Doctoral Thesis

Bioinformatics Analysis of Rat Muscle Microarray Gene Expression and Genetics of Pig Exterior Traits

Author(s):  
Ghaderi-Zefrehei, Mostafa

Publication Date:  
2008

Permanent Link:  
https://doi.org/10.3929/ethz-a-005673271

Rights / License:  
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Bioinformatics Analysis of Rat Muscle Microarray Gene Expression and Genetics of Pig Exterior Traits

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

For the degree of
DOCTOR OF SCIENCES

Presented by
Mostafa Ghaderi-Zefrehei

Master of Science – Isfahan University of Technology, Iran
Born on March 20th, 1976, Iran

Accepted on the recommendation of:

Prof. Dr. Wolfgang Langhans, examiner
Prof. Dr. Haja N. Kadarmideen, examiner
Prof. Dr. Martin Flück, co-examiner

Zurich 2008
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

# LIST OF TABLES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

# SUMMARY

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

# ZUSAMMENFASSUNG

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
</tr>
</tbody>
</table>

# ACKNOWLEDGMENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
</tr>
</tbody>
</table>

# CHAPTER 1

1.1 Gene Expression Microarray Data and Its Uses ...........................................13
   1.1.1 Exploring Physiology of Muscular Activity: The Rat Transcriptomic Experiments 14
   1.1.2 Expression Profiling and t-test Statistics .....................................................15
1.2 Genetical Genomics and Computational Systems Biology ..........................15
   1.2.1 Application of Genetical Genomics and Systems Biology ..........................16
   1.2.2 Deciphering Gene Module Pattern in Microarray Data ...............................17
   1.2.3 DNA Sequence Motif .....................................................................................17
1.3 Quantitative Genetics and Animal Breeding and Its Uses ..........................18
   1.3.1 Application in Genetic Improvement of Pig Reproduction and Type Traits 19
   1.3.2 Agglomerative Hierarchical Clustering and PCA ..........................................20
   1.3.3 Bayesian Complex Major Gene Segregation Analysis ...................................21
1.4 Software and Programming ...........................................................................22
1.5 Objectives of This Thesis ...............................................................................22

# CHAPTER 2

2.1 Abstract ...........................................................................................................24
2.2 Introduction ......................................................................................................25
2.3 Material and Methods .....................................................................................26
   2.3.1 Microarray Experiment .................................................................................26
   2.3.2 Finding Differentially Expressed Genes ........................................................27
   2.3.3 Ranking Genes ................................................................................................30
2.4. Results and Discussion ..................................................................................31
   2.4.1 Differentially Expressed Genes .....................................................................31
   2.4.1 Differentially Expressed Genes Comparison among Experiments ...............38
2.5. Checking the Model ..........................................................................................41
2.6. Implications .....................................................................................................48
2.7. Acknowledgments ............................................................................................48
2.8. Appendix ..........................................................................................................48

# CHAPTER 3

3.1. Abstract ...........................................................................................................52
3.2. Introduction ......................................................................................................52
3.3. Methods ...........................................................................................................54
   3.3.1. Microarray Experiment ...............................................................................54
   3.3.2. Expression QTLs Mapping .........................................................................54
   3.3.3. Gene Co-Expression Network Construction and Connectivity ..................54
3.4. Results and Discussion ..................................................................................57
   3.4.1. Expression QTL Mapping Task ....................................................................57
### Chapter 4

**Abstract**

**Introduction**

**Materials and Methods**

1. Breeding Program and Station-Tested Traits
2. Data
3. Statistical Genetic Analysis

**Results and Discussion**

1. Teat Traits
   1. Count right (ZiRe)
   2. Count left (ZiLi)
2. Reproductive Traits
3. Number of piglets born alive (LGF)
4. Number of weaned piglets (AGF)
5. Number of piglets during the lactation (GSF)
6. Piglets survival (FAR)
7. Genetic Correlations
   1. Swiss-Large-White Breed
   2. Swiss-Landrace
   3. Swiss-Large-White-Sire-Line

**Implications**

**Acknowledgments**

**Appendix A**

**Appendix B**

### Chapter 5

**Bayesian Major Gene Segregation Analysis of Carcass Length Trait Across Three Pigs Breeds**

**Abstract**

**Introduction**

**Materials and Methods**

1. Data
2. Box-Cox Transformation
3. Notation and Assumptions
4. Complex Segregation Analysis
5. Gibbs Sampling Algorithm

**Results and Discussion**

**Implications**

**Acknowledgments**

### Chapter 6

**General Discussion**
List of Figures

Figure 2. 1 Before and after normalization............................................................. 28
Figure 2. 2 Plotting of expression values against its theoretical quintiles............. 28
Figure 2. 3 Insightful plots to visualize state of differentially expressed genes......... 33
Figure 2. 4 Pattern of differentially expressed genes of first experiment at three levels of pFDR ......................................................................................................................... 34
Figure 2. 5 Pattern of differentially expressed genes of second experiment at three levels of pFDR ......................................................................................................................... 36
Figure 2. 6 Pattern of differentially expressed genes of third experiment at three levels of pFDR ......................................................................................................................... 37
Figure 2. 7 General plots regarding to model checking............................................ 41
Figure 3. 1 Postulated a seven-gene network with its adjacency matrix .................. 55
Figure 3. 2 Whole genome genetic mapping and bootstrap tests of Glutathione S-transferase theta 1 over chromosome 10................................................................. 58
Figure 3. 3 Physical map of chromosome 10 between 69 to 73M base pairs......... 59
Figure 3. 4 Gene ontology enrichment of glutathione S-transferase theta 1.............. 59
Figure 3. 5 Cluster of eQTLs mapping of correlated gene expression values and epistatics results for Glutathione S-transferase theta 1................................................................. 61
Figure 3. 6 Plot the scale free topology and mean gene connectivity versus .......... 64
Figure 3. 7 Scatter plots of arrays along the module eigengenes, classical threedimensional scaling plots for and enrichment of differentially expressed genes. .... 65
Figure 3. 8 Connectivity plot variance of gene expression........................................ 67
Figure 3. 9 Sequence logo of discovered motif (consensus sequence).................... 70
Figure 3. 10 Whole genome genetic mapping and bootstrap tests of Cytochrom c oxidase subunit VIII-H gene (up) over chromosome 14............................... 72
Figure 3. 11 Whole genome genetic mapping and bootstrap of solute carrier family 4, member 3 gene over chromosome 20......................................................... 73
Figure 3. 12 Whole genome genetic mapping and bootstrap tests of carbonic anhydrase 3 gene (up) over chromosome 15................................................................. 73
Figure 3. 13 Whole genome genetic mapping and bootstrap tests of ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 1 gene over chromosome 1................................................................. 74
Figure 3. 14 Whole genome genetic mapping and bootstrap tests of fatty acid binding protein 3 gene for fatty acid binding protein 3 gene over chromosome 8............ 74
Figure 4. 1 Estimated breeding values for teat traits.............................................. 108
Figure 4. 2 Estimated breeding values for leg traits............................................... 108
Figure 4. 3 Estimated breeding values for reproductive traits............................... 109
Figure 4. 4 Complete method of clustering of traits with two dissimilarity methods. ... 111
Figure 5.1 Theoretical normal distribution and empirical normal distribution of carcass length distribution in SLW breed................................................................. 116
Figure 5.2 Theoretical normal distribution and empirical normal distribution of carcass length distribution in SLR breed........................................................................ 116
Figure 5.3 Theoretical normal distribution and empirical normal distribution of carcass length distribution in SLWSL breed................................................................. 117
Figure 5.4 Marginal non-parametric posterior distributions of error variance and polygenic variance for SLW breed. ........................................................................ 123
Figure 5.5 Marginal non-parametric posterior distributions of major gene allele frequencies and polygenic heritability for SLW breed ........................................ 123
Figure 5.6 Marginal non-parametric posterior distributions of major additive and dominance effect for SLW breed........................................................................ 123
Figure 5.7 Marginal non-parametric posterior distributions of error variance and polygenic variance for SLR breed. ........................................................................ 125
Figure 5.8 Marginal non-parametric posterior distributions of major gene allele frequencies and polygenic heritability for SLR breed ........................................ 125
Figure 5.9 Marginal non-parametric posterior distributions of major additive and dominance effect for SLR breed........................................................................ 125
Figure 5.10 Marginal non-parametric posterior distributions of error variance and polygenic variance for SLWSL breed................................................................. 127
Figure 5.11 Marginal non-parametric posterior distributions of major gene allele frequencies and polygenic heritability for SLWSL breed........................................ 127
Figure 5.12 Marginal non-parametric posterior distributions of major additive and dominance effect for SLWSL breed................................................................. 127

List of Tables

Table 2.1 In common genes which have differentially expressed in pairwise comparison of all three experiment at 5%pFDR. ................................................................. 44
Table 2.2 In common genes which have differentially expressed in pairwise comparison of all three experiment at 10%pFDR. ................................................................. 45
Table 2.3 In common genes which have differentially expressed in pairwise comparison of all three experiment at 20%pFDR. ................................................................. 47
Table 3.1 Gene ontology results, the candidate gene is bold highlighted........................ 60
Table 3.2 The following loci in the HXBBXH data set have associations with the above Glutathione S-transferase theta 1 data. ................................................................. 60
Table 3.3 Summary of information due to extracting DNA promoter sequence of some hub-acting genes. ............................................................................................. 69
Table 3.4 Estimated PSSM due to hub-acting genes. ........................................................ 70
Table 4.1 Estimated direct additive heritability for different exterior traits in literatures. .................................................................................................................. 77
Table 4.2 Descriptive statistics of traits used in this study............................................. 79
Table 4.3 Estimation of teat traits genetic components across breeds......................... 85
Table 4.4 Estimation of leg traits genetic components for Swiss-Large-White breed. ... 88
Table 4. 5 Estimation of leg traits genetic components for Swiss-Landrace breed........ 89
Table 4. 6 Estimation of leg traits genetics components for Swiss-Large-White-Sire-Line breed........................................................................................................................................ 89
Table 4. 7 Estimation of reproductive traits genetics components for Swiss-Large-White breed........................................................................................................................................ 92
Table 4. 8 Estimation of Reproductive traits genetics components for Swiss-Landrace breed........................................................................................................................................ 93
Table 4. 9 Estimation of reproductive traits genetics components for Swiss-Large-White-Sire-Line breed........................................................................................................................................ 93
Table 4. 10 Estimates of heritabilities (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among traits in Swiss-Large-Whit breed................................................................. 96
Table 4. 11 Estimates of heritabilities (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among traits in Swiss-Landrace breed........................................................................................................................................ 99
Table 4. 12 Estimates of heritabilities (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among traits in Swiss-Large-White-Sire-Line breed........................................................................................................................................ 103
Table 4. 13 Model comparison based on log likelihood ratio test within traits across breeds. ......................................................................................................................................................................................... 105
Table 4. 14 Estimation of principal components based on variance covariance matrix among traits.................................................................................................................................................................................. 111
Table 4. 15 Estimation of principal components based on correlation matrix among traits. ........................................................................................................................................................................................................................ 112
Table 5. 1 General information about the data and records and number of fixed effects which used in this study.................................................................................................................................................................................. 116
Table 5. 2 Estimated non-Parametric posterior distributions of means, standard deviation, right and left bound of 95% highest posterior density (HPD) for carcass length in SWL breed. ................................................................................................................................................................................................ 122
Table 5. 3 Estimated non-parametric posterior distributions of means, standard deviation, right and left bonds of 95% highest posterior density (HPD) for carcass length in SLR breed. ................................................................................................................................................................................................ 124
Table 5. 4 Estimated non-parametric posterior distributions of means, standard deviation, right and left bonds of 93% highest posterior density (HPD) for carcass length in SLWSL breed. ................................................................................................................................................................................................ 126
Table 6. 1 Summary of genetical genomic analysis of candidate genes......................... 133
Summary

This doctoral thesis is based on two related disciplines, namely bioinformatics/computational biology and quantitative genetics. For bioinformatics/computational biology research, gene expression profiles from rat hind leg tissues, collected in experiments related to muscle systems physiology/biology were used. For quantitative genetics research, large volumes of phenotype and pedigree data collected from Swiss national pig breeding company, called Suisag, on teat, reproductive and type traits were used. The overall formulation of research problem with gene expression data in rats was firstly to identify genes that are significantly differentially expressed and up- or down-regulated between different treatments. Secondly, the aim was to identify gene co-expression networks typical of a particular condition/treatment of muscle tissue. The experiment involved subjecting rat muscle to different muscle loading and reloading (active/inactive) experiments. The main messages from chapters 2 and 3, relate to differentially expressed genes and mode of regulation that involved genes in ATP production and fatty acid cycle metabolism. In chapter 3, a novel and innovative research in the area of genetical genomics/systems biology for rat data was conducted. Genetical genomic analysis of some selected candidate genes (obtained from analyses in chapter 2) showed that these genes are more often trans-regulated than cis- regulated. However, a number of hub-acting genes in the identified gene co-expression network shared at least one DNA sequence motif. Moreover, genomic regions on chromosome 3 were identified to harbor many candidate genes. Although ad-hoc high throughput transcriptomic datasets were used, the computational methods and theoretical framework applies to general microarray experimental designs. Most statistical and computational methods were from BioConductor packages and are based on R programming.

On the quantitative genetics side, the research problem that was addressed was estimation of genetic parameters (heritability and genetic correlations) for a range of reproductive and teat traits in Swiss pig populations as well as the statistical methods to cluster highly correlated traits. Results showed that estimated total heritability for exterior, teat, and
reproductive traits were low. It is suggested that the use of maternal effect is not efficient to improve genetic performance of teat traits. Principal component analysis (PCA) and clustering algorithms revealed a hidden data structure in pig data. Irrespective of clustering algorithms and distance functions used, teat traits always were clustered together. The use of mixture models to estimate both polygenic and single major gene components (and its mode of inheritance) for carcass length traits in three pig breeds is a research problem that has been addressed for the first time by this thesis. The data was received from these breeds: Swiss-Large-White (SLW); Swiss-Landrace (SLR) and Swiss-Large-White-Sire-Line (SLWSL). Bayesian complex major gene segregation analysis indicated the possibility of segregating an autosomal major gene for carcass length trait. Therefore, for improving genetics performance of carcass length, including major gene effects in genetic analysis model is highly suggested. Overall, this thesis also developed a strong theoretical background in genetic parameter estimation and segregation analysis and shed some insights into data handling and analyses of animal breeding data.
Zusammenfassung

Rechenverfahren wurden mit Hilfe der “Bionconductor“ Software durchgeführt und basieren auf R Programmierung.

