Marine
Microbial Biodiversity,
Bioinformatics & Biotechnology

Grant agreement n°287589
Acronym: Micro B3
Start date of project: 01/01/2012, funded for 48 month

Deliverable 9.8

Report on final course/jamboree including OSD participants

Version: 1.1
Circulated to: WP9 Training & Dissemination Committee (08/04/2015)
Approved by: Prof. Frank Oliver Glöckner, (12/05/2015)

Expected Submission Date: **30/04/2015**
Actual Submission Date: **11/05/2015**

Lead Party for Deliverable: CNRS
Mail: pascal.hingamp@univ-amu.fr **Tel.: +33 6 34 68 80 91**

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

The Micro B3 project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 287589 (Joint Call OCEAN.2011-2: Marine microbial diversity – new insights into marine ecosystems functioning and its biotechnological potential).

Summary

The deliverable N° 9.8 "*Report on final course/jamboree including OSD participants*" is an action of WP 9 "*Dissemination & Outreach*" under task 9.4 "*Training*". The course/jamboree brought 30 participants (the majority of whom represented OSD 2014 sampling sites) together with 14 lecturers/organizers (representing 8 distinct Micro B3 partners). The week long event, hosted by the EMBL-EBI on the Hinxton Genome Campus in Cambridge UK, took place from March 9th through March 13th 2015. The aim of the course/jamboree was three fold: i) to present participants with the most current OSD 2014 datasets (both metagenomes and metabarcodes, as well as environmental parameters from each sampling site), ii) conceive analyses strategies for the OSD dataset and iii) start collaborative implementation of the analyses strategies. A major outcome of the course/jamboree was the setting up of the OSD Analysis Group formed by both OSD participants and Micro B3 organizers which has set itself the task of writing a consortium paper around the analysis of the OSD 2014 dataset.

Table of Contents

Introduction

As planned in the "*Dissemination & Training Plan*" (D9.2), a final training course was carried out in March 2015 (month 38 of the Micro B3 project). This training course had an explicit triple objective:

1. present the available OSD 2014 data to OSD participants (data location, format and essential metrics)
2. design a collaborative analysis strategy for the OSD dataset
3. kick start analyses during hands-on sessions

The final training course was designated "OSD Analysis Jamboree" in order to stress the fact that, together with classical lectures by Micro B3 organizers, this event would also include collective hand-on work sessions involving both participants and MicroB3 organizers.

OSD Analysis Jamboree preparation

In preparation for the event, on 20th September 2014 an OSD Analysis Core Group was formed to lay out detailed plans for analyses of the OSD dataset. The 21 person strong group (table 1) represented 10 Micro B3 partners involved in the 6 work packages related to OSD data analysis (WP1-WP6).

Anna Kopf née Klindworth, Jacobs Univ.
Antonio Fernandez-Guerra, MPIMM
Carlos Pedrós-Alió, ICM
Chris Bowler, CNRS - ENS
Christian Jeanthon, CNRS - SBR
Daniel Vaulot, CNRS - SBR
Daniele Iudicone, SZN
Declan Schroeder, MBA
Emilie Villar, CNRS - AMU
Frank Oliver Glöckner, Jacobs Univ.
Gunnar Gerdts, AWI
Guy Cochrane, EBI
Ivaylo Kostadinov, MPIMM
Julia Schnetzer, MPIMM
Katja Lehmann, UOXF
Mesude Bicak, UOXF
Pascal Hingamp, CNRS - AMU
Pier Luigi Buttigieg, AWI
Renzo Kottmann, MPIMM
Stéphane Pesant, MARUM
Adriana Zingone, SZN

Table 1: OSD Analysis Core Group members

The OSD Analysis Core Group held monthly teleconferences from September to April, and also had a face-to-face meeting during a satellite Micro B3 Extended Executive Board Meeting on the 13th Jan. 2015 in Bremen. The group monitored the progress of the OSD data production during fall 2014, as well as provided council on the methods and protocols for the data quality controls and primary analyses. A collaborative online document was used to record the group's decisions and action items (see Annex A1).

