



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Introducing evolutionary biologists to the analysis of big data: guidelines to organize extended bioinformatics training courses

**Citation for published version:**

Faria, R, Triant, D, Perdomo-sabogal, A, Overduin, B, Bleidorn, C, Santana, CIB, Langenberger, D, Dall'olio, GM, Indrischek, H, Aerts, J, Engelhardt, J, Engelken, J, Liebal, K, Fasold, M, Robb, S, Grath, S, Kolora, SRR, Carvalho, T, Salzburger, W, Jovanovic, V & Nowick, K 2018, 'Introducing evolutionary biologists to the analysis of big data: guidelines to organize extended bioinformatics training courses' *Evolution: Education and Outreach*, vol. 11, no. 8. DOI: 10.1186/s12052-018-0080-z

**Digital Object Identifier (DOI):**

[10.1186/s12052-018-0080-z](https://doi.org/10.1186/s12052-018-0080-z)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Evolution: Education and Outreach*

**Publisher Rights Statement:**

© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.




CURRICULUM AND EDUCATION

Open Access



# Introducing evolutionary biologists to the analysis of big data: guidelines to organize extended bioinformatics training courses

Rui Faria<sup>1,2,3\*</sup>, Deborah Triant<sup>4</sup>, Alvaro Perdomo-Sabogal<sup>5,6,7</sup>, Bert Overduin<sup>8</sup>, Christoph Bleidorn<sup>9,10\*</sup> , Clara Isabel Bermudez Santana<sup>11</sup>, David Langenberger<sup>12</sup>, Giovanni Marco Dall'Olio<sup>13</sup>, Henrike Indrischek<sup>14,15</sup>, Jan Aerts<sup>16,17</sup>, Jan Engelhardt<sup>15</sup>, Johannes Engelken<sup>2^A</sup>, Katja Liebal<sup>18</sup>, Mario Fasold<sup>12</sup>, Sofia Robb<sup>19</sup>, Sonja Grath<sup>20</sup>, Sree Rohit Raj Kolora<sup>10,15</sup>, Tiago Carvalho<sup>2</sup>, Walter Salzburger<sup>21</sup>, Vladimir Jovanovic<sup>7,22</sup> and Katja Nowick<sup>5,6,7\*</sup>

## Abstract

Research in evolutionary biology has been progressively influenced by big data such as massive genome and transcriptome sequencing data, scalar measurements of several phenotypes on tens to thousands of individuals, as well as from collecting worldwide environmental data at an increasingly detailed scale. The handling and analysis of such data require computational skills that usually exceed the abilities of most traditionally trained evolutionary biologists. Here we discuss the advantages, challenges and considerations for organizing and running bioinformatics training courses of 2–3 weeks in length to introduce evolutionary biologists to the computational analysis of big data. Extended courses have the advantage of offering trainees the opportunity to learn a more comprehensive set of complementary topics and skills and allowing for more time to practice newly acquired competences. Many organizational aspects are common to any course, as the need to define precise learning objectives and the selection of appropriate and highly motivated instructors and trainees, among others. However, other features assume particular importance in extended bioinformatics training courses. To successfully implement a learning-by-doing philosophy, sufficient and enthusiastic teaching assistants (TAs) are necessary to offer prompt help to trainees. Further, a good balance between theoretical background and practice time needs to be provided and assured that the schedule includes enough flexibility for extra review sessions or further discussions if desired. A final project enables trainees to apply their newly learned skills to real data or case studies of their interest. To promote a friendly atmosphere throughout the course and to build a close-knit community after the course, allow time for some scientific discussions and social activities. In addition, to not exhaust trainees and TAs, some leisure time needs to be organized. Finally, all

\*Correspondence: r.macieiradefaria@sheffield.ac.uk; christoph.bleidorn@biologie.uni-goettingen.de; katja.nowick@fu-berlin.de

<sup>A</sup>Johannes Engelken—Deceased

<sup>1</sup> CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

<sup>7</sup> Faculty for Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Königin-Luise-Straße 1-3, 14195 Berlin, Germany

<sup>10</sup> German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

Full list of author information is available at the end of the article



organization should be done while keeping the budget within fair limits. In order to create a sustainable course that constantly improves and adapts to the trainees' needs, gathering short- and long-term feedback after the end of the course is important. Based on our experience we have collected a set of recommendations to effectively organize and run extended bioinformatics training courses for evolutionary biologists, which we here want to share with the community. They offer a complementary way for the practical teaching of modern evolutionary biology and reaching out to the biological community.

**Keywords:** Active learning, Bioinformatics, Evolutionary biology, Extended course, Genomics, High-throughput-sequencing, Programming, Sustainable course

## Background

The principles of evolution not only have gained a central importance in biology but also in other disciplines such as medicine, psychology, agriculture, economy, computation, among others (Meagher 1999; Losos et al. 2013; Carroll et al. 2014). The steady growth of high-throughput research in evolutionary biology during the last decade has made it necessary for life scientists, and in particular evolutionary biologists, to become acquainted with programming and bioinformatics skills to analyze the large and seemingly unmanageable datasets (Marx 2013; Via et al. 2011, 2013; Kwok 2013; Greene et al. 2016; Schneider et al. 2010). Although we have been observing an increased offer in undergraduate and master courses in bioinformatics, there is still a huge demand for specialized training in bioinformatics and computational biology among evolutionary biologists (or life scientists in general) (Schneider et al. 2012; Attwood et al. 2015). Thus, students and researchers (including group-leaders) on these scientific disciplines often lack the necessary skills to make the most from the available data to answer their questions (Brazas et al. 2017).

