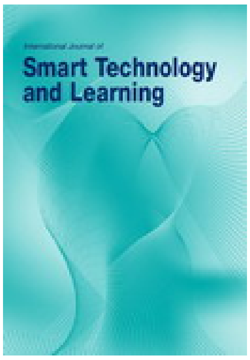

International Journal of Smart Technology and Learning

ISSN online: 2056-4058 - ISSN print: 2056-404X
https://www.inderscience.com/ijsmarttl

Bioinformatics education for undergraduates: the need for project-based and experiential approaches

Michael J. Wolyniak

DOI: 10.1504/IJSMARTTL.2023.10054693

Article History:

Received:	18 February 2021
Last revised:	27 June 2021
Accepted:	03 August 2021
Published online:	17 March 2023

Int. J. Smart Technology and Learning, Vol. 3, No. 2, 2023 107

Bioinformatics education for undergraduates: the need for project-based and experiential approaches

Michael J. Wolyniak

Hampden-Sydney College,
P.O. Box 183,
Hampden-Sydney, VA 23943, USA
Email: mwolyniak@hsc.edu

Abstract: The -omics revolution and advances in DNA sequencing technology have made bioinformatics an essential tool for full participation in the rapidly-evolving life science research community. However, considerable barriers experienced by undergraduate instructors and students have largely prevented the mainstream integration of bioinformatics into life science curricula. To overcome these barriers, several groups have developed collaborative initiatives with the goal of providing instructors with the skills, confidence, and resources that they need to successfully implement bioinformatics course modules to their students in an engaging manner. This review paper considers some of these successful initiatives and offers ideas on how their work can be further expanded to make bioinformatics education a standard practice at the undergraduate level. Overall, the work summarised in this paper highlights the importance of collaboration between undergraduate instructors at different institutions as well as between undergraduate instructors and research-intensive faculty to develop authentic undergraduate research experiences that can benefit all groups involved.

Keywords: bioinformatics; high-throughput sequencing; undergraduate education; barriers; CUREs; course-based undergraduate research experiences.

Reference to this paper should be made as follows: Wolyniak, M.J. (2023) ‘Bioinformatics education for undergraduates: the need for project-based and experiential approaches’, *Int. J. Smart Technology and Learning*, Vol. 3, No. 2, pp.107–117.

Biographical notes: Michael J. Wolyniak is the McGavacks Associate Professor of Biology at Hampden-Sydney College, where he teaches a variety of courses in molecular and cellular biology as well as bioinformatics.

1 Introduction

A constant challenge for all undergraduate instructors in the scientific disciplines is to keep their curricula current in the face of rapid advances in technology, methodology, and discovery. While many instructors will regularly attend workshops and conferences to keep themselves abreast of advances in their fields, it is extremely difficult to gain the professional training necessary to confidently teach about novel technologies and experimental approaches that did not exist during these instructors’ time in graduate school or postdoctoral training. In this review paper, we consider the ways in which undergraduate faculty in the life sciences have adjusted to the rapid rise of bioinformatics and how they have formed partnerships between themselves and with professional researchers to develop exciting and cutting-edge educational experiences for their students.

2 The "Big Data" revolution in the life sciences

Within the past generation, the life sciences have undergone a "big data" revolution in which a knowledge of informatics and computer science is becoming more and more essential for researchers and students alike. The driving force behind this revolution has been the development and refinement of high-throughput DNA sequencing technology (Tripathi et al., 2016; Delseny et al., 2010). Commonly known as "next-generation" sequencing, scientists can now sequence entire genomes of organisms in a fraction of the time and the cost that it would have been a mere ten years ago (Kulski 2016). "Next-generation" sequencing also allows for looking at snapshots of total RNA expression in a given organism or tissue at a specific time, which in turn allows scientists to do sophisticated studies on gene expression and regulation of entire genetic systems. The meteoric rise of sequencing technologies has continued in more recent years with what is being informally called "third generation sequencing", in which genomic sequence results can be generated using a device the size of a flash drive connected via USB connection to a laptop computer (van Dijk et al., 2018). This sequencing revolution has been transformative for the life sciences, overturning previously-established morphological-based findings on natural selection and evolution while uncovering insights about the nature of cell and molecular biology that would have been impossible to discern in the recent past.

