Education resources of the National Center for Biotechnology Information

Peter S. Cooper, Dawn Lipshultz, Wayne T. Matten, Scott D. McGinnis, Steven Pechous, Monica L. Romiti, Tao Tao, Majda Valjavec-Gratian and Eric W. Sayers

Abstract

The National Center for Biotechnology Information (NCBI) hosts 39 literature and molecular biology databases containing almost half a billion records. As the complexity of these data and associated resources and tools continues to expand, so does the need for educational resources to help investigators, clinicians, information special- ists and the general public make use of the wealth of public data available at the NCBI. This review describes the educational resources available at NCBI via the NCBI Education page (www.ncbi.nlm.nih.gov/Education/). These re- sources include materials designed for new users, such as About NCBI and the NCBI Guide, as well as documenta- tion, Frequently Asked Questions (FAQs) and writings on the NCBI Bookshelf such as the NCBI Help Manual and the NCBI Handbook. NCBI also provides teaching materials such as tutorials, problem sets and educational tools such as the Amino Acid Explorer, PSSM Viewer and Ebot. NCBI also offers training programs including the Discovery Workshops, webinars and tutorials at conferences. To help users keep up-to-date, NCBI produces the online NCBI News and offers RSS feeds and mailing lists, along with a presence on Facebook, Twitter and YouTube.

Keywords: *Bioinformatics; education; tutorials; NCBI; databases; GenBank*

INTRODUCTION

The National Center for Biotechnology Information (NCBI) is one of the world's major information hubs for biomedical data [1]. The NCBI hosts 39 molecu- lar and literature databases including the PubMed biomedical literature database and the GenBank nu- cleotide sequence database [2]. These databases are integrated in the Entrez search and retrieval system. NCBI also offers 3D structural (VAST) and sequence similarity search (BLAST) services. All are accessible through the NCBI web site (www.ncbi.nlm.nih .gov), making the NCBI one of the busiest US Government sites on the World Wide Web. NCBI also produces standalone software and databases that

Corresponding author. Peter S. Cooper, NCBI/NLM/NIH, 45 Center Drive, Bethesda, MD 20892, USA. Tel: +1–301–435–5951; Fax: +1–301–594–5166; E-mail: cooper@ncbi.nlm.nih.gov

Peter Cooper is an NIH Staff Scientist who provides general user support, designs and presents workshops and webinars, edits the NCBI News, and is involved in several web design efforts.

Dawn Lipshultz is a Technical Information Specialist with the Lockheed Martin Corporation, under contract to the NCBI, who writes the NCBI News, maintains the NCBI homepage and manages the NCBI presence on Facebook, YouTube, and Twitter.

Wayne Matten is a Senior Scientist with Consolidated Safety Services, Inc., under contract to the NCBI, who provides general user support and creates and maintains video tutorials and help documentation.

Scott McGinnis is an NIH Staff Scientist who oversees the BLAST software technical support and documentation, monitors NCBI web servers, and performs studies in web analytics.

Steven Pechous is a Senior Bioinformatics Scientist with the KEVRIC Corporation, under contract to the NCBI, who provides general user support, presents workshops, develops educational materials, and performs studies in web analytics.

Monica Romiti is the on-site manager for the KEVRIC Corporation, under contract to the NCBI, who provides general user support, GenBank sequence submission and PubMed Central help and is involved in writing help documentation.

Tao Tao is an NIH Staff Scientist who provides general user support, software testing and technical documentation, manages con- ference exhibits, and performs studies in web analytics.

Majda Valjavec-Gratian is a Senior Bioinformatics Scientist with the KEVRIC Corporation, under contract to the NCBI, who provides general user support, GenBank sequence submission and PubMed Central help, and is involved in teaching and preparation of educational material.

Eric Sayers is an NIH Staff Scientist who supervises user support, outreach, education and web analytics.

