

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.

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 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.



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.

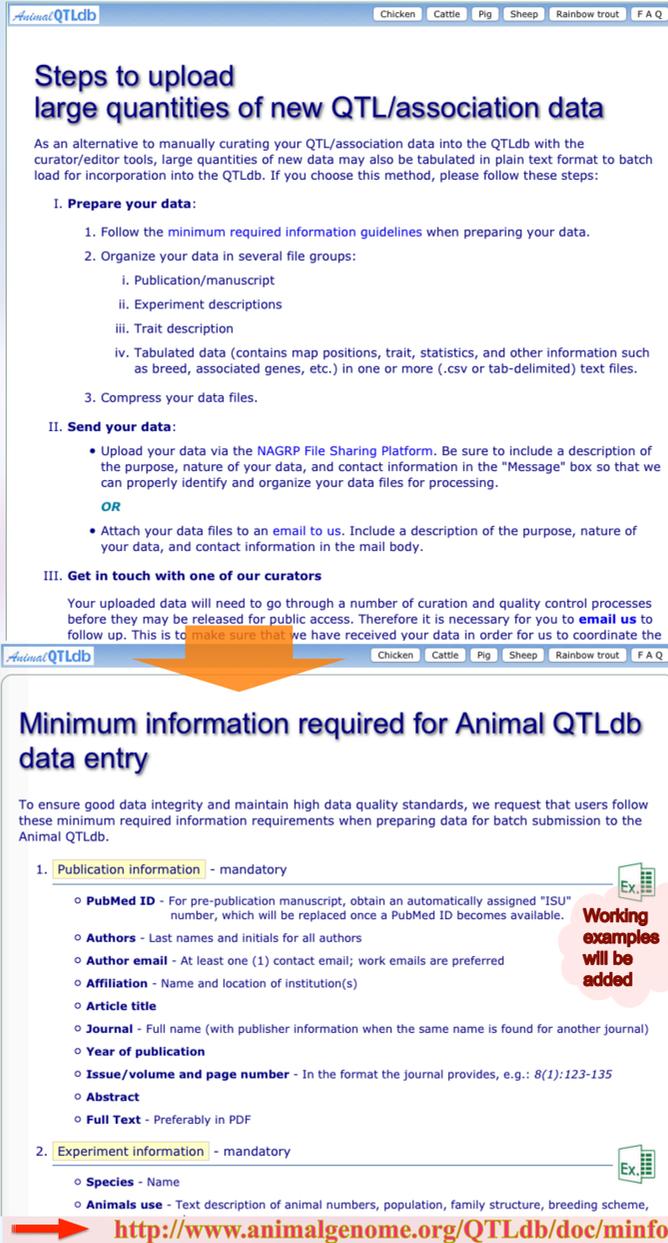
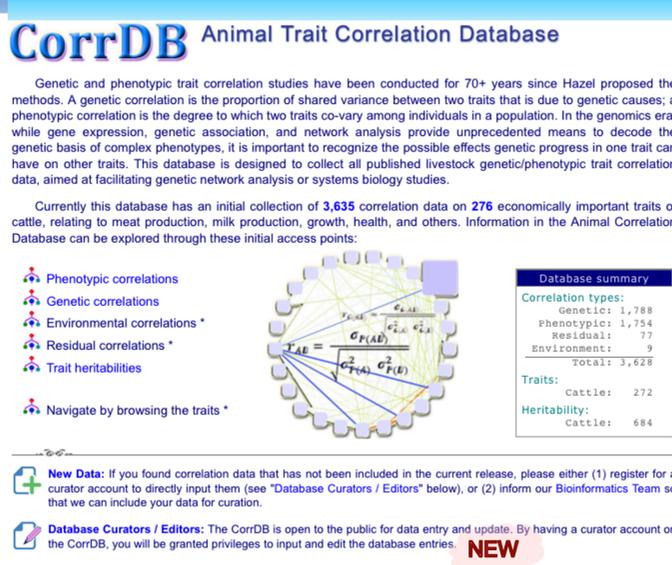
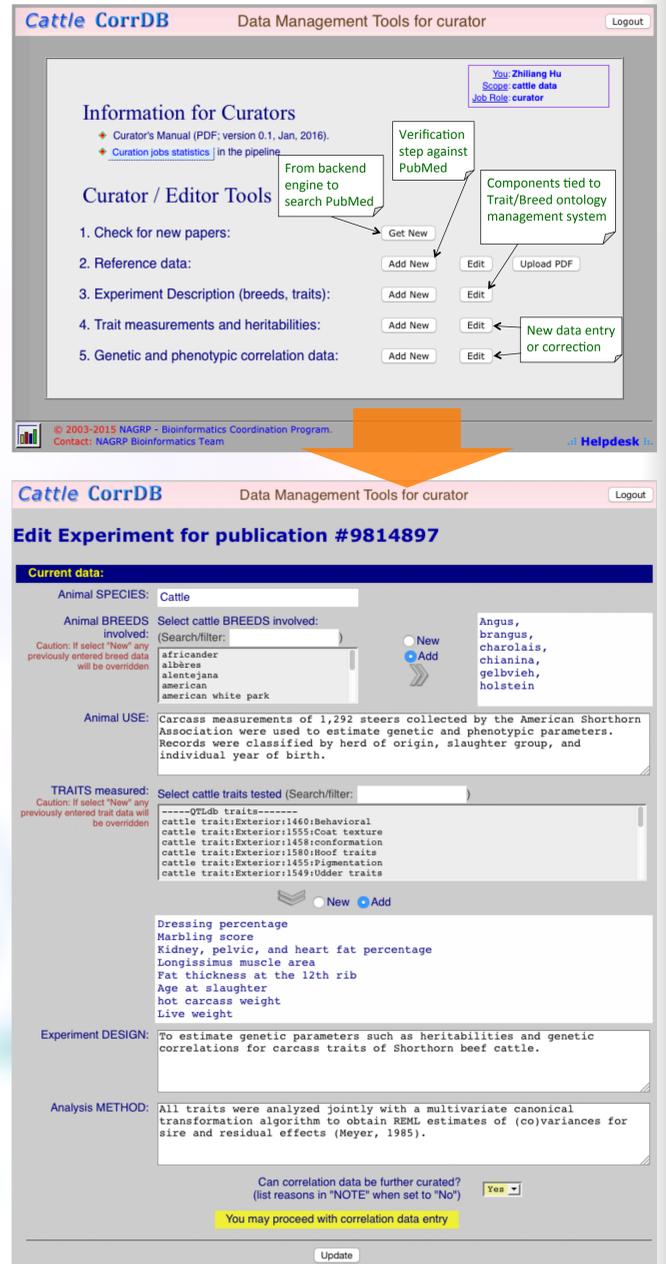


