

# The Radiation Hybrid database

Patricia Rodriguez-Tomé\* and Philip Lijnzaad

EMBL Outstation, The European Bioinformatics Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Received September 18, 1998; Accepted September 22, 1998

## ABSTRACT

Since July 1995, the European Bioinformatics Institute (EBI) has maintained the Radiation Hybrid database (RHdb; <http://www.ebi.ac.uk/RHdb/>), a public database for radiation hybrid data. Radiation hybrid mapping is an important technique for determining high resolution maps. Recently, CORBA access has been added to RHdb. The EBI is an Outstation of the European Molecular Biology Laboratory (EMBL).

## INTRODUCTION

The radiation hybrid mapping technique (1,2) is a method for ordering markers along a chromosome, and gives estimates of physical distances between them. Radiation hybrids are produced by fusing irradiated donor cells with recipient rodent cells. These hybrid cell lines are grouped in so-called panels of clones, each containing different sets of chromosome fragments produced by radiation-induced breakage. The clones are screened by PCR amplification (producing 'scoring data') to establish the presence or absence of a given marker. Nearby loci will tend to show similar retention patterns, the so-called score vectors. Using these results the proximity can be calculated based on a statistical model. The quality measure of marker positions is expressed as a likelihood or LOD score (3,4).

Radiation hybrid methods can be used to map non-polymorphic markers such as sequence tagged sites (STS). Expressed sequence tags (ESTs) are particularly attractive in this respect and are used frequently.

## THE RADIATION HYBRID DATABASE (RHdb)

RHdb is a repository of raw data relevant to radiation hybrid mapping. It was set up in 1995 to support a group of European and US genome mapping laboratories. RHdb stores data on panels, experimental conditions, STSs and experimental results of assays. Maps are kept too, as well as authoring and bibliographic information. Extensive cross-referencing to other databases is another important aspect of RHdb. The database is species independent and contains currently human (since 1995), mouse (since 1997) and rat scores data (since 1998).

The mouse data in RHdb is a collaboration between RHdb at EBI and the Jackson Laboratory Mapping Panels. Both sites are collectors of RH data. Data are exchanged systematically, which means that both databases contain the same set. Nevertheless,

submission procedures are different and users can choose the process most suited to their current needs.

RHdb can be accessed on the World Wide Web at: <http://www.ebi.ac.uk/RHdb/>. This page provides information about the database and access to reports and query tools. The first release (July 1995) contained 1115 assay entries; release 6 (August 1996) contained 28 516 entries, while the most recent release (release 13, September 1998) contains 85 974 entries, showing that growth is substantial.

## Information technology

RHdb is stored and maintained in the relational database management system (RDBMS) ORACLE.

It is important to have a correct data model of the data at hand. The model should be rich, yet simple. It makes the understanding, the querying and the maintenance easier. We have used the object oriented design tool Rational Rose to model our data. The model is described at:

[http://www.ebi.ac.uk/RHdb/SCHEMA/RHdb\\_object.html](http://www.ebi.ac.uk/RHdb/SCHEMA/RHdb_object.html)

## The data

The primary type of data are the assay entries: the scoring results of the PCR amplification of a particular STS (with given primers) on a particular panel by a particular laboratory. Assay entries are given an accession number of the form RH*n*, which is a permanent unique identifier.

There are three other primary types of entries in the database: (i) panels, with information about the authors, distributors and the clones. The panel name serves as the 'accession number'; (ii) maps, with the order, position and LOD scores of assays. They have an accession number of the form CM*n*; and (iii) PCR experimental conditions in free text format. Accession numbers are of the form EI*n*.

An RH entry also includes other types of information: (i) author/laboratory identification (of form CI*n*); (ii) bibliographic information (of form PI*n*), if available; and (iii) a 'flag' describing the type of STS used (EST, genetic marker, CpG island, etc.). These classifications of STSs enable the query of subsets of the data.

