Animal QTLdb and CorrDB updates: integrative development of genetics/genomics databases and tools to meet new challenges

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

Abstract

Since the inception of the Animal Quantitative Trait Loci Database (QTLdb) and the Animal Trait Correlation Database (CorrDB), development of database infrastructure and tools has steadily evolved to meet user needs and the demands of ongoing growth in the amount and types of data. To date, 145,842 QTL/association data on 1,914 different traits in 6 species have been curated from 1,931 publications in 14 years. A total of 4,470 correlation data between 302 different traits in cattle have been used to establish a database model linking the two databases, in an effort to expand livestock genetic/genomic information networks. New developments include CorrDB curator tools integrated with those of QTLdb, and a better structured environment to share trait ontology development resources and centralized curator activity management. The co-development of the two database platforms allows information transfer from genetically/phenotypically correlated traits to their QTL/association locations in the genome, and further to curated genome features such as genes, SNPs, and other types of variations. A new trait modifier creation tool has extended the capacity to manage trait variants in a scalable structure. One of the major goals in the development of these database resources is to facilitate more organized, inclusive, and complete data curation, toward a well-structured "big data" reservoir and improved future utilization of the data.

The statistics

The Animal QTLdb has been widely used over the past 14 years. According to Google Scholar, there have been over **800+** citations to our publications on Animal QTLdb works, and over **1520+** citations to the use of the QTLdb, as of the end of 2017 (1). The success of Animal QTLdb is evidenced not only by continued growth in newly curated data (**Figures 1 and 2**), but also by the growing number of new and improved functions to provide better access to the data for analysis, as well as new and upgraded tools to facilitate data curations and safeguard data quality (2).

The developments

Recent developmental progress is highlighted by the co-development of the Animal QTLdb and CorrDB. The basis for this co-development is the sharing of the system resources and tools developed for management of reference data, trait ontology data, and breed ontology data that form a common foundation for information links between databases (**Figure 3**).

For example, within the QTLdb, when QTL/association data are detected which also have correlation data for the same trait, an alert is displayed with hyperlinks to the CorrDB, listing "correlation data matching the trait" on the line; likewise, a "QTL" link is attached to a trait name within the CorrDB when QTL/association data are detected for the trait (**Figure 4**).

Data exploration can include further organization of the data by "gene-centric views" or "trait-centric views" of QTL/associations (**Figure 5**).
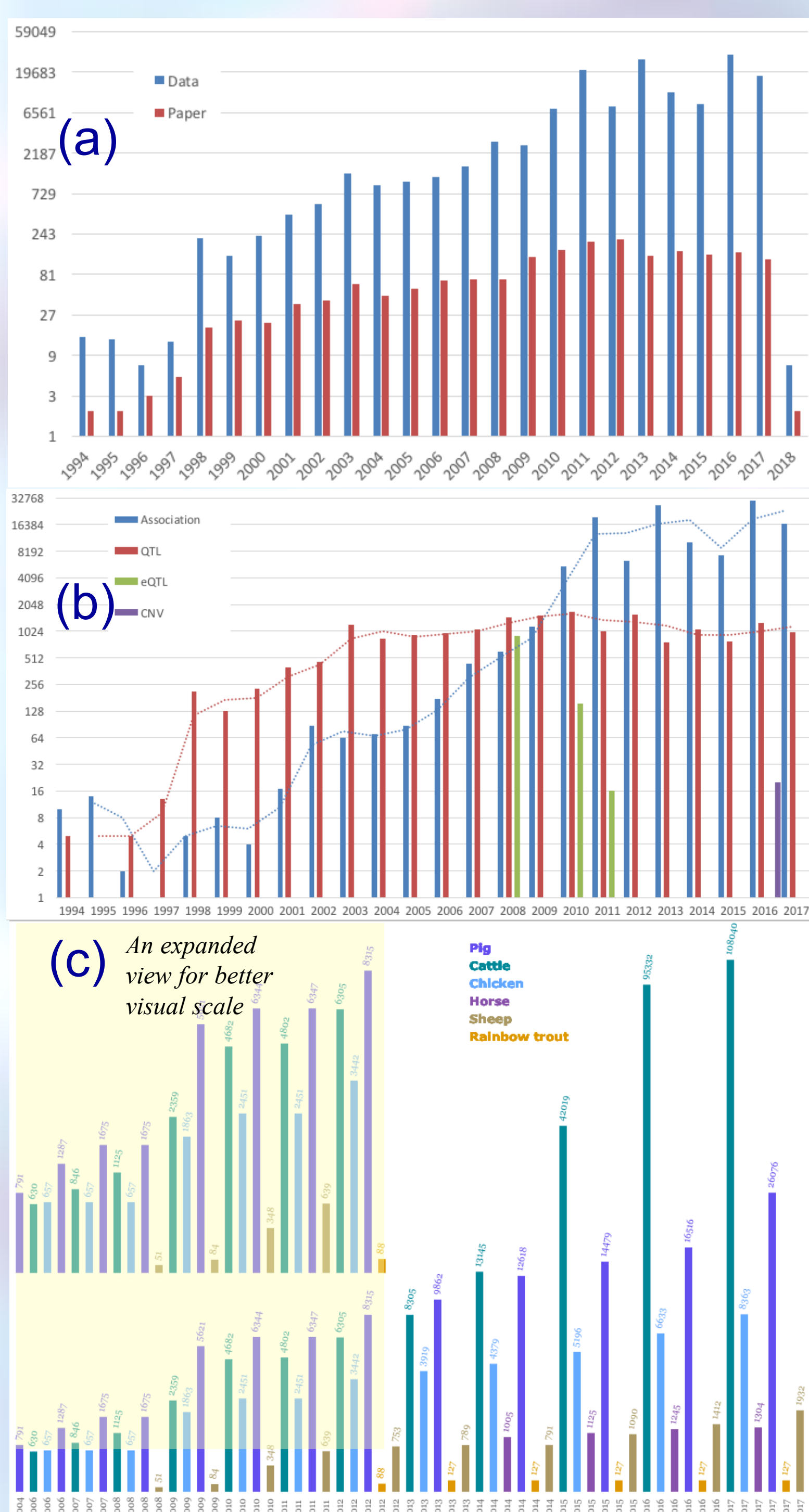
Figure 2

