frontiers
in Education

Bioinformatics on the Road: Taking Training to Students and Researchers Beyond State Capitals

Marcus Braga[1], Fabrício Araujo[2,3,4], Edian Franco[2,5,6], Kenny Pinheiro[2], Jakelyne Silva[1], Denner Maués[2], Sebastiao Neto[2], Lucas Pompeu[2], Luis Guimaraes[2], Adriana Carneiro[2], Igor Hamoy[3] and Rommel Ramos[2]*

[1]Paragominas Campus, Federal Rural University of Amazônia, Paragominas, Brazil, [2]Laboratory of Biological Engineering, Federal University of Pará, Belém, Brazil, [3]Laboratory of Applied Genetics, Federal University of Amazônia, Belém, Brazil, [4]Castanhal Campus, Federal University of Pará, Castanhal, Brazil, [5]Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo, República Dominicana, [6]Instituto de Innovación en Biotecnología e Industria (IIBI), Santo Domingo, República Dominicana

In Brazil, training capable bioinformaticians is done, mostly, in graduate programs, sometimes with experiences during the undergraduate period. However, this formation tends to be inefficient in attracting students to the area and mainly in attracting professionals to support research projects in research groups. To solve these issues, participation in short courses is important for training students and professionals in the usage of tools for specific areas that use bioinformatics, as well as in ways to develop solutions tailored to the local needs of academic institutions or research groups. In this aim, the project "Bioinformática na Estrada" (Bioinformatics on the Road) proposed improving bioinformaticians' skills in undergraduate and graduate courses, primarily in the countryside of the State of Pará, in the Amazon region of Brazil. The project scope is practical courses focused on the areas of interest of the place where the courses are occurring to train and encourage students and researchers to work in this field, reducing the existing gap due to the lack of qualified bioinformatics professionals. Theoretical and practical workshops took place, such as Introduction to Bioinformatics, Computer Science Basics, Applications of Computational Intelligence applied to Bioinformatics and Biotechnology, Computational Tools for Bioinformatics, Soil Genomics and Research Perspectives and Horizons in the Amazon Region. In the end, 444 undergraduate and graduate students from higher education institutions in the state of Pará and other Brazilian states attended the events of the Bioinformatics on the Road project.

Keywords: education, bioinformatic, computer science, computational biology, training

INTRODUCTION

Since the first studies focused on manipulating biological sequences, bioinformatics has played an essential role in the stages of data analysis. In this context, alignment methods for comparing two or more sequences have become popular in implementing software such as Fasta (Pearson and Lipman, 1988) and Blast (Altschul et al., 1990). Since their early versions, these programs have been widely used by the scientific community, regardless of the level of knowledge in bioinformatics, even with other improved software available in both accuracy and execution time, such as Diamond (Buchfink et al., 2015).

However, after the next-generation sequencing platforms (NGS) release in mid-2005 (Schuster, 2008), there has been a significant increase in complete genome sequencing projects. The demand for new bioinformatics solutions capable of handling the data offers fast and accurate results. Despite the challenges imposed by NGS technologies, such as increasing throughput and reducing the size of reads, there are currently platforms capable of generating reads larger than 4 Mb (Fujimoto et al., 2021).

Additionally, the launch of benchtop sequencing platforms such as 454 GS Junior (GSJ), Ion Torrent Personal Genome Machine (PGM) - Life Technologies (Carlsbad, CA), Illumina MiSeq (Jünemann et al., 2013), the evolution of sequencing platforms, the availability of large volumes of biological data and the arising of new analyzes that can be performed, showed that research groups that operate mainly outside the large capitals of Brazil are still not prepared to take full advantage of these technologies because of lack of skilled human resources.

The training of research groups formed by students from different areas and levels: undergraduate, master's, doctoral, and postdoc, has become essential to give small research groups from the countryside, which have significant questions and minimal structure, a chance to participate in the genomic era, improve their research and data analysis, which certainly favors better training of students and national research. The demand for bioinformatics professionals is not exclusive to researchers who work outside the large capitals.

Thus, some professors from the Federal University of Pará and the Federal Rural University of Amazônia decided to act in the training of human resources in bioinformatics, students and researchers, through extension projects in cities from the countryside in order to increase the number of professionals trained in bioinformatics and thus help them to be independent in specific analyzes of their areas.

