

## Accessing HLA Sequencing Data Through the 6ace Database

Roger Horton and Stephan Beck

### 1. Introduction

The chromosome 6 database (6ace) is one of a suite of databases available at the Sanger Centre, which serve the human genome sequencing communities of several chromosomes (1, 6, 20, 22, and X). Data may be retrieved in graphical or textual form using interactive windows and menus or simple command texts. The database and its management system are based on ACEDB and can be accessed via three main routes, a graphics interface, a text interface, and a Web interface. Here, we describe 6ace with particular emphasis on how to access major histocompatibility complex (MHC) and human leukocyte antigen (HLA) associated data.

### 2. Databases

#### 2.1. ACEDB

Initially written as a *Caenorhabditis elegans* DataBase (*1*), ACEDB is an object-based data management system specifically designed for use with genomic data. It is now being used for other genome projects as well as *C. elegans*, including *Drosophila*

From: *Methods in Molecular Biology*, vol. 210: *MHC Protocols*  
Edited by: S. H. Powis and Robert W. Vaughan © Humana Press Inc., Totowa, NJ

*melanogaster*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, and many more.

ACEDB can be used via an Xwindows graphical interface (XACE) (do not confuse with Xace, which is a chromosome-specific database), via TACE (a command line textural interface), or via WEBACE (T. Hubbard, personal communication), which is a Web browser interface. XACE versions allow data to be viewed not only as text fields but also as genetic and physical maps, or by using applications external to ACEDB to view images or sequence alignments. The maps are emulated in the WEBACE versions.

## **2.2. 6ace and MHCDB**

The sequencing effort on human chromosome 6 was initially focused on the MHC, with the data being accumulated in the MHC database, MHCDB (2). These data were later merged in the Chromosome 6 database (3), 6ace, the use of which is described here. Although MHC peptides were formerly included in MHCDB, these are now catalogued in an independent database, MHCPEP, at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia (<http://wehih.wehi.edu.au/mhcpep/>) (4). The 6ace database incorporates data not just from the Sanger Centre Chromosome 6 Project (5,6) but from the whole chromosome 6 community.

## **2.3. 6ace Content**

As there is rapid progress in the sequencing of the human genome, it is difficult to provide an up-to-date picture of the content of any genomic database. **Table 1** provides a snapshot of the situation in early 1998. Tools are available within the database for the user to gain information on the current status of the data at any time.

## **3. Access to 6ace**

There are three means of getting access to 6ace. If you have your own UNIX system, you may copy the database from the Sanger Centre ftp site. If you are a registered user of the computing facili-

**Table 1**  
**Summary of the Content of the 6ace Database**

Class	Subclass	Database content (MHC data in parenthesis)	
		Number	Megabase
Sequence	Genome_sequence	86 (32)	7.53 (1.19)
	- Sequenced_by = SC	59 (13)	6.20 (0.75 ≡ MHC class II)
Locus	Gene	402 (73)	
	Pseudogene	50 (29)	
Allele		662 (662)	
Image		27 (27)	
STS (RH mapped)		2532	
cosmid		499 (264)	
BAC		124 (0)	
PAC		6084 (11)	
YAC		69 (69)	

ties of the Human Genome Mapping Project Resource Centre (HGMP), you may access 6ace on-line. Finally, there is access via the Web to the Sanger Centre, where pages allow interactive use of 6ace. These means of access can be all reached from the Sanger Centre Chromosome 6 Web page (<http://www.sanger.ac.uk/HGP/Chr6/>) by looking under the Chromosome 6 Database entry.

### **3.1. Getting 6ace by ftp from the Sanger Centre**

Releases of the Sanger Centre chromosome-specific databases take place at the beginning of each calendar month with weekly updates. The releases can be obtained from (<ftp://ftp.sanger.ac.uk/pub/human/chr6/RELEASES/>).

The weekly updates are found in ([ftp://ftp.sanger.ac.uk/pub/human/chr6/weekly\\_release/](ftp://ftp.sanger.ac.uk/pub/human/chr6/weekly_release/)).

There are README files in the directories, which give full instruction for installing the databases. Basically, it is required to copy a file of the form:

6ace\_release\_mm-yy.tar.Z

where mm and yy are month and year numbers, respectively. This file needs to be uncompressed (unzipped) and extracted using the UNIX tar command to give a 6ace directory.