An increasing number of focused professional development workshops and courses are available to respond to this huge demand for specialized training in bioinformatics and computational biology (Schneider et al. 2012; Attwood et al. 2015; Brazas et al. 2017). Moreover efforts have been joined by the international research community [e.g. the Bioinformatics Training Network (Schneider et al. 2012)] to create networks to improve education in bioinformatics (such as GOBLET—Global Organization of Bioinformatics Learning, Education and Training—<http://www.mygoblet.org>), where course organizers, trainers and potential trainees can have access to information about available courses, teaching materials as well as trainers in different topics, with the goal of organizing and improving the quality of the state of the art training in bioinformatics (Attwood et al. 2015; Corpas et al. 2015; Consortium GOBLET 2013).

Despite their huge utility, the majority of the available courses is usually short (1–5 days) (Schneider et al. 2012), and often focuses on a single topic (e.g. NGS, Next

generation sequencing) or programming language (e.g. Python). Online courses are another reasonable alternative, but these require substantial self-discipline and suffer from restricted interaction, with limited real-time support from instructors (Brazas et al. 2017). Complementary to short and online courses, extended bioinformatics courses (EBCs—e.g. 2–3 weeks duration) offer an integrative training format suitable for evolutionary biologists embracing such a new and broad field.

All authors of this contribution have been organizers, instructors or teaching assistants of at least one EBC. Our article was mainly inspired by our experience with the organization of the EBC on “Programming for Evolutionary Biology” at the University of Leipzig (<http://evop.bioinf.uni-leipzig.de/>), but also from experience with other courses, such as “Programming for Biology” at Cold Spring Harbor Laboratory (<https://meetings.cshl.edu/courses.aspx?course=C-INFO&year=16>), and “Workshop on Molecular Evolution” and “Workshop on Population and Speciation Genomics” at Cesky Krumlov (<http://evomics.org>). “Programming for Evolutionary Biology” has taken place for the 8th time this year, while “Programming for Biology” started in 1999 and the Cesky Krumlov workshops go back to the Molecular Evolution workshop founded in 1988 in the Marine Biology Laboratory at Woods Hole. Similar to the other workshops, “Programming for Evolutionary Biology” is an annual event organized for about 25 trainees and targeted towards PhD students and postdocs in evolutionary biology or biology in general. All mentioned workshops are 10–18 days long and cover a wide range of topics, including working in a Unix environment, programming/scripting languages, and the application of tools and programming skills to the analysis of genome-, population- and taxa-scale data to answer evolutionary biology research questions.

Based on our experience, a major opportunity of EBCs is to expose trainees to a wide variety of topics while spending enough time on the basic theoretical background of each topic to better guide trainees through “hands on”. Additionally, a comprehensive and complementary set of scripting languages can

be taught that trainees can later apply to their own research. EBCs also allow more time for exercises, discussions, data analyses, working with the trainees' own data in line with a "learning-by-doing" philosophy. The course can be complemented by scientific talks demonstrating the potential of the tools/skills covered in the course in real case studies. Furthermore, there is opportunity for developing a 3–4 day final project, where trainees integrate and put all skills learned in the previous days into practice, by applying them to their own projects and ideas. This usually helps motivate trainees to practice what they have learned when they return home. Finally, EBCs also build strong camaraderie among trainees, instructors and teaching assistants (TAs) to an extent not achieved in short courses, often creating a long-term active community. All in all, well-planned extended courses with interconnected modules can facilitate training in less time compared to several short courses covering the same topics.

However, the organization of EBCs can be challenging. First, much coordination is needed to organize topics, programming languages, instructors, and experienced TAs who can be away from their labs for such long periods. Second, running such courses demands a substantial infrastructure that must be 100% dedicated to the course during the entire time. Third, since costs tend to be high, trainees are often disheartened to attend (especially if they come from countries with limited research funding). Finally, the long duration of a course that covers a large variety of topics and/or programming languages can be overwhelming.

When we tell colleagues about intensive courses that run for full days during 2–3 weeks including weekends, their first reaction is often disbelief. Nevertheless, if well planned, EBCs can be very effective and rewarding. While general guidelines for the organization of bioinformatics training courses have been proposed, these tend to concentrate on short formats (Via et al. 2011), thereby neglecting some idiosyncrasies of longer alternatives. In order to fill this gap, here we propose a set of guidelines for making an EBC in evolutionary biology successful based upon our experience as trainees, TAs, instructors and organizers. Although we focus on organizational aspects specific to the EBC format, for the sake of completeness we also reiterate some guidelines proposed by others that are key to any course format (e.g. Via et al. 2011, 2013). Here we use examples based on our own experience in organizing a course oriented to evolutionary genomics, but the same principles can be applied to related subdisciplines within the broad field of evolutionary biology.

### **Define the learning objectives and design an integrative curriculum**

With an EBC there is opportunity to create an integrative training plan that offers a more varied array of current topics in evolutionary biology (e.g. high-throughput sequencing data analysis, phylogenomics, statistics, visualization, etc.) and programming languages (e.g. Perl, Python, R, Bash, JAVA, etc.). As with any other course format, start organizing an EBC by well-defining the general objectives and training goals (Via et al. 2011; Goldsmid 1981). These could be for instance: *acquiring essential computing skills for evolutionary biology research*, or *understanding the statistical foundations of genomic data analysis*. Then contemplate which subdisciplines (e.g. population genomics), theoretical concepts (e.g. adaptation) and skills (e.g. de-novo assembly) the trainees should learn to reach the general goals. For example, they will have to understand some of the main evolutionary concepts, become familiar with a Unix environment, learn programming languages, develop the necessary statistics background, and understand how to use tools to analyze data produced by next generation sequencing platforms. Other important aspects are mechanisms to ensure reproducibility of research, including proper documentation of source code and how to make custom scripts publicly available.

