

Review

UCSC genome browser tutorial

Ann S. Zweig ^{a,*}, Donna Karolchik ^a, Robert M. Kuhn ^a, David Haussler ^{a,b}, W. James Kent ^a^a UCSC Genome Bioinformatics Group, Center for Biomolecular Science and Engineering, University of California Santa Cruz (UCSC), Santa Cruz, CA 95064, USA^b Howard Hughes Medical Institute, University of California Santa Cruz (UCSC), Santa Cruz, CA 95064, USA

ARTICLE INFO

Article history:

Received 12 July 2007

Accepted 18 February 2008

Available online 2 June 2008

Keywords:

UCSC genome browser

UCSC table browser

Genome tutorial

Genome bioinformatics

Genomics

Bioinformatics

BLAT

Genome-wide data

Comparative genomics

ABSTRACT

The University of California Santa Cruz (UCSC) Genome Bioinformatics website consists of a suite of free, open-source, on-line tools that can be used to browse, analyze, and query genomic data. These tools are available to anyone who has an Internet browser and an interest in genomics. The website provides a quick and easy-to-use visual display of genomic data. It places annotation tracks beneath genome coordinate positions, allowing rapid visual correlation of different types of information. Many of the annotation tracks are submitted by scientists worldwide; the others are computed by the UCSC Genome Bioinformatics group from publicly available sequence data. It also allows users to upload and display their own experimental results or annotation sets by creating a custom track. The suite of tools, downloadable data files, and links to documentation and other information can be found at <http://genome.ucsc.edu/>.

© 2008 Elsevier Inc. Open access under CC BY-NC-ND license.

Contents

Introduction	76
Genome Browser.	76
Viewing tracks	76
Configuring the Genome Browser display	77
Navigating	78
Printing an image	78
Retrieving DNA sequence	78
BLAT	78
BLAT—Try this	79
Custom tracks	79
Custom tracks step 1: Formatting the data	79
Custom tracks step 2: Defining the display characteristics	79
Custom tracks step 3: Uploading your custom track	80
Custom tracks step 4: Sharing your custom tracks	80
Table Browser	80
Table Browser example 1: Searching for genes	80
Table Browser example 2: Joining tables	81
Table Browser example 3: Retrieving DNA sequence	81
Table Browser example 4: Intersecting tracks	81
Sessions	82
Creating a session	82
Sharing a session with others	82
Editing a session	83
In silico PCR	83
PCR—Try this	83

* Corresponding author. CBSE/ITI, 501D Eng. II Bldg, UCSC, 1156 High Street, Santa Cruz, CA 95064, USA. Fax: +1 831 459 1809.

E-mail address: ann@soe.ucsc.edu (A.S. Zweig).

Other tools	83
Proteome Browser	83
VisiGene	83
Gene Sorter	84
Acknowledgments	84
Appendix A. Online resources	84
Appendix B. User's guide	84
Appendix C. Supplementary data	84
References	84

Introduction

This paper provides detailed instruction for using the UCSC Genome Bioinformatics website tools. The paper is divided into sections—one for each of the major tools available on the website. Each section answers these questions about a tool: What does it do? How do I use it? What problem does it help me solve? Additionally, most sections include a hands-on “try this” example that you can use on the UCSC website to experiment with the tool and determine how it might be helpful to you. If you are already familiar with the tools available on the website, you may want to access the Supplementary data for a hands-on example of solving a real problem using the UCSC Genome Bioinformatics website tools.

The major tools discussed in this paper include:

- Genome Browser—graphical view of genes, gene structure, and annotation tracks.
- BLAT—aligning DNA sequence with a reference genomic assembly.
- Custom Tracks—displaying your data in conjunction with existing browser data.
- Table Browser—bulk data manipulation and downloads, intersections and joins between data sets.
- Session—sharing your data with others.
- PCR—getting DNA bracketed by a pair of primers.

Across the top of the website, you will see a dark blue navigation bar—use these links to navigate to the tools discussed in this paper.

Because the suite of tools is continually being enhanced and updated, some of the interface details discussed in this paper may have changed by the time this paper is published.

Genome Browser

The fundamental tool in the UCSC Genome Browser suite of tools is the one that displays the genomic sequence together with annotation tracks, which are mapped to the sequence. Genome annotation tracks include information such as assembly data, genes and gene predictions, mRNA and expressed sequence tag evidence, comparative genomics, regulation, expression, and variation data. This tool is referred to as the Genome Browser (GB) [1,2].

Some of the common uses for the GB are:

- Search for known genes or disease-related genes.
- View orthologous genes in the genomes of other organisms.
- Locate restriction enzymes, STS markers, and BAC end pairs.
- Search for SNPs and other variations in the base genome.
- View base-by-base alignment details of track elements to the genome.
- View microarray expression data.
- Find mRNAs and ESTs from other organisms that map to your assembly.
- Produce figures for scientific publications.

The GB is the right place to start if you are new to the UCSC Genome Bioinformatics website. Navigate to the GB by clicking on the **Genomes** link in the top blue navigation bar. The resulting GB gateway page lists several dozen organisms. Many organisms have more than one assembly

available; as of July 2007 the four most recent assemblies of the human genome are available as separate databases. Each assembly has a date and a name. The date corresponds to the date on which the underlying sequence files were created or released by the sequencing center. The assembly name is based on UCSC nomenclature: the truncated organism name followed by the assembly release number. For example, the most recent mouse assembly for *Mus musculus* is named mm8 (the eighth UCSC *Mus musculus* release). The date on the mm8 assembly is February 2006; the date on which the sequence was released by the sequencing consortium. Human genome assemblies are named **hg** (short for **h**uman **g**enome) followed by a number. The most recent human assembly is named hg18, and the associated date is March 2006. Assemblies for organisms released for the first time after 2003 use a six-letter nomenclature: e.g., bosTau2 for the cow, *Bos taurus*.

To view the annotation tracks display, choose your clade, genome, and assembly, and then search for an item of interest in the genome. You can search for a gene, an mRNA, the name of a researcher, a chromosomal location, and many other things. Check the list under the **Sample position queries** heading on the gateway page for more examples of searchable items. Press the **submit** button after entering your position/search term. If you don't have a search term in mind and you'd just like to see the annotated assembly in visual form, simply press the **submit** button. This will display the assembly showing the default annotation tracks at the default position.

