



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Animal QTLdb

**Citation for published version:**

Hu, Z-L, Park, CA, Wu, X-L & Reecy, JM 2013, 'Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era' *Nucleic Acids Research*, vol. 41, no. Database issue, pp. D871-9. DOI: 10.1093/nar/gks1150

**Digital Object Identifier (DOI):**

[10.1093/nar/gks1150](https://doi.org/10.1093/nar/gks1150)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Nucleic Acids Research*

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era

Zhi-Liang Hu<sup>1,\*</sup>, Carissa A. Park<sup>1</sup>, Xiao-Lin Wu<sup>2</sup> and James M. Reecy<sup>1,\*</sup>

<sup>1</sup>Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames, IA 50011 and <sup>2</sup>Department of Meat and Animal Science, College of Agriculture and Life Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

Received June 30, 2012; Revised October 24, 2012; Accepted October 25, 2012

## ABSTRACT

The Animal QTL database (QTLdb; <http://www.animalgenome.org/QTLdb>) is designed to house all publicly available QTL and single-nucleotide polymorphism/gene association data on livestock animal species. An earlier version was published in the *Nucleic Acids Research Database* issue in 2007. Since then, we have continued our efforts to develop new and improved database tools to allow more data types, parameters and functions. Our efforts have transformed the Animal QTLdb into a tool that actively serves the research community as a quality data repository and more importantly, a provider of easily accessible tools and functions to disseminate QTL and gene association information. The QTLdb has been heavily used by the livestock genomics community since its first public release in 2004. To date, there are 5920 cattle, 3442 chicken, 7451 pigs, 753 sheep and 88 rainbow trout data points in the database, and at least 290 publications that cite use of the database. The rapid advancement in genomic studies of cattle, chicken, pigs, sheep and other livestock animals has presented us with challenges, as well as opportunities for the QTLdb to meet the evolving needs of the research community. Here, we report our progress over the recent years and highlight new functions and services available to the general public.

## INTRODUCTION

Previously (1–3), we have reported on the success of Animal QTLdb, which was developed to house publicly available quantitative trait loci (QTL) data for cattle, chicken and pigs, to provide tools for aligning various genome features

to QTL and to enable comparison of QTL results within species and across experiments. As of 2007, tools had been developed to allow map alignments of QTL against various genome features, such as bacterial artificial chromosome (BAC) end sequences, single-nucleotide polymorphisms (SNPs), Affymetrix or oligo array elements and the human genome via radiation hybrid (RH) map anchor markers. In conjunction with Animal QTLdb, comparisons of QTL across species have been made possible by virtual comparative map (VCmap), a tool co-developed by Iowa State University, Medical College of Wisconsin and University of Iowa (<http://www.animalgenome.org/VCmap>). These efforts have successfully improved the public's ability to retrieve and analyze QTL data. Significant progress has been made over the past few years. First, the database has been expanded to include two more species, sheep and rainbow trout (<http://www.animalgenome.org/QTLdb/notes.php>), to serve a larger research community and to aid comparative mapping efforts. Meanwhile, new QTL data have been actively curated into the database. Since 2007, the number of QTL in the database has increased by 5.5-fold, reaching 17 566 QTL (5920 cattle, 3442 chicken, 7451 pigs, 753 sheep and 88 rainbow trout). Second, the popularity of the Animal QTLdb has been evidenced not only by daily web access records but also by the number of publications that cite use of the database—this number has reached 290 by the summer of 2012 (search for 'animalgenome.org/QTLdb' at <http://scholar.google.com>). Third, the *Journal of Animal Science* has listed the Animal QTLdb as one of the databases in which to deposit new QTL data to meet their publication requirements (<http://www.journalofanimalscience.org/site/misc/JAS-InstructionsToAuthors.pdf>). As a result, an increasing number of volunteer curators have chosen to enter their own data. Fourth, the Animal Trait Ontology (ATO) has been further developed into the Vertebrate Trait (VT) Ontology and livestock Product Trait (PT) Ontology, with relevant terms submitted for the Clinical Measurement

\*To whom correspondence should be addressed. Tel: +1 901 759 0643; Fax: +1 901 759 0643; Email: [zhu@iastate.edu](mailto:zhu@iastate.edu)  
Correspondence may also be addressed to James M. Reecy. Tel: +1 515 294 9269; Fax: +1 515 294 2401; Email: [jreecy@iastate.edu](mailto:jreecy@iastate.edu)

Ontology (CMO) (4). The ability to annotate QTL data to these more precisely defined traits and measurement ontology terms allows for improved accuracy of trait analyses.

In the meantime, the landscape of animal genomics research has been dramatically changed, with the genome sequences for cattle, chicken, horse, pigs and sheep becoming available over only 5–6 years of time. This presents a great challenge, as well as a huge opportunity, for QTL analysis. As all QTL data published in the past were linkage map-based, transfer of their linkage map locations to genome coordinates is needed in order for them to be useful in the genomics context. As gene sets have become available for microarray expression analysis, and high-density (HD) SNP arrays have been generated for whole-genome association (WGA) studies, QTL analysis is no longer the only way to link between genomes and traits. It has been our vision for the Animal QTLdb that it serves as a bridge between genotypes (genes) and phenotypes (traits) (3), which will inevitably necessitate inclusion of SNP/genome-wide association study (GWAS) data. Under this concept, we must bring related experimental results together for examination through a process called meta-analysis. Therefore, continued improvements to the Animal QTLdb are extremely important, as huge amounts of similar data continue to accumulate rapidly.