Each of these concepts and skills could be taught in separate modules. It is extremely important to organize these modules in a logical order and attain a high level of complementarity and cohesion. You can achieve this consistency by re-using datasets or by employing the same programming languages in multiple modules. For example, a first module on "introduction to Unix" will provide the skills for many subsequent modules, since most bioinformatics software and analyses are done in a Unix environment. A module on NGS data analysis could follow in which the trainees learn how to map reads and perform variant calling using the Unix command line. The trainees will later use these variants in a population genomics module where they address real evolutionary biology problems such as determining which genomic regions are involved in adaptation using a Python script that they learned and implemented in a previous programming module. Finally, this set of modules can be complemented by a scientific talk, demonstrating the potential and practical applications of the tools they just learned.

Each module should also be defined by its own set of *specific objectives* that build upon those from earlier modules. The organizers need to communicate the general objectives of the course to each instructor so that they can design their modules by focusing on how it will contribute to the general objectives of the course. While

the specific objectives are unique to each module, they need to be synchronized by the organizers to ensure their interconnectivity and that they contribute to the general goals of the course. Although this is crucial to any course, it assumes a particular importance in EBCs, which usually have a large number of modules. To implement this strategy, it is necessary to reinforce the interaction between the instructors and TAs so that all are aware of each other's abilities or area of expertise. Here the organizers have a crucial role in establishing the first contact and mediating communications during the course planning period.

Once you have defined the general and specific aims of your course, you need to communicate them in the advertisement for your course (see below), which helps to address the appropriate target audience (Via et al. 2013). Create an informative website about your course content and organization, to which the participants refer to before and during the course. After collecting information about the trainees' needs from their application letters, and also taking into account the trainees' feedback from previous years (see below), you can adjust or fine-tune your course modules to meet those needs (Via et al. 2011, 2013).

Finally, as emphasized by Via et al. (2011), it is important to communicate the course objectives at the beginning of the course, in form of an overview, and again during the course in the form of reminders/updates tailored to the specific modules. Knowing and prioritizing learning goals are necessary for trainees so that they can focus and better understand all of the course material (Handelsman et al. 2007).

### **Appoint an interactive and complementary team of instructors and TAs**

General rules for selecting instructors have been proposed by Via et al. (2013). Here we want to reinforce that rather than choosing only known "authorities" in the field as instructors, you should also consider their didactic skills, creativity, originality, social competencies including their joyfulness in teaching and working with other people, as well as gender-balance issues. More than targeting a specific profile for all instructors, the trainees can benefit from the diversity of the instructors' skills and philosophies. Finding a motivated instructor and TA for an EBC without seeing her/him in action may be difficult. Therefore, the best way to recruit them is to ask for recommendations from colleagues and other course organizers (Via et al. 2013), or search in platforms such as GOBLET. Instructors should be aware that such courses are meant to be highly interactive and with a strong hands-on component. Trainees are more engaged and motivated when

they feel as though the instructors connect with them and see them as more than just another faceless trainee (Greene et al. 2016). To promote discussions outside the classroom, encourage instructors to stay for several days, attend other modules in addition to the one they are teaching, have meals together with the trainees, and also attend the social events. Including arrival and departure dates of all instructors in the syllabus allows trainees to plan interactions with a particular instructor.

TAs should be present during the entire course so that they develop a strong familiarity with the trainees and their skill levels. This makes them the "backbone" of the course as they know which topics were parts of previous lectures and how well the trainees acquired previous programming skills. By our experience, aim for a TA:trainees ratio of 1:6 or 1:5, because in a hands-on course the trainees will need a lot of support. Encourage instructors to submit their materials and exercises as early as possible so that TAs can have access to them before the start of the course enabling them to better prepare. The team of TAs should be highly motivated but also patient and socially skilled (see below). A diverse group of TAs with regard to research background, programming skills, nationality, spoken languages and gender makes it easier to approach different kinds of trainees and likewise for trainees to feel comfortable asking for help. It is very important that you know the TAs' skills and limitations in particular topics and programming languages. Previous psychological studies showed that trainees easily differentiate between attitudes of instructors and TAs in the classroom (Kendall and Schussler 2012). Therefore, we strongly recommend that TAs be eager, well organized and enthusiastic. Trainees will easily pick-up on this and will likely be more motivated themselves. If trainees are struggling, they may blame it on the lack of help. Since the job of the TAs is very demanding you could allow them 1–2 free days during the course to avoid exhaustion. To this end, you can implement a resting rotating system such that a given TA can have a break in the module they are less experienced in while still maintaining the optimal trainee-TA ratio.

Try not to completely change your team from year to year, to ensure sustainability. Once an EBC becomes established, continuity can be achieved by inviting the best instructors/TAs from previous years based on their performance and recommendations from other instructors, TAs, organizers and GOBLET, as well as from trainees through evaluations forms (see below). Another successful strategy is to invite former course participants to return as TAs. Because an instructor or TA might unexpectedly make a last minute cancellation, try to have a plan B, for instance, somebody of your set of organizers, instructors and TAs might be able to give a talk on



the respective topic or you improvise a discussion round with all participants about the topic.