Table 1: Web addresses referred to in this review

NCBI Guide and Education pages	
NCBI Homepage (Site Guide)	/guide/
Resources List	/guide/all/
How-to guides	/guide/all/howto/
Education page	/Education/
Bookshelf	
Help Manual	/bookshelf/br.fcgi?book=helpcollect
Handbook	/bookshelf/br.fcgi?book=handbook
Comparative Genomics:	
Chapter 9: BLAST Quickstart	
Chapter 10: PSI-BLAST Tutorial	/bookshelf/br.fcgi?book=comgen
Chapter 10: PSI-BLAST Tutorial	
Chapter 24: Identification of Disease Genes	
Short Courses	/bookshelf/br.fcgi?book=coursework
NCBI Newsletter	/bookshelf/br.fcgi?book=newsncbi
Community	
Facebook	www.facebook.com/ncbi.nlm
Twitter	twitter.com/ncbi/
YouTube Channel	www.youtube.com/ncbinlm

All NCBI pages described here begin with the base address, www.ncbi.nlm.nih.gov. Appending the portion in the right column to this base will retrieve the page identified on the left.

are frequently downloaded and widely used at other sites. Anyone who works with biomedical or biomolecular data will sooner or later visit the NCBI site or use an NCBI product. Knowledge of what data and analyses are available at the NCBI and how to use these effectively are essential for biologists of all kinds. Moreover, keeping up with the rapidly changing data landscape and the corresponding changes in interfaces and tools is a major challenge facing both consumers and producers of bioinformatics services. These two factors have generated an increasing demand for information about NCBI products and services. This review focuses on educational resources produced by the NCBI that are designed specifically to aid users of our products. After a brief look at the new design of the NCBI homepage with its greater emphasis on serving as a site guide, this review will survey the wide range of available educational materials and products. Table 1 presents a list of direct links to Web pages for resources presented in this review.

THE NCBI GUIDE AND THE BOOKSHELF
The NCBI guide

For the past few years the NCBI web site has been changing to make the site easier to use and to expose related resources as part of the ongoing NCBI Discovery Initiative. Many powerful features are now more obvious and more self-explanatory. As part of this effort, the NCBI homepage has been completely redesigned as the NCBI Guide with a listing of all NCBI resources classified by topic alongside sets of directions for many common tasks. The NCBI Guide is an educational resource on its own and serves as a gateway to other materials. This Guide is designed to provide rapid access to major areas of the NCBI web site and to provide help and guidance for selecting the most appropriate databases, tools and other resources for the task at hand. The central section of the NCBI Guide provides a list of 14 categories of NCBI resources: Literature; DNA & RNA; Proteins; Sequence Analysis; Genes & Expression; Genomes & Maps; Domains & Structures; Genetics & Medicine; Taxonomy; Data & Software; Training & Tutorials; Homology; Small Molecules; and Variation. Each one of these categories expands to a list of relevant Resources grouped into Databases, Downloads, Submissions and Tools where appropriate. Associated with each Resource group on a separate tab are step-by-step How-to guides. These How-to guides provide recipes for common tasks associated with a particular set of resources. The How-to guides are short instruction sets for accomplishing common tasks in each of the Resources in the NCBI Guide. Typically, each How-to offers several ways to reach the goal depending on the nature of the starting point. Each path has few steps and

Figure 1: A How-to guide from the NCBI Guide showing how to view the 3D structure of a protein.

is designed to rapidly reach the goal. These How-to guides are distinct from the tutorials described below as they do not provide specific examples and are intended to serve mainly as navigational aids. A sample How-to, the instructions for viewing a 3D structure of a protein, is shown in Figure 1.

The Training & Tutorials category of the NCBI Guide is one access point for all other NCBI educational pages and resources. This category expands to an organized list of databases, analysis tools and downloads. Each item in this list has a brief description and a main heading link that leads directly to the relevant page. NCBI educational resources that are particularly relevant, new or otherwise important are bolded in this list and populate the Quick Links list in the upper right-hand area of the page for easy access.

The NCBI Bookshelf

Many of the resources listed in the Training & Tutorials category of the NCBI Guide point to items maintained and displayed in the NCBI Bookshelf, an online library of full-text books of various kinds made available through agreements initiated by publishers, who also determine the specific edition available at NCBI. Many of these books are on-line versions of standard biomedical textbooks on cell biology, molecular biology and biochemistry, making the Bookshelf a valuable educational resource in its own right. In addition to holding these textbooks with background information, the Bookshelf is increasingly the repository for the

NCBI documentation, tutorials, course materials and newsletters discussed later in this review. The Bookshelf collection may be browsed from the following page: www.ncbi.nlm.nih.gov/books/.

THE NCBI EDUCATION PAGE

The following sections of this review will focus on six categories of educational materials and services for NCBI users: Getting Started, Documentation, Teaching Resources, Courses and Workshops, News and Updates and Community. These six categories also appear on the NCBI Education page, depicted in Figure 2, which serves as the principal starting point for exploring NCBI educational materials.

