



*H2020-WIDESPREAD-2017*  
*Twinning*

## **MiCoBion**

**Microbial Communities in Biomedical and Environmental Areas, and Systems Biology**

Starting date of the project: 01/09/2018  
Duration: 36 months

### **= Deliverable: D1.5 =**

**Organized Advanced course in metagenomics and OMICs data integration**

Due date of deliverable: 31/07/2019  
Actual submission date: 30/07/2019

Responsible WP: Vladimír Beneš, WP1, EMBL  
Responsible TL: Pavel Stopka, CUNI  
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)	



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 810224.*

**AUTHORS**

Author	Institution	Contact (e-mail, phone)
Pavel Stopka	CUNI	<a href="mailto:pavel.stopka@natur.cuni.cz">pavel.stopka@natur.cuni.cz</a>

**DOCUMENT CONTROL**

Document version	Date	Change
V1.0	28/07/2019	First draft prepared
V1.1	30/07/2019	Edits by Project Manager

**VALIDATION**

Reviewers	Validation date
Work Package Deputy Martin Pospíšek	30/07/2019
Project Manager Tomáš Palatý	30/07/2019
Coordinator Jan Tachezy	30/07/2019

**DOCUMENT DATA**

<b>Keywords</b>	Course, metagenomics, OMICs
<b>Point of Contact</b>	Name: Pavel Stopka Partner: CUNI Address: Průmyslová 595, 252 50 Vestec, Czech Republic  Phone: +420 325 873 907 E-mail: <a href="mailto:pavel.stopka@natur.cuni.cz">pavel.stopka@natur.cuni.cz</a>
<b>Delivery date</b>	28/07/2019

**DISTRIBUTION LIST**

Date	Issue	Recipients
30/07/2019	V1.1	Partners via Google Disk, EC

**DISCLAIMER**

Any dissemination of results reflects only the authors' view and the European Commission Horizon 2020 is not responsible for any use that may be made of the information this document contains.

## Executive Summary

This document provides a summary of a completed task within Work Package WP1 *Training*, namely: T1.5 *Advanced course on metagenomics and OMICs-data integration* led by the Charles University Prague – BIOCEV (CUNI, P. Stopka). It contains description of the content of the organized course, presents details about the event's participants and provides details about the course outcomes and benefits brought to CUNI.

## Table of Contents

Table of Contents.....	4
1. Introduction .....	5
2. Course Participants.....	6
3. Course Content.....	6
4. Course Outcomes .....	8
5. Conclusions .....	8
6. Degree of Progress .....	8
7. Dissemination Level.....	8

## 1. Introduction

Deliverable D1.5 Organized *Advanced course in metagenomics and OMICs-data integration* is associated with WP1-training task and is organized at the same time as task T4.4 *Workshop OMICs data integration*. The objective of this task is to improve current knowledge in metagenomics data analysis and OMICs data integration. It was taught by four guests and three CUNI lecturers. The course was scheduled for 3 days and for 30 trainees. The rationale of this event was to organize both the course (T1.5) and the workshop (T4.4) at the same time, so the workshop participants may also provide training for students and young researchers. Monday lectures (June 3<sup>rd</sup>) organized within the T4.4 workshop were accessible for T1.5 course attendees. Tuesday (June 4<sup>th</sup>) course and Wednesday (June 5<sup>th</sup>) troubleshooting, however, were solely devoted to practical training.

## 2. Course Participants

Attendees of the course were young scientists and students from BIOCEV/CUNI (Figure 1): Monday, 3<sup>rd</sup> June - 33 people from BIOCEV (11 students, 22 employees) and 5 from other departments within the BIOCEV framework; Tuesday, 4<sup>th</sup> June and Wednesday, 5<sup>th</sup> June - 26 attendees from BIOCEV (12 students, 14 employees) and 5 from other institutions within BIOCEV.



