CAL-laborate
A collaborative publication on the use of Information Technology in tertiary teaching and learning for the life sciences

June 2003
Contents

Editors:

Anne Fernandez
UniServe Science
Carslaw Building (F07)
The University of Sydney
NSW 2006
Australia
PhySciCH@mail.usyd.edu.au

Trish Walker
LTSN Centre for Bioscience
School of Biochemistry and Molecular Biology
University of Leeds
LS2 9JT
United Kingdom
t.e.walker@leeds.ac.uk

Stacey Sorensen
Department of Synchrotron Radiation Research
University of Lund
S-221 00 Lund
Sweden
Stacey.sorensen@sljus.lu.se

The introduction of web-based formative assessment in medical science: An action research project ................................................................. 1
Robert B. Dow, Queensland University of Technology

Changes in use of technological methods of teaching and learning in undergraduate pharmacology in UK higher education .................... 4
Ian E. Hughes, University of Leeds

Simulation as a teaching alternative: Utopia or reality? ..................... 9
Anders Hyltander, Sahlgrenska University Hospital

Photosynthesis in Silico: A multimedia CD-ROM combining animations, simulations and self-paced modules for photosynthesis education at all tertiary levels .................................................................................. 13
Sharon A. Robinson, Wendy Russell and Andrew Netherwood,
University of Wollongong

Computer aided interactive histology – summary of 10 years .......... 17
Göran Sandberg, Karolinska Institutet

Teaching bioinformatics: A student-centred and problem based approach .................................................................................................. 25
Yun-Can Ai, Zhongshan (Sun Yet-Sen) University, Lars Jermiin and Neville Firth, The University of Sydney

This newsletter was brought to you by: ............................................. 31

CAL-laborate is published by:

UniServe Science
Carslaw Building (F07)
The University of Sydney
NSW 2006 Australia
ISSN: 1443-4482 (print version)
ISSN: 1443-4490 (online version)
http://science.uniserve.edu.au/
© 2003
Teaching bioinformatics: A student-centred and problem based approach

This article was first published in ‘The China Papers’ Issue 1, October 2002 and is reprinted here, with some modification, with permission from the authors.

Abstract

Bioinformatics is one of the fastest growing interdisciplinary sciences of the late 20th and early 21st century. For many students, bioinformatics is such a new discipline that they will not necessarily have the required background knowledge. It is therefore necessary to build bridges for students with diverse backgrounds. This paper describes teaching a bioinformatics course, explains what bridges needed to be built, and provides some examples of how the approaches of student-centred teaching and problem based learning could have been used in teaching this course using WebCT as the delivery tool. To understand the major concepts of bioinformatics, two conceptual frameworks are used: similarity, which enables analysis of predictions about structure/function; and dissimilarity, that allows inference of evolutionary history based on distance. These attributes are used as examples of core modules (the bridges), because all analyses need to be undertaken in appropriate biological context but involve multiple disciplines, including mathematics, statistics, computer science, and information technology, that need to be integrated into biology. The course focuses on three aspects: (i) to assist students with computing or related backgrounds understand the difficult core concepts of biology; (ii) to build critical understanding of the key mathematical and statistical principles that are crucial to bioinformatics; and (iii) to give all students hands-on training by allowing them to explore key issues using dominant bioinformatics platforms. Using these bridges, students should be able to develop the skills of creative experts rather than those of handy technicians.

Introduction

Bioinformatics is a scientific discipline that has emerged recently in response to accelerating demand for an expedient, flexible, and intelligent means of storing, managing and querying large and complex biological data sets. It is an interdisciplinary science with strong links between the life sciences and mathematics, statistics, and computer science, and it is regarded by a wide range of interest groups (governments, universities, and industry) as a crucial area of development – huge investments have been made worldwide, particularly in the USA and Europe. The ultimate goal of bioinformatics is to allow scientific exploration and exploitation of previously uncharted interdisciplinary territories. Stated differently, the aim is to enable the discovery of new biological insights and to create a global perspective, from which unifying principles in biology can be discerned (Mount 2001).