As the life science community has embraced new sequencing technologies, it has also found itself with the need to come to terms with analysis of "big data" sets. While the mining of large sets of data has been a part of several other academic disciplines for a considerable length of time, the -omics revolution spawned by high-throughput sequencing has led to the emergence of computational biology and bioinformatics as legitimate life science subdisciplines (Yang et al., 2008). Scientists with a passion for both biology and computer science have found themselves well-positioned in this emerging field to create the algorithms, platforms, and statistical tools necessary to mine large data sets for information that can unlock the mysteries of life (Magana et al., 2014; Yang et al., 2009). As "big data" becomes more and more essential for effective research in the life sciences, the demand for computational biology and bioinformatics knowledge and experience can be expected to remain high for the foreseeable future.

The development of new and exciting research trends in -omics, computational biology, and bioinformatics has created the need for training new generations of scientists that are prepared to utilise and expand on the progress that has been made with developing and implementing tools for biologically-based "big data" sets (Porter and Smith 2019). This training, in order to be lasting and effective, needs to begin well before arrival in a graduate program and must take place in the undergraduate or even secondary school levels of education. However, a 2019 survey of undergraduate faculty in the United States found that while 95% agree with the idea that bioinformatics should be integrated into undergraduate life sciences education, only 32% were actually teaching courses that involved bioinformatics content (Williams et al., 2019). In this review, we will consider what has been shown through research and experience to be the emerging best practices for effective bioinformatics training, the barriers that can hinder the development of effective training programs, and how the Course-based Undergraduate Research Experience (CURE) movement in life sciences education has been central to the development of strong bioinformatics training programs.

3 What about bioinformatics do undergraduates need to know and learn?

While technological and research advances demand an increased effort to provide bioinformatics training to undergraduates who major in the life sciences, there is not a clear consensus on what core concepts or experiences should be encompassed by such training. A recent survey of 1260 life sciences faculty representing undergraduate teaching institutions of all varieties sought to establish the needs and interests of bioinformatics educators and establish a concept inventory for a bioinformatics-based curriculum (Wilson Sayers et al., 2018). While there was near universal agreement on the need to increase bioinformatics-based opportunities for undergraduates, there was divergence based largely on institution type on what bioinformatics concepts are critical to this goal. Specifically, faculty at smaller institutions emphasised the importance of students mastering online databases and software tools while faculty at more research-intensive institutions also included the ability to write simple computational tools that could solve bioinformatics problems (Wilson Sayers et al., 2018). This divergence likely represents the difference in resource availability in terms of both infrastructure and personnel at these distinct types of institutions. The survey results also prioritised the need to understand and master the concepts of bioinformatics over the experience of engaging research problems in the field, a result which may represent the recent development of bioinformatics and the subsequent lack of direct experience that many current instructors have had with the field.

The Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE) (Dinsdale et al., 2015), used these survey results to devise a set of bioinformatics core competencies that emphasises the early introduction of computational practices to life science education, the ways that computational techniques can be used to answer complex questions in evolution, molecular biology, and medical genetics, and the need for student engagement with authentic "big data" sets to learn how to effectively use bioinformatics to solve research problems (Table 1). These core competencies stand as a broad set of learning objectives that could be used to model a bioinformatics curriculum at any type of undergraduate teaching institution regardless of resource level or size.

