Developing an online programme in computational biology

Heather M Vincent and Christopher Page

Submitted: 12th July 2012; Received (in revised form): 6th August 2012

Abstract

Much has been written about the need for continuing education and training to enable life scientists and computer scientists to manage and exploit the different types of biological data now becoming available. Here we describe the development of an online programme that combines short training courses, so that those who require an educational programme can progress to complete a formal qualification. Although this flexible approach fits the needs of course participants, it does not fit easily within the organizational structures of a campus-based university.

Keywords: computational biology; online Masters; professional development; bioinformatics courses; distance learning; training

INTRODUCTION

Several authors have published articles on the continuing need for education and training in the broad area of computational biology. The International Society for Computational Biology has a long-standing Education Committee (EduComm) that provides advice on the development of educational programmes [1], which are mainly for full-time students. On the other hand, Via et al. [2] explain that, as the high-throughput methods provide us with increasing quantities of biological data for storage and analysis, there is a parallel need for part-time training courses that will enable professional people to develop their skills in computational biology. All of these education and training courses aim to fulfil the requirements of both biologists and computer scientists, providing opportunities for people from different backgrounds to develop their computational and their biological thinking.

Unfortunately it is not always possible for professional scientists to take time away from their place of work to attend an intensive training course, and some people will wish to take a number of courses to extend the range and depth of their learning [3]. A series of online modules can provide alternative means for scientists to fulfil their needs for professional development. When individual modules are linked into a co-ordinated programme of study, the course participant can choose to work towards a formal qualification. Furthermore, online delivery is a particularly appropriate means to deliver a course concerned with the use of online tools and data sources. An advantage to the module provider is the flexibility of the online modules, which can also be used to support research staff on campus. At the University of Manchester some members of staff take online courses alongside the fully distance students. We also use the online exercises in teaching PhD students, who would otherwise need specifically designed training in computational methods. This reuse leads to lower costs per student.

While there is some teaching of computational biology at undergraduate level, most people wishing to acquire skills in this area are either new graduates in a life science or in computer science, or they are professional scientists looking for opportunities for development. The programmes designed for these two groups are referred to as education or training programmes [4]. An education programme is generally intended for full-time students, who will attend for at least a year. These programmes are designed to provide broad coverage of theoretical concepts, and the successful student can graduate with a formal qualification. A training course, which is of a much...
shorter duration, is often more suitable for professional scientists already in employment. However, although the scientists attracted to short training courses do value the mono-thematic and practical focus of these courses, they are also keen to build their research capabilities [3]. The course options are shown above (Figure 1). In the remainder of this article we will describe our work in the development, delivery and management of online modules, which enable a student to work towards a part-time qualification in computational biology. We offered our first online module in 1998, so we now have nearly 15 years experience in developing and updating courses to match the needs for training in this rapidly developing area.

**BACKGROUND**

A graduate seeking training in computational biology must first decide whether this should be full-time or part-time. The full-time route will normally lead to a qualification at MSc or PhD level.

For people in full-time employment, a part-time course is probably the only practical solution. The options here range from self-teaching from online resources, to enrolment on a part-time programme with support from a tutor (Figure 1).

Although there are, now, many excellent web-based resources for self-teaching, a study of distance education shows that, without learner support, success rates are low [5].

At the next level of support, some help on a specific topic may be provided by volunteers. For example, self-learners, and others who do not have access to bioinformatics experts at their place of work, can post questions to online bulletin boards. The BIONET [6] was an early example of such an online community. More recently BioStar uses a sophisticated platform for users to ask and answer questions, and to rate questions and answers [7]. The option to vote on contributions is a key feature, as it means that the work of volunteers is rewarded.

Learners taking an organized training course, which may be face-to-face or online, are fully supported by trained staff [4], so participants can ask for help as required. These courses are of fixed duration, and can be supplemented by the use of the open access resources. An attendance certificate might be provided, but coursework is not normally assessed. Without some assessment it is not possible to provide a pass/fail certificate.

The UK Open University, which can claim to be the first institution to provide distance education with an integrated support system [5], provides an excellent example for anyone developing distance courses for part-time students. When we began running online modules in 1998, we therefore aimed to provide a similar level of learner support at the module level. Tutors answer questions on the course content and, as the modules are accredited, evaluate work submitted for assessment.