In December 2014, the OSD Analysis Core Group sent an invitation out to all the OSD sampling site participants (Annex A2) to join the 9-13th March Jamboree. The event was also advertised on the Micro B3 web site (http://www.microb3.eu/events/workshops/osd-analysis-workshop). The EMBL-EBI host for the event was in charge of the logistics of the event (and to this effect opened a specific web page http://www.ebi.ac.uk/training/course/micro-b3-osd-analysis-workshop-2015).

The OSD Analysis Core Group designed the Jamboree program, with roughly lectures for half the day, followed by hands-on sessions (Table 2). A total of 14 scientists from 8 Micro B3 partners participated in the event as lecturers and jamboree organizers (Table 3, Illustration 1). Out of the 39 applicants the OSD Analysis Core Group selected shortlisted the 30 participants (Table 3), to take part in the Jamboree. In order to help in forming a group with broad research expertise and interests, an extensive application form collected information on scientific questions that the applicants would wish to address with the OSD dataset. As the number of applicants exceeded the number of participants the venue could safely accommodate (Illustration 2), priority was given to applicants who were associated with an OSD 2014 sampling site.

Outcomes of the 9-13th March 2015 OSD Analysis Jamboree

As attested by the OSD Analysis Jamboree participant exit survey carried out within 4 days after the end of the event by the EMBL-EBI host (average 18 responses, Annex A3), the overall rating for the event is above 88% positive (rated either “good” or “excellent”). The “*Numerical ecology approaches*” session was particularly well received (88% rated this lecture as “excellent”). Importantly, 94% of the respondents stated that they will use the resources and tools presented during the Jamboree, indicating that the chosen program effectively covered future needs of the participants. The main criticisms (A3 - Q7) relate to the incomplete availability of the OSD 2014 dataset during the Jamboree. Indeed, following the late early January delivery of sequence data by the sequencing subcontractor, a subset of the data was still being processed for quality checks, cleaning and initial annotations during the actual event, with some analyses becoming available only mid-week or after the event. The free text comments (A3 - Q6) show that the collaborative atmosphere was the striking positive feature of this event.

Indeed, a complementary and collective draft work program for the comprehensive analysis of the OSD 2014 dataset was designed during the jamboree (Table 4).

Shortly after the end of the Jamboree, these analysis proposals were grouped along three main themes: diversity, function, global & ecological patterns. An invitation to OSD participants not present at the jamboree has further extended the breadth of the proposed analyses. Moreover, all jamboree participants were invited to join the OSD Analysis Consortium which has been set up with the aim to write a collective OSD Consortium publication to be submitted in the course of 2015. As such, the aim of the event to put the OSD data into the hands of the OSD participants and to kick-start the OSD dataset analysis is fully attained.

	Monday 09.03	Tuesday 10.03	Wednesday 11.03	Thursday 12.03	Friday 13.03
09:00-10:00		First steps before analysis: quality control of the OSD sequence data (Antonio Fernandez)	Numerical ecology approaches to analysing high-dimensional data & sharing analytical knowledge: GUSTA ME (Pier Luigi Buttigieg)	Biogeography: concepts and modelling (Daniele Iudicone)	OSD Paper (Mesude Bicak)
10:00-10:30 Tea/Coffee Break					
10:30-12:00	Arrival	10-11: Keynote Lecture (Kendrew Lecture theatre) 11: 30 - 13: 00 The M2B3 Standards (Petra ten Hoopen)	MG-Portal (Alex Mitchell)	Oceanographic data (Stephane Pesant)	Summary & OSD 2015 (Pascal & FOG, Anna)
12:00-13:00 Lunch					
13:00-15:30	12:30-13:00 Introduction to the EBI (Laura Emery) Jamboree organisation (Pascal Hingamp Chris Bowler)	14:00: Alex Mitchell gives first overview of how to browse MG-Portal for OSD data 14:10: Stephane Pesant OSD environmental data	hands-on (PLB) - data exploration - checking assumptions - associating responses to their context - hypothesis testing - sharing analytical knowledge	hands-on (FOG)	Departure
15:00-15:30 Tea/Coffee Break					
15:30-17:00	From OSD 2014 to the data we have: Accessing the OSD data at ENA + MB3IS (Ivaylo Kostadinov)	All: Overview • Where is data • Contact persons Discussion of Analysis Themes: • Who wants to do what • What data in which format is needed as input	hands-on (GLM,PH) -co-occurrence -amongst other themes...	hands-on (KL)	
		16:30-17:00 Sharing of analysis results			
19:00 Dinner at The Red Lion, Holiday Inn Express, Duxford				**19:00** Downing College	

