



OPEN ACCESS

EDITED AND REVIEWED BY
Lianghuo Fan,
East China Normal University, China

*CORRESPONDENCE
Renato Augusto Corrêa dos Santos
renatoacsantos@gmail.com

SPECIALTY SECTION
This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

RECEIVED 25 July 2022
ACCEPTED 15 August 2022
PUBLISHED 29 August 2022

CITATION
dos Santos RAC, Verli H and de
Melo-Minardi RC (2022) Editorial:
Original strategies for training and
educational initiatives in
bioinformatics.
Front. Educ. 7:1003098.
doi: 10.3389/feduc.2022.1003098

COPYRIGHT
© 2022 dos Santos, Verli and de
Melo-Minardi. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Editorial: Original strategies for training and educational initiatives in bioinformatics

Renato Augusto Corrêa dos Santos^{1*}, Hugo Verli² and Raquel Cardoso de Melo-Minardi³

¹Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, Brazil, ²Center of Biotechnology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, ³Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

KEYWORDS

bioinformatics, education, training, online course, programming

Editorial on the Research Topic

[Original strategies for training and educational initiatives in bioinformatics](#)

In the Research Topic “*Original strategies for training and educational initiatives in bioinformatics*,” original research (5), perspective (4), opinion (2), a brief research report (1) and an “curriculum, instruction, and pedagogy” (1) article were published. Most of the authors are affiliated in the U.S. (Niepielko and Shumskaya; Tully et al.; Daetwyler et al.; Dow et al.; Robertson et al.; Goller et al.), Africa (Sangeda et al.; Aron et al.), and South America (Ras et al.; Melo-Minardi et al.; Dorn et al.; Braga et al.; da Silva et al.). Several articles involve intercontinental collaborations (Tully et al.; Ras et al.; Sangeda et al.; Braga et al.). This special edition includes manuscripts published in Frontiers in Bioinformatics (Section Computational BioImaging and Protein Bioinformatics) and Frontiers in Education (Section STEM Education).

In their opinion article, Niepielko and Shumskaya reflect on the necessity of data analysis and interpretation skills for life sciences professionals, and that there is a large gap in their training today (e.g., basic statistics knowledge). Authors argue that bioinformatics should be mandatory and taught in a biology course as soon as students have completed classes on general biology and statistics, and that it should be focused on application of statistics in biology, and not in the mathematical theory behind it, making it more accessible. Next, they report their experiences in introducing bioinformatics to undergraduate students at Kean University, including a mandatory course for sophomores of B.S. in Biology, one for students that want to develop more skills in computational biology and that also works well for students majoring in computer sciences, and another for those pursuing a B.S. in Science and Technology in Molecular Biology (Bioinformatics and Genomic Science). They developed their own materials suitable for beginners with a biology background, including activities that are available as Open Education Resources. Given the common lack of motivation for

students, they emphasize the importance of hands-on courses that employ “real, timely, and relevant data” for students’ excitement. Interestingly, at Kean University they offer a 4-day bioinformatics workshop, which includes lectures and hands-on activities.

Goller et al. describe the inquiry-based Molecular Biology Laboratory Education Modules (MBLEMs) developed by the Biotechnology Program at North Carolina State University and how they use bioinformatics tools combined with wet-lab experiments to answer research questions. In this Perspective article, authors explain how MBLEMs are created, a 5D process that includes Designation, Design, Development, Deployment, and Dissemination, and usually take place as part of the program of teaching postdoctoral scholars. Finally, they provide examples of MBLEMs addressing (i) first-year courses for STEM and non-STEM majors, (ii) elective courses for upper-level and graduate students, and (iii) short courses for completing a biotechnology minor, in which bioinformatics tools are included at different levels. A major focus on MBLEMs design is that bioinformatics tools are chosen to offer the best possible pedagogical experience (e.g., user-friendly software), and courses have peer support and employ inquiry-based projects.

