An update on database growth and improvements of Animal QTLdb and CorrDB

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

Department of Animal Science, Iowa State University,

2255 Kildee Hall, 806 Stange Road, Ames, IA 50011

https://www.animalgenome.org/QTLdb
https://www.animalgenome.org/CorrDB

Abstract

The Animal Quantitative Trait Loci (QTL) Database and Animal Trait Correlation Database (CorrDB) have undergone phenomenal data growth and the addition of multiple database tools over the past few years. These efforts have been well-received by the community, as evidenced by the increasing number of database web hits and data downloads every year. To date, there have been 174,241 QTL/associations on 2,062 traits from 6 species reported in 2,214 publications, 18,049 correlations on 629 traits, and 3,083 heritability data on 912 traits reported in 374 publications curated into the databases. The Animal QTLdb and/or CorrDB have been cited by over 1,000 scientific journal papers. New developments since our last update include (1) a more streamlined curation workflow and tools to support collaborative data curation between different curators; (2) more links between data types available for public access, including epistatic and pleiotropic QTL/associations, related data from follow-up studies, combined analyses of multiple studies, re-analyses, confirmation studies, or additional reports from the same experiments; (3) an improved QTL/association enrichment tool that allows user-selected genome regions and trait types to be compared against the rest of the genome to identify possible areas of over-representation; and (4) use of Intermine to facilitate genome-based QTL/association and correlation data linking, searches, and downloads. Future developments of the databases will include support of multiple genome builds within a species, and tools for liftover of QTL/association mapping data across different versions of genome assemblies.

Background

The Animal QTLdb has been recognized worldwide as a valuable resource for livestock animal QTL/association data. The database is visited by some 26,000 unique users each year, with an average data download of 136 GB per user per year. According to Google Scholar, there have been over 1,050 citations to our publications on the Animal QTLdb, and over 2,080 citations to the use of the QTLdb [1]. Over the past 16 years, there have been 178,491 QTL/associations on 2,073 traits from 6 species reported in 2,260 publications, 18,638 correlations on 641 traits, and 3,264 heritability data on 934 traits reported in 344 publications curated into the databases (**Figure 1**). Here we report our recent efforts, including adding more data types and developing tools to facilitate curation and the incorporation of the new data types into the database.

New developments

1) *A new collaborative curation environment*. A curator's working environment in the Animal QTLdb is normally isolated. When multiple curators and editors work together, they often need to collaborate virtually on the same dataset from different locations. We developed the curator/editor environment to allow them to see and work on the same dataset with a visible workflow. **Figure 2** shows how a curator (M) can seek help from Curator/Editor N, where the data status is shared and can be viewed in each person's respective curation environment.

2) *New presentations for new data types*. Although we began curating epistasis and pleiotropy data from QTL/association studies including these data types in 2009 and 2014, respectively, we have recently developed new web tools for viewing these data. **Figure 3** shows two screen shots; one is a publication data summary page where pleiotropic data links are listed (Figure 3a) and the other displays pleiotropic groups in which different traits and animal breeds are involved (Figure 3b). **An example screen shot for epistasis is not shown** because the data summaries and links are similar to that for pleiotropy data shown in Figure 3.

Among the data we make accessible to the public are the links to related publications that may be of interest to users. These related publications include, but are not limited to, follow-up studies, combined analyses of multiple studies, re-analyses, confirmation studies, or additional reports from the same experiments where QTL/associations were previously reported. Publication links appear on the reference data summary page similarly to that for epistasis and pleiotropy data. (**Example screen shots are not shown here**). Users are encouraged to visit the website to see how it works.

3) *An improved version of QTL/association data enrichment analysis tool*. Previously we reported the implementation of a new data enrichment analysis tool in the QTLdb [2], which allowed users to identify regions of the genome and to determine if they may be overrepresented in reported QTL/associations for traits of interest. Although that tool included a feature that allowed users to expand the analysis on a particular chromosome for subsection analysis, it lacked the power and flexibility to do so genome wide. Now we report an improved version of this tool that does genome-wide chromosome subsection analysis for data overrepresentations, and also provides options for users to select traits and/or chromosome section sizes for analysis. **Figure 4** shows the starting page (a) and the results page (b).

4) *Improvements to the Animal CorrDB web interfaces*. We have begun to build some frequently used queries into the user interfaces to improve user experience and allow quicker and more direct access to data. This work is ongoing, but users can see the initial changes to the front page of the CorrDB (**Data are not shown here**; please visit https://www.animalgenome.org/CorrDB to find out more).

Acknowledgments:

Figure 1

A graph showing the increase in curated QTL/association data in the QTLdb through its 40 releases over the past 16 years.
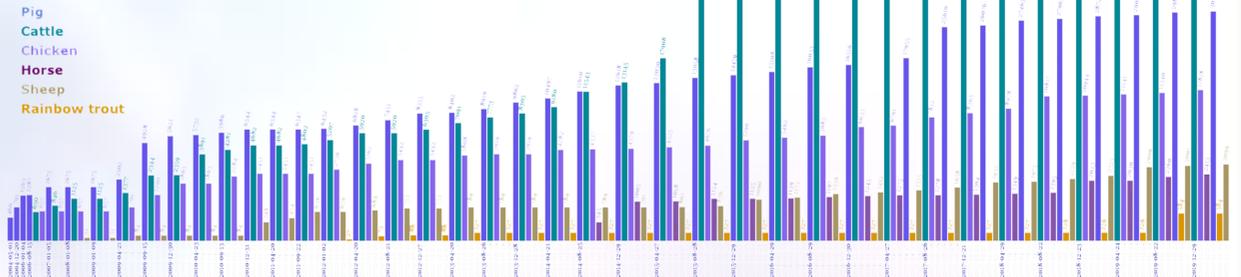
Figure 2

A screen shot showing web forms in a curator environment (a) accessible to Curator M, and a curator/editor environment (b) accessible to Curator/Editor N, in which a data file uploaded by Curator M was processed by Curator/Editor N, and the workflow is visible to both.

Figure 3

A pleiotropic data presentation summary is shown. In (a) is the data summary page of a publication, and pleiotropic data details (b) linked from that on page (a).

Figure 4 (a)

Improved version of the QTL/association data enrichment analysis tool: a user options input page, where users can select traits and/or chromosomal regions of interest.

Figure 4 (b)

A sample enrichment analysis results page, which shows genome-wide tests at both chromosomal level and sub-chromosome section levels. Each chromosome can be expanded (shown for Chr. 2 and 6 in this graph) to view section level test results.

References

1. Google Scholar Citations. (2019). URL: https://scholar.google.com. *Last visited: January 4, 2020.*

2. Z-L. Hu, C.A. Park, and J.M. Reecy. (2019). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Research, Volume 47, Issue D1, 8 January 2019, Pages D701–D710.