IJODeL, Vol. 2, No. 2, (December 2016)

Identification of User-Friendly Bioinformatics Tools for Courses in Open and Distance Learning

Prasad Senadheera[1], Malinda Kumarasinghe[2], Pradeepa Perera[3], Jayantha Wattewidanage[4], T.K. Weerasinghe[5]

[1]Senior Lecturer in Botany, The Open University of Sri Lanka, Sri Lanka, spsen@ou.ac.lk
[2]Lecturer in Education Technology , The Open University of Sri Lanka, Sri Lanka kumarasinha@gmail.com
[3]Lecturer in Botany, The Open University of Sri Lanka, Sri Lanka, ppper@ou.ac.lk
[4]Director of Education Technology and Media center, The Open University of Sri Lanka, Sri Lanka, wvper@ou.ac.lk
[5]Senior Lecturer in Botany, The Open University of Sri Lanka, Sri Lanka, tkwee@ou.ac.lk

Abstract

Bioinformatics has become a vital offspring of biology with a wider spectrum of applications. Many fields of biology now rely on the analysis and interpretations of observations through the genomic scale data using bioinformatics. There is a considerable demand to learn bioinformatics, which can be easily offered in open and distance (e)learning (ODL/ODeL) systems due to its total reliance on information and computer technology (ICT). Generally, course design and content development for new courses in ODL demand substantial resources and time. Use of existing bioinformatics software tools and allied documents can drastically reduce these resource demands and time factors in course designing and updating. Given a large number of proliferating bioinformatics software tools available on the Internet, it requires comprehensive criteria to identify software tools and documentation that are user-friendly for open and distance learners, and needs multidisciplinary knowledge in ODL, Biology, and Bioinformatics. A criteria-based software evaluation was adopted in this study to evaluate the user-friendliness of bioinformatics software, particularly the plant genomic analysis software which was available in the Internet for free usage. Two main criteria used in this selection: Usability and Maintainability. These main criteria consisted of several sub-criteria such as understandability, documentation, installability, learnability, etc. A wider spectrum of Bioinformatics software tools designed for different stages of genomic data analysis, from simple sequence analysis to genome wide analysis, were used in this evaluation process. Based on the criteria, a series of user-friendly bioinformatics tools that can be used at different stages of genomic data analysis were identified. In addition to the criteria of selection, these tools and allied documentation can effectively be integrated with development of a comprehensive ODL-based bioinformatics courses. Further, this outcome facilitates offering hard science concepts through the minimum commitment of resources and time to meet the escalating demand for capacity development in Bioinformatics.

Keywords: Bioinformatics, ODL, ODeL, user-friendly, genomic data, software tools, data analysis

Introduction

Advances in computation technologies and high throughput experimental data analysis methods have created a deluge of biological data, which eventually shaped a new multidisciplinary field of science, called Bioinformatics. Biology and computer sciences are the key fields merged in Bioinformatics, but it requires substantial inputs from the fields of mathematics, statistics, information technology, physics, and chemistry as well. This relatively new field of science has unveiled exciting new potential for advanced research in biological systems. Many fields of biology now rely on analysis and interpretations of observations using genomic scale data of DNA, RNA, protein, and metabolites. Curating, organizing, analyzing, and making this large volume of biological data available to public usage requires intensive computational infrastructure and knowledge. Applications of bioinformatics are varied from fundamental biology experiments to advance medical sciences. Bioinformatics analysis is as complex and diverse as the biology of life.

The field of bioinformatics is expanding at an enormous rate and is playing an increasingly central role (Brooksbank, Cameron, & Thornton, 2010). The bioinformatics infrastructure consists of a set of publicly available databases, software tools to analyze data, and integrated bioinformatics platforms (IBP). These databases are specialized according to the type of data (DNA, Protein sequence and structure, phylogenetic, etc.) they curate and so are the software tools and integration platforms for various analysis requirements. The National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EMBL-EBI) are the most versatile integrated bioinformatics platforms (IBP) that offer excellent teaching and learning material. They have integrated an array of software tools to mine data from literature to genomic and proteomic information. A list of bioinformatics databases can be accessed at Nucleic Acid Research Database Summary Paper Category List (https://www.oxfordjournals.org/our_journals/nar/database/cat/1), and a list of software tools are available at the Expasy bioinformatics resource portal (https://www.expasy.org). Most of these listed tools are freely available online.