### **Find a suitable venue and provide an adequate infrastructure**

Although an adequate venue and infrastructure is a basic need for any course (Via et al. 2013), it assumes a particular relevance for EBCs, because participants will spend most of the 2–3 weeks at the venue you select. You will need at least one room for the entire duration of the course that provides enough space and access to one computer or workstation for each trainee. Because you are coordinating several instructors with potentially different teaching styles and module topics, ask everybody in advance for the computational requirements of their modules, and install and test all necessary software before the course starts. While the use of personal laptops might be desirable for some trainees, their configurations can be very diverse. In our experience it can be extremely time consuming and frustrating to customize them to fit the course format, even if detailed instructions have been sent beforehand. Therefore, ensure that all trainees are working in the same computational system. One way to achieve this is to prepare USB keys with the operating system and all tools pre-installed, which can be used to boot the computers. However, the performance of the computers or workstations may not be sufficient for all needs. You will also need access to a high-performance cluster or server through cloud computing (Brown 2010). Importantly, perform a final test of the entire computing infrastructure before your course starts to avoid any potential problems (Via et al. 2011). Still, even with all the preparation work, do not underestimate the problems that can occur during a longer course. Thus, a dedicated systems administrator who is involved in planning and on-call duty for the duration of your course is essential.

In addition to the general computer infrastructure, consider having two projectors, one for lecture slides and one for demonstrations of code or results, as jumping between slides and a terminal can be confusing for the trainees. To make the long days as comfortable as possible, you could have a separate room for group discussions, scientific talks and review sessions to assure undivided attention to the instructors. And as a welcomed change for the trainees, provide yet another room for breaks where coffee, refreshments and some food are available.

### **Keep costs of your EBC fair**

Most of the budget typically goes into accommodations, meals, rental of the venue, computer infrastructure, and travel costs. In addition, the complicated logistics of

EBCs may require hiring administrative personnel, to negotiate contracts with service providers and handle invoices, and technical support, such as a systems administrator, etc. (see above). While these costs may be more affordable for short courses, they easily result in prohibitive fees in EBCs.

By making arrangements with the host university, high costs for renting the venue and technical equipment can, at least partially, be avoided. Moreover, to keep the budget within reasonable limits, you can try to arrange accommodations within walking distance of the course venue to save money on shuttles or public transportation. Apartments or university dorms might be a more reasonable alternative to hotels, but you might be able to negotiate with a hotel for an extended stay of a large group. Keep in mind that with such a long course the room is like a resting sanctuary to recover from intensive, long days. Thus, a low budget should not come at the expense of a comfortable and peaceful living environment.

To further help with expense, instructors and speakers might be willing to teach without monetary compensation apart from reimbursement for their travel and accommodation. Inviting local instructors can further reduce accommodation and travelling costs. Although instructors may not need monetary compensation for their participation, it is important to pay some honorary to TAs, given their constant and intensive contribution to the success of the course.

Grant money and sponsorships are other options that can reduce the course fees, at least for trainees from distant countries or from labs with very little resources. However, when choosing sponsors, make sure that they do not influence the course content or the selection of trainees.

While diverse and quality food and drinks should be provided for trainees, there are options for decreasing costs. For instance, instead of hiring caterers, the organizers can make coffee and buy cookies and fruits and plan meals in university canteens, to provide everybody with necessary “brain food”. While meals do not always need to be fancy, it is important to budget healthy food and drinks and to include a “proper” dinner into the course fee.

Ultimately, despite the quality of the course and instructors, the associated fees are often a major factor in determining the number of applicants and the viability of any EBC. But cutting costs too much can have a negative impact on the quality of your course and can lead to a reduction in your trainees’ motivation and appreciation for the course. One should also be cautious when offering courses for free as these are often perceived as less rigorous or of less value. Thus, by paying at least a “symbolic” fee, trainees are generally more committed to attend the entire course.

### Advertisement and trainees' selection

After defining the course objectives, designing the training plan, gathering a team of instructors and TAs, setting up the stage and defining the budget, it is time to advertise and select your trainees. Do not only set up a webpage for the course, but announce the course also in specific platforms (such as GOBLET or Evoldir, [http://evol.mcmaster.ca/cgi-bin/my\\_wrap/brian/evoldir/WorksHopsCourses/](http://evol.mcmaster.ca/cgi-bin/my_wrap/brian/evoldir/WorksHopsCourses/)), so that it will be noticed within the field. Further, ask colleagues and alumni to distribute the information via email and by hanging up posters and advertise it in Facebook and other social media.

Ensuring that trainees' expectations match the training goals is key for the success of any course (Via et al. 2011), but it assumes a particular importance in extended courses because the trainees' investment is higher. To select the most appropriate trainees for your course and to help applicants judge how beneficial the course will be for them, clearly define your objectives and target audience in the call for applicants. The call should include the general structure and specific topics of your course, as well as the affiliation and areas of expertise of instructors. Syllabi or material from past years' courses can be helpful for potential applicants to assess whether the course is well-suited for them. Further define how much prior knowledge and experience you expect from your trainees (e.g. is it a beginners or advanced course). To this end, a quiz can be developed for applicants to evaluate their knowledge in the field and to determine if the course is appropriate for them.

A complementary strategy is to ask the applicants to submit a motivation letter outlining their research background, projects, relevant datasets, their current computational skills, why they would be a good fit for the course and how the course would aid their own research. This will allow organizers to assess how well the trainees' expectations match the course objectives.

