three lab sessions, the equivalent of Q1 above), while 26% of the current students fall into Q1 (<10 h work with GEP). Thus, the two pools are roughly similar in distribution across the course types through which the GEP project is offered.

Alumni students were also invited to comment on their GEP experience. In particular, students were invited to respond to the question "Please comment on your GEP experience: What was good about it, and what changes would have made it better?" Most of the student comments were favorable (∼90%). The most common comments from alumni stated that they enjoyed the course, they enjoyed the independence, they appreciated having a real research project, they learned a lot, and the experience made them feel like researchers. Other comments emphasized that the experience was relevant to their future plans, was interesting, and inspired teamwork. Several students commented on their initial confusion and subsequent development of understanding. (See Table 3 for a numerical analysis of the 320 comments, as well as sample comments for each category.) The results of this survey support the conclusion that we can provide a robust learning experience in genomics through a research project. They further reinforce the notion that an in-depth research experience with a considerable time investment produces the most tangible and long-term gains.

DISCUSSION

The importance of providing research experiences for science, technology, engineering, and mathematics undergraduates is well documented (see references cited in the *Introduction*). However, providing such experiences has been challenging due to the substantial resource and personnel requirements for individual mentored research. Classroom-based undergraduate research experiences can provide an alternative,

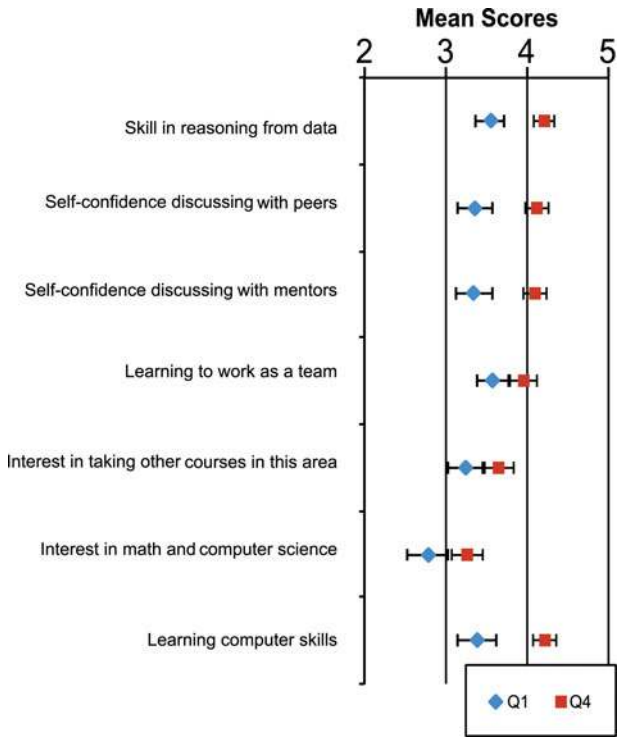

Figure 9. Comparison of student responses on seven additional learning benefit items categorized by the reported number of hours of instruction and lab work on the annotation project (Q1: 1–10 h; Q4: > 36 h). The Q1 group includes 105–107 observations; the Q4 group includes 172–174 observations. (Respondents tallied here were those who answered the specific question.) The error bars represent two SEM.

but despite this, many outstanding pilot projects (see *Introduction*) have not been widely adopted, perhaps because of concerns about the different needs and cultures of institutions of various sizes, missions, and budget, or of students with diverse levels of preparation. The expansion of the GEP to ~100 affiliated schools in the past 3 yr has allowed for the collection and analysis of student outcomes to determine whether specific factors contribute to the knowledge gains (quizzes) and self-reported learning gains in understanding research (survey) observed among GEP students. Institutional characteristics such as public or private funding, student body size, the presence or absence of advanced degree programs in biology, or the prevalence of particular subgroups of students such as first-generation college students or minority students had no significant effect on the overall knowledge and learning survey scores. Interestingly, responses by both current students and GEP alumni indicate that gains are enhanced by increased time invested in the research project in the classroom.

