RatMap—rat genome tools and data

Greta Petersen[1], Per Johnson[1], Lars Andersson[1], Karin Klinga-Levan[2], Pedro M. Gómez-Fabre[1] and Fredrik Ståhl[1,3,*]

[1]Department of Cell and Molecular Biology–Genetics, Göteborg University, Box 462, SE 40530 Göteborg, Sweden, [2]School of Natural Sciences, University of Skövde, Box 408, SE 54128 Skövde, Sweden and [3]School of Health Sciences, University College of Borås, Sweden

ABSTRACT

The rat genome database RatMap (http://ratmap.org or http://ratmap.gen.gu.se) has been one of the main resources for rat genome information since 1994. The database is maintained by CMB–Genetics at Göteborg University in Sweden and provides information on rat genes, polymorphic rat DNA-markers and rat quantitative trait loci (QTLs), all curated at RatMap. The database is under the supervision of the Rat Gene and Nomenclature Committee (RGNC); thus much attention is paid to rat gene nomenclature. RatMap presents information on rat idiograms, karyotypes and provides a unified presentation of the rat genome sequence and integrated rat linkage maps. A set of tools is also available to facilitate the identification and characterization of rat QTLs, as well as the estimation of exon/intron number and sizes in individual rat genes. Furthermore, comparative gene maps of rat in regard to mouse and human are provided.

INTRODUCTION

The rat has for many years been one of the most frequently used model organisms for physiological and pharmacological research (1–3). A large number of inbred rat strains display a variety of complex disorders common in human populations, such as arthritis, hypertension and cancer (4). Factors such as inherent genotype variability and different lifestyles make the identification of the genes responsible for these disorders difficult to study in humans. These circumstances have made the rat a very attractive model for analyzing complex diseases (5). Using rat genome information provides favorable opportunities to find the genes contributing to these polygenic disorders in humans; the starting point usually being a quantitative trait loci (QTLs) study (6).

In order to enable efficient use of the rat as a model organism, the rat genome database RatMap was introduced in 1994 by CMB–Genetics at Göteborg University, Sweden. The primary goal was to collect and present all well-defined rat genes localized to chromosomes. Since then, RatMap has been one of the main sites for rat genome information worldwide. In collaboration with the US rat genome database RGD (7), RatMap is today an official organ of the Rat Genome and Nomenclature Committee (RGNC), dealing with rat gene nomenclatures issues on a daily basis. RatMap is an open database that can be accessed at http://ratmap.org or http://ratmap.gen.gu.se.

This article describes the updating procedures, nomenclature work, data and tools available at RatMap.

Updating procedure and presentation

Most of the rat gene data entered into RatMap have been obtained from published papers and rat genome sources found on the Internet. All data are manually curated before being entered into the database. Users can submit or correct data and reserve rat gene symbols through web-based forms. Owing to the curation efforts, each gene, QTL and DNA-marker found in RatMap has been checked against gene symbols in other mammalian databases to limit the possibility that a gene/marker/QTL is already known under another name. Considerable efforts are made to keep the gene symbol nomenclature accurate and in synchrony with that of human and mouse. Consequently, each rat gene is checked for potential orthology with human and mouse genes before being entered into RatMap. This is a crucial step, since naming a rat gene with a symbol already used in mouse or human implies that this gene actually shares the same ancestral gene in all three species.

The rat gene information in RatMap can be accessed using query forms, gene lists, gene maps and queries based on literature reference. If available, all gene records in RatMap are presented with accession IDs and URLs to other relevant databases such as NCBI PubMed, LocusLink and Unigene (8), GenBank (9), EMBL (10), DDBJ (11) and the genome databases RGD (7), MGD (12) and GDB (13).

Gene, DNA-marker and QTL records also include information on chromosomal position, mode of localization, status

---

*To whom correspondence should be addressed. Tel: +46 31 773 3961; Fax: +46 31 773 2599; Email: fredrik.stahl@gen.gu.se

and references. If referenced, primer pairs, relations between DNA-markers, genes and QTLs are also noted.

Furthermore, RatMap maintains data on rat idiograms and karyotypes, linkage map orientation and rat strain information. RatMap also gives individual labs the opportunity to present their genome data on the web.