In this article, we report our recent progress in redeveloping the Animal QTLdb to meet these challenges.

## MATERIALS AND METHODS

### Data, data curation and data transformation

The QTLdb accepts either curated public data from journal papers or private laboratory reports subject to publication. More than 50 parameters/data types are subject to collection to describe a QTL, as reported earlier (3). We have recently added a number of new data types to enhance our ability to be more inclusive in QTL/association data collection. These new types include ‘association’ data for candidate gene or single marker associations; ‘eQTL’ from microarray-based QTL scan analysis; ‘test scale’ to differentiate genome-wide, chromosome-wide, comparison-wise and experiment-wise QTL/association reports; ‘test model’ to indicate epistatic or maternally or paternally imprinted QTL; new test statistics such as Bayes value and likelihood ratio, etc. We have also added animal breed information for future breed-associated QTL analysis. The backbone maps to record QTL are from USDA-Meat Animal Research Center (MARC; for pigs and cattle), Wageningen University (for chicken), University of Melbourne (for sheep) and the National Center for Cool and Cold Water Aquaculture (NCCCWA; for rainbow trout). Reported QTL genome locations were obtained by interpolating their linkage map positions via anchor markers.

QTL are mapping features recorded as linkage distances. In order for GBrowse to display QTL and for users to easily port QTL data for customized analysis, we established a process to convert the QTL linkage

map locations (centimorgan, cM) to the corresponding physical locations (megabase pair, Mbp). The data conversion is a mathematical process built in a Perl script, whereby interpolation or extrapolation is performed with reference to the nearest common anchoring marker locations on both maps.

The QTLdb has a three-tiered data curation structure so that curators, editors and database administrators can work together and share responsibilities in a workflow to ensure data quality and smooth process control. In the past few years, a set of new data debugging tools, process control mechanisms and functions for the ease of use of the tools have been developed in response to lessons learned during data curation and debugging.

### Platform and software

The QTLdb is built on a RedHat Linux platform with MySQL (version 14.12) as the backend relational database and Apache (2.2.13) as the web server. Perl (5.8.8) was used to program the web interface for user-controlled data presentations and interactive curator tools for data entry. Some lightweight PHP hypertext preprocessor and Javascript codes were also used to develop web functions where needed. An embedded R script was developed for QTL meta-plots.

## RESULTS

Since 2007, we have made 14 database releases with both new data and new functions, at a recent pace of three releases per year. The number of publications curated into the QTLdb has been steadily increasing at a rate of ~30% per year on average (Supplementary Figure S1), which indicates the importance of the research to the community. As the QTLdb is being increasingly used (<http://www.animalgenome.org/log/>), various user requests continue to be received by our Helpdesk, which compels us to further improve the QTLdb in order to better serve the needs of the research community.

### New data types and parameters

#### GWAS and eQTL data

Genome-wide association study (also known as whole-genome association study, WGAS) and expression QTL (eQTL) are newly emerged methods (relative to traditional QTL) to analyze the associations of abundant genetic variants (typically SNPs) with traits of interest (Table 1). Like QTL, GWAS adds value to our understanding of genome–trait relationships. GWAS data are genome map based, whereas QTL data are linkage map based. We have set up genome maps using the most updated version of the genome build available for each species. When a new build is available, we update both the genome maps within the QTLdb and the genome version information page linked from each genome name on the web site. The genome maps are aligned with their respective linkage maps in order to display both types of data in parallel. Two methods were used to align the maps: (i) linearly scale out both linkage and genome maps with the same length base, such that their

**Table 1.** New data types introduced to the Animal QTLdb since our last NAR publication in 2007

Analysis types	Association <sup>a</sup> eQTL <sup>a</sup> QTL
Map types	Linkage map (cM) Genome map (bp) <sup>a</sup>
Test models	Paternally imprinted <sup>a</sup> Maternally imprinted <sup>a</sup> Sex-specific <sup>a</sup> Epistatic <sup>a</sup> Mendelian <sup>a</sup>
Statistical parameters	LOD score LS mean <i>P</i> -value <i>F</i> -statistic Variance Bayes value <sup>a</sup> Likelihood ratio <sup>a</sup>
QTL descriptors	Trait name Breeds <sup>a</sup> Chromosome Flank Marker A2 Flank Marker A1 Peak Mark Flank Marker B1 Flank Marker B2
Significance levels	Suggestive Significant
Test scale	Genome-wise <sup>a</sup> Chromosome-wise <sup>a</sup> Comparison-wise <sup>a</sup> Experiment-wise <sup>a</sup>
QTL term mapping	Vertebrate Trait Ontology <sup>a</sup> Product Trait Ontology <sup>a</sup> Clinical Measurement Ontology <sup>a</sup>
Reference information	Author emails <sup>a</sup> (Other reference parameters omitted)
Species	Cattle Chicken Pig Sheep <sup>a</sup> Rainbow trout <sup>a</sup>

<sup>a</sup>New data types. Note that only necessary parameters are listed to save space.

map locations are visually comparable and (ii) use anchor markers between them to transfer map information. Figure 1b shows how genome and linkage maps are laid out, and how the GWAS and QTL are plotted to facilitate user analysis.