Getting Started

The Getting Started section contains introductory material designed to orient new visitors to the NCBI site. 'About NCBI' is particularly helpful as an introduction to the NCBI site for new users and provides general information about the history, mission and organization of NCBI. In addition, 'About NCBI' offers descriptions of the science behind much of the data at NCBI, including introductions to the specialized techniques that generate modern sequence, genome, variation and expression data, all written for people unfamiliar with these methods. Other sections of 'About NCBI' explore the nature of whole-genome assemblies and associated data for human and
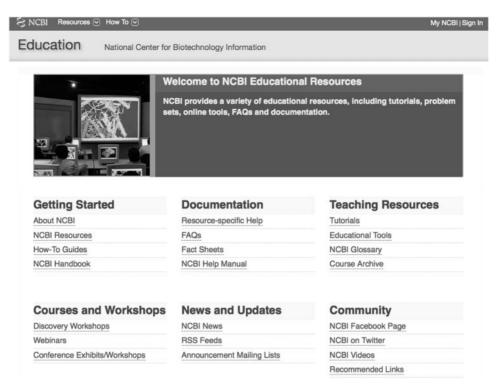
Figure 2: The NCBI Education page is the main portal for access to NCBI educational resources.

model organism genomes, focusing on understanding how genome data explain human biology and the origins of genetic diseases. Also included in the Getting Started section are the NCBI Resource list and How-to guides from the NCBI Guide.

The last major resource under Getting Started is The NCBI Handbook. This NCBI textbook provides detailed background information on most of the NCBI databases and tools—how they are constructed, what data are in them and how to search them. The book provides a valuable introduction to the NCBI site along with its architecture and culture. The book is organized into three main parts: The Databases, Dataflow and Processing, and Querying and Linking the Data. The Handbook also has a recently updated NCBI glossary that is extensively linked from many parts of the web site and offers definitions of NCBI-specific terminology as well as some general bioinformatics, molecular biology and biochemical terms.

Documentation

The Documentation section collects basic help documents that provide detailed information about the purpose, function and user-adjustable features of a particular tool or service, as well as practical aspects of its use and special techniques (tips and tricks). NCBI Documentation linked to the Education page encompasses the Resource-specific Help pages and Frequently Asked Questions (FAQs) associated with the various databases and tools as well as NCBI Fact Sheets on these resources.

The Resource-specific Help and FAQ pages gather together in one place all of the available documentation and FAQs that are also linked to the homepages for the specific resources. For example, help documentation for Entrez is closely associated with the database homepages; either linked on the blue sidebar in database homepages with older styles (e.g. www.ncbi.nlm.nih.gov/gene/) or, in pages with newer styles (e.g. www.ncbi.nlm.nih.gov/pubmed/), in the first column of links with information about using the resource. All resource homepages at NCBI are in the process of moving to this newer style, which includes the NCBI search bar and footer area that is also present on the Guide page. The new footer provides rapid navigation to all major areas of the NCBI site and easier navigation to general help documents such as the NCBI Handbook and Help Manual collection by listing them in the 'Getting Started' column on the left-hand side.

NCBI Fact Sheets are two- or four-page full-color handouts available in PDF format that describe a specific resource and highlight important features and

aspects of its use. The growing collection of fact sheets now includes documents for BLAST, Align 2 Sequences, Primer-BLAST, Books, the Conserved Domain Database, Epigenomics, GenBank, Gene, Map Viewer, the new PubMed design and Variation data (dbSNP). These useful introductions to specific NCBI resources are part of the materials available at the NCBI exhibit booth at scientific meetings described below in the section on Courses and Workshops. These documents can also be downloaded and printed out from the Fact Sheets page linked to Documentation.

The repository and source for much of the resource-specific help is the NCBI Help Manual collection on the Bookshelf, which is also linked separately under Documentation. The NCBI Help Manual collection is a growing set of online books that serve as the major documentation for a particular resource, tool or database. Both the NCBI Handbook and the Help Manual are also linked to the Getting Started section in the NCBI footer that appears on all of the new NCBI pages.

Teaching Resources

Teaching Resources are instructional materials that demonstrate NCBI tools and databases using specific examples that highlight useful features. These are divided into Tutorials, Educational Tools and the Course Archive. Each category leads to a separate page with additional links. The NCBI Glossary is also linked to the Teaching Resources and provides an up-to-date view of technical terms associated with modern molecular biology, genetics, genomics as well as NCBI-specific terms.

