H2020-WIDESPREAD-2017
Twinning

MiCoBion

Microbial Communities in Biomedical and Environmental Areas, and Systems Biology

Starting date of the project: 01/09/2018
Duration: 51 months

= Deliverable: D1.4 =

Report on organized courses Computational Genomics

Due date of deliverable: 30/06/2022
Actual submission date: 22/06/2022

Responsible WP: Vladimír Beneš, WP1, EMBL
Responsible TL: Marc van Ranst, KUL
Revision: V1.1

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

AUTHORS

Author	Institution	Contact (e-mail, phone)
Jan Tachezy	CUNI	tachezy@natur.cuni.cz

DOCUMENT CONTROL

Document version	Date	Change
V1.0	17/06/2022	First draft prepared
V1.1	21/06/2022	Edits by Project Manager

VALIDATION

Reviewers		Validation date
Work Package Leader	Vladimír Beneš	22/06/2022
Project Manager	Tomáš Palatý	21/06/2022
Coordinator	Jan Tachezy	22/06/2022

DOCUMENT DATA

Keywords	Computational Genomics, data
Point of Contact	Name: Jan Tachezy Partner: CUNI Address: Průmyslová 595, 252 50 Vestec, Czech Republic Phone: +420 325 873 908 E-mail: tachezy@natur.cuni.cz
Delivery date	22/06/2022

DISTRIBUTION LIST

Date	Issue	Recipients
22/06/2022	V1.1	Partners via Google Disc, EC

DISCLAIMER

Executive Summary

This document provides a summary of a completed task within Work Package WP1 *Training*, namely: T1.4 *Course in Computational Genomics* led by the Catholic University in Leuven (KUL). It contains description of the content of the organized course, presents details about the events' participants and provides detail about the course outcomes and benefits brought to the Charles University (CUNI).

Table of Contents

1. Introduction

Deliverable D1.4 *Report on organized courses Computational Genomics* is associated with task T1.4. The objective of this task was to improve knowledge in bioinformatic analysis and interpretation of individual DNA, RNA, and amino acid sequences as well as large data analysis. The task consisted of three courses, which were held in an intensive three-day workshop format in February 13-15, 2019; February 12-14, 2020; and June 8-10, 2022. The computational Genomics courses were performed in the computer room PuA at CUNI by experts from KUL and from CUNI.
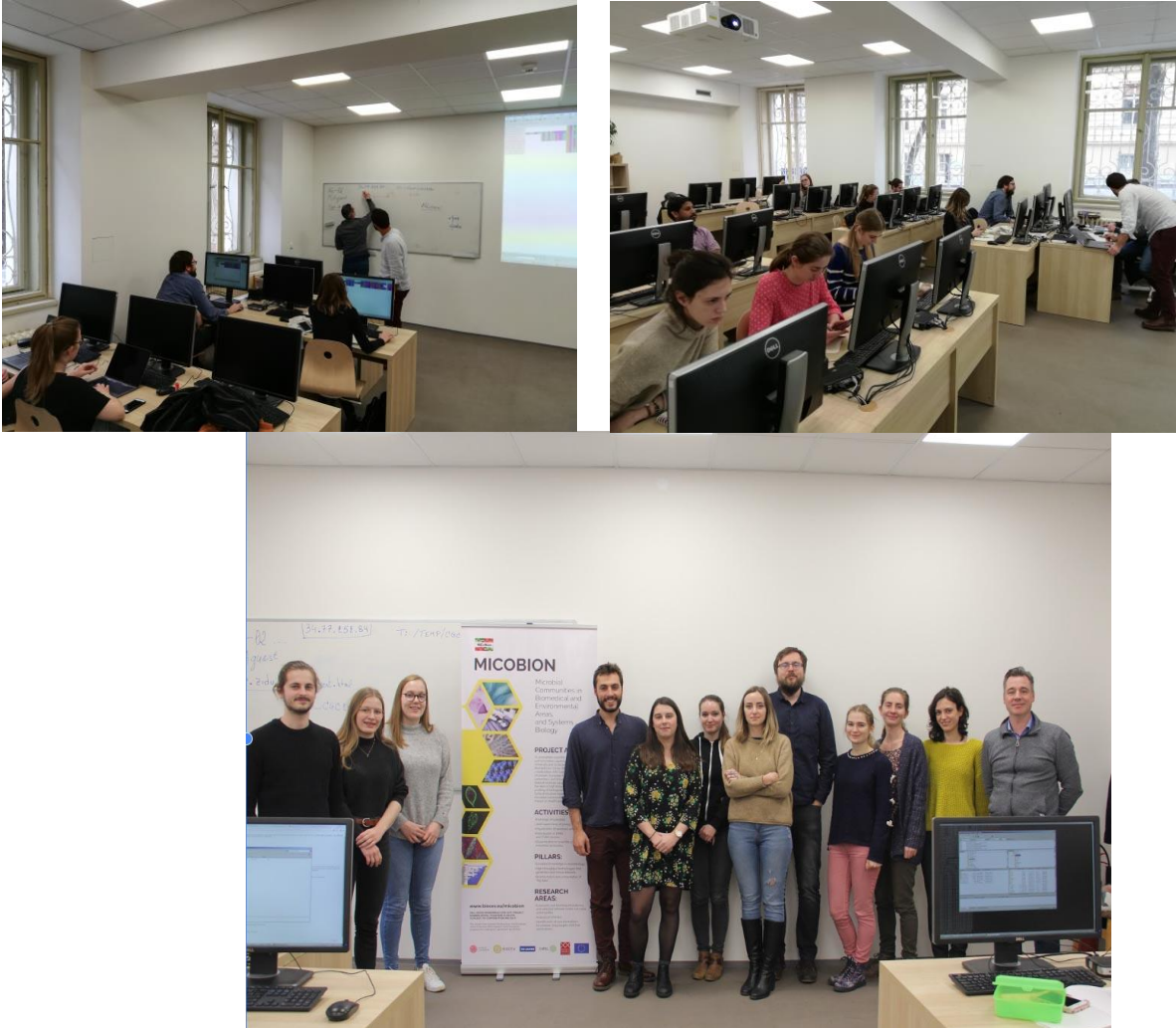
2. Course Participants