### Animal breeds

This information is important for analysis of the source of a QTL, although not all publications include such information. The introduction of breed information was carried out in late 2007 aided by a breed ontology database set up with initial data from the Breeds of Livestock Project at Oklahoma State University (<http://www.ansi.okstate.edu/breeds/>). The display of animal breed information on QTL detail queries was implemented in 2010 (Figure 3b), and the ability to search for QTL by animal breeds was implemented in 2012.

### Genetic test models

The original Pig QTLdb was developed only to accommodate classical, simple QTL association data. With the

development of QTL tests under various genetic models, the following options were added for data curation: ‘paternally imprinted’, ‘maternally imprinted’, ‘sex-specific’, ‘Mendelian’, ‘epistatic’ and ‘associated QTL’ information. The epistasis types include ‘antagonistic’, ‘synergistic’, ‘genetic suppression’, ‘genetic enhancement’, ‘intragenic complementation’, ‘allelic complementation’, ‘interallelic complementation’, etc. Collection of these data is useful to facilitate future genetic network analysis and meta-analysis.

### New test statistics

In addition to logarithm of the odds (LOD) score, least squares (LS) mean, *P*-value, *F*-statistic and variance, we have added options for Bayes value and likelihood ratio. This allows us to be as inclusive as possible for all QTL/association reports.

### QTL alignments to genome maps and cytogenetic G-band maps

#### Genome maps and cytogenetic G-band maps

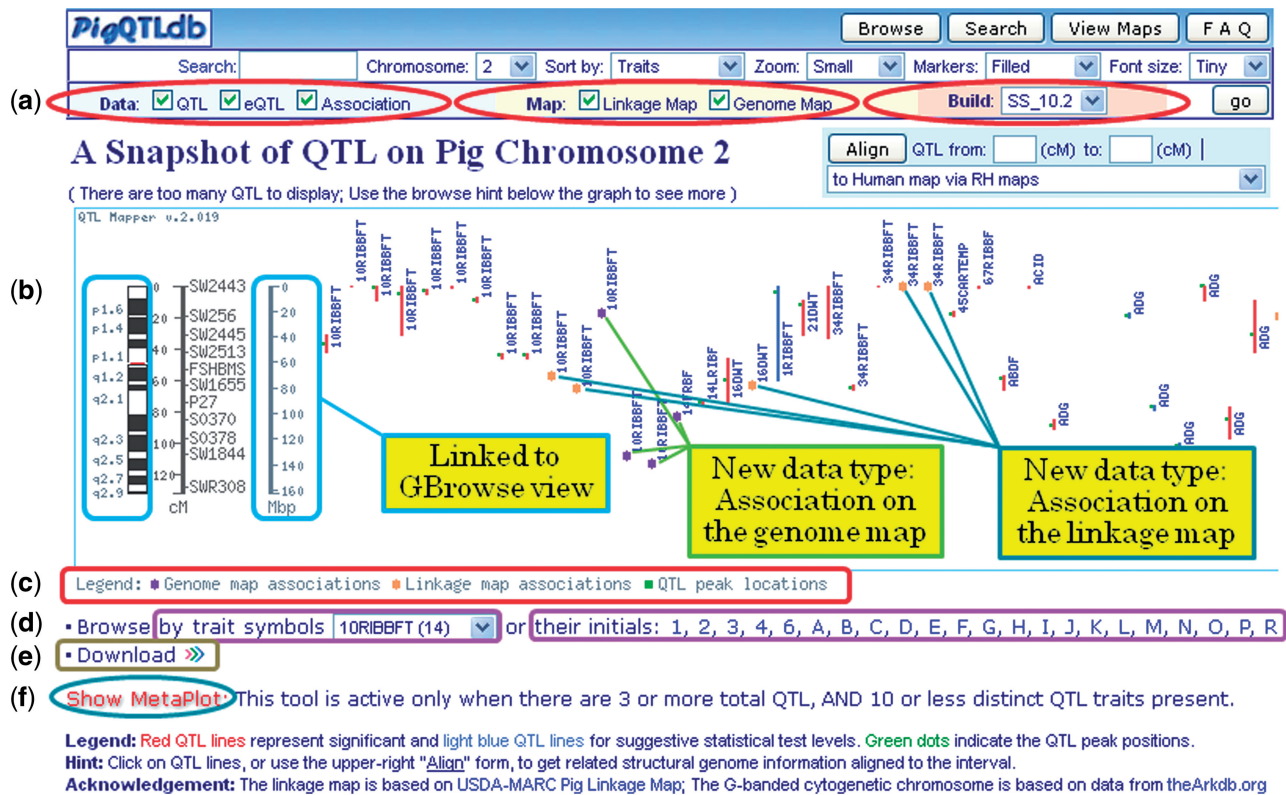
The alignment of QTL/association data is made both within the QTLdb using our own graphic tools, and by using GBrowse in a separate setup. The GBrowse setup is to accommodate QTL alignments in newly available genome assemblies for cattle (*Bos taurus*), chicken (*Gallus gallus*) and pig (*Sus scrofa*). In order to accommodate both linkage maps and genome maps to integratively display QTL data, a number of improvements had to be made. First, the back-end relational database was restructured to store and integrate genome maps parallel to linkage maps. Second, the QTL graph tool was rebuilt so that both genome coordinates (Mbp) and linkage map locations (cM) can be comparatively displayed (Figure 1; also see above).

