



WebWise: Guide to the University of Washington Genome Center Web Site

Kim D. Pruitt

Genome Res. 1998 8: 572-575

Access the most recent version at doi:[10.1101/gr.8.6.572](https://doi.org/10.1101/gr.8.6.572)

References This article cites 1 articles, 1 of which can be accessed free at:
<http://genome.cshlp.org/content/8/6/572.full.html#ref-list-1>

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](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, and the Cellecta logo, which consists of a green molecular structure and the word "CELLECTA" in white capital letters.

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

Cold Spring Harbor Laboratory Press

WebWise: Guide to the University of Washington Genome Center 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 Washington Genome Center (UWGC) web site. Please note that the web address reported in the first WebWise installment (Pruitt 1997) calls up an internal page; the home page address is <http://www.genome.washington.edu/UWGC/>. As with all of the WebWise reviews, you may find it useful to point your Browser window to the UWGC Home page while reading. Redundant links, and those not related to the Human Genome Project, are omitted from the site map (Fig. 1). The main features of this web site, as well as the features of those sites already reviewed, are indicated in Table 1.

General Information

The UWGC web site, although smaller than some of the other Genome Center web sites, presents an aesthetically pleasing and well-organized Home page. The navigation toolbar located at the top of each page offers links to some general information pages, software tool descriptions, and a site map. Contact information, including mailing address, phone number, fax number, and an e-mail address is provided at the bottom of each page. In addition, a Feedback form, reached through the toolbar link, provides a convenient mechanism to send in comments. The links to the map and sequence data, organized by project, are provided in the body of the page.

By following the General Information links, one is brought to a page containing three links: About the Genome Center, Events and Seminars, and Job Opportunities. The content behind the latter two links is pretty self-evident and need not be discussed here. The first link

brings you to a page that indicates some information about the UWGC's funding, its director (Maynard Olson), and its contributions in the area of software development (including Phil Green's phred/phrap base-calling and sequence-assembly software). By following the Tutorial link, one is brought to a collection of pages that provides a basic overview of the Human Genome Project, which includes general descriptions of the Project, basic genetics, mapping, and sequencing as well as a glossary. Note that the toolbar links are not implemented properly on some of the tutorial pages. Follow the Search/Site Map link to get an overview of the pages available at this web site. This link nomenclature appears to indicate that a tool to interactively search the web site is available just one click away; unfortunately this is not the case. However, the site map does appear to be a comprehensive listing of all pages available at this web site (with links directly to each page), which is useful if you are looking for an internal page or if you merely want to have an overview of the contents. Although the "general information" provided here is certainly a valuable addition to the web site, it would be helpful if travel directions, a map, and a staff directory were provided.

Data

The UWGC focuses on two regions of chromosome 7, namely the 7q14 and 7p31.3 regions, and the HLA class I locus on chromosome 6. A general review of the sequencing progress is available by following the Production Summary link from the Home page (<http://www.genome.washington.edu/UWGC/prodsum.asp>). The Production Summary page indicates that ~2 Mb of chromosome 7 and ~1.8 Mb of HLA class I locus sequence have been obtained. The

chromosome 7 figure refers to finished nonredundant sequence, whereas the HLA class I data are not complete and as such still include overlaps and gaps. A favorable addition to the UWGC web site is the inclusion of data quality information, both in the form of a general description and as downloadable phrap-quality files associated with individual sequence files (see below). The Data Quality information page (<http://www.genome.washington.edu/UWGC/dataqual.htm>) presents definitions and examples of sequencing errors detected in the chromosome 7 project.

A more focused overview of the targets is available by following the Project Overview link for each project on the Home page. Although both the chromosome 7 and HLA class I locus project overviews provide a nice summary, they could be more explicit. For instance, the chromosome 7 overview does not include an indication of the size (Mb) of the targets, and the HLA class I locus overview gives no indication that this locus is mapped to chromosome 6.

