

# Bioinformatics Training Network (BTN): a community resource for bioinformatics trainers

*Maria V. Schneider, Peter Walter, Marie-Claude Blatter, James Watson, Michelle D. Brazas, Kristian Rother, Aidan Budd, Allegra Via, Celia W.G. van Gelder, Joachim Jacob, Pedro Fernandes, Tommi H. Nyrönen, Javier De Las Rivas, Thomas Blicher, Rafael C. Jimenez, Jane Loveland, Jennifer McDowall, Phil Jones, Brendan W. Vaughan, Rodrigo Lopez, Teresa K. Attwood and Catherine Brooksbank*

Submitted: 12th July 2011; Received (in revised form): 31st August 2011

## Abstract

Funding bodies are increasingly recognizing the need to provide graduates and researchers with access to short intensive courses in a variety of disciplines, in order both to improve the general skills base and to provide solid foundations on which researchers may build their careers. In response to the development of ‘high-throughput biology’, the need for training in the field of bioinformatics, in particular, is seeing a resurgence: it has been defined as a key priority by many Institutions and research programmes and is now an important component of many grant

Corresponding author. Maria Victoria Schneider, EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. Tel: +44(0)1223494448; Fax: +44 (0) 1223 492 468; E-mail: vicky@ebi.ac.uk

**Maria Victoria Schneider** is the Training Programme Project Leader in the Outreach and Training Team at EMBL–EBI where she is responsible for developing and coordinating the user training activities of EMBL–EBI.

**Peter Walter** is a Web developer in the External Services Team at EMBL–EBI, where he is responsible for the development and maintenance of the EMBL–EBI website and collaborative web portals with which the EMBL–EBI is involved.

**Marie–Claude Blatter** is part of the communication team of the Swiss–Prot group (SIB Swiss Institute of Bioinformatics), where she coordinates teaching/training programs and also activities for the public at large.

**James Watson** is the Senior Scientific Training Officer in the Outreach and Training Team at EMBL–EBI. Prior to this, he was a postdoctoral researcher in the Thornton group looking at protein structure–function relationships.

**Michelle D. Brazas** is Manager of Knowledge and Research Exchange at the Ontario Institute for Cancer Research, where she coordinates the Canadian Bioinformatics Workshop series and other outreach education programs.

**Kristian Rother** is teaching structural bioinformatics, programming, and software development to bioinformaticians at the UAM Poznan. He is also a contributor to the BioPython and PyCogent software libraries.

**Aidan Budd** is a Senior Computational Biologist at the EMBL in Heidelberg, working in the Gibson Team. Working on a range of projects involving short protein linear motifs and phylogenies, he is also responsible for the training and outreach activities of the group.

**Allegra Via** is a researcher in Bioinformatics and Computational Biology at Sapienza University of Rome (Italy) where she teaches the course of Macromolecular Structures in the Engineering of Industrial Nanotechnology Master Degree. She also teaches programming to bioinformaticians and biologists in short training courses.

**Celia W.G. van Gelder** is project leader education of the Netherlands Bioinformatics Centre (NBIC) and coordinator education and bioinformatics teacher at the Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

**Joachim Jacob** organizes bioinformatics training sessions for the Bioinformatics Training and Services Facility (BITS). BITS is part of the VIB, a life sciences institute in Belgium.

**Pedro Fernandes** is the creator and coordinator of the Gulbenkian Training Programme in Bioinformatics, held at the Instituto Gulbenkian de Ciência in Oeiras, Portugal.

**Tommi H. Nyrönen** is a development manager at the Finnish IT Center for Science—CSC where he is responsible for developing services and technology for biomedical sciences.

**Javier De Las Rivas** is PI Scientist at the Cancer Research Center in Salamanca (Spain) leader of the Bioinformatics and Functional Genomics research group.

**Thomas Blicher** is associate professor at the NNF Center for Protein Research (University of Copenhagen) and Center for Biological Sequence Analysis (Technical University of Denmark). He is involved in teaching undergraduate and graduate students as well as developing methods for use of web-based tools in teaching.

**Rafael C Jimenez** is a software engineer on the Proteomics Services Team at the EMBL–EBI. He has a background in biological sciences and software engineering. He participates in EMBL–EBI training activities focused on data integration and proteomics.

proposals. Nevertheless, when it comes to planning and preparing to meet such training needs, tension arises between the reward structures that predominate in the scientific community which compel individuals to publish or perish, and the time that must be devoted to the design, delivery and maintenance of high-quality training materials. Conversely, there is much relevant teaching material and training expertise available worldwide that, were it properly organized, could be exploited by anyone who needs to provide training or needs to set up a new course. To do this, however, the materials would have to be centralized in a database and clearly tagged in relation to target audiences, learning objectives, etc. Ideally, they would also be peer reviewed, and easily and efficiently accessible for downloading. Here, we present the Bioinformatics Training Network (BTN), a new enterprise that has been initiated to address these needs and review it, respectively, to similar initiatives and collections.