These findings demonstrate the utility of a broad-based, centrally organized genomics research project in providing successful classroom research experiences for undergraduates across diverse institutions. They further suggest that it is worthwhile to devote a significant amount of class time to achieve the full benefits of a research experience for a large number of students. The outcomes of student participation in such a program compare favorably with extracurricular undergraduate research experiences in terms of positive influence on students over time. This outcome may in part be the result of features of the project structure that mirror both the scientific process and utilization of the cognitive skills outlined in Bloom's taxonomy; student responses indicate strong gains, particularly in learning about the nature of science and the research process, along with attainment of competence and confidence in research skills, particularly in analyzing conflicting data and achieving a defendable resolution.

The data demonstrate a significant correlation between the amount of time devoted in a course to GEP materials and the perception of the experience by undergraduates as actual research, a result seen among both current students and GEP alumni. Importantly, students who were exposed to as little as 10 total semester hours of GEP work showed significant learning gains, as did students exposed to an average of 45.5 instructional hours (Q4) in a semester. However, the distinction between smaller and larger faculty investments of class time in the research project is revealed when assessing student confidence data; a significant investment of classroom time in the project is necessary to provide an experience that attains the full benefits of a research experience. This time-dependent variation in our learning versus confidence outcomes reflects

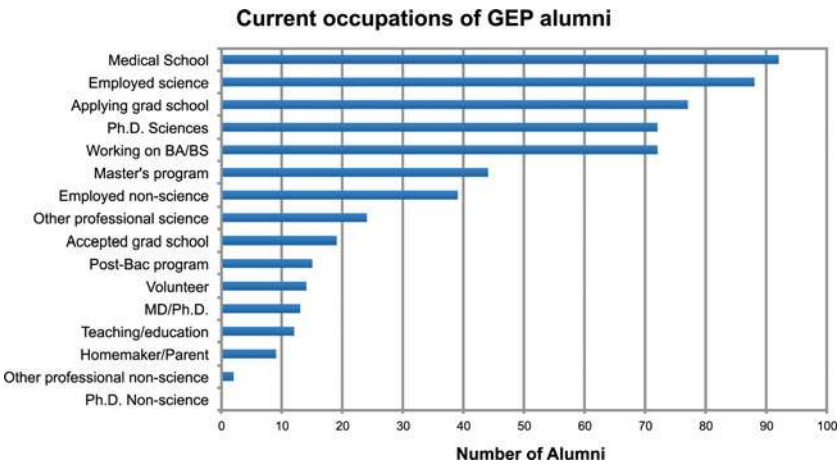

Figure 10. The current occupations of GEP alumni as self-reported on the alumni survey.

