Extension of Animal QTLdb (IV): QTL Meta-analysis On-the-fly

Zhi-Liang Hu [1], Xiao-Lin Wu [2], Carissa A. Park [1], and James M. Reecy [1]

1. Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames, IA 50011
2. Department of Dairy Science, University of Wisconsin, Madison, WI 53706

Abstract

The Animal QTLdb was created to warehouse, compare and facilitate meta-analyses of QTL results from multiple, independent experiments. To date, more than 12,600 cattle, chicken, pig and sheep QTL have been curated into the database. A number of meta-analysis studies have been carried out in several livestock species, but the results are not readily comparable due to the use of different models and scenarios for statistical inference in these studies. Each analysis is not only time consuming, but also not trivial for a biologist to rerun when new data are added. On the other hand, the continuing rapid increase of QTL data demands a better and faster way to combine QTL data in order to speed up the discovery of causal variants that underlie each QTL. We have implemented two simple, non-parametric meta-analysis methods to perform real-time QTL meta-analysis on the fly. One method is a simple plot of QTL counts at each centiMorgan (cM) along a chromosome; the other method is a kernel density distribution plot of such counts. In both methods, the presence of QTL is computed to indicate the most likely QTL locations supported by available data. This function is available on the QTLdb when users visualize a QTL relative to a chromosome of cattle, chicken, pig or sheep. We are actively working to implement more practical and robust meta-analysis methods to facilitate better focused QTL data mining.

Introduction

The Animal QTLdb was developed as an online tool to house all publicly available QTL data from cattle, chicken, pigs, and sheep for cross-experiment comparisons (Hu et al. 2006, 2007a, b). To date, there have been 4,281 QTL for cattle, 2,284 QTL for chicken, 5,986 QTL for pigs, and 84 QTL for sheep have been curated into the QTLdb. As the amount of curated QTL data constantly grows, it is becoming a challenge for researchers to comprehend the large number of QTL records in a meaningful way and make the most efficient use of these data without investing a significant amount of effort.

Since its creation in 2004, the Animal QTLdb has been receiving more than 500,000 hits per year. To date, there have been over 100 citations of Animal QTLdb in journal papers and other publications. However, With the previous versions, users have been mainly relying on their own ways to make a cross-study comparison of QTL results. This was typically done by downloading the QTL data and then re-analyzing them by their own tools. Such a procedure was very time consuming and extremely laborious.

One way to make sense of the large amount of QTL data is to automate the process for users to integrate multiple results from individual studies. There have been a number of computational and statistical approaches used in terms of meta-analysis (Goffinet and Gerber, 2000; Etzel and Guerra, 2003; Khatkar et al., 2004). While studying the congruency between these results is often a complex task to tackle, a tool to assist users for the analysis is useful. We made a couple meta-plots tools, and coupled it with the QTLdb, as a start to provide semi-automated assistance for users to perform their QTL meta-analysis.

Materials and Methods

All QTL data used in this study were extracted from publications that had been curated into the Animal QTLdb (Hu et al., 2007a). The parameters, criteria and methods used for data curation were described previously (Hu et al., 2005).

The web user interfaces for data query, retrieval and display are programmed in Perl/CGI. The "Pile Plot" is based on the number of QTL counted at every centiMorgan (cM) along a chromosome. The plot is produced "on the fly" by a Perl script, employing Lincoln Stein's Perl GD library to generate instant web graphs. The kernel density estimation is carried out "during the fly" with an R program, and the output is ported to a Perl script to produce the web graphs.

All data and tools are hosted on a RedHat Linux server located at the Iowa State University, freely accessible at this URL:

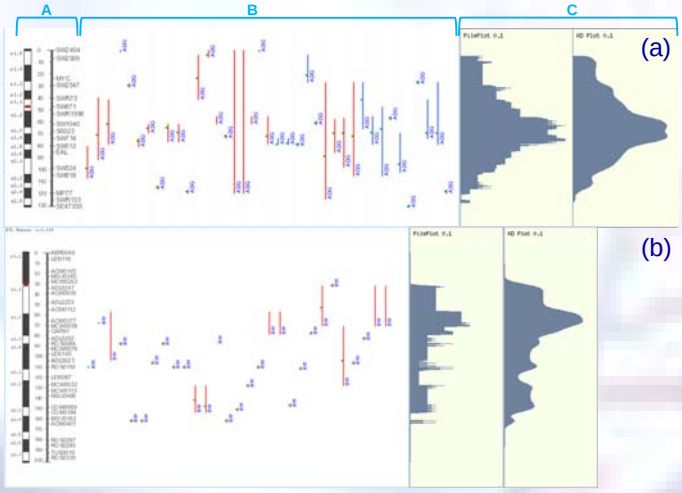
http://www.AnimalGenome.ORG/QTLdb/

Figure 1.

A snapshot of the chromosomal view of SSC 7 within the Animal QTLdb showing QTL for Average Daily Gain (ADG). Note that when only the QTL for a single trait are displayed, a "Show MetaPlot" link appears (red arrow). When this link is clicked, the meta-plots will be displayed to the right of the QTL graphs (see Figure 2).

Figure 2.