Cross-references to other databases are systematically added when they are given by the submitter, or when they can be inferred. There are now over 274 000 cross-references, to the following databases: ATCC; CGM-WUSM; CHLC; GDB; Généthon; Genexpress; IMAGE; KDRI; NHGRI; PAGE; RHalloc;

\*To whom correspondence should be addressed. Tel: +44 1223 494409; Fax: +44 1223 494468; Email: [rhdb@ebi.ac.uk](mailto:rhdb@ebi.ac.uk)

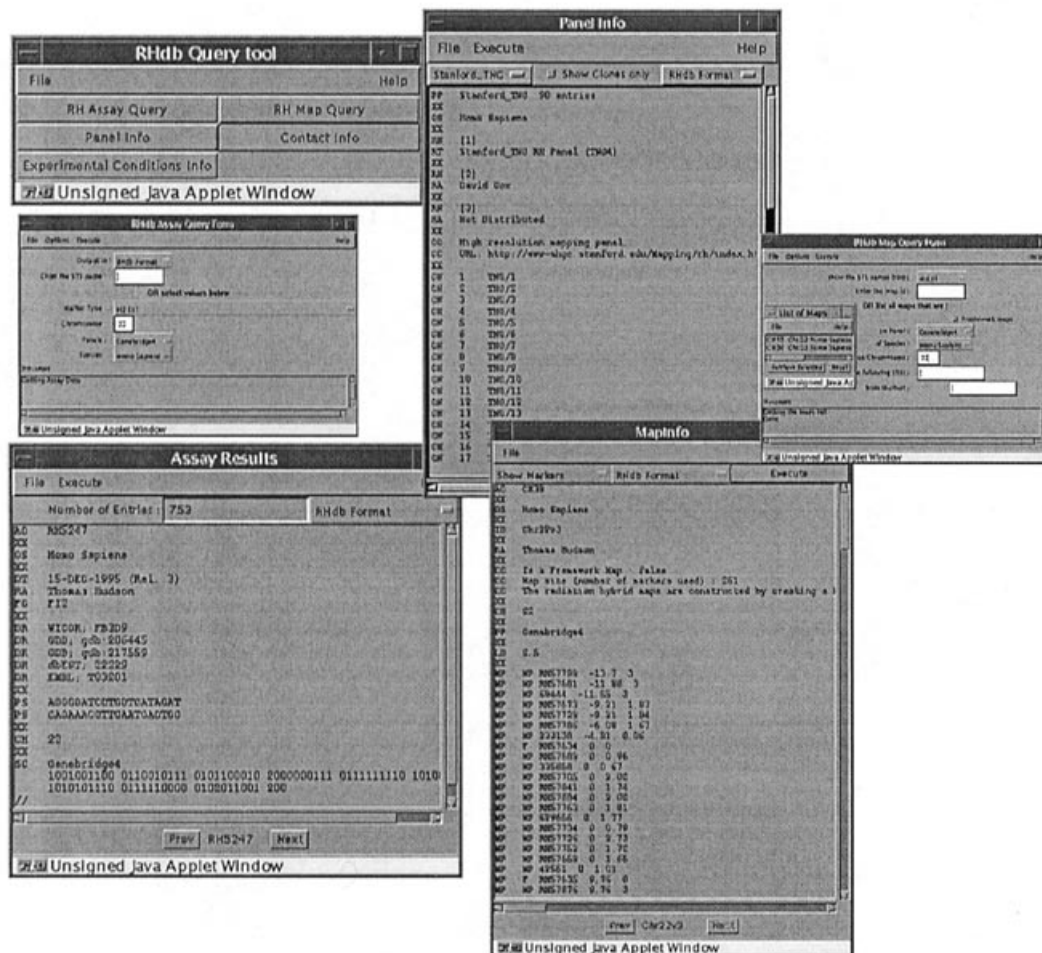


Figure 1. The RHdb browser applet at <http://corba.ebi.ac.uk/RHdb/Clients/RHdbClient>

RHdb; SALK; SangerSTS; TIGR; UCHSC; UT; WICGR; WTCHG; dbEST; dbSTS, Unigene (for Human data), (EBI) ClusterDB.

Details and statistics concerning the data with regard to marker types, chromosomes and database cross-references can be found at [http://www.ebi.ac.uk/RHdb/STATS/rhdb\\_stat.html](http://www.ebi.ac.uk/RHdb/STATS/rhdb_stat.html). The Unigene cross-references are not exported to the flat files release. These references being, by nature of that database, not persistent.

### Data submission

The primary intent of RHdb is as a public repository of data relevant to the (re)construction of maps. Since experimental results often come in large batches, the submission process is largely automatic and able to handle large quantities of data. For flat file submissions, a tagged-field format is used. A full description of the formats can be found at [http://www.ebi.ac.uk/RHdb/rh\\_formats.html](http://www.ebi.ac.uk/RHdb/rh_formats.html). A syntax verification program for this format is provided on the EBI anonymous FTP server at [ftp://ftp.ebi.ac.uk/databases/RHdb/softs/rh\\_submit.tar.gz](ftp://ftp.ebi.ac.uk/databases/RHdb/softs/rh_submit.tar.gz)

Entries should be submitted by e-mail to [rhdb@ebi.ac.uk](mailto:rhdb@ebi.ac.uk).