When selecting trainees, prioritize those applicants who have the targeted experience level, needs and expectations that are in line with the course goals (Via et al. 2013). Form a selection committee among instructors and TAs as their different perspectives can be beneficial to the selection process. For EBCs, focus on applicants with broad interests and needs over applicants who have specific interests on a single topic (e.g. NGS data analysis), which can be more easily acquired in specialized short courses. Further, try to maximize the diversity of trainees in terms of research background, model systems, experience, as well as nationality, cultural background, gender and personality, as this will enrich the learning experience for everyone. For maximal outreach, favor trainees who are willing to share their knowledge with other students back in their home institutions. This will

allow to multiply the effectiveness of our efforts in preparing modern day evolutionary biologist to face global challenges (Carroll et al. 2014).

### Achieve an optimal balance between theory and practice in each module and continuously support your trainees

The primary goal of most trainees is to apply their newly gained knowledge to their own research. Active learning has been shown to significantly increase proficiency in performance (Udovic et al. 2002; Ebert-May et al. 1997). A good balance between theory and practice is a common guideline for any course (e.g. Via et al. 2011) and particularly important when the practical work is new to the trainees. Although there is no magic number, based on our experience we strongly encourage reserving at least 70% of the workshop time to practical exercises, alternating lecture and practical sessions. New concepts can be introduced in a short lecture followed by a practical session with each new concept and exercise building upon the previous one. EBCs also allow the implementation of problem solving strategies to make trainees progressively independent. Encourage them to look into documentation or discussion forums themselves to find solutions to small problems they did not explicitly see in the lectures and then share them with the class to promote discussion. Extra tips for personal work organization (e.g. computational lab books, github), useful programming/statistical practices (e.g. data scrutiny and p-hacking), as well as advise on study design are also important components that can easily be built into an EBC.

Because trainees' scientific background can be quite heterogeneous (Via et al. 2011, 2013; Schneider et al. 2010) (even within the broad field of evolutionary biology), some trainees might be unaware of key concepts while others might have only limited knowledge in statistics or might be unaware of the intricacies of NGS data. Also, the time it takes trainees to finish exercises usually has a long lagging tail. That is why every effort should be made to not lose any of the class early in the course. The difference in trainees' skills can easily increase with duration of the course, which is why keeping track of their progress is critical during EBCs. TAs should try to accommodate various skill levels and distribute the amount of time spent with each trainee, accordingly. TAs should also be able to identify trainees who are struggling. To help with this, a system with sticky note indicators can be implemented. While trainees are working on exercises, they place a colored sticky note on their computer monitor with different colors corresponding to whether the trainee is struggling or has finished the task. This way, the TAs and instructors can easily keep track of which trainees are

finishing quickly and which are having a more difficult time completing the exercises, and therefore may need help. You should also let the trainees know that the course materials and solutions to exercises will be available after the course and also prepare some extra exercises for more advanced trainees.

At the end of each module, trainees should be able to determine if they reached the aims that were initially defined for that module and their own goals. Otherwise, a review session covering a problematic topic could be implemented, if there is enough flexibility in the timetable.

### **Develop a 3–4 day final project to reinforce and integrate the learned EBC contents**

Most participants in these courses hope to develop concrete ideas as to how to apply the course content to their own research projects (Via et al. 2011). In an EBC you can include several days for a final project during which trainees can address their own research questions or implement workflows that use their own or publicly available data. This will greatly reinforce the topics learned during the course as the trainees can integrate many of the methods, analyses and programming languages to which they were just exposed. From the feedback we received from our trainees, the final projects helped most of them to implement and improve their newly learned skills. Although it is not realistic to expect group publications solely based on the work performed during these 3–4 days, trainees can continue working on the final projects after the course, offering an important basis for subsequent research papers (Hinaux et al. 2013).

To promote interest in the final projects, encourage trainees to propose project ideas either using their own data or publicly available datasets. If there are not enough proposals presented, have a set of “back-up” project ideas and datasets prepared. Projects should be as diverse as possible with regard to the questions asked and methods used so that the trainees with different research interests will all be engaged. Assign one TA to guide the 4–5 members of the project team, to define S.M.A.R.T.-like objectives (Doran 1981), to manage job division within the team and also to deal with any problems that might occur. Organizers and instructors should monitor groups’ progress regularly, in order to correct eventual problems and guide them so that they feel they can accomplish the project. Let the group choose a creative title for their project, e.g., “Bad Brains”, “Stressed shrimps”, “S(t)imulated trees”, “Slimy orthologs”), to increase the fun factor. At the end of the course, groups should give short presentations of their projects and celebrate their hard work together.

### **Build a close-knit community, avoid exhaustion and encourage social activities during the course**

Participants will spend a lot of time together; thus it is even more crucial for EBCs than for other courses to ensure an open atmosphere and friendly interaction between organizers, instructors, TAs and trainees so that everybody feels integrated into the group. To make everybody feel welcome, you should not only introduce them to the venue and the city, but also give them a syllabus and timetable as soon as they arrive or prior to arrival if the schedule has been finalized. Create a comfortable learning environment, encourage everybody to wear name tags, and schedule “icebreaker activities” to allow the participants to get to know each other. For example, all instructors, TAs and trainees could bring food/drink items that are traditional in their respective countries to share during breaks. During the first day(s) of the course, we suggest making an extra time slot where everybody can briefly and informally talk about their own research interests. Restrict the length of these presentations to a few minutes and keep them informal by having them during the evening session and by randomizing their order. This will help to become familiar with each other’s scientific background and technical skills and can assist TAs to motivate trainees during the learning process and to encourage them to suggest project ideas.