Inherent in the NIBLSE recommendations is the need to get students to engage with authentic research questions (defined for the purposes of this paper as projects in which students must consider literature, propose hypotheses, and develop experiments to solve problems for which they do not know the answer) and use experiential learning with the goal of getting students to think and act like scientists. Such experiential-based techniques have long been recognised as part of the high-impact pedagogical practices that foster undergraduate student retention and graduation in the STEM disciplines (Kuh et al., 2008) and were fully integrated in the landmark *Vision and Change* report of the American Association for the Advancement of Science which has had a transformative effect on the way life science education has evolved in undergraduate institutions across the United States in the past decade (AAAS, 2011). Specifically, *Vision and Change* recommends the integration of authentic research experiences into the life sciences curriculum as well as the use of active learning techniques such as group activities, projects, and "flipped classrooms" that research has shown improve the ability of undergraduates to develop the skills and passions for science needed for long term success (AAAS, 2011). While it is true that the integration of these learning techniques often results in the need to remove some of the content-based lectures of a traditional bioinformatics course, it is evident that this trade-off is beneficial for students and helps them to develop a skill set that prepares them to handle the diverse challenges that await them in more advanced coursework (Eagan et al., 2013). Thus, it is important not only to develop and implement coursework in bioinformatics for undergraduates but to also do so in an engaging manner that allows students to "learn by doing" and to work with existing bioinformatics tools in the pursuit of answering research questions.

Table 1 NIBLSE-recommended bioinformatics core competencies (adapted from Wilson Sayers et al., 2018)

1	Explain the role of computation and data mining in addressing hypothesis-driven and hypothesis-generating questions within the life sciences.
2	Summarise key computational concepts, such as algorithms and relational databases, and their applications in the life sciences.
3	Apply statistical concepts used in bioinformatics.
4	Use bioinformatics tools to examine complex biological problems in evolution, information flow, and other important areas of biology.
5	Find, retrieve, and organise various types of biological data.
6	Explore and/or model biological interactions, networks, and data integration using bioinformatics.
7	Use command-line bioinformatics tools and write simple computer scripts.
8	Describe and manage biological data types, structure, and reproducibility.
9	Interpret the ethical, legal, medical, and legal implications of biological data.

4 Barriers to effective bioinformatics education

The NIBLSE group also surveyed the same faculty group mentioned above on the barriers that prevented the implementation of bioinformatics coursework (Williams et al., 2019). The results suggested, not surprisingly, that a lack of faculty training and preparation was the most common barrier observed among this group. Given the relatively recent emergence of bioinformatics as a distinct field in the life sciences, it is expected that a significant percentage of current instructors would never have received any sort of formal training in this field and would therefore not feel as if they could adequately implement a successful bioinformatics course (Zhan et al., 2019). However, the study also surprisingly found that the youngest cohort of faculty members were the group least likely to be teaching bioinformatics at their institutions despite having had the most opportunity to receive formal bioinformatics training. It is unclear if this result is due to the relative inability of this newer cohort to influence curricular planning decisions at their institutions or if there are larger problems with the relative attractiveness of bioinformatics coursework to instructors and students alike relative to other life science subdisciplines. To this end, instructors also reported significant issues with low student interest in bioinformatics and student intimidation by the subject material. Other barriers reported by this study included issues in equity and inclusion for students and faculty from underrepresented groups and limitations in time and resources at institutions that place less emphasis on research. These barriers, while not an exhaustive list, indicate that there is much work to be done in the life sciences education community to make bioinformatics a more accessible topic to instructors and students alike that can be successfully integrated into the curriculum at institutions of all sizes and missions.

5 CURES and collaborations for bioinformatics

As mentioned above, the *Vision and Change* report calls for the integration of research experiences into undergraduate curricula at all levels as a high-impact pedagogical practice that can improve student interest and retention in STEM fields as well as provide students with a set of skills that will prepare them to solve a diverse field of complex scientific problems (AAAS, 2011). Over the past several years, a body of literature has accumulated that has verified these ideas and implicates CUREs as key to preparing undergraduate STEM students for a scientific workforce that is ever-evolving and requires strong problem-solving skills in order to adapt to these changes (Eagan et al., 2013; Schultz et al., 2011; Hurtado et al., 2008). CUREs also have been shown to reduce the barriers that have persisted in life science education by providing students from underrepresented groups in STEM equitable access to research opportunities (Carpi et al., 2017; Bangera and Brownell, 2014). Therefore, it seems that an excellent way to overcome some of the barriers mentioned above to the implementation of bioinformatics into the undergraduate curriculum is the development and implementation of bioinformatics-based CUREs into both existing and novel coursework. These CUREs would, of course, need to be tailored to the existing needs of individual courses and curricula at each individual institution. Nevertheless, CUREs to integrate bioinformatics-based authentic research experiences into all levels of the undergraduate life science curriculum are an exciting potential solution to making bioinformatics a more accessible and attractive subdiscipline to both students and instructors.