Bioinformatics courses were initially available for the graduate courses. Due to increasing applicability of Bioinformatics over a wider spectrum of biology, it has been introduced to undergraduate education in conventional university systems over the last couple of decades (Counsell, 2003; Ditty et al., 2010). The interdisciplinary nature and computationally intensiveness of bioinformatics pose challenges in adopting it into undergraduate learning, in general (Koch & Fuellen, 2008). There is also a wider set of efforts geared to formulate instructional curriculum from multidisciplinary teams to integrate concepts of bioinformatics into the undergraduate teaching (Dinsdale et al., 2015).

Open and Distance Learning (ODL) is the most favored system of learning at present due to many reasons (NCES, 2015). Teaching science subjects in ODL has always been a challenging task due to their practical orientation that demands access to substantial infrastructure. Unlike most of the life science subjects, bioinformatics needs limited physical resources such as access to a computer and the Internet. There is a significant complementarity between Bioinformatics and the ODL and dependence on technology assisted learning with many open source online tools and learning material. Moreover, learning outcomes of Bioinformatics are often customized according to the application domain which can be achieved in the flexible learning characteristics of ODL.

The introduction of ODL courses demands considerable resources and time, as it requires course design, content development, and delivery. Updating such courses also need considerable engagement in comparison to the conventional systems. However, the inclusion of bioinformatics software tools and allied documentation in course contents of the ODL can significantly reduce the resources and time spent on course development. Further, it makes these courses more amenable to include rapid changes in the field of bioinformatics. However, the most challenging task is to select the compatible software tools, which are conducive to include in ODL. However, there are no recorded criteria to evaluate these bioinformatics software tools for their compatibility for ODL. General software evaluation criteria contain some key parameters that can be adopted in developing criteria for evaluating compatibility for ODL.

Software tools of bioinformatics are numerous and diverse according to their application. This article evaluated a selected set of bioinformatics software tools, which are used in the main steps of bioinformatics analysis for their compatibility to adopting to undergraduate ODL programs. Of the hundreds of Bioinformatics software tools available for microbiological, animal, human, and plant subjects, particular attention was given to those specialized for plants due to the author's expertise in Plant Sciences.

Objective

This study aimed to develop evaluation criteria for bioinformatics software to be included as ODL material. Based on these criteria, series of bioinformatics software tools that could be directly integrated with course development as learning material was identified.

Methodology

Bioinformatics software tools that are available on the Internet (particularly listed in https://www.expasy.org/resources) and cited commonly in published literature were chosen for the analysis. Authors experience in using bioinformatics tools for plant research and designing ODL courses in bioinformatics was useful in shortlisting a large number of tools. Commonly cited software tools in the published research articles were first chosen for the study. Besides, a selected group of bioinformatics software tools listed in https://www.expasy.org/resources was also studied. This study, conducted in September 2016, included more than 40 software tools and IBP. The analysis criteria mainly consist of usability and maintainability guidelines in Table 1 adopted from the Software Evaluation Criteria (Jackson, Crouch, & Baxter, 2011). A score of 1 was given for every compliance to each criterion and 0 was given when there was no compliance. Total percentage of the score was used in determining the user friendliness for ODL.

Table 1. The Criteria adopted to evaluate the user-friendliness of Bioinformatics software tools for inclusion in ODL Programs on Bioinformatics.

	Usability
Understandability	**Easily understood?** What the software does and its purpose? The software's basic functions? The software's advanced functions? High-level description of how the software works are available. Case studies of use are available. Step by step input guide available in GUI
Documentation	**Comprehensive, appropriate, well-structured user documentation?** Completeness of the documentation in describing the functionality? Clarity of descriptions? Lists resources for further information. Is task-oriented. Consists of clear, step-by-step instructions. Gives examples of what the user can see at each step (e.g. screen shots or command-line excerpts.) Documentation on the project website makes it clear what version of the software the documentation applies to.