Following the chromosome 7 or HLA locus Mapping Summary link brings one to a page that presents a strategy description, an overview of the map, and links to more detailed maps. The overview illustration is quite useful as it indicates the location of the contigs for which more detailed map links are provided. As clones are mapped, they are joined together to form nonredundant contig maps; two and three contig maps are available for the chromosome 7 and HLA class I locus projects, respectively. Links to these map displays are located below the overview illustration. Information about individual clones is presented below that, but the last column of the table displays a "broken gif image" because of a bad link reference in the underlying html code.

To view the contig maps you must

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

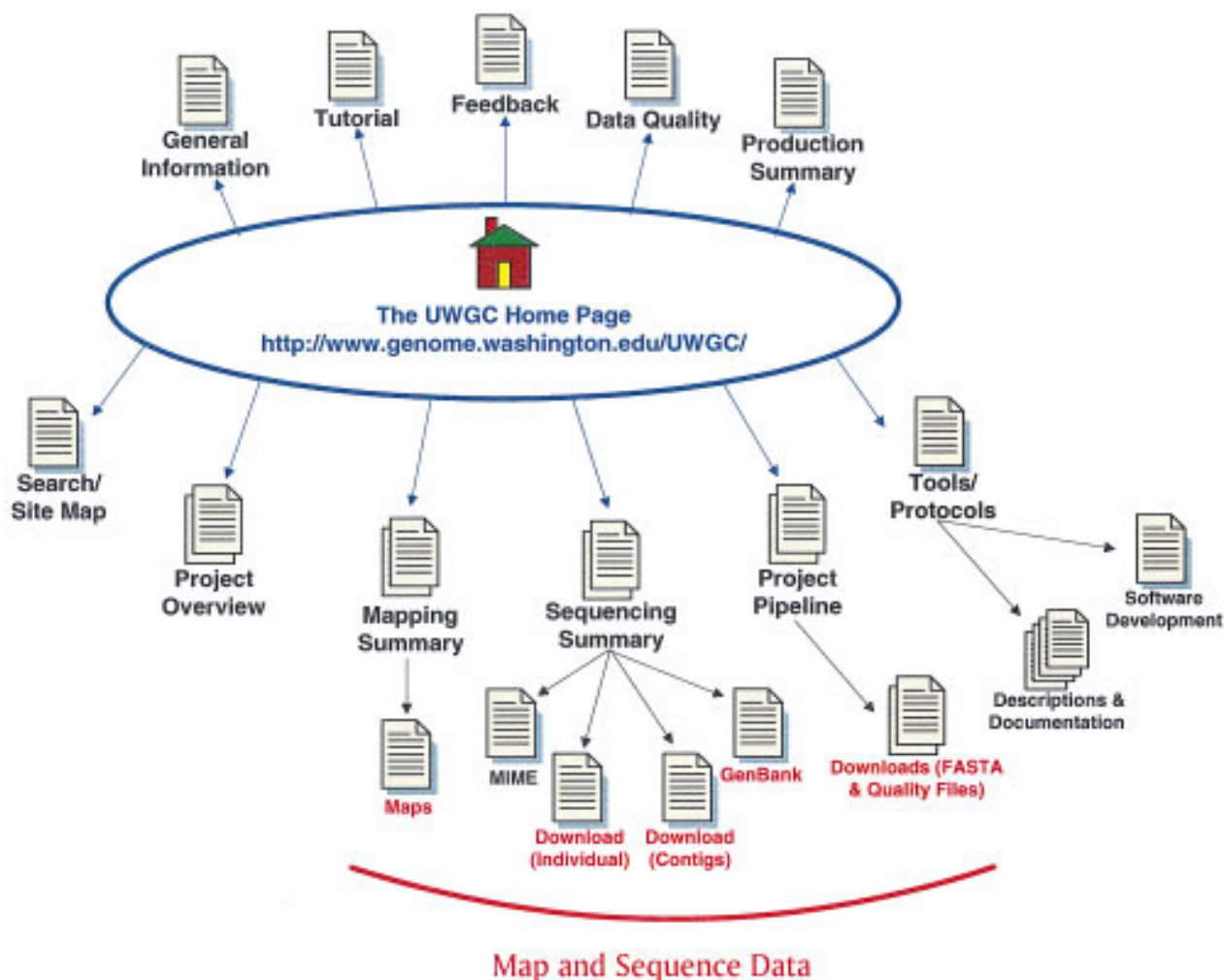


Figure 1 Site map for the UWGC. The main links to pages discussed in the text are illustrated here. Links above the Home Page are to general informational resources; the links located below the Home Page are to the data and additional tools. Some links available on the web site are not indicated here.

