Marine

Microbial Biodiversity,

Bioinformatics & Biotechnology

Deliverable 9.6

Report on Training Pipeline and Courses Held until Mid-Term

Version: final

Circulated to: WP 9 Training & Dissemination Committee 18/12/2013

Approved by: Prof. Frank Oliver Glöckner, 31/12/2013

Expected Submission Date: 31/12/2013

Actual Submission Date: 27/12/2013

Lead Party for Deliverable: EMPA Bremen, Partner # 26

Mail: j.wesnigk@empa-bremen.de **Tel.: +49 421 50 79 20**

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)	

Table of Content

Introduction & Summary

This report contains an overview on the training courses originally planned and adapted to training needs documented in the first months of the project. Then it provides information on the two courses held, including selection and evaluation details, until the mid-term of the Micro B3 project:

1) Micro B3's 1st Theoretical and Practical Training Course: Bioinformatics 1 – Basics and (Meta)Genome Annotation;

2) Micro B3's 2nd Theoretical and Practical Training Course: Multivariate Analysis of Marine Microbial Biodiversity.

The original plan proposed (see next page) was a training pipeline with a total of five courses starting with up to three more theoretical courses on bioinformatics for beginners. This first set is planned to be followed by two hands-on practical courses:

- a multi-disciplinary summer school on standardized sampling for optimally comparable bioinformatic analyses and

- a final course on joint annotation of functional, meta(genomic) and other Omic data focusing especially on data generated during the Ocean Sampling Day (OSD).

A Stakeholders' Training/Information Needs Analysis was done as the first step in order to select topics and formats which are not yet readily available on the market (See D9.22 for final outcome). The training needs part of the survey led to new topics being in more urgent demand than basis bioinformatics training. Therefore, instead of repeats, two different new courses were planned for 2013 and 2014. In total, two courses were held and are reported on in detail in this deliverable: the first bioinformatics course in October 2012 and the statistics & modelling course in June 2013.

Further outcomes of the Training Needs Analysis have been used as basis for the third theoretical and practical course on 'Marine Metagenomics Bioinformatics'. It will be held from 24 to 28 March 2014 at the European Bioinformatics Institute, Hinxton, UK (see www.microb3.eu/events/workshops/training-course-metagenome-analysis).

Outcomes are also used for the summer school, planned for 27 May to 6 June 2014 at HCMR, Crete. It will be multi- and interdisciplinary and cover the whole Micro B3/OSD process from standardised sampling, including environmental parameter and (meta)genomic protocols, to processing, integrating data into databases, and analysis / annotation / interpretation (see www.microb3.eu/events/workshops/micro-b3-summer-school-crete).

Content elements of the final annotation course are also preliminarily planned (see D 9.22). However, they will need to be adapted as technology develops and as decisions for the final OSD research focus areas are still being discussed, i.e. what to sequence in which depth for which purpose (phylogeny, diversity, function). These decisions, to be taken at the General Assembly in April 2014, will be valuable input to fine-tune the OSD annotation course.

Additionally, MG4U project funding could be used for holding the course on 'Basics Bioinformatics for Marine Microbial Genomics' for a second time at MPI Bremen in May 2013 (see www.microb3.eu/events/workshops/mg4u-bioinformatics-training-2013).

Background to Training Course Development

Extract from the Description of Work

Scientific training is planned to be organized into a training pipeline, including one summer school. It will provide bioinformatics expertise to young Micro B3 scientists, and later also to the participants of the Ocean Sampling Day before and after this event. They and the trained new generation will be the seed for disseminating Micro B3 standards and tools to the wider scientific community. Courses or at least modules are planned to provide models for later repetition, and thus will be thoroughly documented. Multipliers will be selected to participate in order to enable this on smaller e.g. regional scales and in other languages where appropriate and useful to generate more long-term sustainability in capacity building (from the DoW, p. 78).

Task 9-4: Training: A dedicated bioinformatics training pipeline consisting of at least three different courses will focus on postgraduate training, developing multiple capacities and bioinformatics skills of PhD students and postdocs. The special feature for the core participants, i.e. young scientists within Micro B3 is that they will be working on their own data during the training, together with experts from the Micro B3 consortium. These project students will have the opportunity to participate in all types of courses building upon each other as well as using internet-based tools like Annotathon in between courses. Training tools such as Annotathon will be adapted to converge with professional distributed annotation tools in WP 5, merging trainees and professionals in the same online community for online self-training and for courses.

One more theoretical and two hands-on course formats are envisaged, which will also be open for external participants based on an open call followed by a selection process on a need-to-know basis. Special emphasis will be put on equal opportunity gender participation and wide geographical distribution with the participation of Mediterranean countries scientists guided by CIESM.

The first course will last five days and bring the Micro B3 students to the same level of bioinformatic understanding, including familiarizing them with Annotathon. Therefore it will focus more on teaching and theory as well as on discussing questions on the background and training needs of the students for the next years. This course will be repeated max. two more times within project year 2 and latest year 3 with a duration of 4 – 5 days. It will be based on experience with training MGE students in bioinformatics, done by RIBOCON, with contributions from EMBL-EBI, MPI and CNRS expertise (D9.6) (from the DoW, pp. 48-49).