Acknowledgments

At this step, I would like to thanks all those people and scholars who provided me the opportunity to proceed along my PhD study and made this thesis possible. It was an enjoyable experience for me. First of all, I wish to express my sincerest gratitude to my father and mother and my great nation (Iran) and government for their continuous support throughout my scholarship. They helped me out financially and spiritually. I extend my gratitude: to Professor Haja. N. Kadarmideen, for his support, guidance and motivation during my PhD degree at ETH Zurich. I am especially grateful his confidence in me to allowing me to work on the different subjects presented in this thesis. To Professor Wolfgang Langhans for his support and the provision of the environment to wrap up my PhD thesis. To Professor Martin Flück for providing access to microarray data, for agreeing to be one of my referees and for his constructive criticism during revision of my thesis. To Dr. Peter Von Rohr for the valuable discussions and input over the years, his insights and advice. I thank him for patiently answering my emails with questions (there were a lot of those), for his collaboration on parts of this work, for his insights, and his friendship. I learnt many things by his shadow, encapsulating many area of programming especially R, Perl, C++ programming, and the concept of compendium in bioinformatics. To Dr.Luc Janss, an amazing brilliant analytical mind with sharp knowledge. He remains as one of the greatest scientist I have ever worked with. His helpful discussions and criticisms first time provided me how to go for spike-based computation. I’m also deeply grateful to Dr. Massoud Malek; he provided encouragement in gene expression and gene networking. I would also like to thank to Dr. Houcine Ilahi, an elegant researcher and scholar and helping me major gene analyses. Furthermore, I would like to thank Dr. Yougjun Li for his enthusiasm to teach me classical genetic evaluation and also his help on ASREML codes and showing me the first time the FORTRAN programming. I would like to issue my best to Abdol-Ghaffar, Masumeh and Azzam Ghaderi-Zefrehei, Haj Ghodrat Karimi Neisiani, Bagher Ahmadi. I would like to express my best to Ulrike Janssen-Tapken, someone with enormous talent for her patient during course of German learning and many other things which she taught me. I have gained by many Professors which amongst them were: Prof. Gaston Gonnet, Jörg Stelling, Van Velck, Steve Horvath.
I would like to express my best to Ruth Hauser for her great help, her help was enormous and unforgettable to me, Margrit Spiri, Senn Markus, Hossein Moradi, Sima Fakheran, Gholam-Reza Sharokhshahi, Nick Corbet, Dominik Doyscher, Tobias Zwyer, Dirk Müller, Hans Ulrich-Gressely, Hossein-Ali Ghofrani. Thanks to Michael Koss for his friendship and for his time by providing me great ideas in wet-lab genotyping and his explanation on molecular genetics, to Abdelhak Mansouri a good friend and his helpful idea and to Myrtha Arnold, by showing me some exciting wet-lab experiments. To Prof. Geary Nori, an elegant scientist and Dr. Katayoun Afshar, Dr. Karim Abbaspour for his Mollana session, Abasse Es-Haghi-Kohpayeh, Moard Alaeifar, Haj Abass Soltan-Mohammadi, Nahid, Neda, Elham, Niousha, Pouya and Parsa Pierali, for their helps and Jean-Philippe Pellet from IBM for his ideas on Bayesian network. I also thank DAGRL, ETH Zurich for funding my visit to CSIRO Rockhampton, Australia for 2 months. I thank the staff scientists and students of CSIRO Rockhampton for their help and friendship. Finally thanks to ETH, an institution truly devoted to education.

Dedicate to: My Brother:

Mohsen Ghaderi-Zefrehei
Chapter 1
General Introduction

The chapters in this thesis can broadly be divided based on species (Rat, Pig) or modeling paradigms (quantitative genetic evaluation, bioinformatics/system biology). All chapters that deal with pig species are quantitative genetic evaluation research and the rest of the chapters contain bioinformatics / systems biology analysis for Rat species. For every chapter, relevant, stand-alone introduction and literature review are provided. This was done to provide better cohesive streamline of the concepts and materials in each chapter. Here we give the general introduction about the chapters and research covered in this thesis.

1.1 Gene Expression Microarray Data and its Uses

Microarray experiment enables the simultaneous measure of the expression levels of genes sets. Such large global-scale gene expression profiling has revolutionized biological, medical, agricultural and even computational researches. In this context, many methods and techniques have been developed which among the commonest are: spotted cDNA, Affymetrix, Agilent, Operon, ClonTech, Amersham Home-Made Array. By these techniques we can now search for biomarker, functional SNP, hub-acting and disease-related genes in a systematic and unbiased manner (Ho, J. W. K. et al.; 2008, Woolfson, A. et al., 2006, Xiao, Y. Y. et al., 2007). In general, in a representative biological gene expression profiling research, we either run experiment over time, which is called time-course gene expression experiment, or experiment is run over different group’s i.e. non-diseased individuals versus disease-suffering individuals. More complex microarray experiment scenarios can be encountered when for example groups are viewed over time. Complex microarray experiment scenario provides complex data structure too. To gain insight into the complex biological questions, researchers must first overcome the challenge involved in analyzing the large and complex datasets that are generated from
complex microarray experiments. To help biologist, statistician and computer scientist focused to come up with scalable and effective knowledge discovery approaches from complex microarray experiments (Gan, X. et al., 2008, McIntosh, T. et al., 2007). Team up between computational researchers and biologist can provide timely proved solutions for biological motivated questions. This indicates how important it is the cohesive interaction between computational expertise and pure biologist in terms of developing new idea in the context of microarray gene expression paradigm. The author after years of work with microarray data analysis has this impression that the role of pure biologist is immensely becoming crucial when it comes to interpret, validate, or reject the computational tasks which are done for analysing microarray data. In this thesis we use the results of home-made array.

1.1.1 Exploring Physiology of Muscular Activity: The Rat Transcriptomic Experiments

Long-term muscle activity depends on energy provision. Adenosine triphosphate (ATP) is the high-energy molecular compound from which the body derives energy. ATP is produced through both anaerobic and aerobic energy metabolic systems. Intensive muscle endurance training, which leads to increase aerobic capacity of skeletal muscle, is physiologically important phenomenon. However, first it is not possible to apply many muscle enduring scenarios for human being and second, likely different muscle endurance training scenarios affect different part of genome or metabolic network topology, which without having global-scale of microarray gene expression, the weight and role of each molecular component cannot be well addressed. Vast majority of difficulty for muscle studies in human can be accustomed in Rat muscle studies, yet with maintaining the strength and implication in human. Having above arguments, Transcriptomic experiments of muscle studies using Rat muscle are physiologically highly relevant. In this context many experiments have been conducted (Cheng, G. et al., 2004, Dapp, C. et al., 2004, Jin, K. L. et al., 2001, Kostrominova, T. Y. et al., 2005, Wittwer, M. et al., 2002). In this study we use microarray gene expression experiments conducted in rats at the Institute of Anatomy, University of Bern, Switzerland, for
understanding the impact of different treatments (of activity and inactivity) on soleus muscle gene expression patterns, with an overall objective of getting insights into corresponding human muscles via comparative transcriptomics.

1.1.2 Expression Profiling and t-test Statistics

We already discussed the concept of microarray. Microarray data is huge and expensive data in nature; therefore, there is a need of using robust gene ranking measure to come up with biologically motivated results. It can happen that using non-robust gene ranking measure would leave researcher with false positive results. In general, one of the first and main step toward analyzing and deciphering of microarray gene expression data is to find set of differentially expressed genes. Statistically speaking, when thousands of genes in a microarray gene expression dataset are evaluated simultaneously by many non-robust gene ranking methods i.e. fold changes and some significance tests, the probability of detecting false positives rises sharply. It means that some of genes which we report to be differentially expressed, they are actually not differentially expressed. Many approaches and procedures have been proposed in the literature to test the null hypothesis. In this thesis, we use moderated t-statistics (Symth, G. K. 2004), as a robust gene ranking measure, for Rat’s muscle microarray gene expression data (Flück, M. et al., 2005).

1.2 Genetical Genomics and Computational Systems Biology

Computational system biology is a complex art in the computational science. This area is rapidly growing because it holds the promise for major discoveries in the Life Sciences. Computational system biology now aims to understand biological processes at the system level. System-level analysis can be grounded on discoveries at molecular-level. Having made great progress in genome sequence projects and range of other molecular genetics projects that accumulate in-depth knowledge of molecular nature of biological system, we are now at the stage to seriously explore the possibility of system-level understanding of biological system solidly grounded on molecular-level knowledge. Both structure of
the system and its molecular-level components plays indispensable role forming symbiotic state of the system as a whole. It is misleading to put attention only system structure, such as network topology, without paying sufficient attention to diversities and functionalities of components and possibility of using other branch of science like genetics. There are numbers of exciting and profound issues that are actively investigated in genetics, such as robustness of quantitative trait loci (QTL). Wisely merging many genetics idea like QTL in computational system biology, will lead to computational system genetics. Expression QTL mapping (eQTL) or genetical genomics can be seen as a successful try in this way. eQTL identifies the genomic locations to which the expression values are linked. If those estimated genomic locations have concordance mode with physical location of gene itself, it is called cis-acting eQTL and if it fallen outside of gene physical location, it is called Trans-acting eQTL. Many researchers have explored this new area of system-level analysis (Kadarmideen, H. N. et al., 2006, 2007). Therefore, by one measure—that of raw output—genetics ideas can be assumed as a new triumph to enrich computational system biology. More likely, in upcoming times computational system genetics will be prime paradigm in system-level understanding of biological questions.

1.2.1 Application of Genetical Genomics and Systems Biology

The main theme in biological sciences in this century is system-level analysis of biological questions. In preceding section, we made it clear that computational system genetics can merge many ideas from genetics into computational system biology. This should advance our fundamental understanding of biological system and lead to the identification of many unseen association in biological system. There are several challenges in which genetical genomics can soundly address them. For example localizing regulatory polymorphisms, when there is wide spectrum of cis-acting regulatory mechanisms, the non-concurrent and inconsistent effects of regulatory elements in different biological samples, and the difficulty in isolating the causal and regulator regions when there are many genes. Kadarmideen, H. N., (2008) covered many
1.2.2 Deciphering Gene Module Pattern in Microarray Data

Microarray data show module pattern or an over-represented pattern. This means, not all genes intend to have identical transcriptional pattern. Likely, genes which react identically to a given perturbing effect are modulated together. The concept of module-based learning network in context of gene expression has received many attentions (Prinz, S. et al., 2004; Segal, E. et al., 2003; Zhang, B. et al., 2005). Segal, E. et al., (2003) defined a new representation of Bayesian network which he called module network. Module network in the context of gene expression, explicitly partitions the genes into modules. Each module stands for a set of genes that have the same statistical behavior, i.e., they share the same set of parents and local probabilistic distribution. By enforcing this constraint on the Bayesian network, they significantly reduced the complexity of model parameter space as well as the number of parameters which generally can be encountered in the Bayesian network. These reductions can lead to more robust estimation of structure and better generalization on unseen structure in expression dataset. By making the modular structure explicit, the module network provides insights into the domain that are often hidden and obscured by the intricate details of a probabilistic learning network algorithms i.e., large Bayesian network structure. Segal, E. et al., (2003) idea is probabilistic-based way of module learning. However, a robust non-probabilistic module network learning method in the context of gene expression introduced by (Zhang, B. et al., 2005). This method has some extra facilities too i.e. the ability of entering extra source of information into learned network. In this thesis, this method of learning module network is used. We show how to use module network in this study to single out hub-acting genes.

1.2.3 DNA Sequence Motif

Exploring and extracting regulatory elements from related DNA sequences is a hot topic these days because it holds the key to identify general mechanisms of gene regulation. In general this analysis is achieved by clustering co-expressed genes of the same species to
find a short pattern of nucleotides upstream of the start sites of the encoded transcript which refers to a common transcription factor binding site (TFBS). This set of TFBS is called the ‘motif’. This time-consuming task offers to identify possible conserved nucleotide regions involved in coordinated gene regulation. Full coverage of many algorithms in this context can be found in (Maximilian Haußler, 2005), though, original statistical mechanical models, or thermodynamics-based motif search algorithms have not covered in it. In this thesis, to decipher regulation program of estimated hub-acting genes, we turn to find set of TFBSs (motif) across aforementioned genes.

1.3 Quantitative Genetics and Animal Breeding and its Uses

The central idea in population genetics theory is the examination of the change in the genetic make-up of a population as time goes on as a result of natural forces i.e., selection, mutation, and similar factors. Many ideas from pure mathematical population genetics were adopted in practical perspectives. This adoption ended up with new area of science, which is called quantitative genetics. Quantitative genetics, encapsulates all methods which currently we use in classical animal breeding, plant breeding, forest breeding, and etc. In other words, quantitative genetics is dealing with the analysis of diversity in complex traits that are caused by the joint effects of many alleles at several genetic loci, as well as non-genetic factors. There have been many pioneers in founding and developing quantitative genetics paradigm. William (Bill) Hill FRS is probably the world’s most eminent quantitative geneticist, with over 200 refereed publications in major journals (Charlesworth, B. et al., 2005). Animal breeding, which adopted quantitative genetic ideas into animal genetic setting, seeks to improve genetic performance of animal and meanwhile try to keep genetic diversity across individuals in the population. The seminal papers from Henderson, C.R. (1984) laid down the solid and robust statistical methodologies for improving genetic progress of interesting traits worldwide. Developing a general mathematical layout of relations between the genotype (the set of genes of an individual) and the phenotype (the set of observed characteristics) which is workable in a given environment with small bias, is the main things which likely
researcher in animal setting is looking for. The revolution in molecular genetic mapping technology and the advent of whole genome sequences have turned animal breeding setting into one of the fastest growing areas of quantitative genetics. Nowadays, many researchers worldwide are adopting DNA-based information into new model to end up with motivated results. New model development is critically dependent on the foundation of knowledge which researcher bears up and likelihood of data.

1.3.1 Application in Genetic Improvement of Pig Reproduction and Type Traits

In general, reproductive and type (exterior) traits in pig are categorized as low-heritable traits (Fukawa, K. et al., 2001; Pramod, K.M., 2002; Serenius, T. et al., 2004). Therefore, it seems reasonable to assume that the selection response for these traits in pig would probably be smaller than high-heritable traits i.e. carcass length. For measuring reproductive traits in practical situations, there are many causes for a reduced response. The commonest are: the course of natural selection working against artificial selection, small additive genetic variance, interaction of genotype by environment and genetic and environmental maternal effects (Holt, M. et al., 2005). In this context, some insights about magnificent role of aforementioned causes can be grasped using model organisms, but it is also more difficult to standardize the environmental causes in large pig production systems. In general, genetic selection for production traits or economically justifiable traits are often more important than reproductive and type traits, and may have unfavorable genetic correlations with reproductive and type traits. Many literatures have addressed this phenomenon (Sernius, T. et al., 2004). Even though, management strategies and environmental causes largely influence reproductive and type traits, they are also under the control of genes. The mechanism in which genes affect aforementioned traits can be deciphered by both quantitative and molecular genetic techniques. We assume quantitative genetic analysis can further provide an insight for committing molecular genetic analysis or not. From quantitative genetics point of view, if trait showed substantial genetic variation, i.e., high-heritable traits, molecular genetic study is likely recommended for that trait, otherwise it would depend on breeding executive board
decisions. Needless to say in general, there are some risks of using knowledge at the DNA level in breeding programs. In this thesis, we take the quantitative genetics approach to run large-scale genetic analysis of additive and maternal components of pig exterior, teat and reproductive traits across three pig breeds data. The data are provided by Suisag, located at Switzerland. In this thesis, we first time try to estimate many genetics components across three pig breeds. From the aforementioned it will be patently clear that knowledge of large-scale genetics components undoubtedly make it possible to build solid index, which will allow a more balanced genetic progress in teat, exteriors and reproductive traits while still maintaining progress in all other traits of interest too. Having clear picture of different estimated genetic components of matrix of genetic covariance among traits, would play key role in grounding unified genetic selection.