In the beginning of the genomics era, bioinformatics was mainly concerned with the creation and maintenance of databases to store digitised biological information, such as nucleotide and amino acid sequences. Development of these types of databases involved not only design issues, but also the development of complex interfaces whereby researchers could both access existing data, and submit new or revised data (e.g. to the NCBI, http://www.ncbi.nlm.nih.gov/). More recently, emphasis has shifted towards the questions of how to analyse large data sets, particularly those stored in different formats in different databases. Ultimately, however, integration is needed (e.g. Chicurel 2002) in order to form a comprehensive picture of normal cellular and sub-cellular activities, so that researchers may study how these activities are globally regulated. The actual process of analysing and interpreting digitised biological data is often referred to...
as computational biology. It is commonly recognized that sub-disciplines within bioinformatics and computational biology include: (i) the development and implementation of tools that enable efficient access, management and use of various types of digitised biological information; and (ii) the development of new mathematical theorems, statistical methods and algorithms to infer relationships among members of large data sets, locate genes within nucleotide sequence, and predict protein structure and/or function.

Teaching bioinformatics is an interesting and challenging task because it requires in-depth knowledge of different disciplinary areas that are all evolving at great speed – this blend of interdisciplinary expertise is rarely available from a single individual, so courses in bioinformatics often need to be planned and taught by academics, who have widely different formal training. This, in turn, poses not only challenges, but also opportunities for the development of novel teaching strategies (Ai 2002). In the present paper, we discuss our strategies for teaching bioinformatics by outlining core themes and concepts in bioinformatics, explaining what bridges we are trying to build, and providing some examples to show how the approaches of student-centred teaching and problem based learning can be used in teaching this new discipline.

Course description

Bioinformatics is the application of computers to the life sciences and in particular to genomics. This makes it possible to study biology at the genome-wide level. In teaching such a course, we aim to focus on the major applications, including data storage, retrieval, and analysis of biological information, rather than on the engineering of new bioinformatics applications. The broad aim of the course is that students will become extensively trained and develop hands-on skills in the use of bioinformatics technologies; and that they will understand and appreciate the enormous potential of bioinformatics and genomics in the contemporary life sciences. Expected outcomes include:

- an awareness of the breadth of bioinformatics resources and applications, including non-sequence-based biological information;
- skills and experience in the use of a core set of programs and databases for nucleotide and amino acid sequence analysis and phylogenetic reconstruction;
- a basic understanding of the theoretical foundation and underlying assumptions of the programs, and their relative strengths/limitations; and
- competence in the evaluation of output from the programs in appropriate biological context.

Curriculum

The details of the curriculum being taught at The University of Sydney, Australia, and at Zhongshan University, China, are shown in Table 1. There are four basic themes to be covered. Within themes 2 and 3, student-centred teaching and problem based learning are used to further student understanding of sequence alignment and phylogenetics.

<table>
<thead>
<tr>
<th>Core Themes</th>
<th>Main Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bioinformatics resources</td>
<td>The relationships among bioinformatics, genomics and proteomics</td>
</tr>
<tr>
<td></td>
<td>Computational sequence analysis (‘in silico biology’)</td>
</tr>
<tr>
<td></td>
<td>Laboratory tools</td>
</tr>
<tr>
<td></td>
<td>Databases I – sequence databases</td>
</tr>
<tr>
<td></td>
<td>Databases II – other biological databases</td>
</tr>
<tr>
<td>2. Sequence similarity and</td>
<td>Sequence comparison and alignment</td>
</tr>
<tr>
<td>structure/function inference</td>
<td>Database searching</td>
</tr>
<tr>
<td></td>
<td>Multiple sequence alignment, motifs and profiles</td>
</tr>
<tr>
<td></td>
<td>Signal detection</td>
</tr>
<tr>
<td></td>
<td>Analysis of deduced products</td>
</tr>
<tr>
<td>3. Sequence dissimilarity and</td>
<td>Structural inference</td>
</tr>
<tr>
<td>evolution</td>
<td>Introduction to phylogenetics</td>
</tr>
<tr>
<td></td>
<td>Simulations, multiple hits and compatibility plots</td>
</tr>
<tr>
<td></td>
<td>Phylogenetic inference using parsimony methods</td>
</tr>
<tr>
<td></td>
<td>Phylogenetic inference using distance-based methods</td>
</tr>
<tr>
<td>4. Integrating themes</td>
<td>Phylogenetic inference using maximum-likelihood methods</td>
</tr>
<tr>
<td></td>
<td>Introduction to functional genomics</td>
</tr>
</tbody>
</table>

