WebWise: Guide to the University of Oklahoma's Advanced Center for Genome Technology Web Site

Kim D. Pruitt¹

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894 USA

This installment of the WebWise series reviews the University of Oklahoma Advanced Center for Genome Technology (ACGT) web site. This sequencing center is involved in sequencing several bacterial and fungal genomes in addition to their human genome sequencing effort. Although there is no intent to minimize the significance or volume of their sequencing effort in other genomes, this article does focus on those pages of the web site that are relevant to the Human Genome Project. The site map displayed in Figure 1 does not include redundant links, and some groupings of related links are depicted as one link (e.g., the several links to collaborators are illustrated as one link on the site map). The main features of this web site, as well as the features of those sites already reviewed, are indicated in Table 1. As with the previous WebWise reviews, you may find it useful to point your Browser window to the ACGT web site while reading.

General Information

The ACGT web site utilizes an organizational approach that facilitates maintenance by reducing the number of pages associated with the web site. One person (the Director, Dr. Bruce Roe) maintains the web site, and displaying numerous links directly on the Home Page both reduces the maintenance burden and reduces the number of "jumps" a visitor must make to reach some of the information. The ACGT Home page (<http://www.genome.ou.edu/>) presents links to

organism-specific data pages at the top of the page. These links are presented as graphic icons and then repeated in text form directly underneath. The last update date for the web page is indicated under the links, as well as in an e-mail link to Dr. Roe. As you scroll down the page you can access a search form (to search the web site), links to additional information about the ACGT, links to local tools, collaborators, and links of more general interest. At the very bottom of the page is a link to the web site statistics; although this type of service is typically more useful to those maintaining a web site, the presentation of the last 10 visitors makes it more interesting to check out.

Some general information about the ACGT is available under the sections titled Information About the Roe Lab and Local Tools and Information. A list of personnel, with e-mail links, is accessible by following the Personnel link. Following the Sequencing Strategy link brings you to a description of the ACGT's sequencing strategy (<http://www.genome.ou.edu/SeqStrategy.html>). And some general information about both the University of Oklahoma and local climate is available, but unfortunately there are no direct links to either a map or travel directions. The selection of links available under the Other Web Sites of Interest header include links to general genome-related resources such as databases, sequence analysis and search tools, and program help documentation. A large listing of online journals is accessible by following the last link in this section.

A special feature of the ACGT web site is the series of Protocols pages (<http://www.genome.ou.edu/proto.html>). Follow the Protocols link from the Home page to reach a list outlining the specific

protocols available. Use the search tool, conveniently provided on this page, to look for a particular type of protocol, or simply browse through the provided links. This set of pages includes instructions and troubleshooting tips for sequencing off a variety of vectors, using Big Dye Terminators, gel preparation, and template isolation, to name a few. In addition, a web version of Bruce Roe's protocol book (Roe et al. 1996) is also available. This book provides numerous procedures for commonly used molecular biology techniques. Overall, these pages provide a useful resource to the research community.

Data

The ACGT's major sequencing effort is on chromosome 22, but it is also sequencing regions on chromosomes 1, 9, and 11 (in the proximity of 1p13.1, 9p22, 9q34, and 11q13). The ACGT has generated >8 Mb of (redundant) human DNA sequence for these regions, as reported at the bottom of the Human Genomic Sequencing Progress page (see below).

To review maps of the sequencing targets, follow the Human and Mouse Genomic link provided at the top of the Home page as a small graphic (the same link is also provided in text form directly below the graphics). The Human and Mouse Genomic Sequencing page (<http://www.genome.ou.edu/human.html>) includes links to maps and a brief summary of the sequencing effort. Clicking on a map link, in the form of a small chromosome graphic, brings one to a page with a larger view of that chromosome image. These graphics depict the general region of the sequencing target relative to the cytogenetic map as well as additional general information

¹Corresponding author.
E-MAIL pruitt@ncbi.nlm.nih.gov; FAX (301) 435-2433.