Due to the large sizes of genome data (e.g. millions of rows of high-density SNP data and increasing), efforts have been made to optimize queries and snap views to improve the MySQL query efficiency and minimize any noticeable transit delays for users.

#### Conversion of coordinates

In order to transfer QTL from linkage maps to genome maps, anchor marker-based coordinate interpolations are used. Briefly, the genome Mbp coordinates corresponding to a QTL on the linkage map are converted from their linkage map locations using the closest anchor marker locations between the linkage and the genome maps as a reference. The error sizes of interpolated Mbp locations vary depending on the anchor markers available and their distances from the target QTL location. The map distance is estimated using a Mbp–cM factor calculated based on the actual linkage and genome map lengths of that particular genome or chromosome. As such, the interpolation is only an estimate. Although the estimates are ‘gross’, the error sizes are not significant relative to the size and test errors of most QTL.

The converted QTL genome locations are ported to GBrowse for display in alignment with NCBI (5) and/or Ensembl (6) annotated genes, as well as with the locations of array elements and HD SNPs. The data are also



**Figure 1.** A screenshot of an Animal QTLdb chromosomal map view with new elements or functions in boxes or in circles: (a) options to view different data types, allowing 'QTL', 'eQTL' and 'association' data to be displayed in various combinations; options to view 'linkage map', 'genome map' or both; and options to view different genome map builds if applicable. (b) Cytogenetic bands and genome coordinates aligned side-by-side with the linkage map. Note that the genome chromosome bar can be clicked to go to viewing QTL in GBrowse. (c) New shapes/colors for different data types. (d) All QTL/traits on a chromosome can be easily retrieved to view by selecting from a pull-down menu or by their first initials. (e) A flexible data download tool that allows users to download data in the current view or by chromosome. (f) A link to trigger MetaPlot function (Figure 5 for details).

available for free download, with warning messages reminding users to verify data before any critical use of the converted coordinates.

### Data display

Display of QTL and association data in the QTLdb native chromosomal view and in GBrowse are illustrated in Figures 1 and 2. Improvements have been made to the QTL data organization by trait hierarchies/trait-type groups to aid data dissemination.

### Development of ATO and mapping to QTL traits

In an effort to standardize trait nomenclature across species, the ATO has undergone active development into VT, PT and CMO (Table 2) as part of a collaboration between Animal QTLdb, Rat Genome Database (<http://rgd.mcw.edu/>), Mouse Genome Informatics (<http://www.informatics.jax.org/>), the French National Institute for Agricultural Research (INRA; <http://www.international.inra.fr>), SABRE Research UK (<http://www.sabre.org.uk>) and EADGENE (<http://www.eadgene.info>). To facilitate a smooth transfer of QTL information to the new trait terms, we have developed an interactive 'ATO to VT/PT/CMO mapping tool' as part of the QTLdb curator tool set, for curators to assign mapping relationships.

We have also undertaken a database restructuring for dynamic yet consistent trait hierarchy management to allow for easier future development. Figure 3a shows an example of the trait mapping results. As a by-product, a new trait hierarchy tree structure display and navigation tool using combined java scripts and W3 cascading style sheets has been created. Figure 4 shows a simple example, including new features built into the tool.

### QTL meta-analysis

#### Meta-plots

Probably the biggest advantage provided by the QTLdb is its utility for metadata analysis. Our preliminary work on making QTL meta-plots was reported earlier (7). Subsequent minor improvements have been made for better browser compatibility, among other issues. Currently available tools include a 'pile plot' (histogram) and a 'kernel density plot'. The pile plot is based on the actual counts ( $y$ -value) of the reported QTLs, plotted at 1-cM ( $x$ -value) bins along the target chromosome. The kernel density estimation is a non-parametric approach to estimate the probability density function of location-wise QTL incidence treated as a random variable. We have also recently modified the rules so that a group of selected traits, rather than the previous