Learnability	**Easy to learn how to use its functions?** How straightforward is it to learn Basic functional tasks? Tool tip indicators present A getting started guide is provided outlining a basic example of using the software. Instructions are provided for many basic use cases. Instructions are provided supporting all use cases. Audio visual guide is available.
Installability	**Straightforward to install on a supported system?** Available online only. Install the software onto a target platform? Configure the software following installation for use? The website has instructions for installing the software.
Copyright/Licensing	**Adoption of appropriate license?** Has an appropriate license been adopted? The software has an open source license.
Community	**Evidence of current/future community?** To what extent does/will an active user community exist for this product? The website has list of important partners or collaborators. Users are requested to cite the project if publishing papers based on results derived from the software. Email newsgroup. Frequently ask questions (FAQ).
Maintainability	
Supportability	**Evidence of current/future developer support?** The website has a page describing how to get support. User doc has a page describing how to get support. The project has an e-mail address. E-mail archives are publicly readable. The website has a search facility. Project resources are hosted in an organization/institution/ Bioinformatics resource platform.
Evolvability	**Evidence of current/future development?** The website describes project roadmap or plans or milestones (either on a web page or within a ticketing system). The website describes how the project is funded/sustained.
Interoperability	**Interoperable with other required/related software?** Uses open standards.
Portability	**Usable on multiple platforms?** User-friendliness of the software can be used on other platforms? The application can be built on and run under MacOSX. The application can be built on and run under Windows.

Results and Discussions

Open and distance learning programs, in general, require considerable time and effort in adopting the contents to suit the ODL learner. This process of course development sometimes becomes a bottleneck for the development of ODL programs and is particularly challenging for the rapidly developing courses in science that need to be continuously updated. There are many methods adopted to successfully overcome this challenge, such as by integrating open source courses and deviating from conventional print-based media to online content, etc. This study presents two aspects to meet the challenge in developing bioinformatics courses. Evaluating criteria can be employed to select the course content while a set of bioinformatics software tools used for allied documentation can be directly integrated as course content in ODL programs. All these recommended courses are freely available online. There are only a few studies done on these aspects of integrating bioinformatics software tools in ODL systems (Ditty et al., 2010; Dinsdale et al., 2015). The software evaluation criteria adopted in this study target the undergraduate bioinformatics courses in ODL but can also be extended to graduate courses as well.

The score assigned in Table 2 is proportional to the user-friendliness of ODL learning. This score was computed by deriving the percentage obtained from a total of 42 criteria used in the evaluation. Bioinformatics software tools were categorized according to the stage of analysis in which they are designed for. With increasing complexity, sequence similarity search, multiple sequence comparison, phylogenic and genomic analyses, and proteomics are the main stages of categorization according to work. Besides those, software tools and few general IBP were also listed for their versatile functionality. These IBP provide open-sourced learning material designed for self-learning.

Usability criteria mainly consist of three sections, namely: understandability, documentation, and learnability. The presence of well-structured information on functionality, variables, and results is very important, particularly for the ODL user, to understand the functionality of the tool. Almost all software evaluated provided an indication of its function but only 60% of the software had user manuals. Availability of data for case study significantly supports students' self-learning of the software tool. Citing resources for further study is equally important for the learner to get comprehensive knowledge and experience. Almost all software carried description on its functions, but only 60% of them had step-by-step guidelines for users, while 44% provided descriptions with screenshots. NCBI, EBI sequence search, Phylip, MEGA, uniprot, ugene provided step-by-step guidelines. Nearly 15 % of software tools which were mainly IBPs had supplemented audio-visual resources with tutorials.