After syntax checking, the data is subjected to additional tests: (i) the number of scores is ascertained to be equal to the size of the panel; (ii) species information is verified with that in the

EMBL/GenBank/DDBJ sequence database (5,6) by using the mandatory cross-references to this database; (iii) similarly, the primers are checked with the actual sequence; and (iv) cross-references between RHdb entries are added.

### Data access

All data types (RH assays, experimental conditions, panels and maps) can be retrieved as ASCII files in a tagged field format described at [http://www.ebi.ac.uk/RHdb/rh\\_formats.html](http://www.ebi.ac.uk/RHdb/rh_formats.html). The files are available on the EBI anonymous FTP server at <ftp://ftp.ebi.ac.uk/databases/RHdb> with the filenames. Incremental updates are made available in the same directory.

### Data query/retrieval

Traditional WWW forms give the user simple query access. It is possible using these forms and the following tools to query RHdb using Unigene and other EST Cluster databases, ids.

More sophisticated access is available by using Java applets which are clients to the RHdb CORBA server. These clients can be found at <http://corba.ebi.ac.uk/RHdb/Clients/>

Figure 1 has an example of the windows available in the browsing client.

```
//...
module Scores {
  typedef sequence <string> Strings;
  struct ScoreDatum { // defines a data type
    string rhassay; // the rhid
    string sts;
    string score;
  };
  typedef sequence <ScoreDatum> ScoreData;
  // ...
  interface ScoresTrader : Traders::Trader { // has all the methods
    ScoreData getTheseScoresData(in Strings ids) raises (RHdb::RHNotFound);
    ScoreData getScoresDataInterval(in string mapID, in string fromMarker,
                                     in string toMarker) raises (ScoresException);
  };
  // ...
  // ...
};
```

Figure 2. A fragment of the IDL. The full specification can be found at <http://corba.ebi.ac.uk/RHdb/Servers/IDL>

The EBI has adopted CORBA as their future medium for the serving of data. CORBA is the Common Object Request Broker Architecture. It was developed by a large consortium of software industries as a standard for distributed object-oriented computing (7).

The central elements of a CORBA service are the server (which implements the functionality) and the ORB (Object Request Broker; it provides the communication). Together, they serve data and execute operations (e.g., retrieving data from a database) in response to requests made by client programs.

The facilities offered by a CORBA server are published in its specification, which is written in IDL (Interface Definition Language). The server implements the objects, which can be used by client programs in any way desired.

A fragment of the IDL specification that is implemented to serve RHdb is shown in Figure 2. For example, the query method

getTheseScoresData(in Strings ids) is an operation of the class (interface) ScoresTrader. The parameters are input parameters only (IDL: in), and are used to specify queries. The method's return type is a list (IDL: sequence) of ScoreDatum, defined earlier in the fragment. In addition, getTheseScoresData(in Strings ids) can raise run-time exceptions of the type RHNotFound (not shown).

A client Java applet that uses this server is shown in Figure 3. It can be found at:

<http://corba.ebi.ac.uk/RHdb/Clients/RHScoresFormat.html> (8).

Detailed information on the server can be had from:

<http://corba.ebi.ac.uk/RHdb/Servers/>

World Wide Web

Reports are automatically generated every night. They can be accessed through the EBI WWW server at [http://www.ebi.ac.uk/RHdb/rh\\_reports.html](http://www.ebi.ac.uk/RHdb/rh_reports.html)

Future developments

CORBA is starting to deliver machine-independent, language-independent, Internet-accessible objects. Although a number of database specific issues have to be resolved, the prospects of this medium are very good.

An exciting application is the transparent integration of heterogeneous, distributed databases. By using CORBA's class inheritance, the aspects of the data that are common to the databases can be abstracted into a superclass, which must be implemented by all databases. The database-specific details can be implemented in subclasses specific to the particular database (typically, the local database).

