An implementation of new approaches to extend livestock trait ontologies for practical curation management of QTL, association, correlation, and heritability data
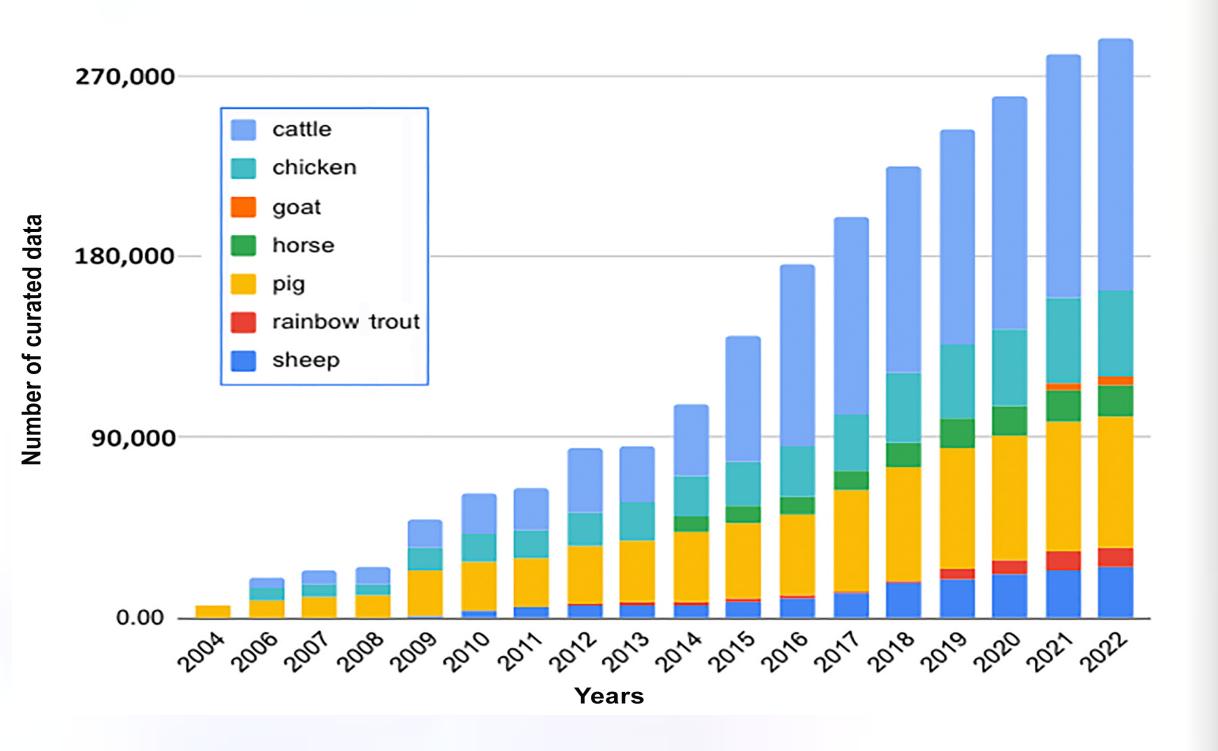
Zhi-Liang Hu, Carissa A. Park, and James M. Reecy

Department of Animal Science, Iowa State University, 2255 Kildee Hall, 806 Stange Road, Ames, IA 50011

Abstract

In genetics studies to improve livestock production, it is essential to accurately identify and record animal production traits. However, it has always been a challenge in research or production environments to unambiguously and consistently name and compare traits of interest that are evaluated over time, at different locations, and by different people. There have been efforts to establish frameworks using controlled vocabularies, concept hierarchies, and defined relationships between terms in an attempt to standardize these unwieldy trait names. However, gaps still remain regarding what level of concept granularity should be retained to formulate proper, precise, and reusable terms for description in a long-term, sustainable data management system. In recent updates to the Animal QTLdb and CorrDB, we have introduced an approach to extend livestock trait ontologies using trait modifiers and quantifiers to classify traits that differ slightly in how they are measured, evaluated, or combined with other factors ("quantifiers" or "descriptors"). Since the additional information cannot be accommodated by extension of formal ontologies, we previously organized these modified traits as "sibling traits". The new structure allows us to manage the extended trait/modifier information at the experiment level, as "trait variants". Here we describe how this system has been implemented in the Animal QTLdb and CorrDB, simplifying the curation and management of such trait information.

