Animal QTLdb: towards a comprehensive database and tool set for livestock genome research

Zhi-Liang Hu, Carissa A. Park, and James M. Reecy

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

Abstract

The Animal QTL Database (QTLdb; http://www.animalgenome.org/QTLdb) has been actively developed over the past 11 years and has become one of the most used database resources in the animal genome research community. With 28 releases to date, the phenomenal growth of the QTLdb is evidenced by ongoing data curation, database expansions, and addition of new functions, among many other improvements. Our efforts are effectively transitioning the QTLdb into a comprehensive database and tool set for the research community. The updates presented here serve as a summary of our progress over the past few years and as an introduction to several new features, including data types, species, trait mapping to ontologies, data alliances, user interface tools and curation tools, and federation with an Animal Trait Correlation Database (CorrDB). With these updates we would also like to solicit ideas and input from the research community regarding how we should focus our development efforts to better serve our users in the long run. For example, what data presentation, download, formatting, integration, and analysis tools might help with your research? What future database development efforts would help us keep pace with genome biology research? With your help we hope to continue to expand the utility of the QTLdb/CorrDB in the coming years.

Figure 1

New steps to allow batch upload of large amounts of QTL/association data for curation in the Animal QTLdb. Criteria for minimum required information must be met for the data to be accepted.

Animal QTLdb

Chicken Cattle Pig Sheep Rainbow trout F A Q

Steps to upload large quantities of new QTL/association data

As an alternative to manually curating your QTL/association data into the QTLdb with the curator/editor tools, large quantities of new data may also be tabulated in plain text format to batch load for incorporation into the QTLdb. If you choose this method, please follow these steps:

I. Prepare your data:

 Follow the minimum required information guidelines when preparing your data. Organize your data in several file groups: Publication/manuscript Experiment descriptions Trait description

Figure 4

CorrDB curation tools are being developed to streamline links between different components of data resources to reduce errors and necessity for manual processes.



Introduction

As a widely used database resource in the animal genome research community, the success of Animal QTLdb is evidenced by continued growth in newly curated data, built-in functions, and external data alliances with major genomics databases (1). Over the past year, we have continued development efforts by starting a new round of improvements. These efforts have been focused on automating processes as much as possible for data curation and data transfer between applications and our alliances, while adding new tools to aid data analysis by users. The emphasis of our current work is also on long-term sustainable development in terms of component reuse and interoperability.

Results

Some of the developmental works shown here are still in progress.

- New steps are in place to allow users to batch upload large amounts of QTL/association data to speed up the curation process. We have developed guidelines regarding the minimum required information for preparation of data before it can be accepted by the database (Figure 1).
- While the Animal QTLdb and the CorrDB are each continually improved (Figure 2), we have started to co-develop them under a federated database model by modularizing components for use in both (Figure 3).
- New curator tools for the CorrDB have been developed to be more

- iv. Tabulated data (contains map positions, trait, statistics, and other information such as breed, associated genes, etc.) in one or more (.csv or tab-delimited) text files.
- Compress your data files.

II. Send your data:

• Upload your data via the NAGRP File Sharing Platform. Be sure to include a description of the purpose, nature of your data, and contact information in the "Message" box so that we can properly identify and organize your data files for processing.

OR

 Attach your data files to an email to us. Include a description of the purpose, nature of your data, and contact information in the mail body.

III. Get in touch with one of our curators

Your uploaded data will need to go through a number of curation and quality control processes before they may be released for public access. Therefore it is necessary for you to email us to follow up. This is to we have received your data in order for us to coordinate the

Animal QTLdb

Chicken Cattle Pig Sheep Rainbow trout FAQ

Minimum information required for Animal QTLdb data entry

To ensure good data integrity and maintain high data quality standards, we request that users follow these minimum required information requirements when preparing data for batch submission to the Animal QTLdb.

1.	Publication information - mandatory	
	 PubMed ID - For pre-publication manuscript, obtain an automatically assigned "ISU" number, which will be replaced once a PubMed ID becomes available. 	Working
	 Authors - Last names and initials for all authors 	examples
	 Author email - At least one (1) contact email; work emails are preferred 	will be
	 Affiliation - Name and location of institution(s) 	added
	• Article title	
	• Journal - Full name (with publisher information when the same name is found for an	other journal)

Full hame (with publisher information when the same hame is found for another • Year of publication

- Issue/volume and page number In the format the journal provides, e.g.: 8(1):123-135
- Abstract

Release 28 (Dec 29, 2015)

UCSC

CNV

WAS .

CACTOTGACOTAGCTGGGCTT

Full Text - Preferably in PDF

5. Gene	ic and phenotypic correlation data: Add New Edit Or correction
© 2003-2015 N/ Contact: NAGRP	AGRP - Bioinformatics Coordination Program. Bioinformatics Team
Cattle Corr	Data Management Tools for curator
Edit Experir	nent for publication #9814897
Current data:	
Animal SPEC	ES: Cattle
Animal BREE involution: If select "New previously entered breed will be overri	Select cattle BREEDS involved: Angus, /ed: (Search/filter:) New brangus, 'any data dden africander oAdd charolais, albères alentejana merican gelbvieh, american white park holstein
Animal U	SE: Carcass measurements of 1,292 steers collected by the American Short Association were used to estimate genetic and phenotypic parameters. Records were classified by herd of origin, slaughter group, and individual year of birth.
TRAITS measu Caution: If select "New previously entered trait dat be overri	red: 'any a will dden Select cattle traits tested (Search/filter:)) QTLdb traits cattle trait:Exterior:1460:Behavioral cattle trait:Exterior:1555:Coat texture cattle trait:Exterior:1555:Coat texture cattle trait:Exterior:1458:conformation cattle trait:Exterior:1580:Hoof traits cattle trait:Exterior:1549:Udder traits
	New • Add
	Dressing percentage Marbling score Kidney, pelvic, and heart fat percentage Longissimus muscle area Fat thickness at the 12th rib Age at slaughter
	Live weight

intuitive to use and require fewer manual steps by curators (Figure 4).

