

## Feature Meeting Report

# GCAT-SEEKquence: Genome Consortium for Active Teaching of Undergraduates through Increased Faculty Access to Next-Generation Sequencing Data

Vincent P. Buonaccorsi,\* Michael D. Boyle,\* Deborah Grove,<sup>†</sup> Craig Praul,<sup>†</sup> Eric Sakk,<sup>‡</sup> Ash Stuart,<sup>§</sup> Tammy Tobin,<sup>||</sup> Jay Hosler,\* Susan L. Carney,<sup>¶</sup> Michael J. Engle,<sup>#</sup> Barry E. Overton,<sup>@</sup> Jeffrey D. Newman,\*\* Marie Pizzorno,<sup>††</sup> Jennifer R. Powell,<sup>‡‡</sup> and Nancy Trun<sup>§§</sup>

\*Juniata College, Huntingdon, PA 16652; <sup>†</sup>Pennsylvania State University, State College, PA 16802; <sup>‡</sup>Morgan State University, Baltimore, MD 21251; <sup>§</sup>Ramapo College of New Jersey, Mahwah, NJ 07430; <sup>||</sup>Susquehanna University, Selinsgrove, PA 17870; <sup>¶</sup>Hood College, Frederick, MD 21701; <sup>#</sup>Mount Aloysius College, Cresson, PA 16630; <sup>@</sup>Lock Haven University, Lock Haven, PA 17745; <sup>\*\*</sup>Lycoming College, Williamsport, PA 17701; <sup>††</sup>Bucknell University, Lewisburg, PA 17837; <sup>‡‡</sup>Gettysburg College, Gettysburg, PA 17325; <sup>§§</sup>Duquesne University, Pittsburgh, PA 15282

To transform undergraduate biology education, faculty need to provide opportunities for students to engage in the process of science. The rise of research approaches using next-generation (NextGen) sequencing has been impressive, but incorporation of such approaches into the undergraduate curriculum remains a major challenge. In this paper, we report proceedings of a National Science Foundation–funded workshop held July 11–14, 2011, at Juniata College. The purpose of the workshop was to develop a regional research coordination network for undergraduate biology education (RCN/UBE). The network is collaborating with a genome-sequencing core facility located at Pennsylvania State University (University Park) to enable undergraduate students and faculty at small colleges to access state-of-the-art sequencing technology. We aim to create a database of references, protocols, and raw data related to NextGen sequencing, and to find innovative ways to reduce costs related to sequencing and bioinformatics analysis. It was agreed that our regional network for NextGen sequencing could operate more effectively if it were partnered with the Genome Consortium for Active Teaching (GCAT) as a new arm of that consortium, entitled GCAT-SEEK(quence). This step would also permit the approach to be replicated elsewhere.

## THE CHALLENGE OF TRANSFORMING UNDERGRADUATE EDUCATION

The Vision and Change deliberations on transforming undergraduate biology education recently articulated a need to

DOI: 10.1187/cbe.11-08-0065

Address correspondence to: Vincent P. Buonaccorsi (buonaccorsi@juniata.edu).

© 2011 V. P. Buonaccorsi *et al.* CBE—Life Sciences Education © 2011 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.



engage students in the process of science and to present science as a vibrant and active field (American Association for the Advancement of Science, 2010). Extensive pedagogic research concludes that participation in open-ended research endeavors fosters a sense of ownership over a biological subject, and enhances teaching and learning in biological sciences (Teagle Foundation, 2007). Developing innovative cross-disciplinary approaches and empowering faculty with the tools to implement novel strategies remains a challenge at all levels of undergraduate education.

In the last 5 yr, the rise of next-generation (NextGen) sequencing approaches in addressing biological problems has been spectacular, but incorporating NextGen sequencing data for active teaching in the undergraduate curriculum remains a major challenge. For faculty at small and medium-sized institutions of higher education, high teaching loads, lack of access to state-of-the-art equipment, and budgetary constraints typically conspire to inhibit faculty from considering NextGen sequencing in their own experiments. High capital costs, extraordinarily high rates of technological change, and daunting computational and analytical requirements make the technology exceptionally challenging to assimilate into the undergraduate curriculum.