1.3.2 Agglomerative Hierarchical Clustering and PCA

In coming time geneticist will be observing an explosive growth in the availability of various high-dimensional phenotypic data. Reduction of high-dimensional data which comply with extracting useful knowledge from data can immensely help geneticist to have better implication about the measured data. Two general approaches which can be used in this sense are hierarchical clustering algorithms and principal component analysis (PCA). In general there are two types of hierarchical clustering: Divisive and Agglomerative methods. Divisive methods start with the assumption that all objects (in our case traits) are part of a single cluster. The algorithm splits this large cluster step by step until each object (trait) is a separate cluster. Agglomerative methods start inversely. Each cluster consists of one object (trait). The clusters are combined step by step. In each step those two clusters with the smallest dissimilarity or the highest similarity are merged. Iteration continues until all objects (traits) are in one single cluster. In other words, agglomerative methods build the tree bottom-up, successively grouping together clustering deemed most similar. Principal component analysis (PCA) is a technique that is useful for the compression and classification of data. The purpose of PCA is to reduce the dimensionality of the dataset by finding a new set of traits, smaller than the original set of traits that nonetheless retains most of the original dataset information. By information we mean the variation present in the original dataset, given by the
correlations between the original traits. The new traits, called principal components (PCs), are uncorrelated, and are ordered by the fraction of the total information each retains. In other words, PCA analysis is a linear function which maps the original variables to new variables. So far, there weren’t any study to address jointly analysis of mixed pig’s traits data using either PCA or clustering algorithms. In this thesis, as an appendix of chapter 4, we resort on applying different clustering algorithms and PCA to reveal hidden structure for teat, exterior and exterior traits.

1.3.3 Bayesian Complex Major Gene Segregation Analysis

In quantitative genetics, it is of prime interest that the mode of inheritance of the trait to be fully understood. Central limit theorem play crucial role in underlying theory of complex segregations analysis. In the context of animal genetics, Bayesian complex segregation analysis was first introduced by Hoeschele, I. (1988). Almost a decade latter on, Janss, L.L.G. et al., (1995, 1997) come up with Gibbs sampling within a Bayesian paradigm to search for a Mendelian major gene affecting meat quality traits in a crossed F2 population based on a mixed inheritance multiplicative model. Using Bayesian paradigm (Zeng, W. et al., 2003) detected a major gene affecting height of loblolly pine in some half-diallel progeny populations in forest tree breeding. Sometimes we may encounter with failing to find evidence of Mendelian/ Imprinting transmission of major gene, however, this does not necessarily address the absence of a major gene. In general, depending on quantitative trait, the inheritance of major gene can be complex. Also it is possible that inheritance of quantitative trait to be governed most likely with more than one major gene, as well as gene-environment (GxE) interactions could add more complexity. In this thesis, one of the major goals of Bayesian segregation analyses is to detect a Mendelian major gene for carcass length trait across three pig breeds population. Gibbs sampling algorithm is used to obtain model parameter estimates. By our knowledge there have not been any studies in pig population which covered carcass length trait in terms of major gene analysis. Also this is first time that we turn to estimate 12 posteriors high-dimensional parameters for carcass length trait. This remains as one of main novelty in this thesis. We argue that having high-dimensional posterior distribution
of major gene is important, since the implications of segregation analysis on phenotypic data, can give clues to either commit DNA-based on mapping the putative major gene influencing carcass length trait or not.

1.4 Software and Programming

In this thesis, different softwares and language programming are used. For Rat species we mostly rely on Bioconductor, R language and to some extent Perl scripts. We use Limma package. For pig species we use SAS, ASReml, and finally Maggic software. To handle the pedigree structure, we use SQL. Running Maggic package requires having knowledge in FORTRAN and also shell programming.

1.5 Objectives of This Thesis

In chapter 1, a general introduction about the themes and ideas which are used in this thesis are investigated. In chapter 2, we fit linear model to find set of differentially expressed genes with robust gene ranking measure across three rat microarray experiments. This chapter also considers Metabolomics data derivation from transcriptomics data, in which enzyme concentration optimization is performed when there are no co-segregating metabolites with gene expression or when only available data are gene expressions. Chapter 3 relies on the data which is used in chapter 2. However, in this chapter Genetical genomic approach, gene co-expression networking, Bioinformatics database searching and DNA motif detection are applied to get insight into genes which have functional effect on rat hind limb muscle. Genetical genomics is applied for set of candidate genes. These set of candidate genes are due to results of chapter 2. The key objective for this chapter is to winnow down hub-acting genes or set of genes which show more connectivity with respect to other genes. We will call gene(s) with the highest total connectivity value (summing over gene in-degree and out-degree) as hub-acting gene(s). In this chapter, the possibility of supporting Genetical genomics approach for the results of gene co-expression network and vice versa in the context of Microarray data is investigated. By our knowledge, this sort of genomics analysis, has not been done so far
in the context of Rat hindlimb muscle Microarray gene expression data, therefore, the results of this chapter likely can shed some light on unseen part of Rat hind limb muscle genome. Chapter 4 presents large-scale additive and maternal genetic computation for pig performance traits (exterior, teat and reproductive traits). This computation is done for three Swiss pig breeds. For carrying out genetic computations, we use two kinds of modeling: Univariate and pair-wise genetic analysis. In univariate genetic analysis, we apply four models per trait. Also in this context, model comparison using asymptotic theory is done. To our knowledge, this is first time that large scale model comparison in the context of pig performance trait across three pig breeds is conducted. Also, by our knowledge, this is the first time, that large scale estimation of maternal genetic component for pig performance traits is done. Having large scale estimated genetics components for aforementioned traits, can have direct or indirect effect on economical breeding program. The objective of this chapter is to provide genetic parameter estimates that can be used in a national Swiss pig breeding program. As an appendix of this chapter, different unsupervised agglomerative hierarchal clustering algorithms accompanied by two principal component analyses (PCA) are used to explore pig exterior traits. By doing so, we can explore how traits are structurally related to each other and how much information are exist in original dataset. Chapter 5 deals with estimation of many genetics components of an autosomal major gene which sought to influence carcass length trait across three pig breeds. The breeds are the same breeds which are used in chapter 4. We use mixed inheritance multiplicative model in the context of Bayesian paradigm. To estimate the model parameters, we use Gibbs sampling. To our knowledge, this is the first time that large scale estimation of posterior distributions not only for carcass length trait, but also in major gene effects context is carried out. One of novelty of this study mostly backs to much attention which was put on dominance effect estimation. The objective of this chapter was to see the portion of genetic variation of these traits due to major gene effects and compare the breeds. Chapter 6 presents some general discussion about results which obtained from the materials and chapters which are covered in this thesis. This thesis can be seen as collection of new methods which researchers are using in diverse areas, it is hope that this thesis would be of very tiny contribution in areas and subjects which covered up.
Chapter 2

Expression Profiling of Genes Involved in Muscular Activity and Inactivity in Hind Legs of Rats for Comparative Transcriptomics

2.1 Abstract

DNA microarray technology allows simultaneous determination of the expression level of thousands of genes. It creates rich information in terms of expression profile of genes under study. In this study, gene expression profile of rat hind limb muscle data in three different experiments was used to investigate the differentially expressed genes. The first experiment was 35-days-cage-controls versus 35-days-suspension, the second experiment was 14-days-cage-controls versus 14-days-suspension and the third experiment was 14-days-suspension + 1day-reloading versus 14-days-suspension + 5 days-reloading. In total, matrix of expression of whole three experiments contained 1184 genes profiled in 39 arrays. A nested linear model with appropriate design matrix was defined in LIMMA for gene expression. Different contrasts were done within different and across all experiments to find genes which showed different expression profile. Moderate t-statistics was used to indicate mode of differentially expressed genes and to rank them. Parallel inference for genes posed course of dimensionality in term of adjust for multiple testing, e.g., control family-wise error rate (FWE) or positive false discovery rate (pFDR). Three thresholds (cut off) for of pFDR (5%, 10%, and 20%) were used to end up with differentially expressed genes. The results indicated that the experiment three led to large number of genes which showed differentially expressed pattern than either experiment one and two. In threshold 5% pFDR, in experiment three 138 genes, turned out to be differentially expressed, with the majority of these genes showing up-regulated
style (73 genes). In experiment three mode of up-regulated style was dominant in the same level of threshold with respect to other experiments. Just one gene (which encodes Glutathione-S-transferase) showed differentially expressed pattern over all three experiments at threshold 20% of FDR. The results showed that based on moderated t-statistics criterion, three experiments (14-days-suspension + 1day-reloading versus 14-days-suspension + 5 days-reloading) had profound effect on expression pattern of all pros. These different experiments affected genes which were most likely involved in ATP production and fatty acid cycle metabolism. Hence, the implications of this result are that the supply of energy source is of greater importance for longstanding work and exercise than protein and other mediator nutrient molecules. These results can have significant impact on adaptive animal and human physiology.

### 2.2 Introduction

DNA microarray technology allows simultaneous determination of the expression level of thousands of genes. This information is provided as matrix of expression. Rows of this matrix represent number of genes and number of columns represents the expression level of gene in given time or assay or condition and so forth. One of the primary and fundamental goals for this data is to detect differentially expressed genes over time or over conditions. Recently Storey, et al., (2007) proposed a new approach aimed at optimally performing many hypothesis tests in a high-dimensional study. This approach which estimates the optimal discovery procedure (ODP) has recently been introduced and theoretically shown to optimally perform multiple significance tests in particular for identifying differentially expressed genes in microarray experiments. Having two groups of arrayed individuals (group 1 and group 2) when there is asymmetry in the differential expression, i.e. more genes are over-expressed in group 2 than in group 1, ODP is a promising tool to extract biological signals. The t-statistics and its diverse extensions play key role in gene expression analysis. Opgen-Rhein, et al., (2007) succinctly categorized t-statistics based on gene ranking criteria into four categories. They also introduced new gene ranking criterion so-called shrinkage t-statistics. They emphasized that the improved SAM statistic (Wu, B. L. 2005) generally provides very poor rankings of genes. Also
they indicated that moderated t-statistics and shrinkage t-statistics are the only two methods that perform optimally in all three simulation settings. In comparison, SAM with two test statistics (t-statistics and Wilcoxon Rank Sums test) with empirical Bayes for real and simulated expressed dataset, it was concluded that while SAM is the most powerful method in the analysis of the simulated data set, it performs worse in the applications to the real data sets (Smyth, G.K. 2004). Based on these arguments empirical Bayes based on gene ranking criterion (moderated t-statistics) was used for this study. This method promises to identify most genes which are really expressed across conditions. On the other hand, one of the main ideas of this study was to jointly analyze three different experiments. This can be easily formulated using the above (Smyth, G.K. 2004) approach. Enormous flexible contrasts can be defined for each motivating hypothesis. This study is based on three gene expression experiments conducted in rats at the University of Bern, Switzerland for understanding the impact of different treatments (of activity and inactivity) on soleus muscle gene expression patterns, with an overall objective of getting insights into corresponding human muscles via comparative transcriptomics.

2.3 Material and Methods

2.3.1 Microarray Experiment

In this study, gene expression profiles of rat hind limb muscle (rat soleus muscle) data in three different experiments were used to investigate the differentially expressed genes pattern and regulation of this muscle. The soleus muscle is a powerful muscle in the back part of the lower leg (the calf). It runs from just below the knee to the heel, and is involved in standing and walking and running. The first experiment was 35-days-cage -controls vs. 35-days-suspension, the second experiment was 14-days-cage-controls vs14-days-suspension and the third experiment was 14-days- suspension-plus-1-day reloading vs. 14-days-suspension- plus + 5 days reloading. These sets of experiments were done to produce transcript profiles of genes which most likely respond to these treatments in rat hind limb muscle. Doing this type of gene transcript profiling is not only important to get
insight into rat’s hind limb muscle but also can be used in comparative genomic studies. This was the case for these experiments as one of the main ideas was to make comparisons between human and rat soleus muscles from system biology viewpoints. For all of these three experiments, a home-made array was used to make profiles of genes in the rat hind limb muscle. This home-made array contained 1200 spots. Every spot in the array was one gene. For gene expression profiling less number spots were used. In total, matrix of gene expression of all three experiments contained 1184 genes profiled in 39 arrays. For this dataset, background data also was obtained but because of increasing variability in gene expression levels, the applied expressed data wasn’t background modified. For information in this regards, readers are requested to consult (Fluck, M. et al., 2005). Skeletal muscle atrophy is a change that occurs in muscles of adult animals as a result of the conditions of disuse (e.g., immobilization, denervation, muscle unloading), aging, starvation, and a number of disease states (i.e., cachexia), and characterized by a decrease in protein content, fiber diameter, force production, and fatigue resistance. This study was designed to explore different types of conditions i.e. muscle loading reloading, gravity on producing muscle atrophy, since these different conditions imply different types of molecular triggers and signaling pathways for muscle atrophy.

2.3.2 Finding Differentially Expressed Genes

To achieve consistency between arrays in terms of expression values (variance stabilization), between array normalization was used. For this variance stabilization/ data calibration, different methods have been suggested in literature (Bolstad, B.M. et al., 2003; Smyth, G.K. et al., 2003; Yang, Y.H. et al., 2002). Because in our analysis, we did not involve with any kind of formal platform (say, two-channels spotted cDNA array, Affymetrix and etc.) and the data was a result of raw expressed data of home-made arrays, we used robust estimation of variance-stabilizing and calibrating transformations for microarray data. The method calibrates for sample-to-sample variations through shifting and scaling, and transforms the intensities to a scale where the variance is approximately independent of the mean intensity. Figure 2. 1 shows the differences between, before and after normalization.
Figure 2. 1 Before and after normalization.

The variance stabilizing transformation is equivalent to the natural logarithm in the high-intensity range, and to a linear transformation in the low-intensity range. The problem
with this method is that it is slow in nature which is due to the nested iteration loops of
the numerical optimization of the likelihood function and the heuristic that identifies the
non-outlying data points in the least trimmed squares regression (Limma). However, we
did not have such problem first since we did not have any missing values (so there was
not any need to missing values imputations) and the second we did not have quite large
number of arrays in our analysis. **Figure 2.** shows the results on raw expression data. It
patently shows that between arrays normalization must be done before analysis. A nested
linear model to end up with a set of a differentially expressed gene was as follows:

\[
Y = \beta_0 + \sum_{m=1}^{2} \beta_{mi}(x_{mi}) + \sum_{l=1}^{l=3} \beta_{li}(x_{li}) + \epsilon_i
\]

Where: \(Y\) was the vector of gene expression for a given gene \(\{Y_i, i = 1,...,39\}\) that is, 39
expression values per gene; \(\beta_0\) was grand mean; \(\beta_{mi}\) was coefficients for experimental
effect for a given gene; \(\beta_{li}\) was coefficients for nested treatment within experiments for a
given gene; \(x_{mi}\) was incidence matrix for experiments with 39 by 2 dimensionalities for a
given gene; \(x_{li}\) was incidence matrix for nested effect with 39 by 3 dimensionality for
given gene; \(\epsilon\) was vector of unknown random errors for given gene \(\{\epsilon_i, i = 1,...,39\}\). It
was assumed that errors are normally, identically and independently distributed. Also it
was assumed that error variance is constant across population groups of model. In this
analysis model, the “linear” in model referred to the linearity of the mean of \(Y\)
(expression values for a given gene) in the parameter \(\beta\) for fixed values of \(x\). Linear
model advantage in analysis of microarray data can be addressed as: (1) analyzing all
arrays together, (2) combining information in an optimal way, (3) combined estimation of
precision, (4) extensible to an arbitrarily complicated experiments, (5) design matrix
flexibility which specifies RNA targets used on arrays and (6) flexibility of contrast
matrix which specifies which comparisons are of interest. There may be more or fewer
contrasts than coefficients for the linear model, although if more, then the contrasts will
be linearly dependent. In our analysis, the contrasts for mean of expressions within three
experiments were obtained, although across experiments contrasts were obtained as well.
Above model was written in the sense that an intuitive mapping from statistical model to its corresponding R codes can be easily constructed. However, we can write down above model in more general way as follows:

$$Y_{ijk} = \mu + E_i + T_{j(i)} + e_{ijk}$$

Which: $Y_{ijk}$ is $k_{th}$ expression value in $j_{th}$ treatment in $i_{th}$ experiment, $\mu$ is grand mean, $E_i$ is effect of experiment, $T_{j(i)}$ is effect of treatment within experiment and $e_{ijk}$ is error term. However, in our view, the earlier parametric model is intuitively less demanding and suitable in the context of gene expression study.