We also realized that many people taking a course for professional development need a certificate. Sometimes the certificate is necessary to show to an employer funding the course that the student had successfully completed the assessments. In other cases the student might be planning a career move, so a formal qualification would be useful. We therefore began work on the development of a distance programme, which can lead to a Masters qualification in computational biology. For this web-based programme, we deliver content from a Learning Content Management System (LCMS). In the early years we used WebCT, then a research project at the University of British Columbia [8]. We now use Moodle [9].

A distance programme is more than a series of modules as students require additional guidance on the route that will best support their career aims [10]. This article describes our work in the design, development, running and updating of a distance programme in computational biology.

---

Figure 1: Types of course available for people seeking professional development in computational biology, organized by level of tutor support.
DEVELOPING A DISTANCE MASTERS PROGRAMME

We will begin with an overview of our approach to teaching before discussing the design of the whole programme and the issues arising at the implementation stage.

Designing the modules

The overall design of our modules is now the result of over a decade of research, and has been refined in response to feedback from students.

In a module we aim to provide background theory, practical exercises to re-enforce theoretical understanding and assessments that will provide the bases for feedback, further discussion and progression. In combining theory with a large practical component, we aim to combine the best elements of the more traditional education programmes and focussed training courses.

We assess understanding by providing several short exercises at the start of the module. In many ways computational biology is an ideal subject for online teaching, as the practical exercises arise naturally from the course material. For example, our course in microarray data analysis is based around a series of exercises using Bioconductor [11]. Following an introductory section on microarrays and experimental design, students download a dataset and follow practical instructions for carrying out checks on data quality. Questions and answers are posted on the module bulletin board.

During this exercise students are asked to identify an anomalous array from the dataset, and to consider the possible reasons for the anomaly. The aims of the exercise, from a teaching perspective, are partly to ensure that students can use Bioconductor, but also to assess understanding of the technology for data capture.

Students use the exercise upload area in Moodle to submit brief notes on their conclusions. The tutor marks the exercises and posts individual feedback, so that students have the opportunity to correct any mis-understanding at an early stage. The tutor also posts a summary of the exercise, emphasizing the importance of understanding the method for data capture. The learning design is summarized in Figure 2.

To consolidate learning from the exercises, students are then given a short project. These module projects develop understanding, giving the best students the opportunity to demonstrate their research skills within the module. In the overall design of the programme, the module projects provide those students who wish to progress to the Masters degree with practice in project management and in writing a short paper.

As we are interacting with students scattered across the globe, we have to ensure that none of our exercises will require high bandwidth. We therefore test download times from outside the University network. Also, we check that recommended software will run on commonly used platforms, and make any specific requirements clear in advance.

Programme design

Our focus is on the provision of accredited modules for professional development. The credits awarded by the University allow the successful student to build towards the award of a formal qualification. We attract people with a range of professional aims, so we need to build flexibility into our programme. We also need to provide guidance on the most appropriate routes through the programme.

The organizers of an accredited programme must ensure that graduates leave with the set of skills expected by employers, so we provide a number of core modules. The designers of any new Masters programme need to consider what to cover in the core modules. For those working in computational biology the EduComm has recently presented the results of a survey of curricula for bioinformatics programmes, and has identified a consensus set [1]. The most important topics identified are programming, statistics and databases, which are also core modules in our programme. The optional modules that we
offer have changed as new methods have been developed and new datatypes have become available. For example, our next module will cover methods for Next Generation Sequence analysis. Our approach to course updating is described in more depth below.

As online courses attract people with a much wider range of entry qualifications and experience than is normally the case for a more conventional, full-time, Masters programme, we need to put in place a robust process for approving exemptions for those who already covered the objectives specified for the core modules [12].

Our overall programme therefore includes a range of core and optional modules, a literature project and a research project. The topic for the literature or the research project is chosen in consultation with the student, and each project is individually supervised.

A successful student may choose to leave the programme with a module certificate, providing evidence of the skills developed in a single topic. Other people join the programme in order to gain a qualification. A student who wishes to graduate with a Masters degree must pass six modules and a research project. Others may choose instead to leave with a Postgraduate Diploma (six modules and a literature project), or a Postgraduate Certificate (four modules).

Programme development and administration

The programme accreditation process takes some time, so it is important to begin planning at an early stage. The programme leader will also need to assemble a small team of collaborators, who will help to define the content of each module.