have Adobe Acrobat Viewer software installed on your computer. This software is freely available and can be conveniently downloaded by following the link to the Adobe web site provided toward the top of the Mapping Summary page. The maps are displayed in a frames formatted window, but a link to a non-frame view is also supplied. A brief description of the mapping strategy is provided in the top frame, and the map itself is displayed in the bottom frame. Unfortunately, the top frame exhibits some annoying behavior—clicking on a toolbar link displays the new page in the top frame rather than on a new page. Because this is an Adobe Acrobat display, you can use the Adobe toolbar to manipulate the map view. For instance you can zoom, move horizontally and vertically, print, and carry out text

searches. Movement in the display can be accomplished by using the web Browser scroll bars or by simply performing a click-and-drag operation. For the most part, these map displays are large and include the numerous sub-clones isolated to map the region, a restriction map, and a graph of clone depth across the map. The map displays could benefit from the inclusion of a map legend. The labels provided on one of the maps indicates that the sequencing clone tiling path is represented by the lavender shaded boxes that appear directly under the restriction map.

From the Home page, follow the Sequencing Summary link to access the sequence data for a given project. Note that sequence data are not yet available for the HLA class I locus project, although the Production Summary page

indicates that ~1.8 Mb of phase 1 and phase 2 sequence has been generated. In addition, only “finished” submitted sequences are available to download from the chromosome 7 Sequencing Summary page. The chromosome 7 sequence data are available for both joined contigs and individual clones. Two contigs, representing ~2.1 Mb, are available for the chromosome 7 targets; however, it is not intuitively obvious how these contigs relate to the two defined targets or to the map displays. By opening up a second Browser window and looking at both the Sequencing Summary and Map Summary pages one can determine that the two chromosome 7 DNA sequence contigs correspond to the 7q31.3 contig map, with one gap remaining between the two DNA contigs. It is not clear whether any DNA sequence data are pre-

Insight/Outlook

Table 1. Features of the UWGC Web Site

Center		GSC	SC	SHGC	W/ MIT	BCM	JENA	UTSW	UWGC	
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	e	f		
		Not Available			●	●	●		●	●
	Search the Maps		●					●		
	Search for Sequences	●	●			●		●		
	Search the Web Site	●		●	●	●		●		
Software	Documentation	a	b	c	d	e	f			
		FTP Site	a	b	c	d	e	f		
		Web Page Link	a	b	c	d	e	f		
		Contact the Site					●	●		●
	Available from:									

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 Centre; (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.

sented for the 7p14 target; certainly no DNA contigs are available.

The sequence data are not viewable in the Browser window, but the contig DNA data can be downloaded to your local computer by simply clicking on the contig links and selecting the directory path in the download window. The files downloaded consist of a single joined contig in FASTA format. No annotation information is provided either on the web page or in the downloaded files. As you scroll down the Sequencing Summary page you are presented with a

large table listing individual clones, presumably for the 7q31.3 target, for which sequence data are available. One would again have to point two Browser windows to both the Mapping Summary and Sequencing Summary pages to determine the relationship between the clones listed on the Sequencing page and those indicated on the higher resolution maps. The large Sequencing Summary table includes links to the GenBank records, to the ASN.1 records (the MIME link), and to downloadable FASTA-formatted sequence files (the

FILE link). The MIME link appears to use an outdated link format—it calls up a page with scrambled text rather than the ASN.1 format. Although a statement at the bottom of the page indicates that the downloadable FASTA sequence files may be more up to date than the GenBank files, there is no explicit indication of the update frequency of these files.