Figure 4

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



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

- 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.
- 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.

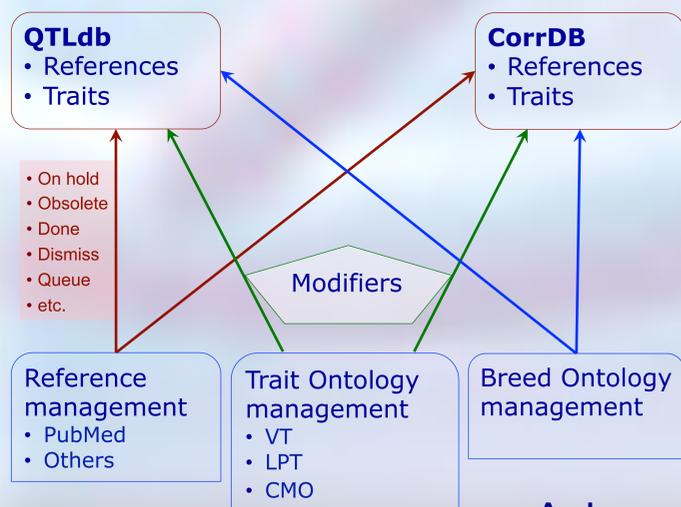
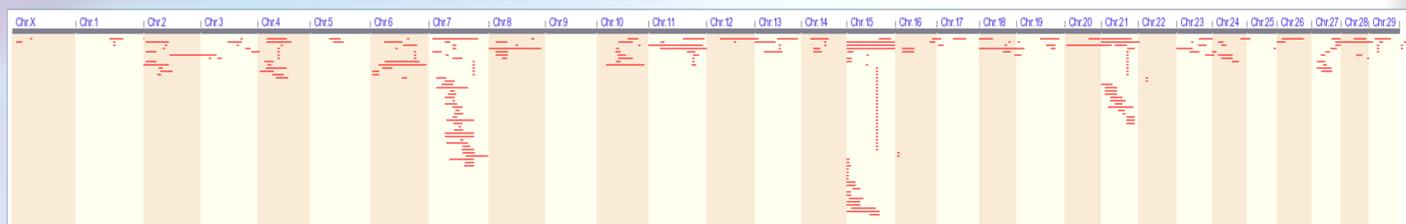


Figure 3 (left)

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.

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.



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