The GB annotation track display page can be divided into several sections (Fig. 1). The browser image forms the data display and is surrounded by a variety of control sections, including navigation controls above and below, a chromosome ideogram on human and some other assemblies showing the location of the window on the chromosome, and track controls grouped by type at the bottom of the page.

The browser image is highly configurable and can easily be exported for use in publications and presentations. At the very top of the image is the Base Position track which shows a series of numbers. These numbers correspond to the base positions for the portion of the chromosome being displayed. The annotation tracks are stacked up in the image below the Base Position track.

Most but not all genomic assemblies are completely assembled onto chromosomes; some are based on scaffolds or contigs. The GB displays these scaffold- and contig-based assemblies in the same manner as the chromosome-based assemblies.

Viewing tracks

If this is your first visit to the GB, and you did not enter a search term, the annotation tracks image will display the default annotation tracks which are an important subset of the entire set of tracks available. To view other annotation tracks, enable them from the track controls by choosing a visibility other than **hide** from the drop-down menu under the track name. A short period of experimentation will familiarize you with the visibility choices. After choosing the annotation tracks you would like to see, press the **refresh** button under the browser image to view them within the tracks image. Each annotation track has a page containing a description of the track, details about track construction, methods used in the experimentation, validation, credits, and references. Access the description page

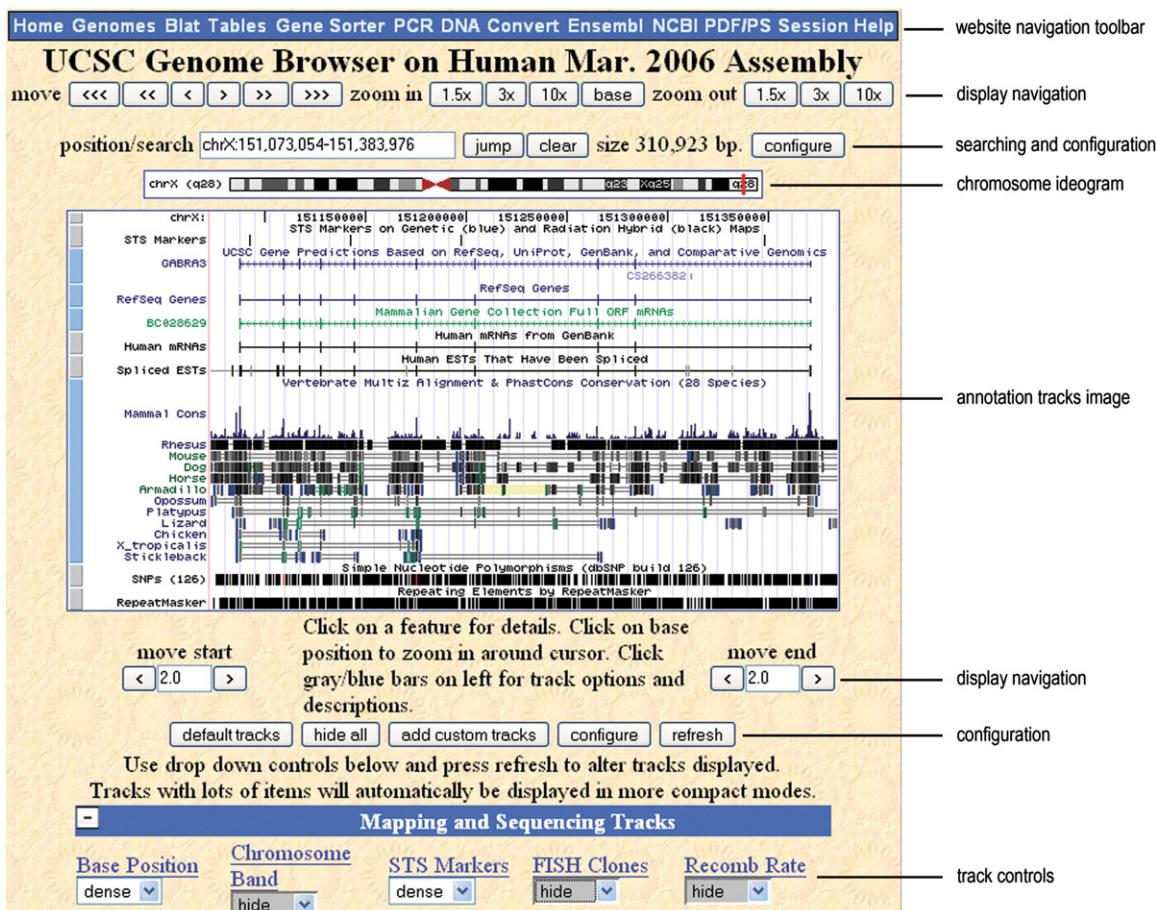


Fig. 1. The UCSC Genome Browser display for the hg18 assembly with the default tracks at the default position. At the top of the page is the website navigation toolbar. Below that are two rows of buttons for navigating within the display of the annotated genome. For the human assemblies and some other genomes a chromosome ideogram is displayed next—the red mark shows precisely what portion of that chromosome is enlarged in the tracks image. Below that, the genome annotation tracks image displays the annotated genome in a horizontal orientation with the short arm of the chromosome on the left. Below the annotation tracks image there are more buttons for configuring aspects of the display. In some assemblies there is a chromosome color key. Finally, at the bottom of the page, the track controls are grouped together by type. Use these controls to add or remove individual annotation tracks to or from the display.

for a track either by pressing the small blue or gray bar to the extreme left of the track display, or by clicking on the track name in the track control section.

Many of the track description pages also include a configuration section where it is possible to make changes to some aspects of the track display for that track. Examples of things that can be configured on a track-wide basis include: setting data thresholds; including or excluding data from certain sources; choosing from a set of data labels; coloring track elements by codons; and choosing graph type, height, range, and scale. The items that are configurable depend on the track and data type.

In addition to the description of the track, there are also details available for each individual element in a track. Depending on the track, these details may be quite extensive including such things as: summary information from other databases, primers, alignment to the genome, and off-site links to even more information. To see these details, click directly on the element in which you are interested. As always, to return to the GB, click on the **Genome Browser** link from the navigation bar.