The current release of the ACEDB code will also be required. This can be obtained from (<ftp://ftp.sanger.ac.uk/pub/acedb/>). The current version (4.5) of the code is in a directory `ace4_5/`.

### **3.2. Using 6ace at the HGMP**

Details of registration and use of HGMP computing facilities can be obtained at (<http://www.hgmp.mrc.ac.uk/>). If suitable computer facilities are available, this enables the use of the graphical interface to 6ace. The HGMP database currently uses the monthly 6ace releases, but not the weekly updates. The HGMP also organizes training courses in the use of ACEDB.

### **3.3. Web Access to 6ace**

The Sanger Web pages provide both graphical and textural access to 6ace. Major developments of this route of access are currently being made so as to provide better access to Sanger Centre data, in particular with WEBACE.

The WEBACE home page is at (<http://webace.sanger.ac.uk/>) and has a link to the full set of Sanger Centre ACEDB databases including 6ace. The user is then guided through the class structure of 6ace, and, if a suitable browser is being used, both graphics and textural objects can be displayed. This means of access will become increasingly popular with the expansion of applications available.

For those with a knowledge of TACE, the text version of ACEDB, another page ([http://www.sanger.ac.uk/HGP/db\\_query/query.shtml](http://www.sanger.ac.uk/HGP/db_query/query.shtml)) gives the user access to the QUERY THE CHROMOSOME SPECIFIC DATABASES page. Clicking on the “Open access” button leads to a TACE query form. Consult the help pages for a description of how to use TACE.

## 4. Abbreviations and Conventions

In describing the use of 6ace, the following abbreviations and conventions will be used here. In the majority of cases, they apply to the use of XACE.

CMW; Class Menu Window, the first menu window that appears in XACE.

MKS; Main KeySet window generated LMB DC on a Class in the CMW.

LMB, RMB; Left and Right Mouse Button, respectively.

SC, DC; Single and Double Click, respectively.

“text”; Text as it appears on the screen in 6ace.

‘text’; Text to be entered from the keyboard by the user.

↵; The keyboard RETURN key.

The convention within ACEDB (apart from in the CMW and some of its menus) is that **bold text** is a link to further data and if clicked (LMB DC; LMB SC simply highlights the entry) will open up a further window.

Quit; This has two meanings in XACE. In all windows except the CMW, it may be present as a Quit button or may be selected from the RMB pull-down menu. When selected, it closes the window concerned.

In the CMW selected from the RMB menu, it closes and exits ACEDB after the warning: “Do you really want to quit acedb? Yes No”.

In TACE, typing ‘quit’ closes 6ace immediately.

## 5. Structure of ACEDB Databases

The structure of data storage within an ACEDB database can be likened to a tree. The main menu access is via the trunk. This leads

to classes that are the branches of the tree. Within classes, there are a number of objects like twigs on the branches. If the structure is followed through, eventually a tag field is reached, likened to a leaf, which contains an item of data. The skill in using the database is to know your way around this tree structure.

The structure in each class is contained within its Model. Looking at the Models can be of great help when trying to access and interpret data.

It should be remembered that many of the commonly encountered objects belong to subclasses. Thus, finished sequences are put in the `Genome_sequence` subclass, which is part of the `Sequence` class distinguished (or filtered) by having the `Properties` tag “`Genomic_canonical`”. Similarly, `Gene` and `Pseudogene` are subclasses of `Locus`, filtered by the `Type` tag “`Gene`” and “`Pseudogene`”, respectively.

In XACE, a list of classes and subclasses can be accessed by clicking (LMB DC) on the ▼ after “In class:” in the CMW. Model is itself a class that can be displayed and selected from. In TACE, typing “Classes” lists the classes and the number of objects they each contain. The command “Model”, followed by a class name, will show the model structure.

WEBACE restricts the accessible classes to those on the menu lists. Browsing a class from the main Classes page generates a further page listing objects in the selected class. At the foot of this page is a “View...model” clickable button, leading to a display of the model structure.