Many universities in the UK have recently re-organized to bring smaller units together into larger Faculties [12]. Despite this, computational biology still crosses organizational boundaries, with life scientists and computer scientists accommodated in different Faculties. A successful course will therefore require collaboration across these boundaries. This collaboration is necessary during programme design, module design, in module teaching and it is essential to the organization of research projects.

A programme that is designed to provide opportunities for professional development to people from a range of different backgrounds needs to include a broad range of modules in different specialist areas, so the design of the overall programme requires input from those engaged in research across the breadth of the subject. Some experts might write a complete module, but it is generally more appropriate to ask different people to contribute module sections, collaborating at the module level. Collaboration in module design is particularly important in the computer science modules, which require built-in examples that will be relevant to biologists.

The amount of time needed to design and write an online course is rarely recognized. We estimate that it will take an experienced writer at least 50 working days to produce the first version of a module. We then expect to carry out a substantial revision after the first run, and to make additional adjustments after each subsequent run. This is necessary to ensure that we continue to cover the most recent methods. Others have suggested that the preparation of a completely new module can take up to 12 months [13]. A large part of this time is spent in preparing suitable practical exercises. It is difficult for someone who already has a full teaching and research load to find the time to write a new online module, so it may be necessary to seek external support. We have been very fortunate in the level of support that we have received from the Biotechnology and Biological Sciences Research Council (BBSRC) [14].

Once the module has been written, teaching effort must be sustainable. Campus-based universities measure staff–student contact time in terms of lectures, practicals and tutorials, whereas distance teaching may not fit into similar defined blocks of time. The accurate costing of delivery of a distance course is necessary in order to make accurate comparisons with on-campus teaching, but current cost models are inaccurate [15].

Supervision of research projects may also require collaboration across Faculties. One example would be a student developing a new computational method for biological data analysis. He needs advice on software development, and also on similar tools already available, so he is has joint supervision from different Faculties.

People working in computational biology will probably already have inter-Faculty links, which make good starting points for collaboration on programme and module development. Collaboration based on teaching can then lead to new opportunities for collaborative research.

After all the work necessary to obtain accreditation and funding, keeping the programme running
requires further sustained commitment from the members of the programme team. Online students are members of our university, so we need to ensure that everyone attending online has access to the same levels of support as students who attend on campus. On-campus students can often help each other in finding and interpreting information, but this informal support is not readily available to distance students. We need to be sensitive to these needs, and to provide help in navigating systems that are designed with the assumption that all students are studying full-time, on campus. For example, student registration is organized around the September start date for full-time students. Payments are recorded, students are registered for the whole year, and student cards are prepared for collection from an office on campus. To run a flexible programme we need to enable students to register at different points through the year, for periods of less than a full year, and cards must be posted to students. At present, flexible enrolment requires manual intervention in the registration process. This means that additional effort is needed, and additional administrative overheads incurred, in order to ensure that distance students are not disadvantaged at the registration stage.

The support system must operate from the initial contact from a potential applicant to writing references for graduates. In practice the first contact will probably be with our web pages, which must be designed to provide full details about the modules together with clear routes through the programme. In the Programme Handbook we also aim to make the links between the different modules clear. However, as noted above, the potential applicants to an online programme come from a wide range of different backgrounds. We therefore need to provide personalized advice on the options that will best suit the needs of individual applicants. The importance of supporting students not just at module level, but also in their wider development, is discussed by Stevens and Kelly [10]. The more successful students have specific reasons for studying, so a personal tutor who can help individuals to identify learning goals will also help the student in the development of reflective learning. Stevens and Kelly [10] describe practice at the Open University, which has been developed to support the career development of part-time learners. We too have recognized the importance of this programme-level support, so we aim to provide advice on career options in computational biology. We provide personal advice by email, phone or Skype by arrangement with the student.

This personal guidance includes advice on the choice of a research project. Many of our students are already active in research, and so will be able to propose a research topic. In such cases the role of the personal advisor is to evaluate the proposal and to advise the student on any necessary revisions. The advisor will then consult the members of the computational biology community to find an appropriate internal supervisor. Students who are not already working in a research environment will discuss their aims and ambitions with the personal advisor, who will then help them to choose from the projects available within the university.