Configuring the Genome Browser display

There are many other ways to customize the annotation tracks image. Press the **configure** button to view the configuration page. This button is found both above and below the tracks image. At the top of the configuration page, you can change the width and font size of

the tracks image. The items in the next section of the page are used like so:

- **Display chromosome ideogram above main graphic.** Uncheck this box if you do not want the chromosome ideogram to be shown above the GB tracks image.
- **Show light blue vertical guidelines.** Uncheck this box if you do not want to see the light blue vertical guidelines in the background of the tracks image.
- **Display labels to the left of items in tracks.** Uncheck this box if you do not want to see the track labels and element labels on the left side of the tracks image.
- **Display description above each track.** Uncheck this box if you do not want to see the track name displayed above each track in the tracks image.
- **Show track controls under main graphic.** Uncheck this box if you do not want to see the track controls listed below the tracks image. If you choose to hide the track controls, you will then need to use the configuration page to turn individual annotation tracks on and off.
- **Next/previous item navigation.** Check this box if you would like to be able to navigate between consecutive elements within a track. In the tracks image, you will see gray-shaded double arrowheads on either side of the track name. Click on a double arrowhead to navigate to the next off-screen element in that track. This configuration item is off by default.
- **Next/previous exon navigation.** Check this box if you would like to enable navigation from one exon or alignment block to the next within a gene track. In the tracks image, you will see open double

arrowheads directly on the gene or other track element. Click on a double arrowhead to navigate to the next exon in that gene. This configuration item is off by default.

- **Enable track re-ordering.** Check this box if you would like to change the order in which the annotation tracks are displayed in the tracks image. This is useful when constructing images for use in publications. In the track list section at the bottom of the configuration page, you can enter different numerical values for each track or group of tracks. These changes cause a reshuffling of the individual tracks within a group, or of entire groups with respect to one another. Additionally, you can move a track to a different group by choosing a new group name from the drop-down box in the **Group** column. This configuration item is off by default.

After making changes on the configuration page, press the **submit** button to return to the GB to view the resulting tracks image.

Navigating

The **position/search** box always contains the chromosome name (or a scaffold or contig if this assembly is not chromosome based) and the start and end position of the chromosomal region that is shown in detail in the annotation tracks image. There are several ways to view a part of the chromosome other than that on display. Use the **move** buttons near the top of the GB page to keep the same number of bases displayed in the image, but shift the entire display right or left. **Move** buttons with chevrons pointing to the right shift the entire display toward the right end (long arm) of the chromosome. Buttons with more chevrons move further along the chromosome than buttons with fewer chevrons. Exactly how far the image moves along the chromosome depends on the width of your display image and how closely zoomed in you are.

You can also move only the right or left end of the tracks image while leaving the other end where it is, effectively squeezing or stretching it. To do this, use the **move start** or **move end** buttons located immediately below the tracks image. Increase the number of base pairs in the display by moving the start position to the left—enter a number in the box on the left side and press the button with the single left chevron below the **move start** label. The start position will shift left by the number of light blue gridlines specified in the box.

When you come across a track element that you would like to look at in more detail, you will want to zoom in. To do this, use the **zoom in/out** buttons near the top of the web page. Each of the zoom buttons keeps the display centered, and zooms the start and end points in or out. To zoom in to an element that is not in the center of the display, click the Base Position track directly above the element of interest. This will recenter the display around your element of interest and zoom in. The default is a 3x zoom; change that by editing the Base Position configuration page.

When the display is zoomed in to base granularity, the individual nucleotides are listed across the Base Position track (Fig. 2). To the left of this listing, you will see a directional arrow. If this arrow points right (the default) then the nucleotide listing is read from left to right ($5' \rightarrow 3'$) and is correct for elements on the sense strand. If your element of interest is on the antisense strand, press the arrow to complement the bases, and then read them from right to left ($3' \leftarrow 5'$). Note that this does not change the direction of the annotation tracks displayed below the Base Position track. You will be able to determine which strand your element of interest is on by looking for directional chevrons on the element in the tracks image. Annotation tracks such

as gene tracks, mRNAs, ESTs, etc. quite often superimpose chevrons onto each element in the track or place them onto introns between them. Right-facing chevrons denote that the element is on the sense strand.

Printing an image

The GB provides a way to produce a copy of the tracks image suitable for publication or printing. Before printing, you may wish to add a title to your tracks image. Click on the small blue or gray bar to the extreme left of the Base Position track and enter a title into the **title** box. You can also choose to add the assembly name and chromosomal position to this track. Once you have configured the tracks image, click on the PDF/PS link in the navigation menu. Images saved in PostScript can be printed at a high resolution and/or edited. Images saved in PDF format can be viewed using Adobe Acrobat Reader.

Retrieving DNA sequence

To capture the underlying DNA sequence for the chromosomal position showing in the annotation display, click on the **DNA** link from the navigation bar. This page contains configuration options for the DNA output format. To display extra bases upstream or downstream of the sequence, enter the number of bases in the corresponding text box. The Sequence Formatting section lists options for adjusting the case of all or part of the DNA sequence. To choose one of these formats, click the corresponding option button or check box. Press the **get DNA** button when all of the options are set.

If you would like to retrieve DNA sequence for specific track elements (e.g., only the exons of a specific gene or set of genes), or for multiple locations, you can do this using the Table Brower tool (see Table Brower section of this paper).

BLAT

BLAT (BLAST-Like Alignment Tool) is a sequence alignment tool [3]. It has the ability to align both DNA and protein sequence to the underlying genome. BLAT on DNA works by keeping an index of the entire genome in memory—it is very fast. BLAT on DNA sequence is designed to quickly find sequences of 95% or greater similarity, of a length of 40 bases or more. It may miss more divergent or shorter sequence alignments. BLAT on protein sequence finds sequences of 80% or greater similarity, of a length of 20 amino acids or more.

Some common uses of BLAT include:

- finding the genomic coordinates of mRNA or protein within a given assembly
- determining the exon structure of a gene
- displaying a coding region within a full-length gene
- isolating an EST of special interest as its own track
- searching for gene family members
- finding human homologs of a query from another species.