At the UNIX level the model structures are all contained in the `face/wspec/` directory in the `models.wrm` file, whereas `subclasses.wrm` contains each subclass with its parent class on an “`Is_a_subclass_of`” tag together with the filter used. The appearance of data represented in a graphical form in XACE is determined by the Methods. These Methods can be viewed in XACE by selecting “Query” from the pull down menu found by holding down the RMB in the CMW. Entering ‘find method’ in the yellow “Query:” box lists the Methods. ‘Find Method’ can also be used in TACE even though TACE cannot display graphical objects.

## 6. 6ace in Use

### 6.1. XACE

Perhaps the most familiar version of the database, XACE contains a menu of commonly used classes, presented as a list in a window (CMW). List Items are clickable (LMB DC) and lead to windows containing KeySets (*sic*) of the selected class. The KeySets list entries are in turn clickable (if depicted in bold type) and can eventually lead to the display of the particular object of interest. Pull-down menus are available via the RMB, and windows buttons lead to other options. **Fig. 1** shows the 6ace CMW (first to appear when 6ace is opened) with “Genome\_sequence” selected to give the MKS, and the object E1448 further selected in text mode to show full details of the entry for this cosmid. Note that, when there is a choice, XACE will try to display an object graphically. The text version of E1448 was selected by clicking on the toggle button at the top left of MKS “Show As...:” to select Text. An alternative for an object is to highlight the object name LMB SC and then select “Show as text” on the menu.

LMB SC on “Genome\_sequence” and then entering ‘E1448 ↵’ in the yellow “Search:” box would generate a KeySet with just one entry and open the E1448 map directly.

### 6.2. TACE

In contrast to XACE, TACE provides textural access to ACEDB databases. It uses the query language that is also employed by the graphical version. On UNIX systems, TACE can be opened with the command ‘tace’ if paths have been set correctly. It may also be used via the Sanger Centre Web pages at ([http://www.sanger.ac.uk/HGP/db\\_query/query.shtml](http://www.sanger.ac.uk/HGP/db_query/query.shtml)) as described in **Subheading 3**.

TACE is not immediately user-friendly, in that the operator needs to know some commands to type in order to get started, whereas the XACE version has the advantage of menus and clickable entries in object lists, but, to the initiated, TACE provides a rapid tool for the interrogation of the database. Typing ‘?’ at the TACE prompt (>)

AGEDB 4.6 Human Chromosome 6

Search: \* In Class: ▼ Ready Genome-Sequence Help...

Map Gene

Cluster\_id BAC CEPH\_grid  
 Gene\_Class PAC Grid  
 Image STS Lab\_grids  
 Locus YAC PAC\_grid  
 Motif Pool Polygrid  
 OMIM Sequence

KeySet Author Paper  
 Model Laboratory Person  
 View Journal Url

Global Search: Long Search

Show As.: Text Biblio More Info Quit  
 Query... Select/Modif... Export... Other Help

86 items 1 selected

8RI	dJ12409	dJ299C21	dJ487J7	U89335
A1	dJ130C2	dJ324L9	dJ509L4	U89336
cICR2046	dJ155D22	dJ340C1	dJ514K20	U89337
cICF0811	dJ162C6	dJ341I10	dJ522O2	Z15025
cyR30H3_13B	dJ167A14	dJ344F17	dJ528L19	Z15026
D84401	dJ172K2	dJ359M14	JW19	Z15027
dJ29K1	dJ179P9	dJ365E2	3448	027
dJ50J22	dJ187N21	dJ381E2	F1121	p797a11
dJ66H14	dJ188H10	dJ396A12	HSAF19664	pp412
dJ67M12	dJ193R12	dJ427M4	HSU91328	pm30
dJ76C18	dJ230J17	dJ431A14	HLRG33IP03	pm56
dJ93H18	dJ244F11	dJ436G17	LC11	pm67
dJ93N13	dJ257A7	dJ443E24	mnc0295.com	017
dJ94G16	dJ265J14	dJ451B15	014	018
dJ111M5	dJ271G9	dJ453D15	019A	019
dJ121G13	dJ292F10	dJ467D16	019A014	020

Sequence: E1448

Attach... Quit

```

DNA E1448 27919
DNA_contig E1448
External_refs Clone_type Cosmid
Length 27919
Structure Subsequence -----> 7
Overlaps F1121
DB_info Database Sanger_Finished E1448
EMBL Z80898 HSE1448
DB_annotation EMBL <see below 1>
Origin From_Author Beck Stephan
Finishing_group 33
Species H.sapiens
Human
Chromosome Chr-6
Sequenced_by SC
Status -----> 8
Map 6p21.3ctg With With_cosmid E1448
Cosmid E1448
Gene HLA-DQB1
Analysis_details Analysis_directory "humpub/analysis/projects/Chr_6/E1448/950
930
Seq_contig E1448
Analysis_summary Average_of_fraction 0.43
  