Figure 1: The course trainees

## 3. Course Content

*Advanced course in metagenomics and OMICs-data integration* was held on 3 – 5 June 2019, at CUNI in BIOCEV lecture hall (**Monday, June 3<sup>rd</sup>**) and BIOCEV seminar rooms. Prior to the meeting we (P. Stopka, J. Kreisinger at CUNI) have generated large datasets from several experiments including metagenomic, proteomic and metabolomic datasets. The idea was to teach attendees essential approaches in microbiome analyses. All participants had their own notebooks with packages downloaded prior to the course.

### Tuesday (June 3<sup>rd</sup>):

The course began with practical introduction to microbiome data manipulation in R-software. CUNI Lecturer dr. Jakub Kreisinger introduced basic algorithms to gene mapping (bacterial 16S RNA), to data normalization and cleaning-up redundant sequences (name of the block: ***R tools for microbiota profiling***). This step let to people analysing the data on their computers. The course was based on two levels (slower and faster), students could either map and normalize the data themselves or (in case they were slower) they could follow the instruction and skip their own work and the use pre-processed data to carry on further analyses on microbial (alpha/beta) diversity. The training was reasonably slow so each participant had chance to follow and perform all procedures. When any kind of data incompatibility or system problems were encountered the scripts and packages were transferred on flash disks.



Figure 2: Pavel Stopka helping to setup data-networks



Figure 3: dr. J. Kreisinger and dr. Michal Vinkler are demonstrating how to upload metagenomic data

Following J. Kreisinger's training, another lecturer (**Andres Gomez: *Marker discovery from microbiome data, using machine learning, ecological tools and network theory***) continued with advanced analysis of metagenome. The trainees were taught how to detect 'structures' in hierarchically organized large datasets using hierarchical clustering and deep-machine-learning algorithms (e.g. *randomForest* app.) and how to interpret structures within datasets with various types of PCA analyses. This hands-on-training was very successful because all participants



managed to run this analysis on their notebooks at the same time as the training continued and again either with raw or pre-processed data.

The last step of training aimed to teach attendees how to integrate datasets. All attendees were advised to upload proteomic and metabolomic data from one experiment and dr. Denis Laloe trained the participants how to detect similarities between the two OMICs datasets using multivariate approaches (***Challenges in big data integration; Multivariate approaches – training session***).

### Wednesday, 5<sup>th</sup> June:

This part of the course was prepared such that students were involved in troubleshooting and made their routines applicable to their own data. We used three separate rooms including the seminar room above (Figure 1, Figure 2, Figure 3) so each trainee had chance to separately discuss their scientific problems with one of the mentors.

## 4. Workshop Outcomes

**SCRIPTS:** All participants of the course received R-SCRIPTS (statistical routines) which could be immediately used for OMICs data elsewhere. This was particularly important for those attendees that were beginners and could not follow all steps of the training in time.

**TROBLESHOOTING:** Each participant had individual discussions with other participants which helped to improve the precision and quality of analyses within already existing projects.

**EXPERIENCE:** All attendees of the course had unique opportunity to meet international lecturers because this course was organized at the same time and in the same BIOCEV building as T4.4 workshop. This made possible that young researchers had chance to discuss their topics similarly as scientists do in conferences.

## 5. Conclusions

T1.5 *Advanced course on metagenomics and OMICs-data integration* was successfully completed. Internationally sound scientists were invited and three CUNI scientists took part (Erica Leder, Dennis Laloe, Dustin Penn, Andres Gomez, Pavel Stopka, Jakub Kreisinger, Michal Vinkler) who provided hands-on training on data integration within their field related to metagenomics and data integration. The training was successful because each participant received R-scripts and pre-processed data so each of them had chance to analyse data in real time and because training was held at the same time as T4.4 workshop there was a synergic effect between these two tasks.

## 6. Degree of Progress

The deliverable is 100% fulfilled.

## 7. Dissemination Level

The Deliverable D1.5 *Organized Advanced course on metagenomics and OMICs-data integration* document is public.