The front pages of Animal QTLdb and CorrDB, showing the current data statistics and data release status, which are updated upon each visit to the websites.
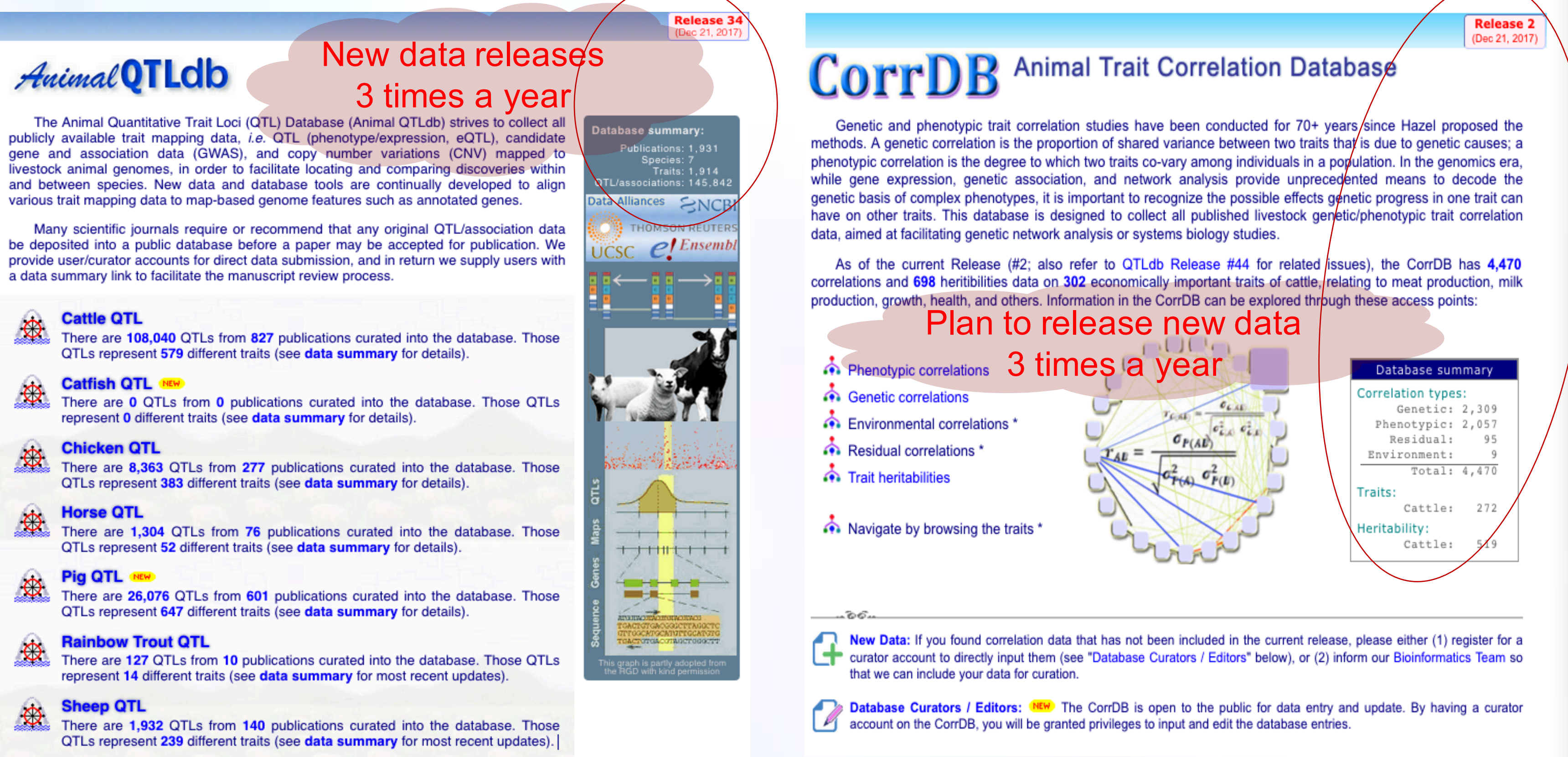