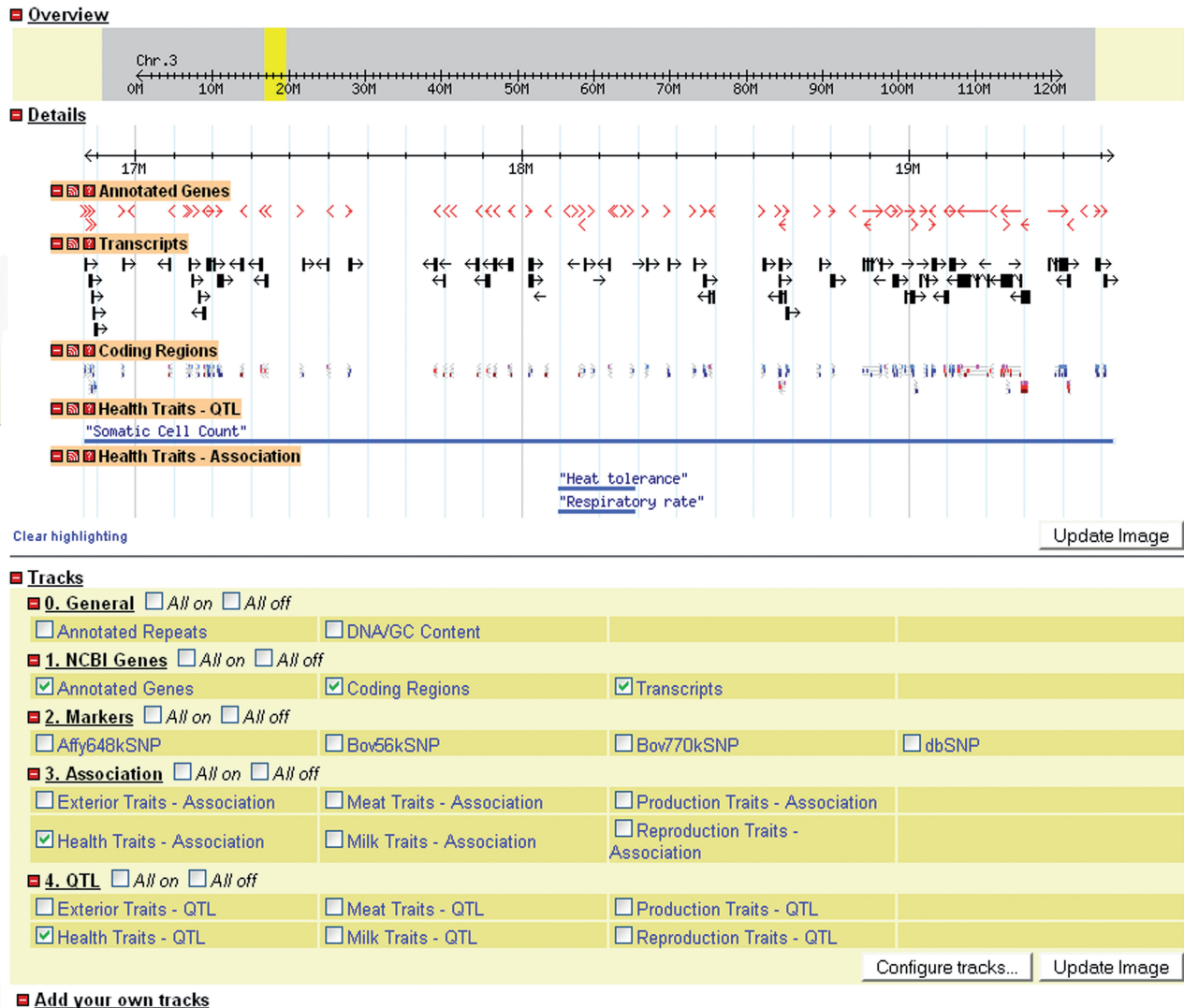


Figure 2. A snapshot of GBrowse display of QTL and association data on part of bovine chromosome 3 in alignment with annotated genes and transcripts on UMD3.1 genome build of *Bos taurus*.

limit of one trait, can be subjected to meta-plot analysis. For example, the trait for a meta-plot can be either a trait analyzed across multiple experiments or a group of similar traits abstracted to describe a scenario (Figure 5).

### Improved curator and editor tools

With the increased complexity and diversity of data types within the QTLdb, the post-curation data quality control (QC) and tasks required to run them through a release pipeline have introduced new challenges. We have improved the tools and methods for how this is handled, using redeveloped curator/editor tools and improved QC procedures.

#### Curator/editor 'realms'

Originally, a curator was given an account to curate data within a species. Additional accounts had to be created for each species for which a curator needed to have access. The various accounts were unwieldy, creating

inconvenience for curators and editors. We have unified the curator accounts so that one only needs to login once, then has options to go into any of the available species' 'realms' for his or her work. The most important improvement made to the curator/editor tools is the implementation of a number of QC rule sets and codes to alert the curator when a data integrity problem is identified. The problematic data are not accepted until the problem is fixed. At the database level, the problematic data are recorded rather than denied acceptance, but only flagged for their interim status. As such, multiple curators/editors may be able to have access to problematic data (as long as access rights are properly granted) in order to work together to solve the problem.

#### Automated PUBMED search and data pre-load

To reduce the curator's workload and keep track of incoming and curated data, an automated procedure was implemented in Perl script to perform periodic searches of PubMed via NCBI eUtil portal (<http://www.ncbi.nlm.nih>).