Table 2: Program of the March Jamboree

Jamboree Participants	Jamboree Organizers
Linda Amaral-Zettler	Chris Bowler, CNRS-ENS
Dhwani Desai	Pier Luigi Buttigieg, AWI
Hugo Doré	Laura Emery, EMBL-EBI
Bernardo Duarte	Antonio Fernàndez-Guerra, MPIMM
Maribeth Gidley	Frank Oliver Glöckner, Jacob's Univ.
Kelly Goodwin	Pascal Hingamp, CNRS-AMU
Tim Kahlke	Daniele Iudicone, SZN
Mahrous Kandil	Anna Kopf, Jacob's Univ.
Janset Kankuş	Ivaylo Kostadinov, MPIMM
Gipsi Lima Mendez	Renzo Kottmann, MPIMM
Catarina Magalhaes	Katja Lehmann, UOXF
Clara Magalhães Loureiro	Alex Mitchell, EMBL-EBI
Loïs Maignien	Stéphane Pesant, MARUM
Francesca Malfatti	Petra ten Hoopen, EMBL-EBI
Jonas Mortelmans	
Anastasis Oulas	
Paraskevi Polymenakou	
Grazia Marina Quero	
Riccardo Rosselli	
Pablo Sanchez	
Christopher Sinigalliano	
Jennifer Tolman	
Maria Paola Tomasino	
Alessandro Vezzi	
Emilie Villar	
Shiao Wang	
Bryan Wilson	
Adriana Zingone	

Table 3: March Jamboree participants

1. Abundance of specific functions across OSD samples
e.g.: AMOA, nitrogen cycle, DMSP, phosphorous cycle, oxygenase
2. Minimum Entropy Decomposition (MED) + oligotyping
Phase I: MED
Phase II: Oligotyping of user-picked taxonomic bins
3. Abundance of fecal / pathogen markers
4. Co-occurence network
5. Estimation of taxonomic composition without a reference DB
6. Picocyanobacteria abundance (PetB)
7. Metagenome traits
8. Harmful algae
9. Functional profiling; mapping onto KEGG pathways + MAPLE
10. Latitudinal biodiversity gradients
11. Ocean comparisons (grouping by Longhurst Provinces): dispersal, connectivity, emerging biogeography
12. Contig binning
13. Giant virus abundance
14. Assembly free metagenomic comparisons

Table 4: OSD data analyses proposed during the OSD Analysis Jamboree

Illustration 1: Participants of the March 2015 OSD Analysis Jamboree at EMBL-EBI

Illustration 2: Jamboree hands-on session in the EMBL-EBI computer room

Annexes

Annex A1: OSD Analysis Core Group online documentation

- 1 - Objectives
- 2 - Timeline
- 3 - Datasets (description/overview)
- 4 - Analyses
    - Primary analyses (preparing raw data: cleaning, normalising, annotating)
    - Environmental Data
    - Illumina Pre-processing Workflow
    - Illumina pre-processing workflow
    - Barcodes (rRNA amplicons)
    - Metagenomes
    - Secondary analyses (answering biological questions)
    - Diversity
    - Function
    - Ecological Traits
    - Co-occurrence
- 5 - OSD Analysis Training/Jamboree Organisation
- 6 - OSD Analysis Manuscript
- 7 - Core Group Participants contacts
- 8 - Email to OSD participants on OSD sequencing update and March Analysis Training/Jamboree
- 9 - Email to OSD Analysis Consortium
- 10 - Email to OSD Community
- 11 - Email to OSD Analysis Proposal Contributors
- 12 - Telecons, Agenda & Actions
    - October 2, Thursday 1pm
    - November 12, Wednesday 12pm UK time
    - December 17, Wednesday 12pm UK time
    - January 13th 2015, 9-12am, EEB Satellite OSD analysis meeting
    - February 17th 2015, Tuesday 4pm UK Time
    - March 3rd 2015, Tuesday 5pm UK Time
    - April 13th 2015, Monday 3pm UK Time - OSD Paper Taskforce Telecon
    - April 17th, Friday 3pm UK Time - OSD Paper Core Group Telecon
    - Thursday, 30 April at 3pm UK time - OSD Paper Core Group Telecon