2.3.3. Ranking Genes

For a given contrast null hypothesis for a given gene was defined as gene expression value which is not significantly different from zero. Against this hypothesis, alternative hypothesis is defined as gene expression value which is significantly different from zero. And finally the null hypothesis is rejected if moderated t-statistic is greater or less than what is expected by chance at some predetermined significance level. Moderated t-statistics defined as

$$\tilde{t}_g = \frac{\bar{M}_g}{\tilde{s}_g}$$

which $\bar{M}_g$ was the fraction of average fold change; $n$ was the number of gene expression values for interested in contrast $\tilde{s}_g$ was square root of

$$\tilde{s}_g^2 = \frac{s_g^2d_g + s_0^2d_0}{d_g + d_0}$$

where $s_g^2$ and $d_g$ and $s_0^2$ and $d_0$ were variance of gene, freedom of gene, variance of mean and degree freedom of mean respectively. In moderated t-statistics variability of estimate is adjusted by a prior variance as above. Because of framework of linear model, genes which were not differentially expressed (or had equal means across conditions/contrast), the moderated t-statistics for these genes were zero. In other words, genes which had negative moderated t-statistics were considered as down-regulated genes and genes with positive moderated t-statistics were considered as up-regulated genes. Statistically speaking, estimated moderated t-statistics, like many other test statistics, is a random value; therefore, there is corresponding p-value for it. As a
result, we can end up with a vector of 1184 p-values for interesting contrasts. In this step we should tackle with course of multiple comparisons. Many methods of multiple-comparison adjusted p-values were used to handle the multiple comparison issues in the context of microarray data. Less conservative multiple-comparison corrections are included (Benjamini, Y. et al., 1995, 2001; Hommel, G. 1988) respectively. Hochberg's and Hommel's methods are valid when the hypothesis tests are independent or when they are non-negatively associated (Sarkar, S.K. 1998). Hommel's method is more powerful than Hochberg's, but the difference is usually small and the Hochberg p-values are faster to compute. In this study, Benjamini, Y. et al., (1995) method was used. The vector of p-values adjusted for multiple testing (which some time called q-values) was sorted in ascending way. For interesting contrasts, three thresholds (cut offs) were applied arbitrarily. These thresholds were 0.05%, 10% and 20% of probability of false discovery rate (pFDR) or q-values. The genes which were included in each one of these categories were investigated in terms of mode of regulation (up or down regulation). This was done for all three cut off vales for each contrast.

2.4. Results and Discussion

2.4.1. Differentially Expressed Genes

Since a cell always uses a small fraction of its genes to respond to specific perturbation, we would expect a small number of genes to show differential expression pattern. One way to show this is to make a histogram of pFDR of involved genes. If the histogram of pFDR (q-values) followed uniform distribution, it does mean no gene is being turned out to be differentially expressed. One of the other ways to understand the proportions of genes which have been differentially expressed is plotting gene ranking criteria against their behavior in null hypothesis distribution. Figure 2.3 summarized results of all three experiments in terms of differentially expressed genes. First row of this figure belongs to the first experiment, the second row belongs to the second experiment and third row belongs to the third experiment. This was done in the sense that across experiment
comparisons (column-wise) would be readily done too. The left column is a histogram of pDFR (q-values) for the first, second and third experiments. The middle column is an empirical density plot of moderated t-statistics for the contrast of the first, second and third experiments as points and fitted normal density as line. The right column is the ordinate axis on log scale of the first, second, third and log of posterior odd (B) against difference of log intensity(M) or the average fold change in the first, second and third experiments. The calculation of the empirical density of the data is a fundamental element in constructing the log-density plots. As a result, first empirical density of moderated t-statistics was calculated and then a plot was created. In practice, the empirical density is usually calculated using a histogram (scaled to have unit area) or its generalization, a kernel density estimate. The plots constructed using kernel density estimation is the more appealing in the gene expression area, and so we concentrated on this version of the methodology. First box in the first column of Figure 2.3 shows small distortion from uniform distribution and indicates some genes were differentially expressed. First box in the second column of Figure 2.3 shows a closer look at the computational task for the first experiment, indicating that the plot of empirical density of moderated t-statistics (points) and fitted normal density (line) approximately overlap each other. In the left tail of the curve a small deviation can be observed. As it is indicated, as log of density is becoming small, moderated t-statistic show more deviate from its distribution in null hypothesis (solid line). However the first box in the second column in Figure 2.3 indicates that estimated moderated t-statistics with some deviations follows the optimal normal density curve. In the first box in the third column in Figure 2.3 the shape of log of posterior odd (B) against difference of log intensity (M) is usually parabolic which is shown. Most genes have an average value around zero, and these have the minimum values of B. In general, the larger the average expression of a gene, the larger is the chance of high B. The actual B-level of genes with a high averages depends on the variance, so that the outer ends of parabola will be rather fuzzy. Two genes with exactly the same average might be far apart in B.
Figure 2.3 Insightful plots to visualize state of differentially expressed genes.

The results indicated that smallest number of differentially expressed genes turned out to be in the first experiment with respect to the other two experiments in all three thresholds. From 1184 genes, at threshold 5% of pFDR, just one gene (D64045, phosphatidylinositol 3-kinase) turned out to be differentially expressed. The protein of this gene's mRNAs is involved in virtually all of the intracellular transducers/effectors/modulators and cytoplasmic actions. The result revealed that this gene was up regulated. Figure 2.4 shows the results. At threshold 10% pFDR, 9 genes turned out to be differentially expressed. From these nine genes, 6 genes showed an up-regulated pattern. Genes that encoded cytoplasmic proteins and nuclear proteins are implicated.
Figure 2.4 Pattern of differentially expressed genes of first experiment at three levels of pFDR.

The profile of genes involved in cell adhesion receptors/proteins was affected differentially i.e., the mRNA of DNA binding and chromatin proteins was changed (increased) with suspension. These up-regulated genes were mostly involved in protein, amino acid and nucleic acid metabolism. For example, (1) cytoplasmic proteins, (2) DNA binding and chromatin proteins, (3) extracellular secreted proteins and (4) other intracellular transducers/effectors/modulators were among the polypeptides which were coded by these genes. In other words, in switching from control to suspension treatment, the soleus muscle has increased a kind of genes which encode polypeptide involved in protein and acid nucleic metabolism. The genes for which the expression level was diminished with suspension were (U08141) sodium/calcium exchanger NCX2, (X64589) G2/M-specific cyclin B1 (CCNB1) and (U82591) growth-related c-myc-responsive protein RCL. At threshold 20% FDR, 16 genes were differentially expressed of which 13 genes showed an up-regulated pattern. In other words, the number of down-regulated genes decreased. This result indicates that even at high level of 20% of pFDR, only 1.4% percent of all genes showed differently expressed behavior, most of them (82%) showed up-regulated pattern. This indicates either the experiment was not strong enough to induce any signal for triggering gene expression or genes did not show any response to the experimental effect. Contrary to a similar diagram in the first experiment, the histogram of pFDR in the second experiment (Figure 2.3) represented that more genes
were differentially expressed. **Figure 2. 3** showed more distortion from uniform distribution than corresponding pFDR plot in the first experiment and indicates some more genes were differentially expressed. Plotting gene ranking criteria (in our case moderated t-statistics) against their behavior in null hypothesis distribution in the second experiment shows more deviation from null hypothesis than in experiment one. We could also say that the contrast which was made to show differentially expressed genes in the second experiment showed more differentially expressed genes than in the first experiment. At 5% of pFDR, 15 genes did show differentially expressed pattern. Among these differentially expressed genes, 12 genes showed down-regulated behavior and the rest showed up-regulated pattern. In 5% of pFDR, the expression level of (J02592) Glutathione-S-transferase, (M60921) B-cell translocation gene 2 and (M92340) interleukin 6 receptor beta chain were increased (up-regulated) by switching from 14-days-cage-controls to 14-days-suspension. Among these genes, (M92340) interleukin 6 receptor beta chain is a gene which encodes different polypeptides which are involved in different functions. It was indicated by switching from 14-days-cage-controls to 14-days-suspension, rat hind limb soleus muscle increased the transcriptional level of genes which are involving in membrane cell mediation processes. Also, the levels of most differentially expressed genes (80%) were diminished by switching from 14-days-cage-controls to 14-days-suspension. **Figure 2. 5** shows the correspond results. When the time of suspension was increased, this not only made that different genes showed differentially expressed pattern, but also that the mode of differentially regulated gene changed inversely. At threshold 10% of pFDR 18 genes did show differentially expressed pattern. Even though threshold was two folded but it didn’t identify more of differentially expressed genes, as was the case for the first experiment. In this experiment, contrary to the first experiment, most identified differentially expressed genes showed a down-regulated mode.
The trend of mode of differentially expressed genes in this experiment showed opposite behavior in terms of mode of differentially expressed genes with respect to the first experiment. Some genes like (U14007) Aquaporin 4 and (M62781) Insulin-like growth factor-binding protein 5 were differentially expressed. Actually these genes encode polypeptides which are functioning as membrane channels and transporters, extracellular transport/carrier proteins and extracellular secreted proteins. About the 71% of differentially expressed genes lost their transcriptional profile by switching from 14-days-cage-controls to 14-days-suspension. In other words, the number of up-regulated genes increased by an increasing cut off value. The question whether the same gene was differentially expressed over experiments is investigated later. At level of threshold of 20% of pFDR, 19 genes showed a differentially expressed pattern, among which 13 genes showed a down-regulated pattern. At 20% of pFDR, the level of transcriptional profile of 33% of differentially expressed genes increased by switching from 14-days-cage-controls to 14-days-suspension. These 33% were due to the (X78461) potassium channel RB-IRK2 gene which encodes different polypeptides, voltage-gated ion channels and plasma membrane proteins. About 67% of the differentially expressed genes lost their expression profile by switching from 14-days-cage-controls to 14-days-suspension. For this experiment it was indicated that as the level of cut off values increased, the ratio of genes which lost their expression profile (down-regulated genes) by switching from 14-days-
cage-controls to 14-days-suspension, decreased. Different plots in the last row of Figure 2. 3 patently clear that this experiment gave arise to more differentiated genes than the other experiments. The main difference that can be seen between this diagram with corresponding diagrams in the experiments described above is the strong tendency of expression values toward zero in such way that observed expression values sharply accumulated around zero. As it was expected, contrast within this experiment culminated in quite more differentially expressed genes.

![Bar chart showing differentially expressed genes](image)

**Figure 2. 6** Pattern of differentially expressed genes of third experiment at three levels of pDFR.

The modes of differentially expressed genes were more or less similar to the trend observed in experiment two. Most differentially expressed genes showed a down-regulated differentially expressed mode. At threshold of 5% of pFDR in the third experiment, 138 genes turned out to be differentially expressed. Among these differentially expressed, 65 genes showed a down-regulated mode and the rest showed an up-regulated mode. At 10% of pFDR, 178 showed differentially expressed behavior. On this level, the number of up-regulated genes increased in comparison to the number of down-regulated genes, and this was the case at the 20% threshold of pFDR. **Figure 2. 6** shows the results. In 14-days-suspension + 1day-reloading the number of 6 rats was arrayed and in 14-days-suspension + 5 days-reloading the number of 5 rats was arrayed. In comparison with the other experiments, the dimensionality of the gene expression
matrix for this experiment was less than two of the aforementioned experiments and this could be a little bit problematic in nature. Some genes like: (U82591) growth-related c-myc-responsive protein RCL and (X64589) G2/M-specific cyclin B1, (X57523) antigen peptide transporter 1, (U34959) transducin beta-2 subunit and GTP-binding protein G(i)/G(s)/G(t) beta subunit 2 (GNB2) which were involved in signal transduction, highly up-regulated in our analysis. In other words, the level of their transcriptional levels increased by switching from 14-days-suspension + 1day-reloading to 14-days-suspension + 5 days-reloading. In the same manner, (D16237) M-phase inducer phosphatase 2 (MPI2) and cell division control protein 25 B (CDC25B) which were involved in cell cycle highly up-regulated. Based on available gene ontology information at 5% of pFDR, quite many genes which were involved in nucleotide metabolism like: (J05167, M91597, D10754, D13376, 4835, J02791, M30581, X56133, J02701, D21800, Z14117, M17086, M22412, D10041, M22413, U59809, U09540, D30040 up-regulated. This was also the case for genes which were involved in electron transport: (D13124) ATP synthase, (D30647) Acyl-Coa dehydrogenase, (U22830) P2 purinoreceptor subclass 2Y, (M34728) sterol carrier protein 2 (SCP2).