The insert above is a reminder of the training pipeline as it was originally planned. As less PhD students are working in the Micro B3 consortium than anticipated and as the Training Needs Analysis provided a substantial group of new topics, instead of repeats of the 'bioinformatics for beginners' format, two new courses were devised by the Micro B3 training team. Having more experienced researchers in the trainee groups also led to a stronger focus in the selection and evaluation on further dissemination of the methods taught. The pipeline concept is still visible through future courses building on the first ones held and through re-inviting the same trainees, esp. from Micro B3 partners and participants in the Ocean Sampling Day (OSD).

Selection and Formal Evaluation of Micro B3 Training Courses

Online questionnaires were designed for each course for selection and for evaluation. The latter were explained to the participants on the last day to be filled in immediately. It can be found in the Annex 1 and 2. The overall evaluation results are on the Micro B3 website under

www.microb3.eu/events/workshops/first-bioinformatics-training/evaluation and

www.microb3.eu/events/workshops/statistics-training-2013/evaluation

Some further free-text questions and some with 'yes,no, maybe' reply options, were added for future use and dissemination purposes and are reported on only below. The idea was to get input for further trainings, motivate and find out about further spreading of the information obtained by the participants within their home labs. An overview of replies is given below separately for each course.

Micro B3's First Theoretical and Practical Training Course: Bioinformatics 1 – Basics and (Meta)Genome Annotation

Background of this Course

From 8 to 12 October 2012, 11 participants of the Micro B3 project came to Bremen, Germany for a training in sequence analysis and (meta)genome annotation. The course was held at the Max Planck Institute (MPI) for Marine Microbiology in Bremen and taught by the Micro B3 partners RIBOCON and CNRS with contributions of expertise from the Microbial Genomics and Bioinformatics Research Group at MPI Bremen. Dr. Jörg Peplies from Ribocon had organised the course with support from the project secretariat and from EMPA during selection and for the evaluation sheet.

After three days of introduction into fundamental sequence analysis ranging from databases to pair-wise and multiple sequence analysis performed by the Microbial Genomics and Bioinformatics Research Group of MPI Bremen, Pascal Hingamp from CNRS UMR 7256 with Aix-Marseille Université introduced the e-learning portal Annotathon with hands-on experience based on real sequences from the Tara Oceans expedition.

Rationale & Target Audience

Within WP 9 (Outreach & Dissemination), the Micro B3 project offers dedicated bioinformatics training for developing multiple capacities and bioinformatics skills of PhD students, postdocs, and senior scientists.

The rationale of the initial five days bioinformatics training was to bring the Micro B3 scientists to the same level of bioinformatics understanding, including the first steps in (meta)genome annotation using the web-based training platform Annotathon. This introductory theoretical and practical training course is supplemented by advanced and hands-on formats later in the project.

Selection of Participants

Participants were selected to cover a range of research fields and institutions in marine sciences. Major criteria were that the knowledge gained in the course is of immediate use for the participant and that the participant is prepared to disseminate the knowledge to his colleagues and/or students in his home institution. Both had to be made transparent in the application. Furthermore equal gender distribution was strived for. Finally four female and seven male participants were selected, all from Micro B3 partner institutions representing mostly biodiversity and biotechnology interests (WP 6 and 7). Also one participant from the IT company Interworks, active in WP 5 took part. Most, i.e. eight trainees already had a PhD degree, two had a Bachelor of Science degree and one a Master's degree: two of these were PhD students.

Table 1: List of participants of Micro B3 Bioinformatics I Training Course

Title	Gender	Partner	Country
BSc	male	INTERWORKS	Former Yugoslav Republic of Macedonia
PhD	male	CNRS/ENS	France
PhD	male	Genoscope/Roscoff	France
PhD	male	HCMR	Greece
PhD	male	MATIS	Iceland
PhD	female	BIOMERIT	Ireland
PhD	female	ICM-CSIC	Spain
PhD	male	CSIC for PharmaMar	Spain
PhD	male	UGRO	The Netherlands
PhD	female	TUBITAK	Turkey
BSc	female	MBA	United Kingdom
Total 11, only 3 (PhD) students	**4 female, 7 male**	**2 industry partners 2 close to WP 7**	

Course Outline and Schedule for Micro B3 Bioinformatics I Training

Bioinformatics 1 was a purely computer-based training course. During the first three days the trainees were taught in the basics of DNA sequence analysis in lectures, demonstrations, and hands-on sessions. They learnt how to deal with the corresponding software tools and how to interpret their results. In the practical parts, the course focused on resources like sequence databases, Linux and the command line as a toolbox for daily bioinformatics work, as well as use of common software tools for sequence similarity searches, alignments, etc..

The aim was to provide basic bioinformatics and computational skills and understanding for the handling and later annotation of biological sequence data generated by today's sequencing technologies. On the last two days, the course focused on aspects of *in silico* annotation of (meta)genomic sequence data, including initial open reading frame prediction, functional predictions, and taxonomic assignment through phylogenetic tree inference. The last two days focused on the e-learning portal Annotathon.

