

# RHdb: the Radiation Hybrid database

Patricia Rodriguez-Tomé\* and Philip Lijnzaad

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

Received September 18, 2000; Accepted September 27, 2000

## ABSTRACT

Since July 1995, the European Bioinformatics Institute (EBI) has maintained 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. RHdb is also served by CORBA servers. 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 is a repository of raw data relevant to radiation hybrid mapping. RHdb stores data on panels, experimental conditions, STSs and experimental results of assays. Maps are also kept, 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 currently contains human, mouse and rat scores data. Release 18 (September 2000) contained 126 540 RH entries (for 100 285 unique STSs). 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.

The Radiation Hybrid database 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.

## Information technology

RHdb is stored and maintained in the relational database management system (RDBMS) ORACLE 8. The model used to describe the data stored in RHdb can be accessed at [http://corba.ebi.ac.uk/RHdb/schema/RHdb\\_object.html](http://corba.ebi.ac.uk/RHdb/schema/RHdb_object.html). It is important to have a correct model of the data stored. This model should be semantically rich, yet simple, in order to facilitate understanding, querying and maintenance of the database. We have used the object-oriented design tool Rational Rose to model our data; the model is then mapped onto the underlying RDBMS.

## 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 and 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*
- (iii) PCR experimental conditions in free text format. Accession numbers are of the form EI*n*.

All accession numbers in RHdb are unique. They are never reused, even if data are deleted.

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.
- (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 342 700 cross-references to related databases.

Details and statistics concerning the data with regard to marker types, chromosomes and database cross-references are built every night for each species in the database.

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

### 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://corba.ebi.ac.uk/RHdb/howto/howto\\_submit.html](http://corba.ebi.ac.uk/RHdb/howto/howto_submit.html). Entries should be submitted by e-mail to [rhdb@ebi.ac.uk](mailto:rhdb@ebi.ac.uk). A web form is also available to help users with small amount of data ([http://corba.ebi.ac.uk/RHdb/submission\\_form.html](http://corba.ebi.ac.uk/RHdb/submission_form.html)).

After syntax checking, the data are 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.
- (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://corba.ebi.ac.uk/RHdb/formats/rh\\_formats.html](http://corba.ebi.ac.uk/RHdb/formats/rh_formats.html). The files are available on the EBI anonymous FTP server at <ftp://ftp.ebi.ac.uk/pub/databases/RHdb>. Releases are made on a regular basis. Between two releases, incremental updates are made available in the same directory.

### Data query/retrieval

Statistics on the data are generated every night and are available from the RHdb home page. Traditional WWW forms give the user simple query access (<http://corba.ebi.ac.uk/RHdb/queries/>). More sophisticated access is available by using Java applets which are clients to the RHdb CORBA server (7). These clients can be found at <http://corba.ebi.ac.uk/RHdb/Clients/>. The CORBA server gives access to the full and current database.

The EBI has adopted CORBA as its 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 (<http://www.omg.org/>). The central elements of a CORBA service are the server (which implements the functionality) and the ORB (the Object Request Broker which 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. All information about the RHdb CORBA servers and source code example is available at [http://corba.ebi.ac.uk/RHdb/RHdb\\_CORBA/](http://corba.ebi.ac.uk/RHdb/RHdb_CORBA/).

### How to contact RHdb at the European Bioinformatics Institute

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 1223 494468

### REFERENCES

1. Cox,D.R., Burmeister,E., Price,R., Kim,S. and Myers R.M. (1990) Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, **250**, 245–250.
2. Walter,M.A., Spillet,D.J., Thomas,P., Weissenbach,J. and Goodfellow,P.N. (1994) A method for constructing radiation hybrids of whole genomes. *Nature Genet.*, **7**, 22–28.
3. Hudson,T.J., Stein,L.D., Gerety,S.S., Ma,J., Castle,A.B., Silva,J., Slonim,D.K., Baptista,R., Kruglyak,L., Xu,S.-H. *et al.* (1995) An STS-based map of the human genome. *Science*, **270**, 1945–1954.
4. Gyapey,G., Schmitt,K., Fizames,C., Jones,H., Vega-Czarny,N., Spillet,D., Museler,D., Prud'Homme,J.-F., Dib,C. and Auffray,C. (1996) A radiation hybrid map of the human genome. *Hum. Mol. Genet.*, **5**, 339–346.
5. Baker,W., van den Broek,A., Camon,E., Hingamp,P., Sterk,P., Stoesser,G. and Tuli,M.-A. (2000) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **28**, 19–23. Updated article in this issue: *Nucleic Acids Res.* (2001), **29**, 17–21.
6. Benson,D., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18.
7. Rodriguez-Tomé,P., Helgesen,C., Lijnzaad,P. and Jungfer,K. (1997) A CORBA Server for the Radiation Hybrid Database. *Proceedings for the 1997 ISMB Conference*. AAAI Press, pp. 250–253.