Figure 3

A conceptual diagram showing that reference data, trait ontology data and breed ontology data are shared and managed with tools developed to serve both databases. This enables the sharing of the standardized data sets to form the basis for data links between the two databases. Note that the QTLdb is already linked with NCBI and Ensembl databases through mapped genes; therefore this development aligns with our long-term goal of building federated databases.

Figure 5

Examples of a "gene-centric view" and a "trait-centric view" of the QTL/association data. Note that the views may be interchangeable by selecting the proper radio button during the search, as long as there are linked data to facilitate the view.

Figure 1

The curated data increase in the Animal QTLdb. (a) The increase in total number of QTL/association data and number of publications curated; (b) the increase in different association data types; and (c) the increase in data by species. All data on the Y-axes are logarithm transformed for improved readability.
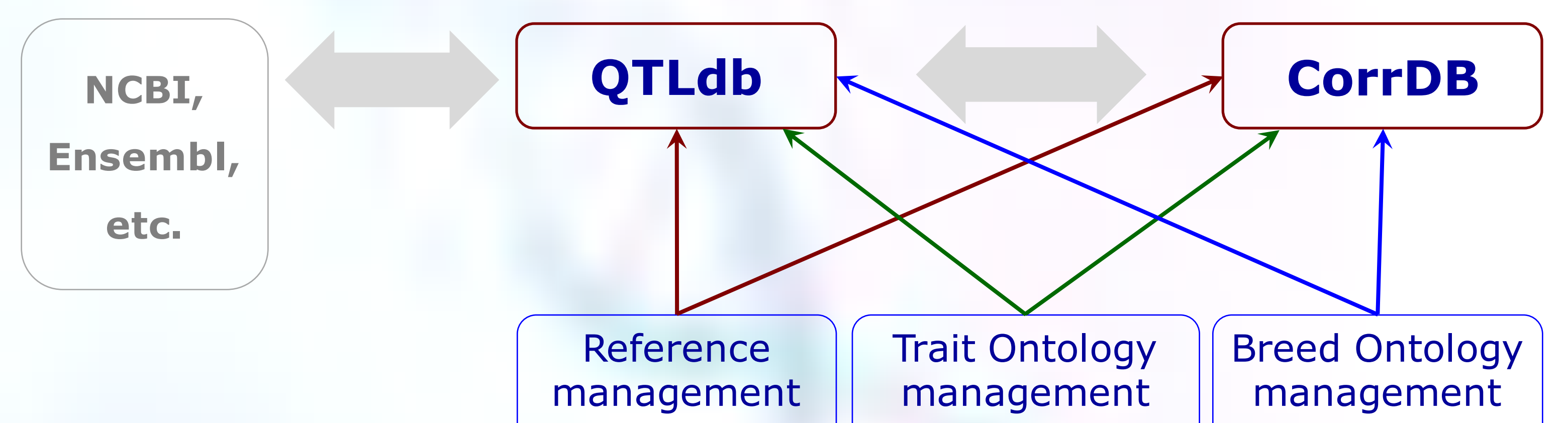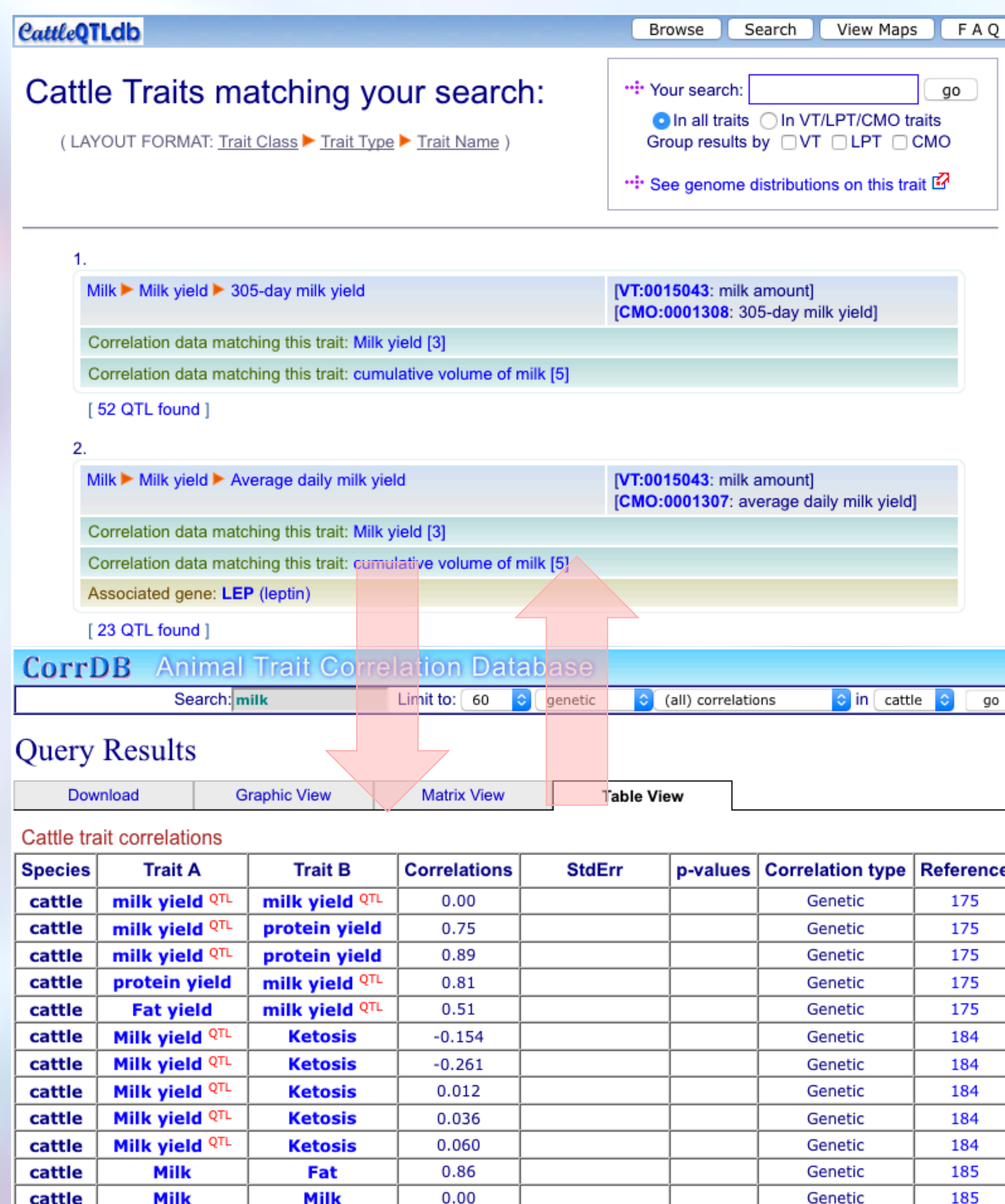

Figure 4

Links between QTLdb and CorrDB are made using common trait terms standardized through the Vertebrate Trait (VT) Ontology, Clinical Measurement Ontology (CMO), and Livestock Product Trait (LPT) Ontology. Note that in the data display a link to the other database is either stated (as in the QTLdb, top), or flagged with a label (CorrDB, bottom).

The future

Improvement of QTLdb/CorrDB database tools is a continual process. We strive to put in place long-term sustainable resources that facilitate data re-use and increase the productivity of livestock genetic improvement.

References

1. Google Scholar Citations. (2018). URL: https://scholar.google.com. *Last visited: January 4, 2018.*
2. Z-L. Hu, C.A. Park, and J.M. Reecy. (2018). Development of Animal QTLdb and CorrDB: Resynthesizing Big Data to Improve Meta-analysis of Genetic and Genomic Information. *11th World Congress on Genetics Applied to Livestock Production (WCGALP), Auckland, New Zealand, February 11 – 16, 2018.*