As bioinformatics is a useful tool for several research areas, the training demands tend to have customized specificities, which in traditional courses are not met. Furthermore, the lack of knowledge of different analyzes and sometimes regarding the possibilities limits the research groups. For this reason, the training has always been designed considering the types of research of the groups/institutions covered in those cities, which is a great innovation in terms of capacity building in bioinformatics, with hands-on activities since the wet lab to dry lab.

CAPACITATION EVENTS

In the last 4 years, five events were held in satellite cities, where the institutions offered the minimum physical structure, with computing and biology laboratories, auditoriums, local support staff, and well-established demands that can be answered by the team working in training. Mainly, the courses include theoretical and practical activities to allow participants to test their knowledge.

The project activities started in 2017, with the Computational Biology Applied to Agribusiness Meeting. The target audience were undergraduate students from the Federal Rural University of Amazônia (UFRA), from agricultural science courses, such as Agronomy, Animal Science and Forestry Engineering, and the newly created Information Systems course, in the city of Paragominas, in the southeast of State. This city is one of the largest agribusiness centers in Pará and received a green seal due to good management practices and exploitation of the environment. Thus, it is common to observe research in the region to evaluate the soil recovery time, management possibilities, including chemical and molecular analyses, where it is possible to collaborate with bioinformatics fully. This was the only course with exclusively theoretical content carried out by the team as the students were still at the beginning of their courses.

In 2019, the team expanded contact with researchers from the UFRA (Paragominas) who worked with soil analysis. Thus, it was possible to plan a special event where students had the opportunity to experience, in a biology laboratory, how a microorganism is isolated from soil samples. After, in courses at the computer lab, they had access to techniques on how to assemble prokaryote genomes and, finally, their annotation. Interestingly, the participants demonstrated that they already had different backgrounds, both biological and computational, which was important in exchanging experiences.

As part of the commitment to the dissemination of science and the training of new professionals, the project financed by the Coordination for the Improvement of Higher Education Personnel (CAPES) called PROCAD Amazônia, which involves collaboration between the Federal University of Pará (UFPA), Federal University of Minas Gerais (UFMG) and UFRA, organized the event entitled Training in Bioinformatics - Belém Stage, which aimed to immerse and train undergraduate and graduate students in bioinformatics through basic and advanced courses that addressed the standard topics of the area. The courses were given from January 13th to 17th, 2020, at the Federal University of Pará, in Belém, capital of the state of Pará.

The training was taught by graduate students who are part of the PROCAD project, under the supervision of the project's professors and researchers. The content was aimed at developing the theoretical-practical skills necessary to learn of the main bioinformatics tools and the analysis of biological data in genomics and transcriptomics to meet local demands. The training was structured with 20% theoretical and 80% practical content.

The training was split into five modules from 4 to 8 h in duration, among which we can mention: Introduction to the Linux Environment (4 h), Introduction to Programming with Python (4 h), Genome Assembly (8 h), Analysis of Transcriptomic Data: RNA-Seq (8 h) and Machine Learning Techniques: Clustering (8 h).

The Introduction to Linux Environment course had the following contents: introduction to the shell, basic Linux commands, and terminal pipeline development. Introduction to Python course had as content the installation of the python environment, main algorithms, reading biological data files, analysis methods, and data processing. The Genome Assembly course addressed topics such as data pre-processing, genome assembly algorithms, assembly of prokaryote and eukaryotic

genomes and analysis, and evaluation of assembly results. Analysis of Transcriptomic Data course contained topics such as obtaining RNA-seq data, algorithms and methods for analyzing gene expression data, differential expression analysis, and technique for data evaluation. Finally, the Machine Learning Technique course focused on clustering techniques and covered topics such as data pre-processing, clustering algorithms and their uses, metrics for evaluating results and techniques for presenting results.

The training was carried out at the Computer Laboratory of the Faculty of Biotechnology of UFPA, which has 24 computers (Intel® Core™ i5-4590S Processor, 4GB, HD 500GB, 23" LED monitor), where hands-on activities of the courses were taken.

For the selection of the participants, a web platform was developed where they could register. This made it possible for undergraduate and graduate students from different universities (public and private) to participate in the training. Ninety applications were received, from which 39 students were selected due to limited computer equipment. For the final selection of the participants, the following criteria were established: 1) be signed into an undergraduate or graduate program related to the areas of biotechnology and bioinformatics; 2) have basic computer knowledge; 3) be conducting research or graduate work in the areas of bioinformatics.