The Tutorials are step-by-step resource- or problem-oriented documents, Web pages or videos that explain how to use a resource or perform a task. Web Tutorials are NCBI Web pages or chapters from the Bookshelf. Some of these show how to use a specific resource, such as the Genome Workbench tutorial. Others, such as the Disease Genes book chapter, follow a path or tell a scientific story that highlights specific tools and databases. The Disease Genes chapter [3], the BLAST QuickStart chapter [4] and the 3D structures Web pages are based on the NCBI Mini-courses, one of several previous NCBI instructor-led course offerings. The Problem Sets are additional material from NCBI courses and are printable standalone problem sets or examples. These include the problem sets from the general NCBI Field Guide course and the detailed handouts from the problem-based and resource-oriented NCBI Mini-courses. Although these courses are no longer taught, the materials in Tutorials are maintained so that they work as shown with the current NCBI Web pages and underlying data. Information on current NCBI course offerings is presented below in the Courses and Workshops section. Video Tutorials are short videos and other animated presentations on how to use the NCBI site. Many of these demonstrate aspects of the literature services, PubMed and PubMed Central. In addition there are a growing number of videos such as the ones entitled 'Downloading records from Entrez' and 'Retrieving sequences for an organism' that are animated renderings of the How-to guides from the NCBI Guide. These show how to perform a specific task in the NCBI molecular databases.

There are also three interactive Educational Tools—originally developed as part of the NCBI training courses—that are useful learning aids. The Amino Acid Explorer displays and sorts amino acids by their physical and chemical properties, shows the consequences of nucleotide mutations and lists the functions of each amino acid in various conserved protein domains. The PSSM Viewer provides insight into important functional elements within proteins through interactive displays of position-specific score matrices (PSSMs) derived either from the NCBI CDD database or produced by standalone PSI-BLAST. The Ebot tool from one of the NCBI technical workshop courses is a demonstration program that translates a user-defined query into a data-gathering pipeline using the Entrez Programming Utilities (E-utilities) interface. The tool generates a downloadable Perl script that can be executed locally to retrieve data.

The NCBI has a long tradition of offering instructor-led workshops. These have included the Mini-courses and Field Guide, computer-oriented technical workshops and the team-taught collaborative course for medical librarians and information specialists. The Course Archive contains these course materials in their original forms and includes the web pages, slide sets and handouts. Materials here are no longer maintained and reflect the state of the NCBI data and web site on the last course date,

which was in the spring of 2008 for most courses. Nevertheless, many of these materials are still useful. Those from the technical workshops, the course for librarians and the mini-courses can serve as self-guided interactive tutorials, and all can serve as starting points for developing new teaching materials.

Courses and workshops

The current NCBI educational outreach program consists of NCBI-taught Discovery Workshops, Webinar broadcasts and a conference presence including tutorial sessions and live help at the NCBI exhibit booth.

Discovery Workshops will be offered beginning in the fall of 2010. The 2-day program consists of a set of four, 2.5-h hands-on sections emphasizing different aspects of the NCBI site. Each module uses specific examples to highlight important features of a set of related resources and tools and shows how to accomplish common tasks using these. The four modules focus on the following areas: Sequences, Genomes and Maps; Proteins, Domains and Structures; Using NCBI BLAST; and Human Variation and Disease Genes. These training courses are offered three times a year at the National Library of Medicine on the main campus of the National Institutes of Health in Bethesda, MD and four times a year in other areas of the United States. The off-campus courses are distributed across the eight regions of the National Network of Libraries of Medicine (http://nnlm.gov) so that each region hosts a course every 2 years. The Discovery Workshops page has more information and a list of upcoming programs.

In addition to the in-person training courses, NCBI offers short web broadcasts, or Webinars, that are presented using the Adobe Connect web meeting software. Using this delivery mechanism, the only software needed to participate is a web browser. The sessions provide hands-on practice using the NCBI web site guided by an NCBI instructor. Webinars are presented to universities, medical schools and government research facilities in the United States. Multiple classrooms at multiple sites can participate in each Webinar. Current Webinar titles include: NCBI Overview; What's New at NCBI; An Update on NCBI BLAST; and Genomes Update. A separate Webinars page has more details, upcoming broadcasts, links to materials and requirements for hosting a course.