### QTL #10651 Description:

QTL Trait Information	
QTL Trait: <b>Carcass weight</b>	Vertebrate Trait Ontology: n/a
	Product Trait Ontology: <b>Dressed carcass weight</b>
QTL Symbol: <b>CWT</b>	Clinical Measurement Ontology: <b>Carcass weight</b>

QTL Map Information	
Chromosome:	1
QTL Peak Location:	148 (cM)
QTL Span:	139.3-154.6 (cM)
Flanking markers	Upper, "Suggestive": n/a
	Upper, "Significant": BMS599
	Peak: BMS4014
	Lower, "Significant": #
	Lower, "Suggestive": n/a
(c) Analysis type:	QTL
Model tested:	Mendelian
Test base:	Genome-wise
Threshold significance level:	Significant
Test Statistics	LOD Score: n/a
	Least Square Mean: n/a
	P-values: 4.69
	F-Statistics: n/a
	Variance: n/a
	Bayes value: 6
	Likelihood Ratio: n/a
	Dominance effect: 0.0132
	Additive effect: n/a
	Candidate Gene: n/a
Links: NCBI GeneDB   Edit   Map view	

QTL Experiment in Brief	
Animals:	Twenty-nine generations of commercial American Angus.
Breeds that this QTL is associated with:	Angus
Design:	Genome scan with 402 markers at 8 cM intervals on 14 cattle production traits using 3391 animals from 48 half-sib families.
Analysis:	General pedigrees based genetic model analysis using le Markov
Software:	GridQTL
Notes:	Not a cross based segregation analysis
Links:	Edit

Reference	
Authors:	McClure MC, Morsci NS, Schnabel RD, Kim JW, SJ, Chapple
	University of Missouri,
	loci influencing productive traits in
	commercial Angus cattle.
Year:	2010
Journal:	Anim Genet.
Issues:	41 (6): 597-607
Links:	PubMed   Abstract   Paper   Edit

Additional Information	
Comments:	SNP scan

**Figure 3.** A snapshot of an Animal QTLdb data details page, showing new parameters and features added to the database. The new parameters subject to collection into the database include data analysis types (c), test models (c) and animal breeds (b). The ATO is now linked to VT, PT and CMO (a).

**Table 2.** Definitions of VT, PT and CMO—ontology realms within which the livestock ATO is undergoing development for unified trait term standards (8)

- The VT is a controlled vocabulary for the description of traits (measurable or observable characteristics) pertaining to the morphology, physiology or development of vertebrate organisms.
- The CMO is designed to be used to standardize morphological and physiological measurement records generated from clinical and model organism research and health programs.
- The PT is a controlled vocabulary for the description of traits (measurable or observable characteristics) pertaining to products produced by or obtained from the body of an agricultural animal or bird maintained for use and profit.

gov/books/NBK25501/). We built a local search and track database, with which we keep a record of which papers have been curated, which have high priority in the queue, which are being reviewed or are on hold, and which are not applicable. This effectively facilitates collaboration by multiple curators. Recently, we have established a ‘curators’ mailing listserv’ to share experiences and lessons, to inform curators of the curation queue status, etc.

**Database release procedures and tools**

We have implemented a nine-step data release procedure, which includes a number of post-curation data fixes [such as flanking marker/location (cM) validation, curation evidence updates, pre-release integrity check-ups, etc.] and administrative actions. The latter includes rolling new data into public portals, exporting data for external collaborative database synchronization, such as to NCBI (<http://www.ncbi.nlm.nih.gov/gene>) and Thomson Reuters Digital Resources (<http://wokinfo.com/news/new/>) and generation of download files in each of the

four supported formats (raw, GFF, SAM and BED), generation of new database statistics and updating the web portal, rebuilding of GBrowse database tracks for each species and finally, retrieval of updates from NCBI on newly assigned Gene Database unique identifiers for dynamic inter-database links. Supplementary Figure S2 shows the conceptual work/data flow pipeline implemented in the QTLdb data curator tools, where the roles of curators, editors and database administrators are sketched along with the data flow. The same procedure is used for problematic data debugging, rollback or data obsolescence (not shown).

With each release, new functions are introduced, although their actual implementations are made seamlessly over time. That is, whenever a new function is developed, tested and passed, it is quietly rolled into public portals. Quality assurance is fulfilled by routine maintenance and problem fixes. New functions are normally announced collectively along with data releases.

**Miscellaneous improvements**

Many ‘small’ improvements and bug fixes were made over the past 5 years. These include, but are not limited to: (i) improved literature search tools (structured search for QTL publications that helps users to better target what they look for); (ii) a pull-down menu to list unique traits on a chromosome view (with QTL counts for each trait and an option to display selection); (iii) improved tool bar on chromosome view (to allow display choices of QTL/eQTL/association data or linkage/genome maps); (iv) improved QTL search on chromosome view (to allow the display of combinations of QTL per user’s choice); (v) intelligent alert of search results if not found within a species but similar results exist in another species (QTL ID, traits, etc.); (vi) genome-wide view of QTL by trait types; (vii) improvement of QTL map image quality (facilitates use for publications); (viii) customized data download of user’s searched/browsed dataset (at chromosomal level) and (ix) data downloads, exports and sharing (in the formats of GFF, BED and SAM, as well as tab-delimited plain text files for each individual chromosome and for the whole genome of a species). Data are also exported into specific formats designed to work with