Navigate to the BLAT tool by clicking on the **BLAT** link in the top blue navigation bar. Configure the BLAT page by choosing the genome and assembly to which you would like to align your DNA or protein sequence. Let the BLAT tool know what type of sequence you have: choose this from the drop-down list under **Query type**. Alternatively, you can leave this set to the default: **BLAT's guess**; the tool almost always guesses correctly. To order the search results based on how

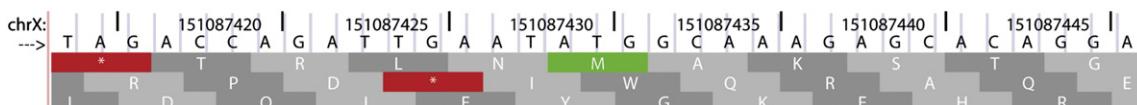


Fig. 2. Base Position track with nucleotides, codon translation, and directional arrow.

Table 1

Sample DNA and protein sequence from human genome to BLAT to chimp genome assembly (FASTA format)

```
>sample_DNA_sequence
ATGAGCTTTGATAACAACTACCATTGGTGGCCAGGGCTATGCCAAAGGAGG
CCTGGGCTGCAGCTATGGCTGTGGCTTAGCGGCTATGGCTATGGCTGCTGCTG
ACTGCCCATGGTGTATGAAAGATCTGGTTCTGGCTGCTTCCTGA

>sample_protein_sequence
MSFDNNYHGGQGYAKGLGCSYCGLGSYGYACYCPWCYERSWFSGCF
```

close your sequence matches the genomic sequence, choose one of the score options in the **Sort output** menu. The score is determined by the number of matches vs mismatches in the final alignment of the sequence to the genome.

Configure your sequence in FASTA format as shown in [Table 1](#) for submission to the BLAT tool. FASTA is a very simple plain text format for displaying nucleotide or protein sequence. For each record, there is one header line that begins with ">" and contains a description or name of the record, followed by one or more lines whose letters represent the DNA or protein sequence. You can submit up to 25 sequences at the same time if they are of the same type and are preceded by unique header lines. Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10,000 or fewer letters will be processed. Press the **submit** button after configuring the BLAT page and pasting your sequence in the large text box.

The results from the BLAT program will be displayed on the BLAT search results web page. If you chose to view your results as hyperlinks, you will see two links for each result. The **browser** link will take you to the GB with your BLAT result aligned to the other tracks in the image. The **details** link will show you a base-by-base alignment of your sequence with the underlying genomic sequence. If you chose to view your results in pattern space layout (psl) format, you will see the results listed in psl format either with or without a header line. The output from the **psl no header** option can be used to create a custom track in the Genome Browser (see the Custom Tracks section of this paper).

BLAT—Try this

Open the UCSC Genome Browser to the home page and navigate to the BLAT tool. Configure the tool like so:

Genome:	Chimp
Assembly:	Mar 2006
Query Type:	DNA
Sort Output:	query, score
Output Type:	hyperlink

Copy the DNA part of the FASTA file from [Table 1](#) and paste it into the text box. This DNA is from the human hg18 assembly. However, we are BLATting it on to the Chimp assembly. Press the **submit** button. The BLAT tool finds five hits for this DNA sequence on five different chromosomes in the Chimp assembly. The best hit is on the sense strand of chromosome 21 at this location: 30,344,132–30,344,278. Click on the **details** link for this match to see the base-by-base alignment details between the human DNA you entered into the BLAT tool (the query), and the chimp DNA it matches (the target) (see [Table 2](#)). Click on the **browser** link for this match to open the GB to this location in the Chimp assembly. The BLAT results will be displayed as a custom track in the GB.

Custom tracks

Display your own experimental results or annotation sets in the UCSC GB by creating a custom track. Your custom tracks will be displayed in the GB together with the existing standard annotation tracks. This is

useful for visualizing your data in conjunction with the existing annotation data. For example, you may have a file with human SNP locations that have been discovered in your laboratory. You can create a track showing the location of those SNPs and use the Table Browser to intersect your track with the existing SNP track to find out which SNPs have not yet been reported.

Depending on how you create your custom track, your data are either kept on the UCSC servers for some finite period of time (at the writing of this paper, custom tracks are removed from the server after 48 hours of non-use), or you can host it on your own server for more privacy and an indefinite span of time. Either way, the data are not made available to the general public unless you specify it.

To create and display a custom track, follow these steps.

Custom tracks step 1: Formatting the data

Create an annotation file by formatting your data into one of the following formats:

- General Feature Format (GFF)
- Gene Transfer Format (GTF)
- Pattern Space Layout (PSL) (you can get this format from the output of the BLAT tool)
- Browser Extensible Display (BED)
- Wiggle format (WIG)

More detail about these file types can be found in the **Data File Format** section of the FAQ at: <http://genome.cse.ucsc.edu/FAQ/FAQformat>.

Custom tracks step 2: Defining the display characteristics

In its most basic form, a custom track data file need only contain the data set with positional information formatted in one of the above file formats; the **browser** and **track** lines discussed in this section are optional.

Add one or more optional **browser** lines to the beginning of your annotation file to configure such things as the position to which the GB opens when your track is displayed, and the visibility of other annotation tracks. For example, if the following line is prepended to your data:

```
browser position chr4:1-20000
```

the GB will initially display the first 20,000 bases of chromosome 4 when your custom track is opened. If this line is prepended to your data:

```
browser hide all
```

the GB will hide all other annotation tracks except your custom track. You can choose to hide individual tracks and show others using **browser** lines.

After the **browser** lines and before your formatted data, add a **track** line to define display attributes for your custom track. In this

Table 2

Side-by-side alignment of human DNA from [Table 1](#) BLATed onto the Chimp genome

```
00000001 atgagctttgataacaactaccatggtggccagggtatgccaaaggagg 00000050
>>>>> ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| >>>>>
30344132 atgagctttgataacaactaccatggtggccagggtatgccaaaggagg 30344181
00000051 cctgggctgcagctatggctgtgttagcggtatggctatgcctgtct 00000100
>>>>> ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| | >>>>>
30344182 cctgggctgcagctatggctgtgttagcggtatggctatgcctgtct 30344231
00000101 actggccatgggttatgaaagatcttggtttctggctgtctctga 00000147
>>>>> ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| >>>>>
30344232 actggccatgggttatgaaagatcttggtttctggctgtctctga 30344278
```

Human DNA (query) is listed on each of the top three rows; chimp DNA (target) is listed on the bottom rows. Where there is no vertical line, there is disagreement between the bases in the query and the target.

definition line you can set such attributes as track name, description, colors, etc. For example,

```
track name="my track name" description="this track
contains my annotation data" visibility="full"
htmlUrl="http://myurl.com"
```

This assigns a name and description to the data set that follows it. It also ensures that the custom track will be displayed when the GB opens. Additionally it copies the text from the specified URL and places it on the description page for this custom track.

