# A Frame-Work for Developing the Pig QTL Database

Zhiliang Hu   and   Max Rothschild

Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University

## Abstract

One application of genetic linkage maps in livestock species is mapping loci underlying the genetic differences of economically important traits. As a result of active mapping of quantitative trait loci (QTL) in pigs during the past decade, hundreds of QTL in pig genome have been reported for growth, meat quality, reproduction, disease resistance and other traits. We are developing a pig QTL database to allow easy search and comparison of publicly available QTL data. A few approaches have been taken to accommodate the complex need for QTL information storage, organization and presentation. We have introduced a "trait ontology" concept to standardize the way animal traits are named and to simplify the way that the traits may be organized, in order for comparisons of QTL data to be possible. The database schema is based on a relational database, making use of existing pig map databases and other publicly available databases resources. The pig QTL database is also designed to include data representing major genes and markers having large effect on economically important traits. Efforts are undertaken to make it part of the integrated functional genomics resources for pigs.

## Introduction

Genetic linkage maps for pigs (Archibald et al., 1995; Rohrer et al., 1996; Marklund et al., 1996) have provided a useful tool for identification of genes that contribute to continuous variation. In the past a few years significant progress has been made to map economically important Quantitative Trait Loci (QTL) (Rothschild, M. F. and G. S. Plastow, 1999; Bidanel and Rothschild, 2002). Comparative mapping and candidate gene research have also facilitated use of abundant information from human and mouse genome studies and identification of chromosomal segments containing genetic information controlling variation in traits such as meat quality, fertility, and disease resistance in pigs. To date, there have been over 60 publications on studies for QTL and major genes in pigs and as a result hundreds of QTL and candidate genes have been identified (for review, see Bidanel and Rothschild, 2002).

The ultimate goal of QTL studies is the identification of the gene(s) responsible for the phenotypic variation observed in a particular trait. However, it is a challenging task to combine results from different QTL studies for positional candidate gene searches with information scattered in many publications, each described in different ways. Our purpose is to build a pig QTL database, for easy search and comparison of QTL results from different studies, derived from different populations, and obtained with perhaps different testing methods. With the available comparative mapping information, it is possible to consolidate the existing pig QTL results to one QTL map. With overlapping QTL results from different studies, it will be easier to confirm and narrow down QTL regions that may help to speed up the positional search for underlying genes.

## Material and Methods

1. Platform: Digital UNIX 4.0D running on a Compaq 522u work station. Underlining relational database: Ingres II.

2. Approach: We plan to program the web interface in generic DBI/DBD with Perl/CGI. The graphic tool to present QTL will also be embedded in CGI. Image-map technique is to be used for embedding hyper-links from attributes on the graph to various database locations. The purpose for this approach is for easy transfer or adapt from the current database schema and tools to other available ones when needed.

## Results

Several approaches have been taken to accommodate the complex need for QTL information storage, organization and presentation.

1. "Trait ontology" – a standardized way to describe traits

In the farm animal industry, it is not uncommon to encounter different names to describe the same production trait. In animal science, physiologists, geneticists and nutritionists often have their own terms within their specialty for the same trait. We have therefore introduced a "trait ontology" concept to standardize the way animal traits are named and to simplify the way that the traits may be organized, in order for comparisons of QTL data to be possible, and the trait data to be manageable with a relational database.

Table 1 gives a snap shot of a table organizing 169 pig production traits/parameters into a hierarchy containing five (5) trait classes and twenty-six (26) trait types. The classification of traits in this way also helps to categorize the QTL identified and provide a convenient way for researchers to analyze the data.

2. Considerations in database structure design

QTL analysis methods have been an active research area while QTL studies were carried out in the past decade. As a result, the QTL data in the public domain are not only scattered but also do not follow a commonly agreed standard for QTL detection and description. This requires the database schema to be comprehensive in order to be inclusive for information to be stored.