Examples of the way in which we support our part-time students are provided in Table 1.

We also provide generic advice within our LCMS. The organization of the different sections available to students is shown in Figure 3.

The Programme Handbook and the Study Centre are open to all students, while a module is only open to students registered for the module. There are private areas for individual research projects, which may be shared with an external supervisor. We use the shared bulletin boards to provide programme information and to support discussion across the programme. For example, we have a 'Job Centre' bulletin board, which we use for posting information on any opportunities that might be of interest to our students and graduates.

### Updating course content

In the period over which we have been offering courses there have been huge changes in both data types and in methods for analysis. In response we have developed new courses, and have also regularly revised existing material.

Particular care must be exercised when developing and revising material, especially when these tasks are performed in a collaborative environment. Failure to rigidly enforce some form of centralized version control will inevitably lead to multiple, disparate and inconsistent copies of the module material on numerous machines, requiring tedious and error-prone merging of the various versions. Numerous possible methods may be adopted to prevent this situation arising, and we have explored a number of options.
Table 1: Examples of ‘student pathways’ [10]

<table>
<thead>
<tr>
<th>Student 1</th>
<th>Student 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recent PhD in Zoology. Aims to gain computational skills to enhance experimental work.</strong></td>
<td><strong>BSc and extensive work experience in computer science. Aims to apply for posts in computational biology.</strong></td>
</tr>
<tr>
<td><strong>Discussion, recommendations and outcome</strong></td>
<td><strong>Discussion, recommendations and outcome</strong></td>
</tr>
<tr>
<td>Begin by developing good programming practice</td>
<td>Keen to work towards the MSc, and has time to take two modules in first semester. Recommend background reading in molecular biology</td>
</tr>
<tr>
<td>Introduction to software development in Java</td>
<td>Introduction to Bioinformatics AND Bioinformatics for Systems Biology</td>
</tr>
<tr>
<td><strong>Now a new mother, so takes a break</strong></td>
<td>Needs to gain a better understanding of data types. Recommend Biocomputing, in which biologists and computer scientists work together on group projects.</td>
</tr>
<tr>
<td>Bioinformatics for Systems Biology</td>
<td>Biocomputing AND Maths for Metabolic Modelling</td>
</tr>
<tr>
<td>Returns to work and is interested in taking an introduction to modelling.</td>
<td>We agree that experience with transcriptome data would be useful. Bioinformatics for Transcriptomics is not running this semester, but taking just one module leaves time for project planning.</td>
</tr>
<tr>
<td></td>
<td>Discuss project requirements</td>
</tr>
<tr>
<td></td>
<td>Bioinformatics for Transcriptomics</td>
</tr>
<tr>
<td>Recognizes that knowledge of R will be important to career options.</td>
<td>Makes contact with a research lab near place of work. We arrange joint project supervision, with one supervisor in this lab and the other in Manchester.</td>
</tr>
<tr>
<td>Maths for Metabolic Modelling</td>
<td>Research project</td>
</tr>
<tr>
<td>Understanding of databases is a core skill in computational biology, so recommend a database course.</td>
<td>Databases and data modelling</td>
</tr>
<tr>
<td><strong>Completes Postgraduate Certificate</strong></td>
<td><strong>Graduates with a Masters degree</strong></td>
</tr>
<tr>
<td>Now working with microarray data, so decides to take an additional module</td>
<td>Bioinformatics for Transcriptomics</td>
</tr>
</tbody>
</table>

Our part-time students have work and family commitments, which may have an impact on their module options. The examples above are chosen to illustrate the advice given on preparation for part-time study, the study time required for a module, the availability of module options and the pathways followed.
over the period during which we have been developing material:

(i) Manual merging into a central master copy of a module. This is typically managed by a single technician, who is given responsibility for maintaining the master copy and collating changes and additions made by academics, with such changes being passed to the responsible technician via email or other electronic means. While this method can work in some situations—usually ones where only a single academic and technician are working together—it is very error prone and labour intensive, as it relies on both the technician and the academic keeping track of changes and communicating them correctly.

