JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION, May 2010, p. 38-41
Copyright © 2010 American Society for Microbiology DOI: 10.1128/jmbe.v11.i1.133

Perspectives

"Shovel-ready" Sequences as a Stimulus for the Next Generation of Life Scientists

Michael D. Boyle*
Department of Biology, Juniata College, Huntingdon, PA 16652

Genomics and bioinformatics are dynamic fields well-suited for capturing the imagination of undergraduates in both research laboratories and classrooms. Currently, raw nucleotide sequence is being provided, as part of several genomics research initiatives, for undergraduate research and teaching. These initiatives could be easily extended and much more effective if the source of the sequenced material and the subsequent focus of the data analysis were aligned with the research interests of individual faculty at undergraduate institutions. By judicious use of surplus capacity in existing nucleotide sequencing cores, raw sequence data could be generated to support ongoing research efforts involving undergraduates. This would allow these students to participate actively in discovery research, with a goal of making novel contributions to their field through original research while nurturing the next generation of talented research scientists.

Reforming undergraduate science education continues to be the focus of many stakeholders including government agencies, foundations, professional societies, faculty, and students. In biology education over the past decade, reforms have been driven, in part, by the perception that US students are losing their competitive advantage in science, technology, engineering, and mathematics [STEM] (2, 13, 15). Most of these reforms have focused on changing the traditional emphasis on lectures and memorization to more student–centered approaches (2, 13, 15). The new emphasis promotes pedagogical approaches that utilize small groups, case studies, and open-ended laboratories with a focus on critical thinking and problem-solving (2).

The value of involving undergraduates in basic research with faculty mentors is well documented across STEM disciplines and among different types of academic institutions. Evidence supporting the effectiveness of these differing pedagogical strategies is provided in large surveys of student attitudes conducted by the National Survey for Student Engagement (NSSE). Analysis of these data has led to the conclusion that undergraduate research is a high-impact educational practice (9). Similar conclusions have also been drawn from smaller studies using additional measures such as: continuation in science at the graduate level and improved basic knowledge in biology when compared to peer groups who did not participate in research (12, 16). Focus groups also found that students doing research had a greater interest in and appreciation of science (10).

Reports like Bio2010 (2) and the current NSF-AAAS Transforming Undergraduate Biology Education initiative (http://www.visionandchange.org/index.php/inv_conference/) have been quick to recognize that biology is a rapidly expanding field driven in significant part by innovations in technology (e.g. PCR, microarrays, DNA sequencing). These reports also recognize the value of undergraduate research and recommend that more undergraduate students should be engaged in research involving cutting-edge techniques. Implementing these recommendations is challenging because of the high cost and limited access to critical equipment and associated resources.

Bioinformatics has been recognized as a practical approach that allows undergraduates access to research projects using existing nucleotide sequence available in public genome data bases. This bypasses one of the major hurdles – the expense of facilities and modern equipment to generate analyzable data. Empowering faculty with effective pedagogical approaches to use existing sequence data has been the focus of several groups including the American Society for Microbiology in collaboration with the Department of Energy (http://www.facultyprograms.org/page02a.shtml), the Genome Education Partnership (5, 11, 14), and others (6, 7, 8). In a related initiative ("Teaching Big Science at Small Colleges"), a consortium of small colleges (Barnard, Carleton, Vassar, and Williams) is actively developing a new genomics curriculum that includes inquiry-based units that can be integrated into existing courses (http://serc.carleton.edu/genomics/index.htm). These groups have provided resource materials and/or training workshops to facilitate introduction of bioinformatic exercises and modules into the classroom.

Recently, Silverthorn and colleagues (17) evaluated how effectively teaching modules developed as part of the integrative themes in physiology curricular project could be incorporated into existing courses of interested faculty who had not been directly involved in the generation of these teaching materials. The results of their analysis identified significant problems with integrating these teaching modules

---

*Corresponding author. Mailing address: Department of Biology, Juniata College, Department of Biology, 1700 Moore Street Huntingdon, PA 16652. Phone: (814) 641-3553. Fax: (814) 641-3685. E-mail: Boyle@juniata.edu.

into existing courses at other institutions (17). The problems with integration could be attributed to multiple factors including lack of academic support, class size, insufficient time to implement change, and other work-load issues, as well as other curricular concerns, IRB and personal factors (17). As a consequence, about half of the faculty who initially agreed to test the modules in their courses withdrew before implementing the study. Another 22% changed a module before using it. Among the insights provided in this study was the importance to faculty of comfort with materials they use to teach (17). This may explain why a sense of ownership in the development of teaching material is important.