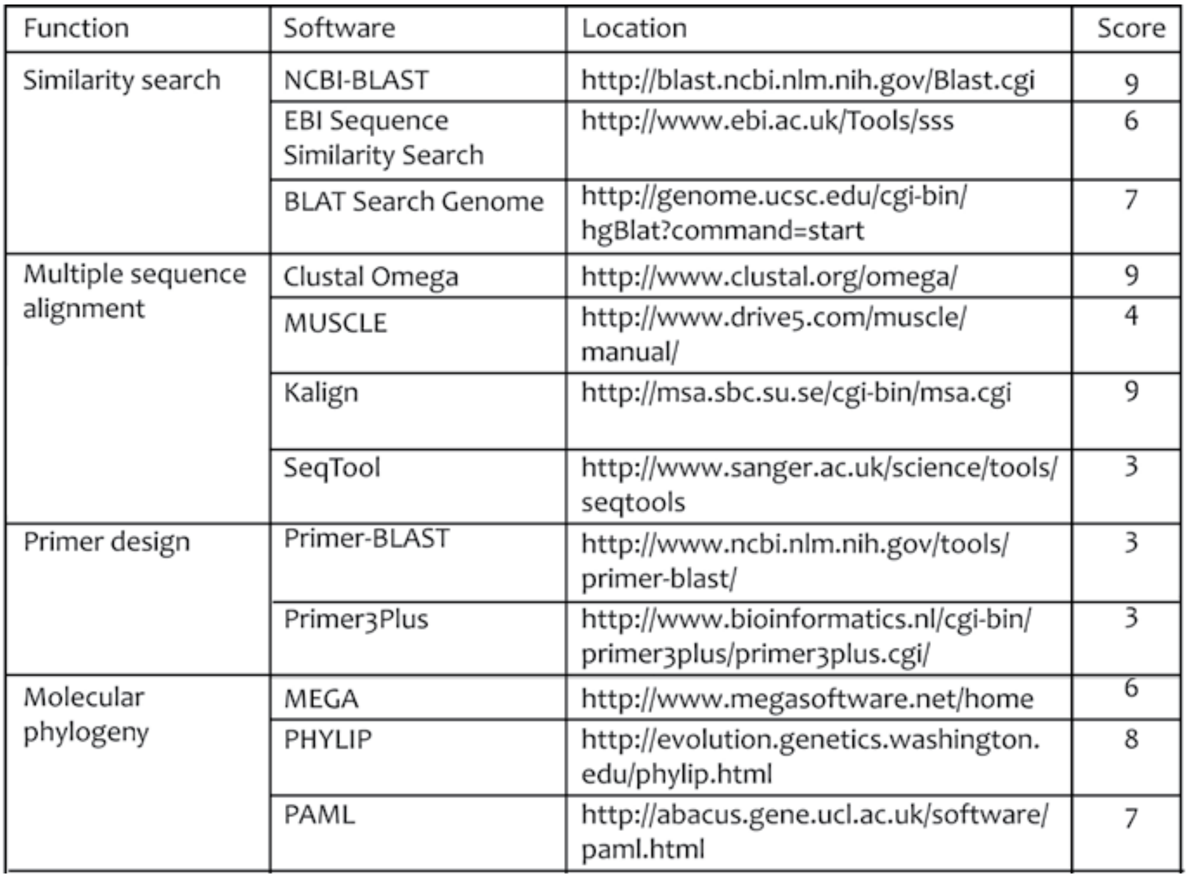
Open and distance learners rely heavily on peer assisted-learning (Bandarage, et al., 2015). Therefore having a space for the users and incorporating a link for frequently ask questions (FAQ) in the website which is associated with the software tool can enhance learning and problem-solving skills. Only 40% and 42% of the software tools have space for users and link to FAQ, respectively. NCBI-blast, MUSCLE, Primer-blast and MEGA are examples of tools which have an active user community. Only 12% of the software, including GENE and NCBI-blast, had email groups.

Maintainability of the software indicates if it can easily be used in the personal computer (PC) environment and whether there is continuous support or updating. The rapid development of data analysis techniques in Bioinformatics warrants continuous support by the software developers to update the tools and user awareness equally. More than 75% of the software tools and IBP provide updated versions routinely and are compatible with a variety of PC software platforms. Only 40%