2.4.1. Comparison among Experiments

Comparison experiments can reveal genes that respond to different experiments. In this section we made comparisons among experiments to find commonly differentially expressed genes and to understand their mode of regulation. Within three thresholds of 5%, 10% and 20% of pFDR (Table 2.1, Table 2.2 and Table 2.3), a comparison was made to find firstly which genes were commonly differentially expressed across experiments and secondly what was their mode of regulation i.e. up-regulated or down-regulated mode. Table2.1, Table 2.2 and Table 2.3 show these comparisons. As presented in experiment one, just one gene turned out to be differentially expressed. At this level of pFDR, the comparison among experiments one and two and experiments one and three did not end up with in common differentially expressed genes (Table 2.1). Comparing experiment two with experiment three ended up with 8 in common differentially expressed genes. These genes i.e. heart fatty acid-binding protein (H-FABP)
and Glucose-6-phosphate dehydrogenase were involved in energy metabolism and lipid metabolism. Later investigations showed that these genes were down-regulated in both experiments two and three. In comparison of experiment one to experiment two at 10% pDFR level (Table 2.2) in total, two genes turned out to be differentially expressed (Glutathione-S-transferase; Insulin-like growth factor-binding protein 5). Comparing experiment one with experiment three, two genes (growth-related c-myc-responsive protein RCL; G2/M-specific cyclin B1 (CCNB1) jointly turned out to be differentially expressed. Taking a look at the mode of these genes indicated that both of these genes were down-regulated in experiment one and up-expressed in experiment three. In comparison of experiment two to experiment three, a number of genes turned out to be differentially expressed. Most of these genes which are involved in carbohydrate metabolism like ATP synthase, Glucose-6-phosphate dehydrogenase and Fructose-1 were down-regulated in experiment two but up-regulated in experiment three. This was also the case for other genes which function in the metabolic pathway of fatty acids like fatty acid binding protein 2, heart fatty acid-binding protein (H-FABP) and more. Just one differentially expressed gene showed the diverse way of regulation, that is, up-regulation in experiment two and down-regulation in experiment three (interleukin 6 receptor beta chain). As Table 2.3 shows, comparisons at 20% pDFR ended up with a diverse mode of regulation for the same genes across second and third experiments. These comparisons indicated that experiments work differently for the same set of genes. In other words, this indicates that regulation program of these genes work inversely in terms of these two experiments. In this study we used moderated t-statistics because fold change is the simplest method for identifying differentially expressed genes (Derisi, J.L. et al., 1997; Draghici, S. 2002; Schena, M. et al., 1996). Normalized expression measures are typically computed on the log scale. An estimated difference in means on the log scale can be converted to an estimated fold change on the original scale. For example, for given contrast between two treatments we estimated $\tau_1 - \tau_2$ to be 1.514. This translates into an estimated fold change of $\exp(1.514) = 4.54$. This means that treatment 1 is estimated to increase the expression of the gene by a multiplicative factor of 4.54 relative to its level under treatment 2. Standard t-statistics can be used in gene expression data. Standard t-test is robust to variance heterogeneity across genes but it may have low power due to
small degrees of freedom. In addition, the variances estimated from each gene are not stable. Small estimated variances obtained by chance can produce larger t-tests even when the corresponding fold changes are small. This statistics test is modified in gene expression in different number of ways based on variance heterogeneity, homogeneity across genes, number of degree of freedom, variance stability across genes and so on. The regularized t-test combines information from gene-specific and global average variance estimates by using a weighted average of the two as the denominator for a gene-specific t-test (Baldi, P. et al., 2001). The B statistic (Lonnstedt, I. et al., 2002) allows for gene specific variances but it also combines information across many genes using an empirical Bayes approach and thus should be more stable than the ordinary t-statistic. The volcano plot (Cui, X.Q. et al., 2003) is a scatter-plot of the –log10-transformed p-values from the gene-specific t-test against the log₂ fold change. This method is a dependent method. It mainly goes back to the kind of model which is fitted to data and range of fold change which the researcher is interested in. From a mathematical point of view most of these approaches are taken into moderated t-statistics which make it to be a robustly competitive gene ranking measure across range of microarray platforms (Opgen-Rhein, R. et al., 2007). In the appendix a general formula is driven to make a bridge from expression level to metabolic level with assumption of some general considerations. This derived formula can provide researchers with insightful information about metabolic level when rich area of metabolic information is scant.
2.5. Checking the Model

![General plots regarding to model checking.](image)

**Figure 2.7** General plots regarding to model checking.

Usually it is not known in advance whether an applied model in the study is appropriate or not. Therefore, it is necessary to conduct a search focused on residuals to look for evidence that either the necessary assumptions are violated or not. In model building, a residual is what is left after the model is fitted. In our analysis, the true errors are assumed to be independently normally distributed with a mean of zero and a constant variance. If the model is appropriate for the data, the observed residuals which are estimates of the true errors should have similar characteristics. One of convenient method to detect this matter is to plot the residuals against the predicted values. If the assumptions of linearity and homogeneity of variance are met, there should be no relationship between the predicted and residual values. We should be suspicious of any observable pattern. Systematic patterns between the predicted values and the residuals suggest possible violations of the linearity assumption. If the assumption was met, the residuals would be randomly distributed in about the horizontal straight line through zeros. **Figure 2.7** shows plotting of predicted error against predicted gene expression values. As it can be seen, an overall pattern can not be detected. This indicates that the applied model has been well fitted to the data and most model assumptions have been met. Next we proceed to check the model by plotting the residuals due to the model. This way we can tell
whether or not the assumptions of the model have been met. The distribution of residuals may not appear to be normal for reasons other than actual non-normality: misspecification of the model, non-constant variance, a small number of residuals actually available for analysis, etc. Therefore, we should pursue several lines of investigations. One of the simplest methods is to construct a histogram of the residuals. One of the assumptions of fitted models was errors are independently, identically normally distributed. Figure 2.7 shows roughly this situation though some deviation from this assumption can be seen. Three different experiments were analyzed together. It was expected that errors of these three experiments be different from each other. This is why graphs in Figure 2.7 do not exactly follow a normal distribution. However, it is unreasonable to expect the observed residuals to be exactly normal - some deviation is expected because of sampling variation. Even if the errors are normally distributed in the population, sample residuals are only approximately normal.
<table>
<thead>
<tr>
<th>Comparisons</th>
<th>Gene names</th>
<th>Gene annotations</th>
<th>Classification</th>
<th>Mode of regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 days cage controls; 35 days suspension VS 14 days cage controls; 14 days suspension (Experiment one VS Experiment two)</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>Up in both experiments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Down in both experiments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Down in Expr1 Up in Expr2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Down in Expr1, Up in Expr2</td>
</tr>
</tbody>
</table>
Table 2. In common genes which have differentially expressed in pair-wise comparison of all three experiment at 5% pFDR.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>Gene names</th>
<th>Gene annotations</th>
<th>Classification</th>
<th>Mode of regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Up in both experiments</td>
</tr>
<tr>
<td>35 days cage controls; 35 days suspension VS. 14 days cage controls; 14 days suspension. (Experiment one VS Experiment two)</td>
<td>J02592</td>
<td>Glutathione-S-transferase</td>
<td>FL08 SC00</td>
<td>J02592</td>
</tr>
<tr>
<td></td>
<td>M62781</td>
<td>Insulin-like growth factor-binding protein 5</td>
<td>FF00 SF00</td>
<td>M62781</td>
</tr>
<tr>
<td>35 days cage controls; 35 days suspension VS. 4 days suspension + 1 day reloading; 14 days suspension + 5 days reloading (Experiment one VS Experiment three)</td>
<td>U82591</td>
<td>growth-related c-myc-responsive protein RCL</td>
<td>FC00 SN00</td>
<td>null</td>
</tr>
<tr>
<td></td>
<td>X64589</td>
<td>G2/M-specific cyclin B1 (CCNB1)</td>
<td>FC01 SC00 SN00</td>
<td>null</td>
</tr>
<tr>
<td>14 days cage controls; 14 days suspension VS. 4 days suspension + 1</td>
<td>X64827</td>
<td>Cytochrome c oxidase subunit VIII-H (heart/muscle)</td>
<td>FL00 SM00</td>
<td></td>
</tr>
<tr>
<td>GenBank Accession</td>
<td>Description</td>
<td>FL01</td>
<td>FL02</td>
<td>FL03</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>D13123</td>
<td>ATP synthase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X07467</td>
<td>Glucose-6-phosphate dehydrogenase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M35992</td>
<td>Fatty acid binding protein 2</td>
<td>FT10</td>
<td>FK05</td>
<td>SC00</td>
</tr>
<tr>
<td>J02773</td>
<td>heart fatty acid-binding protein (H-FABP)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M16235</td>
<td>hepatic triacylglycerol lipase</td>
<td>FL03</td>
<td>FL04</td>
<td>FL08</td>
</tr>
<tr>
<td>M35991</td>
<td>Fatty acid binding protein 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J05167</td>
<td>Solute carrier family 4</td>
<td>FI05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M86240</td>
<td>Fructose-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M92340</td>
<td>interleukin 6 receptor beta chain</td>
<td>FR02</td>
<td>FE00</td>
<td>SB00</td>
</tr>
<tr>
<td>M22413</td>
<td>carbonic anhydrase 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** In common genes which have differentially expressed in pair-wise comparison of all three experiment at 10% pFDR.
<table>
<thead>
<tr>
<th>Comparisons</th>
<th>Gene names</th>
<th>Gene annotations</th>
<th>Classification</th>
<th>Mode of regulation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>35 days cage controls; 35 days suspension VS. 14 days cage controls; 14 days suspension (Experiment one VS Experiment two)</td>
<td>J02592</td>
<td>Glutathione-S-transferase</td>
<td>FL08 SC00</td>
<td>Up in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td></td>
<td>M62781</td>
<td>Insulin-like growth factor-binding protein 5</td>
<td>FF00 SF00</td>
<td>Down in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td>35 days cage controls; 35 days suspension VS. 4 days suspension + 1 day reloading; 14 days suspension + 5 days reloading (Experiment one VS Experiment three)</td>
<td>U82591</td>
<td>growth-related c-myc-responsive protein RCL</td>
<td>FC00 SN00</td>
<td>Up in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td></td>
<td>X64589</td>
<td>G2/M-specific cyclin B1 (CCNB1)</td>
<td>FC01 SC00 SN00</td>
<td>Down in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td></td>
<td>Y00697</td>
<td>Cathepsin L</td>
<td>FU05 SL00</td>
<td>Down in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td></td>
<td>U66274</td>
<td>Neuropeptide Y5 receptor</td>
<td>FR04 FV03 SB00</td>
<td>Down in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td></td>
<td>D13126</td>
<td>neural visinin-like Ca2+-binding protein type 3</td>
<td>FT08 SC00 SB00</td>
<td>Down in both experiments</td>
<td>Down in both experiments</td>
</tr>
<tr>
<td>J02592</td>
<td>Glutathione-S-transferase</td>
<td>FL08 SC00</td>
<td>Up in both experiments</td>
<td>Down in both experiments</td>
<td>Up in Expr1 Down Expr2</td>
</tr>
<tr>
<td>Accession</td>
<td>Gene Name</td>
<td>Accession</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X64827</td>
<td>Cytochrome c oxidase subunit VIII-H (heart/muscle)</td>
<td>FL00 SM00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D13123</td>
<td>ATP synthase</td>
<td>FI04 FL03 SM00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X07467</td>
<td>Glucose-6-phosphate dehydrogenase</td>
<td>FL02 FL03 SC00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M35992</td>
<td>Fatty acid binding protein 2</td>
<td>FT10 FK05 SC00 SB00</td>
<td>M35992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J02773</td>
<td>heart fatty acid-binding protein (H-FABP)</td>
<td>FT10 SC00</td>
<td>J02773</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M16235</td>
<td>hepatic triacylglycerol lipase</td>
<td>FL03 FL04 FL08 SF00</td>
<td>M16235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M35991</td>
<td>Fatty acid binding protein 1</td>
<td>FT10 SC00</td>
<td>M35991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J05167</td>
<td>Solute carrier family 4</td>
<td>FI05 SB00</td>
<td>J05167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M86240</td>
<td>Fructose-1</td>
<td>FL01 SC00</td>
<td>M86240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M92340</td>
<td>interleukin 6 receptor beta chain</td>
<td>FR02 FE00 SB00 SF00</td>
<td>M92340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M22413</td>
<td>carbonic anhydrase 3</td>
<td>FL00 SC00</td>
<td>M22413</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J02592</td>
<td>Glutathione-S-transferase</td>
<td>FL08 SC00</td>
<td>J02592</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X78461</td>
<td>potassium channel RB-IRK2</td>
<td>FI03 SB00</td>
<td>X78461</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

14 days cage controls; 14 days suspension VS 4 days suspension + 1 day reloading; 14 days suspension + 5 days reloading (Experiment two VS Experiment three)

Table 2.3 In common genes which have differentially expressed in pair-wise comparison of all three experiment at 20% pFDR.
2.6. Implications

In this study, three experiments were conducted to understand the impact of different loading, reloading and gravity effects on rat soleus muscle gene expression. The first time transcriptional values of 1184 genes of hind limb rat muscle data in three experiments were jointly analyzed. This joint analysis of genes removed most noise and systematic effects in such a way that comparisons of genes can be done with committing less error rate. It was indicated that quite a large number of genes turned out to be differentially expressed in the third experiment. Also, it was indicated that these different experiments affected genes which most likely were involving in ATP producing cycles and fatty acid metabolism. As a result from this kind of muscle exercise the supply of a source of energy is of greater importance for longstanding work and exercise than protein and other mediator nutrient molecules. Also it was attempted to understand which genes have been differentially expressed across all experiments. Glutathione-S-transferase (J02592) turned out to be a single gene showing this behavior. This gene was used to get clues through its regulations using expression QTLs mapping (eQTLs).

2.7. Acknowledgments

Thanks to Dr. Dirk Müller (dirk.mueller@inf.ethz.ch) from computational system biology at the ETH for his words on proposed idea at appendix and his corrections.

2.8. Appendix

Since we did not have corresponding metabolic data per gene, we tried to make a bridge from “transcriptomic” level to “Metabolicmic” level by considering some assumptions. The following idea can be roughly called reverse engineering. Consider for gene A we have its expression values (transcriptomic level) across two biological condition. This biological condition could be cancer and normal tissues or any other biological motivated experiment. The idea is here that if gene A up/down-regulated across these two biological states, what would happen to its corresponding enzyme activity which actually is functioning at metabolic level and how we can somehow optimize it. Since in this study we were dealing with rat’s muscle array data and muscle enduring exercise we looked at
lactate hydrogenase and oxidase enzymes and their related genes, but there weren’t any gene(s) which actually were differentially expressed with an compelling role for these enzymes, as a result we did not test the following idea for our data. We assume for similar data like our data, the following idea can be used as far as they are dealing with optimizing genes with encoding hydrogenase and oxidase enzymes. Also, it should be interesting when rich biological data would be at premium. Before proceeding, let’s make a rough sketch to give the big picture of the idea:

\[ \text{Gene} \xrightarrow{r_{TC}} \text{mRNA} \xrightarrow{r_{TL}} \text{E} \xrightarrow{r_{deg}} \phi \Rightarrow \text{Steady-State} \]

In the above sketch: \( \text{Gene} \) = gene name which is profiled over the course of experiments, \( r_{TC} \) = rate of transcription of gene to mRNA, \( \text{mRNA} \) = gene expression values (the only thing which we do have at hand!), \( r_{TL} \) = rate of producing enzyme (E) \( E \) = Enzyme, \( r_{deg} \) = rate of degradations, \( \phi \) = indicates degradation and that the degraded species product is NOT balanced (= not modeled). The rate of enzyme activity due to this enzyme can be written as follows:

\[
V_E = \frac{V_E^{\max} \cdot C_s(f(\text{ATP})))}{K_m + C_s} \sim C_E
\]

And \( V_E^{\max} = k_{cat} \cdot C_E \) as result and at steady-state we can write:

\[
\frac{\partial C_E}{\partial t} = 0 \iff r_{TL} - r_{deg} = 0
\]
\[
\iff r_{TL} = k_{TL} \cdot C_{mRNA} = k_{deg} \cdot C_E = r_{deg}
\]
\[
\Rightarrow C_E = \frac{k_{TL} \cdot C_{mRNA}}{k_{deg}} \sim C_{mRNA}
\]

This formula tells that provided we measure stationary expression, levels of enzyme protein are proportional to the absolute mRNA level we measure. Note that this does not necessarily mean that the corresponding enzyme activity scales linearly with the protein level. Post-translational modifications (e.g. phosphorylations) or metabolic regulation (e.g. by ATP, substrate activation or product inhibition) will frequently let the actual enzyme activity deviate from the linear scaling with \( C_E \) even if the protein level is proportional to the mRNA concentration. Moreover, we have completely neglected possible changing
impacts of translational regulation and/or alternative splicing when a gene is up- or down-regulated. But it may well be that the flux change through a metabolic pathway for treatment vs. control is controlled by mechanisms acting on the final enzyme itself without any necessary modification at the transcriptional level (= of its CmRNA). Because we used rat muscle enduring experiments, we suggest that information regarding gene (A) can be found at KEGG assuming that gene (A) is involved in lactate dehydrogenase, oxidase metabolism. However, more than one gene can be found, so this follows:

1. Using KEGG, determine the gene names (for Rattus norvegicus) of all genes encoding the enzymes present (=highlighted in KEGG) in Rattus norvegicus in the following pathways: glycolysis, TCA cycle, respiration (esp. cyt c oxidase), lactate formation (lactate dehydrogenase?) . These are not too many enzymes (maybe, say, 30). One would use the online version of KEGG, select rat and the corresponding pathways and then write down the corresponding genes names manually. That is probably quicker than trying to set up an automated routine, at least if I extrapolate from my programming style.