Time	08.10.12	09.10.12	10.10.12	11.10.12	12.10.12
8.30-10.00	Introduction biological data databases	Sequence comparison similarity, identity substitution matrices	Multiple alignments patterns, profiles databases Motif representation and search algorithms	Introduction to the e-learning portal for marine DNA sequence annotation Annotathon 1: ORFs	Annotathon 4: phylogenetic tree inference
10.00-10.30	Break	Break	Break	Break	Break
10.30-12.00	Linux for beginners I	Linux for beginners II	Linux for beginners III	Annotathon 2: functional hypotheses	Annotathon 5: of trees & taxonomy
12.30-13.00	Lunch break	Lunch break	Lunch break	Lunch break	Lunch break
13.00-15.00	Contextual data aquisition, storage MIxS, Ontologies	Dynamic Programming	Multiple alignment tools, MAFFT, Muscle	Annotathon 3: multiple alignments	Annotathon 6: team Micro B3 is primed to annotate!
15.00-15.30	Break	Break	Break	Break	END
15.30-17.00	Important databases Navigation, retrieval	BLAST [n,p,x] Command line BLAST output parsing	Pfam, Prosite, InterPro Command line HMMer	Relax - free time	
17:00-18:30	Icebreaker Dinner at MPI-Bremen			Social Event	
18:30 - 20:00	Special guest lecture by Jack Gilbert from Argonne National Lab			Social Event	

Formal Evaluation of Micro B3 Bioinformatics I Training

An online questionnaire was designed and explained to the participants on the last day to be filled in immediately. The result of questions 1 - 9 are summarized below and can also be found on the Micro B3 portal under

http://www.microb3.eu/events/workshops/first-bioinformatics-training/evaluation

Some further questions were added for internal purposes, esp. to get input for further trainings and find out about dissemination in the home labs by the participants. A summary of replies is given below. Only 10 out of 11 participants filled in the questionnaire, and some did not answer all questions; in these cases the number of replies is indicated.
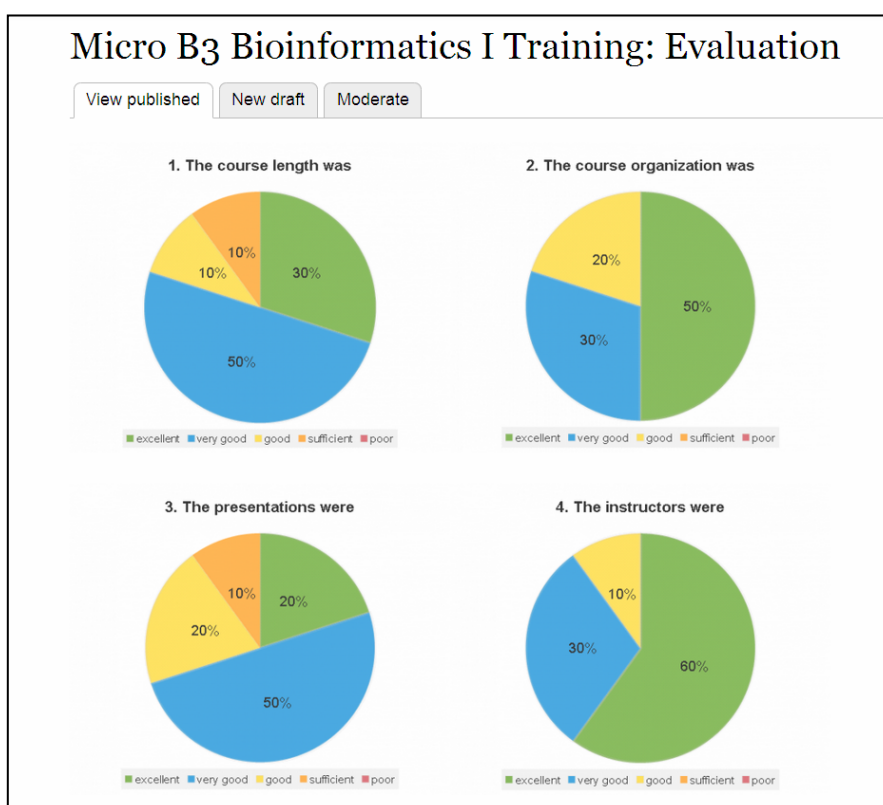
Figure 1: 70% or more trainees answered with excellent or very good; the instructors got the highest marks with 60% rating them excellent.
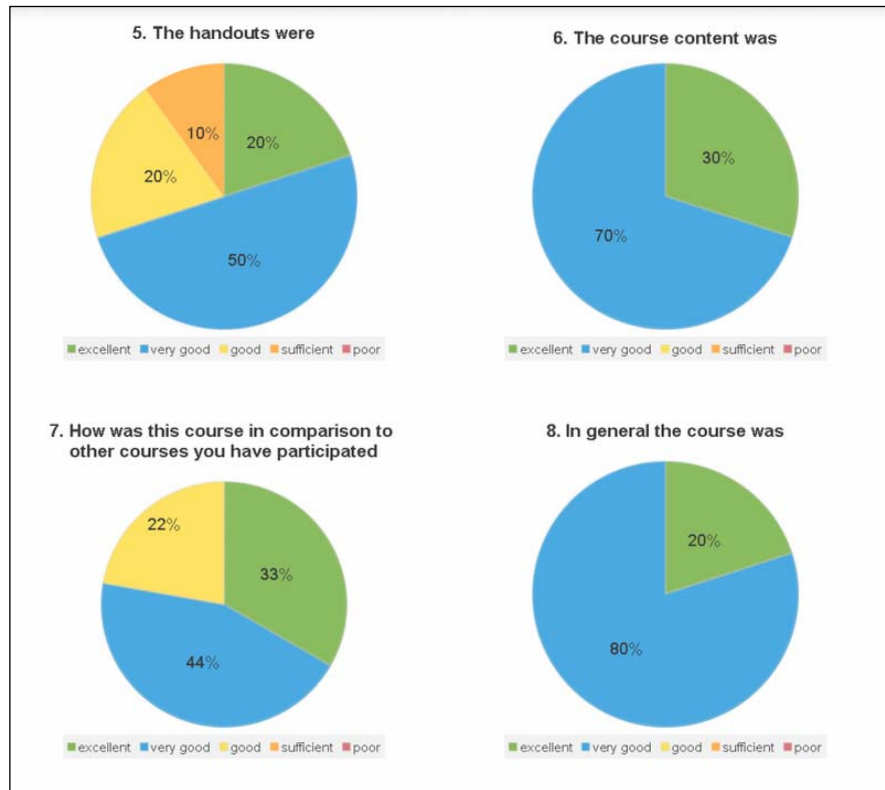
Figure 2: 70% or more answered with excellent or very good, overall 100 % rated the course (8 replies) and its content (6 replies) as very good or even excellent.