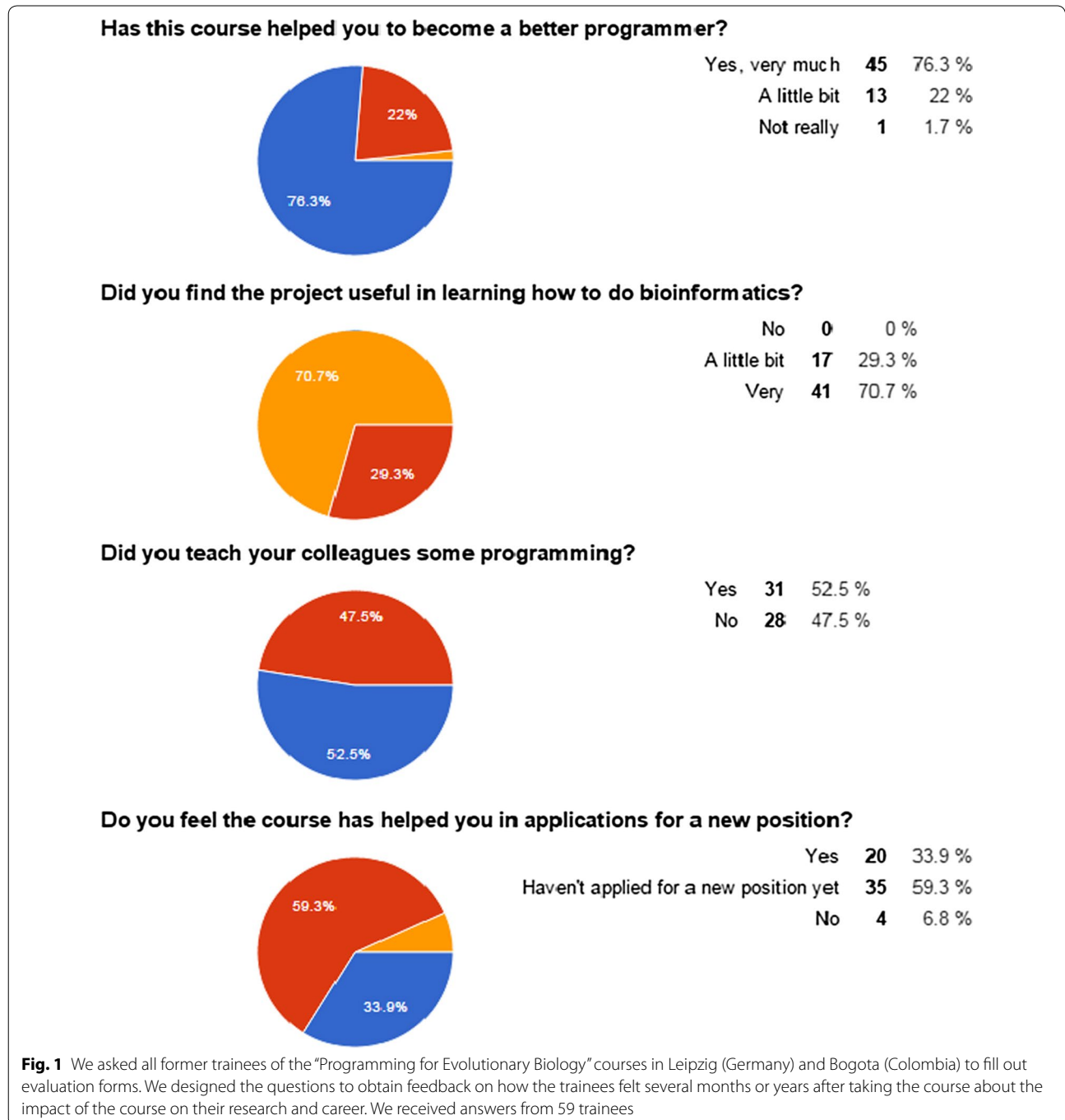
Because everyone will spend several intense weeks and face new challenges you run the risk of exhausting your trainees and TAs with an EBC. To avoid this, create “time oases” to give your trainees much-needed breaks, both mentally as well as physically, from long days of sitting in lectures and in front of the computer. For example, allow your trainees a sufficiently long break at noon to go out for lunch by themselves. Furthermore, daily dinner for the entire group can be provided, thus generating a non-academic space where instructors, TAs and trainees can interact. Also include activities that promote social interaction between all participants outside the classroom, for instance an excursion, a sport event, or a dinner at the end of the course. However, keep things informal and apart from a few organized activities leave it up to the trainees to suggest additional activities, which might include mingling in pubs. We are aware of several long-running courses that have one or more established social activities (“traditions”) that uniquely identify them (e.g. a walk in the castle, sauna or a night swim). Importantly, all these social interactions can contribute to fostering a network that extends beyond the duration of the course. A course becomes “legendary” when it not only helps to build strong scientists but also evolves some of its own rituals, customs and myths.



**Aftermath: continue improving and networking**

Establishing EBC courses is like an evolutionary process requiring constant adaptation to the ever changing training needs by trying out new modules or changes to modules and selecting the most successful changes. To assess these changes, trainees should have ample opportunity to provide feedback through anonymous evaluation forms at the end of the course (Via et al. 2011, 2013). These evaluation forms can be used to gather feedback on any

aspect of the course, from the course structure, the content of each module, the projects, TA's and instructors' performance, or any other aspect of the organization itself, such as accommodation, venue and infrastructure. The main goal of this evaluation is to help organizers identifying the most positive and negative aspects so that they can be kept or corrected in future editions. In addition, it is also important to assess the long-term impact of the course, one or 2 years later (Fig. 1) (Via



**Fig. 1** We asked all former trainees of the “Programming for Evolutionary Biology” courses in Leipzig (Germany) and Bogota (Colombia) to fill out evaluation forms. We designed the questions to obtain feedback on how the trainees felt several months or years after taking the course about the impact of the course on their research and career. We received answers from 59 trainees

et al. 2013). By this you can gain a clearer perspective of the effectiveness of the course in terms of providing the students fundamental skills for solving problems of their own research projects and a more integrative idea about the usefulness of your course by comparing with other courses they may have attended since then. The evaluations are also useful for organizers to confirm current or identify new skills needed in bioinformatics applied to evolutionary biology problems, so that the course content can reflect these trends, and remain useful through time. Furthermore, evaluations can be used to assess the most efficient advertisement strategy of the course. However, it is important to highlight that these evaluation forms should not be a substitute for direct feedback. Trainees, TAs and instructors should be welcome to give their personal opinions at any stage of the course.