Annex A2: Invitation to join the OSD Analysis Jamboree

Subject: Update on OSD sequencing and March Analysis Jamboree
Date: Thu, 18 Dec 2014 17:09:23 +0100 (CET)
From: Dr. Anna Kopf née Klindworth <aklindwo@mpi-bremen.de>
To: Osd contact <osd-contact@microb3.eu>

Dear OSD participants,

We wish to keep you updated on four OSD fronts: 1/ Sample sequencing status, 2/ OSD Analysis Core Group, 3/ March 2015 OSD Analysis Jamboree, 4/OSD 2015

1/ Sequencing Status Protocol A (Sterivex Filtration):
The preparation of metagenome shotgun libraries by LGC is on schedule, with first results expected also in early January. Concerning prokaryotic barcodes, delivery of the 16S sequencing results by our sequencing partner LGC (initially expected this December) has been postponed by the company until mid-January. Moreover, we are pleased to announce that we managed to secure additional 18S sequencing from the sterivex filters. Please also have a look at attached spreadsheet.

Sequencing Status Protocol B (LifeWatch Collaboration using different type of filters):
OSD sample DNA extracts have been sent to Italy in order to produce eukaryotic 18S barcodes through the OSD/Lifewatch collaboration. PCR amplifications followed by sequencing are about to start, with sequences expected to be ready by the end of January.

2/ The “OSD Analysis Core Group” (see OACG tab in attached spreadsheet) has been set up to coordinate a global analysis of the OSD dataset, hopefully leading to an OSD consortium publication. In the short term, the OACG’s task has been to design & build a “pre-processing pipeline” to demultiplex, trim, quality control and submit all OSD sequence datasets to the ENA international archives as soon as the raw data is delivered by our sequencing partners. In order to ensure that the full OSD dataset is a uniform dataset supporting cross-comparisons, data from both OSD pilots and OSD June 2014 will be processed and submitted to ENA in one single batch, hopefully completed by mid-February (note that all OSD sequence datasets submitted to ENA will be made public immediately as stated in the Fort Lauderdale rules for sharing data from large-scale biological research projects http://www.genome.gov/pages/research/wellcomereport0303.pdf). The next task for the OACG will be to push these pre-processed OSD datasets through “primary analysis” pipelines (e.g. SILVAngs, BioVel MG-Traits, EBI MG-Portal) in order to provide the OSD community with a first set of standardized results, including OTU clustering, abundance estimates, and taxonomic/functional annotations.

3/ We are pleased to announce that an OSD Analysis Jamboree will take place on the 9-13th March 2015 at the EBI Training Centre in Cambridge, UK. The jamboree will bring together OSD participants and members of the OSD Analysis Core Group with the aim of transforming the OSD primary analysis results described above into biological and ecological meaning. OSD participants wishing to participate in the jamboree will be invited to register their interest early in January. Because of the size of the venue, the maximum number of participants will be limited to 30. More information will be made available at:
http://www.microb3.eu/events/workshops/osd-analysis-workshop

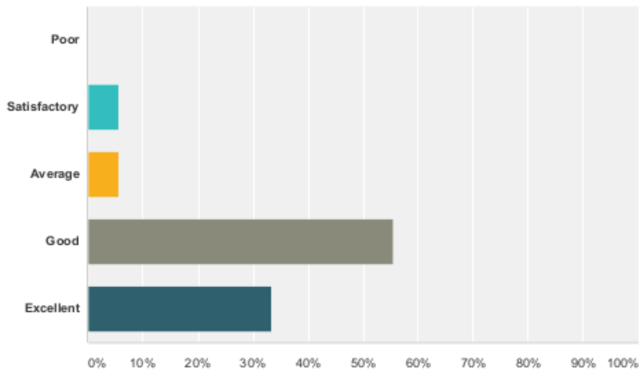
4/ Please note that OSD 2015 will happen, however, we are not sure to what extent. It all depends on our resources. We have another extended executive board meeting of the Micro B3 project in January and we plan to finalize the plan for OSD 2015 during that meeting. For now I would kindly ask you to simply save the date for the next OSD on June 21st 2015.