Introduction

The curated data in the Animal Quantitative Trait Loci Database (QTLdb) and Animal Trait Correlation Database (CorrDB) have been steadily increasing over the past 18 years (1; Figure 1). Since the data curated into the database are from thousands of scientific papers published in more than 200 journals, it has been a continual challenge to develop, improve, and maintain a sustainable database structure for sustainable developments. Proper curation and management of trait information from livestock production and research activities remain a challenge. Previously, we implemented "sibling traits" (2, 3) to extend complex trait names while including one or more "modifiers", but this method also introduced new challenges. Recently, we introduced a new "trait variant" concept and structure to target the problem, and we report initial success with this approach.

Results

Concepts. A trait term may be "modified" by a property, or "descriptor", such as intramuscular fat content (IMF) measured in different muscles or using different methods. IMF is the base trait, the added properties are modifiers, and the extended name of the base trait plus modifier(s) is called a trait variant. We introduced quantifiers to further define the use of a modifier in a specific situation (Table 1).

1. Curation of complex traits with modifiers has been implemented at the experiment level (Figure 2).

2. A curation tool was developed to allow modifier and quantifier information to be easily appended to base traits (Figure 3) to form (new) trait variants.

3. Conversion of "sibling traits" to "trait variants" using the new scheme and procedures generated positive impacts on the Animal QTLdb and CorrDB (Table 2).

4. Significant reduction of complex traits to manage after transitioning from "sibling traits" to "trait variants" demonstrates improved efficiency using this method (Table 3).

While the modifier factor partitioning approach provides possibilities for a more scalable system, it also opens additional opportunities for complex trait curation and management in the future.

Table 1.
Table 1. "Modifiers" and "quantifiers" used in the implementation of a new trait variant management system where trait variants are curated at the experiment level. This scheme helped to relieve curation and data management burdens caused by long and unwieldy lists of "sibling traits".

	Modifiers	Quantifiers
1	Analysis	adjusted, calculated, estimated
2	Anatomy location	above, anterior, at, below, by, dorsal, in, of, on, posterior
3	Environment	challenge, confinement, stress
4	Herd	calves, cows, ewes, heifers, layers, sows
5	Instrument	manufacturer, name, type
6	Measurement	amount, area, character, color, composition, count, length, maximum, response, speed, weight
7	Parity	count
8	Pedigree	dam, daughter, maternal, paternal, sire
9	Stage	adult, end, feeder, finisher, gestation, lactation, nursing, parturition, start, weaning, yearling
10	Time	after, age, at, basis, before, by, duration, weight
11	Treatment	challenge, drug, fast, feed, freeze, thaw, trim

Acknowledgements

Figure 2. A conceptual graph showing data structure differences between "sibling traits" (modified as part of ontology extensions) (a), and "trait variants" (modified with extended info) created at the experiment level (b), their use in QTL data annotations, and their effects on ontology trait data management.

(a) Sibling traits created with modifiers for QTL annotation

(b) Trait variants created with modifiers per experiment

Discussion

Gkoutos et al. (4) demonstrated the use of a decomposition strategy using the Phenotype and Trait Ontology (PATO) to dissect the terms in the Human Phenotype Ontology (HPO) into their entity/quality properties. While this was effective for human medical data, this strategy is not sufficient for accurate dissemination of trait information in livestock animals. Our approach using modifiers/quantifiers demonstrates the possibility of partitioning complex traits using additional trait descriptor information and provides a better structure for curation management of trait details.

Table 2. Number of experiments and annotated data affected in the QTLdb and CorrDB due to trait management changes from "sibling traits" to "trait variants" in 2022.

Databases	Affected data types	Cattle	Chicken	Goat	Horse	Pig	Rainbow trout	Sheep	Total
QTL/association	Total base traits (BT)	678	370	25	65	692	28	265	2,123
	BT with variants	28	10	2	1	33	6	13	93
	New trait variants	112	114	4	1	110	6	71	418
	Experiments affected	123	342	2	1	93	1	39	625
	Annotated data affected	10,010	648	10	16	4,906	174	463	16,227
Correlation	Total base traits (BT)	373	106	33	36	252		76	876
	BT with variants	42	13		1	17		18	91
	New trait variants	181	52		1	101		90	425
	Experiments affected	40	21		1	18		22	102
	Annotated data affected	1,392	135		10	3,143		893	5,573
Heritability	Total base traits (BT)	395	112	2	53	285		96	943
	BT with variants	43	13	1	1	18		18	94
	New trait variants	170	52	3	1	97		90	413
	Experiments affected	45	19	1	1	19		22	107
	Annotated data affected	163	9	3	1	203		36	415