2. Check which of the above genes are present on the microarray and group them according to the above pathways (you may lump TCA cycle and respiratory enzymess). We need at least one enzyme in the respiratory chain plus the lactate dehydrogenase to make the inference we wish! if this is not possible, there is no point continuing this avenue.

3. if the above precondition is fulfilled we can check:

(i) Whether we get consistent trends in each pathway (ideally all genes should be found to be up- or down-regulated, at least a majority should exhibit a similar behaviour)

(ii) Calculate the ratio of the expression levels for both the treated and the control case: e.g.

\[ \frac{C_{mRNA,Lac-Dehydrogenase}}{C_{mRNA,cyt_c-oxidase}} \]

(our core enzyme comparison) does it change significantly? would we get a similar trend in the ratio (up or down) when we substitute other respiratory enzymes or possibly TCA enzymes (may deviate from the behavior of the respiratory enzymes since the TCA serves many functions) in the denominator? we can make a similar
comparison using glycolytic enzymes (single ones or an average of multiple genes, which may strengthen the credibility of our result, provided the exhibit the same trend) and use the ratio between lactate formation and glycolysis to see whether glycolysis is up-regulated under exercise conditions or whether only the distribution (=split ratio) of glycolytic flux at the pyruvate node towards lactate and the TCA/respiration changes. Overall, the above derived formula says that we can get into the concentrations of enzyme which is encoded by a gene only by measured expression values of that gene. This is either erroneous or sedative. Erroneous since maybe there are much hidden things which affect the expression values and we cannot get in them. Sedative since just by having just expression values we can get in to its enzymes concentration that it governs.
Chapter 3
System Genetics Approach for Analysing Rat Muscle Expression Data

3.1. Abstract

Ensembles of different methods (expression QTL mapping, weighted co-expression gene network, DNA motif detection and Bioinformatics database search) were used to get insight into rat’s hindlimb muscle gene expression data. Extra source of information was used to mediate eQTL toward system genetic analysis. Without generalizing the results, we identified following genes: M54987, M59742, X01454, M17523, X03639, U62779, M59786, X51707, U37026, D90219, M35862 out of 1184 genes to act likely as most connected genes (hub-acting genes) and to be associated with well-functioning of rat hindlimb muscle. Increasing muscle loading duration culminated in identifying the hub-acting gene M35862. Expression QTL mapping for candidate genes ended up with Trans-acting regulatory region, addressing complex regulation for aforementioned candidate genes. This study showed that differentially expressed genes are slightly likely to act as hub genes across dataset. Furthermore, it strongly indicated that different sources of information can be insightful in system genetics analysis or in post-era genomic paradigm, especially when the richness of genomic data is at premium.

3.2. Introduction

For finding differentially expressed genes, many statistical methods (gene ranking measures) have been developed (Cui, X. Q. et al., 2003; Motakis, E. S. et al., 2006; Wu, B. L. 2005). Some of these approaches were discussed in chapter 1. Irrespective of kind of gene ranking method, it is the set of differentially expressed genes or more varying genes which are generally used for learning different kinds of gene networks. Many gene networking algorithms have been developed to explore different biological questions in this area (Horvath, S. et al., 2006; Segal, E. et al., 2003; Zhang, B. et al., 2005). Some of these methods, like weighted co-expression gene network, were tailored toward finding
hub-acting genes (Horvath, S. et al., 2006; Zhang, B. et al., 2005). Finding hub-acting
genes (or set of genes which show more association with respect to other genes) can
provide insight into a biological motivated question. Some module-based gene network
learning methods (e.g. Zhang, B. et al., 2005), which gain from undirected graph theory,
allows using an extra source of information in learning gene co-expression network. This
extra source of information could be, individual health states, biologically-proved gene
functions, single nucleotide polymorphism (SNP) or sex of arrayed individuals. Also, this
method like many other methods (e.g. Segal, E. et al., 2003) permits to construct and
module learning method is a probabilistic predictive module learning method. For
example, they found that cell cycle module appears to regulate the histone module, and
this was confirmed experimentally. Therefore, module-oriented gene networking methods
have a crucial role in finding biological signal in high-throughput dataset. The skill in
system genetics (system biology + genetics) is how a researcher can integrate difference
sources of non-well related information in an insightful way. Expression QTL mapping
eQTL) can be seen as a successful effort in this way. Joint analysis of DNA marker
genotype and gene expression values was originally invented by Jansen, R. C. et al.,
(2001) who named it as Genetical Genomics. By mapping eQTLs, two kinds of eQTL are
distinguished: cis-acting eQTLs and trans-acting eQTLs. If the peak likelihood of eQTL
grid search falls on gene’s physical position itself, it is called cis-acting eQTLs. If this
peak falls in other part of genome, it is called trans-acting eQTLs. If this
peak falls in other part of genome, it is called trans-acting eQTLs. The trans-regulatory
eQTL is more likely to reveal a complex network of transcriptional regulation. Recently,
several studies have conducted genetical genomics experiments ranging from human
which covered in (de Koning, D. J. et al., 2005) to pigs (Kadarmideen, H. N. et al., 2007).
In general, having co-segregation of expression values and genetical markers are key in
running eQTL study. However, Kadarmideen, H. N. et al., (2006) successfully used a
cross species eQTL analysis to map the regulatory region. We argue that this kind of
analysis makes sense in gene expression data analysis by knowing that in many datasets
all credible sources of information for eQTL study can rarely be achieved. For example,
in our dataset, we had only transcriptomics information. Therefore, implementing direct
eQTL for our data was not immediately possible. By assuming some underlying gene
regulations process, which acts identically in same and across species, it is possible to use
genomics information from other datasets. This is why we resorted to the aforementioned
method. The study in this chapter is based on the three gene expression experiments
which discussed at chapter 2. In this study, we used a weighted co-expression gene network to find probable hub-acting genes for aforementioned data. Also, we will use eQTL mappings for interesting (candidate) genes.

3.3. Methods

3.3.1. Microarray Experiment

The experiments description for the gene expression dataset used here can be found in chapter 2.

3.3.2. Expression QTLs Mapping

Since we had not any co-segregated genetical marker, a simple query in related gene expression database in GeneNetwrok resources was done to find the most biologically related genes to our candidate genes. We used the MDC/CAS/ICL Kidney 230A (Apr05) MAS5 database for above the purpose (for more information about this population reader consult WebQtl site http://www.webqtl.org/). Using publicly available data on gene expression, SNP linkage maps and all the related software’s freely available at WebQTL server (www.genenetwork.org), we ran eQTL mapping to get insights into systems genetics of candidate genes. To detecting eQTLs, interval mapping was done to compute eQTLs either genome-wise and chromosome-wise. We used 1000 permutation and 1000 bootstrap tests to assess strength and consistency of linkage for expression values. Clustering of candidate genes was done (only for glutathione S-transferase theta 1 candidate gene) with 100 other genes, which were mostly correlated (based on Pearson correlation).

3.3.3. Gene Co-Expression Network Construction and Connectivity

In this study, we investigated weighted co-expression gene network (Zhang, B. et al., 2005) for rat hindlimb muscle gene expression data. We give a short explanation of the way that this method works mostly from graph theory point of view. Generally, when graph theory is applied in gene expression study, node or vertex stands for gene and edge indicates the relation among genes. Graph $G$ represents a pair consisting of a vertex set
A separating set (or vertex cut) of the graph $G$ is a set $S \subseteq V(G)$ in which $G - S$ has more than one component. For $G \neq K_n$ ($K_n$ is a complete graph on $n$ vertices) the connectivity of $G$ is $k(G) = \min \{|S|: S \text{ is a vertex cut} \}$. By definition, $k(K_n) = n - 1$, in which $n$ stand for number of vertices. The graph $G$ is $k$-connected if there is no vertex cut to size $k - 1$ (i.e $k(G) \geq k$) (More information can be found at Douglas, B. 2001). This general concept of connectivity would be helpful for understanding gene connectivity and also adjacency matrix. Let’s suppose that we have seven genes with given postulated structure as in Figure 3.1. The adjacency matrix of this structure is also shown. Note that Figure 3.1 indicates a directed graph which deviates from weighted co-expression network assumption, but it is done to make the concept of adjacency matrix and connectivity clear. In an adjacency matrix, the non- zero matrix represents existing edges. The number of edges to which a vertex is connected is called degree (K). The adjacency matrix of Figure 3.1 indicates that summing the columns and rows corresponds to the number of in-degree and out-degree edges of a gene incident, respectively.

![Figure 3.1 Postulated a seven-gene network with its adjacency matrix.](image)

For example the number of in-degree for gene 2 is 1, but out-degree of this gene is 2 (sum over in-degree and out-degree of a given gene is equal to its total degree which can be seen as connectivity, too). Biologically it can be postulated that gene 2 is somehow regulator of two other genes (3, 4) though in biology finding regulators gene(s) is a formidable task. Note that in making a adjacency matrix, the relationship of a gene with
itself is assumed to be 0. It is worthwhile mentioning that the adjacency matrix due to an
undirected graph is symmetric but for directed graph it is asymmetric. In weighted co-
expression gene network, adjacency matrix plays fundamental role. In gene weighted co-
expression gene networks, Pearson correlations in mRNA levels are normally used to
define connectivity and to group genes with similar expression profiles into modules,
thereby highlighting the higher-order properties of the transcriptome. To clarify, let’s
suppose that the absolute value of the Pearson correlation between two genes \((i,j)\) is
indicated like follows : \(\rho_{i,j} = |\text{cor}(i, j)|\). As shown, this measure of similarity among two
genes is handled in the following way: \(a_{ij} = \text{power}(\rho_{i,j}, \beta) = |\rho_{i,j}|^\beta\), which is a kernel
of weighted co-expression gene network (Zhang, B, et al., 2005). In terms of genes
adjacency matrix and connectivity, the above function can be written as: Connection
strength:=|correlation|^\beta. The main problem for this function is how to determine the
threshold parameter \(\beta\). Selecting higher and lower values for this parameter may lead to
either a dense or sparse graph (considering genes as node of a graph). This value must be
specified in such way that it would be biologically meaningful. Zhang, B, et al., (2005)
indicated that scale-free topology criterion can be used to select \(\beta\) in such way that it will
be biologically significant. The results of aforementioned adjacency function \((a_{ij})\) is used
as an input to make the so-called topological overlap matrix. Topological overlap of two
nodes (genes) reflects their relative interconnectivity. For a network represented by an
adjacency matrix \(A = [a_{ij}]\) which \(a_{ij} \in \{0,1\}\) the TOM is given by
\[
\omega_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}}
\]
where, \(l_{ij} = \sum_{u \neq i, j} a_{iu} a_{uj}\) denotes the number of nodes to which
both i and j are connected, and k is the number of connections of a node, with
\(k_i = \sum_{u \neq i} a_{iu} \) and \(k_j = \sum_{u \neq j} a_{ju} \). Since \(a_{ij} \in [0,1]\). Zhang, B. et al., (2005) found that
\[
l_{ij} = \sum_{u \neq i, j} a_{iu} a_{uj} \leq \min(\sum_{u \neq i} a_{iu}, \sum_{u \neq j} a_{ju}) = \min(\sum_{u \neq i} a_{iu}, \sum_{u \neq j} a_{ju}) - a_{ij}. \]
Thus, \(l_{ij} + a_{ij} \leq \min(\sum_{u \neq i} a_{iu}, \sum_{u \neq j} a_{ju})\). Since \(1 - a_{ij} \geq 0\), they found that \(\omega_{ij}\) is a number between 0 and 1. The 2 reasons for
adding \(1 - a_{ij}\) to the denominator in the topological overlap matrix can be found in (Zhang,
B. et al., 2005). Topological overlap matrix is symmetric, i.e, \(\omega_{ij} = \omega_{ji}\) and its diagonal
elements are set to 0 (i.e. \(\omega_{ii} = 0\)). The rationale for considering this similarity measure is
that nodes that are part of highly integrated modules are expected to have high topological
overlap with their neighbors. In this study, above order of topological overlap matrix is used as an input (dissimilarity measure or distance function) to the average hierarchical linkage algorithm (unsupervised) to end up with modules. The m-th order topological overlap measure allows one to trade-off sensitivity versus specificity when it comes to defining pair-wise interconnectedness and network modules (Zhang, B. et al., 2005). We used R codes snippet which is provided (Zhang, B. et al., 2005) for this task.

3.4. Results and Discussion

Because a cell always uses some small fraction of its genes to respond to the changes which it involves, we would expect quite a small number of genes to show a differential expression pattern. Comparison between experiments can reveal genes that respond to different experiments. It was tried to understand which genes have been differentially expressed across whole experiments. Glutathione-S-transferase (J02592) turned out to be single gene which shows this behavior. In general, with a differentially expressed approach one can just address that a set of genes for given contrast show differentially expressed pattern and we can not address anything in terms of regulation program. Due to this reason, Glutathione-S-transferase (J02592) and five more extra genes (as candidate genes) were used to get a clue thorough its regulation program using expression QTLs mapping. The following are the results.

3.4.1. Expression QTL Mapping Task

We used a rat population to study Glutathione-S-transferase (Mu 2) gene via Glutathione S-transferase theta 1 with the assumption that genetic networks and systems genetics/biology are conserved between members of the same species. This approach has indeed been applied in the form of comparative systems genetics (Kadarmideen, H. N. et al., 2007). Glutathione S-transferase theta 1 gene is located in chromosome 20 of rat genome. Our effort has led to the identification of two hotspot trans-acting eQTLs in chromosome 3 and 10, based on the interval mapping analysis of the dense rat marker genetic map, which was available in webQTLs. Figure 3. 2 shows the results for chromosome 10. This figure shows that the aforementioned gene is controlled by others genes (Trans-acting eQTL). Figure 3. 3, Figure 3. 4 and Table 3. 1 supply more information. This information pinpoints to the complexity which this gene is involved. In
a regression analysis, we have identified 14 loci in chromosome 3 and 10 that had significant main effects on expression values of Glutathione S-transferase theta 1. Many of these 14 loci, however, were identified primarily, or solely, from strong epistatic interactions detected among specific alleles at different genetic loci (right panel of Figure 3.5). Indeed, a fundamental conclusion to be drawn from these data is that Glutathione S-transferase theta 1 expression values is significantly affected by complex genetic interactions at multiple loci. The genome-wide interaction analysis identified (Figure 3.5) numerous pairs of loci that, in particular allelic combinations, had epistatic effects on Glutathione S-transferase theta 1 expression values much greater than either of the loci alone. (Additional higher order interactions may also be present; however, our current sample size is not large enough to reliably detect these effects.) This epistasis is most clearly seen in the analysis of the amplitude of the Glutathione S-transferase theta 1, in which particular alleles of the loci, each had relatively little effect on the Glutathione S-transferase theta 1 expression individually, but in specific allelic combination produced a significant effect.

Figure 3.2 Whole genome genetic mapping and bootstrap tests of Glutathione S-transferase theta 1 over chromosome 10.
Figure 3.3 Physical map of chromosome 10 between 69 to 73M base pairs.

A better understanding of such epistatic interactions will be essential to fully understand the complex nature of Glutathione S-transferase theta 1.