```

will list a help file of the commands and their function. Typing ‘classes’ lists the classes in the database and the number of objects each contains. To get the output obtained in XACE above for E1448 would require just the following two commands (at the > prompt, with output lines beginning //):

```
>find genome_sequence E1448
//Found 1 objects in this class
//1 Active Objects
>show
```

The data generated are of such a volume as to be virtually unreadable. To overcome this, write the data to a file using the command >write E1448.ace and examine your UNIX directory for the output.

### 6.3. WEBACE

Here, the window lists of XACE are replaced by selectable lists in windows within the Web pages. 6ace can be selected by clicking “acedb6” and then the “Browse DB” button. To obtain the E1448 details given above, it is necessary to make selections as follows (as illustrated in **Fig. 2**): “—>Genome\_sequence” below “Sequence” and click the “Browse” button on the Classes page and then select “E1448” in the window generated. There is also the option of Text, Graphics (passive), and Graphics (active). Selecting the first followed by clicking the “show object(s)” button generates a “Sequence: E1448” page with the data as seen in XACE above. Links are shown in hypertext, and where the volume of data is large in a tag field, it may be in a compressed form beneath an “Expand” hypertext.

---

Fig. 1. (*opposite page*) Text display of Sequence E1448 in 6ace (XACE). Class Menu Window (*top left*) with the class “Genome\_Sequence” selected; the Main KeySet for Genome\_Sequence (*top right*) with the sequence name E1448 selected, and the “Show As...:” button toggled to Text; and the resulting Sequence: E1448 text window showing some of the database entries for the sequence.

[Data release policy and guidelines and conditions on use of data](#)  


[Home](#) | [Help](#) | [Projects](#) | [Database Searches](#) | [Software](#) | [Tutorials](#)  
 Netscape: aeddb6: classes

**The Sanger Centre**

Netscape: aeddb6: classes

**Webcam - AceDB on the web**

**Please select a database:**

- acedb1
- acedb20
- acedb22
- acedb6**
- acedbx
- cgcacc
- duel
- exprace
- humace
- maldb

Netscape: aeddb6: Sequence: E1448

**Sequence : E1448**

DNA E1448 27919  
 DNA\_contig E1448  
 External\_refs Clone\_type Cosmid  
 Length 27919

**Classes**

- Laboratory (25)
- Locus (1903)
- > Gene
- Map (217)
- Motif (187)
- OMIM (252)
- PAC (553028)
- Pool (303)
- Probe (29)
- STS (6138)
- > STRP
- Sequence (62901)
- > Genome\_Sequence
- > cDNA\_Sequence
- YAC (88983)

**Genome\_Sequence**

- dJ1487J7
- dJ509L4
- dJ514K20
- dJ522O2
- dJ528L19
- DV19
- E1448**
- F1121
- HSAF19664
- HSU91328
- H\_RG33IP03
- LC11
- mhc0295.com
- O14
- O19A

- Graphics (passive)
- Graphics (active)

## 7. Examples

It is impossible here to give a complete set of examples of all the functions of 6ace, but the user is encouraged to explore the databases to gain experience. A novice user need not fear causing damage, as only those with write access can initiate changes. Only when users are running a local copy of the database and this is opened with TACE are they likely to have write access without specifically asking for it.

The MHC, also known as the HLA complex, is located on the short arm of human chromosome 6 (6p21.3). It is the most polymorphic and gene dense region of the human genome. The examples selected here, therefore, have been chosen to illustrate different types of variation and polymorphism present in HLA sequences.