When asked for a percentage rating, how well the course addressed the expectations of participants, the replies can be summarized in the following way: for 70 % of participants expectations were met to 90% or more; 20 % were content with 80 %; 10 % with 70% of the course overall.

Future Use and Dissemination of Knowledge of Micro B3 Bioinformatics I Training

The following group of questions was based on text replies and selection of topics for different purposes. The idea was to learn what trainees would do with the knowledge obtained and in which further issues they are interested for their upcoming work.

Trainees were asked to select topics most useful to their work in the next approximately 12 months. Topics were named in the following order (in brackets the sums for 'high' or as specified):

- Web based tools (11)
- Locally installed systems (11)
- Command line (11)
- NGS sequence analysis (10)
- Quality assurance & quality control (8)
- Sequence annotation (8)
- Bioinformatic theories & concepts (7 high, 4 medium)

- Analysing and reporting (meta-)genome sequences (7 high, 2 medium)

Participants were also asked to name modules focusing on useful topics to learn as a next step which could be interesting for future courses. The issues are listed according to how often they were named. Several of these will be covered in upcoming courses.

The top three topics named were:

1. Linux
2. Dynamic programming
3. Sequences & phylogeny

Within these groups issues detailed were for example 'more in depth training in command lines' and 'learning to code basic scripts (re topic 2) or 'training in pipelines for analysing NGS sequencing data', including phylogeny, statistical analysis, sequence-editing.

When asked about further dissemination of the knowledge obtained, the majority of 10 participants stated that they will disseminate some of the course content in their home labs, nine in an informal way and five also formally. Choice of language was seven for English and six preferring their local language. Furthermore, six participants were willing to act as tutors in case a repetition is planned.

The listed issues were named in descending order as potential training elements of high use to colleagues / working groups:

1. Sequence analysis, multiple alignment, annotation, incl. Annotathon
2. Use of BLAST in command line or other online tools, Linux-based
3. Databases and web-based tools
4. Understanding of systems (concepts, different processes, approaches and tools).

Comments – Highlights and Future Ideas

Also a free text box was provided for any comment related to the course. The trainees used it to make many useful comments, of which a selection is summarized below. Points of critique were discussed with the course organizers directly, if they were rather specific, and some are listed below as ideas for future courses.

Positive Points

1. Instructors – valuable, excellent, explanations clear and willing to help out, Database navigation, data retrieval, command line BLAST HMMer and Prosite was great.
2. Linux for beginners was very useful.
3. Interaction with tools - immediate interaction and training for the specific software tools used for analysis in bioinformatics was extremely valuable.
4. Almost all the practical sessions were really helpful.

5.   Annotathon - to learn the pipeline to work with metagenomes has been very useful.

6.   Theoretical part – valuable asset in understanding the high-level concepts and establishing common language and common ground.

Ideas for Future Courses

1.   Database navigation, data retrieval, command line BLAST HMMer and Prosite was very useful. Always working on concrete examples could have made it even better (2 times).

2.   Having more courses on phylogeny especially using ribosomal DNA would be useful.

3.   In future courses it might be better to first give a classical lecture/tutorial after which the students should resolve a 'real' question using Linux.

4.   Dedicating the last day to processing own data with the tools learned. Participants would need to be notified that they should bring some of their own data, but having a hands-on day-long session to start analyzing own data while being able to ask questions would be very useful.