**Keywords:** *Bioinformatics; training; end users; bioinformatics courses; learning bioinformatics*

## INTRODUCTION

As biological data have become increasingly central to the life sciences, there is a need both for professional bioinformaticians and for ‘data savvy’ life scientists who are comfortable accessing, submitting and analysing data. Over the past 10 years, higher education institutions have begun to offer undergraduate and postgraduate courses in bioinformatics and computational biology. This is addressing the need for professional bioinformaticians but does not address the very different needs of the growing and diversifying number of life scientists who need to be able to access, submit and analyse biological data. Bioinformatics training during life science undergraduate studies could partially address this need, but is not yet universal in life-science curricula. The demand for a variety of short, focused courses, ranging from introductory to advanced levels, has therefore grown, and continues to grow, enormously [1]. This growth is also partly related to rapid technological advances in experimental techniques, especially those that characterize modern ‘high-throughput biology’, and the parallel evolution of the bioinformatics tools and resources necessary to manage and analyse the vast quantities of data these are producing [2].

**Jane Loveland** works in the HAVANA group on vertebrate genome sequence annotation in the Wellcome Trust Sanger Institute. She co-ordinates the Open Door Workshops with Wellcome Trust Advanced Courses.

**Jennifer McDowall** is a Scientific Training Officer in the External Services Team at EMBL-EBI where she is responsible for developing and teaching patent-related training activities.

**Phil Jones** is a senior software engineer on the InterPro project, a database of protein families, domains, regions, repeats and sites at the EBI. He has a background in both biological sciences and software engineering, and holds a PGCE in secondary level science education.

**Brendan W. Vaughan** is the web coordinator in the External Services Team at EMBL-EBI where he is responsible for coordinating the development and maintenance of the main EMBL-EBI website along with numerous collaborative web portals with which the EMBL-EBI is involved.

**Rodrigo Lopez** is the head of the External Services Team at EMBL-EBI.

**Teresa K. Attwood** is professor of Bioinformatics at the University of Manchester; she coordinates and teaches the introductory module of the Bioinformatics Masters course, and has written several bioinformatics reference and textbooks.

**Catherine Brooksbank** is the Head of Outreach and Training at EMBL-EBI where she coordinates EMBL-EBI’s public relations and user training programmes.

Currently, short courses (typically from 1 to 5 days in length) in various bioinformatics topics are delivered worldwide. They generally represent a compromise between content depth and time, where workshops cater to heterogeneous target audiences and provide focused bioinformatics training in relatively short time-frames. Anyone involved in organizing and delivering such courses faces common (and not trivial) challenges [3]. The Bioinformatics Training Network (BTN), a community-led project that arose in response to these challenges, aims to provide platform supported, pragmatic solutions for the exchange of expertise, training materials and training experiences. Here, we describe this effort and invite anyone involved in bioinformatics training to engage with, and contribute to, this community effort.

## WHAT CAN YOU DO WITH THE BTN?

The BTN website (<http://www.biotnet.org/>) is a centralized resource for the bioinformatics community to share training materials and techniques, and to disseminate awareness and spark new discussions and initiatives. The global community is welcome to

browse the site, and we encourage, in particular, those who are interested or involved in bioinformatics training to join the BTN. All materials in the BTN are made freely available under the Creative Commons Attribution-Share Alike 2.5 License (<http://creativecommons.org/licenses/by-sa/2.5/>). The site has been developed by the European Bioinformatics Institute (EMBL-EBI <http://www.ebi.ac.uk/>), using the Open Source content management system Drupal (<http://drupal.org/>). To facilitate searching, a key feature of the site is the use of tag words. These are currently added as free text when users upload their materials, but ultimately will be compiled into a controlled vocabulary reflecting the most frequently used terminology and using available established ontologies when appropriate. All terms used as tags are searchable, allowing users to sort information at all levels. This facilitates efficient exploration, leading users more rapidly to relevant information (e.g. tag 'PhD': by clicking on the term PhD, users obtain details of all trainers registered with the site who provide training for PhD students). The five main areas of the BTN website are described in detail below.