Motivated by cancellation during the COVID-19 pandemic of a successful in-person annual hackathon where students were able to solve “real-world” problems and interact with researchers at UT Southwestern, Daetwyler et al. conceived the “U-Hack Med Gap Year” internship, a paid and completely remote experience designed to overcome the lack of exposure to research in a lab. In their perspective article, they described the administrative challenges, how they overcame several of them, in particular those limiting online employment, and the restrictions that remained (e.g., international interns with temporary VISA were not allowed). They described how the recruitment process was and steps involved, and overall subjects of 10 projects that involved themes such as the visualization of 3D data, machine learning, development of a platform in the medical field, and genome sequencing. Finally, they summarized the key features of this novel internship program: a full involvement of the interns in the research lab and in mentor-mentee process including interaction with the principal investigator, graduate students and postdocs of the host laboratory, their access and training to use institutional computing clusters, a biweekly meeting to provide them high-level training, career sessions, interaction with other interns, and support for achievement of the program learning goals.

Aron et al. described a successful sustainable training environment for Bioinformatics carried across African countries by the Pan-African bioinformatics network for H3Africa (H3ABioNet). It initially combined in-person courses, online classes, and hackathon, and later switched to a mixed-model of online and distance learning, where local institutions and instructors supported learning. First

courses in this model occurred successfully and effectively (starting in 2016), and remained active even under restrictions imposed by the COVID-19 pandemic. Authors highlighted how important hackathons organized by the H3ABioNet are to continuous applied skills development and peer learning by African scientists, engineers, and systems administrators, including the online edition that occurred imposed by COVID-19. Finally, they described how the H3ABioNet internship program (during which interns analyze their own data) has strengthened knowledge and skills of individuals, initiatives for building additional career-building skills, and bioinformatics communities through online training and support, being gender-inclusive, and promoting curated and accessible training materials. In summary, H3ABioNet initiatives have reached a very wide audience of bioinformatics users, scientists, and engineers, and together constitute a remarkable example for next multidisciplinary training in resource-limited settings worldwide.

An original research article by Dow et al. demonstrated how the U.S. Department of Energy (DOE) Systems Biology Knowledgebase (KBase) can be used by bioinformaticians and biologists to exploit several data and analysis tools. In particular, the manuscript shows how educators can use the platform in the classroom to teach bioinformatics. Authors demonstrated that educators can leverage Narratives that run workflows available in KBase, or create modified versions of existing ones to adapt to their desired learning concepts, and produce teaching resources that are FAIR, reproducible, and shareable. As part of their study, authors (mostly represented by instructors of the Educators KBase Working Group) presented results of surveys that assessed students’ perception of use of KBase resources used in several courses in American institutions such as Genomics, Metagenomics, Microbiology, and Molecular Biology, and also educators’s perspectives, showing that such resources were useful and valuable for both groups. Additionally, a study case of a metagenomics course at the North Carolina State University and results of a survey showed Kbase as a useful resource for learning new knowledge and metagenomics analysis pipelines. Finally, authors provide the link to KBase documentation, where interested educators can learn how to use or construct a new narrative or join the community for support, networking, and collaboration.

In another opinion article, Ras et al. report a set of best practices for delivering Bioinformatics training in Low to Middle Income Countries (LMICs), motivated by known difficulties associated with the organization of effective training, such as access to facilities, infrastructure, and internet, and the lack of local expertise. They divided their best practices into three categories: Planning, Development, and Implementation. In planning, they emphasize the importance of inclusion of topics of interested from the involved countries (e.g., use biodiversity data); getting support from key stakeholders, given

that a particular issue in LMICs is limitation of financial and institutional supports; fostering collaboration between countries and regions with different levels of development of bioinformatics capacity and expertise; to have a project plan, so that both administrative and training goals are achieved on time and effectively. Regarding “Development,” authors provide suggestions on organizations that have developed materials and learning resources and emphasize the importance of support, resources, and guides over the process, highlighting how important local trainers and researchers are for the success of events, in particular when these events involve international collaboration; importantly, to consider delivering introductory courses in the language spoken by trainees. Finally, regarding the “Implementation” phase, they comment on how to provide an inclusive course during the selection process, giving tips about timing for visa, accommodation and transportation, in case of face-to-face events. Importantly, authors consider socio-political situations that can lead to postponements and cancellations, so organizers must take them into account to insure costs associated with eventual reimbursements.