Be prepared for OSD 2015: www.oceansamplingday.org

Best wishes and season’s greetings to all!
Anna on behalf of the OSD Core Team
__
Attached: one spreadsheet file with 2 sheets: i) list of DNA extracts earmarked for sequencing (indicating 16S, 18S, metagenome) named “Overview_Sequencing”, ii) list of OSD Analysis Core Group members named “OACG member”

Annex A3: Exit survey of the OSD Analysis Jamboree

Questions 1 to 3 pertain to administrative matters (Q1 “*Name*”, Q2 “*Email*”, Q3 “*May we contact you by email in future to take part in usability testing/ interviews/surveys or other user research activities?”*). Results for questions Q4 to Q11 were the following:

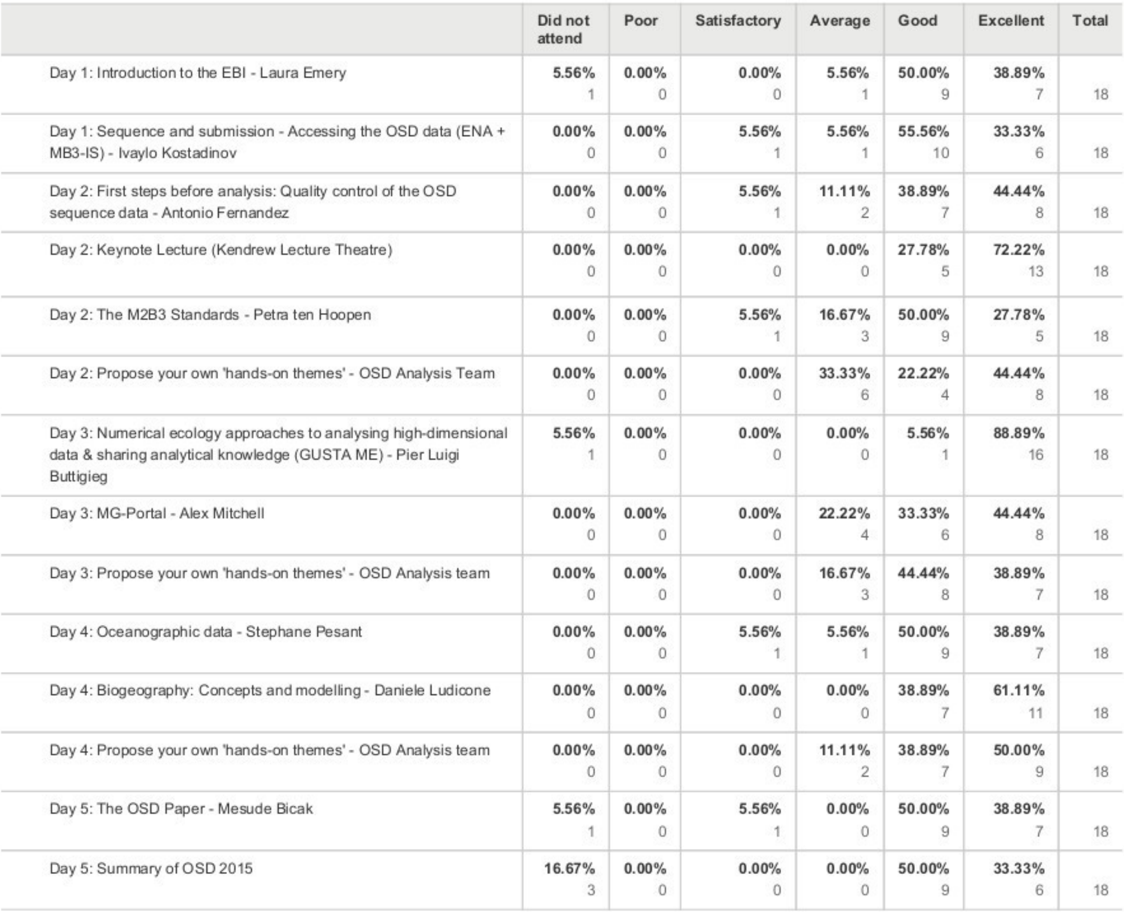


Q4 **Please tell us your overall rating for the entire course.**

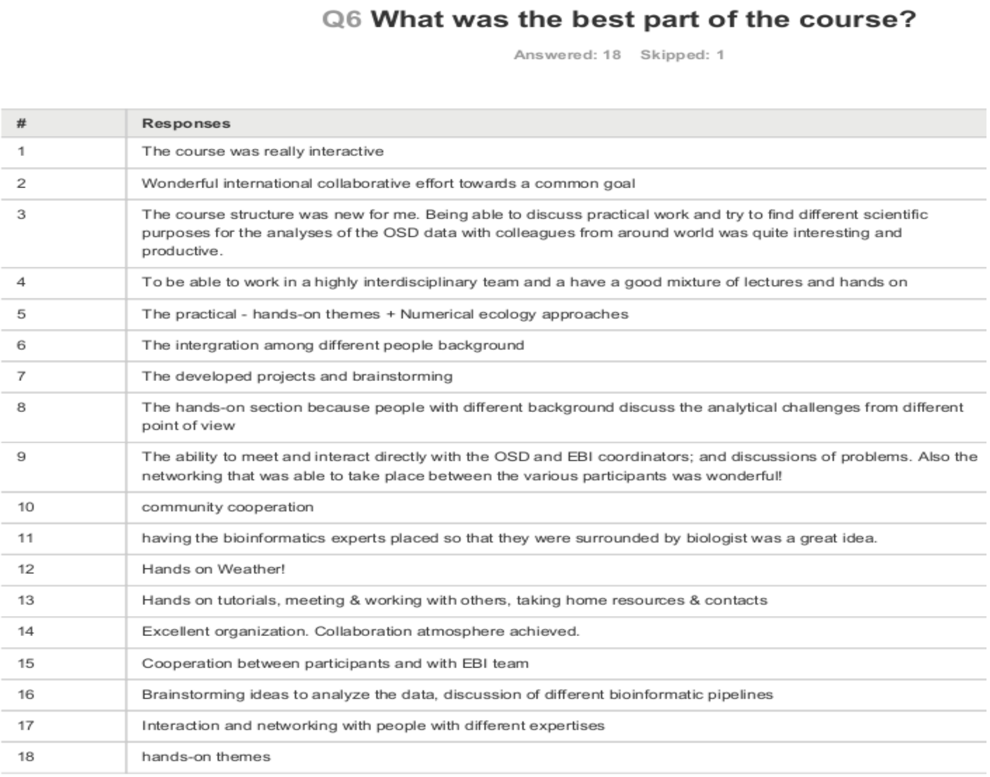
Answered: 18 Skipped: 1

Q5 **Please rate each section of the course.**

Answered: 18 Skipped: 1



	Did not attend	Poor	Satisfactory	Average	Good	Excellent	Total
Day 1: Introduction to the EBI - Laura Emery	5.56% 1	0.00% 0	0.00% 0	5.56% 1	50.00% 9	38.89% 7	18
Day 1: Sequence and submission - Accessing the OSD data (ENA + MB3-IS) - Ivaylo Kostadinov	0.00% 0	0.00% 0	5.56% 1	5.56% 1	55.56% 10	33.33% 6	18