If you have included more than one data set in your annotation file, insert a **track** line at the beginning of each new set of data. However, each annotation file should only contain one set of optional browser lines.

Custom tracks step 3: Uploading your custom track

Now that your annotation file has been configured, you will need to move it to the UCSC GB. To access the custom tracks upload page, press the **add custom tracks** button from the GB gateway page or annotation track display page. There are three methods for loading your custom track into the GB: file, URL, and direct text entry.

If the annotation file you created resides on your computer and you would like to upload it onto the UCSC server, use the file method. Press the **Browse** button located directly above the **URL/data** text box, and then locate the annotation file on your computer. The annotation file may be compressed by any of the following programs: gzip (.gz), compress (.Z), or bzip2 (.bz2). Compressed files must include the appropriate suffix in their name. Uncompressed files are also accepted, but your Internet Browser may time out while attempting to upload very large files.

To load your track via URL, first configure your annotation file (see Steps 1 and 2) and place it on your web server. The GB supports both HTTP and FTP (passive-only) protocols. Enter the URL for your annotation file in the **URL/data** text box. To load more than one track at a time using this method, enter several URLs at once, placing each URL on a separate line. In this case, the data are hosted on your server but are also loaded into a temporary database on the UCSC server that is accessible only by you.

The third method of uploading your data is to type or paste the annotation text directly into the **URL/data** text box.

Whichever method you use, press the **submit** button when you are finished. This will take you to the **Manage Custom Tracks** page. To view your custom track in the GB, from the **Manage Custom Tracks** page click on the chromosome number in the Position column of your custom track.

Custom tracks step 4: Sharing your custom tracks

Unless you can control file-level access on your own computer via user permissions, all users of your computer will be able to see your custom tracks, but they will be hidden to users of other computers. If you have used the URL upload method to create your custom tracks, you can share them with others on other computers or at other locations. They will have access to your data and be able to view your custom tracks in the UCSC GB. To allow this, you will create another URL that links your annotation data to the GB, and give it to them. The URL that you will share must contain three pieces of information:

- The species or genome assembly on which your annotation data are based. To specify a particular genome assembly for an organism, use the **db** parameter, **db=database_name**, where *database_name* is the UCSC code for the genome assembly. An example of this is: **db=hg18** (Human, March 2006 assembly).
- The genomic position to which the GB should initially open. This information is of the form **position=chr_position**, where *chr_position* is a chromosome number, with or without a set of coordinates. Examples of this include: **position=chr22**, **position=chr22:15916196-31832390**. Note that if the annotation data

file includes a position line, that value will override this one; the browser will open to the position defined in the annotation data file.

- The URL of the annotation file on your website. This information is of the form **hgt.customText=URL**, where *URL* points to the annotation file on your website.

Combine the above pieces of information into a URL of the following format (the information specific to your annotation file is in italics):

```
http://genome.ucsc.edu/cgi-bin/hgTracks?db=db_name&
position=chr_position&hgt.customText=URL
```

Provide this complete URL to others with whom you would like to share your custom tracks. Advise them to paste the URL into the **URL/data** text box on the **Add Custom Tracks** page, and then press the **Submit** button.

The Sessions section of this paper discusses a way to share custom tracks together with specific UCSC tracks of your choosing.

Table Browser

The data underlying each annotation track in the GB are kept in one or more database tables. The UCSC Genome Bioinformatics website offers complete access to these tables through the Table Browser (TB) tool. The TB tool provides a graphical interface that allows users to query the database tables with basic and advanced structured query language (SQL) queries. However, knowledge of SQL is not necessary. The TB provides a convenient alternative to downloading and manipulating the entire massive database. Navigate to the TB by pressing the **Tables** link in the top blue navigation bar.

Using the TB, you can:

- retrieve the DNA sequence or annotation data underlying GB tracks for the entire genome, a specified coordinate range, or a set of accessions
- apply a filter to set constraints on field values included in the output
- generate a custom track that can be graphically displayed in the GB
- conduct both structured and free-form SQL queries on the data
- combine queries on multiple tables or custom tracks through an intersection or union and generate a single set of output data
- display basic statistics calculated over a selected data set
- display the schema for a table and view a list of all other tables in the database that are connected to the table
- organize the output data into several different formats for use in other applications, spreadsheets, databases, or custom tracks.

Here we use a few examples to demonstrate the power of the TB tool.

Table Browser example 1: Searching for genes

Assume you have several hundred RefSeq Gene Accession names and you would like to know their exact position on the most recent human genome assembly. You would like to get the output in two ways: as a list of chromosomal locations, and visually in the GB. You can accomplish all of this using the TB.

Navigate to the TB and configure it like so:

clade	Vertebrate
genome	Human
assembly	Mar. 2006
group	Genes and Gene Prediction Tracks
track	RefSeq Genes
table	refGene
region	genome

By setting the **region** to genome, you will ensure that the TB tool searches the table that contains the entire set of RefSeq Gene data: refGene. If you know, on the other hand, that all of your RefSeq Genes are on chromosome X, you can speed your search by restricting it to only this chromosome. Enter **chrX** into the position text box and press

the **lookup** button. In this example, we will set the **region** to genome. Note that if you are searching a large table, a full genome search may take such a long time that your Internet Browser times out before the results are completely generated. In this case, you should consider breaking your search apart into individual chromosomes.

To upload your RefSeq IDs, press the **paste list** button next to the **identifiers (names/accessions)** label. On the next page, paste the list of identifiers for which you would like to search, one per line or separated by white space. For this example, we will use a list of only 10 RefSeq Genes, but the TB can easily handle much larger lists. See **Table 3** for a list of the RefSeq Genes that can be used in this example. Press the **submit** button to return to the main TB page. Press the **summary/statistics** button to see a list of statistics regarding your selection. A few of the more interesting statistics include the number of items in your selection (10 in this case) and what percentage of the entire genome those items cover (0.01% in this case). The **Back** button on your Internet Browser will return you to the main TB page.

