

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:

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

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.

Upload MIQAS compliant QTL/association data
(.xls/.xlsx/.txt/.doc/.docx files)

Curator: ZhiLiang Hu
Species: Rainbow trout

Following the minimum required information, you should have your data with these info:

- Publication or Manuscript
- Experiment information
- Trait (definition/description)
- QTL/Association data (include mapping/statistics/etc)

Uploaded files in your data realm:

- 30635785_supp1.xlsx (1.99 MB, Jun 4, 2019, Curator: Cari Park, Curated: 2019-06-10)

NOTE:
1. Your uploaded files will be on one of our curator/editor's desk.
2. New uploads with the same file name will override the existing ones.

Select files to upload:

File 1: [Choose File] No file chosen
File 2: [Choose File] No file chosen

[Upload Now]

User/Curator uploaded data for curation

NOTES TO EDITORS:

- You can claim/download a data set for curation if available; come back to mark it off when done.
- The uploaded data are supposed to be prepared following the minimum requirement. Contact the data owner in case of problems.

Data owner:	File name	Uploaded	Size	Curator	Curated
ZhiLiang Hu (zhihanghu@gmail.com)	30635785_supp1.xlsx Curated QTL IDs: 170066-170079	Jun 4, 2019	1.99 MB	Cari Park	2019-06-10

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).

Publication # 24906442 Details

Authors: Mahdi Saatchi, Robert D Schnabel, Jeremy F Taylor and Dorian J Garrick
Affiliation: Iowa State University

Title: Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds

Journal: BMC Genomics, 2014, 15:442 DOI: 10.1186/1471-2164-15-442

Links: PubMed | Abstract | Paper (Related study: none)

Number of QTL/association reported by traits:

- QTL for Body weight (birth) reported for chromosome(s): Asso:24792, Asso:24793, Asso:24794, Asso:24795, Asso:24796, Asso:24797, Chr.14: Asso:24798, Asso:24799, Asso:24800, Asso:24801, Asso:24802, Chr.15: Asso:24803, Chr.19: Asso:24804, Chr.20: Asso:24805, Asso:24806, Asso:24807, Asso:24808, Chr.26: Asso:24809, Chr.29: Asso:24810.

QTL/Association Reported by chromosome:

Chr.X (4), Chr.1 (11), Chr.2 (10), Chr.3 (4), Chr.4 (6), Chr.5 (16), Chr.6 (56), Chr.7 (34), Chr.8 (3), Chr.9 (5), Chr.10 (4), Chr.11 (5), Chr.12 (4), Chr.13 (5), Chr.14 (30), Chr.15 (7), Chr.16 (6), Chr.17 (10), Chr.18 (5), Chr.19 (5), Chr.20 (26), Chr.21 (6), Chr.22 (1), Chr.23 (2), Chr.24 (3), Chr.25 (4), Chr.26 (6), Chr.27 (2), Chr.29 (8).

Pleiotropic QTL found:

- BTA2_6, on Chr.2 : 7 QTL: involving 6 traits in 2 breeds.
- BTA5_48-50, on Chr.5 : 6 QTL: involving 4 traits in 2 breeds.
- BTA5_106, on Chr.5 : 3 QTL: involving 3 traits in 1 breeds.
- BTA6_37-42, on Chr.6 : 47 QTL: involving 8 traits in 7 breeds.
- BTA7_33, on Chr.7 : 21 QTL: involving 6 traits in 6 breeds.
- BTA14_23-26, on Chr.14 : 26 QTL: involving 5 traits in 3 breeds.
- BTA15_38-39, on Chr.15 : 3 QTL: involving 3 traits in 1 breeds.
- BTA20_4, on Chr.20 : 22 QTL: involving 8 traits in 4 breeds.
- BTA29_30, on Chr.29 : 3 QTL: involving 3 traits in 1 breeds.