The Genome Consortium for Active Teaching (GCAT; Campbell *et al.*, 2006) was developed a decade ago to meet similar challenges in relation to the use of microarrays in undergraduate biology education. GCAT offers highly discounted microarray chips and array scanning and a supporting network of faculty expertise to educators working with undergraduates. In one decade, the effort has trained over 360 faculty and 24,000 undergraduates in the use and interpretation of microarray data. The newly trained students were enrolled in primarily undergraduate institutions, including those that historically serve underrepresented populations. GCAT has met many of the goals of the *BIO2010* report (Campbell *et al.*, 2007), and recently expanded its focus into synthetic biology (Wolyniak *et al.*, 2010). Now that NextGen sequencing is rapidly superseding microarray technology for a variety of technical and economic considerations, GCAT and others recognized the need to find cost-effective and innovative strategies to facilitate active teaching of NextGen technology at the undergraduate level. Understanding the advantages and limitations of continually evolving transformative technologies like NextGen sequencing is essential preparation for future life scientists, medical professionals, and, indeed, a scientifically literate citizenry, as the age of personalized medicine moves toward becoming reality. In addition, analyzing raw sequence data provides students with learning opportunities that underscore interdisciplinary concepts central and relevant to studies of all forms of life.

## THE MEETING

In this paper, we report proceedings of a workshop from a National Science Foundation (NSF)-funded incubator grant for research coordination networks for undergraduate biology education (RCN/UBE). The network aimed to collaborate with a centrally located genome-sequencing core fa-

cility at the Pennsylvania State University (PSU) to enable undergraduate students and faculty in the mid-Atlantic region to access state-of-the-art sequencing technology. Initial network participants included Juniata College, Susquehanna University, Duquesne University, Hampton University, Morgan State University, Ramapo College of New Jersey, Gettysburg College, Lycoming College, Lock Haven University, Mount Aloysius College, Bucknell University, and Hood College, with the genome-sequencing facility at PSU supporting the data acquisition and dissemination aspects of the initiative. The meeting was held July 11–14, 2011, at Juniata College and PSU and included the individuals who helped write the incubator grant and invited speakers Malcolm Campbell (Davidson College), Anton Nekrutenko (PSU), Istvan Albert (PSU), and Bill Morgan (College of Wooster). Through presentations and periodic whole-group discussions, we worked together to exchange ideas to develop a structure to approach the problem of introducing NextGen sequencing to undergraduates.

During the course of the meeting, a number of parallels emerged between the thinking of the participants and the philosophy of GCAT. Members of both groups valued the academic freedom provided by their ability to choose and direct their own research and scholarly activities. Both groups recognized the value of communal support from colleagues at similar small institutions to help compensate for lack of a critical mass of peers on each campus. Both groups recognized the strategic need to partner with other groups, like the microarray manufacturers in GCAT's initial plan, or genome-sequencing facilities, such as PSU. All of these considerations suggested that the mid-Atlantic network for NextGen sequencing could operate more effectively and enable the approach to be replicated elsewhere if it partnered with GCAT as a new arm of that consortium. GCAT has established an efficient dissemination strategy through its website and listserv, and many of the members currently using microarrays will be poised to transition to NextGen sequencing as it replaces gene chip technology. Our shared values and the success of the RCN/UBE grant led to an agreement with Malcolm Campbell for our RCN/UBE to become GCAT-SEEK(quence) and to complement another GCAT initiative in the emerging field of synthetic biology (GCAT-SynBio; Wolyniak *et al.*, 2010).

At our network meeting, we formed a nascent community of biologists from distinct areas (e.g., molecular biology, environmental science, plant biology, microbiology) aiming to develop parallel research studies in the scholarship of teaching and learning and discovery science. The specific goals of the workshop were to: 1) learn lessons from the GCAT model; 2) learn the scope of NextGen sequencing technology, applications, and analysis; 3) develop common learning goals for students using this technology and develop appropriate methods of assessment and; 4) develop goals and an administration plan for the network.

Malcolm Campbell presented the keynote address on GCAT, describing lessons learned from administration of the consortium. In particular, he emphasized the importance of an undergraduate focus, inclusion of minority-serving institutions, assessment of educational activities, advertisement of the network, faculty development, and taking on the most difficult problems related to a technology to make the

network valuable. The GCAT model also recognized the importance of the investigator retaining ownership in the direction of his or her research. This was one of the key reasons for adopting a model in which an investigator requested the raw sequence data related to their research expertise and passion. (For a detailed discussion of these considerations, see Boyle [2010].) The GCAT-SEEK network will periodically request proposals using its listserv.