Insight/Outlook

Figure 1 Site map for the University of Oklahoma's ACGT. The main links to pages discussed in the text are illustrated here. Links at the top of the Home page are to general informational resources; the links located at the bottom are to the data and additional tools. Some links available on the web site are not indicated here.

such as gene, marker, and clone names. The type and amount of this general information is not completely uniform from page to page, but the overview maps do succeed in providing the "big picture." All of the maps are image mapped, although they lead to slightly different things. By clicking on the yellow highlighted regions of the chromosome 1, 9, and 11 maps a listing is called

up of the accession numbers that correspond to the maps. The chromosome 1 page also includes a description of the project, and although three accession numbers are listed they are not, unfortunately, linked to a public database or ACGT's FTP site. Unfortunately, you cannot simply select a clone on a map to view the sequence of that particular clone. However, the accession numbers

are listed in relative order, which strengthens the link between the map and sequence data.

Both low- and high-resolution maps are provided for the chromosome 22 project; these maps are also image mapped so one can simply point and click to zoom into a higher resolution view, and then to access links to the DNA sequence. The first map view pro-

Table 1. Features of the ACGT Web Site

Center			GSC	SC	SHGC	W/MIT	BCM	JENA	UTSW	UWGC	CGM	ACGT	
Map Data	Static Map		●				●		●	●	●		
	Image Mapped		●	●	●			●	●			●	
	Tabular List				●	●	●						
	Clones Linked to Sequence			●	●	●	●		●			●	
Sequence Data	Download data from FTP or Web Site		●	●	●	●	●	●	●	●			
	Download a Database			●									
	Links to Public Databases			●		●	●	●		●	●	●	
	Update Frequency	Daily	●	●	●	●	●	●	●			●	
		Weekly											
		Unknown*								●	●		
	Sequence Annotation	Graphic						●	●		●		
		Text	●					●	●		●		
Not Available		●	●	●	●	●			●		●		
Search Services	Similarity Searches	Performance	a	○	○	○	○	○	○	○	○	○	
			b	○	○	○	○	○	○	○	○	○	
			c	○	○	○	○	○	●	○	○	○	○
			d	●	●	○	○	○	○	○	○	○	○
			e	○	○	○	○	○	○	○	○	○	○
			f	○	○	○	○	○	○	○	○	○	○
	Quality of Output	a	○	○	○	○	○	○	○	○	○	○	
		b	●	○	○	○	○	○	○	○	○	○	
		c	○	○	○	○	○	○	○	○	○	○	
		d	○	●	○	○	○	○	○	○	○	○	
Not Available				●	●	●			●	●	●		
Search the Maps			●					●					
Search for Sequences		●	●			●		●			●		
Search the Web Site		●		●	●	●		●			●		
Software	Documentation	a	○	●	○	○	○	○	○	○	○	○	
		b	●	○	○	○	○	●	○	○	●	○	
		c	○	○	○	○	○	○	○	○	○	○	
		d	○	○	○	○	○	○	○	○	○	○	
	Available from:	FTP Site	a	○	○	○	○	○	○	○	○	○	○
			b	○	●	○	○	○	○	○	○	○	○
			c	○	○	○	○	○	○	○	○	○	○
			d	○	○	○	○	○	○	○	○	○	○
		Web Page Link	a	○	○	○	○	○	○	○	○	○	○
			b	○	●	○	○	○	○	○	○	○	○
			c	○	○	○	○	○	○	○	○	○	○
			d	○	○	○	○	○	○	○	○	○	○
			e	○	○	○	○	○	○	○	○	○	○
			f	○	○	○	○	○	○	○	○	○	○
Contact the Site					●	●			●	●			