Attendees of the course were scientists and students from the Charles University (CUNI) and the Catholic University of Leuven (KUL). In 2019, there were 19 out of 20 applied participants. Seventeen participants were from CUNI and two from KUL. The summary of the event is provided below and can also be found on the website: https://www.biocev.eu/cs/o-nas/projekty/micobion.2/course-in-computational-genomics.168.

Computational Genomics 2019

The Course Computational Genomics 2020 was attended by twelve scientists and students, eight from the Charles University and four from the Catholic University of Leuven. Details of the course content are provided below.

Computational Genomics 2020

—

The third course in Computational Genomics was postponed due to COVID-19 pandemic to June 8-10, 2022. This last course was attended by twelve participants from the Charles University and a one from Catholic University of Leuven.

Computational Genomics 2022

3. Course Content

The first course presented introduction into Computational Genomics and provided overview on available tools in various web sites that can be used for analyses of nucleotide and amino acid sequences. The main aim was to provide such information/knowledge that each participant who passed the course would be able to use all these tools from his/her local computer with help of provided bookmarks.

Program 2019:

Day 1
General introduction
- DNA/protein sequence databases
- Sequence submission
- Nucleotide sequencing projects
- DNA/protein primary and secondary structure analysis
- Comparative genomics

Day 2
- Protein domain/motif analysis

- Protein sorting/localization
- Protein secondary structure prediction
- Protein tertiary structure visualisation

Day 3
- Phylogenetic analysis: MEGA

The course Computational Genomics 2020 included introduction of Linux based commands for communication with server at KUL.

Program 2020:

Day 1
Introduction to bioinformatics
Nucleotide and protein databases
    Biological databases
Sequence retrieval from databases
Database alert servers
Genomics
    Human genome project
    Other genome projects
    Comparative genome analysis
DNA sequence analysis
Sequence format conversion
Contig assembly
Open reading frame prediction

Day 2
Protein sequence analysis
    Protein domain/motif analysis
    Protein sorting/localisation
    Hydrophobicity plots
    Reverse translation
Alignment scoring/substitution matrices
Pairwise DNA/protein homology database searches

Day 3
Multiple sequence alignment
Editing/correcting multiple sequence alignments
Phylogenetic analysis
Secondary/tertiary structure prediction
RNA secondary structure prediction
Protein secondary structure prediction
Protein tertiary structure visualisation

Program 2022:
Course Computational Genomics 2022 was more focused on Linux system, communication with server via command lines, and extension to the area of next gen sequencing, particularly analysis of data generated by Illumina and MinION technology.

Day 1:
Introduction to Linux
Sequence retrieval/databases/sequencing manipulations and corrections
Homology/similarity
Alignments (online/server)
Alignment corrections and scoring
Homology/similarity searches
BLAST (online/server)
2D RNA protein structures / 3D / transmembrane domains

Day 2:
Introduction to next-gen sequencing
Basic next-gen illumina dataset analysis
Basic Oxford nanopore technologies MinION dataset analysis

Day 3:
Phylogenetic tree reconstruction (NJ, ML, visualization)

4. Course Outcomes

Professional Experience: For young students and researcher, which are not specialist in bioinformatics, the first two courses provided useful tools and experience in initial analyses of DNA, RNA and proteins sequences. This is an important step for searches and selection of gene/proteins of interest, which may serve for example for detail phylogenetic analyses or which precede experimental on-bench investigations. The participant gained during the course the large set of bookmarks with useful servers for bioinformatics, and they learned how to use these servers. Next, they gained experience how to analyse DNA/RNA/proteins sequences using bioinformatic tools that required communications via command lines, and introductory for analysis of next-gen big data.

5. Conclusions

T1.4 *Course in Computational Genomics* was successfully completed concerning the content of the courses, their numbers, and participation of KUL and CUNI lecturers. The number of participants in the first course met our expectations, in following two courses the numbers were lower (12 and 13) which reflected difficulties for participants to attend the course due to COVID-19 pandemic and difficulties to organize the last course, which was several times postponed because of the pandemic from the original date February 2021 to June 2022.

6. Degree of Progress

The deliverable was 100% fulfilled.

7. Dissemination Level

The Deliverable 1.4 Course in Computational Genomics document is a public deliverable.