Table 1. Bioinformatics course content

Two conceptual frameworks – similarity and dissimilarity

Genomics generates unprecedented amounts of data, whereas bioinformatics focuses on converting these data into knowledge. Obviously, data generation and analysis are inter-dependent. In the context of genomics, bioinformatics is often concerned with sequence analysis, which can usually be considered within two conceptual frameworks – similarity and dissimilarity. Analysis of similarity enables predictions about structure/function, whilst dissimilarity analysis allows inference of evolutionary history based on distance. Analyses involve disciplines including mathematics, statistics, computer science, and information technology, but need to be undertaken within the appropriate biological context. Bioinformatics represents a challenge for people coming from different disciplines – bridges are needed to help them communicate with each other.
Building bridges by using core modules

Bioinformatics requires the skills of a diverse group of specialists. Biologists want to be computational biologists, while informaticians want to be biological informaticians. They all need to be trained in theory as well as software development and applications. Core modules must cater for diverse backgrounds and offer the required bridges. We focus on three groups of students: (i) students with a computing or related background must be assisted in learning core concepts of biology; (ii) students lacking adequate mathematical/statistical expertise must be taught key principles about methods of relevance to bioinformatics; and (iii) all students must be given hands-on training in the use of predominant bioinformatics platforms. Using these educational bridges, students should become creative experts rather than simply handy technicians.

Similarity analysis module

The inference of structure and function based on biological sequence similarity is a fundamental concept in bioinformatics, and represents a good example of a bridging module. The Dotplot method is a particularly informative way to do similarity analysis between two sequences. However, although the method is conceptually simple, it can be difficult for non-biology students to understand the biological significance. Conversely biology students may be confused by the mechanisms used to reduce noise in the plot. In this case we bridge an abstract visualisation strategy directly to its biological basis to illustrate insertion/deletion/duplication, etc. This bridge provides a strong foundation so that all students can understand gene evolution. This is crucial if students are later to use or even develop sophisticated software to explore whole genome scale events.

Dissimilarity analysis module

Phylogenetics is based on the concept that taxonomic lineages diverge from one another over time and, therefore, that estimates of dissimilarity or distance between representatives of the taxonomic lineages may tell us how removed they are from one another in an evolutionary sense. Many computer programs developed to infer phylogenetic trees use a simple matrix, Q, which defines the rate of substitutions from one nucleotide to another (Figure 2).

However, the significance of this matrix is often difficult to understand for students who lack a strong mathematics background, and this can impact on their ability to use software correctly and effectively. Students with a strong mathematical background, on the other hand, will feel instantly familiar with the properties of this matrix but they may not be able to understand how it can be used to model the evolutionary process at the level of DNA. This example clearly illustrates why there is a need to build interdisciplinary bridges for students with limited interdisciplinary training and knowledge. By taking advantage of the mixed disciplinary background of a typical student cohort (e.g. biology, mathematics, statistics, computer science), it is possible to bridge the interdisciplinary gaps by asking groups of students to address problems that require knowledge from different disciplines. This has the added advantage that students are actively involved in the teaching/learning process and, in particular, that they learn to effectively communicate their knowledge to students with a different academic background. To underpin the complex knowledge that these students are acquiring, students are provided further online tutorial practice with simulations and constructions of phylogenetic trees based on their understanding of the rate matrix (Figure 2), which in turn promotes an understanding of the evolutionary history in terms of phylogenetics and the key principles and methods that are critical to bioinformatics.

Figure 1. Similarity analysis using the Dotplot method, showing visualization of genetic insertion/deletion/duplication events on a large scale
CAL-laborate, June 2003

$$Q = \begin{bmatrix}
-\sum_{y \neq x} q_{xy} & \mu a \pi_C & \mu b \pi_G & \mu c \pi_T \\
\mu g \pi_A & -\sum_{y \neq x} q_{xy} & \mu d \pi_G & \mu e \pi_T \\
\mu h \pi_A & \mu i \pi_C & -\sum_{y \neq x} q_{xy} & \mu f \pi_T \\
\mu j \pi_A & \mu k \pi_C & \mu l \pi_G & -\sum q_{xy}
\end{bmatrix}$$

Figure 2. Each element of this matrix, $q_{xy}$, represents the instantaneous rate of substitution from nucleotide $x$ to nucleotide $y$ during an infinitesimal amount of time ($dt$) ($x$, $y = A, C, G, T$) — for further details, see Swofford et al. (1996).