Quantitative trait loci—nomenclature and presentation

At RatMap, specific attention is paid to QTLs since these are the focus of much of the present rat genome research. In the QTL curating process, RatMap has been responsible for developing a specific nomenclature. Thus each new rat QTL is given an appropriate symbol, such as 'Bp' (for blood pressure) and a number ('Bp22') which reflects the order in which the QTL was entered into the RatMap database. This means that all new published QTLs are given a unique symbol unless it is clearly stated that a certain QTL refers to a previous study. As a consequence of this nomenclature work, a given QTL symbol is not necessarily found in the original paper since many symbols are originally introduced by the RatMap curators. Sometimes different QTL investigations of a similar trait cover the same chromosomal region. However, if not explicitly stated in the original references, these QTLs will still be given separate symbols.

All QTLs can be retrieved from the RatMap query form. In order to give an overview of the different QTLs, they have been brought together into the framework of a single integrated rat linkage map. This linkage map is composed of individual linkage reports, including all QTL reports in RatMap and presently contains 8640 DNA-markers. A query tool, simply called 'QTL', relating to this integrated QTL map permits queries by QTL type and/or chromosomal position. All QTLs that intersect a chromosomal region selected by the user are automatically retrieved. This tool has been applied as a model for a bovine QTL browser (P. Polineni *et al.*, manuscript in preparation) and could be extended to QTL studies in other species as well.

The QTL service also includes an instrument for finding gene candidates contributing to the QTL phenotypes. A first version of this tool, called Candidate Gene Capture (CGC), covers 37 rheumatoid arthritis QTLs (L. Andersson *et al.*, manuscript in preparation). Genomic regions in rats associated with rheumatoid arthritis are combined with rat/human gene homology data, descriptions of phenotypic gene effects and selected keywords. Each keyword is assigned a value that is used for ranking genes based on their textual description given by the OMIM database (8).

RatMapped

RatMapped is a newly developed tool combining the integrated linkage map with gene sequence positions obtained from BLAST alignments of 4170 rat gene sequences (obtained from GenBank) and 2962 rat DNA-markers (obtained from RatMap) compared with the complete rat genome sequence (14). The integrated linkage map has made it possible to include rat QTLs into RatMapped, which means that rat genes, polymorphic DNA-markers and QTLs are all presented in a single table with positions in base pairs and cM. The contents can be surveyed by selecting a specific chromosome or searching a gene or a DNA-marker. The resulting web presentation can also be downloaded in text format. Links to RatMap, RGD or GenBank are given for all genes, DNA-markers and QTLs.

BACFinder

BACFinder is a web-tool based on BLAST alignments between 11607 rat sequences obtained from GenBank (8055 mRNA sequences and 3552 DNA sequences) and 13914 rat bacterial artificial chromosome (BAC) sequences obtained from the Baylor College of Medicine (14) (J. Fernandez-Banet *et al.*, manuscript in preparation). BACFinder allows the user to search for a specific gene by its name or accession ID, or find the GenBank sequences that align within a specific BAC. The results are shown graphically, which makes it easy to see how sequences and BAC clones are related and in tables with more detailed information about alignment length, similarity percentage, *E*-value, etc. Each alignment can be viewed by using either GenBank sequences or the rat BACs as primary matrices. This makes it possible to deduce the number and approximate position of exons within rat genes as well as exon and intron sizes since a single mRNA sequence usually makes multiple alignments with a fully sequenced BAC.

Comparative maps

In order to take full advantage of the rat as a model for human complex diseases, rat genomic data need to be closely linked to other model organisms and human (15). Based on 893 orthologous gene pairs, a detailed comparative map between rat and mouse was put together and made available as a separate tool (16). This mouse–rat prediction map, called mouseGAPP, is also integrated into the RatMap database and can, optionally, be retrieved through the general RatMap query form. Since the mouse map contains about five times as many chromosomally localized genes as that of the rat, the position of more than 6000 genes could be predicted in the rat based on this study with accuracy estimated to exceed 95% (16).