6 Examples of networked collaborations and CURE initiatives

Like bioinformatics education, CUREs and other collaborative ventures come with their own barriers to development, including logistical concerns and lack of appropriate pedagogical training (Shaffer et al., 2014). Luckily, several collaborative initiatives have emerged to help instructors new to the practice of both CUREs and bioinformatics to work with more experienced peers to develop the skills and the confidence necessary to implement CURE modules in their own classrooms. Another exciting aspect of these CURE initiatives in many cases is the crowdsourced nature of their projects. In these initiatives, undergraduates can potentially find themselves working on research in parallel with students in classrooms around the United States and beyond to solve large-scale problems that can contribute to the research literature while providing students with a potentially transformative educational experience that may inspire them to develop careers in the life sciences. Below are several examples of some of these transformative collaborative initiatives that are changing the face of undergraduate bioinformatics education:

Genome Consortium for Active Teaching using Next Generation Sequencing (GCAT-SEEK): GCAT-SEEK (Buonaccorsi et al., 2017; Buonaccorsi et al., 2011) (http://gcat-seek.weebly.com) is a network formed in response to the significant infrastructure barriers facing the majority of undergraduate instructors with respect to the design and implementation of projects involving high-throughput sequencing. The network is itself a "next generation" project, having evolved from the original GCAT network which focused on making microarray technology accessible to undergraduate classrooms (Campbell et al., 2006). Through the use of annual hands-on workshops for both instructors and students as well as affordable access to sequencing runs of samples generated by undergraduates, GCAT-SEEK provides both the opportunity for instructors to introduce bioinformatics research into their classrooms and the support necessary to give instructors the resources and training they need to implement and sustain these efforts. Workshop participants bring their own project ideas and goals to the workshops so that the result of the immersive training experience is a CURE module that is ready to be used in their unique class environment. Collaborations with research universities and use of bulk run discounts reduce the logistical barriers the majority of undergraduate faculty face when planning "big data" experimentation. These logistical barriers include the expense of running sequencing projects for several different student projects as well as limited access to state-of-the-art sequencing instrumentation at non-research intense institutions. Importantly, the community aspect of GCAT-SEEK allows for the development of collaboration and joint project ideas that let students at different institutions work with each other in the generation and analysis of "big data" sets that would not be possible by themselves. Such collaborations often result in increased levels of processed data for faculty at research-intense universities and authorship opportunities for undergraduates at institutions of all types.

DNA Subway: A common barrier to bioinformatics curricular implementation is the sheer number of specific online tools for the processing and analysis of "big data" sets combined with a lack of knowledge of how to use these tools in concert to properly conduct fruitful research that would benefit students (Williams et al., 2019; Cummings and Temple, 2010). DNA Subway (Hilgert et al., 2014; Williams et al., 2014) (http://dnasubway.cyverse.org) is a tool developed by the iPlant Collaborative (Merchant et al., 2016) to provide a classroom-friendly interface for students to explore high-throughput sequencing analysis tools. Using the various "subway lines" developed to gather analytical tools together in a logical order, students can predict and annotate genes in a given sequence, discover gene homologs across other sequenced genomes, identify species, examine RNA-seq data to explore gene frequencies in a given data set, and perform metabarcoding analysis. DNA Subway seeks to remove the issues of complexity surrounding the myriad of online analytical tools used by the bioinformatics research community and organise and present such tools in a manner that eliminates accessibility barriers to students.