Two examples of QTL meta-plots, (a) and (b). The first displays QTL for Average Daily Gain (ADG) on pig chromosome 4 (a), in which one can see how the peak of meta-QTL is supported by the number of reported QTL that overlap; while in (b) is shown the QTL meta-plot for Body Weight (BW) on chicken chromosome 5, where the pattern of supporting evidence by the number of reported QTL is not very obvious. In both cases, the power of the QTL meta-plot is demonstrated. (**Legend**: **A**: G-banded and linkage maps of chromosomes; **B**: Reported QTL plots; **C**: QTL meta-plots).



Results

The QTL meta-analysis results are visually presented in the form of what we call meta-plots. Currently, we have made two types of meta-plots available: a pile plot and a kernel density plot. To use the tool, one needs to browse to the species/chromosome of interest and search for a particular trait for analysis, for example, Average Daily Gain (Figure 1). Only when a single trait is shown in the resulting page, do the live link to meta-plots become available (Figure 2).

◆ **Pile Plot**: The pile plot is based on the actual counts (y value) of the presence of reported QTLs at a one-centiMorgan (cM; x value) bin along the target chromosome. The count is transformed with Log10 for better scalability (Figure 2, C).

◆ **Kernel Density Plot**: The kernel density plot is also based on the actual QTL counts. However, the kernel density is estimated for the best probability of the curve height, assuming QTL data is a random variable. The outcome of the estimates gives the effect of smoothing out the plot.

The utility of the meta-plot is obvious in visually locating the most probable chromosome region(s) for the QTL under examination, as shown in Figure 2(a). It is worth noting, however, that in situations where the original QTL distribution is not very obvious, as shown in Figure 2(b), the role of the meta-plot is more evident.

Discussions

Meta-analysis is the application of statistical procedures to collections of findings from individual studies for the purpose of integrating, synthesizing, and making sense of them. Notably, several statistical meta-analysis methods have been proposed. In this study, we only adopted two of the simplest approaches without making things more complex at this time. While we have demonstrated the power of QTL meta-plotting tools to assist users, we wish to discuss some shortcomings of the tools, and to highlight some of our development goals.

Firstly, the current meta-plot tools all reported QTLs equally, therefore assuming each reported QTL has the same statistical power. In reality this may not be true because each reported QTL has passed different statistical analysis threshold tests. For example, each reported QTL experiment may have followed different designs, used different family structures and number of animals, QTL may be detected using different statistics, such as p-value, LOD score, F-test estimates, etc. While it is difficult to take each and every one of these parameters into consideration, our meta-plots provided a simple, fast and reliable method, for users to instantly meta-plotting QTL of interests, in order to obtain a general view of chromosome-wise QTL distribution, without dealing with these hard-to-overcome problems.

Secondly, the current meta-plots do not distinguish "suggestive QTL" or "significant QTL" which have already been recorded in the Animal QTLdb. We are in a process to define appropriate weight factors for optimal description of the QTL data by meta-plots. Hopefully an updated version of the meta-plots will accommodate this feature soon.

Thirdly, the current version does not allow users to select a subset of QTL data, meaning that all QTL included for meta-plots are those data that have been stored in the QTLdb. Users do not yet have the ability to decide which QTL to include or exclude based on their judgment of the nature and quality of the original QTL data. This feature would be useful for a more targeted analysis.

In order for the QTL meta-analysis to be more subjective and robust, challenges must be addressed to improve the QTLdb in terms of data collection, curation, and categorization. For example, an ideal meta-QTL data set should contain all details of each and every QTL experiment (e.g., the number of animals, sexes, generations, groups, breeds, origins) and all test statistics (e.g., significance tests, simulation tests, confidence intervals, etc). While it might be impossible to retrospectively obtain every piece of such data from experiments done multiple years ago by some groups, it at least provides some basis for improvement, for better QTL meta-analysis.

References

Etzel, C. J., and Guerra, R. (2002) Meta-analysis of genetic-linkage analysis of quantitative-trait loci. Am. J. Hum. Genet. 71, 56-65.

Goffinet, B. and Gerber, S. (2000). Quantitative trait loci: A meta-analysis. Genetics 155, 463-473.

Hu, Zhi-Liang and James M. Reecy (2007b). Animal QTLdb: beyond a repository - a public platform for QTL comparisons and integration with diverse types of structural genomic information. Mammalian Genome, 18: 1-4.

Hu, Zhi-Liang, Eric Ryan Fritz and James M. Reecy (2007a). AnimalQTLdb: a livestock QTL database tTool set for positional QTL iInformation mining and beyond. Nucleic Acids Research, 35(Database issue):D604-D609.

Hu, Zhi-Liang, Svetlana Dracheva, Wonhee Jang, Donna Maglott, John Bastiaansen, Max F. Rothschild and James M. Reecy (2005). A QTL resource and comparison tool for pigs: PigQTLDB. Mammalian Genome. 16(10):792-800.

Khatkar, M. S. Thomson, P. C., Tammen, I., Raadsma, H. W. (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet. Sel. Evol. 36, 163-190.

Acknowledgements

This research is supported by the USDA NAGRP Bioinformatics Coordination Project and the USDA-NRI Grant number 2007-04187.