Figure 3.4 Gene ontology enrichment of glutathione S-transferase theta 1.

Results such as these underscore the complexity of regulation of this gene and argue against any approach that devotes all analysis of Glutathione S-transferase theta 1 to single gene effects. This genetic analysis revealed previously undetected complexity in the Glutathione S-transferase theta 1 regulations and points to the presence of many as yet undiscovered genes that contribute to regulate Glutathione S-transferase theta 1 in rat.
### Genes in cofactor metabolism based on gene ontology enrichment

<table>
<thead>
<tr>
<th>Entrez Gene ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>361071</td>
<td>succinate-Coenzyme A ligase, ADP-forming, beta subunit (predicted)</td>
</tr>
<tr>
<td>29554</td>
<td>pyruvate dehydrogenase E1 alpha 1</td>
</tr>
<tr>
<td>309004</td>
<td></td>
</tr>
<tr>
<td>362404</td>
<td>succinate-Coenzyme A ligase, GDP-forming, beta subunit</td>
</tr>
<tr>
<td>25260</td>
<td>glutathione S-transferase theta 1</td>
</tr>
</tbody>
</table>

### Genes in main pathways of carbohydrate metabolism based on gene ontology enrichment

<table>
<thead>
<tr>
<th>Entrez Gene ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>361071</td>
<td>succinate-Coenzyme A ligase, ADP-forming, beta subunit (predicted)</td>
</tr>
<tr>
<td>29554</td>
<td>pyruvate dehydrogenase E1 alpha 1</td>
</tr>
<tr>
<td>362404</td>
<td>succinate-Coenzyme A ligase, GDP-forming, beta subunit</td>
</tr>
<tr>
<td>63938</td>
<td>3-hydroxyisobutyrate dehydrogenase</td>
</tr>
</tbody>
</table>

### Genes in polyphosphate-glucose phosphotransferase activity based on gene ontology enrichment

<table>
<thead>
<tr>
<th>Entrez Gene ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>83617</td>
<td>homeodomain interacting protein kinase 3</td>
</tr>
<tr>
<td>Z25579</td>
<td>mitogen activated protein kinase kinase 12</td>
</tr>
<tr>
<td>286993</td>
<td>TAO kinase 1</td>
</tr>
<tr>
<td>60450</td>
<td>megakaryocyte-associated tyrosine kinase</td>
</tr>
</tbody>
</table>

Table 3.1 Gene ontology results, the candidate gene is bold highlighted.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Locus</th>
<th>Additive Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>D3Rat180</td>
<td>-0.074</td>
</tr>
<tr>
<td>3</td>
<td>D3Cebr2s4</td>
<td>-0.076</td>
</tr>
<tr>
<td>3</td>
<td>Cat</td>
<td>-0.076</td>
</tr>
<tr>
<td>10</td>
<td>D10Cebr44s3</td>
<td>0.080</td>
</tr>
<tr>
<td>10</td>
<td>D10Rat102</td>
<td>0.080</td>
</tr>
<tr>
<td>10</td>
<td>D10Mit2</td>
<td>0.082</td>
</tr>
<tr>
<td>10</td>
<td>D10Rat29</td>
<td>0.081</td>
</tr>
<tr>
<td>10</td>
<td>D10Rat70</td>
<td>0.086</td>
</tr>
<tr>
<td>10</td>
<td>D10Rat28</td>
<td>0.086</td>
</tr>
<tr>
<td>10</td>
<td>D10Cebrp1016s5</td>
<td>0.090</td>
</tr>
<tr>
<td>10</td>
<td>D10Utr11</td>
<td>0.089</td>
</tr>
<tr>
<td>10</td>
<td>D10Rat26</td>
<td>0.080</td>
</tr>
<tr>
<td>10</td>
<td>Slc4a3</td>
<td>0.079</td>
</tr>
<tr>
<td>19</td>
<td>Agt</td>
<td>-0.076</td>
</tr>
</tbody>
</table>

Table 3.2 The following loci in the HXBBXH data set have associations with the above Glutathione S-transferase theta 1 data.

Although severe disruption in regulation of Glutathione S-transferase theta 1 may be caused by mutations in hotspot genes, it is possible that the broad variety of Glutathione
S-transferase theta 1 regulation observed in rat is the result of polymorphisms in multiple interacting loci such as the ones described here. We didn’t check the direction effects of eQTLs in hotspots. In other words, the proportions of genes with positive and negative eQTL effects at each hotspot weren’t contrasted using suitable methods like a chi-square test.

The result of the study can pinpoint to some difficulties of gene expression profile using rat brain to get insight into role of glutathione-S-transferases classes in alcohol preferences studies (Saba, L. et al., 2006). Bjork, K. et al., (2006) showed that glutathione-S-transferases of the lpha (Gsta4) transcript levels are increased in several brain regions of AA rats compared with the ANA rats. They found a correlation between haplotype variant and gene expression of the Gsta4 gene. The opposite correlation between alcohol drinking phenotype and Gsta4 transcript levels in iP/iNP and AA/ANA rats, respectively, does not support a simple relationship between Gsta4 expression and alcohol preference. However, low GST expression, reduced life span, and memory performance in ANA rats may be a useful model of brain aging and neurodegeneration. Given the fact that the majority of our identified gene (result of regression and eQTL mapping for loading reloading experiment) seems to be trans-regulated, we have not finished the search for common transcriptional regulatory mechanisms for this group of genes. On the other hand, validations of cis-regulated genes are not trivial task. Doss, S. et al., (2005) argued that of the 28 predicted cis-acting eQTL in their study, 18 confirmed
cis-regulation as shown by the cis–trans test. Of the remaining 10 genes, seven tested as trans-acting according to the above definition. There could be several possible explanations for the failure to confirm the regulations program of remaining genes (8) in their work. Although the power of the genetical genomic analyses described extends beyond gene discovery, including revealing genetic interactions among different loci and providing a landscape of the genes underlying a given complex gene regulations, identification of the trans-acting genes themselves is still of great interest. One approach towards this end would be to develop congenic rat lines carrying several of the trans-acting eQTLs founded in this study here. From this congenic rat, we can determine if each locus has main regulations effects in isolation (or effects in interacting pairs) and if so, begin to isolate the genetic interval responsible for this regulation effect, with the eventual target being the identification and testing of candidate genes that might be responsible for regulation phenomenon. One other thing which can be justified by having such congenics rat lines is to detect the role of miniscule environment effect in gene regulations. Hopefully public availability of rat genomic sequence (from several intensive strains) in the near future promises this prospect considerably. The variation in transcript level can be determined by variation among alleles, that is, some property of the functional gene itself (cis-acting eQTL). Alternatively, variation may occur in a factor contributing to transcription rate, specificity, or stability of the transcript (trans-acting). In most cases, these alternatives may be distinguished by genetic or physical mapping of the transcribed gene. In the case of cis-regulation, the genomic location of the eQTL will be mapped with the gene. If variation is trans-regulated, the site controlling the variation may be located elsewhere. The strength of inference in distinguishing these alternatives lies in concluding that the locations of the gene and the eQTL are significantly different. However, mapping of eQTL to non-homologous chromosomes in different genetic backgrounds suggests that variation in the regulation of transcript abundance occurs predominantly in trans-acting loci. The well annotated complete genome sequence of rat is now available and therefore most likely inferences can be made about whether the eQTL identified correspond to the physical location of the gene (cis-regulation) or to the genetic loci of its trans-regulator. However, if most of the transcript-level regulation occurred in cis-, a high level of homology among the genetic location of eQTL would have been expected. In yeast, only 25% of the genes for which eQTL were identified co-localized to the position of the eQTL (Brem, R. B. et al., 2002). The left panel Figure 3. 5 shows clustering between the aforementioned gene with more than 100 other genes in
database which had high level of Pearson correlation with it. Those can help to unravel more extra regulatory work. As literatures address, Glutathione-S-transferase class are interesting genes in aging and alcohol-preferences studies. Result of this study can help researcher to focus on the aforementioned chromosomes if they are looking for regulatory element of Glutathione-S-transferase class. An extra study was done with the aforementioned method of eQTLs mapping for five more candidate genes (Figure 3. 10 through Figure 3. 14). Using above arguments, it can be seen that regulation program of these genes are in trans-acting phase and as a result, it is complex. This indicates that for more extra understanding of regulating program of these genes, more work like metabolic QTLs is required. However, eQTL analysis certainly has some restrictions. One of the main problems is due to sample size. The number of eQTL (irrespective of cis or trans-acting eQTLs mode) is likely to be downwardly biased estimated due to linkage and limited sample size. There is also likely a bias in the estimation of eQTL effect distribution as only eQTL with relatively large effects are likely to be detected and some QTL effects may represent the joint effects of multiple closely linked genes. The location of eQTL is rarely precise, and the precision of genetic map or physical location, although typically greater, may also have a significant error of estimation. In such situations, on the basis of the limits of an eQTL interval, two eQTL may appear to overlap, although this could be due to effects of different genes. Also to make more restrictions of this method, eQTL studies involve very high-dimensional computations; controlling family-wise error is not trivial and permutation analysis is too time consuming. Also biologically, it is interesting to ask if the transcripts are cis-regulated or trans-regulated and current methods to test whether a transcript is cis- and trans- regulated is ad-hoc. All methods described so far are transcript-based. A model which incorporates all transcripts and metabolic simultaneously is expected to be more powerful. One of the main problems is applying the identical cut off values (thresholds) for both cis-acting eQTLs and trans-acting eQTLs, though by now, no statistical method has been introduced to handle this dilemma. Analysis of epistasis may only detect a part of gene interactions and there could be many other hidden interactions between detected and undetected eQTLs. Some more issues in this regards have been addressed in the following works (Kendziorski, C. et al., 2006; Wei, et al., 2007). Bayesian analysis is attractive in eQTL analysis since we can put priors on the cis- and trans-regulations and calculate their posterior probabilities, which is recommended to be used in this context.
3.4.2. Results of Gene networking Task

In finding differentially expressed genes at 20 % of pFDR or q-value, 233 genes showed differentially expressed pattern. This set of genes was used as an extra source of information for learning weighted co-expression gene network. However, the biological goals of the analysis were a) to identify gene modules and b) to relate network connectivity to differentially expressed genes coded as 1 if the gene was differentially expressed and 0 otherwise. Use of unsupervised clustering (i.e. lack of any priori knowledge about data) based on co-expression measure, gave arise to modules enriched for biologically important processes and most likely illuminates that these modules are a robust feature of the molecular architecture in rate muscle loading reloading data. Topological overlap can serve as an important filter to counter the effects of spurious or missing connections between network nodes. As it was indicated, topological overlap matrix has a different order (Zhang, B. et al., 2005). Figure 3. 6 shows the general way of learning weighted co-expression gene network. The left panel shows a threshold to pick up β value due to free topology measure. Also this figure indicates genes that have been clustered into 6 modules. We selected β =6 which indicates threshold or cutoff value. The results of weighted co-expression gene network analyses are highly robust with respect to the choice of the parameter β. Figure 3. 7 shows more information about these estimated modules. Also, the right panel of Figure 3. 7 shows enriched differentially expressed genes in different modules defined as the branches of the dendrograms.

Figure 3. 6 Plot the scale free topology and mean gene connectivity versus.
Figure 3.7 Scatter plots of arrays along the module eigengenes, classical three-dimensional scaling plots for and enrichment of differentially expressed genes.

For example, right panel of Figure 3.6 shows the dendrograms of our network applications. Genes of proper modules are assigned a color (e.g. turquoise, blue etc). The module definition depends on how the branches cut off the dendrogram. For more information about how to pick up the right cut off or threshold, reader referred to (Zhang, B. et al., 2005). Of the 1184 genes, 608 were clustered into 5 proper modules, and the remaining genes were colored in grey. Some of proper modules were significantly enriched for genes with the following specific gene ontology classes: yellow (73 genes, non-enriched); turquoise (344 genes, enriched in large scale in signal transducer activity); green (68 genes, non-enriched); brown (90 genes, enriched in physiological process/protein signaling pathway; blue (108 genes, enriched in plasma membrane/ATP binding). These results were obtained using EASE software which can be found at: http://david.niaid.nih.gov/david/ease1.htm. To study the relationship between connectivity and differentially expressed genes (coded as 0 and 1), we focused on the turquoise module since it was enriched (relatively) with differentially expressed gene. Other proper modules were enriched to some extent with differentially expressed genes as well. In the turquoise module, intra-modular connectivity and differentially expressed genes was not highly correlated in the dataset. The overall connectivity for each gene is the sum of the connection strengths (|correlation|^β) between that gene and all other 1184 genes in the network. The intra-modular connectivity was calculated for each gene as the sum of the connection strengths between that gene and all other genes in its module. It is being suggested that highly connected “hub-acting” genes are thought to play crucial role in organizing the behavior of biological modules (Horvath, S. et al., 2006). As it is obvious, in identified modules turquoise is highly enriched with differentially expressed
genes. Therefore, we set out to identify the turquoise hub-acting genes. We hypothesized that intra-modular hub-acting genes, may be associated with well-functioning of rat muscle. We obtained hub-acting genes not only for whole modules networks, but also for turquoise module, too. Estimated hub-acting genes in this study were: M54987; M59742; X01454, M17523; X03639; U62779; M59786; X51707; U37026; D90219; M35862. The last seven of them were clustered together in turquoise module and the rest were distributed across other modules. The estimated list of module genes may be further screened and nailed down based on (i) biological plausibility based on external gene (ontology) information, (ii) the availability of protein biomarkers and metabolite signatures for further validation, (iii) the availability of comparable genomics models, like the mouse, for further validation, and/or, (iv) the gravity- based on studies in hard-data measuring scenario like human studies, i.e., the opportunity for multiple genome comparisons. It is worthwhile to say that in learning co-expression gene network differentially expressed genes were not cohesively clustered together. In other words, different modules have been enriched with different fractions of the differentially expressed genes. We can interpret this matter as follows. Aforementioned experiments in chapter 2, likely affected many different metabolic pathways. If they would have affected a small number of metabolic pathways, differentially expressed genes should likely be lumped over in one or two modules. Such thing did not happen here. As we can see, differentially expressed genes were scattered and were enriched across whole genome (modules). This implies one of hidden potentials of co-expression gene network in addressing the number of metabolic networks which are affected by an experiment or any other perturbation effects. Grey module, which is a non-proper module by definition, is sharply enriched with differentially expressed genes. This again implies more difficulty in correctly addressing the number of metabolic pathways which affected by aforementioned experiments, since grey module contains genes which were not modulated with other genes in one of proper modules.
To correlate individual gene expression profiles with the entire module, the expression profile of the module genes was summarized by the first module eigen-gene, which is defined by using the singular value decomposition of the expression data (Results not shown here). In comparing with list of highly differentiated genes in third experiment, the estimated hub-acting genes were not categorized as highly differentiated genes. It is worth to be mentioned that most of estimated hub-acting genes in this study belonged to the third experiment. Figure 3.8 includes different plots regarding some different identified modules (i.e connectivity of modules network against mean and variance of expression values). Also the way that cut off values is specified is shown. The clustering coefficient and connectivity measure showed high correlation in turquoise across identified modules. Also no indicated estimation was obtained for Yellow module.