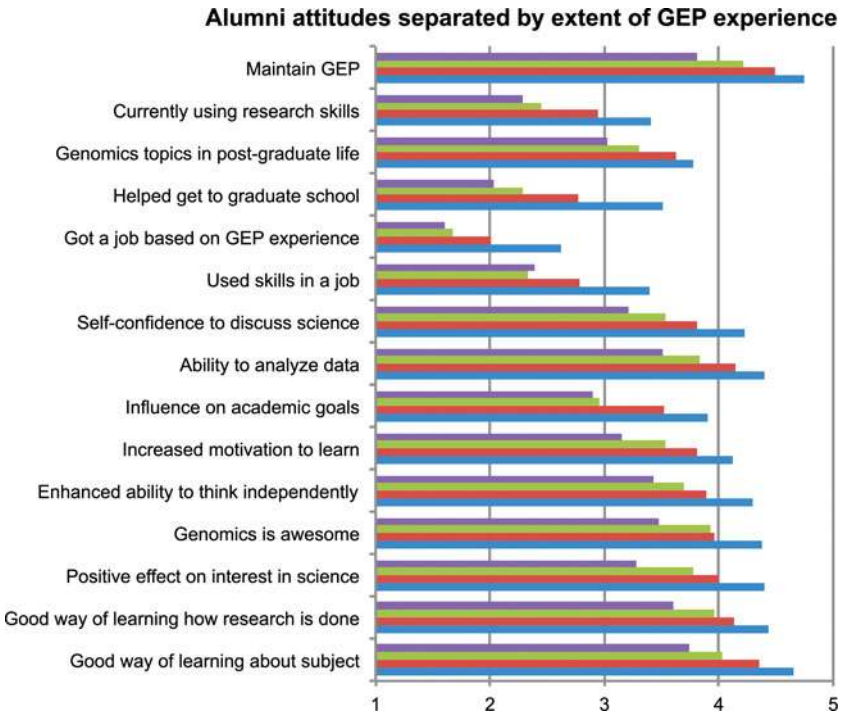

Figure 11. Alumni student attitudes separated by extent of GEP experience. Purple: "Took up just 1–3 labs"; green: "Course spent quarter"; red: "Course spent half or more"; blue: "Course devoted primarily to the GEP project." Scale is 1–5, with 1 being "strongly disagree" and 5 being "strongly agree." See Table S2 for numerical values.

the classic question of "content versus experience" that is at the heart of many pedagogical debates in science education: Is the student educational experience compromised by sacrificing emphasis on content for a more inquiry-based and experiential learning? We would argue, based on our above findings, that the use of a research project as a significant component of a biology course has significantly bolstered student confidence and competence, as evidenced by GEP alumni student comments and by them continuing their careers in the sciences. An increase in the science-literate workforce is one of the most tangible goals as we respond to calls for increased undergraduate research opportunities.

A notable outcome of implementation of the GEP curriculum as compared with a SURE experience is the same or larger reported student learning gains in understanding the practice of science. The differences in scores between GEP and SURE students may be influenced by the fact that students in a research-based classroom environment are in a large community of peer learners. Undergraduate students involved in a summer research experience at a university are often working with a postdoc or graduate student and few direct peers. The parallel research experiences provided as part of a research-based course allow for peers to assist and support one another in deeper and more meaningful ways than they might otherwise have the opportunity to do. The simultaneous act of mentoring and being mentored contributes to the opportunities that students have to gain practice and confidence in their abilities as scientists. Additionally, not all undergraduate students involved in a SURE program have a substantial stake in the research project. Indeed, student comments on their GEP experiences often stress the importance of being responsible for their own projects, while participating as a team member. Finally, more than 90% of the implementations of the GEP curriculum required students to report their results in some manner to a broader scientific community, either to their class, to others at their institution, or to those assembled for regional or national scientific meetings. This opportunity for students to participate in the scientific community, as scientists do, contributes much toward shifting students from their accustomed roles as observers of the scientific process to active and competent participants in the scientific process.

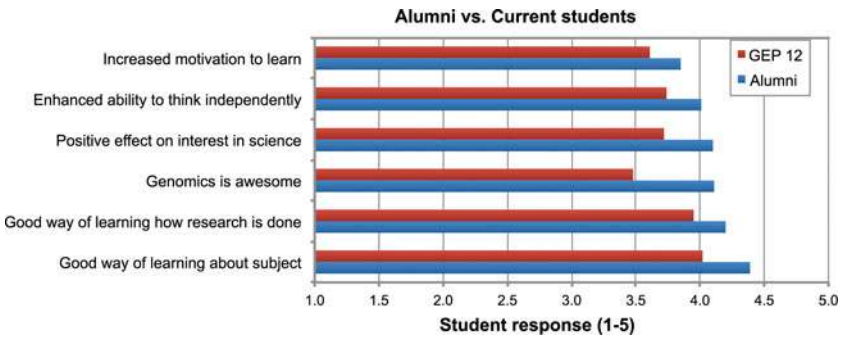

Figure 12. Comparison of alumni and current student responses to six attitudinal survey questions. Scale is 1–5, with 1 being "strongly disagree" and 5 being "strongly agree."

Table 3. Frequency analysis of alumni responses to the question "Please comment on your GEP experience: What was good about it, and what changes would have made it better?"