### 7.1. Maps and Images in XACE

This example uses maps and images to illustrate the representation of variation in 6ace. A variation is either a substitution, insertion, or deletion of bases in the DNA sequence. XACE has the advantage over TACE of being able to display graphical objects such as the ACEDB genetic and physical maps. As WEBACE emulates XACE, it is also able to display these maps. Selecting “Maps” from the CMW lists all the genetic maps in 6ace. Of particular relevance to those with an interest in the HLA region will be “Chr\_6” (the map of the entire chromosome) and “6p21\_3ctg” (the map of the MHC region). These appear superimposed in **Fig. 3**. This shows, from left to right, the chromosome 6 ideogram flanked by a scale, in kb (kilobases), and the chromosome bands. The boxes (red) along the ideogram indicate regions selected for sequencing or where sequencing is already in progress. (Note: all colors given here are

---

Fig. 2. (*opposite page*) Text display of Sequence E1448 in 6ace (WEBACE). The three pages used to select the 6ace database (acedb6; top left), the Genome\_Sequence class (top center), and the E1448 sequence text (top right), together with the resulting Sequence: E1448 text page (bottom).



the default colors generally used in the XACE and WEBACE versions of 6ace. In XACE, it is possible for the user to alter these colors). The box associated with the MHC region (6p21\_3ctg) has been selected using LMB SC. This selection changes the box color to pale blue, and the contig name appears highlighted in blue. The zoomed in view of 6p21\_3ctg shows the status of the 4000 kb MHC region. To the left of the scale, the approximate positions and sizes of all yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC), P1-based artificial chromosome (PAC), and cosmid clones are shown. A LMB SC on any clone or clone name will highlight the corresponding clone name or clone in pale blue. To the right of the scale the sequencing tiling path and corresponding clone names are shown. The key to the color-coded status of each clone can be viewed by LMB DC on the Status bar (green) at the foot of the map (grey, sequencing in progress; red, sequencing finished; black, analyzed and submitted to a public database). The mirrored tiling path to the right indicates where a particular clone is being or has been sequenced (pale blue, Sanger Centre; purple, others). LMB DC on any clone will bring up a text window with full details of its sequencing and with links to sequencing laboratory location and names of investigators involved. Finally, on the far right hand side, the names of genes are displayed alongside the clones in which they are encoded. The limitation of space in the magnification shown here results in the truncation or exclusion of some gene names. This may easily be overcome by repeated LMB SC on the “Zoom in” button until the required magnification is displayed. During zooming, the middle mouse button can be used to center on the object to be enlarged by SC on it or to scroll by SC at the top or bottom of the window.

Switching between different maps is very easy. The text window produced above (LMB DC on any clone) will contain a tag “DNA”

---

Fig. 3. (*opposite page*) The 6ace Chromosome 6 (*left*) and MHC (*right*) maps. The position of the latter (6p21\_3ctg) is highlighted in the former. For further details see text.



followed by the clone name in bold text. This is a link to the physical map of the clone DNA sequence. LMB DC on this bold text brings up a map as shown in **Fig. 4** (main picture), in this case for clone E1448. On the left is a locator bar (light red when whole map is viewed, but changing to green when zoomed in) next to a box representing the clone (black outline; LMB DC on this box returns to the sequence text window). Names of genes within the clone are superimposed on the locator bar, as is that for *HLA-DQB1* here. To the right is a vertical bar (yellow) and scale in bases for the sequence in view. All the other details in the map relate to features of the DNA annotated using the HPREP analysis suite (G. Micklem and R. Durdin, personal communication). Those to the left of the yellow bar and scale represent features of the reverse complement, whereas those to its right represent features of the positive strand. The features include significant matches determined using Basic Local Alignment Search Tool (BLAST) (7) to entries in the protein databases (BLASTX; light blue boxes), to entries in the nucleotide databases (BLASTN; yellow boxes), and to vertebrate mRNA (brown boxes). At the far right, similarity to expressed sequence tag (EST) clusters are shown as pale violet boxes. Repeat sequences determined using RepeatMasker2 (8) (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>) are also shown (Alus as green boxes, others in dark blue). Further information can be gained for all of the features, either by LMB SC on a box that brings up brief details in the banner line (light blue) at the top of the map, or by LMB DC that will call up a text window giving fuller details.