A rat–human comparative map, called humanGAPP, has now been introduced at RatMap. In this map, almost 1000 rat–human orthologous gene pairs were used as anchor points to create a framework from which the complete human–rat comparative map was built. At least two consecutive orthologous gene pairs on the same pair of a human and a rat chromosome were used to define an evolutionary conserved segment (ECS). In all, 130 human/rat ECSs were found. In contrast to the mouse–rat comparative map, this human–rat map is arranged with the human chromosome as a framework using the human base pair positions obtained from the UCSC Genome Browser (17) as a matrix. Thus each gene is ordered and defined by its start codon. Supporting data, such as gene symbol, gene description, OMIM accession ID, GDB accession ID (13) and LocusLink ID were downloaded from NCBI (8). The human framework data is linked to rat genome information from RatMap. All rat and human genes within any ECS, including the homologous gene pairs, are accessible through a web-based interface. For rat genes aligned with BACs (obtained from the BACFinder database), details on alignment and links to the corresponding rat BACs are presented.

It should be noted that both the mouse–rat and the human–rat comparative maps are independently developed at RatMap and based on the occurrence of manually inspected orthologous gene pairs. Thus, they serve as an independent complement to similar comparative maps built on other assumptions.

IMPLEMENTATION

The RatMap website is built on Apache HTTP server 1.3 using PHP 4.0 (PHP: Hypertext Preprocessor) as the scripting language and MySQL(T) v.3.23.39 as the database management system.

CITING RATMAP

Please use the following format when citing RatMap: RatMap, CMB–Genetics, Göteborg University, Sweden (http://ratmap.org)

REFERENCES

1. Gill,T.J.,III, Smith,G.J., Wissler,R.W. and Kunz,H.W. (1989) The rat as an experimental animal. *Science*, **245**, 269–276.
2. Szpirer,C., Szpirer,J., Klinga-Levan,K., Stahl,F. and Levan,G. (1996) The rat: an experimental animal in search of a genetic map. *Folia Biol. (Praha)*, **42**, 175–226.
3. Jacob,H.J. (1999) Functional genomics and rat models. *Genome Res.*, **9**, 1013–1016.
4. Günther,E. (1999) Rat models of disease. *Rat Genome*, **5**, 75–96.
5. James,M.R. and Lindpaintner,K. (1997) Why map the rat? *Trends Genet.*, **13**, 171–173.
6. Lander,E.S. and Botstein,D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
7. Twigger,S., Liu,J., Shimoyama,M., Chen,D., Pasko,D., Long,H., Ginster,J., Chen,C.-F., Nigam,R., Kwitek,A. *et al.* (2002) Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res.*, **30**, 125–128.
8. Wheeler,D.L., Church,D.M., Edgar,R., Federhen,S., Helmberg,W., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E. *et al.* (2004) Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res.*, **32**, D35–D40.
9. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2004) GenBank: update. *Nucleic Acids Res*, **32**, D23–D26.
10. Kulikova,T., Aldebert,P., Althorpe,N., Baker,W., Bates,K., Browne,P., van den Broek,A., Cochrane,G., Duggan,K., Eberhardt,R. *et al.* (2004) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **32**, D27–D30.
11. Miyazaki,S., Sugawara,H., Ikeo,K., Gojobori,T. and Tateno,Y. (2004) DDBJ in the stream of various biological data. *Nucleic Acids Res.*, **32**, D31–D34.
12. Bult,C.J., Blake,J.A., Richardson,J.E., Kadin,J.A. and Eppig,J.T. (2004) The Mouse Genome Database (MGD): integrating biology with the genome. *Nucleic Acids Res.*, **32**, D476–D481.
13. Letovsky,S.I., Cottingham,R.W., Porter,C.J. and Li,P.W.D. (1998) GDB: the Human Genome Database. *Nucleic Acids Res.*, **26**, 94–99.
14. Rat Genome Sequence Project (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
15. O'Brien,S.J., Wienberg,J. and Lyons,L.A. (1997) Comparative genomics: lessons from cats. *Trends Genet.*, **13**, 393–399.
16. Gomez-Fabre,P.M., Helou,K. and Stahl,F. (2002) Predictions based on the rat–mouse comparative map provide mapping information on over 6000 new rat genes. *Mamm. Genome*, **13**, 189–193.
17. Karolchik,D., Hinrichs,A.S., Furey,T.S., Roskin,K.M., Sugnet,C.W., Haussler,D. and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.