Comment	Frequency	Examples[a]
Enjoyment (literally saying "I enjoyed the course" or a thought much like it).	42	I enjoyed getting to experience the ownership of a portion of a project (my fosmid) while still being able to have the opportunity to work in a group setting and learn with my peers. I enjoyed participating in the GEP program because it allowed myself and a partner to really take a hands on approach to genomic education. The project was ours, had our name on it and we really felt like we contributed to the science. The process was long and at times difficult but overall, I enjoyed being part of it. I really enjoyed the feeling of participating in a meaning [*sic*] research project. In addition, I think it encouraged active participation in the class and encouraged partners to work diligently together to achieve a common goal.
Research-like (comments about how either the topic or the work made the student feel like a researcher as opposed to a passive student).	10	Working on near-independent projects in the small-group setting really set the course apart from other college experiences; it is certainly the closest a course ever came to emulating an actual research project. I liked that we were doing real, primary research, working directly with original sequence data. It was nice to actually be doing science "first-hand."
Important or significant (similar to faculty comments that it was useful to be engaged in significant research).	6	I liked that what I did actually counted! It was very rewarding to take a class in which your work truly has an impact on the scientific community in comparison to most other classes where the learning is essentially to achieve personal means of learning the information and performing well on the examinations. I thought it was great to contribute to something practical and useful.
Confusion to clarity (comments that the early part of the course was confusing or frustrating but then learning led to clarity and understanding).	6	From what I recall, it was initially confusing as I got familiar with the websites and software required to analyze fosmids. Once I got going, though, it was like solving a puzzle, and I was satisfied with my work when the project was over. I can't think of anything in particular I would like to change. Frustration can easily occur when looking for a your fosmid. However, struggling through this research is also the best part because it is the best way to learn and understand your research. Constantly it was a trial and error for figuring out which key to use or tab, but once I knew how to use it, it was very easy but for the first couple of times it was confusing.
Relevant (was relevant to science or to grad school or career).	8	I think genomics is becoming a more relevant field and every student with an interest in bioinformatics, cell/molecular biology, or genetics should be exposed to the GEP course. I really enjoyed learning how to use all of the online bioinformatics resources. They have been extremely useful in my graduate studies. I think a stronger emphasis on annotation should be made due to the advances in sequencing technology [that] will probably make finishing obsolete in the future. The GEP experience was very helpful in introducing me to the field of bioinformatics and its associated tools. I consistently make use of the skills I gained while taking the course now, during the completion of my PhD program. It also helped to improve communicating scientific data, as I presented my results during the class and will also be a coauthor in an upcoming scientific article.
Realistic (the research was on a "real" or "authentic" problem as opposed to a scripted lab).	16	I liked that we were doing real, primary research, working directly with original sequence data. Gave a chance to get students involved in real-world work in the field of molecular biology and bioinformatics. Interesting and a great way to problem solve. Having a hand in generating real primary research as an undergraduate gave me so much more independence in my view of science. I did feel like I owned my project and was contributing to the greater field of knowledge.

(Continued).

Table 3. Continued

Comment	Frequency	Examples[a]
Independence (working and thinking independently).	22	It really involved a lot of independent thought and problem solving ability. I enjoyed the challenge and I enjoyed the fact that I was contributing to the knowledge base.
		I truly enjoyed working independently on our GEP project. It was the only course I had in my undergraduate career that had novel research findings and did not require work with a partner. It was somewhat difficult to keep a "notebook" during the course, though I believe that this is a reflection of how parts of research are becoming more computational and are not as conducive to a traditional daily log of research.
		Working independently (separate from the professor) on a project where the answers were not previously known was an invaluable lesson about the scientific process.
Learned a lot.	14	I learned more about genomics in this class than I did in any other class I took in school.
		I really did learn an immense amount of information while taking the course. Before taking this course, I had no idea how to annotate genes of any specimen and had never heard of websites such as pubmed, NCBI, flybase, etc. This GEP course really exposed me to a whole new topic within what we had learned in high school. I really enjoyed being able to understand the complex, yet stimulating GEP information in this course.
		I learned a lot about genomics, bioinformatics tools and about research. I think that it was a great experience.
Interesting.	5	The experience that I received increased my interest in bioinformatics.
		It was interesting to learn how to use the interactive tools and fun to discover new genes. It made me feel as if I was a part of something important.
		At first, I was somewhat confused as to what it was that we (my partner and I) had to do, but with time it became more and more interesting and fun.
Teamwork is important.	4	Teamwork was good and the feeling of contributing to science.
		Working as a team with my partner was fun. I learned a lot about sequencing.
		Great experience of working together as a team and the results were perfect.
Negative (comments about time and clarity of instruction).	36	We should have spent more time doing the project. It was frustrating to spend all that time learning how to complete the project and then only doing one fosmid.
		I feel there needed to be more time spent introducing and really explaining the subject matter. I felt thrown into it and it only tangentially related to my course material.
		The entire process should have been explained more. I was not sure what I was doing.