At the peak of initial shutdowns at the beginning of the COVID-19 pandemics (March, 2019), [Tully et al.](#) founded the Bioinformatics Virtual Coordination Network (BVCN), aiming to provide bioinformatics training resources so that wet-lab researchers could learn how to run their bioinformatic analyses while in isolation. While this was the initial objective, the open and mutable lessons in their Github repository allow learners to access and adapt the material to their own needs. In their perspective manuscript, the authors describe how they implemented the education program (which includes several topics, such as programming languages and bioinformatic analyses), all tools and platforms incorporated in lesson plans, how they planned to foster a diverse and inclusive community (including a code of conduct addressing concerns on acceptable and unacceptable behaviors), and even the evolution of topics that emerged over the process, including discussions on variations in approaches used to address research questions (such lesson discussions are made publicly available on the BVCN Youtube channel).

To map the interest and capacity for doing bioinformatics and related research in Tanzania (Sub-Saharan Africa), including human resources and infrastructure, [Sangeda et al.](#) employed self-administered online surveys to staff of public and private institutions in the country. The main purposes of such surveys were to enable leveraging of existing resources and building sustainable expertise in the country. As results, they demonstrate that most participants were early career researchers but even though the level of interest in bioinformatics is high, there is a low level of skilled human resources and a lack of infrastructure. Authors raise the importance of investing in training of undergraduate and graduate students, giving a

special emphasis to promotion use of digital resources. They encourage collaboration among local institutions and with global partners and stakeholders, highlight the importance of funding and investments from the government for growth and success of local bioinformatics, and launch other communities like the Tanzania Society of Human Genetics and Tanzania Genome Network to promote the use of bioinformatics.

[Robertson et al.](#) provide a background on the curriculum and teaching initiatives with focus on data science and analysis skills in the context of high-throughput (HT) technologies. Since these technologies have application in several areas of biology and have been advancing at an enormous pace recently, authors emphasize the importance of teaching the required basis for HT analysis to undergraduate students. They mention successful educational initiatives that gave students the opportunity to analyze data and get involved in research and refer to the High-throughput Discovery Science & Inquiry-based Case Studies for Today's Students (HITS), which focuses on development of curriculum materials based in case studies. HT case studies are complex enough and have untold stories, providing an enriching place for students to dive into data, develop quantitative skills, hypothesis development, and make discoveries; importantly, they include both the focus on experimental approaches and on quantitative data analysis, highlighting the interdisciplinary nature of the process of HT analysis in biology. Finally, they summarize the goals and achievements of the HITS network model to integrate HT discovery into curricula, that include (i) case fellows, that give opportunity to groups committed to create and validate HT case studies; (ii) an HITS annual conference, that provides an opportunity to be exposed to HT, to interact with experts, and to promote collaboration; and (iii) the HITS steering committee that provides guidance for network expansion, ensures inclusivity, and helps to disseminate products and recruit faculty.

Motivated by the lack of clear definition of bioinformatics and the associated difficulties in training and teaching topics in this steadily growing discipline, for both undergraduates in life sciences and computer science, a brief research report by [Dorn et al.](#) described the learning model implemented in the Southernmost Brazilian School of Bioinformatics (“Escola Gaúcha de Bioinformática,” or EGB). The school brings together graduate and undergraduate students from different fields and backgrounds in a place where they can learn theory and practice (including lectures and hands-on activities) and interact among themselves and with professionals in different areas of bioinformatics, including the analysis of biological sequences and structure of macromolecules. EGB introduces non-experts to bioinformatics, with focus on transdisciplinarity. Importantly, even though most participants are from the state where the school takes place (Rio Grande do Sul), it receives students from many Brazilian states (including regions where bioinformatic research and courses are scarce) and other Latin American countries.