Micro B3's Second Theoretical and Practical Training Course: Multivariate Analysis of Marine Microbial Biodiversity

Background of this Course

The course was held at Jacobs University Bremen, Germany from 17 to 21 June 2013. Dr. Alban Ramette, Dr. Pier Luigi Buttigieg, both from partner AWI, and Dr. Antonio Fernandez-Guerra from MPI were the trainers. Dr. Jörg Peplies from Ribocon had organised the course with support from the project secretariat at Jacobs University and from EMPA during selection and for the evaluation sheet.

Rationale & Target Audience

Within WP 9 (Outreach & Dissemination), the Micro B3 project offered one dedicated statistical training for developing the multivariate analysis skills of PhD students, postdocs, and senior scientists contending with complex ecological datasets. This five-day multivariate statistical training course aimed to provide participants with a functional understanding of the concepts and tools currently being used in the multivariate analysis of ecological datasets. The course contained a brief, high-level review of standard multivariate statistical methods. Practical sessions demonstrated their use based on freely-available software (mainly PAST and R). No advanced statistical or mathematical prior knowledge was required.

The conceptual component was complemented by hands-on analysis of selected data sets typically resulting from ecological surveys (samples-by-species, and samples-by-environment tables). By the end of the course, participants should have acquired enough conceptual and practical skills to independently perform routine multivariate analyses as well as to interpret and evaluate such analyses presented in the published literature. Some sessions were reserved for volunteers to present datasets from their own projects on which to test their newly-acquired analysis skills.

Selection of Participants

In total 40 applications were received, from partners of Micro B3, other European marine research institutes, but also from the United States and South Africa.

The application questionnaire had asked:

- 12 questions related to the general experience of applicants with general statistical concepts ranging from data organisation to scripting languages;

- 7 questions related to familiarity with hypothesis testing and data resampling schemes covering topics ranging from Bootstrapping to ANOSIM;

- another 7 questions related to knowledge and application of advanced analysis tools from dimensionality reduction (e.g. Principal Component Analysis) to Time-Series Analysis.

Based on an analysis of the data provided by each applicant concerning their statistical backgrounds, the course leader used multivariate statistical tool (Principal Component

Analysis) to reduce the dimensionality of the table of parameters. By this mean it was attempted to determine the most homogenous group possible and to avoid mixing total beginners with researchers experienced in all or most of the proposed tools and concepts. An optimum 'middle ground' from the levels mentioned in the applications was established and used to better tailor the course content to the participants' profiles. Thus the analysis helped to characterise the different areas of expertise of the whole group.

The following criteria were then applied to the potential trainee group selected this way:

- focus on Micro B3 members and OSD participants, (only one per partner)
- equal gender representation
- need to know (course content of immediate use)
- willingness to spread knowledge.

Finally 21 were selected (for details see Table 2).

Table 2. List of Participants of Micro B3 Statistics Training Course

Title	Gender	Partner/Home Inst.	Country
PhD student	Female	*University of Vienna*	Austria
PhD	Male	*Ruđer Bošković Institute*	Croatia
PhD student	Male	CNRS	France
PhD student	Female	ENS/CNRS	France
Prof./Lect.	Male	HCMR	Greece
PhD student	Female	AWI	Germany
Phd student	Female	AWI	Germany
BSc	Male	UniHB	Germany
Phd student	Female	UniHB	Germany
PhD	Male	MATIS	Iceland
PhD	Male	IAMC	Italy
PhD	Male	*University of Bergen*	Norway
Prof./Lect.	Male	*Sultan Qaboos University*	Oman
PhD	Female	*CIIMAR*	Portugal
PhD student	Female	*University of Cape Town*	South Africa
PhD	Male	ICM-CSIC	Spain

Phd student	Male	AZTI-Tecnalia	Spain
PhD student	Female	*University of Amsterdam*	The Netherlands
PhD student	Male	UGRO	The Netherlands
Prof./Lect.	Female	Istanbul University	Turkey
PhD	Female	MBA	United Kingdom
Total 21 participants	**10 female, 11 male**	*Italics: no Micro B3 partner*	

Note: from the local partners AWI and University Bremen (Uni HB) two participants each were accepted in accordance with the course leader and because no additional costs arose.

Course Outline and Schedule for Micro B3 Statistics Training

Time	17.06.13	18.06.13	19.06.13	20.06.13	21.06.13
9:00 -10:30	Data exploration, missing data, outlier detection, normal distribution, data transformation	Simple ordination, NMDS, ANOSIM, Eigenanalysis	Contextual interpretation: dataset, variable preparation, indirect gradient analysis	Direct gradient analysis, redundancy analysis, canonical correspondance analysis	Integrating spatial data into ecological modeling, recommendations
10:30 – 11:00	Break	Break	Break	Break	Break
11:00-12:00	Linear regression, variance-covariance, cluster analysis, dissimilants coefficients	Principal Component Analysis, Principal Coordinate Analysis, Correspondance Analysis	Mantel test, Monte Carlo permutation, Mantel correlogram	Variable selection, variation partitioning	Examples and case-studies
12:00 – 13:30	Lunch break	Lunch break	Lunch break	Lunch break	Lunch break
13:30 – 15:30	Practicals in PAST	Practicals in PAST	Practicals in PAST	Practicals in PAST	Examples and case-studies
15:30 – 16:00	Break	Break	Break	Break	END
16:00 - 18:00	Practicals in R	Practicals in R	Practicals in R	Practicals in R	
18:00 - 20:00			Participants data	Social Event	
				Social Event	

In the first two days, the basic statistical and mathematical concepts needed to understand multivariate statistics were presented. Special emphasis was given to data visualization, outlier detection and variable distribution transformation. Further, similarity coefficient calculation was introduced alongside clustering and ordination approaches which allow representation of multivariate structures in lower-dimensional, and hence more easily interpretable, spaces. Days 3 & 4 focused on the ecological interpretation of community changes between samples when metadata is available.