[a]Verbatim sample quotes are shown for each category.

Clearly, we have not yet fully explored questions of scale. Those GEP-affiliated courses that strive for a full research experience are typically laboratory courses that enroll between five and 32 students, or independent study/research courses that enroll two to five students working together. Some GEP faculty members have used the project to provide an introduction to bioinformatics and gene annotation within a midlevel course in genetics enrolling 100–200 students. The instances to date devote only ~10 h to the GEP project, so presumably have a lower impact, as documented above. Anecdotal observations suggest that one trained individual for every six to seven novices is required to avoid the inevitable frustrations that come with learning any new computational system, and implementation on a large scale therefore requires a cadre of instructors. While the participation of one or more senior scientists who can provide perspective and context for the research itself is an important part of the cohort of mentors, much of the mentoring required is procedural and is often best supplied by peer instructors (undergraduate TAs) who can easily communicate the ins and outs of computational infrastructure. How large a class could grow and still show significant student gains is unknown. Given appropriate staffing, and a willingness to devote class/lab time to the effort, it should be possible to expand the research experience to hundreds of students. How best to organize a large cohort to preserve the sense of a learning community and ensure that all students gain a sense of increased efficacy in doing science, has not yet been explored.

What sorts of projects are well suited for a research-based laboratory course that can be implemented at a range of different institutional types? Ideally, one would like to see a range of projects of this type available, to provide a good match with the different research interests of faculty members. Our thoughts have been shaped by discussions with other faculty members who are experimenting with this style of teaching, particularly with HHMI professors Utpal Banerjee (see Chen *et al.*, 2005; Call *et al.*, 2007; Evans *et al.*, 2009), Graham Hatfull (see Hanauer *et al.*, 2006; Jacob-Sera *et al.*, 2012), and Scott Strobel (see Bascom-Slack *et al.*, 2012). The following concepts, highlighted by the work of Hatfull *et al.* (2006, see Table 3 therein), fit many undergraduate programs but apply to bioinformatics particularly well.

The most important characteristic to consider in designing a research-based course is the development of a *parallel project*, one that allows the instructor to teach the students a common set of research tools and approaches while providing individual projects for the students. Ideally, the findings of the group as a whole provide added value to the individual results (e.g., the data will be more meaningful when brought together). For instance, given a starting DNA sequence assembly, a region of interest can be readily divided up among a class into smaller, overlapping projects, and the results subsequently used to reassemble the whole. The combined work of GEP students has allowed us to improve and provide careful annotation of megabase regions of selected eukaryotic genomes, something that could not be efficiently accomplished in any other way. Next, *technical simplicity* must be considered to maintain safety, provide compatibility with scheduling constraints, and ensure a high probability of success. A genomics research project has low start-up costs, requiring only computers with Internet access. There are no safety issues, and access to the project can be provided 24/7. There are no scheduling demands as such—no overnight incubations, no generation times to wait for, and so on. Many experiments, being electronic, can be repeated quickly in real time, so failure does not result in any significant costs or time penalties, and there is a high probability that one will learn something of value in the process. Third, a genomics research project can be *designed at many levels*, depending on the level of the students. For example, college freshmen generally do very well in annotating a prokaryotic genome or the genome of a phage (Hatfull *et al*., 2006; Caruso *et al*., 2009; Harrison *et al*., 2011) while upperclassmen are ready to annotate and analyze a eukaryotic genome with its additional complexities, including multi-exon genes (Leung *et al*., 2010; Shaffer *et al*., 2010). Once students are engaged in the project, they find that there are always more questions to ask. Students taking computer science as well as biology can generate their own programs to address questions of interest. A research course can also be *designed to teach students skills* in accessing various databases and using bioinformatics tools appropriately, with homework or quizzes to test those skills. We find that it is important to utilize various checkpoints (e.g., requiring a report on the first attempt to annotate a gene) to provide feedback to students and develop their confidence. Also, if one *introduces students to the basic tools of the trade* (e.g., EMBOSS, 2013; NCBI, 2013; UCSC Genome Browser, 2013; etc.), the skills they develop can of course be applied subsequently to analysis of other genes and genomes. Importantly, all research has as its goal *publishable original findings*, and this aspect gives students an added sense of responsibility, knowing that other scientists will build on their work; the prospect of a coauthorship provides a tangible reward for their efforts. Even if time does not permit commitment to the research goal, introducing students to the research tools and challenging them to analyze raw data are of value. Finally, the effort involved in *completing one's own project*, while part of a group receiving similar training and struggling with similar challenges, appears to generate student ownership of the projects and creates a dynamic partnership overall.