Show pleiotropic QTL groups:

The publication:
Mahdi Saatchi, Robert D Schnabel, Jeremy F Taylor and Dorian J Garrick (2014): Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics, 15:442

The Pleiotropic Group Information:

- Pleiotropy Group ID: 5
- Pleiotropy Group Name: BTA2_6
- Chromosome: 2
- Pleiotropic QTL: QTL24522, QTL24560, QTL24660, QTL24689, QTL24710, QTL24747, QTL24752, QTL24753, QTL24754, QTL24755, QTL24756, QTL24757, QTL24758, QTL24759, QTL24760, QTL24761, QTL24762, QTL24763, QTL24764, QTL24765, QTL24766, QTL24767, QTL24768, QTL24769, QTL24770, QTL24771, QTL24772, QTL24773, QTL24774, QTL24775, QTL24776, QTL24777, QTL24778, QTL24779, QTL24780, QTL24781, QTL24782, QTL24783, QTL24784, QTL24785, QTL24786, QTL24787, QTL24788, QTL24789, QTL24790, QTL24791, QTL24792, QTL24793, QTL24794, QTL24795, QTL24796, QTL24797, QTL24798, QTL24799, QTL24800, QTL24801, QTL24802, QTL24803, QTL24804, QTL24805, QTL24806, QTL24807, QTL24808, QTL24809, QTL24810, QTL24811, QTL24812, QTL24813, QTL24814, QTL24815, QTL24816, QTL24817, QTL24818, QTL24819, QTL24820, QTL24821, QTL24822, QTL24823, QTL24824, QTL24825, QTL24826, QTL24827, QTL24828, QTL24829, QTL24830, QTL24831, QTL24832, QTL24833, QTL24834, QTL24835, QTL24836, QTL24837, QTL24838, QTL24839, QTL24840, QTL24841, QTL24842, QTL24843, QTL24844, QTL24845, QTL24846, QTL24847, QTL24848, QTL24849, QTL24850, QTL24851, QTL24852, QTL24853, QTL24854, QTL24855, QTL24856, QTL24857, QTL24858, QTL24859, QTL24860, QTL24861, QTL24862, QTL24863, QTL24864, QTL24865, QTL24866, QTL24867, QTL24868, QTL24869, QTL24870, QTL24871, QTL24872, QTL24873, QTL24874, QTL24875, QTL24876, QTL24877, QTL24878, QTL24879, QTL24880, QTL24881, QTL24882, QTL24883, QTL24884, QTL24885, QTL24886, QTL24887, QTL24888, QTL24889, QTL24890, QTL24891, QTL24892, QTL24893, QTL24894, QTL24895, QTL24896, QTL24897, QTL24898, QTL24899, QTL24900, QTL24901, QTL24902, QTL24903, QTL24904, QTL24905, QTL24906, QTL24907, QTL24908, QTL24909, QTL24910, QTL24911, QTL24912, QTL24913, QTL24914, QTL24915, QTL24916, QTL24917, QTL24918, QTL24919, QTL24920, QTL24921, QTL24922, QTL24923, QTL24924, QTL24925, QTL24926, QTL24927, QTL24928, QTL24929, QTL24930, QTL24931, QTL24932, QTL24933, QTL24934, QTL24935, QTL24936, QTL24937, QTL24938, QTL24939, QTL24940, QTL24941, QTL24942, QTL24943, QTL24944, QTL24945, QTL24946, QTL24947, QTL24948, QTL24949, QTL24950, QTL24951, QTL24952, QTL24953, QTL24954, QTL24955, QTL24956, QTL24957, QTL24958, QTL24959, QTL24960, QTL24961, QTL24962, QTL24963, QTL24964, QTL24965, QTL24966, QTL24967, QTL24968, QTL24969, QTL24970, QTL24971, QTL24972, QTL24973, QTL24974, QTL24975, QTL24976, QTL24977, QTL24978, QTL24979, QTL24980, QTL24981, QTL24982, QTL24983, QTL24984, QTL24985, QTL24986, QTL24987, QTL24988, QTL24989, QTL24990, QTL24991, QTL24992, QTL24993, QTL24994, QTL24995, QTL24996, QTL24997, QTL24998, QTL24999, QTL25000.
- Pleiotropy Type: Co-occurrence
- Breeds:
 - Limousin
 - Maine-Anjou
- Traits:
 - Marbling score [MARBL]
 - Yield grade [YGRADE]
 - Body weight (birth) [BW]
 - Body weight (weaning) [WWT]
- Comments: Breed-specific pleiotropic QTL (associated with more than one trait but segregating in only one)

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.