### **Trainers and organizers**

This section lists individuals who have registered to contribute to the site and are involved in bioinformatics training; this includes not only those delivering training sessions directly, but also those involved in the organization and planning of such training. The rationale behind this is 2-fold: first, to build up a community in which trainers can identify each other and share areas of expertise, and in which they can liaise with, and get help from, other trainers; second, to function as a hub for course organizers to identify potential trainers for specific subjects and levels.

### **Training materials**

This section is the main focus of the BTN site. This section is a direct response to the often-voiced desire to have a common repository of well-documented, open-source training materials. It is specifically intended to help bioinformatics trainers to share, adapt and develop materials, and to make them freely available to the community at large. It is hoped that both having a centralized repository and fostering a culture of sharing will help new ideas and best practices to be readily and rapidly disseminated. To this end, trainers have the opportunity to rank each

other's materials, in a similar way to the Amazon ranking system. This includes a comments box to allow trainers to state exactly what was rated and the reason for the given rankings. This functionality aims to stimulate constructive feedback and, consequently, to help improve the quality and timeliness of materials available from the site (Figure 1).

Training materials are grouped into 10 major categories, according to the types of data handled by different tools and resources: Genes and Genomes; Proteins and Proteomes; RNAs and Transcriptomes; Interactions and Pathways; Small molecules and Metabolomes; Ontologies; Literature and Text mining; Programming and Operating Systems; Biostatistics; and Other. Where appropriate, materials can be tagged to more than one section (e.g. 'drug discovery' is relevant to both proteins and small molecules). These categories have been derived from extensive discussions among the contributing authors. However, as the site grows and the community enlarges, the categories will doubtless evolve to reflect wider community needs and interests. Special care will be taken to monitor the content of the 'Other' section, to avoid this growing amorphously, and making it more and more difficult to find relevant materials within it.

Anyone can readily access and download materials from the BTN site without a requirement for registration. However, to upload information, registration is necessary. To register, contributors need only provide a brief motivation: this is required in order to ensure that requests are genuine. Apart from a textual description, uploaded materials can include files with any of the following extensions: ppt, pptx, doc, docx, xls, xlsx, pdf, txt. Other formats (e.g. videos) can be incorporated as external hyperlinks. At present, the maximum file size allowed per upload is 64 MB.

While aiming to provide a comprehensive repository of training materials, the BTN nevertheless wishes to avoid redundancy. Therefore, for those who want to share materials but already have a location from which these are available, the BTN site offers the ability simply to add a link to remote material rather than requiring them to be uploaded again. The materials, links and classifications (especially the 'Other' section) will be reviewed periodically by appointed BTN members to provide the greatest functionality to the community, while retaining system security and ensuring that the BTN works within system capacities.

The screenshot shows the Bioinformatics Training Network (BTN) website. The main navigation menu includes Home, Trainers & Organisers, Training Materials (highlighted), Training Facilities, Courses & Events, Useful Links, and About Us. The 'Training Materials' section is titled 'TRAINING MATERIALS' and contains a 'Document Categories' table with the following categories: Gene & Genomes, Protein & Proteomes, RNAs & Transcriptomes, Interactions & Pathways, Small molecules & Metabolomes, Ontologies, Literature & Text mining, Programming & Operating systems, Biostatistics, and Other.

Below the categories, there are two columns: 'Most Recently Updated' and 'Most Recently Rated'. The 'Most Recently Updated' column lists materials such as 'Introduction to Biopython' (updated 4 days 5 hours ago), 'Microarray Data Analysis' (5 days 2 hours ago), 'Ensembl presentations and course booklet' (1 week 7 hours ago), 'Ensembl tutorials' (1 week 7 hours ago), and 'R Basics' (1 week 5 days ago). The 'Most Recently Rated' column lists materials such as 'History of bioinformatics' (5 stars), 'Introduction to Biopython' (4.5 stars), 'Introduction to protein sequence analysis' (5 stars), 'Python programming for beginners - handouts and exercises' (5 stars), and 'EMBL-EBI resources and tools quick list' (5 stars).

The footer of the page includes logos for the Swiss Institute of Bioinformatics, VIB, nbic, Netherlands Bioinformatics Centre, bioinformatics.nl, EMBL-EBI, and SLING. The copyright notice reads: © 2010-2011 Bioinformatics Training Network. All Rights Reserved.

**Figure 1:** Screen dump of one of the BTN web pages showing the section about “Training Materials” containing a top level list of topics (categories), a list of the most recently updated materials (on the left) and a list of the most rated materials (on the right).

Finally, to minimize the risk that the BTN accumulates dysfunctional or outdated materials, automatically generated e-mails will be sent to the submitter 2 years after the submission date asking them to either confirm that their materials are still valid and useful or to agree to the deletion of their materials from the site.