Genome Solver: Inspired by advances in sequencing technology and the prospects of discovery fostered by the Human Microbiome Project (HMP), Genome Solver (https://genomesolver.qubeshub.org) is a clearinghouse for instructors and students interested in working with prokaryotic genomic datasets and developing collaborations with other instructors to enhance the scientific experience for students (Rosenwald et al., 2012). Genome Solver identifies itself as a community of practice for faculty and students, and it is this collaborative spirit that drives its mission of providing accessibility to prokaryotic genetic exploration. It has accomplished this through serving as a repository for pedagogical materials as well as a forum for communication between instructors and research experts on how to best approach genomics education in different pedagogical contexts. The resources on Genome Solver go beyond the set-up and implementation of genomics research projects for the classroom and also include the classroom activities and communication tools that will allow students to develop the skills to effectively disseminate their work to a larger audience.

Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES): The SEA-PHAGES initiative (http://www.seaphages.org) is a project of the Howard Hughes Medical Institute (HHMI) that seeks to introduce CUREs into the first year of the undergraduate life science curriculum through the use of bacteriophage discovery and analysis (Jordan et al., 2014). Roughly 100 undergraduate institutions across the United States are engaged in the full SEA-PHAGES curriculum, which involves isolating a novel bacteriophage from the environment, getting a genomic DNA sequence of the bacteriophage, and utilising bioinformatics tools to annotate the resulting sequence and officially document the results in GenBank. Scientifically, the goal of SEA-PHAGES is to document the diversity of bacteriophages in the environment to get a better sense of the pressures that govern their evolution. An important feature of SEA-PHAGES is the collaborative nature of the network both between different participating institutions and with Dr. Graham Hatfull and his laboratory at the University of Pittsburgh that manages the project. With respect to bioinformatics, the genomic annotation process relies on simple online analytical tools and strong training support for instructors from the SEA-PHAGES team. In this way, SEA-PHAGES is an excellent teaching tool for instructors with little to no prior bioinformatics experience that seek to learn how to introduce bioinformatics fundamentals to their students in an engaging manner. Over the years, various SEA-PHAGES faculty have developed their own extensions to the core network curriculum to enhance the bioinformatics tools that their own students can explore. One such tool, the Phage Evidence Collection and Annotation Network (PECAAN – https://discover.kbrinsgd.org), has greatly simplified the annotation process by consolidating the many online tools for the development of genomic annotations to a single location in a manner similar to DNA Subway.

Genomics Education Partnership (GEP): Like SEA-PHAGES, the GEP (https://thegep.org) was designed to take the work of a large research laboratory, in this case the work of Dr. Sarah Elgin of Washington University in St. Louis, and crowdsource the data analysis to hundreds of undergraduate classrooms in an effort to both accelerate the data analysis process and to provide authentic and engaging research experiences to a large group of undergraduate students (Elgin et al., 2017; Shaffer et al., 2014; Lopatto et al., 2008). Work in the GEP focuses on the genomic annotation on several *Drosophila* species, and yearly immersive training sessions for instructors provide the skills and support necessary for them to successfully implement GEP materials in their own classrooms. As with SEA-PHAGES, the GEP set of resources has grown immensely over the past several years on the backs of veteran network instructors who have developed several enhancements to the core GEP workflow as well as curriculum appropriate for introductory-level students.

NIBLSE incubators: The NIBLSE initiative, as shown in this paper, has played a vital role in recent years in bringing together undergraduate bioinformatics instructors from all different types of institutions and working to identify both the barriers facing the mainstream introduction of bioinformatics education to the undergraduate curriculum and the prospective solutions that can be developed to eliminate such barriers. One such project that combines NIBLES and the Quantitative Undergraduate Biology Education and Synthesis (QUBES) group is an Incubators initiative (https://qubeshub.org/community/groups/niblse/resourcecollection/incubator) that seeks to increase the amount of high-quality bioinformatics-related teaching material available for public use for the undergraduate classroom (Ryder et al., 2020). Incubators are short-term communities in which unpublished bioinformatics teaching materials are refined by members of NIBLSE so that they can be utilised by as wide of a group of instructors as possible and published as an open education resource in which the creator of the materials can receive scholarly credit for their work. The work involved in the refinement of these teaching materials forges collaborations between faculty members and seeds the development of new projects that can be pursued by members of the network. The result of the NIBLSE Incubators is the continual generation of novel and high-quality pedagogical modules that can be employed in a wide variety of classrooms to present bioinformatics in an engaging and active manner.