The red circles indicate features that are available at this web site or the quality of a given feature within a general range of better (a) to worse (f). Sequence data are assessed for their availability from an FTP site, availability in a database (such as ACeDB), whether archived sequences are linked directly to the public database records, the frequency of FTP site update, and whether any sequence annotation is provided in either a text or graphic format. Each web site is scored for the availability of various search services, including the ability to carry out similarity searches against the sequences in their database or perform a key word search of the map data, sequence data, or web site. Documentation and availability of software tools developed by the center are also indicated. (GSC) Washington University's Genome Sequencing Center; (SC) The Sanger Center; (SHGC) Stanford Human Genome Center; (W/MIT) Whitehead/MIT Genome Center; (BCM) Baylor College of Medicine Human Genome Center; (JENA) The Institute of Molecular Biology Genome Sequencing Center Jena; (UTSW) McDermott Center for Human Growth and Development at University of Texas Southwest; (UWGC) University of Washington Genome Center; (CGM) Washington University Center for Genetics in Medicine; (ACGT) University of Oklahoma Advanced Center for Genome Technology.

(*) Although the genome sequencing centers have agreed to update sequence data on a daily basis, the update frequency is not explicitly indicated on the web site.

vides an overview of the chromosome, and targeted regions—named by gene or syndrome—are indicated with red rectangles. The red rectangles are linked to higher resolution maps that provide the clone names, color-coded sequence status, and tiling path. These maps, in turn, are linked to a Sequencing Progress page that presents an ordered list of accession numbers corresponding to the map.

This approach to providing map and sequence data facilitates the integration of these data and is undoubtedly appreciated by anyone trying to assemble such data in a given targeted region. For the most part, there is good agreement between the contig maps and Sequencing Progress pages; however, a couple of discrepancies between the maps and Sequencing Progress reports were noted. For

example, some of the clones that are included on the chromosome 22 Cat-eye syndrome (CES) map (<http://www.genome.ou.edu/maps/ces.html>) are omitted from the list of accession numbers (http://www.genome.ou.edu/maps/ces_gb.html). Although this is probably simply a reflection of which clones have been sequenced and deposited in GenBank (and unfinished clones are not

Insight/Outlook

listed), the more disturbing observation is that some clones are listed on the CES Sequencing Progress page but are not included on the CES contig map. This is a strong indication that these pages need to be updated; some of the Sequencing Progress pages indicated a last update in 1997.

An up-to-date report of the sequence data is provided on the Human Genomic Sequence Progress page (http://www.genome.ou.edu/hum_totals.html). This page is updated daily and can be reached via the Human link toward the top of the Home page or the Human Genome Sequencing Progress link shortly below that (or from the Human and Mouse Genomic page). This Sequencing Progress page, which lists all of the clones for which sequencing is in progress or finished, presents a table that includes clone type, clone name, accession number, level (status), size, quality scores, update date, and chromosome (cytogenetic position). Accession numbers, linked to the GenBank record, are provided for level 2 and 3 (phase 2 and 3) sequences. Only a few links are provided for more preliminary data that are still at level 1 (phase 1). The ACGT web site does not provide links to locally stored sequence data, that is, all of the sequence data links are to records deposited in GenBank. Although this table is not tightly linked to the map data, one can use the chromosome information to roughly determine which clones correspond to which contig map; but again the maps are not all fully consistent with this up-to-date sequence progress table.

A link to the ACGT's FTP site is provided on the Home page, under the Local tools and Information section. Surprisingly, the FTP site does not appear to include the human sequence data. It does include sequence data for some of the bacterial sequencing projects and data targeted toward specific individuals but nothing that one would identify as human data. Once the sequence generated for a clone has reached phase 3 level, there is little difference between the sequence data stored at GenBank versus that stored by a sequencing center. However, for clones that are still very much "in progress," there may be considerable difference, for a period of several days, between the sequence deposited in GenBank and the sequence data available at the sequencing center.

As we all know, science progresses at a rapid pace and can be a competitive arena; there are some that would argue in favor of making all updated preliminary data immediately available via the center's web site.

Tools

The ACGT web site does include a useful search tool on both the Home page and the Protocols page. The Home page search tool successfully found relevant pages when tested with a map region (CES), a clone name (77h2), and a software tool name (Sheetwriter). The search tool on the Protocols page also performs well but should be used specifically to search that series of pages.