Similar projects to the *Drosophila* work described here could be constructed around any number of model organisms to provide an accurate annotation of genomic regions of special interest and/or to address specific questions concerning the organization and evolution of genes and genomes (for additional examples, see Kerfeld and Simons, 2007; Ditty *et al*., 2010; Goff *et al*., 2011; Banta *et al*., 2012; Singer *et al*., 2013). A consortium clearly has cost/benefit advantages, and we believe that more national projects of this type should be supported by agencies interested in improving undergraduate science education in the United States in partnership with those interested in the daunting task of improving the utility of large data sets such as genomic sequences. Both the need for and the value of bona fide research experiences for undergraduates are well established, and the success of the GEP experience confirms that such experiences can be provided for students economically in terms of both time and money. The continued success of such efforts generally encourages us to believe that such in-class research will become an increasingly common part of the undergraduate science experience.

ACKNOWLEDGMENTS

We thank the many students who have participated in GEP-affiliated courses since 2006, particularly those who have served as TAs. We also thank Frances Thuet for her work in setting up the assessment websites and helping to collect those data; the many Washington University undergraduates and the staff members of the Genome Institute who have served as TAs in the GEP workshops and courses; and the additional Washington University staff who have helped to organize and facilitate these meetings. S.C.R.E. thanks Graham Hatfull (University of Pittsburgh), Scott Strobel (Yale University), and Utpal Banerjee (University of California–Los Angeles) for thoughtful discussion of research projects for undergraduates. This project has depended on materials received through the Drosophila Genomics Resource Center and continual access to FlyBase (2013), as well as tools provided by the NCBI (2013). All GEP materials, except the online quizzes, are available under a "share-alike" type license at the GEP website: http://gep.wustl.edu/about/legalnotices. See section on *Faculty Implementation Strategies* and the Supplemental Material (Text S1) for suggestions on the use of these materials. College and university faculty interested in joining the *Drosophila* research project should contact S.C.R.E. at selgin@biology.wustl.edu.

This work was supported by grant 52007051 from the Howard Hughes Medical Institute to S.C.R.E. under the Professors Program, by grant 2U54 HG00307910 from the National Human Genome Research Institute (Richard K. Wilson, principal investigator), and by Washington University in St. Louis. None of the above funders had any role in the design or conduct of the study; nor in the collection, analysis, or interpretation of the data; nor in the preparation, review, or approval of the manuscript.

REFERENCES

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC. http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf (accessed 4 July 2013).

Anderson LW, Krathwohl DR (eds.) (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Boston, MA: Allyn & Bacon.

Banta LM *et al.* (2012). Integrating genomics research throughout the undergraduate curriculum: a collection of inquiry-based genomics lab modules. CBE Life Sci Educ *11*, 203–208.

Bascom-Slack CA, Arnold AE, Strobel SA (2012). Student-directed discovery of the plant microbiome and its products. Science *338*, 485–486.

Bauer KW, Bennett JS (2003). Alumni perceptions used to assess undergraduate research experience. J High Educ *74*, 210–230.

Baylor College of Medicine (2012). Human Genome Sequencing Center Drosophila modENCODE Project. www.hgsc.bcm.edu/drosophila-modencode-project (accessed 27 January 2014).

Bennett D *et al.* (2007). Experiences That Matter: Enhancing Student Learning and Success, NSSE Annual Report 2007. http://nsse.iub.edu/nsse_2007_annual_report (accessed 5 August 2013).