In Figure 1 is shown a relational database schema in which QTL related information are stored and related to each other. All QTL related data are organized into five (5) tables: (1). Experimental design; (2) Related traits and their ontology hierarchy; (3) QTL description; (4) QTL sources (QTL effects): (5) Information source – publications.

3. Integration of QTL map with existing map databases

The pig QTL database is designed to include data representing both major genes and markers having large effect on economically important traits. It is our intention to make this "QTL database" merge functional genomic data with the mapping data, therefore serve the purpose of database integration. This will be largely achieved

## Results (cont.)

through making use of existing pig map databases and other publicly available databases resources. In Figure 2 is shown a conceptual database model in which the pig QTL database is linked with other existing public databases. It also shows that similar works may be done for other livestock species.

4. Platform, interface, and other considerations:

In order to maximize the utility of public databases, we designed the database to be interfaced with World Wide Web for both database curators/editors and the users. In this way we make the database a public tool to use. Also, we try to present the QTL/mapping data in an integrated way. In Figure 3 is shown how shall the QTL data may be presented along with linkage data and physical mapping data. On these mapping representations, hyper-links will be embedded in the graph so that dynamic links are built under each feature on the map. The database system is web based on an UNIX platform. Standard DBD/DBI with CGI developed in Perl will make it easily transferable or adapted into similar environment.

Table 1. A snap shot of a trait ontology table showing classification of 169 pig production traits in 5 trait classes and 26 trait types.

| Level 1 TraitClass | Level 2 TraitType | Level 3 Trait |
|---|---|---|
| 1 Tissue Quality | 1 fatness | 1 abdominal fat<br>2 back fat<br>3 back fat above muscle dorsi<br>..... |
| | 2 meat_color | 24 color a<br>25 color b<br>26 color L<br>..... |
| | 3 pH | 30 pH 24 hours post mortem<br>31 pH 45 minutes post mortem<br>..... |
| | ..... | |
| 2 Health | 11 pathogen | 94 Salmonella count in feces day 1 post infection<br>95 Salmonella count in feces day 7 post infection<br>..... |
| | 12 Immune capacity | 98 number of leucocytes<br>99 mitogen-induced proliferation<br>100 IgG levels pre/post-vaccination of K88 antigen<br>..... |
| | 13 Disease resistance | 105 resistance to E.coli K88 infection<br>..... |
| ..... | ..... | |
| 5 Reproduction | 22 Litter size | 155 total number born<br>156 total number born alive<br>157 fully formed pigs<br>..... |
| | ..... | |
| | 26 endocrine | 169 Plasma FSH concentration |

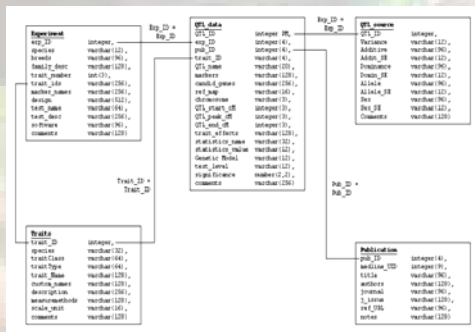Figure 1.A relational database schema showing how the QTL related information are managed



Figure 2. A conceptual database model showing how the pig QTL database shall be linked with other existing public databases and how similar work may be done for other livestock species.
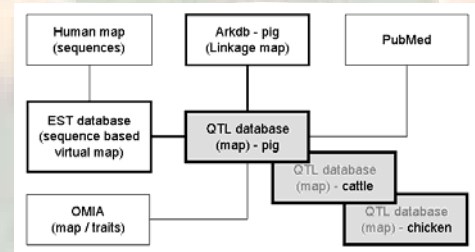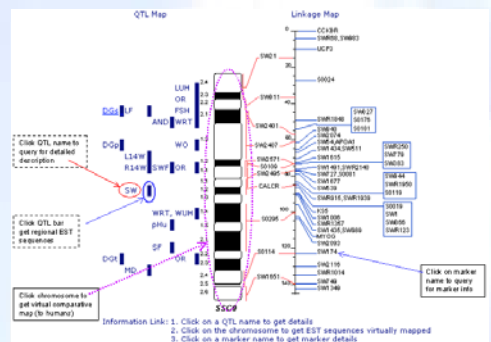


Figure 3. A schematic diagram showing how shall the QTL data may be presented along with linkage data and physical mapping data.