To view your genes in the GB, select **custom track** from the drop-down menu next to **output format**. Press the **get output** button. On this page, you can configure your custom track: give it a name, a description, and more. In this case, we will accept the default values. Your custom track is created for you with the results of your TB query by making a BED file. The default is for the TB to create one BED file per gene. Leave the default setting and press the **get custom track in genome browser** button. This will open the GB and your custom track (the results of your TB query) to either the position you last visited in this genome assembly or the position you specified in your TB search. Because we searched on the entire genome in this example, the GB will open to the position we last visited in this genome assembly. If you do not see any items displayed in your custom track, you are not looking at a location that contains any of the 10 RefSeq Genes for which you searched. All 10 genes in this are located on one chromosome band of chromosome 17: q23.1. To navigate to the custom track, enter **17q23.1** into the **position/search** box, and then press the **jump** button. This will cause the display to switch to this location, and you should see your custom track with the 10 RefSeq genes. Turn on the RefSeq annotation track to confirm the correlation between this track and your custom track.

Table Browser example 2: Joining tables

In this example we will assume that, as in the previous example, you do not know where your RefSeq Genes are located. However this time we demonstrate how to use the TB to find the chromosomal positions of the sample RefSeq Genes, and their associated UCSC Gene IDs. Starting on the main TB page, reconfigure it so that the **clade**, **genome**, **assembly**, **group**, **track**, **table**, and **region** are the same as in the first example. Use the same list of 10 RefSeq Gene identifiers. This time, for **output format**, choose **selected fields from primary and related tables**. If you wish to save the results to a file, in the **output file** text box, enter the name of the file to which you would like your results saved. In this example, because the output file will be small, we will allow it to open directly in the browser by not entering a file name. Press the **get output** button.

On the next web page, we choose the fields of interest from the tables. This is similar to a SQL join statement. Several tables are related to the

Table 3

Ten RefSeq Gene Accession IDs for use in the Table Browser examples

NM_024612
NM_004859
NM_001015509
NM_016077
NM_030938
NM_016261
NM_003161
NM_016125
NM_022070
NM_000717

refGene table by a common field. You can get any information you like from these related tables for the accession IDs that you entered on the main page of the TB. For this example, we will gather the following information: gene name, chromosome, and the position of transcription start, transcription end, coding start, and coding end. Additionally, we will join the refGene table to the UCSC Gene table to find out the corresponding UCSC Gene IDs of our RefSeq Genes of interest.

Place a checkmark next to the following fields in the hg18.refGene table field list: name, chrom, txStart, txEnd, cdsStart, cdsEnd. From the list of linked tables, place a checkmark next to the kgXref table. This table contains gene names cross-referenced across several gene sets. Scroll to the bottom of the page and press the **Allow Selection From Checked Tables** button. In the hg18.kgXref section of the page, place a checkmark next to the kgID field. This is the field of the kgXref table that contains the UCSC Gene ID names. Press the **get output** button. This will save the query results to your computer and give it the name you specified. Open the file to see contents similar to those shown in **Table 4**. The data in the file are tab-separated and therefore easily imported into spreadsheets or other data manipulation programs.

Another quick way to use the TB to get a list of positional information for a list of genes is to choose **hyperlinks to Genome Browser** as the **output format**. This will display a page of hyperlinks associated with each gene name (**Table 5**). Click on a hyperlink to go to that position in the GB.

Table Browser example 3: Retrieving DNA sequence

In this example we will use the TB to create a file containing the genomic DNA sequence of the 10 RefSeq Genes. Configure the TB as before for these items: **clade**, **genome**, **assembly**, **group**, **track**, **table**, and **region**. Ensure that the RefSeq IDs are still present in the **identifiers** list. Choose **sequence** from the **output format** drop-down list. Choose a name for your **output file** and enter it into the text box. Press the **get output** button. From the next page choose genomic sequence, and press the **submit** button. From this page, configure the attributes of the sequence. You can add upstream or downstream sequence, remove introns, choose only coding exons, etc. You can also choose to split the output into separate FASTA records—the default is one FASTA record per gene. Additionally you can choose to see parts of the sequence in uppercase or lowercase. It is also possible to mask out repeat sections to either lowercase or to Ns. After you have configured the output to your preferences, press the **get sequence** button.

The output file will contain your sequence formatted as specified. Preceding each block of sequence is a header line that starts with a ">", and then an identifying name, which usually includes the gene name. Assuming you accepted all of the default values, the first part of the output will look similar to the output shown in **Table 6**.

Table Browser example 4: Intersecting tracks

The TB tool can also be used to intersect two annotation tracks with each other. This is useful if you are interested in seeing which elements are unique to one track, which areas of the genome are devoid of elements in both tracks, if any of the track elements are located in gap regions, etc. In this example, we will find out if there are additional genes in the UCSC Gene track that are not found in the RefSeq Gene track. To do this, we will intersect the UCSC Gene track with the RefSeq Gene track limiting the intersect to the region that we have been working with: chr17 q23.1. Configure the main page of the TB like so:

clade	Vertebrate
genome	Human
assembly	Mar. 2006
group	Genes and Gene Prediction Tracks
track	UCSC Genes
table	knownGene

Table 4

Results of Table Browser join for RefSeq Genes and UCSC Gene tables

hg18.refGene .name	hg18.re fGene.c hrom	hg18.refGen e.txStart	hg18.refGe ne.txEnd	hg18.refGe ne.cdsStar t	hg18.ref Gene.cds End	hg18.kgXref .kgID
NM_024612	chr17	54997667	55040484	54997814	55039315	uc002ixn.1
NM_004859	chr17	55051831	55129099	55052274	55125995	uc002ixq.1, uc002ixr.1, uc002ixp.1,
NM_001015509	chr17	55129448	55139638	55129581	55131016	uc002ixs.1
NM_016077	chr17	55129448	55139638	55129581	55130121	uc002ixt.1
NM_030938	chr17	55139810	55273235	55163589	55272054	uc002ixu.1
NM_016261	chr17	55291632	55325078	55292464	55323145	uc002ixw.1, uc002ixx.1,
NM_003161	chr17	55325224	55382566	55325327	55378931	uc002ixy.1
NM_016125	chr17	55384505	55396899	55385119	55396806	uc002iy.1, uc002iyd.1,
NM_022070	chr17	55475333	55511074	55475705	55511057	uc002iyk.1
NM_000717	chr17	55582130	55591686	55582177	55591567	uc002iy.1

Press the **define regions** button after the region text box to enter the exact chromosomal location of the positions for which you would like to restrict your search. Enter the region for the chromosome band chr17q23.1 in BED format: chr17 54900001 55600000. Press the **submit** button to return to the main TB page. If you have a single region, as in this example, you can also use the position search box.