Of the 39 selected students, 60% were graduate students researching in the areas of bioinformatics and 40% were graduate students. Of the total, 80% were students from public higher education institutions and 20% from private institutions. To receive the certificate, students must have participated in at least 90% of the training hours.

As a result, 95% of the participants obtained the participation certificate for having completed the minimum hours and performed all the activities required in training. Due to the great demand for training, the organizing team decided to carry out other introductory and specific training courses, which, as a result of the pandemic caused by COVID-19, were carried out virtually. This allowed the development of activities on a national and international level. In addition to the students enrolled in the event, seven instructors participated in the courses, and three research professors coordinated the training.

Online Events

During the COVID-19 pandemic, between June and August 2020, the project team was asked to conduct training in bioinformatics, especially for students from the research group Nucleus for Research in Applied Computing, at UFRA. These students would start to work in research in bioinformatics. The objective was to train critical and scientific thinking in Computer Science undergraduate students who were starting their research in bioinformatics and, in the future, to encourage their participation in graduate programs. The training program was divided into two modules and addressed topics such as Introduction to Bioinformatics, NGS Sequencing, Biological Sequences Alignment, and Genome Assembly. The training was carried out through the Google Meet platform and was theoretical and practical.

In order to meet the demands of the computing area, in 2020, a course was held to present Artificial Intelligence techniques in the Machine Learning approach and its possible applications in bioinformatics. The Introduction to Machine Learning course was offered with theoretical and practical aspects.

The course was designed over 3-month, between October and December of 2020, carried out, 100% virtually, through the Google Meet platform. The course was held in four modules: in the first unit, called Introduction to Computational Intelligence, an overview of the area was given; the second unit was Machine Learning Fundamentals and addressed types of learning, classic machine learning problems and machine learning algorithms; the third unit dealt with Artificial Neural Networks, giving an overview of the model; and the fourth unit addressed Deep Machine Learning and Deep Neural Networks, presenting models such as Convolutional Neural Networks, Recurrent Neural Networks, Autoencoders, Generative Adversarial Networks, Attention Mechanisms, and others.

50% of the training was carried out with practical activities for the construction of intelligent models and applications in various areas of science, including bioinformatics. Frameworks like Spyder (<https://www.spyder-ide.org/>), Colab (<https://colab.research.google.com/>), Jupyter (<https://jupyter.org/>) and Orange (<https://orangedatamining.com/>) were used for the hands-on activities. The course had 80 participants, 60 of which were linked to various higher education institutions in the state of Pará and 20 linked to institutions in five other states in Brazil. Among the participants, there were undergraduate and graduate students and five professors. This initiative served to introduce intelligent techniques that can be used in bioinformatics research in their respective application scenarios.

Capacitation in Numbers

Training in bioinformatics and computational biology, through the "Bioinformatics on the Road" project, reached dozens of students in recent years, as shown in **Table 1**.

International Actions

Despite not being part of the "Bioinformatic on the road" project, international initiatives were encouraged for collaborators. In April 2019, the course "Bacterial Resistome: from wet laboratory to computational biology" was organized at the Technological Institute of Santo Domingo (INTEC). Fifteen students from different Dominican universities participated in this training, with the primary objective of showing the process for the study of antibiotic-resistant bacteria, from sample collection, DNA extraction, and, finally, the process of analysis and data interpretation through the use of bioinformatics tools. The training had 30 h, divided into 20% theoretical hours and 80% practical hours.

Also, in April 2019, the lecture "Computational Biology" was organized at INTEC, which 30 students from public and private universities attended. In this lecture, the main themes related to the processing, treatment, analysis, and interpretation of genomic data through several bioinformatics tools.

In May 2020, the lecture entitled "Rights and Challenges of Bioinformatics and Computational Biology in the Dominican

TABLE 1 | List of events held in the context of the "Bioinformatics on the Road" project with information on the place where it took place, year and total number of students and teachers.