Figure 1. A graph showing the increase in curated data in the Animal QTLdb and CorrDB over the past 18 years. Data numbers are log transformed to improve visualization.
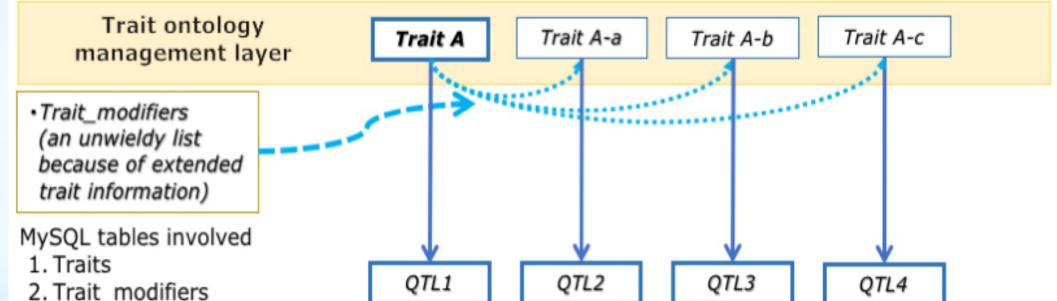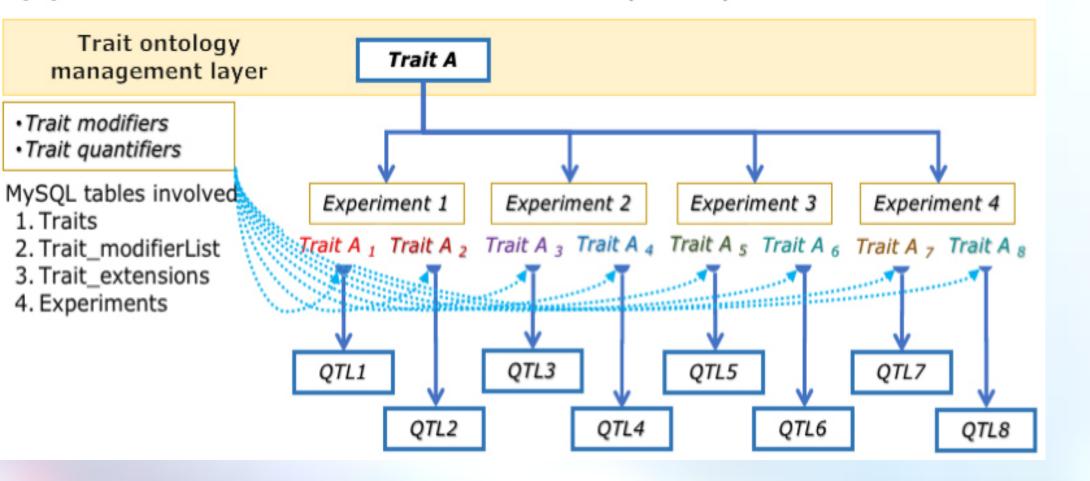

Figure 3. A screenshot of a curation web form showing part of the experiment curation environment. It shows how this implementation allows trait variants to be created from their base traits using controlled vocabulary lists to define modifiers/quantifiers.

Table 3. Total number of trait changes due to the database transition from using "sibling traits" to "trait variants" in 2022.

	Sibling traits	Trait variants	Change
QTL/association	2,272	418	-81.6%
Correlation	902	425	-52.9%
Heritability	1,061	413	-79.9%
Average			-71.5%

References

1. Hu, Z.L., Park, C.A., Reecy, J.M. (2022) Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res, 50, D956-D961

2. Hu, Z.L., Park, C.A., Reecy, J.M. (2016) Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res, 44, D827-833.

3. Fabian, G., Wächter, T., Schroeder, M. (2012) Extending ontologies by finding siblings using set expansion techniques. Bioinformatics, 28, i292-300.

4. Gkoutos, G.V., Mungall, C., Dolken, S., et al. (2009) Entity/quality-based logical definitions for the human skeletal phenome using PATO. Annu Int Conf IEEE Eng Med Biol Soc, 2009, 7069-7072.