Online hands-on use of major Bioinformatics platforms

The third kind of bridge is the hands-on use of major online bioinformatics services, including national and international resources such as ANGIS BioManager (Australia), NCBI (USA), EMBL (Europe), ExPASy (Sweden), and S-Star.org (Multinational) (see Useful Links below). There are hundreds of programs available through BioManager. NCBI was one of the first online bioinformatics resources and is still a world leader. EMBL is a major European bioinformatics resource. ExPASy is the protein sequence analysis expert system, which contains a large number of modular programs. S-Star.org is a pioneering international bioinformatics education initiative, which involves institutions from around the world (Ping et al. 2002). Students are required to use these platforms in their own time, after receiving teacher-guided help in class during tutorial sessions.

Student-centred teaching by using WebCT

Bioinformatics is well suited to the development of student-centred teaching scenarios. WebCT has proven particularly useful in this regard since it caters for presentation of a diverse array of teaching materials, including links to useful bioinformatics resources, QuickTime ‘lecture movies’, practical self-assessment quizzes, assignments, etc. The materials are easy to use for the students and can be updated in a cost effective manner. Recorded ‘lecture movies’ are instantly replayable, making them particularly useful for revision and for foreign students learning the course materials in English. Online evaluation promotes active student learning. Figure 3a shows the Bioinformatics course WebCT site Home page and Figure 3b shows the Bioinformatics course WebCT Learning Resources page.

The learning resources page includes a link to ANGIS BioManager, a service that is widely used in Australia for bioinformatics research, and increasingly also for teaching; BioManager is used throughout the teaching and learning process of this course. Online practical and tutorial activities are provided in laboratory classes, where students are provided extensive training in the programs provided through BioManager (Figure 4). With respect to phylogenetics, emphasis is placed on building an understanding of molecular evolution by using programs developed in-house, such as hetero, zeta and TrExML (Jermiin 2003a, 2003b; Wolf et al. 2000). In these learning scenarios, teachers act as supporters and facilitators rather than hands-on demonstrators, so that students are challenged, stimulated and motivated to become experts in analysing real problems.

Figure 3a. WebCT site Home page for the Bioinformatics course at The University of Sydney
Online problem based learning

During the course, students are required to work in groups to explore and evaluate bioinformatics resources on the Internet, such as ExPASy and NCBI. Each student group is then required to give a seminar to the class, outlining the function, features and methods of the resource, and the group’s evaluation of its utility. This type of teamwork encourages students to work cooperatively, and it establishes a foundation for a data mining project that is undertaken individually. This task is intended to challenge the students to actively learn core skills and reinforce key bioinformatics concepts and generic skills, by requiring them to undertake and report an analysis of a real DNA sequence. Support for sequence analysis, academic writing and reference citation is provided in the form of an example journal manuscript (Apisiridej et al. 1997) that is available on the course WebCT site. These activities form the basis...
for group-wide and individual assessment of performance, and provide a rational basis for constructive feedback during the course. Several criteria are considered to promote, and allow assessment of, both academic and generic skills (see Table 2 for example).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Available Marks</th>
<th>Aspects Considered (marks allocated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical approach, interpretation and content</td>
<td>17</td>
<td>Appropriate selection and use of programs (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluation of results in appropriate biological context (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Demonstrated understanding of use of bioinformatics tools to gain biological insight (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adequacy of supporting information (e.g. parameters used) (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Citation of relevant references and bibliography (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Independence (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clarity and logic (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard (e.g. spelling, grammar, clarity of figures and tables) (4)</td>
</tr>
<tr>
<td>Presentation</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Assessment of the data mining project

Talented students can be further challenged by asking them to write a proposal for a project grant application and/or engage in scholarly debate in class. This provides extremely useful generic skills for students who may become academics in future.

Summary

We have outlined some of the strategies that we have used in teaching bioinformatics, an example of an emerging interdisciplinary science, so as to take advantage of a combination of contemporary teaching strategies and online information technology. The approaches of building bridges, student-centred teaching, and problem based learning have been used to successfully promote student’s active learning.