Unfortunately, this web site does not yet include a BLAST server to search the human DNA sequence data. It does provide the capability to carry out local BLAST searches on the bacterial genome pages, so hopefully that is an indication that this service will also be provided for the human sequence data in the future.

The Other Web Sites of Interest section of the Home page includes some pointers to software tools and documentation at other sites. Many of these are listed on the Online Sequence Analysis and Search tools page (<http://www.genome.ou.edu/osast.html>). In addition to these useful pointers, two software tools are made available by the ACGT. The Informatics page (<http://www.genome.ou.edu/informatics.html>; follow the Informatics link from the Home page) provides links to documentation on software modified by the ACGT (SheetWriter and PrimOU), other Programming Related Sites, and local copies of software documentation (the Finger-Printed Contigs and Image Gel Visualization and Tracking programs). A link to the Swish web site is included on this page; try this resource if you wish to add search capability to your own web site. The SheetWriter program, originally written at Washington University and modified at the ACGT, creates a sample sheet for gel loading. The program can be downloaded from the ACGT FTP site (<ftp://ftp.genome.ou.edu/pub/programs/sheetwriter>) by following the link provided at the bottom of the documentation page or by following the link to the FTP site provided on the Home page (in the Local Tools and Information section). The PrimOU pro-

gram available here was originally written at the University of Texas Southwest and modified at the ACGT. PrimOU is a primer-picking program and it can also be downloaded from the FTP site (<ftp://ftp.genome.ou.edu/pub/programs/>); a link is provided at the bottom of the documentation page.

Conclusions

The ACGT, although not the largest sequencing center, is committed to providing a significant amount of human genome sequence data. Its web site does provide both map and sequence data, and its provision of image-mapped figures significantly facilitates user integration of these two types of data. However, at the time of this review, some discontinuity between the map and sequence data was noted; possibly caused by either infrequent updates or human error. Although the large table of sequence data is kept up to date by automatic nightly updates, the map pages and corresponding sequence progress pages do not appear to be updated frequently. There is no getting around the fact that it does require a substantial time and financial commitment to build and maintain a web site; and this requirement only grows with the addition of services and tools. However, automation of some of the maintenance tasks (such as updating data pages) can reduce the long-term maintenance costs significantly while also enhancing the usefulness of the data. A sequencing center's web site provides a great deal of essential information to the general research community, and it is important that all of the information be reliably up to date.

Although the ACGT does not currently supply a BLAST search service of its human sequence data, the fact that it is supplying this service for the bacterial genomes gives one hope that it will be available for the human data in the not-so-distant future. With regard to BLAST servers, there are two views. Some may feel it is unnecessary to provide these servers on sequencing centers' web sites, as even the unfinished data are submitted and rapidly available at the public databases. So, why should a sequencing center expend the time and money to supply this resource? By making the map and sequence data available on a web site, the sequencing centers are already providing a useful public service.

Provision of a BLAST server increases the utility of this public service because a visitor to the web site then has an alternate (and convenient) method to access the most current version of the sequence data generated at a given sequencing center. The sequence data in and of itself imparts limited information; it becomes more meaningful only after surrounding it with labels such as "it maps to 22p11.2," or "it is near the CES region," or "it has high sequence identity to the gene I work with." If one knows that a sequencing center is working in the same region that one is interested in, it is more convenient to be able to gather all of the most up-to-date relevant information, including identifying sequence similarity with BLAST searches, at the sequencing center's web site.

REFERENCES

Roe, B.A., J.S. Crabtree, and A.S. Khan. 1996. *DNA Isolation and Sequencing*. John Wiley & Sons, New York, NY.

Next Month: The Joint Genome Institute

WebWise: Guide to the University of Oklahoma's Advanced Center for Genome Technology Web Site

Kim D. Pruitt

Genome Res. 1998 8: 763-767

Access the most recent version at doi:[10.1101/gr.8.8.763](https://doi.org/10.1101/gr.8.8.763)

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Focused on your science.

USC
SCIENTIFIC
SIDE BY SCIENCE

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>