To foster long-term contact, provide your trainees access to a server where course material will be accessible after the course and a list of email addresses for everybody to stay in touch. This will allow trainees to contact TAs or instructors when they have questions or need bioinformatics support. Besides this one-to-one interaction, it is important to promote the participation in online forums and the use of social media (e.g. Facebook groups) for networking, sharing new tools, asking for help or advice, and to be aware of job opportunities. Furthermore, former trainees can use these platforms to coordinate meetings at conferences of common interest, or plan a workshop in following years with participants from all course editions, building an active and strong community.

At last, it is important that trainees realize that the course is not really the end, but is instead the beginning. It is never too much to reiterate that they will need to apply the skills they have learned during the course, or otherwise they will lose most of them. It is also important to stress that not every trainee has to become a bioinformatician or a computer scientist! Ultimately, each person has his or her own goals, motivations and interests and after a successful course, we will always end up with a more diversely qualified cohort of evolutionary biologists ready to face the challenges of big data.

### Concluding remarks

Every course format has its pros and cons and the most effective training plan will depend upon your general objectives and targeted trainees (Via et al. 2013). Thus, it is not our goal to present EBCs as a competing alternative but rather complementary to other types of courses. However, our experience reveals that EBCs can be very valuable for scientists when embracing completely new

skills and topics. We have compiled a set of guidelines that, in our opinion, are key to the organization and management of a 2–3 weeks workshop in bioinformatics applied to evolutionary biology, including the importance of defining well the objectives, topics, content, structure of the course; to choose wisely the trainees, TAs, instructors, and the course venue; to provide high-quality training achieving a good balance between theory and practice; to be fair and open about course fees; to facilitate interaction and feedback between all parties involved; and to allow substantial room for social interactions within (e.g. group projects) and outside (e.g. by organizing lunch programs, evening events and excursions) the course schedule, as well as after the course (e.g. conferences, social media). Since EBCs are still rarely offered (out of 94 courses that were listed on EvolDir between February and April 2018 only five were longer than 1 week and included genomics and/or bioinformatics topics), we hope that these guidelines will inspire others to create and help with organizing similar courses. While each of these guidelines have individual merits, it is clear that together they should comply with the triple-S equilibrium: studying-socializing-sleeping.

### Abbreviations

EBC: extended bioinformatics courses; GOBLET: Global Organization of Bioinformatics Learning, Education and Training; NGS: next generation sequencing; SMART: Specific Measurable Accepted Realistic Time Bound to clearly define/communicate goals; TA: teaching assistant.

### Authors' contributions

Rather than relying solely on the opinion of the course organizers, instructors and TAs were also invited to join as authors. The incorporation of their advice and experience in this kind of courses was fundamental for the completeness and of the final version guidelines. This manuscript was written collaboratively and all authors participated in designing the manuscript structure and writing the first draft. RF, DT, WS and KN further worked on it towards the final version. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. <sup>2</sup> IBE, Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Pompeu Fabra University, Doctor Aiguader 88, 08003 Barcelona, Spain. <sup>3</sup> Present Address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK. <sup>4</sup> Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA. <sup>5</sup> TFome Research Group, Bioinformatics Group, Interdisciplinary Center of Bioinformatics, Department of Computer Science, University of Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany. <sup>6</sup> Paul-Flechsig-Institute for Brain Research, University of Leipzig, Jahnallee 59, 04109 Leipzig, Germany. <sup>7</sup> Faculty for Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Königin-Luise-Straße 1-3, 14195 Berlin, Germany. <sup>8</sup> Edinburgh Genomics, The University of Edinburgh, Ashworth Laboratories, The King's Buildings, Charlotte Auerbach Road, Edinburgh EH9 3FL, Scotland, UK. <sup>9</sup> Museo Nacional de Ciencias Naturales, Spanish Research Council (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain. <sup>10</sup> German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. <sup>11</sup> Grupo Rnómica Teórica y Computacional, Departamento de Biología, Universidad Nacional de Colombia, Bogota, Colombia. <sup>12</sup> ecSeq Bioinformatics GmbH, Sternwartenstraße 29, 04103 Leipzig, Germany. <sup>13</sup> Computational Biology, Department of Target Sciences, GlaxoSmithKline, Gunnels Wood

Road, Stevenage SG1 2NY, England, UK. <sup>14</sup> Computational EvoDevo Group, Department of Computer Science, University of Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany. <sup>15</sup> Bioinformatics Group, Department of Computer Science, University of Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany. <sup>16</sup> Visual Data Analysis Lab, ESAT/STADIUS, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium. <sup>17</sup> imec, KU Leuven, Leuven, Belgium. <sup>18</sup> Department of Education and Psychology, Comparative Developmental Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany. <sup>19</sup> Stowers Institute for Medical Research, Kansas City, MO, USA. <sup>20</sup> Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany. <sup>21</sup> Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland. <sup>22</sup> Department of Genetic Research, Institute for Biological Research “Siniša Stanković”, University of Belgrade, Blvd. Despota Stefana 142, 11060 Belgrade, Serbia.