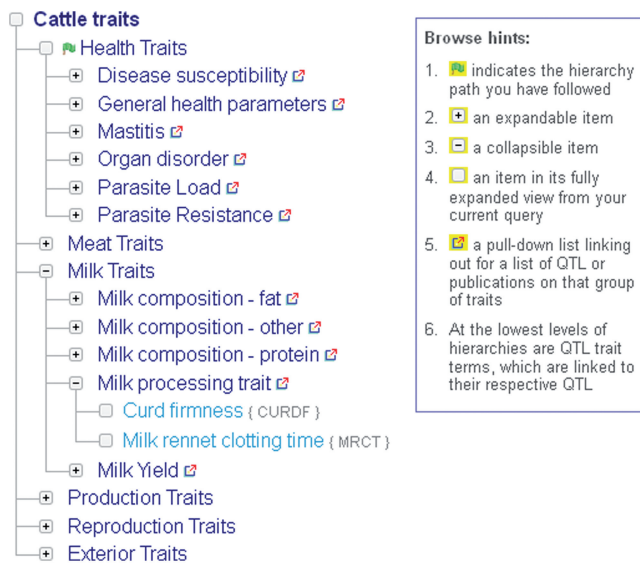


Figure 4. A new hierarchy display tool showing a partial cattle trait tree, along with main features embedded in the display.

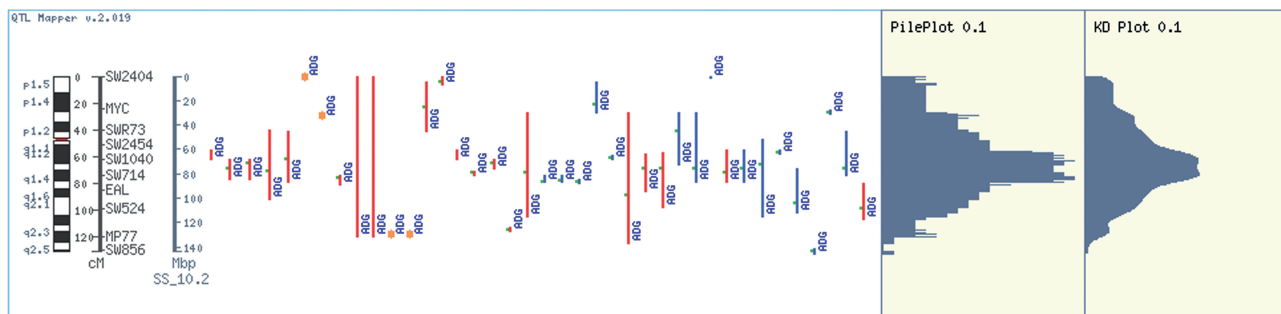


Figure 5. An example of QTL meta-plots, showing a pile plot and a kernel density (KD) plot, for average daily gain (ADG) QTL on pig chromosome 4. The meta-plot peak is supported by the number of reported QTL that overlap. Note this function is active only when there are more than three QTL in a chromosome view (Figure 1).



the NCBI Gene Database and Thomson Reuters Digital Resources, respectively.

## DISCUSSION AND FUTURE DIRECTIONS

A QTL is a map feature that describes a location in a genome where genes underlying quantitative traits of interest may reside. Therefore, the development of QTLdb and incorporated tools and utilities is all map-centric. Numerous improvements to the QTLdb have been made over the past 5 years. Although the most notable changes include the addition of sheep and rainbow trout data, as well as a number of new parameters and features seen on the web portal, the most significant 'upgrade' has been the inclusion of GWAS data. As such, the QTLdb is in fact an animal QTL/association database.

Ideally, a well-designed database should not only completely fit the scope of data that are subject to curation (entry) and employ well-structured database management but also allow for meta-analysis, where data can be easily fit into an analysis grid. The addition of the meta-plot tools is only a starting point, because there are numerous research questions to address before more tools may be implemented (8). Moreover, the nature of QTL/GWAS/eQTL experiments presents some challenges for glitch-free database development in terms of ideal fit of data into database metrics. This is because the genetic data structure has multiple dimensions, and each of these dimensions may follow different standards for data representation and recording. As genetics and genomics rapidly evolve, database development also needs to keep up with the challenges, not only for proper recording of these data, but also for building platforms for the data to be comparable, so that general conclusions may be drawn through meta-analysis. We have made efforts to define the minimum information necessary (9) to describe a QTL or association for effective database management. This effort has been geared at setting up standards for data submission, as well as toward the inclusion of related data that are usually not publishable, but may be useful for meta-analysis. Recently, *Nature Genetics* (10) has requested that authors, when reporting GWAS data, also report the co-localization of trait-associated variants identified by other methods. In particular, they ask authors to publish or enter into databases the genotype frequencies, association *P*-values, etc. for data that may or may not reach genome-wide significance thresholds. This reflects a consensus within the community that having more information available will be useful for future combined meta-analyses. We have enhanced our efforts toward developing and implementing minimum information for QTL and association studies (MIQAS) through the continued improvement of the QTLdb.

Complete inclusion of all published data is crucial for the QTLdb to avoid any possible bias in the representation of a true QTL. Since a comprehensive data update requires extensive time and effort, we have endeavored to open the QTLdb to the public for data entry, and welcome users to volunteer their data to the database. Although it takes time to catch up with the incorporation of newly

available data, it also takes time to roll out new features and utilities for complex database development. This is because developmental processes need to go through trials, system integration, debugging, overall tests and possibly rollback and application of new patches, etc. Often, changes to an external database require us to re-develop certain functions (as when we went through changes in coordination with the transition at NCBI from Locus Link to Gene Database). One recent example is our cooperation with NCBI to implement the deletion of a QTL record when one becomes obsolete at the QTLdb site (primary data source). It is thus important to work closely with partner databases and check for any functional failure and/or incompatibility as a matter of routine.