Press the **intersect** button to set up the intersection between the two annotation tracks. Configure the next page like so:

Group track table	Genes and Gene Prediction Tracks RefSeq Genes RefSeq Genes (refGene)
-------------------------	--

Select the button next to the **All UCSC Genes records that have no overlap with RefSeq Genes** choice. Press the **submit** button to return to the main TB page. Choose custom track as the **output format**. Press the **get output** button. From this page, press **get custom track in genome browser** button. The GB will display your newly created custom track at the genomic position that you entered in the **define regions** step. If you also open the RefSeq Gene track and the UCSC Gene track, you will see that your new custom track contains the seven genes from the UCSC Gene track that are not found in the RefSeq Gene track. Note that the seven genes are annotated with the corresponding UCSC Gene name. As in all annotation tracks, click on an element in your custom track to see details about that element.

Sessions

The Sessions tool allows you to configure the GB with specific track combinations, including custom tracks, and save the configuration options. Multiple sessions may be saved for future reference, for comparison of scenarios, or for sharing with your colleagues. Saved sessions persist for 1 year after your last access, unless deleted.

Table 5
RefSeq Genes with chromosome and position information

NM_024612	at chr17:54997668-55040484
NM_004859	at chr17:55051832-55129099
NM_001015509	at chr17:55129449-55139638
NM_016077	at chr17:55129449-55139638
NM_030938	at chr17:55139811-55273235
NM_016261	at chr17:55291633-55325078
NM_003161	at chr17:55325225-55382566
NM_016125	at chr17:55384506-55396899
NM_022070	at chr17:55475334-55511074
NM_000717	at chr17:55582131-55591686

Custom tracks within sessions persist for at least 48 hours after the last time they are viewed.

Creating a session

It is easy to create a session to save or share. Simply configure the GB as you wish, and then navigate to the Sessions tool by clicking on the **Session** link in the top blue navigation bar. To ensure privacy and security, you must log in to the UCSC genomewiki site and create a username and password. Once you have created your genomewiki username and password, you will not need to repeat this step; you can choose to remain logged in and in the future, the session tool will automatically recognize you. After creating a username and logging in, you will be returned to the Sessions tool automatically. Scroll down to the **Save Settings** section of the page. Type a name into the **Save current settings as named session** box. Choose whether you would like to share your sessions with others or not. If you leave the checkmark in the **allow this session to be loaded by others** checkbox, then others will be able to view your GB settings (including your custom tracks) if you provide them with your username and session name. Your session will not be automatically available to the general public; you will have to provide the details for others to view it. This ensures the confidentiality of your private data. After naming and choosing whether to share your session or not, press the **submit** button.

Once you save a session for the first time, it will be available to you (and others, if you share it) for 1 year. After that, as long as you access your session at least once a year, it will persist on the server until you delete it. If your session contains tracks that expire (such as custom tracks, BLAT results, or genome graph tracks), those tracks will expire at their normal rate, and your session will persist without those tracks.

Instead of saving your session as a named session, you can create a file from your session settings. Save this file to your machine, or post it to a URL for access or sharing. To do this, configure the GB as you would like to see it in your session, navigate to the Sessions tool, and scroll down to the **Save Settings** section. Type a name into the **Save current settings to a local file** box. Press the **submit** button to save or display the file.

Sharing a session with others

There are several ways to allow others to access your session. If you have created a named session that you would like to share, just provide your username and session name to the person with whom you would like to share. Advise the person to navigate to the Session tool and type those two names into the **Load settings from another user's saved session** section, press the **submit** button, and then the

Table 6

First few lines of DNA sequence for the RefSeq Gene NM_030938 extracted using the Table Browser

```
>hg18_refGene_NM_030938 range=chr17:55139811-55273235 5'pad=0 3'pad=0
revComp=FALSE strand=+ repeatMasking=none
TGGAGCCCAGCGCGGGGTGTGAGAGTCCTGAAGGAGCAGCTCCAGGATC
CTGAGATCCGGAGCAGCCGGGGTCTGGAGCGGtaggtaaactaactcagg
gaattgggaccgaaggctcgaggctgtccctcagagcttgcgtttccgg
agctctcccaacttaggcattgcggatctggggactgtgtatccgcacctc
cagcagtgttttaggacgcgtctgactggatgtctcacccctggatgtctc
acgctggcggtatccaggatttttggacacccttcatgttgtgttccccc
```

The line that begins with the ">" is the header line. The nucleotides in uppercase are the 5' UTR sequence. The lowercase nucleotides are the beginning of the first intron.

Genome Browser link on the top blue navigation bar. This will take them to the GB where they can view your session as it was when you saved it. An even easier way to do this is to use the **Email** link next to the session you created in the **My Sessions** section of the Sessions tool. Pressing this link will create an email message containing a URL that will take a user directly to the GB displaying your session. Send the email to the person with whom you would like to share your session.

If you have created your session by saving your settings to a local file, you have two choices for sharing. You can send email to others with the file as an attachment or refer them to a URL. If sending the file, advise the person to save the file to their machine, and then open the Sessions tool and browse to that file in the **Load settings from a local file** section. Choose that file, press the **submit** button to load the session and then **Genome Browser** link to go to the GB to view your session.

If you saved your session settings to a local file that is available from a web server, you can provide a URL to others to share your session. Construct the URL like so (the information specific to your annotation file is in italics):

```
http://genome.cse.ucsc.edu/cgi-bin/hgSession?
hgS_doLoadUrl=submit&hgS_loadUrlName=your_URL
```

where *your_URL* is the URL of your settings file, e.g., <http://www.mysite.edu/~me/mySession.txt>. In this type of link, you can replace "hgSession" with "hgTracks" to proceed directly to the GB.

Editing a session

After you have created and saved a session, you can modify it in any way. The actual session will not change until you resave it with the original name. If you have shared your session, the person with whom you are sharing will not see any changes that you make until you resave your session with the same name and they reload it. If the person with whom you are sharing makes a change while viewing your session, they can save it as a new session under their username, but they cannot effect a change in your original session.

You may want to warn the person with whom you are sharing your session that if they are currently using the GB and they choose to open your session, they will lose all of their current GB settings. To avoid this, they should simply save their own work as a session with a different name before opening your session.