The final component of the NCBI educational outreach program is the NCBI exhibit presence and associated tutorials at scientific conferences. NCBI has a traveling exhibit booth that is part of the marketing exhibit program at four major scientific meetings a year. NCBI Fact Sheets described above in the Documentation section of this review are available at the booth as well as current articles describing the NCBI services. More importantly, two NCBI Public Services staff members are available at the exhibit booth to provide live help on the NCBI site. Staff members provide a range of services: answering simple questions, troubleshooting problems with searches, and helping with designing search and data gathering strategies. At least twice per year, NCBI staff provides tutorial lectures as part of the exhibit program on a specific subset of NCBI tools such as Human Genome Resources or BLAST. The Conferences page presents the NCBI exhibit schedule and has more details. In addition, all NCBI training events and exhibits are announced through the NCBI news outlets and social networking sites presented in the next two sections.

NEWS AND UPDATES

News and Updates comprises mechanisms for keeping up-to-date with changes to NCBI resources, and these mechanisms include the NCBI Newsletter, announcement lists and RSS feeds. The NCBI News portal on the Guide page has short summaries of announcements and other news items of interest to NCBI visitors. These items are also released on the NCBI-announce mailing list and RSS feed. The NCBI-announce mechanism is the master outlet for news and announcements. It excerpts and summarizes information from 11 other RSS feeds and 18 topic-specific mailing lists produced separately for individual resources. Links to the RSS Feeds page as well as the page for subscribing to the NCBI-announce mailing list or any of the topic-specific lists are under the News and Updates section of the Education page.

The NCBI Newsletter is a monthly publication produced on the Bookshelf. It provides a summary of new releases, changes and upcoming

events relating to the NCBI services. The newsletter not only compiles the postings to the NCBI- announce mailing list and RSS feed, but also offers occasional feature articles. An article may demonstrate and provide usage tips for new and updated features and tools, such as the new PubMed interface, or present aspects of biologically interesting data sets available from the NCBI, such as build 37 of the human genome. The complete collection of past NCBI Newsletter issues is also on the Bookshelf.

Community

The Community section includes links to social networking sites and YouTube as well as to other outside sites that are supplemental to the NCBI services. NCBI recently established a presence on the social networking sites Facebook and Twitter. While rapid dissemination of news and announcements is one aspect of these Facebook and Twitter accounts, these two services also provide community networking opportunities for users of NCBI services through the growing lists of Facebook fans and Twitter followers. On Facebook, visitors can comment and post on the NCBI wall providing a potential forum for NCBI users. NCBI also has a YouTube channel that features video recordings of public presentations such as those associated with the GenBank 25th and NCBI 20th anniversary celebrations. How-to animations and other video tutorials will be added to the channel soon. Recommended Links lists links to various sites with related or supplemental content to the NCBI are also a part of Community. Sites linked here include educational pages from other NIH institutes and outside government and university sites.

SUMMARY

The NCBI web site, the premier gateway to increasingly large and complex biomedical data, is undergoing revolutionary changes as part of the Discovery Initiative to make services easier to use and to provide more relevant and useful results. The rapidly changing landscape of biomolecular data means that education and training remain an essential component of improvements to the NCBI site no matter how intuitive and powerful the search tools and resources become. NCBI continues to meet this educational challenge by providing a wide variety of educational materials, hands-on training

opportunities, and rapid communication and discussion through social networking outlets. The main access points for these educational products and services are the redesigned NCBI Guide and Education pages. The Public Services section will continue to develop and create learning tools and resources to support effective access to wealth of interconnected data at the NCBI and support scientific discovery.

Key Points

- The NCBI homepage has been redesigned as the NCBI Guide that lists all NCBI resources in 14 categories alongside How-to guides for common tasks.
- The NCBI Education page (www.ncbi.nlm.nih.gov/Education/) is the primary portal for accessing NCBI educational materials and resources.
- NCBI provides a number of teaching resources including web, print and video tutorials, interactive tools and an archive of course material.
- NCBI offers on-site training through the new Discovery Workshops as well as live Webinars offered at specific institutions and exhibits at conferences.
- NCBI provides information about new and updated services through the online NCBI News, RSS feeds, mailing lists as well as on Facebook, Twitter and YouTube.

References

1. Sayers EW, Barrett T, Benson DA, *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2010;**38**:D5–16.

2. Benson DA, Karsch-Mizrachi I, Lipman DJ, *et al*. GenBank. *Nucleic Acids Res* 2010;**38**:D46–51.

3. Bhagwat M. Identification of disease genes: example-driven web-based tutorial. *Methods Mol Biol* 2007;**396**:371–93.

4. Wheeler D, Bhagwat M. BLAST QuickStart: example-driven web-based BLAST tutorial. *Methods Mol Biol* 2007;**395**:149–76.