The meeting included talks on NextGen technology and bioinformatics. Deb Grove and Craig Praul (codirectors of the PSU Genome Core Facility) detailed the latest sequencing technologies, applications, and costs associated with their Ion Torrent PGM, Roche 454FLX, and SOLiD 5500XL platforms. These are massively parallel DNA-sequencing machines capable of providing hundreds of millions to tens of billions of nucleotides of DNA sequence data in about a week. The resulting data must be processed using specialized bioinformatics techniques that may require high-powered computers. It was determined that up to 50% of costs could be cut when individual researchers cooperate to share sequencing runs. Unique DNA sequence adapters (bar codes) may be ligated onto an investigator's DNA fragments before sequencing. This process allows each investigator's samples to be individually labeled, pooled with other samples, and automatically separated after bulk sequencing. PSU Biochemistry and Molecular Biology faculty Anton Nekrutenko and Istvan Albert framed challenges and approaches to NextGen data analysis. Anton Nekrutenko described the Galaxy bioinformatics analysis framework that he and his colleagues developed (Blankenberg *et al.*, 2010; Goecks *et al.*, 2010). He emphasized the importance of the evolutionary underpinnings of bioinformatics analysis, and how comparison is key to understanding genomes. Istvan Albert, director of the Bioinformatics Consulting Center at PSU, suggested that bioinformatics analysis is challenging, because it is a highly interdisciplinary science incorporating information technology to manage data, computer science to analyze data, statistics to find meaningful patterns, and biology to form relevant hypotheses. He stressed that bioinformatics cannot be learned passively from a book but requires active-learning approaches and student commitment.

The challenge facing undergraduate faculty in introducing bioinformatics was addressed by Ash Stuart (Ramapo College of New Jersey), Eric Sakk (Morgan State University), and Bill Morgan (College of Wooster), who is working on a related initiative in genomics education. The effectiveness of interdisciplinary approaches, open-ended inquiry, case studies, online student learning communities, undergraduate conferences, exchange of students between schools, and interaction among students in different disciplines on a single campus all had the potential to improve student communication, collaboration, and leadership skills. It was stressed that one of the desired skills students should acquire was adaptability, because the field of bioinformatics changes so rapidly. Having a forum to discuss the impact of genomic sciences and bioinformatics analysis using examples from the daily news was discussed as a way to connect the science to a broader societal context. Bill Morgan discussed his progress toward a free, online genomics textbook focused on interactive case studies with mathematical sidebars and modeled after the text, *Discovering Genomics, Proteomics, and Bioinformatics*, by

Campbell and Heyer (2006). He has assembled a group of faculty members from throughout the United States with expertise in genomics to help coordinate the development of learning modules in 10 topic areas. The workshop participants were particularly interested in reviewing the learning objectives for the genome-sequencing topic. It was suggested that a template with learning objectives, protocols, and assessment instruments be developed for the next-generation sequencing module that would encourage faculty adoption because it would allow customization of its activities to the datasets of individual investigators.

No innovation in education can be considered successful if it is not subjected to rigorous evaluation and assessment in the context of defined learning objectives. The workshop participants discussed the learning objectives the network should have, and formed an assessment leadership team chaired by Tammy Tobin (Susquehanna University) and Jay Hosler (Juniata College). This team will guide development of appropriate instruments to monitor student outcomes for network participants, as well as to support individual faculty in assessing the impact of students working with raw sequence data in individual classes. Core learning outcomes proposed for the GCAT-SEEK network were the ability for instructors and students to do the following:

1. Explain each step in the generation and analysis of NextGen sequence data.
2. Discuss the basic biology assumptions that underlie sequence analysis (e.g., evolution, structure, and function).
3. Evaluate the strengths and weaknesses of the methods used in NextGen sequencing, including the impact that data quality has on bioinformatics analysis.
4. Construct a testable hypothesis and experimental design that uses NextGen sequencing and bioinformatics tools.
5. Choose and justify the appropriate methods for a specific NextGen sequencing application.

