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 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 Trail 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 only a small number of “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 modifiers 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.

Discussion
Gicoutos 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/property relationships. 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 1. Number of trait changes due to the database transition from using “sibling traits” to “trait variants” in 2022.

Table 2. Total number of trait changes due to the database transition from using “sibling traits” to “trait variants” in 2022.

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

Table 4. Total number of trait changes due to the database transition from using “sibling traits” to “trait variants” in 2022.

References