For successful development of a comprehensive and shared tool set for multiple curators and users like the Animal QTLdb, details are extremely important. We deem nothing so minor as to not be relevant to the advancement of science or users' desire to easily find information. This applies to the development of curator tools, user interfaces, web layout and cosmetics as well. In terms of presentation of relevant information, we have attempted to make the user interface as brief, slick, interactive and to-the-point as possible. For example, there are more than two dozen new features or functions that we have implemented to improve the user experience. Some may seem like only small or minor improvements, but altogether they help to keep the QTLdb a user-friendly and steadily useful tool.

Due to the complexity and size of genome data, as well as the relatively fast update pace of new genome builds, keeping track of the updates and versioning of genome builds within the QTLdb has been a challenge, especially considering that it also involves mapping of many features. We envision that while QTL/eQTL/association studies will continue to be map based, future development of the database will be more sequence centered and gene or genetic network analysis oriented. As such, future QTLdb development will involve not only mapping features but also genetic factors contributing to our understanding of the connection between genes and traits.

## AVAILABILITY

The database contents and online tools are all freely available at <http://www.animalgenome.org/QTLdb/>. The Animal QTLdb welcomes users to directly deposit their data by applying for a curator account (<http://www.animalgenome.org/QTLdb/app>). We also maintain a frequently asked questions (FAQ) page to serve as a user guide to database functions (<http://www.animalgenome.org/QTLdb/faq>).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures S1 and S2. Supplementary data to this manuscript can be found at: <http://www.animalgenome.org/repository/pub/ISU2012.1004/>.

## ACKNOWLEDGEMENTS

We thank Jill Maddox from the University of Melbourne, Australia and Yniv Palti from the National Center for Cool and Cold Water Aquaculture (NCCCWA), for initiation of the sheep and rainbow trout QTL data curation, respectively. Deep appreciations are due to Andy Law from Roslin Institute for kindly providing the cytogenetic G-band measurements of cattle, chicken, pigs and sheep; to Wonhee Jang and Donna Maglott from NCBI for their efforts to streamline the QTL updates into the NCBI Gene Database; to Daniel Auld from Thomson Reuters for QTL data streamlining to their Digital Resources and to James Koltes for his useful suggestions from a user's perspective. User feedback, various requests, constructive criticisms and suggestions received through the Helpdesk or related collaborations over the past years have been invaluable in our determination of the most useful new developments and in improvement of QTLdb ease of use.

## FUNDING

The USDA NRSP-8 National Animal Genome Research Program, Bioinformatics Coordination Project and partly by the USDA-NRI [2007-04187]. Funding for open access charge: USDA-NIFA Research Funds to the National Research Service Program, NRSP-8, National Animal Genome Research Program.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Hu,Z.-L., Dracheva,S., Jang,W., Maglott,D., Bastiaansen,J., Rothschild,M.F. and Reecy,J.M. (2005) A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm. Genome*, **16**, 792–800.
2. Hu,Z.-L., Fritz,E.R. and Reecy,J.M. (2007) Animal QTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Res.*, **35**, D604–D609.
3. Hu,Z.-L., Park,C.A., Fritz,E.R. and Reecy,J.M. (2010) QTLdb: A Comprehensive Database Tool Building Bridges between Genotypes and Phenotypes. *Invited Lecture at the 9th World Congress on Genetics Applied to Livestock Production*, 1–6 August 2010, Leipzig, Germany.
4. Shimoyama,M., Nigam,R., McIntosh,L.S., Nagarajan,R., Rice,T., Rao,D.C. and Dwinell,M.R. (2012) Three ontologies to define phenotype measurement data. *Front Genet.*, **3**, 87.
5. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39** (Suppl. 1), D52–D57.
6. Flicek,A., Amode,M.R., Barrell,D., Beal,K., Brent,S., Denise,C.-S., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.
7. Hu,Z.-L., Wu,X.L. and Reecy,J.M. (2011) Extension of Animal QTLdb: QTL meta-analysis on the fly. In: *Poster paper published on ACM Conference on Bioinformatics, 1–3 August 2011*. Computational Biology and Biomedicine (ACM-BCB), Chicago, IL, USA.
8. Wu,X.-L., Gianola,D., HuZ-L. and Reecy,J.M. (2011) Meta-analysis of quantitative trait association and mapping studies using parametric and non-parametric models. *J. Biomet. Biostat.*, **S1**, 001.
9. Aerts,J., Carre,W., De Koning,D.J., Hu,Z.-L., Burt,D., Law,A. and Reecy,J.M. (2008) MIQAS - Minimal Information for QTL and Association Studies. *Plant & Animal Genome Conference XVI*, 12–16 January 2008. Town & Country Convention Center, San Diego, CA, USA.
10. Nature Genetics Commentary. (2012) Asking for more. *Nat. Genet.*, **44**, 733.