## Discussion

1. There are already some public databases that have QTL data components, such as the mouse QTL for animal models of obesity (http://www.obesity. chair.ulaval.ca/qtl.html) and rat QTL database (http://ratmap.org/qtler/). These QTL databases are built as a component to the specific database and difficult to be used as a stand-alone database tool for use.

It has been a phenomenon in genome research that similar developments are repeated at different locations. While we realize there has been lacking a "universal" tool for everybody, we also realize that not everyone makes their database design, database schema and codes easily accessible by the public. By developing the pig QTL database, it is also our goal to make every effort to minimize the redundant works and make it a public effort.

2. There exist some specialized QTL databases such as Bone Density QTL Database at the University of Tennessee Health Science Center (http://132.192.64.52/qtl/). However its design only deals with simple situation and its content is only a collected list of published data. Other database such as Rat QTL (http://rgd.mcw.edu/qtls/), Maize QTL database (http://www.agron. missouri.edu/body/qtl.html), Gramene genome database (http://www.gramene. org/), and the Barley genome database (http://obesitygene.pbrc.edu/), etc. However the QTL is only a component of the respective genome databases and hardly available for use in building our pig QTL database. There are also database components and functions that may be of our use and we are working with, such as Anubis in the Arkdb (http://www.thearkdb.org/), Rat QTL map hosted in Sweden (http://ratmap.org/qtler/).

3. In developing a pig QTL database, we are not developing a static QTL information storage place but also trying to provide a generic tool for continuing QTL data update, analysis, and comparison, in pigs. As for a long term goal, our continued efforts will be made to make it part of the integrated functional genomics resources for pigs in the future.

## Acknowledgements

## References

Archibald, A.L., C.S. Haley, J.F. Brown, S. Couperwhite, H.A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, A.-K. Winters, M. Fredholm, N.J. Larsen, V.H. Nielsen, D. Milan, N. Woloszyn, A. Robin, M. Dales, J. Riquet, J. Gellin, J.-C. Caritez, G. Burgaud, L. Ollivier, J.-P. Bidanel, M. Vaiman, C. Renard, H. Geldermann, R Davoli, D. Ruyter, E. J.M. Verstege, M.A.M. Groenen, W. Davies, B. Hoyheim. A. Keiserud, L. Anderson, H. Ellergren, M. Johansson, L. Marklund, J.R. Miller, D.V. Anderson Dear, E. Signer, A J. Jeffrey, C. Moran, P. Le Tissier, Muladno, M.F. Rothschild, C.K. Tuggle, D. Vase, J. Helm, H.-C Liu, A. Ragman, T.-P. Yu, R.G. Larson and C.B. Schmitz. 1995. The PiGMaP consortium linkage map of the pig (Sus scrofa). Mammalian Genome 6:157-175.

Bidanel, J. P. and Max F. Rothschild, 2002. Current Status of Quantitative Trait Loci Mapping in Pigs. Pig News and Information, 2002, Vol. 23, No. 2

Marklund, L., M. J. Moller, B. Hoyheim, W. Davies, M. Fredholm, R. K. Juneja, P. Mariani, W. Coppeiters, H. Ellegren, and I. Andersson. 1996. A comprehensive linkage map of the pig based on a wild pig-Large White intercross. Animal Genetics 27:255-269.

Rohrer, G. A., L. J. Alexander, H. Zhiliang, T. P. L. Smith, J. W. Keele, and C. W. Beattie. 1996. A comprehensive map of the porcine genome. Genome Research 6:371-391.

Rothschild, M. F. and G. S Plastow, 1999. Current Advances in Pig Genomics and Industry Applications. Feburary 1999 AgBiotech Net.