At the foot of the map, immediately to the right of the scale, is an exon and intron diagram for the *HLA-DQB1* gene (dark blue outline boxes and lines). The strength of ACEDB, in its ability to present a

---

Fig. 4. Representation of variation in 6ace. The sequence map of E1448 (*top left*) showing annotation of substitutions and of indels (for full description, *see text*); the DQ-pocket Image with residues color-coded according to variability (*inset, right*); a Blixem output of aligned *HLA-DQB1* allele sequences (*bottom*) compared with that of *HLA-DQB1\*0402*.

combination of results visually, is illustrated by the concentration of features showing matches to the protein and DNA databases and the absence of repeat elements in the region of this gene. The presence of allele sequences is recorded as boxes (red) to the right of, and at the same horizontal level, as the gene exons. Holding down RMB on these allele-display boxes reveals a pull-down menu from which may be chosen “show multiple DNA alignment of just this kind of homologies”. This selects an external application, Blixem (9) which enables the display of aligned DNA sequences of the gene’s alleles (Fig. 4, foot). This application highlights the allelic variation in this gene. The allele sequences of HLA class I genes were taken from Arnett and Parham (10) and those of the class II region from Marsh and Bodmer (11).

Central features of the map are two horizontal bar graphs representing GC content and variation rate. The former (red) shows GC content calculated for intervals of 250 bases on a scale arranged around a midpoint of 50%. The latter shows the variation rate, using the same intervals, between E1448 and an overlapping sequence from a different haplotype (p797a11). The GC content is highest in the sequence adjacent to the second exon in DQB1, which coincides with the presence of a CpG island (yellow box) directly to the left of the bar graph. It is also of interest to note that the variation reaches local maxima of up to 10% (which is the highest level of sequence variation reported in the human genome so far) and that it extends far beyond the transcriptional unit of DQB1 (12). The three gaps in the variation bar graph are due to major indels (cerise boxes), which, in this case, are all of retroviral origin (long terminal repeat [LTR] sequences). The relatively high GC content around the gene is also emphasized. The nature of individual variations are also shown as text, far right, and as small boxes (red) immediately to the left of the gene diagram, where they often form a continuous bar because they are so numerous.

Although all these features can be seen in text form in TACE, only the maps in XACE and WEBACE allow their spacial visualization.

6ace via XACE also provides a means to view the structures and structural variability of HLA molecules using either xv (John Brad-

ley; ftp://ftp.cis.upenn.edu) or RasMol (*13*) if these applications are installed on the computer system in use. With the former, images constructed using Prepi suite (*14*), in which residues are color-coded according to their variability from constant (white) to most variable (red) according to their Shannon entropy, (*15*) can be displayed. With the latter, a 3D interactive display enables the rotation of images of protein tertiary structure.

To access these, LMB DC on the class Image in the CMW and then select from the MKS in which external image files are listed. Select (LMB DC) from the MKS and in the Image object window LMB DC on “Pick\_me\_to\_call”. Selecting, thus, for the “DQ\_pocket” image calls the structure shown as an inset in **Fig. 4**, in which the positions within the structure of the DQ antigen, which are most variable, are banded in red.

In summary, **Fig. 4** illustrates the different ways in which HLA variation and polymorphism data can be accessed and viewed in 6ace.

## **7.2. Obtaining the DNA Sequences of All HLA-DQB1 Alleles**

Obtaining DNA sequences in a format compatible with analysis software is a common problem in bioinformatics. The 6ace database contains allele sequences for genes in the MHC region imported from external databases (*10,11*). This example shows how to obtain the sequences for all the alleles of an HLA gene in fasta format.

### **7.2.1. XACE**

In the CMW enter HLA-DQB1\* in the yellow “Search:” box and LMB DC on ▼ after “In class:”. From the menu window, which appears LMB DC on Allele, to bring up a MKS of all *HLA-DQB1* alleles. These are, of course, objects in the Allele class. What we want are the Sequence objects associated with them. To get a KeySet of these, LMB SC on “Query...” in the MKS, and in the resulting “Query:” yellow box type ‘follow sequence ↵’. A further KeySet

window will appear with the Sequences listed. RMB on the “Export...” button in this window gives a menu from which select “DNA in FASTA format”. Enter a file name in the yellow “File:” box (note that ACEDB will add the extension .dna to this), check that the green “Directory:” box contains an entry you are happy to write to, and press ↵. You should get a message similar to “# I wrote 28 sequences Continue”. Click “Continue”. Examine your UNIX directory for a file containing all the allele DNA sequences in FASTA format beginning as follows:

```
>HLA-DQB1*0201
agagactctcccgaggatttcgtgtaccagtttaagggcatgtgctactt
caccaacgggacagagcgcgtgctcttgtgagcagaagcatctataacc
gagaagagatcgtgcgcttcgacagcgacgt.....
```

### 7.2.2. TACE

In TACE, the following session leads to the writing of a file, `dqb1.dna`, containing all the allele sequences:

```
> find allelele HLA-DQB1*
// Found 34 objects in this class
// 34 Active Objects
> follow sequence
// Autocompleting sequence to Sequence
// Found 28 objects
// 28 Active Objects
> dna dqb1.dna
// 28 object dumped// 28 Active Objects
>
```

Note that although 34 alleles are listed, only 28 appear to have a DNA sequence attached.