Acknowledgements

This paper is based on a presentation given in 2002 while attending a professional development program at The University of Sydney, and was critically reviewed as well as extensively edited for language usage by Associate Professor Mary Peat. The program Teaching Sciences in English: a professional development course for Chinese university science teachers is a collaborative project between the China Scholarship Council and The University of Sydney, coordinated by Associate Professor Mike King (Faculty of Education) and Associate Professor Mary Peat (Faculty of Science). Yun-Can Ai is a recipient of The National Universities Distinguished Teachers by the Chinese Ministry of Education and The Guangdong Provincial Outstanding Talents by the Guangdong Provincial Education Department. Our warmest thanks go to all of the students who provided valuable insight into the course.

Useful links

ANGIS (Australian National Genomic Information Service)

EMBL (European Molecular Biology Laboratory)
http://www.embl-heidelberg.de/

ExPASy (Expert Protein Analysis System)
http://www.expasy.org/

NCBI (National Center for Biotechnology Information)

S-Star.org (BioInformatics Education Alliance)
http://www.s-star.org/

References


Jermiin, L. S. (2003a) Hetero: program to simulate the evolution of nucleotide sequences on a tree, with the inclusion of some more realistic substitution models (ANSI C code available soon).


This newsletter was brought to you by:

UniServe Science

Whilst originally set-up through a Federal Government Grant in 1994, UniServe Science is now funded by The University of Sydney through the College of Sciences and Technology, and the Faculty of Science. UniServe Science caters for Biochemistry, Biology, Chemistry, Computer Science, Geography, Geology, Mathematics & Statistics, Physics and Psychology.

The activities of UniServe Science include:
- collecting information about teaching materials, finding out what is being produced in this country and overseas;
- disseminating all this information, by newsletters and electronic means;
- maintaining a web site which also includes information about discipline-specific teaching resources and links to other relevant sites;
- setting up other exchanges of information by means of the Internet;
- recruiting representatives from every science department in the country in order to establish a nationwide network of direct personal contacts;
- undertaking visits to other institutions and giving talks; and
- holding symposiums, forums, seminars and workshops.

In addition to these, UniServe Science offers some services to local secondary science educators.

http://science.uniserve.edu.au/

Council for Renewal of Higher Education

The Swedish Council for Renewal of Undergraduate Education was established by the Swedish Parliament on 1st July 1990, and became a permanent National Agency in 1993. Since July 1, 1995 the Council is an integral part of the National Agency for Higher Education. Since October 1999 postgraduate education is included in the Council’s responsibilities which has led to a change of name. The purpose of the Council for Renewal of Higher Education is to promote and support efforts to develop quality and renewal of higher education. The three main activities of the council are:
- to award grants for development activities;
- to collect and disseminate information on planned, current, and completed development activities of a fundamental and innovative nature concerning undergraduate education in Sweden and abroad; and
- to evaluate the development activities the Council has funded.

Some specific objectives of the Council are to:
- support the integration of environmental studies in Sweden undergraduate education;
- support changes in curricula and pedagogy in Engineering and Natural Science programs in order to recruit more female students to these programs; and
- support the use of IT in teacher training.

Currently funded projects include ‘Hypermedia and Communications for Active Learning’, ‘New Forms of Assessment in Mathematics and Computer Science’, ‘Interactive Distance Education in Multimedia’ and ‘Computer Assisted Education in Radiographic Techniques for Dental Students’.

http://www.hgur.se/
http://www.hsv.se/english/

Learning and Teaching Support Network

The LTSN has been established by the UK higher education funding bodies to promote high quality learning and teaching in all subject disciplines in higher education. The network supports the sharing of innovation and good practices in learning and teaching including the use, where appropriate, of communications and information technology.

The network consists of:
- 24 subject centres;
- a Generic Centre (GC) and a Technologies Centre; and
- a program director and coordinator.

Many of the new subject centres build on existing Computers in Teaching Initiative (CTI) Centres and involve learning and teaching support networks created by other funding council initiatives, such as the Fund for the Development of Teaching and Learning (FDTL). The new centres are the main points of contact within subject communities for information and advice on good practices and innovations in learning, teaching and assessment, and will provide support for the many networks which already exist.

The GC provides strategic advice to the sector on generic learning and teaching issues, disseminates good practice in the development and deployment of new methods and new technologies, and acts as a knowledge broker for innovation in learning and teaching.

http://www.ltsn.ac.uk/