Techniques allowing explanatory parameters within sample metadata were introduced, e.g., physico-chemical parameters, spatial coordinates, time, to complement analysis of sample-by-species response tables. During the final day, the instructors provided recommendations concerning the combination of analytical approaches into a coherent statistical framework.

Formal Evaluation of Micro B3 Statistics Training

For the second training course, an online questionnaire was designed (see Annex 2) and the participants were asked to fill it in during the final day of the course. The results are summarized below and the statistical part can also be found on the website under www.microb3.eu/events/workshops/statistics-training-2013/evaluation.

18 replies were obtained to all questions from 21 participants overall.
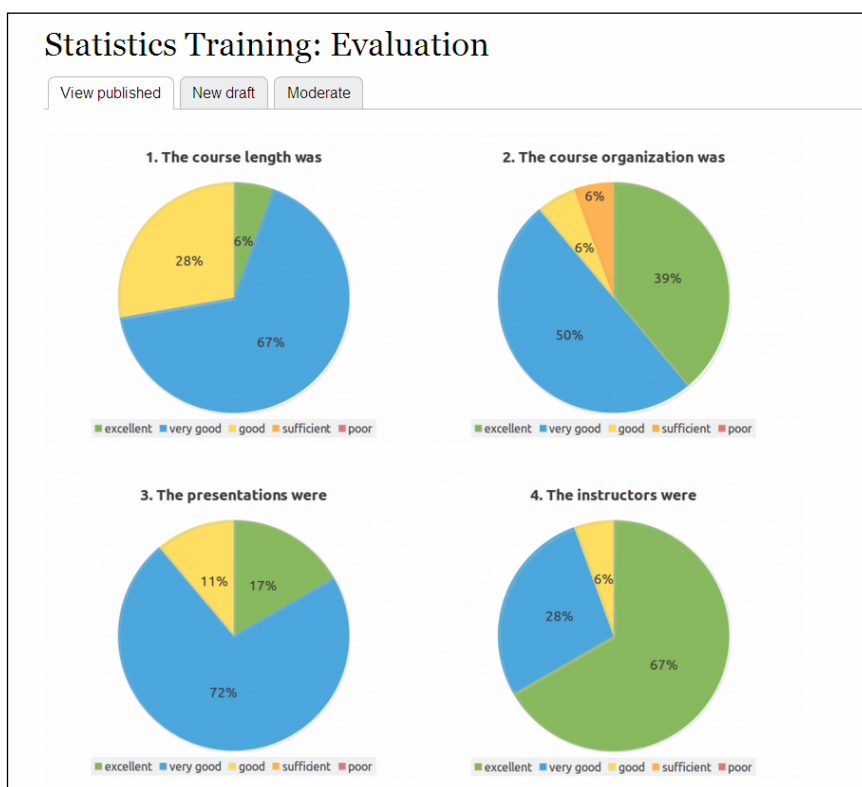
Figure 3: 73% or more trainees answered with very good or excellent; the instructors got the highest marks with 67 % rating them excellent.

When asked for a percentage rating, how well the course addressed the expectations of participants The replies can be summarized in the following way: for 70 % of participants they were met to 80% or more; 10 % each were content with 70, 60 and 50 % of the course overall.
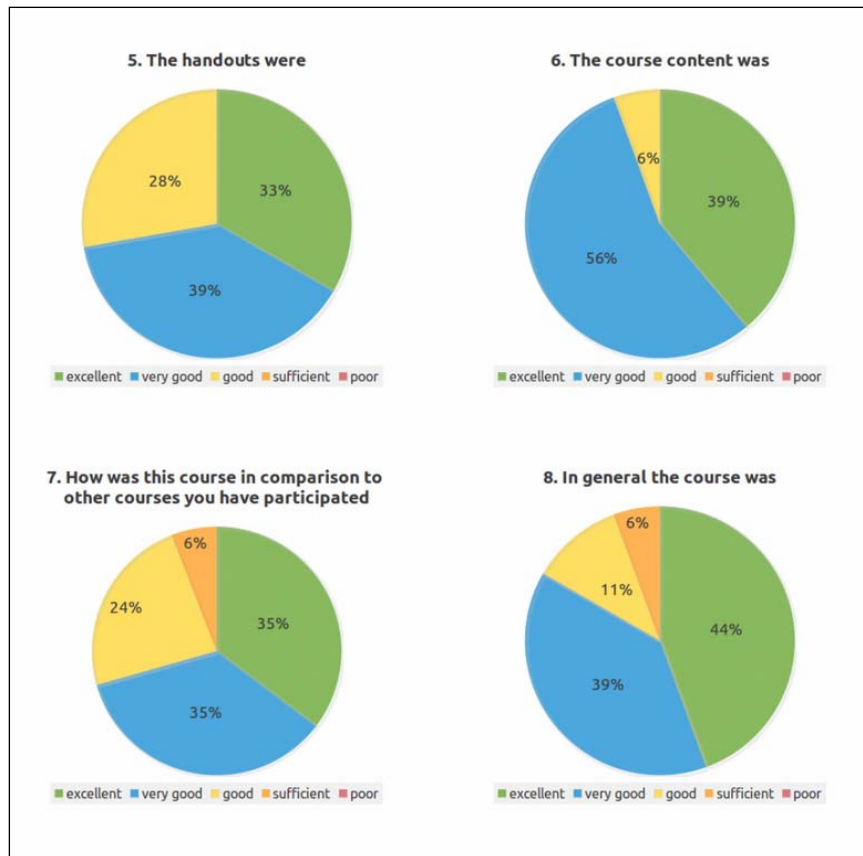
Figure 4: 70% or more answered with very good or excellent, especially for handouts and content; 6 % found this course only sufficient in general and as compared to other courses.