To delete a session, log into the Session tool, scroll to the name of your session in the **My Sessions** section, and then press the **delete** button.

In silico PCR

The UCSC PCR tool searches genomic sequence with a pair of primers and returns all sequences in the database that lies between and includes the primer pair. Use this tool to design primers for your research—it will inform you if your choice of primers is unique in the genome. If the primers are in a region that is too common, the tool will return nothing. Navigate to the PCR tool by clicking on the **PCR** link in the top blue navigation bar.

Configure the PCR tool by choosing the **Genome** and **Assembly** on which you are working. The sequence for each primer must be at least 15 bases long. The **Reverse Primer** must be on the opposite strand and pointing back toward the forward primer. If your reverse primer sequence is from the same strand, check the **Flip Reverse Primer** checkbox—this will reverse complement the sequence of your reverse primer. Enter the **Max Product Size**. This is the maximum total genomic sequence length that the PCR tool should look for; primer hits that exceed this length will not be displayed in the output. Use **Min Perfect Match** and **Min Good Match** to fine-tune the 3' end of your primers.

After entering your primers and configuring the tool, press the **submit** button. If there is at least one match, the resulting page displays all hits in FASTA format. The FASTA body is capitalized in areas where the primer sequence matches the genomic sequence and in lowercase elsewhere. The results page also includes the melting temperatures for each of the primers. The FASTA header includes a link to the GB, and opens it at the location of that hit. If you follow this link to the GB, note that unlike the BLAT tool, the PCR tool does not create a custom track.

PCR—Try this

To try the PCR tool for yourself, see Table 7 for a set of primers to paste into the tool on the website. Choose Mouse and Feb 2006 for the **Genome** and **Assembly**. Leave all of the default values in the three numerical boxes. Check the **Flip Reverse Primer** checkbox. Then press the **submit** button. You will see one FASTA file representing the single match to your primer pair search (see Table 8). It is the 746 bp sequence from the mm8 assembly located at chr1:99,605,552–99,606,297. The '+' shown between the start and end position (chr1:99605552+99606297) indicates that the genomic sequence is on the sense strand.

Other tools

There are several other useful tools at the UCSC Genome Bioinformatics website. We mention them briefly here.

Proteome Browser

The Proteome Browser is similar to the Genome Browser tool; however, it displays proteins and associated annotation tracks. The Proteome Browser provides a wealth of protein information presented in the form of graphical images and links to external websites. Navigate to the Proteome Browser tool by pressing the **Proteome Browser** link in the left-hand sidebar menu on the UCSC GB home page. Alternatively, you can open the Proteome Browser to a specific protein directly from the GB. Click on any gene in the UCSC Genes track. From the details page for that gene, click on the Proteome Browser hyperlink in the **Sequence and Links to Tools and Databases** section of the page. This will open the Proteome Browser to the protein for that gene.

VisiGene

The VisiGene tool is a browser for viewing *in situ-hybridization* images. It allows you to examine cell-by-cell as well as tissue-by-tissue expression patterns. Images from mouse and the frog *Xenopus tropicalis* are available directly or as links from mouse and human browser gene-details pages. The tool serves as a virtual microscope,

Table 7

Forward and reverse primers for the mm8 assembly for use in the PCR tool

Forward Primer: TACITCCCTTTCTAGTTTACAG

Reverse Primer: AGTGCTAGTCTCAGTGTATCA

Table 8

FASTA output from the PCR tool for the primers in Table 7

allowing you to retrieve images that meet specific search criteria, and then interactively zoom and scroll across the collection. Navigate to the VisiGene tool by pressing the **VisiGene** link in the left-hand sidebar menu on the UCSC GB home page.

Gene Sorter

The Gene Sorter tool is used to explore gene families and the relationships among genes. This tool displays a table of genes within a selected genome that are related to one another. Several different relationships may be explored such as protein-level homology, similarity of gene expression profiles, gene ontology terms, or genomic proximity. The Gene Sorter supports searches on a variety of terms and phrases, including the gene name, the SwissProt protein name, a GenBank accession, or a word or phrase present in a gene's description. The tool provides several output formats, including a simple tab-delimited format that may be imported into a spreadsheet or a relational database.

Acknowledgments

The project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract Number N01-CO-12400. The content of this publication does not necessarily reflect the views of policies of the Department of Health and Human Service, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The UCSC Genome Bioinformatics website is also funded by grants from the National Human Genome Research Institute (NHGRI) (Grant P41 HG002371), and the Howard Hughes Medical Institute.

(HHMI). We acknowledge the work of the UCSC Genome Bioinformatics technical staff that maintain and enhance the databases and software, the many collaborators who have contributed annotation data to the project, and our loyal users for their feedback and support.

Appendix A. Online resources

UCSC Genome Bioinformatics Website: <http://genome.ucsc.edu>

Frequently Asked Questions: <http://genome.ucsc.edu/FAQ>

GB Discussion List: <https://www.soe.ucsc.edu/pipermail/genome>

Downloadable Data Files: <http://hgdownload.cse.ucsc.edu>

Appendix B. User's guide

Genome Browser: <http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#GetStarted>

<http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#FineTuning>

BLAT: <http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#BLATAlign>

Custom Tracks: <http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#CustomTracks>

Table Browser: <http://genome.ucsc.edu/goldenPath/help/hgTablesHelp.html>

Sessions: <http://genome.ucsc.edu/goldenPath/help/hgSessionHelp.html>

Proteome Browser: <http://genome.ucsc.edu/goldenPath/help/>

pbTracksHelpFiles/pbTracksHelp.shtml
VisiGene: <http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html>

html#VisiGeneHelp

Genome Graphs: <http://genome.ucsc.edu/goldenPath/help/hgGenomeHelp.html>

Gene Sorter: <http://genome.ucsc.edu/goldenPath/help/hgNearHelp.html>

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2008.02.003.

References

- [1] R.M. Kuhn, et al., The UCSC Genome Browser Database: update 2007, Nucleic Acids Res. 35 (2007) D668–D673 [Database issue].
 - [2] A.S. Hinrichs, et al., The UCSC Genome Browser Database: update 2006, Nucleic Acids Res. 34 (2006) D590–D598 [Database issue].
 - [3] W.J. Kent, BLAT—the BLAST-like alignment tool, Genome Res. 12 (4) (2002) 656–664.