### Acknowledgements

We would like to thank all trainees of the courses “Programming for Evolutionary Biology” and “Programming for Biology” and of the “Workshop on Molecular Evolution” for their evaluations, advice, support and criticism, which helped us to improve our EBC and inspired us to write this manuscript. We are very grateful for receiving financial support from the Volkswagen Foundation, which helped tremendously to start the “Programming for Evolutionary Biology” course and allowed us to prove that it is possible to run such an extended course effectively. We also would like to thank the Research Academy Leipzig from the University of Leipzig, the German Academic Exchange Service (DAAD), and Illumina, from whom we received financial support in subsequent years. The extended course in Colombia was also supported by the equipment donation from the DAAD to the Faculty of Science at the Universidad Nacional de Colombia and partially founded by the Universidad Nacional de Colombia. We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University.

### Competing interests

The authors declare that they have no competing interests.

### Availability of data and materials

Not applicable.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

Not applicable.

### Funding

This manuscript was written without any funding.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 January 2018 Accepted: 9 July 2018

Published online: 26 July 2018

### References

Attwood TK, Bongcam-Rudloff E, Brazas MD, Corpas M, Gaudet P, Lewitter F, Mulder N, Palagi PM, Schneider MV, van Gelder CWG, et al. GOBLET: the

- global organisation for bioinformatics learning education and training. *PLoS Comput Biol*. 2015;11(4):e1004143.
- Brazas MD, Brooksbank C, Jimenez RC, Blackford S, Palagi PM, De Las Rivas J, Quellet BF, Kumuthini J, Korpelainen E, Lewitter F, van Gelder CW. A global perspective on bioinformatics training needs. *BioRxiv*. 2017. <https://doi.org/10.1101/098996>.
- Brown T. Running a next-gen sequence analysis course using Amazon Web Services. Living in an Ivory Basement. 2010. <http://ivory.idyll.org/blog/ngs-course-with-aws.html>. Accessed 2 May 2018.
- Carroll SP, Jørgensen PS, Kinnison MK, Bergstrom CT, Denison RF, Gluckman P, et al. Applying evolutionary biology to address global challenges. *Science*. 2014;346(6207):1245993.
- Consortium GOBLET. The global organisation for bioinformatics learning, education & training (GOBLET). *EMBnet J*. 2013;19:10.
- Corpas M, Jimenez RC, Bongcam-Rudloff E, Budd A, Brazas MD, Fernandes PL, Gaeta B, van Gelder C, Korpelainen E, Lewitter F, et al. The GOBLET training portal: a global repository of bioinformatics training materials, courses and trainers. *Bioinformatics*. 2015;31(1):140–2.
- Doran GT. There's a S.M.A.R.T. way to write management's goals and objectives. *Manag Rev*. 1981;70(11):35–6.
- Ebert-May D, Brewer C, Allred S. Innovation in large lectures: teaching for active learning. *Bioscience*. 1997;47(9):601–7.
- Goldsmid CA. Why formalize the aims of instruction? *Teach Sociol*. 1981;8(3):263–89.
- Greene AC, Giffin KA, Greene CS, Moore JH. Adapting bioinformatics curricula for big data. *Brief Bioinform*. 2016;17(1):43–50.
- Handelsman J, Miller S, Pfund C. *Scientific teaching*. New York: W.H Freeman and Company; 2007.
- Hinaux H, Poulain J, Da Silvam C, Noirot C, Jefferym WR. De novo sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. *PLoS ONE*. 2013;8(1):e53553.
- Kendall KD, Schussler EE. Does instructor type matter? Undergraduate trainee perception of graduate teaching assistants and professors. *CBE Life Sci Educ*. 2012;11:187–99.
- Kwok R. Computing: out of the hood. *Nature*. 2013;504:319–21.
- Losos JB, Arnold SJ, Bejerano G, Brodie ED III, Hibbett D, Hoekstra HE, et al. Evolutionary biology for the 21st century. *PLoS Biol*. 2013;11(1):e1001466.
- Marx V. Biology: the big challenges of big data. *Nature*. 2013;498:255–60.
- Meagher TR. Evolution and today's society. *Bioscience*. 1999;49:923–5.
- Schneider MV, Watson J, Attwood T, Rother K, Budd A, McDowall J, Via A, Fernandes P, Nyronen T, Blicher T, et al. Bioinformatics training: a review of challenges, actions and support requirements. *Brief Bioinform*. 2010;11(6):544–51.
- Schneider MV, Walter P, Blatter M-C, Watson J, Brazas MD, Rother K, Budd A, Via A, van Gelder CWG, Jacob J, et al. Bioinformatics Training Network (BTN): a community resource for bioinformatics trainers. *Brief Bioinform*. 2012;13(3):383–9.
- Udovic D, Morris D, Dickman A, Postlethwait J, Wetherwax P. Workshop biology: demonstrating the effectiveness of active learning in a non-majors biology course. *Bioscience*. 2002;52:272–81.
- Via A, De Las Rivas J, Attwood TK, Landsman D, Brazas MD, et al. Ten simple rules for developing a short bioinformatics training course. *PLoS Comput Biol*. 2011;7(10):e1002245.
- Via A, Blicher T, Bongcam-Rudloff E, Brazas MD, Brooksbank C, et al. Best practices in bioinformatics training for life scientists. *Brief Bioinform*. 2013;14:528–37.