(ii) Many version control systems have been developed within the software development industry to address the problems of collaborative development. We have used several systems, including CVS [16] and Subversion [17], to support collaborative course development, effectively automating the manual merging process discussed above. A single central repository was used to hold the current master copy of the material, and collaborators worked on their own local copies, updating the central repository with their changes and keeping their local copies up to date with the central repository. This technique avoids many of the manual merging errors or omissions typically encountered when maintaining the material entirely manually, and if members of staff are well trained to frequently update both the central repository with their changes, and their local copies with changes made to the central repository by others, it works quite well. Unfortunately, our experience is that this seldom works as well as it could, with the often unintuitive behaviour of such version control systems being confusing and frustrating for staff, and an explosion of working copies, each containing slightly different versions of the material. Once such a situation arises, bringing the central repository and all working copies back to a synchronized state can be a very difficult task.

(iii) Centralized versioning inside a Content Management System (CMS) avoids the problems encountered in the previously discussed options. Many CMS systems are available, and a complete discussion of these systems is beyond
the scope of this article. The methodology we currently use involves a combination of open source software, including the MediaWiki wiki software [18] developed by the MediaWiki Foundation, and the Manchester Course Processor [19], a suite of tools designed and developed by Christopher Page at the University of Manchester. Technicians and staff collaboratively develop and maintain course material within a wiki, and then use the Manchester Course Processor tools to extract that material in forms suitable for distribution to students. The wiki ensures that all changes to the material are tracked, a single definitive version of the material is always maintained, and staff can see how their material will appear to the student as they work. The supporting tools allow staff to extract material developed in the wiki and convert it into self-contained, training packages through a simple web-based interface.

Whichever method is ultimately adopted, it is important to ensure that all staff members comply with it fully, otherwise changes may be lost or omitted from released module material.

DISCUSSION

There is general agreement within the community of the continuing need for training courses in bioinformatics. These courses are available in a range of different formats, from web pages suitable for unsupported learning to face-to-face courses designed for those who require focussed, monothematic, training.

Our aims in offering a co-ordinated programme are to provide flexible, accredited, online modules for those looking for focussed, supported, learning. This is a multi-layered approach, so there are issues to address at different levels.

Others have made recommendations on the design of training courses in computational biology [2, 4], and most of the recommendations also apply to the design of a training course to be delivered by distance learning. The tensions between research and teaching [20] are possibly more evident in the distance learning context because distance teaching does not fit within the established categories for university teaching. However, the role of personal tutor is well established in campus-based universities, and we have adapted the approach for our student cohorts.

Flexibility in access to learning resources is recognized as the key reason for a decision to study by distance learning. Our tutors understand this, and so organize tutorials at times convenient to the student group. However, flexible enrolment may not be supported by university systems that are designed for students studying full-time [12]. An important factor in the success of our programme has been the excellent technical and administrative support available for our LCMS, Moodle, and its associated systems. This support has made it possible to integrate our rather different student cohorts into the mainstream systems.

Our LCMS has also been important in providing a flexible platform for distance students, which allows us to provide access to both taught modules, of fixed duration, and open access to online support systems. We include general information on study skills, and also personal advice on the available routes through the programme. These underlying support systems also provide the links between our single topic modules, to help our students to develop a wider view of current work in computational biology. In this way we aim to help course participants to develop their research capabilities. By offering the programme by distance learning we make the learning and the qualifications accessible to active researchers, who cannot take time away from their full-time posts.

The importance of regular updating of course material is recognized within the community of computational biologists [3], but the necessary costs are not always fully accepted by university management. However, in a university, a course written for distance learning can also be used to support research staff and students on campus, providing opportunities to aggregate demand. Lentell [12] argues that recent re-organizations to universities in the UK have made the development of distance learning policy more difficult to achieve. With more responsibility devolved to Faculty level, the future may be in the development of specialist distance courses, which are also valuable for internal training.

Bioinformatics Education Online: http://octette.cs.man.ac.uk/bioinformatics/index.html
Key points

- Online courses are very suitable for supporting professional development.
- By linking online training courses into a full Masters programme, we can help participants to develop their research capacity.
- The development costs of online courses are high, but aggregation of demand can make specialist courses viable.
- Currency and quality of the courses can be assured by careful control of course versions.
- In a university with systems designed for full-time students, workarounds may be necessary to ‘fit’ distance learners into administrative frameworks designed for students on campus.

FUNDING

The development of some modules was supported by funds awarded through the Biotechnology and Biological Sciences Research Council’s Modular Training for Industry programme.

References