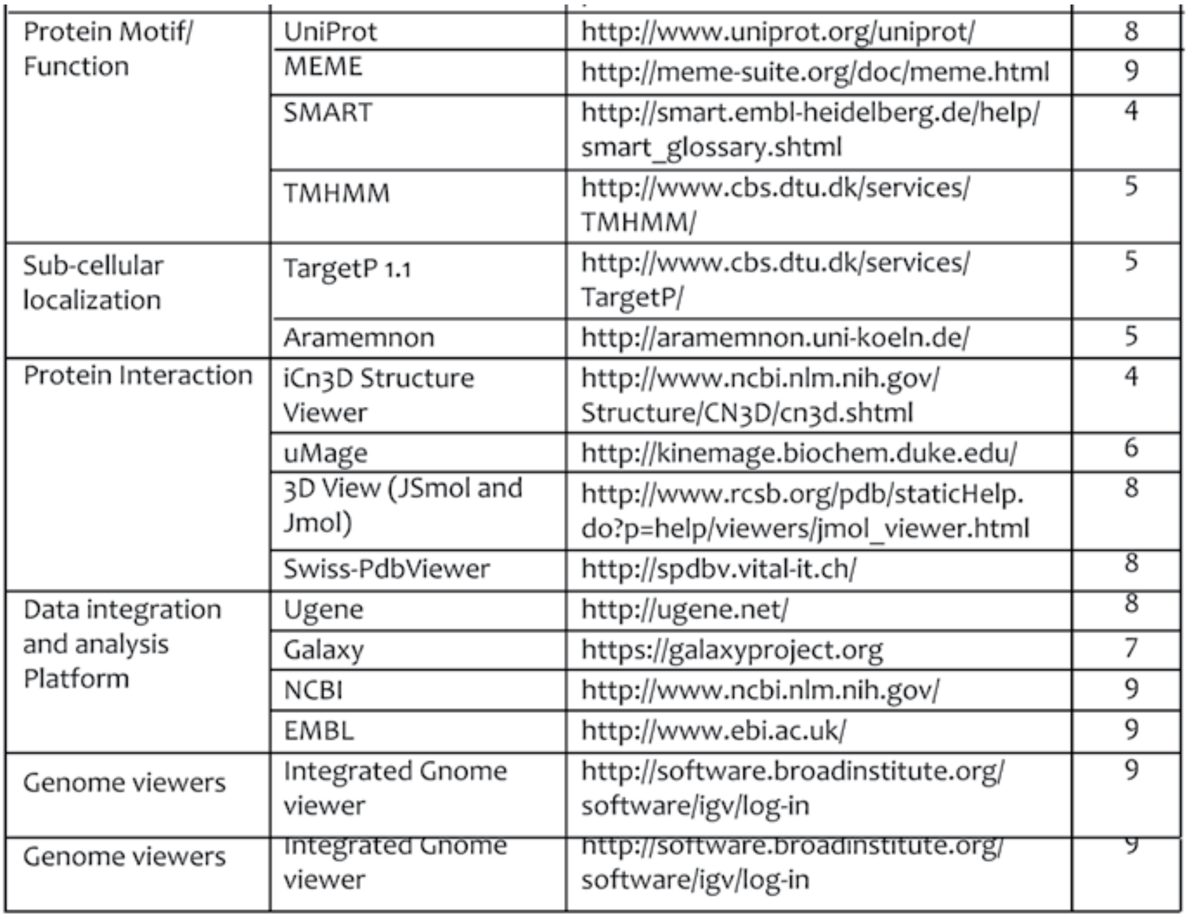
and 8% of software tools studied had FAQ and publicly readable email archives, respectively. More than 80% of software tools meet the evolvability and portability criteria. It is noteworthy that most free Bioinformatics tools which are particularly designed for handling large volumes of data are inclined to LINUX-based systems. As such, it is recommended for the aspiring bioinformatician to be familiar with LINUX systems.

Nevertheless, the lack of descriptive user documents or low score (Table 2) does not undermine the software tool. Some of the highly used software tools (e.g., Aramemnon and TMHMM) have user manuals with few information, as they are intended for experienced users. But novel users tend to learn about them through intuition, peer interaction, or studying use-cases published elsewhere. There are many technical social media forums which can be used to share and enrich the learning experience. In the case of software tools that have a low score, ODL facilitators also require providing more learning support to make them more conducive for the ODL.

Table 2. Bioinformatics software tools and Integrated Bioinformatics platforms for Open and distance learning courses. Score indicates user-friendliness in adopting into learning material for ODL



Function	Software	Location	Score
Similarity search	NCBI-BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi	9
	EBI Sequence Similarity Search	http://www.ebi.ac.uk/Tools/sss	6
	BLAT Search Genome	http://genome.ucsc.edu/cgi-bin/hgBlat?command=start	7
Multiple sequence alignment	Clustal Omega	http://www.clustal.org/omega/	9
	MUSCLE	http://www.drive5.com/muscle/manual/	4
	Kalign	http://msa.sbc.su.se/cgi-bin/msa.cgi	9
	SeqTool	http://www.sanger.ac.uk/science/tools/seqtools	3
Primer design	Primer-BLAST	http://www.ncbi.nlm.nih.gov/tools/primer-blast/	3
	Primer3Plus	http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/	3
Molecular phylogeny	MEGA	http://www.megasoftware.net/home	6
	PHYLIP	http://evolution.genetics.washington.edu/phylip.html	8
	PAML	http://abacus.gene.ucl.ac.uk/software/paml.html	7