Day 2: First steps before analysis: Quality control of the OSD sequence data - Antonio Fernandez	0.00% 0	0.00% 0	5.56% 1	11.11% 2	38.89% 7	44.44% 8	18
Day 2: Keynote Lecture (Kendrew Lecture Theatre)	0.00% 0	0.00% 0	0.00% 0	0.00% 0	27.78% 5	72.22% 13	18
Day 2: The M2B3 Standards - Petra ten Hoopen	0.00% 0	0.00% 0	5.56% 1	16.67% 3	50.00% 9	27.78% 5	18
Day 2: Propose your own 'hands-on themes' - OSD Analysis Team	0.00% 0	0.00% 0	0.00% 0	33.33% 6	22.22% 4	44.44% 8	18
Day 3: Numerical ecology approaches to analysing high-dimensional data & sharing analytical knowledge (GUSTA ME) - Pier Luigi Buttigieg	5.56% 1	0.00% 0	0.00% 0	0.00% 0	5.56% 1	88.89% 16	18
Day 3: MG-Portal - Alex Mitchell	0.00% 0	0.00% 0	0.00% 0	22.22% 4	33.33% 6	44.44% 8	18
Day 3: Propose your own 'hands-on themes' - OSD Analysis team	0.00% 0	0.00% 0	0.00% 0	16.67% 3	44.44% 8	38.89% 7	18
Day 4: Oceanographic data - Stephane Pesant	0.00% 0	0.00% 0	5.56% 1	5.56% 1	50.00% 9	38.89% 7	18
Day 4: Biogeography: Concepts and modelling - Daniele Ludicone	0.00% 0	0.00% 0	0.00% 0	0.00% 0	38.89% 7	61.11% 11	18
Day 4: Propose your own 'hands-on themes' - OSD Analysis team	0.00% 0	0.00% 0	0.00% 0	11.11% 2	38.89% 7	50.00% 9	18
Day 5: The OSD Paper - Mesude Bicak	5.56% 1	0.00% 0	5.56% 1	0.00% 0	50.00% 9	38.89% 7	18
Day 5: Summary of OSD 2015	16.67% 3	0.00% 0	0.00% 0	0.00% 0	50.00% 9	33.33% 6	18



