Experience on a Brazilian Workshop of Python for Biological Data (BWPBD) in 2017

Renato Augusto Corrêa dos Santos[1,*], Leonardo Martins Silva[2], Franciele Grego Esteves[3], Diego Mauricio Riaño Pachón[4], Alisson Hayasi da Costa[5], Ana Ciconelle[6], Bruno Ferreira de Souza[7], Flávia Vischi Winck[4], Raniere Gaia Costa da Silva[8]

1 School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, SP, Brazil.
2 Center for research and molecular diagnosis of genetic diseases, Department of Biophysics, Universidade Federal de São Paulo (UNIFESP), São Paulo, SP, Brazil.
3 Center of Study of Social Insects, Department of Biology, Institute of Biosciences of Rio Claro, São Paulo State University (UNESP), Rio Claro, SP, Brazil.
4 Institute of Chemistry, University of São Paulo, São Paulo, SP, Brazil.
5 Department of Computer Science, Federal University of São Carlos (UFSCar), São Carlos, SP, Brazil.
6 Institute of Mathematics and Statistics (IME), University of São Paulo (USP), São Paulo, SP, Brazil.
7 TIKAL TECH, 1439 Paulista Avenue Room 53, São Paulo, SP, Brazil.
8 eScience Lab, School of Computer Science, University of Manchester and Software Sustainability Institute.

* Corresponding author: renatoacsantos@gmail.com

In Brazil, there is increasing trend in the number of scientific projects involving “Omics” analyses, that is, the global analysis of each class of biological components, such as all proteins from an organism (Proteomics), all RNA molecules (Transcriptomics), among others. These approaches require bioscientists with a minimal knowledge on algorithms and computer programming. Python has been widely employed in the development of many softwares in Biology for different applications, including the analysis of atom motions in DNA molecules and application of Machine Learning to extract their features. However, as a high-level scripting language, Python is also very useful for solving daily tasks in biology such as parsing and merging long text files, or getting and handling data. Moreover, it is a relatively easy language to learn compared to low-level ones, facilitating the learning process by bioscientists. In this scenario, short courses covering computer science skills are relevant not only because they work as a starting point for people interested in improving their data science skills, but also bring basic concepts that improve their communication skills with bioinformaticians and data analysts. Here, we report our experience in a Brazilian Workshop of Python for Biological Data (BWPBD), that took place at the Institute of Chemistry of the University of São Paulo (São Paulo, Brazil) in November, 2017. Main steps in this work were i) selection of members for the organizing committee and the participants and establishment of the institutional support; ii) event broadcast and communication with participants iii) data selection, management, and preparation (adequacy to work as input for practical sessions); iv) development of practical sessions with Python programming and Jupyter notebook including also lectures from invited speakers, v) workshop evaluation through main feedback from participants provided by Google Form in the event last session. We received around 90 individual applications and we selected 17 participants from several universities in São Paulo State, Brazil. The organizing committee comprised graduate and undergraduate students in the fields of biology and computer science who were working with massive biological data, in particular the “omics” biological data. The selection criteria for the participants favored undergraduate and graduate students without any background in computer science or programming skills, and that were involved in projects with data manipulation for whom programming skills would make their work easier. The main drawbacks during the workshop were: intermittent problems with internet WiFi connection during the practical sessions; short time to provide deep demonstration of applications of programming to solve biological problems; and few post-workshop interactions of the students on Facebook group created for this purpose. A second edition of the workshop (2018) is currently under organization and development. We expect to improve our approach, comparing to the first edition, such as including more examples on data handling in files, and also exploring statistics (getting statistics metrics, executing hypothesis tests, data exploitation/ visualization). Thus, our first experience will contribute to improve our future workshops making them enriching environments for maximizing learning of novel programming and communication skills, which is very important to establish an efficient communication with data analysts and bioinformaticians. Lastly, we expect these workshops to enlarge the participants scientific networks, and that initiatives such as this can promote ways of facilitating the processing of biological data and their better understanding.