In their “curriculum, instruction, and pedagogy” article, [Braga et al.](#) describe the course “Bioinformatics on the Road,” another Brazilian initiative from professors in two different public universities, the Federal University of Pará and the Federal Rural University of Amazônia. Motivated by the lack of skilled human resources in bioinformatics, in particular in the countryside and apart from the main national bioinformatics hubs, the scope of this course involves theoretical and practical workshops to graduate and undergraduate students with focus on the areas of interest in the regions where they take place. They provide details on capacitation events over the last 5 years (starting in 2017), their content, computer resources supporting them, and lastly the demographic information about participants, that included graduate and undergraduate students. Like most scientific events happening in 2020, practical and theoretical sessions started to be carried out online from June 2020. At least 400 students benefited from events described in the article, that happened in two cities or online. Interestingly and as part of the results, authors reported that the initiatives not only provided ways to decentralize and strengthen bioinformatics in the countryside, but also connected interested students to research groups working with bioinformatics, either as scientific initiation projects (undergraduate) or in dissertation and theses involving themes (graduate students).

In their original research article, [da Silva et al.](#) compare the in-person bioinformatics course CVBioinfo (“Summer Course in Bioinformatics”) with its online version WOB (“Online Workshop in Bioinformatics”) introduced in 2020 due to the COVID-19 pandemics. They highlighted the main structural differences between the events, and described the overall subjects covered in both editions and the correlation of interests of participants. They also summarized demographic data for participants, emphasizing an increase in the number of women and expansion of their geographical locations in the online edition (WOB). In summary, authors reported to have achieved a more inclusive and accessible event as an online edition, which was also less complex to organize, but also reported limiting factors such as lower interaction among participants, increased distraction, and difficulties in internet connection.

Closing our special edition, [Melo-Minardi et al.](#) reported an approach for evaluating an online distance university extension course—OnlineBioinfo—including its resources, activities, instructional design and student drop out, using learning analytics with machine learning methods. Course modules include an introduction to computer science related topics, basic programming with Python, algorithm complexity, and algorithms for bioinformatics, and most enrolled students belong to life sciences courses such as biology, biomedicine, and pharmaceutical sciences, also including computer science or information systems undergraduate students. Based on data from 245 students, authors were able to predict with high accuracy, even based on information gathered from module zero, whether a student would drop the course.

They could also identify types of exercise (e.g., review or programming) that could better predict students’ completion of the course. [Melo-Minardi et al.](#) also provide enriching information on the perception of students concerning different topics in programming, that can be used to improve the relationship between students’ difficulties and knowledge of each topic in their own course or serve as basis to similar initiatives.

In summary, this collection highlights important contributions to the field of bioinformatics education, and suggests future directions for research to advance the training of the current and future generations of bioinformaticians, and the curricula of professionals and students in several disciplines dealing with biological data. Importantly, our Research Topic covered authors spread across different regions of the globe. We conclude this editorial with some observed trends in the published articles:

- *The importance of building sustainable expertise in bioinformatics, including sustainable educational networks;*
- *Interpretation of “real-world” data, including COVID-19 data;*
- *The influence of COVID-19 pandemics, including movement of courses and internships that were previously in-person to an online version;*
- *The emergence of short bioinformatics courses, specially in Brazil;*
- *Materials developed and courses provided online and with manageable costs, increasingly available in repositories and whose instructors/developers are more concerned about making them FAIR (Findable, Accessible, Interoperable, and Reusable);*
- *Changes in administration of courses, internships, and training to a model adapted to online resources;*
- *Initiatives to provide access to bioinformatics skills in Low-Income countries, including recognition of challenges regarding access to technology, internet connection, and infrastructure.*

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

We thank Patricia Carvajal-López (European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge,

United Kingdom) who helped to compose the original proposal and descriptive page of our Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.