### 7.2.3. WEBACE

This facility is not yet available in WEBACE.

## 8. Conclusion

The examples above give an idea of the variety of HLA data available in the 6ace database, but there is much that has not been described, such as the ability to query the database and to use analysis and gene finding tools, to which the user may progress with experience. Those working on the HLA region may also want to take advantage of other databases, such as IMGT (ImMunoGeneTics; <http://www.ebi.ac.uk/imgt/>) a database of nucleotide sequences of important genes of the immune system, or the HLA Peptide Binding Prediction database (**16**) ([http://bimas.dcrtnih.gov/molbio/hla\\_bind/](http://bimas.dcrtnih.gov/molbio/hla_bind/)). There is a wealth of information available to those willing to explore.

## Acknowledgments

The writers would like to acknowledge the contribution made to the 6ace database by their colleagues at the Sanger Centre and, in particular, to the members of the Chromosome 6 Project group (<http://www.sanger.ac.uk/HGP/Chr6/>) and also to the wider community of those working on the chromosome and in the MHC region for free access to the data now included in the database. The curation of the 6ace database is funded by the Wellcome Trust.

## References

1. Durbin, R. and Thierry-Mieg, J. (1994) The ACEDB genome database, in *Computational Methods in Genome Research* (Suhai, S., ed.), Plenum Press, New York, pp. 45–55.
2. Newell, W. R., Trowsdale, J., and Beck, S. (1996) MHCDB database of the human MHC (release 2). *Immunogenetics* **45**, 6–8.
3. Theaker, A. J., Maslen, G. L., Scott, C. E., et al. (1997) The chromosome 6 database at the Sanger Centre. *DNA Seq.* **8**, 167–172.
4. Brusica, V., Rudy, G., Kyne, A. P., and Harrison, L. C. (1998)

- MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* **26**, 368–371.
5. Avis, T., Clark, E. K., Flack, T. L., et al. (1997) The chromosome 6 sequencing project at the Sanger Centre. *DNA Seq.* **8**, 131–136.
  6. Mungall, A. J., Humphray, S. J., Randby, S. A., et al. (1997) From long range mapping to sequence-ready contigs on human chromosome 6. *DNA Seq.* **8**, 113–154.
  7. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
  8. Smit, A. F. A. and Green, P., unpublished results.
  9. Sonnhammer, E. L. L., and Durbin, R. (1994) A workbench for large scale sequence homology analysis. *Comput. Appl. Biosci.* **10**, 301–307.
  10. Arnett, K. L., and Parham, P. (1995) HLA Class I nucleotide sequences. *Tissue Antigens* **46**, 217–257.
  11. Marsh, S. G. E. and Bodmer, J. G. (1995) HLA Class II region nucleotide sequences. *Tissue Antigens* **46**, 258–280.
  12. Horton, R., Niblett, D., Milne, S., et al. (1998) Large-scale sequence comparisons reveal unusually high levels of variation in the *HLA-DQB1* locus in the class II region of the human MHC. *J. Mol. Biol.* **282**, 71–97.
  13. Sayle, R. A. and Milnerwhite, E. J. (1995) Rasmol-biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–376.
  14. Islam, S. and Sternberg, M. J. E., unpublished results (<http://www/sbg.ic.ac.uk/prepi/>).
  15. Shannon, C. E. and Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
  16. Parker, K. C., Bednarek, M. A., and Coligan, J. E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* **152**, 163–175.



<http://www.springer.com/978-0-89603-548-5>

MHC Protocols

Powis, S.H.; Vaughan, R.W. (Eds.)

2003, XII, 340 p., Hardcover

ISBN: 978-0-89603-548-5

A product of Humana Press