AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond

Zhi-Liang Hu*, Eric Ryan Fritz and James M. Reecy*

Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames, IA 50011-3150, USA

Received August 15, 2006; Revised October 13, 2006; Accepted October 20, 2006

ABSTRACT

The Animal Quantitative Trait Loci (QTL) database (AnimalQTLdb) is designed to house all publicly available QTL data on livestock animal species from which researchers can easily locate and compare QTL within species. The database tools are also added to link the QTL data to other types of genomic information, such as radiation hybrid (RH) maps, finger printed contig (FPC) physical maps, linkage maps and comparative maps to the human genome, etc. Currently, this database contains data on 1287 pig, 630 cattle and 657 chicken QTL, which are dynamically linked to respective RH, FPC and human comparative maps. We plan to apply the tool to other animal species, and add more structural genome information for alignment, in an attempt to aid comparative structural genome studies (http://www.animalgenome.org/QTLdb/).

INTRODUCTION

Animal genome research has primarily been focused on elucidation of genes that underlie diverse phenotypes using different mapping methods. Rapid progress has been seen in the movement from genome maps, to trait maps and to eventual gene discovery in the past decade. This progress is highlighted by the availability of large numbers of quantitative trait loci (QTL) identified in the mouse, rat, cattle, pigs, chicken, sheep and other species. This large amount of QTL information provides a useful bridge to link genome information with phenotype. However, it must be acknowledged that a bottleneck exists between mapped QTL and gene discovery (1). The ultimate goal of QTL mapping is to identify the genes that underlie polygenic traits and to gain a better understanding of their physiological and biochemical functions. However, success has been limited in utilization of the wealth of QTL information for marker assisted selection (2). This is partly due to the inability to link QTL information to versatile genomic information.

One challenge is to find ways to integrate and visualize the QTL data that is produced by different types of QTL analyses, from different laboratories, and with different software tools (3). Development of the PigQTLdb (4) was a significant step towards comparison and consolidation of QTL study results from different laboratories within one species (pig).

In this paper, we report the current status of our continued efforts to integrate available structural genomics information for positional QTL information mining in a comprehensive QTL database tool set called Animal QTLdb, which has also been revamped to house QTL data from multiple species.

MATERIALS AND METHODS

Data

The QTLdb accepts data curated from published papers, or from private laboratory reports that are in the process for publication. Raw data for >50 parameters/data types is collected. These data includes QTL location (chromosome, location,
location span), flanking markers (at suggestive and/or significant test levels), peak markers, test statistics (LOD score, least square means, P-values, F-values, variance), QTL effects (dominance effect, additive effect), candidate genes, traits and their measurements, etc. The data are organized into related tables in a relational database. The database structure is designed to enable dynamic links to external data resources such as NCBI (Figure 5). The backbone map for displaying marker locations was developed with mapping data mainly from USDA-Meat Animal Research Center (MARC) [for pigs: (12); for cattle: (13)] and Wageningen University [for chicken: (14)]. Markers defining a QTL that were not on the MARC or Wageningen maps were interpolated to form a consensus map with the comparative framework of ArkDB (15).

Data curation
QTLdb curation tools were designed such that a three-tiered data curation structure was implemented, which involves the roles of curators, editors and administrators. This implementation scheme was developed to ensure a smooth workflow and quality control during the entry of new data. The roles of Curators are to gather needed information from literature and scientific reports, translate them into database parameter format of the QTLdb, organize the raw data into an archive for possible ‘data debug’ and perform the data entry. The roles of Editors are to review the new data entered by Curators, exam the data integrity and scan for errors, including those of grammar, wording and formatting. The roles of Administrators are to manage the database privileges of the Curators and Editors, make new data releases and coordinate data curation and data debug processes.

Software
We used MySQL (version 12.22) as the underlying relational database and Apache 1.3.31 as the worldwide web server. Perl (5.8.5) was employed to program the common gateway interface (CGI) to present and interpret data, synthesize maps, make dynamic links to various data resources and facilitate the interaction between the web portal and the database. The curator/editor tools are also programmed in Perl/CGI. Lincoln Stein’s Perl GD library (Lincoln, 2000) was employed to draw QTL map graphs ‘on the fly’. All data and tools are hosted on a RedHat Linux server located at Iowa State University.