We are currently investigating this approach together with Infobiogen in France (9). The end goal is transparent database

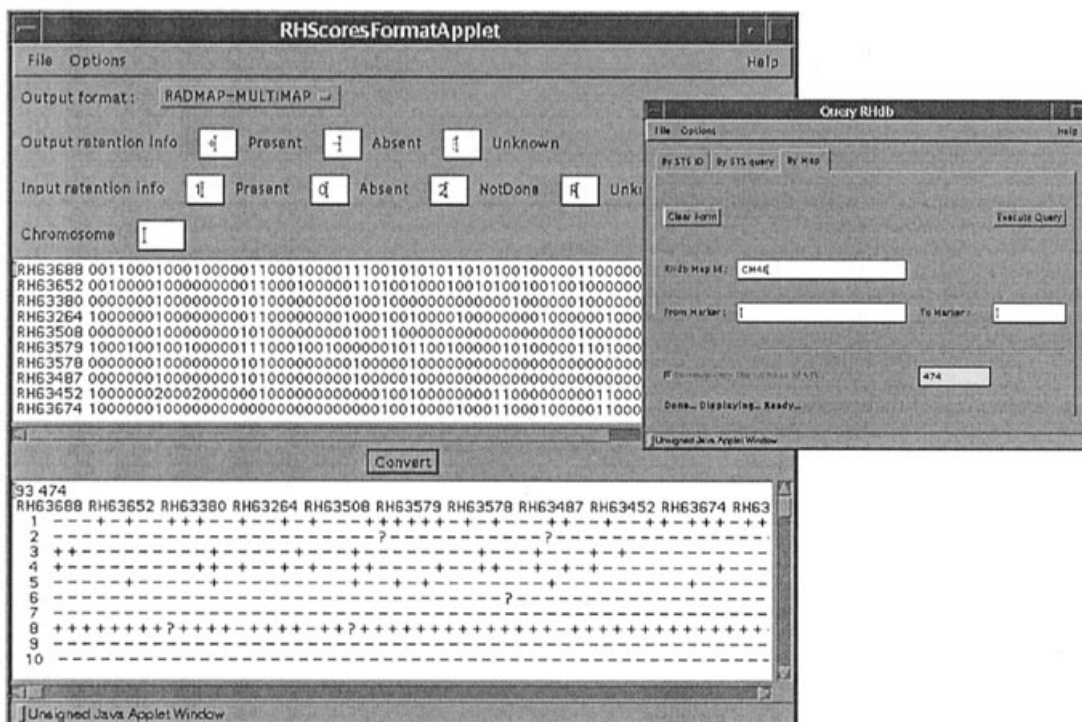


Figure 3. A Java applet using the IDL presented in Figure 2. After retrieving the scores data, it can be reformatted to use with one of the RH mapping programs.

access to chromosome maps, be they radiation hybrid maps, genetic, cytogenetic or physical maps. We are also active in the Object Management Group (OMG) Life Science Research (LSR) Task Force, chairing a working group on maps.

## HOW TO CONTACT RHdb

### Internet

home page: <http://www.ebi.ac.uk/RHdb>

FTP server: <ftp://ftp.ebi.ac.uk/pub/databases/RHdb>

Email: [rhdb@ebi.ac.uk](mailto:rhdb@ebi.ac.uk) (enquiries and submissions)

### Postal address

RHdb, EMBL Outstation—the EBI, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. Fax: +44 (0)1223 494468.

## ACKNOWLEDGEMENTS

We would like to thank P. Deloukas (Sanger Centre, Hinxton) and M. James (Wellcome Trust Centre for Human Genetics, Oxford) for their continuing support.

## REFERENCES

- 1 Cox,D.R., Burmeister,E., Price,R., Kim,S. and Myers R.M. (1990) *Science*, **250**, 245–250.
- 2 Walter,M.A., Spillet,D.J., Thomas,P., Weissenbach,J. and Goodfellow,P.N. (1994) *Nature Genet.*, **7**, 22–28.
- 3 Hudson,T.J., Stein,L.D., Gerety,S.S., Map,J., Castle,A., Silva,J., Slonim,D.K., Batista,R., Kruglyak,L., Xu,S.H., *et al.* (1995) *Science*, **270**, 1945–1954.
- 4 Gyapay,G., Schmitt,K., Fizames,C., Jones,H., Vega-Czarny,N., Spillet,D., Muselet,D., Prud'Homme,J.F., Dib,C., Auffray,C., *et al.* (1996) *Hum. Mol. Genet.*, **5**, 339–346.
- 5 Stoesser,G., Sterk,P., Tuli,M.-A., Stoehr,P.J. and Cameron,G. (1997) *Nucleic Acids Res.*, **25**, 7–13.
- 6 Benson,D., Lipman,D.J. and Ostell,J. (1994) *Nucleic Acids Res.*, **22**, 3441–3444.
- 7 <http://www.omg.org/>
- 8 Rodriguez-Tomé,P., Madsen,R., Andersen,H. and Helgesen,C. (1998) *Trends Genet.*, in press.
- 9 <http://www.infobiogen.fr/index-english.html>