In the context of genome sequencing and bioinformatic analysis of data, student involvement and perceived ownership have been reported to enhance student interest and performance (10). To this end, Howard Hughes Medical Research Institute's (HHMI's) Science Education Alliance program (www.hhmi.org/grants/sea/), the Department of Energy's (DOE's) "adopt-a-genome" project (7), the Genomics Education Partnership (5, 11, 14), and others (6, 10) have sought ways to put raw original genome sequence in the hands of undergraduates. The HHMI program focuses on support for a limited number of institutions, first to isolate bacteriophages in their local environment and then to determine the genome sequence for selected phage. The DOE program offers undergraduate institutions an opportunity to conduct original research involving the annotation and bioinformatic analysis of raw data from recently-sequenced genomes or to nominate an organism to be sequenced (7). The Genomics Education Partnership currently focuses on projects centered around the dot chromosome of Drosophila, and participating undergraduates learn to move from preliminary sequence to high-quality finished sequence, and to annotate genes and other features (5, 11, 14). At UCLA, Kerfeld and colleagues developed an undergraduate research initiative (http://www.lsic.ucla.edu/ugri/) that provides large numbers of students (n = ~2,000/year) the opportunity to participate in a sequencing project that is an integral link between lower and upper level courses in the biological sciences. The impact on students of this initiative has been extensively assessed using a variety of methodologies, demonstrating that the initiative has a positive outcome for student learning (10).

Central to each of these programs is easy access to > 2,000 free bioinformatic analysis sites to allow the data to be analyzed. [For a comprehensive listing of bioinformatic software programs, see http://www.hsls.pitt.edu/guides/genetics/obrc/.] If faculty (or students) are currently unfamiliar with bioinformatic tools, existing professional development support is available through a number of well-established workshops (e.g. http://bioquest.org/bedrock/) and online courses and additional resources (http://www.ncbi.nlm.nih.gov/Class/minicourses/). Consequently, any teaching strategy designed to use existing nucleotide sequence for developing undergraduate classroom activities and research projects can be seamlessly adapted to the use of novel sequence data.

While programs providing novel sequence data are a valuable starting point for engaging and training future life scientists, they currently support only a limited number of projects. Furthermore, the direction of the research is predetermined by the organisms sequenced (in the DOE's 'adopt-a-genome-"project), a focus on bacteriophages (in the HHMI program), or by the region of the Drosophila genome (in the Genomics Education Partnership study). The value in providing sequence as a teaching tool and for undergraduate research could be extended – and potentially be much more effective – if it could be linked to the existing research of faculty at undergraduate institutions. Investigator-requested sequence would empower many individuals who could use nucleotide sequence in their individually selected research areas to ask new scientific questions and test hypotheses with the help of undergraduates.

Recognizing the value of this approach, Cofactor Genomics initiated a classroom project that provides investigator-requested sequence, on a competitive basis, for educational uses in high school and undergraduate science classes (http://www.cofactorgenomics.com/home/about/cofactor-gives-back/). Cofactor received more applications than it could support, which would further suggest that many undergraduate faculty and high school science teachers recognize the educational value of having their students work with raw sequence data. The information that students extract from analysis of novel sequence information would represent true discovery and, as shown previously, could lead to publications and presentations to the scientific community (1, 18, 19) as well as enhancing the community-wide knowledge present in genome data bases.

With recent technological advances and innovations in high-throughput genome sequencing, a single run on a state-of-the-art sequencer would generate sufficient data to provide an undergraduate college professor a decade's worth of mining operations for students. Currently, there are many government laboratories and NIH-supported sequencing cores that could easily provide 10-20 gigabases of sequence a year to support a wider genome initiative in the context of undergraduate education. An existing organization, which focuses on innovations of teaching biology at the undergraduate level, such as the Genome Consortium for Active Teaching [GCAT] (4), could act as the review body to prioritize sequencing requests. The potential criteria for selection might include the novelty of the project, the number of students that the project would impact, and the commitment of the investigator to contribute the research findings to an appropriate database, and would encourage students to present and publish their findings. In addition, the investigator would be asked to provide an assessment plan to measure the impact of the experience on the students who were involved directly (via primary data analysis) or indirectly (via classroom activities).

Many areas of research in the life sciences are utilizing sequencing approaches at the cutting edge of their research fields. Consequently, engaging undergraduates in the analysis of sequence data is worthwhile and relevant to multiple subdisciplines within the life sciences. This diversity of application of core sequencing technology can be readily modeled at small colleges like Juniata. We have ten faculty members in the Biology Department. Four of us have identified specific research projects that could benefit from project-specific sequence data and two others have indicated a possible future interest. The four potential projects that have been proposed (3) and for which we would like to have raw sequence are:

- Understanding vertebrate adaptation and speciation in a natural fish model
- Transcriptome analysis of amphipod populations under going differential predation selection
- Measuring yeast retrotransposition using whole genome analysis
- Genomic analysis of invasive and noninvasive variants of S. *pyogenes* using a whole genome subtractive hybridization approach

Each research area parallels the teaching responsibilities of the faculty involved, and the sequence data would be used for both primary research purposes and classroom learning opportunities. The impact on the student of seeing these tools being applied by a number of different faculty from diverse fields would predictably reinforce common core themes and concepts in biology. The predicted value on student learning and integration of concepts across courses in biology is, in turn, a testable hypothesis.