7 Future directions

The meteoric rise of the -omics subdisciplines and the need to examine life science research questions from a systemic perspective makes the mainstreaming of bioinformatics education at the undergraduate level more essential than ever. The initiatives described here detail the significant advances that have been made towards this goal through collaboration and the development of innovative means to present bioinformatics to students not as a set of facts or protocols to be memorised but rather as a toolbox that can be implemented in ways tailored to the solving of a diverse array of scientific problems. By presenting bioinformatics in a project-based manner, students can engage with its concepts and tools and develop the confidence to work with the “big data” sets that are becoming more and more central to life science research (Pucker et al., 2019).

How can we continue the momentum provided by the initiatives described here? One suggestion is to bolster ongoing efforts to introduce basic ideas in computer science and computational scientific approaches at the secondary school level as a means to help dispel the fears that many undergraduates express about tackling bioinformatics-related material (Kovarik et al., 2013; Barr and Stephenson, 2011). While the concepts and problems that students would be expected to tackle in secondary school would not be as sophisticated as what would be introduced in the undergraduate years, the primary intent of such efforts would be to introduce at an earlier age the centrality of computational practices and “big data” analytics to the life sciences. By providing accessible and engaging means to learn about bioinformatics and computational biology, students would enter their undergraduate years not only prepared to handle bioinformatics in a project-based manner but would also show a higher level of enthusiasm for the subject than what is currently observed in the typical undergraduate classroom.

Another critical development that is necessary is to provide the necessary professional development and flexibility to all instructors but especially to junior members of faculty to innovate in their curricula. Since the NIBLSE survey results reveal that junior faculty members are the least likely cohort to actually implement bioinformatics in the classroom despite being more likely to have had first-hand experience and training with bioinformatics tools than their senior counterparts, more must be done to explore the reasons for this discrepancy. If the problem is a lack of freedom to innovate in the curriculum and to introduce novel ideas, then the networks described here must turn their attentions to developing campaigns designed to reach department chairs to impart on them the importance of providing bioinformatics training to their students and allowing new faculty members to bring their experiences with this subdiscipline into their curricula. If the problem is one of concerns for lack of time or experience, then these networks must work to show these faculty members the benefits of collaboration and working with colleagues across the United States and beyond to bring exciting and, in many cases, crowdsourced research opportunities to their students in a way that gives the instructor confidence in their techniques and their chosen classroom model. It is the aforementioned efforts of these bioinformatics education networks that have pushed forward the exciting pedagogical models described in this paper. Without a doubt, these and other new networks in the years to come will be the driving force for the continued integration of bioinformatics as a staple of the undergraduate life science curriculum.

8 Conclusion

The examples shown here are indicative of an exciting collaborative effort in the life sciences community between bioinformatics researchers and the instructors responsible for training the next generation of data scientists. By partnering in the name of increasing processing of biological "big data" as a training enterprise, undergraduate students are receiving unparalleled opportunities to learn about the data-based foundations of modern biology and to prepare themselves for the careers that are emerging to meet the challenges posed by these foundations. This review paper suggests strong utility to this pedagogical model and is hopeful that the CURE approach to undergraduate education becomes standard practice across all disciplines.

References

American Association for the Advancement of Science (AAAS) (2011) *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC.

Bangera, G. and Brownell, S.E. (2014) 'Course-based undergraduate research experiences can make scientific research more inclusive', *CBE-LSE*, Vol. 13, No. 4, pp.602–606.

Barr, V. and Stephenson, C. (2011) 'Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community?' *ACM Inroads*, Vol. 2, No. 1, pp.48–54.

Buonaccorsi, V., Peterson, M., Lamendella, G., Newman, J., Trun, N., Tobin, T. et al. (2017) 'Vision and change through the Genome Consortium for Active Teaching using Next-Generation Sequencing (GCAT-SEEK)', *CBE Life Sci Educ*, Vol. 13, Nos. 1/2, Pmid:24591495.