Q6 What was the best part of the course?

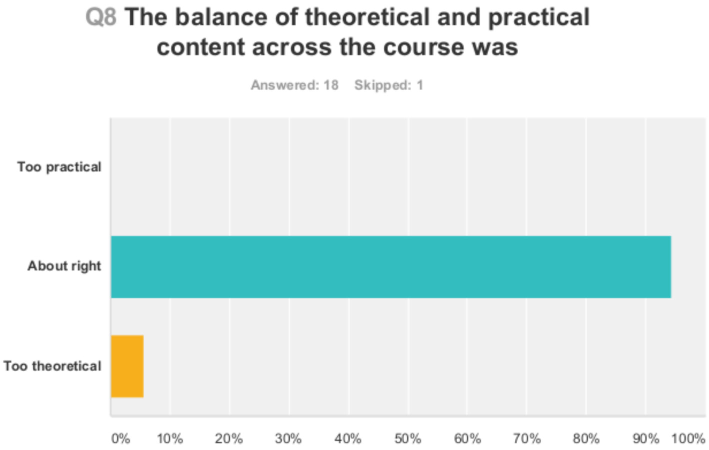
Answered: 18 Skipped: 1

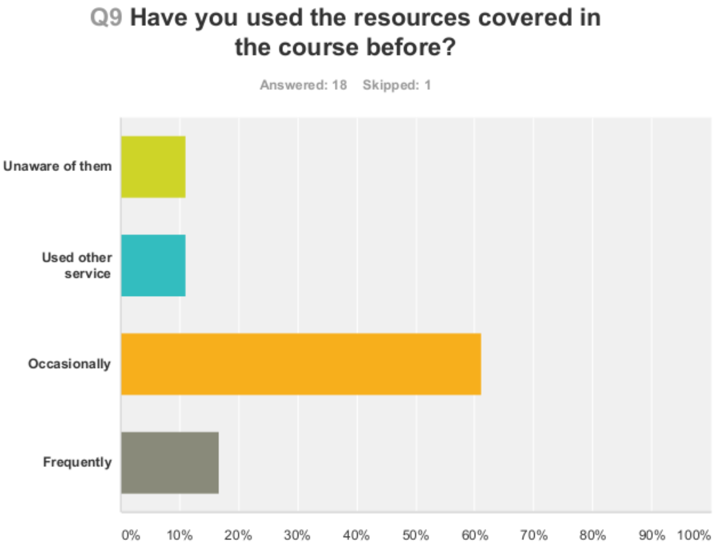
#	Responses
1	The course was really interactive
2	Wonderful international collaborative effort towards a common goal
3	The course structure was new for me. Being able to discuss practical work and try to find different scientific purposes for the analyses of the OSD data with colleagues from around world was quite interesting and productive.
4	To be able to work in a highly interdisciplinary team and a have a good mixture of lectures and hands on
5	The practical - hands-on themes + Numerical ecology approaches
6	The intergration among different people background
7	The developed projects and brainstorming
8	The hands-on section because people with different background discuss the analytical challenges from different point of view
9	The ability to meet and interact directly with the OSD and EBI coordinators; and discussions of problems. Also the networking that was able to take place between the various participants was wonderful!
10	community cooperation
11	having the bioinformatics experts placed so that they were surrounded by biologist was a great idea.
12	Hands on Weather!
13	Hands on tutorials, meeting & working with others, taking home resources & contacts
14	Excellent organization. Collaboration atmosphere achieved.
15	Cooperation between participants and with EBI team
16	Brainstorming ideas to analyze the data, discussion of different bioinformatic pipelines
17	Interaction and networking with people with different expertises
18	hands-on themes