Thus, with a relatively modest "stimulus" investment within the existing scientific infrastructure (sequencing cores and bioinformatic software already in the public domain), individual faculty at undergraduate campuses could allow their existing passions for a scientific research area to be moved to leading edge discovery research. Inevitably, analysis of the data will result in developing new hypotheses, and the excitement of discovery and the implications of new knowledge would permeate into the classroom. The involvement of undergraduates in such an exciting environment would be greatly facilitated by using data relevant to the research passion of the teacher, rather than merely imported or recycled from some unrelated area. The enthusiasm of faculty mentors is a powerful magnet for engaging undergraduates in a research project. Providing undergraduate educators with "shovel-ready" sequence that supported their research goals would energize them and expose undergraduates to the excitement, passion, and love of science in the context of state-of-the-art technology. This, in turn, would allow us to meet the President's challenge, offered during his address to the National Academy of Sciences in April of 2009, "to use [their] love and knowledge of science to spark the same sense of wonder and excitement in a new generation".

ACKNOWLEDGEMENTS

The author would like to thank his colleagues in the Biology Department at Juniata College, especially Dr. Vince Buonaccorsi and Dr. Andrew Fletcher for their contributions in developing the ideas and content of this article.

REFERENCES

1. **Anders, K.R., J.R. Kudrna, K.E. Keller, B. Kinghorn, E.M. Miller, D. Pauw, et al.** 2009. A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. BMC Genetics. **10:**36.
2. BIO2010: Transforming undergraduate education for future research biologists. 2003. National Research Council. The National Academies Press, Washington, D.C.
3. **Boyle, M.D., Bennett, R., Keeney J., Buonaccorsi V.** Transforming Undergraduate Education Through Increased Faculty Access to NextGen Sequencing Runs. Abstract #90. Vision and change in undergraduate education. (2009). http://visionandchange.org/files/2010/03/V-CPoster-Abstracts_2009.pdf
4. **Campbell, A.M., M.L. Ledbetter, L.L. Hoopes, T.T. Eckdahl, L.J. Heyer, A. G. Rosenwald, et al.** 2007. Genome consortium for active teaching: Meeting the goals of BIO2010. CBE Life Sci. Educ. **6:**109–118.
5. **Chen, J., G.B. Call, E. Beyer, C. Bui, A. Cespedes, A. Chan, et.al.** 2005. Discovery-based science education: Functional genomic dissection in Drosophila by undergraduate researchers. PLoS Biol. **3:**e59.
6. **Drew, J.C. and E.W. Triplett.** 2009. Whole genome sequencing in the undergraduate classroom: Outcomes and lessons from a pilot course. J. Microbiol. and Biol. Ed Ed.9: (2008). http://www.microbelibrary.org/Edzine/details.asp?id=2733&Lang [accessed 12 December 2009].
7. **Kerfeld, C.A.** 2009. The joint Genome Institute's Microbial Genome Annotation Program for Undergraduates. FASEB J. **23:**84.2.
8. **Kerfeld, C.A. and R.W. Simons.** 2007. The undergraduate genomics research initiative. PLoS Biol. **5:**e141.
9. **Kuh, G. D.** 2008. High-impact educational practices: What they are, who has access to them, and why they matter. AAC&U Publications, Washington, D.C.
10. **Levis-Fitzgerald, M., N. Denson, and C.A. Kerfeld.** 2005. Undergraduate students conducting research in the life sciences: Opportunities for connected learning. Proceedings of the 2005 Association for the Study of Higher Education Conference, 17-19 Nov., 2005, Philadelphia, P.A.
11. **Lopatto, D., C. Alvarez, D. Barnard, C. Chandrasekaran, H.M. Chung, C. Du, et al.** 2008. Undergraduate research. Genomics Education Partnership. Science. **322:**684–685.
12. **Lopatto, D.** 2007. Undergraduate research experiences support science career decisions and active learning. CBE Life Sci. Educ. **6:**297–306.
13. **Steen, L.A. (ed).** 2005. Math & Bio 2010: Linking Undergraduate Disciplines. Mathematical Association of America, Washington, D.C.
14. **Pfund, C., S. Miller, K. Brenner, P. Bruns, A. Chang, D. Ebert-May, et al.** Professional development: Summer institute to improve university science teaching. Science. **324:**470–471.

15. **PKAL.** PKAL Report on Reports II: Recommendations for Urgent Action, 2006. Transforming America's Scientific and Technological Infrastructure. Project Kaleidoscope, Washington, D.C.
16. **Russell, S.H., M.P. Hancock, and J. McCullough.** 2007. The pipeline. Benefits of undergraduate research experiences. Science. **316:**548–549.
17. **Silverthorn, D.U., M. Patti. P.M. Thorn, and M.D. Svinicki.** 2006. It's difficult to change the way we teach: Lessons from the Integrative Themes in Physiology curriculum module project. Adv. Physiol. Educ. **30:**204-214.
18. **Slater, S.C., B.S. Goldman, B. Goodner, J.C. Steubal, S.K.Farrand, E.W. Nester, et al.** 2009. Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. J Bacteriol. **191:**2501–2511.
19. **Slawson, E,E,, C.D. Shaffer, C.D. Malone, W. Leung, E. Kellmann, R.B. Shevchek, et.al.** 2006. Comparison of dot chromosome sequences from D. melanogaster and D. virilis reveals an enrichment of DNA transposon sequences in heterochromatic domains. Genome Biol. **7:**R15.