DEVELOPMENTS AND RESULTS
Our progress on development of the QTL database is highlighted with new features for integration of several kind of structural genomic information and inclusion of QTL data from multiple species.

QTL comparisons within a single species
The PigQTLdb has been used to successfully house, compare and manage pig QTL data. A number of enhancements on the database tools were made to efficiently locate, interpret and compare QTL results from different studies. These enhancements include better inter-links between ‘browse’, ‘view maps’ and ‘search’ functions, standardized database summaries, etc. (see Frequently Asked Questions #8 at http://www.animalgenome.org/QTLdb/faq.html).

Expand the QTLdb to include data from more species:
potential for comparative QTL analysis
Although there have been a number of similar QTL databases available in various species, our QTLdb has a few unique features over its peers. First, it is designed to house more information to better describe a QTL. Second, it is designed to include trait-associated qualitative variations in the database. Third, it is designed to have the capacity for dynamic cross-platform, cross-internet database links. The value of a good tool is in its practical utility. Toward this end, we have modified the database structure and interface tools, to allow inclusion of QTL data from multiple species. We are now in the process of adding cattle (both dairy and beef) and chicken QTL data. To date, we have collected 630 cattle QTL, 657 chicken QTL and 1287 pig QTL, which represent 89, 112 and 246 cattle, chicken and pig traits, respectively. These QTL were obtained from a total of 180 published journal papers. We continue to actively curate cattle, chicken and pig QTL data into the database and will periodically make database releases to general public.

Trait ontology and its database management
Previously, we introduced a ‘trait ontology’ to manage diverse pig production traits with a controlled trait vocabulary management system for cross-experiment QTL comparisons [(4); Table 1]. We have now improved our trait ontology management by starting to implement tools to manage directed acyclic graph (DAG) data over the existing simple hierarchical terminology management through an Animal Trait Ontology (ATO) Editor [a.k.a. Collaborative Ontology Builder, COB; (16)]. We realize that the development of Trait Ontology is a long-term effort, which involves active community inputs. We have setup a collaborative platform for contributors in the community to work together on the update and management of the ATO (http://www.animalgenome.org/bioinfo/projects/ATO/).

Genomic mining of QTL region with comparative genome information
High-density RH comparative maps have added new opportunities for researchers to search for genes that underlying a QTL with its fine mapped markers. The MySQL database structure within the QTLdb was modified to house the RH-human
comparative maps for cattle, pigs and chickens. We have successfully incorporated the porcine (5) and cattle (6) RH maps and comparative mapping data into the QTLdb. Chicken RH data (7) is in the process being added. Web tools were programmed to extend the QTL display function to show maps of comparative information alignments (Figure 1). Figure 2 shows an example where a pig QTL on chromosome 2 is aligned to the porcine RH map as well as the human genome using the QTLdb tools. The map visualization is made intuitive for end-users to retrieve such information easily via the QTLdb web viewer. With the aid of this tool, map alignments will be further refined as more sequences become available from the chicken/cattle/pig genome-sequencing project. Toward this end, we also plan to add the chicken/cattle/pig genome track information to the QTL maps when they become available.

**Aid for selection of positional candidate clones**

The dissection of promising genomic regions for candidate genes remains a challenge as the size of such regions normally spans from 2 to 20 cM (on average, 1 cM is ~1 million bases). The recent availability of the 16X-complexity pig BAC fingerprint map provided a new resource for positional cloning (11). We have performed a genome-wide alignment of the pig QTL maps with the Sanger BAC FPC maps by blast analysis of the marker sequences against the BAC end sequences (BES), and subsequently created a map alignment with significant blast hits. To date, we have produced 112 anchor points to link the two maps for all 19 pig chromosomes based on sequence blast matches, and this alignment is continuously improved by adding more anchor sites once new data becomes available, to improve alignment quality. For cattle, we have utilized the BAC–RH map alignment information from the Illinois–Texas high-resolution whole-genome cattle–human comparative map (6).

Tools were created for end-users to search with QTL location information (cM) for underlying BAC clones, and related information for further obtaining the sequences. The map alignments will be further refined as more sequences become available from the respective animal genome-sequencing project. Figure 3 shows a snap shot of porcine SSC 8 alignment with the FPC maps using the QTLdb tools.