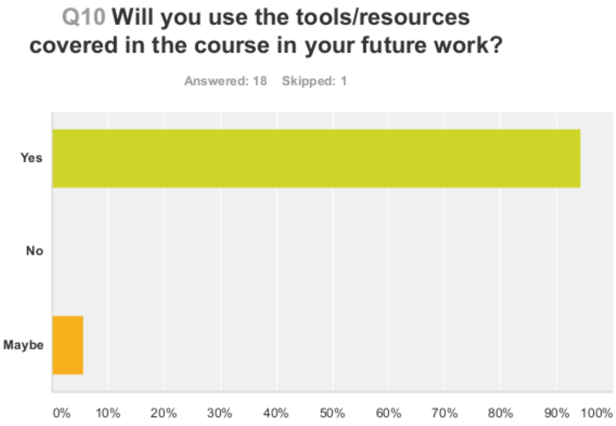
Q7 What was the worst part of the course?

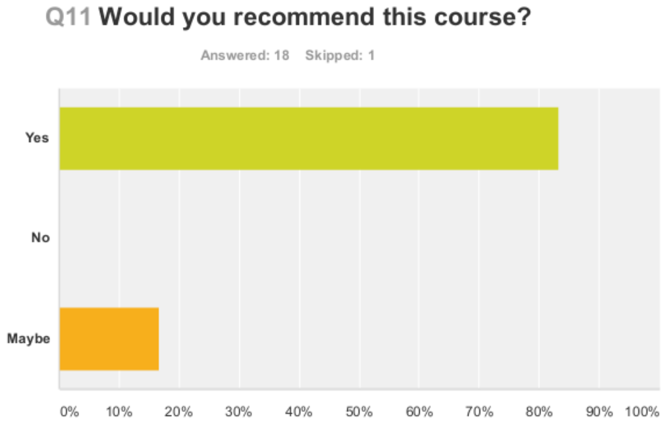
Answered: 18 Skipped: 1

#	Responses
1	The programram was not certain
2	Not being able to drink tea or coffee in the computer room!
3	For me the course didn't have a worst part.
4	the computers were quite slow, almost everybody were using their own computer with the wifi
5	...
6	None
7	None
8	None
9	\initial access to certain data file was limited or unavailable. We had hoped to have more data to work with during the actual course. I realize all the time constraints, but perhaps if we had been made aware of this ahead of time, we could of planned to work on other data sets. Also, computer issues were a constant plague. I experienced numerous freezes and run losses.
10	Needed more time to work with data
11	We very much needed more bandwidth. We were here to work, with real data,and in many ways could not because of lack of adequate resources. And we could not do any work after hours. The hotel internet was awful. EBI should work out something so that training people get better bandwidth
12	Being in off-site hotel with very low Internet bandwidth (night work)
13	The wifi at the hotel!
14	Too long lectures, although I reckon that it was necessary to put everyone in the same page
15	The computer (under BioLinux) was extremely slow - almost unusable
16	None
17	We weren't able to optimize the time for our proposed hands-on, a little bit of dispersion
18	Missing data


Q8 The balance of theoretical and practical content across the course was
Answered: 18 Skipped: 1
Too practical
About right
Too theoretical
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


Q9 Have you used the resources covered in the course before?
Answered: 18 Skipped: 1
Unaware of them
Used other service
Occasionally
Frequently
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


Q10 Will you use the tools/resources covered in the course in your future work?
Answered: 18 Skipped: 1
Yes
No
Maybe
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


Q11 Would you recommend this course?
Answered: 18 Skipped: 1
Yes
No
Maybe
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%