Title of Tutorial:

Introduction to Python programming for bioscientists

Abstract for Tutorial:

Programming skills are increasingly necessary for scientists working with biological data analysis and bioinformatics. Python has been widely used in biology and it is a high-level programming language, which makes it relatively easy to learn compared to others. In this tutorial, we introduce the first steps in analyzing biological data using Python in digital notebooks, which facilitates code documentation, real-time visualization of results, and sharing. The practice will be carried out on Google Meet and Google Colab. Basic programming tasks will be presented, including variable assignment employing the main data structures (e.g., strings, lists and dictionaries), data types (e.g., numbers and sequences of characters), and operations (e.g., loops, comparisons, and decision structures). To provide a real example of the Python application in bioinformatics, we will use SARS-CoV-2 amino acid alignment data in a case study to apply data structures and methods from the Biopython toolkit to obtain information about different COVID-19 variants.

Learning Objectives for Tutorial:

- Introduce Google Colab digital notebooks;
- Present the basic logic and data structures in Python;
- Provide hands-on experience in analyzing biological sequences using Biopython.

Maximum number of attendees participating

30 attendees

Draft Schedule of the tutorial including coffee breaks, including draft talk titles or draft content to be covered in each section

11 am - 1 pm	Hands-on session: Introduction to Google Colab and Jupyter notebooks; Introduction to Python data structures and programming logic;
1 pm - 2 pm	Lunch
2 pm - 3:15 pm	Hands-on session: Introduction to Python data structures and

	programming logic;
3:15 pm - 3:30 pm	Coffee break
3:30 pm - 4:45 pm	Hands-on session: Importing sequences and analyzing alignments with Biopython, including examples employing the main methods to manipulate sequence alignments
4:45 pm - 5:00 pm	Coffee break
5 pm - 6 pm	Hands-on case study: comparison of sequences from relevant genomic loci using basic data structures Analysis of alignment of Spike Protein from several SARS-CoV-2 variants, including the identification of conserved regions (e.g., domains) and mutations that are unique to each variant, with Bio.AlignIO from the Biopython toolkit.

Identify and highlight blocks of hands-on content in your submission:

1. Teaching how to use Google Colab / Jupyter digital notebooks and showing students how they can be used for documentation as well as for visualizing and sharing scientific results;
2. Teaching introductory topics in programming logic and data structures;
3. Providing an example of how to use basic programming skills to analyze biological sequences using SARS-CoV-2's Spike Protein as a case study (problem-based session involving learned data structures, e.g., strings).

Draft List of Tutorial Speakers with titles and affiliations:

Dr. Renato Augusto Corrêa dos Santos (renatoacsantos@gmail.com)

University of Campinas, UNICAMP

ORCID: https://orcid.org/0000-0003-0826-5479

Ph.D. candidate Hemanoel Passarelli Araujo (Hemanuel.passarelli@gmail.com)

Federal University of Minas Gerais (UFMG)

ORCID: https://orcid.org/0000-0003-4783-6641

Undergraduate student Vinícius Henrique Franceschini dos Santos (vinicius6.santos@usp.br)

University of São Paulo (USP)

ORCID: https://orcid.org/0000-0001-9276-0370

Master student Pedro de Carvalho Braga Ilidio Silva (ilidio@alumni.usp.br)

University of São Paulo (USP)

ORCID: https://orcid.org/0000-0002-5750-0204

Intended audience and level:

Researchers at different education levels with interest in learning programming skills for bioinformatics; level of programming skills: beginners.