CattleQTLdb Browse Search View Maps F A Q

Whole genome enrichment analysis by χ^2 tests of QTL/associations discovered throughout the genome

The design of this trait enrichment analysis tool aims at allowing evaluation of all reported QTL/associations of selected traits throughout a genome, to determine if a trait or traits are over-represented in one or more regions of the genome. The setup of the analysis is based on an underlying assumption, that the selected traits are related. For example the traits may belong to a given trait type, or are from a given trait ontology path.

In this implementation we used Chi-squared analysis of the frequencies of reported QTL/associations classified by traits and their chromosome locations. The contingency p-values (p) are estimated to indicate the degree of over-representations (enrichments) of QTL/associations. The false discovery rates (FDR) are also estimated using the Benjamini-Hochberg procedure. The size of Chi-squares in each contingency class are graphically indicated with bars of varying lengths drawn with transformed Chi-square values.

In addition, some available trait correlation data are appended when available, to provide additional supporting information to help users to evaluate the results.

Step 1: Select at least 1 trait check all uncheck all

- 305-DAY MILK YIELD (235)
- ABOMASUM DISPLACEMENT (166)
- AGE AT FIRST CALVING (107)
- AGE AT PUBERTY (10611)
- AVERAGE DAILY GAIN (965)
- BODY DEPTH (592)
- BODY WEIGHT (128)

Step 2: Input at least 1 targeted chromosome segment for analysis.

Format:
"chr#:coordA-coordB bp" OR
"chr#:coordA-coordB Mbp"

2:1200000-28000000 bp
3:1200000-28000000 bp
6:1.2-28.0 Mbp
8:1.2-28.0 Mbp

[Start]

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.

Whole genome QTL/associations enrichment analysis by χ^2 tests

Interested genome regions:
2:1200000-28000000 bp
3:1200000-28000000 bp
6:1.2-28.0 Mbp
8:1.2-28.0 Mbp

Select at least 1 trait (Filter the list by:)

- 305-DAY MILK YIELD (235)
- ABOMASUM DISPLACEMENT (166)
- AGE AT FIRST CALVING (107)
- AGE AT PUBERTY (10611)
- AVERAGE DAILY GAIN (965)
- BODY DEPTH (592)
- BODY WEIGHT (128)

Genome builds:
UMD_3.1

Data types:
all

[Submit]

Data scope:

Selected traits: 7
Chromosomes to flag: 6
Test method: χ^2 test

The purpose of this analysis is for users to examine potentially over-represented traits of interest in user-defined genome regions (chromosomes/sub-sections of a chromosome) in comparison to the rest of the genome. This analysis is designed to provide clues to facilitate discovery of co-located traits that may show something in common, or to indicate areas for further data mining.

Results: Where QTL/associations are found in part of the genome versus the rest of the genome to discover possible regional over-representations

- QTL/association representations by each chromosome
- QTL/association representations by user-defined genomic section sizes (seen upon expansion of each chromosome)
- 7 traits tested:
 - Average daily gain
 - Body weight (birth)
 - Body weight (weaning)
 - Body weight (yearling)
 - Body weight (mature)
 - Body weight
 - Body weight gain

Chromosome	Traits	df*	Total χ^2	p-values	FDR	Size of χ^2
Chr.X	(7 traits combined analysis)	29	11.53333	9.984213e-01	1	
Chr.1	(7 traits combined analysis)	29	1.72987	1.000000e+00	1	
Chr.2	(7 traits combined analysis)	29	6.21541	9.999977e-01	1	
Chr.3	(7 traits combined analysis)	29	11.40039	9.985827e-01	1	
Chr.4	(7 traits combined analysis)	29	9.49865	9.997612e-01	1	
Chr.5	(7 traits combined analysis)	29	18.14605	9.412419e-01	1	
Chr.6	(7 traits combined analysis)	29	26.45241	6.01228e-01	1	
Chr.7	(7 traits combined analysis)	29	1.72987	1.000000e+00	1	
Chr.8	(7 traits combined analysis)	29	2.65472	1.000000e+00	1	
Chr.9	(7 traits combined analysis)	29	9.49865	9.997612e-01	1	
Chr.10	(7 traits combined analysis)	29	1.72987	1.000000e+00	1	
Chr.27	(7 traits combined analysis)	29	4.92062	9.999999e-01	1	
Chr.28	(7 traits combined analysis)	29	7.88015	9.999666e-01	1	
Chr.29	(7 traits combined analysis)	29	11.53333	9.984213e-01	1	

*: Represent region(s) that contain user selected chromosome sections.

References

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