**Improved curator/editor tools**

We have built into the QTLdb a set of data curation tools. The addition of curator/editor tools to the QTLdb made it a more useful tool for the research community by allowing multiple researchers to work together. A number of constraints and checkpoints are enforced during the curation process for data integrity, consistency and error checks.
addition, a three-layer quality control mechanism is implemented, with a curator layer, an editor layer and an administrator layer, which helps to minimize the errors, including those of grammar, wording or data formatting. Figure 4 shows the data flow and workflow for curators/editors/administrators that are built into the QTLdb. The tool also helps to keep track of which curator is responsible for which entry, etc. This tool-set uses a server–client model so that it is possible for multiple curators and/or editors from different locations to work together. We have tested the tools with a number of independent curator/editor testers, locally and remotely, and released the tool for public input of new data by registered curators (http://www.animalgenome.org/ QTLdb/app.html).

Public data and private data
The QTLdb is also built in such a way that it can house both ‘public’ and ‘private’ data. Any new data entered is kept as ‘private data’ by default until the data owner (curator) wishes to release it to the ‘public’. This feature gives general users the flexibility to use the QTLdb as a research tool as well as a data-publishing tool. The ‘private data’ can be compared to public data through an authenticated user interface. However, the data is not available for comparisons through the public web layers. Being used as a research tool, the QTLdb helps researchers to examine their data against the public data and comparative data in order to draw further inferences with their data set, before they wish to publish the data. In another scenario, if published data is found to be problematic, we can turn that data to ‘private’ for further scrutinization before release to the public again.

AVAILABILITY
The database contents and tools are all available online at the URL: http://www.animalgenome.org/QTLdb/. The online user-guide can be found at http://www.animalgenome.org/ QTLdb/faq.html. For those who wish to enter their own data, this link http://www.animalgenome.org/QTLdb/app. html can be used to apply for access, or login directly once an application is already granted.

DISCUSSIONS AND FUTURE DIRECTIONS
There are several QTL databases in the public domain for rat, mouse and cattle (http://bovineqtl.tamu.edu/) (17–21). The following features make our QTLdb stand out from its peers. First, the Animal QTLdb is designed for and includes data from, multiple species, and is built to have potential for cross-species comparisons through maps or queried by using the trait name/gene name as key words. Secondly, mechanisms are implemented in the Animal QTLdb to allow comparative mapping analysis within species and make it possible to be developed for such comparison between species. Its ability to align pig/cattle/chicken QTL maps to the human genome provides a powerful utility for animal genome researchers to draw inferences from the information-rich human genome. Third, the Animal QTLdb is open to the public for new data entry and existing data management. The ability to share data management privileges will lead to greater participation of scientists from diverse research fields. The general public will have more incentives to contribute their input, and in return, to make the tool more useful. Fourth, by design, the Animal QTLdb allows the most comprehensive QTL description by inclusion of all existing parameters, therefore the QTLdb is more applicable to various situations where diverse analytical methods and parameters may be accommodated. This makes it the most objective comparison tool.

Mapping the location of the QTL is an extremely important first step in QTL identification. While it is clear that location is not sufficient to identify a QTL gene, it is an integral part of the investigation and substantially limits the number of candidate genes that may be identified (22). QTL MatchMaker is
ACKNOWLEDGEMENTS

The authors would like to thank Stacey Meyers, Jonathan Beever, Denis Larkin and Harris Lewin from the University of Illinois for providing the pig, cattle RH maps respectively; Sean Humphray, Carol Scott and Jane Rogers from Sanger Institute for their help in aligning the pig BAC FPC against the linkage maps; Behnam Abasht and Susan Lamont from Iowa State University for kindly providing the chicken QTL data, Martien Groenen from Wageningen University for their help with the chicken consensus linkage maps and Max Rothschild as the US Pig Genome Coordinator to bring some needed resources together. In addition, we would like to thank Svetlana Dracheva, Donna Maglott and Wonhee Jang from NCBI for their efforts to incorporate the QTL data through NCBI pipelines. Gratitude also goes to Gary Rohrer from USDA-MARC and Cathy Ernst of MSU, for their diligent work curating some of the pig QTL data and testing the curator tools. Funding to pay the Open Access publication charges for this article was provided by NRSP-8 National Animal Genome Research Program.

Conflict of interest statement. None declared.

REFERENCES

of bovine quantitative trait loci. BMC Bioinformatics, 7, 283.