Additional sequence data are available from the Project Pipeline pages for both the chromosome 7 and HLA class I projects. This page presents a table of links to sequence and phrap quality data in all phases of the production pipeline. Data are organized by status (e.g., Shotgun, Editing, Finishing, Annotation), but a definition of the status terms is not included. Although these status indicators may be self-evident to the large-scale mapping/sequencing community, they are still beneficial to the larger research community to provide an explicit definition of status terms. For example, it is not intuitively obvious how the sequence data available to download would differ between the Finishing and Annotation status levels. In contrast to the chromosome 7 project, very little data can be downloaded from the HLA class I locus Project Pipeline page. The Production Summary page indicates that ~1.8 Mb of unfinished sequence has been generated for this target; however, only one clone is available to download on the Project Pipeline page. This inconsistency raises the question of where the 1.8 Mb of HLA class I locus sequence can be found.

Tools

The UWGC has developed some widely used data analysis tools, the most notable of which are the MCD map assembly software and the PHRED/PHRAP/CONSED base-calling and assembly package. Descriptions of these tools are provided on the Tools/Protocols page, which can be reached from any page by following the toolbar link (<http://www.genome.washington.edu/UWGC/methods.htm>). The Tools/Protocols page provides links to descriptions of methods, protocols, and several data analysis tools, including map assembly, sequence alignment, base-calling, and sequence assembly. The descriptions provide a general overview of each tool that would be of use to someone who does not have experience with the tools.

Specific contact information is provided on these general description pages, and some pages include links to more extensive documentation. In addition, convenient links to download software tools are also provided on some of the description or documentation pages.

The Data Analysis section of the Tools/Protocols page includes a link to the Genome Software Development page (<http://bozeman.mbt.washington.edu/index.html>). This page adopts a different design style and reiterates many of the links available on the preceding page. It does offer a few additional links that may be of interest. For instance, a collection of bioinformatics links can be accessed here (scroll down to the Bioinformatic Jump Page section) as well as documentation of cDNAdb. This software package provides an interface to BLAST and allows one to interactively select interesting alignments to save.

For the most part, the UWGC web site does not provide interactive web interfaces to search or analysis tools. For instance, a BLAST server is not provided to query against the sequence data. Nor does the UWGC provide a tool to search for maps, sequences, or web pages. The only web interface noted is that provided for Repeat Masker, a tool that screens a submitted sequence for repetitive element sequences and returns the masked sequence. From the Tools/Protocols page, follow the Repeat Masker link, then the Submit Data to Server link to access this tool.

Conclusions

The UWGC web site adopts some of the nicer features one hopes to see in any web site, including logical organization, consistent use of a navigation toolbar, and an interesting design style. The site is very easy to navigate and it is a simple matter to find the mapping and sequencing data. Unfortunately, the UWGC does not supply some of the additional features one hopes to find at a genome sequencing center web site, such as a BLAST server and a tighter connection between the maps and the sequence data. It is also notable that they choose not to present the DNA sequence data directly via the Browser window in addition to providing a downloadable file. Obviously, the downloadable file is essential, but for some users the option

to look at the sequence in a Browser window and use their computer's cut-and-paste function is more convenient.

UWGC has made a significant contribution to the Human Genome Project on multiple fronts. In addition to their contribution toward sequencing chromosomes 6 and 7, they have developed several software tools that are now widely used in the community. Their approach to presenting map data is perhaps a reflection of this innovative atmosphere. Presenting genome maps on the web is challenging, as the maps are often quite large and complex. Some centers elect to display the large maps, and one must then use both horizontal and vertical scrolling to view the image. Some centers split up the maps into several more manageable pieces, but one must then hop back and forth between the "big picture" illustration and the higher resolution maps to obtain a comprehensive overview. The UWGC web site presents an alternative approach; map data are presented in Adobe Acrobat PDF format. This is an interesting alternative to the map dilemma. The Adobe PDF format does offer the advantage of having a single display that can be manipulated (scrolled, zoomed, and printed) to obtain the general overview. However, the UWGC displays do not take full advantage of hyperlink options. Thus, the clones are not linked to the DNA sequence data; this lack of interconnection between the data makes it difficult to determine, at a glance, the correspondence between the map and sequence data.

REFERENCES

Pruitt, K.D. 1997. *Genome Res.* 7: 1038-1039.

Next Month: Washington University Center for Genetics in Medicine