Protein Motif/ Function	UniProt	http://www.uniprot.org/uniprot/	8
	MEME	http://meme-suite.org/doc/meme.html	9
	SMART	http://smart.embl-heidelberg.de/help/ smart_glossary.shtml	4
	TMHMM	http://www.cbs.dtu.dk/services/ TMHMM/	5
Sub-cellular localization	TargetP 1.1	http://www.cbs.dtu.dk/services/ TargetP/	5
	Aramemnon	http://aramemnon.uni-koeln.de/	5
Protein Interaction	iCn3D Structure Viewer	http://www.ncbi.nlm.nih.gov/ Structure/CN3D/cn3d.shtml	4
	uMage	http://kinemage.biochem.duke.edu/	6
	3D View (JSmol and Jmol)	http://www.rcsb.org/pdb/staticHelp. do?p=help/viewers/jmol_viewer.html	8
	Swiss-PdbViewer	http://spdbv.vital-it.ch/	8
Data integration and analysis Platform	Ugene	http://ugene.net/	8
	Galaxy	https://galaxyproject.org	7
	NCBI	http://www.ncbi.nlm.nih.gov/	9
	EMBL	http://www.ebi.ac.uk/	9
Genome viewers	Integrated Gnome viewer	http://software.broadinstitute.org/ software/igv/log-in	9

Genome viewers | Integrated Gnome viewer | http://software.broadinstitute.org/ software/igv/log-in | 9

Use of resources in IBP is more favorable for learning as they provide comprehensive documents for the usage of software tools. Many of these sites harbour tutorials, including audio-visual guidance that is favorable for ODL. It is also important to count the data usage in using these online learning resources. Open and distance learners mostly rely on their network connectivity at a cost. IBP, such as Ugene that can be installed on personal computers, may save the cost of data usage but may also require Internet connectivity to run some of the analyses.

With a seemingly endless stream of biological data being generated across sectors of life science, there is an escalating demand for professionals who are competent in the analysis of these data (Levine, 2014). Given the extensive reliance of both ODL and bioinformatics on information and communication platforms, there is a need for ODL institutes to offer academic programs in Bioinformatics to meet this demand. The flexibility and openness of ODL will facilitate those who missed the opportunity to learn Bioinformatics in secondary and undergraduate education or those mid-career developers who need to develop their knowledge and skills in Bioinformatics while working.

Conclusions and Recommendations

To be current in the rapidly changing field of bioinformatics requires continuous updating of bioinformatics programs. Effective integration of software tools and allied open learning resources facilitate Bioinformatics ODL courses to be more responsive to changes, easy to update, and conducive to lifelong learning. Software evaluation criteria provide a foundation to screen bioinformatics software for their ODL-friendliness. Software tools with a score higher than six may directly be integrated into ODL courses. Rest of the tools with score less than six can be integrated into ODL, provided that they need more effort and time in adopting into ODL requirements. On the whole, the outcome of this study presents an approach to offer courses based on hard science concepts in ODL with a minimum commitment of resources and time to meet the escalating demand.

References

Bandarage, et al. (2015). Annual academic sessions -2015. Open University of Sri Lanka.

Brooksbank, C., Cameron, G., & Thornton, J. (2010).*The European Bioinformatics Institute's data resources*. Nucleic Acids Res 38suppl 1: D17–25doi: 10.1093/nar/gkp986.

Counsell, D. (2003). *A review of bioinformatics education in the UK*. Brief Bioinform 4: 7–21. doi: 10.1093/bib/4.1.7.

Dinsdale, E., Elgin, S.C.R., Grandgenett, N., Morgan, W., Rosenwald, A., Tapprich, W. & Triplett, E.W. (2015). 'NIBLSE: A network for integrating Bioinformatics into life sciences education', *CBE-Life Sciences Education*, 14(4), p. 3. doi: 10.1187/cbe.15-06-0123.

Ditty, J.L., Kvaal, C.A., Goodner, B., Freyermuth, S.K., &Bailey, C., et al. (2010). *Incorporating genomics and bioinformatics across the life sciences curriculum*. PLoSBiol 8: e1000448. doi:10.1371/ journal.pbio.1000448.

Jackson, M., Crouch, S., & Baxter, R. (2011). *Software evaluation: Criteria-based assessment.* Software Sustainability Institute. https://www.software.ac.uk/

Koch, I. & Fuellen, G. (2008). *A review of bioinformatics education in Germany.* Brief Bioinform 2008;9:232–42.

Kreeger, K. Y. (2000). Hot careers – A look at what's in demand in life sciences. *The scientist,* 14 (9), p. 29

Levine, A. G. (2014). An explosion of bioinformatics careers. *Science.* Retrieved from https://www.sciencemag.org/careers/features/2014/06/explosion-bioinformatics-careers.

NCES. (2015). Digest of education statistics 2014. Table 311.22. *Number and percentage of undergraduate students taking distance education or online classes and degree programs, by selected characteristics: Selected years, 2003–04 through 2011–12.* Retrieved from http://nces.ed.gov/programs/digest/d14/tables/dt14_311.22.asp?current=yes.