### Training facilities

By sharing our own trainer experiences, it emerged how challenging it can be to get a quick overview of available IT-training facilities and appropriate venues (worldwide). This section was created to allow course organizers and providers to list facilities suitable for hosting bioinformatics training events. At the moment, this is a simple list containing the venue name, location, a brief description, and the relevant

contact details. In the future, as the community grows, a map view will also be developed to allow rapid identification of suitable venues, together with the contact details of the responsible organization/institute.

### Courses and events

Raising awareness of training events is often a significant task for course organizers. Although the primary intention of the BTN site is to address the needs of trainers and course organizers, it emerged from our discussions that being able to identify other courses in the same field would allow potential trainees to locate courses running elsewhere if, for example, a course of interest was fully booked/cancelled, or if his/her expectations were not matched by the learning objectives of the chosen

course. This section is not intended to be a holding place for event websites. Instead, it will act as a bill-board for upcoming events, listing a minimal set of information to allow users to decide whether particular events are relevant to them, plus a hyperlink to the actual event website. This could be useful if, for example, during a course, trainers could point to the BTN as the place where trainees could access all the materials, as any uploaded materials could be searched for using the relevant course-name tag. This would have the additional effect of providing indirect access to the course for remote trainers, allowing them to see what type of material is being taught and where. At present, the list of events can be sorted by title, venue or date. Ultimately, the aim is also to integrate this into a calendar view option.

### Useful links

This section lists the URLs of bioinformatics training providers, promotes awareness and allows informal networking and exchange of experiences. It also represents a platform to discuss new ideas and initiate white papers on best practices.

## BIOINFORMATICS TRAINING: WHAT IS OUT THERE AND HOW THE BTN FITS

There are several initiatives to make bioinformatics course materials available. We review these here and summarize how the BTN differs from these initiatives. Our analysis reveals considerable variation in the functionalities provided by these sites, and in the maintenance and updating mechanisms for training materials. The major challenges that we face when looking for bioinformatics training materials prompted us to look for the following characteristics:

- (i) Accessibility: is it possible to download the materials without logins or fees?
- (ii) Date stamp: is it easy to work out when the materials were created and updated?
- (iii) Peer review: are the materials reviewed or rated by other trainers?
- (iv) Search functionality: how easy is it to retrieve training materials relevant to your specific area of interest?
- (v) Trainer/contact information: can I find out who developed the materials and get in touch with them?
- (vi) Purpose: when relevant, is it easy to link the materials to the course(s) in which they were used?

- (vii) Community engagement: can I upload my own training materials and make them available for other trainers to use?

Table 1 lists the online bioinformatics training resources that we reviewed. Some repositories might be described as interactive e-learning web portals than repositories of materials, but as these are also used by trainers to source training materials, we felt that it was appropriate to include them. When reviewing these sites, large numbers of dead links were encountered and often it was difficult to access information about when was the last content updated.

In light of these diverse endeavours, the BTN aims to provide a complementary, cross-institute portal for identifying training materials, and to help share and promote the recognition of, and feedback to, such institutional initiatives. It also provides an opportunity to broadcast enterprises relevant to individuals involved in bioinformatics training including those devoted to education (which we define here as longer courses, such as Master's or undergraduate programmes), journal and publication opportunities in which bioinformatics training materials are the focus (e.g. tutorials), and funding opportunities for bioinformatics education. The BTN targets trainers (in contrast to other training repositories) by lifting duplication of effort from their shoulders. If constructive collaborations can be built among the BTN's community, then projects may ultimately evolve whose whole is greater than the sum of their parts; this philosophy is the foundation on which the BTN is built.

## WHAT HAPPENS NEXT: TOWARDS AN OPEN, FREE AND COLLABORATIVE EVOLUTION

The BTN website is active at [www.biotnet.org](http://www.biotnet.org). It differs from other repositories of training materials in that its focus is on trainers, rather than trainees. Other functionalities unique to the BTN include the ability to review the materials and rank them according to a star-rating system. We invite all those involved in bioinformatics training either to register and join, or simply to browse and provide feedback, so that the site can be improved and can evolve new functionalities. By openly and freely sharing our bioinformatics training content, wisdom and experiences, we can meet the growing training needs and challenges of present research environment.