Buonaccorsi, V.P., Boyle, M.D., Grove, D., Praul, C., Sakk, E., Stuart, A. et al. (2011) 'GCAT-SEEKquence: genome consortium for active teaching of undergraduates through increased faculty access to next-generation sequencing data', *CBE-LSE*, Vol. 10, No. 4, pp.342–345.

Campbell, A.M., Eckdahl, T.T., Fowlks, E., Heyer, L.J., Hoopes, L.L.M., Ledbetter, M.L. and Rosenwald, A.G. (2006) 'Genome consortium for active teaching (GCAT)', *Science*, Vol. 311, No. 5764, pp.1103–1104.

Carpi, A., Ronan, D.M., Falconer, H.M. and Lents, N.H. (2017) 'Cultivating minority scientists: undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM', *J Res Sci Teach*, Vol. 54, No. 2, pp.169–194.

Cummings, M.P. and Temple, G.G. (2010) 'Broader incorporation of bioinformatics in education: opportunities and challenges', *Brief. Bioinfor*, Vol. 11, No. 6, pp.537–543.

Delseny, M., Han, B. and Hsing, Y.I. (2010) 'High throughput DNA sequencing: the new sequencing revolution', *Plant Science*, Vol. 179, No. 5, pp.407–422.

Dinsdale, E., Elgin, S.C., Grandgenett, N., Morgan, W., Rosenwald, A., Tapprich, W., Triplett, E.W. and Pauley, M.A. (2015) 'NIBLSE: a network for integrating bioinformatics into life sciences education', *CBE-LSE*, Vol. 14, No. 4, le3.

Eagan, M.K., Hurtado, S., Chang, M.J., Garcia, G.A., Herrera, F.A. and Garibay, J.C. (2013) 'Making a difference in science education: the impact of undergraduate research programs', *Am Educ Res J*, Vol. 50, No. 4, pp.683–713.

Elgin, S.C., Hauser, C., Holzen, T.M., Jones, C., Kleinschmit, A., Leatherman, J. and The Genomics Education Partnership (2017) 'The GEP: crowd-sourcing big data analysis with undergraduates', *Trends Genet*, Vol. 33, No. 2, pp.81–85.

Hilgert, U., McKay, S., Khalfan, M., Williams, J., Ghiban, C. and Micklos, D. (2014) 'DNA Subway: making genome analysis egalitarian', *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, pp.1–3.

Hurtado, S., Cabrera, N.L., Lin, M.H., Arellano, L. and Espinosa, L.L. (2008) 'Diversifying science: Underrepresented student experiences in structured research programs', *Res High Educ*, Vol. 50, No. 2, pp.189–214.

Jordan, T.C., Burnett, S.H., Carson, S., Caruso, S.M., Clase, K., DeJong, R.J. et al. (2014) 'A broadly implementable research course in phage discovery and genomics for first-year undergraduate students', *Mbio*, Vol. 5, No. 1, e01051–13. Pmid:24496795.

Kovarik, D.N., Patterson, D.G., Cohen, C., Sanders, E.A., Peterson, K.A., Porter, S.G. and Chowning, J.T. (2013) 'Bioinformatics education in high school: implications for promoting science, technology, engineering, and mathematics careers', *CBE-LSE*, Vol. 12, No. 3, pp.441–459.

Kuh, G.D. (2008) *High-impact educational practices: What they are, who has access to them, and why they matter*, Association of American Colleges and Universities, Washington, DC.

Kulski, J. (Ed.) (2016) *Next Generation Sequencing: Advances, Applications and Challenges*, BoD–Books on Demand.

Lopatto, D., Alvarez, C., Barnard, D., Chandrasekaran, C., Chung, H-M., Du, C. et al. (2008) 'Undergraduate research: Genomics Education Partnership', *Science*, Vol. 322, pp.684–685.