Future Use and Dissemination of Knowledge of the Micro B3 Statistics Training

The following answers are based on selection of topics and text descriptions of different future uses. The idea was to learn what trainees would do with the knowledge obtained and also which further issues they are interested in for their upcoming work. Main outcomes are highlighted below.

Figure 5: Question on usefulness of statistical concepts, ecological interpretation, and combination into one coherent statistical framework for ecological modelling overall, and as specified into sub-topics. For answers see text below.

The following answers were given per group of questions; in brackets also the range per sub-topic is indicated (see Annex 2 for full questionnaire).

- Basic statistical and mathematical concepts incl. visualization, for representation of multivariate structures: 17 yes (range 16-17 yes in sub-topics);

- Ecological (contextual) interpretation of community changes based on metadata: 13 yes (range 8-12 in sub-topics);

- Combination of analytical approaches into coherent statistical framework: 8 yes, 8 maybe, 2 no (range 6-7 yes, 8-9 maybe, 3 no in sub-topics).

Answers show that the first group covering basic concepts was of most immediate use to the trainees, then the second group on ecological biterpretation, where sometimes 'maybe' was selected. In the last group 'maybe' was mentioned as often or more than 'yes', showing that the combinatorial approaches are not (yet) of interest to all trainees.

18

The last set of questions was intended to investigate means and topics of interest to trainees for further spreading once they returned home. There was a very high willingness to share and promote use of the course content, so that it can almost be seen as a train-the-trainer seminar.

The majority of 18 participants stated that they will disseminate some of the course content in their home labs, most in an informal way and six also or only formally. 13 will spread only certain elements and six stated that they will disseminate all of the content. Choice of language was 15 for English and 12 preferring their local language.

As elements from the training of high use to colleagues / working group, a lot of items were listed by the trainees covering most course elements. These will not be repeated here. Several stated that all of the course content was relevant and applicable to work being performed in their home institutions.

Interesting are the planned forms of dissemination, for which a few quotes from the replies are presented below:

> 'I am planning on writing a couple of blog posts on the course. This will be a summary on the content. I will try also to distribute the scripts and notes to colleagues who are interested in the content. Also, I might organise discussion sessions with interested colleagues.'
>
> 'I am the point of reference so I will be able to feed through much of what I have learned to the projects which require multivariate statistical approaches. The guide images and lectures will be very helpful for me to explain the concepts to my colleagues.'
>
> 'It is my intention to do a short presentation showing the overall topics of this course, as well as to communicate the possibility of having a similar course there. Many of my colleagues might be interested in this type of analysis, and I would really like to share what I've learned with them. Unfortunately as a PhD student I don't feel confident enough to teach it myself, but I think it is possible for me to give a brief overview to my colleagues and to make them enthusiastic about it.'

Comments – Highlights and Future Ideas

Finally many useful comments were obtained by all course participants, showing how strongly they engaged with the topic and the lecturers, who got a lot of positive feedback, also directly.

One recurring issues seemed to be the balance between the two programmes/tools PAST and R, where the group seemed to have been split between finding especially the R session not long enough, or too early for optimum understanding. This may reflect different levels of exposure of the participants to different software solutions to address statistical analyses. The presentation of the IEAG online guide, which is still under development within Micro B3, was seen as very positive.

Ideas for Future Courses

Suggestions for improvement covered:

- More time for presentation and exchange between trainees (poster session or booklet on their areas of work).

- Summaries of key features on each subject (assumptions and rules) at the end of each session.

- Individual practical exercise so trainees would have to produce an output and see what they actually understood (probably asking for an extension of course length).

- The wish to work on own data or at least one case study, selected by the trainers or from one participant, to combine several of the elements learnt and follow all steps on this.

As a final note, some further quotes from the overall comments are provided:

> 'I felt like I received a lot of useful knowledge and contacts with other researchers facing similar analysis challenges. Super helpful!'
>
> '…. because most of the time more than one instructor was in the room, questions could be (and were) answered from different points of view, or in different words (and schemes), which aided understanding.'