Figure 2

The Animal QTLdb and the CorrDB have been co-developed in order to share underlying components.

AnimalQTLdb

The Animal Quantitative Trait Loci (QTL) Database (Animal QTLdb) strives to collect all publicly available trait mapping data, i.e. QTL (phenotype/expression, eQTL), candidate gene and association data (GWAS), and copy number variations (CNV) mapped to livestock animal genomes, in order to facilitate locating and comparing discoveries within and between species. New data and database tools are continually developed to align various trait mapping data to map-based genome features such as annotated genes.

Many scientific journals require or recommend that any original QTL/association data be deposited into a public database before a paper may be accepted for publication. We provide user/curator accounts for direct data submission, and supply users with a data Updated 3 summary link to facilitate the manuscript review process.

times a year Cattle QTL There are 42,019 QTLs from 646 publications curated into the database. Those

- QTLs represent 482 different traits (see data summary for details).
- Chicken QTL
- There are 5,196 QTLs from 235 publications curated into the database. Those QTLs represent 321 different traits (see data summary for details).
- Horse QTL
- There are 1,125 QTLs from 59 publications curated into the database. Those QTLs represent 34 different traits (see data summary for details).
- Pig QTI There are 14,479 QTLs from 507 publications curated into the database. Those QTLs represent 592 different traits (see data summary for details).

Rainbow Trout QTL

- There are 127 QTLs from 10 publications curated into the database. Those QTLs represent 14 different traits (see data summary for most recent updates).
- Sheep QTL
- There are 1,090 QTLs from 109 publications curated into the database. Those QTLs represent 215 different traits (see data summary for most recent updates).



Animal Trait Correlation Database CorrDB

Genetic and phenotypic trait correlation studies have been conducted for 70+ years since Hazel proposed the abase summary: methods. A genetic correlation is the proportion of shared variance between two traits that is due to genetic causes; a Publications: 1,566 Species: 6 Traits: 1,658 phenotypic correlation is the degree to which two traits co-vary among individuals in a population. In the genomics era, while gene expression, genetic association, and network analysis provide unprecedented means to decode the QTL: 64,036 genetic basis of complex phenotypes, it is important to recognize the possible effects genetic progress in one trait can Data Alliances SNCE have on other traits. This database is designed to collect all published livestock genetic/phenotypic trait correlation THOMSON REUTER data, aimed at facilitating genetic network analysis or systems biology studies. e! Ensembl

Currently this database has an initial collection of 3,635 correlation data on 276 economically important traits of cattle, relating to meat production, milk production, growth, health, and others. Information in the Animal Correlation Database can be explored through these initial access points:



- New Data: If you found correlation data that has not been included in the current release, please either (1) register for a curator account to directly input them (see "Database Curators / Editors" below), or (2) inform our Bioinformatics Team so that we can include your data for curation.
- Database Curators / Editors: The CorrDB is open to the public for data entry and update. By having a curator account on the CorrDB, you will be granted privileges to input and edit the database entries. NEW

Figure 3 (left)

Three components (references, trait ontology, and breed ontology) are being modularized in order to be shared among additional databases. The goal is to make them independent database resources with standard data portals.

transformation algorithm to obtain REML estimates of (co)variances for sire and residual effects (Meyer, 1985).
Can correlation data be further curated? (list reasons in "NOTE" when set to "No")
You may proceed with correlation data entry
Update

Analysis METHOD: All traits were analyzed jointly with a multivariate canonical

Results (cont'd)

To overcome the space limitations on the current whole-genome view of the QTL/association data, we have developed a new tool to allow an unlimited number of QTL/association data to be plotted to present a better overview of the data under examination (Figure 5). This tool has been implemented for cattle as an option on the existing genome view page. It will soon be available for all species and allow inputs from multiple data access points (e.g., from the trait hierarchy or from literature search data summary).

Future improvement

We realize that the improvement of database tools like QTLdb/CorrDB is a continual process. We plan to bring development of the QTLdb/CorrDB to the next level by adopting them into InterMine (2), an open source data warehouse built specifically for the integration and analysis of complex biological data.

References

1. Zhi-Liang Hu, Carissa A. Park, and James M. Reecy (2016). Developmental progress and current status of the Animal QTLdb. Nucleic Acids Research (Database issue; Advance Access): doi: 10.1093/nar/ gkv1233).



Figure 5 (below)

A whole-genome plot tool for an overview of QTL/association data built in the QTLdb. The advantage of this tool is that it's scalable in size such that large numbers of QTL/association data can be displayed, whereas the previous tool could display only a limited number of data due to its panel layout.

2. Smith RN, Aleksic J, Butano D, Carr A, Contrino S, et al. (2012). InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. Bioinformatics. 28(23):3163-5.



Acknowledgments: This work was supported in part by USDA Grant 2013-67015-21210 and USDA NRSP-8 Bioinformatics Coordination project.