Magana, A.J., Taleyarkhan, M., Rivera Alvarado, D., Kane, M., Springer, J. and Clase, K. (2014) 'A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research', *CBE Life Sci Educ*, Vol. 13, pp.607–623.

Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D. and Antin, P. (2016) 'The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences', *PLoS Biology*, Vol. 14, No. 1, e1002342.

Porter, S.G. and Smith, T.M. (2019) 'Bioinformatics for the Masses: The Need for Practical Data Science in Undergraduate Biology', *Omics*, Vol. 23, No. 6, pp.297–299.

Pucker, B., Schilbert, H.M. and Schumacher, S.F. (2019) 'Integrating molecular biology and bioinformatics education', *J Integr Bioinform*, Vol. 16, No. 3, 20190005.

Rosenwald, A.G., Russell, J.S. and Arora, G. (2013) 'The genome solver website: a virtual space fostering high impact practices for undergraduate biology', *J Microbiol Biol Educ*, Vol. 13, pp.188–190. Pmid:23653812.

Ryder, E.F., Morgan, W.R., Sierk, M., Donovan, S.S., Robertson, S.D., Orndorf, H.C. et al. (2020) 'Incubators: building community networks and developing open educational resources to integrate bioinformatics into life science education', *Biochem Mol Biol Educ*, Vol. 48, No. 4, pp.381–390.

Shaffer, C.D., Alvarez, C.J., Bednarski, A.E., Dunbar, D., Goodman, A.L., Reinke, C. et al. (2014) 'A course-based research experience: how benefits change with increased investment in instructional time', *CBE-LSE*, Vol. 13, pp.111–30.

Schultz, P.W., Hernandez, P.R., Woodcock, A., Estrada, M., Chance, R.C., Aguilar, M. and Serpe, R.T. (2011) 'Patching the pipeline reducing educational disparities in the sciences through minority training programs', *Educ Eval Policy Anal*, Vol. 33, pp.95–114.

Tripathi, R., Sharma, P., Chakraborty, P. and Varadwaj, P.K. (2016) 'Next-generation sequencing revolution through big data analytics', *Frontiers in Life Science*, Vol. 9, No. 2, pp.119–149.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D. and Thermes, C. (2018) 'The third revolution in sequencing technology', *Trends in Genetics*, Vol. 34, No. 9, pp.666–681.

Williams, J.J., Drew, J.C., Galindo-Gonzalez, S., Robic, S., Dinsdale, E., Morgan, W.R. et al. (2019) 'Barriers to integration of bioinformatics into undergraduate life sciences education: a national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction', *PLoS One*, Vol. 14, No. 11, e0224288.

Williams, J., McKay, S., Khalfan, M., Ghiban, C., Hilgert, U., Lauter, S., Jeong, E-S. and Micklos, D. (2014) 'DNA subway – an educational bioinformatics platform for gene and genome analysis: DNA barcoding, and RNA-Seq', *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*, pp.17–22.

Wilson Sayres, M.A., Hauser, C., Sierk, M., Robic, S., Rosenwald, A.G., Smith, T.M. et al. (2018) 'Bioinformatics core competencies for undergraduate life sciences education', *PLoS One*, Vol. 13, No. 6, e0196878.

Yang, J.Y., Yang, M.Q., Zhu, M.M., Arabnia, H. and Deng, Y. (2008) 'Promoting synergistic research and education in genomics and bioinformatics', *BMC Genomics*, Vol. 9(suppl 1), pp.25–28.

Yang, M.Q., Niemierko, A., Yang, J.Y. and Jin, Y. (2009) 'Promoting inter/multidisciplinary education and research in bioinformatics, systems biology and intelligent computing', *Int J Comp Biol Drug Des*, Vol. 2, No. 3, pp.207–220.

Zhan, Y.A., Wray, C.G., Namburi, S., Glantz, S.T., Laubenbacher, R. and Chuang, J.H. (2019) 'Fostering bioinformatics education through skill development of professors: Big Genomic Data Skills Training for Professors', *PLoS Comp Biol*, Vol. 15, No. 6, e1007026.