Annexes

Evaluation questionnaires for the first two Micro B3 training courses

Annex 1.

Evaluation Form of Micro B3 Bioinformatics I Training

Micro B3 Evaluation sheet

Micro B3's 1st Training Course in Bioinformatics, October 8 – 12, 2012, Bremen

Evaluation sheet *

	excellent	very good	good	sufficient	poor	no answer
1. The course length was	○	○	○	○	○	○
2. The course organization was	○	○	○	○	○	○
3. The presentations were	○	○	○	○	○	○
4. The instructors were	○	○	○	○	○	○
5. The handouts were	○	○	○	○	○	○
6. The course content was	○	○	○	○	○	○
7. How does this course compare to other courses you participated in	○	○	○	○	○	○
8. Overall the course was	○	○	○	○	○	○

My expectation for this course was met (0-100%) *

- Select – ▾ %

Comments

Please state what you found valuable in the course, aspects which could be improved, or other ideas/comments you would like to share with the organisers and/or instructors.

Comments *

Please select those topics useful to your work in the next approximately 12 months *

	yes	maybe	no
Bioinformatic theories & concepts	○	○	○
NGS sequence analysis	○	○	○
Sequence annotation	○	○	○
Using compute farms	○	○	○
Web based tools	○	○	○
Locally installed systems	○	○	○
Command line	○	○	○
Relational databases	○	○	○
Workflow tools	○	○	○
Quality assurance & quality control	○	○	○
(Meta)genome mining for specific functions	○	○	○
Bioinformatics for metabolic analysis	○	○	○
Analyzing and reporting (meta) genome sequences	○	○	○
Statistical analyses for biodiversity	○	○	○
Genomic based modeling	○	○	○

Future modules

Which training modules could be of interest for future courses?

Please name modules focussing on useful topics for you to learn as a next step (max. 5 topics):

When you get back home, how will you disseminate the knowledge from this course? *

	yes	no
formally	○	○
informally	○	○
all of the content	○	○
only certain elements (indicate which ones in the box below)	○	○
in English	○	○
in your local language	○	○
not at all	○	○

Most useful elements

Please name elements from this week's training of high use to your colleagues / your working group:

If we repeat this course, would you be willing to act as a tutor?

○ yes

○ no

If "yes" please send us your name by e-mail to mb3pm@microb3.eu.

Submit

Annex 2.

Evaluation Form of Evaluation of Micro B3 Statistic Training

Micro B3 Statistics Training: Evaluation sheet

Micro B3 Statistics Training 2013, June 17– 21, Bremen

Evaluation sheet *

	excellent	very good	good	sufficient	poor	no answer
1. The course length was	○	○	○	○	○	○
2. The course organization was	○	○	○	○	○	○
3. The presentations were	○	○	○	○	○	○
4. The instructors were	○	○	○	○	○	○
5. The handouts were	○	○	○	○	○	○
6. The course content was	○	○	○	○	○	○
7. The practical sessions were	○	○	○	○	○	○
8. How does this course compare to other courses you participated in	○	○	○	○	○	○
9. Overall the course was	○	○	○	○	○	○

My expectation for this course was met (0-100%) *

[– Select – ▼] %

Comments

Please state what you found valuable in the course, aspects which could be improved, or other ideas/comments you would like to share with the organisers and/or instructors.

Comments *

[   ]

Please select those topics useful to your work in the next approximately 12 months *

	yes	maybe	no
Basic statistical and mathematical concepts incl. visualization, for representation of multivariate structures:	⊙	⊙	⊙
- data exploration and transformation, missing data, outlier detection, normal distribution,	⊙	⊙	⊙
- linear regression, variance-covariance, cluster analysis, dissimilants coefficients	⊙	⊙	⊙
- Simple ordination, NMDS, ANOSIM, Eigenanalysis	⊙	⊙	⊙
- Principal Component, Coordinate and Correspondance Analysis	⊙	⊙	⊙
Ecological (contextual) interpretation of community changes based on metadata:	⊙	⊙	⊙
- dataset, variable preparation, indirect gradient analysis	⊙	⊙	⊙
- Mantel test and correlogram, Monte Carlo permutation	⊙	⊙	⊙
- direct gradient, redundancy and canonical correspondance analysis	⊙	⊙	⊙
- variable selection and variation partitioning	⊙	⊙	⊙
Combination of analytical approaches into coherent statistical framework:	⊙	⊙	⊙
- integrating spatial data into ecological modeling	⊙	⊙	⊙
- examples and case-studies	⊙	⊙	⊙

Future modules

Which training modules could be of interest for future courses?

Please name modules focussing on useful topics for you to learn as a next step (max. 5 topics):

When you get back home, how will you disseminate the knowledge from this course? *

	yes	no
formally	○	○
informally	○	○
all of the content	○	○
only certain elements (indicate which ones in the box below)	○	○
in English	○	○
in your local language	○	○
not at all	○	○

Most useful elements

Please name elements from this week's training of high use to your colleagues / your working group:

Submit