**Table 1:** Online resources that provide information and materials relevant to those involved in bioinformatics training

Name	Nature of effort (collaborative versus single Institution)	URL	Login	Freely download/access materials <sup>a</sup>	Updates clearly indicated	Reviewing/ranking materials option	Course materials searchable	Trainer/contact information	Info about training facilities	Links to courses and events
Bioinformatics Training Network	Collaborative effort	www.biotnet.org	Yes (only to upload and review)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
European Molecular Biology Network	Collaborative effort	www.embnct.org http://www.embnct.org/en/QuickGuides	No	Yes	Yes	NA	No	NA	No	Yes
EMBER Consortium	Collaborative effort	http://www.ember.man.ac.uk	Required	Yes	Unclear	No	No	Yes	No	No
International Centre for Genetic & Engineering and Biotechnology	ICGEB	http://net.icgeb.org/	Yes (login as guest)	Yes	Yes	No	No	Yes	No	Yes
The Bioinformatics Training Resource	BTR	http://www.med.nyu.edu/tcr/tcr/btr/	No	Yes	No	NA	No	No	No	No
Bioinformatics Training and Service facility	BITS	http://www.bits.vib.be/training	No	Yes	Yes	No	No	Yes	No	Yes
MITOPENCOURSEWARE	OCW	http://ocw.mit.edu/courses/	No	Partially	Unclear	No	Yes	No	No	Yes
Color Base Pair: Bioinformatics resources	ColorBase	http://www.colorbasepair.com/index.html	No	Yes	Yes	NA	No	No	No	Yes
Bioinformatics organization, Inc.	Bioinformatics Org	http://www.bioinformatics.org/edu/	Yes	No	Yes	No	Yes	Yes	No	Yes
OpenHelix	OpenHelix	http://www.openhelix.com/	No	Partially	Yes	No	Yes	No	No	No
European Proteomics Association	EuPA	www.europa.org/	No	Yes	Unclear	No	No	No	No	Yes
Canadian Bioinformatics Workshops	Institution based	http://www.bioinformatics.ca/workshops/open_access/	No	NA	Yes	NA	No	Yes	No	Yes
NCBI Training & Tutorials	NCBI	http://www.ncbi.nlm.nih.gov/guide/training-tutorials/	No	Yes	Unclear	No	No	No	No	No
eProxemis	SIB	http://e-proxemis.expasy.org/	Yes	NA	Unclear	No	No	No	No	No
MyBio	Collaborative effort	http://mybio.wikia.com/wiki/Bioinformatics_education	Yes	Yes	Unclear	NA	Yes	No	No	No
Virtual Academy of Bioinformatics	BioInfoBank	http://wab.bioinfo.pl/	No	Yes	Yes	No	Yes	Yes	No	No
Babraham Bioinformatics	Babraham Institute	http://www.bioinformatics.bbsrc.ac.uk/training.html	No	Yes	Unclear	No	No	No	No	No
OpenWetWare	Collaborative effort	http://openwetware.org/wiki/Wikimomics	Yes	Yes	Yes	No	Yes	No	No	No

<sup>a</sup>By materials, we mean files (e.g. Microsoft PowerPoint; Adobe PDF), links, web pages or any other content containing information that can be re-used by trainers.

**Key Points**

- The Bioinformatics Training Network (BTN) is a global community-led project aimed to provide pragmatic solutions for the exchange of expertise, training materials and training experiences.
- Constructive collaborations can be built by sharing experiences also at the level of Bioinformatics training. All materials in the BTN are made freely available under the Creative Commons sharealike license.
- A free centralized repository and registry of training materials (and related Bioinformatics information) ensures efficient use and re-use of efforts made across the globe as well as a direct vessel to feedback and improve such resources.
- The BTN also provides an opportunity to broadcast enterprises relevant to individuals involved in bioinformatics training, including education, journal and publication opportunities in which bioinformatics training materials are the focus as well as funding opportunities for bioinformatics education.

**FUNDING**

This work was supported by funds awarded to the EMBL-European Bioinformatics Institute by the

European Commission under SLING, grant agreement number 226073 (Integrating Activity) within Research Infrastructures of the FP7 Capacities Specific Programme.

**References**

1. ELIXIR Training Strategy Committee. Recommendations for a coordinated approach to bioinformatics user training in Europe. [http://www.elixir-europe.org/bcms/elixir/Documents/reports/WP11-Training\\_Strategy\\_Committee\\_Report.pdf](http://www.elixir-europe.org/bcms/elixir/Documents/reports/WP11-Training_Strategy_Committee_Report.pdf) (26 August 2011, date last accessed).
2. Kallioniemi O, Wessels L, Valencia A. On the organization of bioinformatics core services in biology-based research institutes. *Bioinformatics* 2011;**27**(10):1345.
3. Schneider MV, Watson J, Attwood T, *et al.* Bioinformatics training: a review of challenges, actions and support requirements. *Brief Bioinformatics* 2010;**11**(6):544–51.