Bloom BS, Krathwohl DR (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, by a Committee of College and University Examiners. Handbook I: Cognitive Domain, New York: Longmans, Green.

Boomer S, Lodge D, Dutton B (2002). Bacterial diversity studies using the 16S rRNA gene provide a powerful research-based curriculum for molecular biology laboratory. Microbiol Educ *3*, 18–25.

Brenner S (2002). Life sentences: ontology recapitulates philology. Genome Biol *3*, comment1006.

Brodl M (2005). Tapping recent alumni for the development of cutting-edge, investigative teaching laboratory experiments. Bioscene *31*, 13–20.

Campbell AM, Ledbetter ML, Hoopes LL, Eckdahl TT, Heyer LJ, Rosenwald A, Fowlks E, Tonidandel S, Bucholtz B, Gottfried G (2007). Genome Consortium for Active Teaching: meeting the goals of BIO2010. CBE Life Sci Educ *6*, 109–118.

Call GB *et al.* (2007). Genome-wide clonal analysis of lethal mutations in the *Drosophila melanogaster* eye: comparison of the X chromosome and autosomes. Genetics *177*, 689–697.

Caruso S, Sandoz J, Kelsey J (2009). Non-STEM undergraduates become enthusiastic phage-hunters. CBE Life Sci Educ *8*, 278–282.

Chen J *et al.* (2005). Discovery-based science education: functional genomic dissection in *Drosophila* by undergraduate researchers. PLoS Biol *3*, e59.

Clark AG *et al.* (2007). *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature *450*, 203–218.

Ditty JL *et al.* (2010). Incorporating genomics and bioinformatics across the life sciences curriculum. PLoS Biol *8*, e1000448.

Drew J, Triplett E (2008). Whole genome sequencing in the undergraduate classroom: outcomes and lessons from a pilot course. J Microbiol Biol Educ *9*, 3–11.

Dunbar D, Harrison M, Mageeney C, Catagnus C, Cimo A, Beckowski C, Ratmansky L (2012). The rewards and challenges of undergraduate peer mentoring in course-based research: student perspectives from a liberal arts institution. Perspect Undergrad Res Mentoring *1*, 1–8.

Elwess NL, Latourelle SM (2004). Inducing mutations in paramecium: an inquiry-based approach. Bioscene *30*, 25–35.

EMBOSS (2013). EBI: European Bioinformatics Institute home page. www.ebi.ac.uk/Tools/emboss (accessed 6 July 2013).

Emerson JA, Key SCS, Alvarez CJ, Mel S, McNeil G, Saville KJ, Leung W, Shaffer CD, Elgin SCR (2013). Introduction to the Genomics Education Partnership and collaborative genomics research in Drosophila. In: Association for Biology Laboratory Education: Tested Studies for Laboratory Teaching, Proceedings of the Association for Biology Laboratory Education, vol. 34, pp. 135–165. www.ableweb.org/volumes/vol-34/v34reprint.php?ch=6 (accessed 27 January 2014).

Evans C *et al.* (2009). G-TRACE rapid Gal4-based cell lineage analysis in *Drosophila*. Nat Methods *6*, 603–605.

FlyBase (2013). FlyBase: A Database of *Drosophila* Genes and Genomes home page. http://flybase.org (accessed 6 July 2013).

GenBank (2013). NCBI GenBank home page. www.ncbi.nlm.nih.gov/genbank (accessed 6 July 2013).

Genomics Education Partnership (GEP) (2013a). GEP home page. http://gep.wustl.edu (accessed 3 July 2013).

GEP (2013b). UCSC Genome Browser Mirror Home Page. http://gander.wustl.edu/ (accessed 6 July 2013).

Goff SA *et al.* (2011). The iPlant collaborative: cyber infrastructure for plant biology. Front Plant Sci *2*, 34.

Goins GD, White CD, Foushee DB, Smith MA, Whittaker JJ, Byrd G (2009). A multifaceted pipeline to success for undergraduates pursuing bioscience degrees. In: Successful Models for Effectively Retaining and Graduating Students, New York: Thurgood Marshall College Fund.