These proposed learning goals are also posted on the web ([www.gcat-seek.org](http://www.gcat-seek.org)). The goal is to have best teaching practices established and validated through appropriate assessment and distributed to all of the network participants and their colleagues.

## GCAT-SEEK: VISION AND APPROACH

Following whole-group discussion, it was determined the agreed purpose of GCAT-SEEK is to 1) bring functional genomic methods into the undergraduate curriculum, primarily through independent and classroom-based student research using centralized core facilities to make NextGen sequence data accessible to undergraduates; 2) create a clearinghouse of information for educators to use when teaching NextGen sequencing and related topics; 3) create a large database of raw data and analyzed results for pedagogical use by GCAT-SEEK members; and 4) develop a global network of educators who are using functional genomics and NextGen sequencing in the undergraduate curriculum. GCAT-SEEK specifically aims to obtain group discounts at regional research-intensive core facilities, to negotiate software discounts, and to garner support for mini-grants to help cover the cost of the

initial sequencing runs for network participants. The network aims to support its members through online listservs, periodic workshops, and meetings, following an approach similar to that successfully used by other GCAT groups (Campbell *et al.*, 2006). The network approach will add efficiency by coordinating projects and partnering investigators with appropriate sequencing platforms. Given that even the smallest purchasable unit of NextGen sequence will often contain a great excess of information for any given project and that many projects can be combined using bar codes (as discussed in the introductory paragraph), an organized staging for related samples from different investigators was envisioned. This may reduce cost of data acquisition to a few thousand dollars from departmental budgets or mini-grant programs. Furthermore, additional cost efficiency for network participants can be achieved through coordination of the synthesis and maintenance of a database of highly purified bar-coded primers for metagenomic analysis.

### WHAT'S NEXT?

At a time of severe budget scrutiny, the efficiency and cost-effectiveness of the proposed approach is apparent. Sequencing cores are not running at capacity, and technological advances are lowering sequencing costs. Bioinformatics programs, databases, and computing requirements for many types of projects are all either already in the public domain or well within the budgets of even the smallest undergraduate colleges. Given sufficient interest, regional replication of some elements of GCAT-SEEK in the future should be considered as a means of lowering travel costs for meetings and workshops for participants and allowing students to more easily visit genome core facilities. Thus, with a modest investment, this program can start to meet the challenges of training the next generation of life scientists by engaging undergraduates in the process of science presented in the context of modern technology.

### ACKNOWLEDGMENTS

We thank Anton Nekrutenko, Istvan Albert, and Bill Morgan for their excellent and insightful presentations. This meeting was supported by NSF Award DBI-1061893 to M.D.B.

### REFERENCES

- American Association for the Advancement of Science (2010). Vision and Change in Undergraduate Biology Education: A Call to Action. Washington DC: American Association for the Advancement of Science. <http://visionandchange.org/files/2011/02/Vision-and-Change-low-res.pdf> (accessed 21 September 2011).
- Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protocols Mol Biol*, chapter 19, unit 19.10, 1–21.
- Boyle MD (2010). Shovel-ready sequences as a stimulus for the next generation of life scientists. *J Microbiol Biol Educ* 11, 38–41.
- Campbell AM, Eckdahl TT, Fowlks E, Heyer LJ, Mays Hoopes LL, Ledbetter ML, Rosenwald AG (2006). Genome Consortium for Active Teaching (GCAT). *Science* 311, 1103–1104.
- Campbell M, Heyer LJ (2006). *Discovering Genomics, Proteomics, and Bioinformatics*, 2nd ed., Cold Spring Harbor Laboratory Press/Benjamin Cummings: San Francisco, CA.
- 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.
- Goecks J, Nekrutenko A, Taylor J (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11, R86.
- Teagle Foundation (2007). Student learning and faculty research: connecting teaching and scholarship, a Teagle Foundation white paper. [www.teaglefoundation.org/learning/pdf/2006\\_acls\\_whitepaper.pdf](http://www.teaglefoundation.org/learning/pdf/2006_acls_whitepaper.pdf) (accessed 21 September 2011).
- Wolyniak MJ *et al.* (2010). Building better scientists through cross-disciplinary collaboration in synthetic biology: a report from the Genome Consortium for Active Teaching Workshop 2010. *CBE Life Sci Educ* 9, 399–404.