# 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 codevelopment 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.

#### Figure 2

AnimalQTLdb

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.

Traits: 1,91

ces SN

e! Ensem

https://www.animalgenome.org/QTLdb

The Animal Quantitative Trait Loci (QTL) Database (Animal QTLdb) strives to collect al

publicly available trait mapping data, i.e. QTL (phenotype/expression, eQTL), candidate

gene and association data (GWAS), and copy number variations (CNV) mapped to

ivestock animal genomes, in order to facilitate locating and comparing discoveries within

and between species. New data and database tools are continually developed to align

be deposited into a public database before a paper may be accepted for publication. We

provide user/curator accounts for direct data submission, and in return we supply users with

Many scientific journals require or recommend that any original QTL/association data

various trait mapping data to map-based genome features such as annotated genes.

a data summary link to facilitate the manuscript review process

New data releases

3 times a year

https://www.animalgenome.org/CorrDB

Release 2 (Dec 21, 2017)

#### CorrDB Animal Trait Correlation Database

Genetic and phenotypic trait correlation studies have been conducted for 70+ years/since Hazel proposed the methods. A genetic correlation is the proportion of shared variance between two traits that is due to genetic causes; a phenotypic correlation is the degree to which two traits co-vary among individuals in a population. In the genomics era, while gene expression, genetic association, and network analysis provide unprecedented means to decode the genetic basis of complex phenotypes, it is important to recognize the possible effects genetic progress in one trait can have on other traits. This database is designed to collect all published livestock genetic/phenotypic trait correlation data, aimed at facilitating genetic network analysis or systems biology studies.

As of the current Release (#2; also refer to QTLdb Release #44 for related issues), the CorrDB has 4,470 correlations and 698 heritibilities data on 302 economically important traits of cattle, relating to meat production, milk production, growth, health, and others. Information in the CorrDB can be explored through these access points:



## 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-

#### 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.



### centric views" of QTL/associations (Figure 5).

#### 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).

CattleQ1	Ldb				Bro	owse S	earch	View Maps	5 FAQ	
0	<b>T</b>	a fach la anna				. [				
Cattl	e Traits m	•••• Your search:go								
( LAYOUT FORMAT: <u>Trait Class</u> ► <u>Trait Type</u> ► <u>Trait Name</u> )						O In all traits ○ In VT/LPT/CMO traits Group results by ○ VT ○ LPT ○ CMO				
			·· <b>!·</b> See genome distributions on this trait ௴							
1.										
N	1ilk ► Milk yield ► 30	95-day milk yield	[VT:0015043: milk amount] [CMO:0001308: 305-day milk yield]							
c	Correlation data mate	bing this trait: Milk y	[enrere		o day ini	int yiolog				
	Correlation data mate			nilk (5)						
		anny this trait. Cumu	anve volume of t							
[	52 QTL found ]									
2.										
N	1ilk ► Milk yield ► Av	verage daily milk yiel	[VT:0015043: milk amount] [CMO:0001307: average daily milk yield]							
C	Correlation data mate	hing this trait: Milk y	ield [3]							
C	Correlation data mate	hing this trait: <mark>cumu</mark>	lative volume of r	nilk [5]						
А	ssociated gene: LE	P (leptin)								
1	23 QTL found ]									
Corrl	-	Trait Correl	lation Date	abaso						
COITI	Search: m					-112		in Casti		
	Search		Limit to: 60	genetic	<b>)</b> (	all) correlati	ons	ᅌ in 🛛 cattl	le ᅌ go	
Query	Results									
Dow	vnload G	iraphic View	Matrix View		Table Vie	w				
Cattle tra	ait correlations	•								
Species	Trait A	Trait B	Correlations	Std	Err	p-values	Correl	ation type	Reference	
cattle	milk yield QTL	milk yield QTL	0.00				G	enetic	175	
cattle	milk yield QTL	protein yield	0.75				G	enetic	175	
cattle	milk yield QTL	protein yield	0.89				<u> </u>	enetic	175	
cattle	protein yield	milk yield QTL	0.81					enetic	175	
cattle	Fat yield	milk yield QTL	0.51					enetic	175	
cattle	Milk yield QTL	Ketosis	-0.154					enetic	184	
cattle		Ketosis	-0.261					enetic	184	
cattle cattle	Milk yield QTL Milk yield QTL	Ketosis Ketosis	0.012					enetic	184	
cattle			0.036					enetic enetic	184 184	
cattle	Milk viold VIL	K OTOCIC	1 11050							
cattle cattle	Milk yield QTL Milk	Ketosis Fat	0.060					enetic	185	

#### 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.

CattleQTLd	b				Browse	Search	View Maps	FAQ		
Gene se	arch	n results:								
		Gene centr Trait centric view	ic view	Search: leptin						
				Sort results by Trait	t ᅌ		GO	J .		
	1.	Gene Symbol:	LEP							
		Gene Name:	Leptin							
		Also known as: leptin (obesity homolog, mouse); obesity factor								
		96 traits are associated with this gene, with 307 QTL/association found								
	2.	Gene Symbol:	LEPR							
		Gene Name:	Leptin receptor							
		Also known as:	OB receptor							
		4 traits are associated with this gene, with 6 QTL/association found								
	N	Map them by chrom	osomes: Chr.3,	Chr.4.						
CattleQTLdk	0				Browse	Search	View Maps	FAQ		
Gene sea	arch	results:								
	<u> </u>	Gene centric view		Search: leptin						
ļ	o Trait centric view Sort results by Trait ♀ GO									
	Trait name:         305-day milk yield           Ontology terms:         VT:0015043, CMO:0001308									
	1 gene are associated with this trait, with QTL/association found									

## 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.

"LEP" Leptin evidenced by: QTL:18821;

Trait name: Age at first calving

Ontology terms: VT:0002683, CMO:0002510

1 gene are associated with this trait, with QTL/association found

"LEP" Leptin evidenced by: QTL:18820; QTL:31175;

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.

Acknowledgments: This work was supported in part by USDA Grant 2016-67015-24767 and USDA NRSP-8 Bioinformatics Coordination project.