Evento	City	Year	Students	Professors
Computational Biology Applied to Agribusiness Meeting	Paragominas	2017	10	6
Bioinformatics and Biotechnology Workshops	Paragominas	2019	300	8
Training in Bioinformatics - Belém Stage	Belém	2020	39	3
Training in Bioinformatics - NPCA	Online	2020	15	4
Introduction to Machine Learning: Theoretical and Practical Aspects	Online	2020	80	5

Republic" was given as one of the main activities of the week celebrating the 54th anniversary of the Faculty of Sciences of the Universidad Autónoma de Santo Domingo (UASD).

In July 2020, the project team participated in the "International Symposium for Research and Scientific Solutions in Times of Crisis, COVID-19 and Beyond: Food Safety, Health, Education, Environment and Economy," organized by the Ministry of Higher Education Science and Technology of the Dominican Republic (MESCYT). On occasion, the lecture: "Bioinformatics and computational biology as new ferments to face a crisis in the Dominican Republic" was given.

Lessons Learned

As a basic premise for all events, we lack of qualified personnel to work with bioinformatics and one training only cannot change it. For this reason, with decentralized training, which goes to the origin of the demands, we establish connections with local researchers and also with other collaborators from outside the State, in order to make the environment more collaborative, continuing the growth of production scientific research that is dependent on bioinformatics methods.

Some students who attended the events presented their projects, and participated in courses, also had the opportunity to keep in touch with graduate professors from UFPA and UFRA. As a result, they are already attending graduate programs. Research groups in the state capital selected some students to work in scientific initiation due to their knowledge in computing and already have a glimpse of computational biology and bioinformatics.

Graduate students were also present, especially in Belém do Pará, whose central theme was genomics. Thus, they could first contact bioinformatics and even use this knowledge in their thesis and dissertations.

The lesson that attracted the most attention regarding the Bioinformatics on the Road project is that it has collaborated with the growth of research groups that today already generate independent scientific productions with their local collaborators. Thus, we consider that the events were positive for presenting scientific thinking in the context of bioinformatics, but that it is being used in several areas to establish collaborations and increase local scientific production.

Next Steps

The COVID-19 pandemic showed us that distance training is not only possible but essential, as it opens up frontiers that were previously an obstacle, such as geographic and financial. However, in the northern region of Brazil, access to the internet from the countryside cities to carry out synchronous events is not trivial. Thus, our next steps will be the release of the project's channel on the YouTube platform to disseminate the content of online training, including exercises and answers on platforms that allow us to share the content, using active methodologies applied to remote learning.

CONCLUSION

Bioinformatics remains a promising area but with little training initiatives compared to the demand, which is much more evident outside the state capitals, which is one of the factors that justifies the tremendous demand for courses where there were many places available.

On the other hand, the offering of courses based on the interaction between bioinformatics and research themes studied by local researchers becomes more attractive to students, who end up realizing that the challenge is "only" to study bioinformatics, a way to achieve their goals and not a new research area, which leads to an increase in the number of interested people and subscribers.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

MB and RR designed and organize the steps of the project. MB, FA, EF, KP, JS, DM, SN, LP, LG, AC, IH, and RR were lecturers. MB, FA, EF, KP, JS, DM, SN, LP, LG, and AC were practical activities tutor.

FUNDING

project since its inception. Federal University of Pará Dean of Extension and Federal Rural University of Amazon Dean of Extension for the support. Federal University of Pará Dean of research (PROPESP/UFPA) for the financial support to the manuscript production.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* 12 (1), 59–60. doi:10.1038/nmeth.3176

Fujimoto, A., Wong, J. H., Yoshii, Y., Akiyama, S., Tanaka, A., Yagi, H., et al. (2021). Whole-genome Sequencing with Long Reads Reveals Complex Structure and Origin of Structural Variation in Human Genetic Variations and Somatic Mutations in Cancer. *Genome Med.* 13, 65. doi:10.1186/s13073-021-00883-1

Jünemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating Benchtop Sequencing Performance Comparison. *Nat. Biotechnol.* 31, 294–296. doi:10.1038/nbt.2522

Pearson, W. R., and Lipman, D. J. (1988). Improved Tools for Biological Sequence Comparison. *Proc. Natl. Acad. Sci. U S A.* 85 (8), 2444–2448. doi:10.1073/pnas.85.8.2444

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2021.726930/full#supplementary-material

Schuster, S. C. (2008). Next-generation Sequencing Transforms Today's Biology. *Nat. Methods* 5 (1), 16–18. doi:10.1038/nmeth115619