Hanauer DI, Jacobs-Sera D, Pedulla ML, Cresawn SG, Hendrix RW, Hatfull GF (2006). Teaching scientific inquiry. Science *314*, 1880–1881.

Harrison M, Dunbar D, Mageeney C, Lopatto D (2010). Peer mentoring in an introductory biology laboratory. CUR Q *31*, 9–14.

Harrison M, Dunbar D, Ratmansky L, Boyd K, Lopatto D (2011). Classroom-based science research at the introductory level: changes in career choices and attitude. CBE Life Sci Educ *10*, 279–286.

Hatfull GF *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an education platform. PLoS Genet *2*, e92.

Hathaway RS, Nagda BA, Gregerman SR (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study. J Coll Stud Dev *43*, 614–631.

Hingamp P, Brochier C, Talla E, Gautheret D, Thieffry D, Herrmann C (2008). Metagenome annotation using a distributed grid of undergraduate students. PLoS Biol *6*, e296.

Jacob-Sera D *et al.* (2012). On the nature of mycobacteriophage diversity and host preference. Virology *434*, 187–201.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Hausslet D (2002). The human genome browser at UCSC. Genome Res *12*, 996–1006.

Kerfeld CA, Simons RW (2007). The Undergraduate Genomics Research Initiative. PLoS Biol *5*, e141.

Laursen S, Hunter A, Seymour E, Thiry H, Melton G (2010). Undergraduate Research in the Sciences: Engaging Students in Real Science, San Francisco: Jossey-Bass.

Leung W *et al.* (2010). Evolution of a distinct genomic domain in *Drosophila*: comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*. Genetics *185*, 1519–1534.

Locks A, Gregerman S (2008). Undergraduate research as an institutional retention strategy: the University of Michigan model. In: Creating Effective Undergraduate Research Programs in Science, ed. R Taraban and RL Blanton, New York: Teachers College Press, 11–32.

Lopatto D (2004). Survey of Undergraduate Research Experiences (SURE): first findings. Cell Biol Educ *3*, 270–277.

Lopatto D (2006). Undergraduate research as a catalyst for liberal learning. Peer Rev *8*, 22–25.

Lopatto D (2007). Undergraduate research experiences support science career decisions and active learning. CBE Life Sci Educ *6*, 297–306.

Lopatto D (2009). Science in Solution: The Impact of Undergraduate Research on Student Learning, Tucson, AZ: Research Corporation for Science Advancement. web.grinnell.edu/sureiii/science_in_solution_lopatto.pdf (accessed 27 January 2014).

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

Nagda BA, Gregerman SR, Jonides J, Hippel WV, Lerner JS (1998). Undergraduate student-faculty research partnerships affect student retention. Rev High Educ *22*, 55–72.

National Center for Biotechnology Information (2013). NCBI home page. www.ncbi.nlm.nih.gov (accessed 6 July 2013).

National Research Council (2003). BIO2010: Transforming Undergraduate Education for Future Research Biologists. Washington,

DC: National Academies Press. http://books.nap.edu/catalog/10497.html (accessed 6 July 2013).

Oleksyk TK *et al.* (2012). A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. GigaScience *1*, 14.

President's Council of Advisors on Science and Technology (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering and Mathematics, Washington, DC: Executive Office of the President. www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf (accessed 4 July 2013).

Riddle NC, Shaffer CD, Elgin SCR (2009). A lot about a little dot: lessons learned from *Drosophila melanogaster* chromosome 4. Biochem Cell Biol *87*, 229–241.

Riddle NC *et al.* (2012). Enrichment of HP1a on Drosophila chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. PLoS Genet *8*, e1002954.

Seymour E, Hunter A-B, Laursen SL, DeAntoni T (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. Sci Educ *88*, 493–534.

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.

Singer SR *et al.* (2013). Keeping an eye on biology. Science *339*, 408–409.

UCSC Genome Browser (2013). UCSC genome bioinformatics site. http://genome.ucsc.edu (accessed 6 July 2013).

U.S. News & World Report Staff (2011). U.S. News Ultimate College Guide 2011. New York: U.S. News & World Report, L.P.