3.4.2. Motif Search for Rat Muscle Hub-Acting Genes

We put forward this hypothesis that estimated hub-acting genes in this study related in terms of gene regulations to an over-represented DNA sequence(s)/ transcriptional factor binding sites (TFBSs) or not. To check in this matter, we first identified promoter DNA sequences of hub-acting genes using the following link (http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=searchPromForm). The marked promoter sequence size was from -700 to 299 relative to transcription start site. Later on, we used the following tool (http://cosmoweb.berkeley.edu/) to estimate TFBSs or motif. Table 3.3 indicates some general information about the extracted DNA sequence promoter of some of hub-acting genes (we were not able to extract DNA sequence promoter for whole hub-acting genes out of relevant databases i.e., Genbank). Table 3.4 and Figure 3.8 show the results due to motif discovering task (estimated PSSM and motif log respectively). The results
indicate the investigation of hub-acing genes of at least in one TF share the same regulation program. We used passive information (information which did not immediately exist in our dataset) to govern our estimated hub-acting genes. Conditional on the assumptions made in this study, motif searching task, revealed a conceal gene regulation pattern of hub-acting genes. Table 3. 3 has two columns regarding to gene names, of which the leftmost are the gene names which were available in our dataset and the two columns right after it, are the corresponding gene names which turned out during databases searching. Both names are given to clarify this matter.
<table>
<thead>
<tr>
<th>Gene IDs in our Dataset</th>
<th>Gene IDs</th>
<th>Gene Name</th>
<th>Species</th>
<th>Chromosome Location</th>
<th>Strand</th>
<th>Promoter ID</th>
<th>Quality</th>
<th>Transcription Start Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>M59742</td>
<td>80150</td>
<td>Gabt1</td>
<td>rat, Rattus norvegicus</td>
<td>4q42</td>
<td>+</td>
<td>103351</td>
<td>3.1: refseq,predicted</td>
<td>147028783</td>
</tr>
<tr>
<td></td>
<td>80150</td>
<td>Gabt1</td>
<td>rat, Rattus norvegicus</td>
<td>4q42</td>
<td>+</td>
<td>103352</td>
<td>3.2: refseq</td>
<td>147031830</td>
</tr>
<tr>
<td></td>
<td>76374</td>
<td>Tshb</td>
<td>rat, Rattus norvegicus</td>
<td>2q34</td>
<td>-</td>
<td>99332</td>
<td>1: known,curated</td>
<td>191367556</td>
</tr>
<tr>
<td></td>
<td>76374</td>
<td>Tshb</td>
<td>rat, Rattus norvegicus</td>
<td>2q34</td>
<td>-</td>
<td>99331</td>
<td>2: known</td>
<td>191367370</td>
</tr>
<tr>
<td></td>
<td>76374</td>
<td>Tshb</td>
<td>rat, Rattus norvegicus</td>
<td>2q34</td>
<td>-</td>
<td>99330</td>
<td>2: known</td>
<td>191367413</td>
</tr>
<tr>
<td></td>
<td>76374</td>
<td>Tshb</td>
<td>rat, Rattus norvegicus</td>
<td>2q34</td>
<td>-</td>
<td>137088</td>
<td>3.2: refseq</td>
<td>191362537</td>
</tr>
<tr>
<td>X01454</td>
<td>77938</td>
<td>Scn2a1</td>
<td>rat, Rattus norvegicus</td>
<td>3q24</td>
<td>+</td>
<td>100989</td>
<td>3.2: refseq</td>
<td>448402000</td>
</tr>
<tr>
<td></td>
<td>67449</td>
<td>n/a</td>
<td>rat, Rattus norvegicus</td>
<td></td>
<td>-</td>
<td>89915</td>
<td>1: known,curated</td>
<td>90993794</td>
</tr>
<tr>
<td></td>
<td>67448</td>
<td>n/a</td>
<td>rat, Rattus norvegicus</td>
<td></td>
<td>-</td>
<td>89914</td>
<td>2: known</td>
<td>90993398</td>
</tr>
<tr>
<td></td>
<td>67447</td>
<td>n/a</td>
<td>rat, Rattus norvegicus</td>
<td></td>
<td>-</td>
<td>89913</td>
<td>2: known</td>
<td>90993400</td>
</tr>
<tr>
<td></td>
<td>67450</td>
<td>n/a</td>
<td>rat, Rattus norvegicus</td>
<td></td>
<td>-</td>
<td>89916</td>
<td>4: withRNA</td>
<td>91001519</td>
</tr>
<tr>
<td></td>
<td>80551</td>
<td>Cacna1c</td>
<td>rat, Rattus norvegicus</td>
<td></td>
<td>-</td>
<td>103782</td>
<td>2: known</td>
<td>152283545</td>
</tr>
<tr>
<td></td>
<td>80551</td>
<td>Cacna1c</td>
<td>rat, Rattus norvegicus</td>
<td></td>
<td>-</td>
<td>103783</td>
<td>4: withRNA</td>
<td>152294555</td>
</tr>
<tr>
<td>U62779</td>
<td>85859</td>
<td>Scn2b</td>
<td>rat, Rattus norvegicus</td>
<td>8q22</td>
<td>+</td>
<td>109443</td>
<td>3.2: refseq</td>
<td>46333827</td>
</tr>
<tr>
<td></td>
<td>87720</td>
<td>Nppe</td>
<td>rat, Rattus norvegicus</td>
<td>9q35</td>
<td>-</td>
<td>111432</td>
<td>3.1: refseq,predicted</td>
<td>85094253</td>
</tr>
<tr>
<td></td>
<td>87720</td>
<td>Nppe</td>
<td>rat, Rattus norvegicus</td>
<td>9q35</td>
<td>-</td>
<td>138551</td>
<td>3.2: refseq</td>
<td>85090053</td>
</tr>
<tr>
<td></td>
<td>73372</td>
<td>Mak</td>
<td>rat, Rattus norvegicus</td>
<td>17p12</td>
<td>+</td>
<td>96171</td>
<td>3.2: refseq</td>
<td>29519593</td>
</tr>
</tbody>
</table>

Table 3.3 Summary of information due to extracting DNA promoter sequence of some hub-acting genes.
We argued that DNA motif sequence analysis could likely support estimated hub-acting gene(s). Our explanation is that estimated hub-acting genes are genes which are involved in metabolic process; therefore, likely they should be regulated gene(s) rather than regulator gene(s). If it is, likely in upstream region of these gene(s), TFBSs must be located, in which, a TF binds to. The results of motif search supported the idea since the Position Specific Scoring Matrix (PSSM) due to estimated TFBSs of hub-acting genes is obtained with substantial conserved information in bit base (log2). That means, there is likely a common over-represented pattern (TFBS) or motif across whole hub-acting genes. However, we didn’t check which genes are responsible for regulating hub-acting gene(s) in our dataset, but we could say that, for a more in-depth of understanding the way that loading/reloading experiments affect transcriptomic level of rat hindlimb muscle, estimated hub-acting genes deserve to be explored more. To identify a potential novel gene target, we looked for the most highly connected gene(s) that have not been extensively studied in rat dataset. This led us to study the M35862 (Rat male germ cell-associated kinase (mak) mRNA, complete cds) gene, because it had the highest gene connectivity value in our dataset. Results of chapter 2 in this thesis, also shown that M35862 has been down-regulated just in experiment three. Therefore, we proposed that

Table 3. 4 Estimated PSSM due to hub-acting genes.

<table>
<thead>
<tr>
<th>Base</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>0.78</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.87</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>1</td>
<td>0.11</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.13</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3. 9 Sequence logo of discovered motif (consensus sequence).
M35862 is likely functioning in muscle cell when longstanding muscle loading, re-loading experiment is carried out. In other words, since M35862 was affected from contrast within third experiment relative to other contrasts which were made for other experiments, we assumed that it could present a compelling target for more study in rat muscle study especially when long-time study is involve. Mutation within this gene can likely be associated with many rat muscle problems, since its gene ontology classes indicated that this gene is involved in many cell functions (i.e. protein metabolism; macromolecule biosynthesis). There was no extra data to see whether highly connected hub-acting genes of the rat muscle dataset in turquoise module tend to be highly connected for the same genes in other dataset or not. We argue that having extra data of the same genes can help to see whether these modules are reproducible or not. If they are, it can be said that gene co-expression modules are highly preserved in across different dataset for the same genes. Also, to see whether these modules are detectable in another dataset or not, the dataset must be efficiently and sufficiently large and contain gene expression data from different microarray platforms, allowing for array platform independent conclusions. This reveals that identified modules are highly preserved or not. If module preservation is achieved, this can suggest that the module may be involved in biological processes that are shared across many experiment perturbations. Also, examining the expression pattern of identified module genes across large samples including rat muscle dataset and other tissues is useful. If interested in module - in our case turquoise - genes go up or down together across a wide range of tissue types, wide-range tissue comparisons in terms of modules are possible. It is indicated that weighted co-expression gene network can alleviates the multiple testing problem inherent in microarray data analysis (Horvath, S. et al., 2006). Instead of relating thousands of genes for a given experiment, it focuses on the relationship between a few modules (in this study, there were five proper modules) and the experiment. Because the modules may correspond to biological pathways, focusing the analysis on module eigengenes (and equivalently intramodular hub-acting genes) amounts to a biologically motivated data reduction scheme. This study showed that that different source of information can be used to get insight into biological question. It was indicated that multi-biological tasking hub-acting gene M35862 expressed only in third experiment, likely is the gene which must be studied for more interested in biological questions. Results of motif detection indicated that majority of hub-acting genes share in common the same TFBS, that is, they are likely under the same regulations program. Results of eQTLs mapping provided complex scenario for candidate genes. We propose
an that extra study using combination of Genetical genomics and weighted co-expression gene network approaches can provide better insight into estimated hub-acting genes. This can provide much more support for weighted-co-expression analysis. In this way, weighted co-expression gene network can be thought of as biologically motivated gene mining approach, in which we can narrow down the dense loci accumulated in trans-acting regulatory region. As a result, compelling regulatory network can be likely deduced.

### 3.6. Appendix of Extra eQTLs for Candidate Genes

For 5 more extra candidate genes, eQTL was done. For some genes in our analysis which we didn’t find them at Webqtl database, most related genes were queried in aforementioned database.

![Figure 3. 10 Whole genome genetic mapping and bootstrap tests of Cytochrom c oxidase subunit VIII-H gene (up) over chromosome 14.](image-url)
Figure 3.11 Whole genome genetic mapping and bootstrap of solute carrier family 4, member 3 gene over chromosome 20.

Figure 3.12 Whole genome genetic mapping and bootstrap tests of carbonic anhydrase 3 gene (up) over chromosome 15.
Figure 3.13 Whole genome genetic mapping and bootstrap tests of ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 1 gene over chromosome 1.

Figure 3.14 Whole genome genetic mapping and bootstrap tests of fatty acid binding protein 3 gene for fatty acid binding protein 3 gene over chromosome 8.
Chapter 4
Large Scale Estimation of Additive and Maternal Genetic Components for Exterior and Reproductive Traits in Pig Population

4.1. Abstract

The main aim of this study was to estimate additive and maternal genetic components, genetic correlations across breeds and to make comparison of profile of breeds in terms of estimated genetic components. Six exterior traits (teat and leg) and 4 reproductive traits within three different Swiss pig breeds were analyzed. The data were fitted to 4 separate models incorporating different random terms using restricted maximum likelihood methods. By fitting a genetic covariance among additive and maternal genetic effects, estimated additive heritability increased. Comparison of models within traits and breeds based on log-likelihood ratio test indicated the maternal genetic effect in exterior traits was higher than in reproductive traits. The genetic correlations among teat traits were positive and substantial across breeds (>90%). Low level genetic correlation between reproductive traits and exterior traits were observed. The results indicated that including maternal effect in breeding programs is not recommended for teat traits. Also direct genetic improvement of reproductive traits in pigs will be slower than exteriors traits.

4.2. Introduction

Success of any breeding program depends on how many different economically important traits are considered in an optimal manner so as to improve efficiency of production. Studies addressing pigs breeding infrastructure have mainly focused on the genetic gain of production traits, such as growth rate, meat percent, feed efficiency, and piglet production, because these traits are economically important (Olesen, I. et al., 2000). However, from genetic and economical point of view, these can be considered as rather “short-term goal” traits. Pork production has moved on from these ‘short term’ goals to ‘long terms goals’ to consider health, exterior and reproductive performance of animals. Historically, health, reproduction and meat quality characteristics were not considered as feasible breeding objectives due to low heritabilities, difficulty of measurement, lack of
trait definition and lack of economic importance, mainly it would take many years or generations to improve their performance genetically. But this trend is changing as measurement as well as genetic / genomic techniques have dramatically improved in the last few years. Secure breeding programs do need to consider short-term and long-term goals simultaneously and in doing so, genetic components of these short and long goals trait must be calculated. In most studies, reproductive traits and exterior traits have abbreviated into one word: longevity (Torres, R. D. A. et al., 2005; Serenius, T. et al., 2005). It means better performance of these traits leads to better longevity. Longevity is a function of long-term and short term goal traits which affect the efficiency of production system. Estimating different additive and maternal genetic components of these traits is important for genetics evaluations programs. Johnson, Z. B. et al., (2002) showed that maternal effects maybe important for some traits in some breeds and should be included in genetic evaluation systems to provide unbiased estimators of direct breeding values. Robison, O. W. (1972) reported that there is substantial evidence that maternal effects account for a significant portion of the variance for most traits in swine, including those that are manifested relatively late in life. Improvement of maternal response in addition to direct response can lead to greater overall response (Roehe, R. et al., 1993). However maternal genetic effect is genetically neglected for long-term goals traits (exterior and reproductive trait) which are usually considered as traits with low direct additive heritability. Maternal effects have been published for some traits in swine (Johnson, Z. B. et al., 2002) but only a very few studies have reported these effects for exterior and reproductive traits. Understanding the maternal influence on these traits could be crucial for unbiased genetic parameters. Genetic correlation among exterior and reproductive traits across breeds has not been comprehensively reported. Selection of the best sires and dams is desirable in order to increase progeny performance for exterior and reproductive traits and to improve the maternal performance of females. The objective of this work was to estimate direct additive and maternal genetic heritability of a range of reproductive and exterior traits in three Swiss pig breeds comparing 4 separate models and to estimated genetic correlations among traits. Genetic correlation can show the way that two traits behave and as a result the underlying genes which are responsible for penetrance of a trait(s) can be pinpointed. In this study to make better picture of traits, genetic correlation analysis was done within breeds to see the problem of genetic performance of each breed in terms of their measured traits. Advanced statistical models are used and different
models are compared with respect to their goodness of fit for analysing these traits can be seen at the Table 4.1. This table contains some information across number of literatures.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Estimated Range</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fore legs</td>
<td>.04-.32</td>
<td>.18</td>
</tr>
<tr>
<td>Fore Legs front view</td>
<td>.06-.47</td>
<td>.27</td>
</tr>
<tr>
<td>Fore leg bones</td>
<td>.06-.47</td>
<td>.27</td>
</tr>
<tr>
<td>Fore leg pasterns</td>
<td>.31-.48</td>
<td>.40</td>
</tr>
<tr>
<td>Fore leg claws</td>
<td>.04-.21</td>
<td>.13</td>
</tr>
<tr>
<td>Hind Legs</td>
<td>.04-.21</td>
<td>.13</td>
</tr>
<tr>
<td>Hind leg rear view</td>
<td>.06-.47</td>
<td>.27</td>
</tr>
<tr>
<td>Hind leg hocks</td>
<td>.01-.23</td>
<td>.12</td>
</tr>
<tr>
<td>Hind leg pasterns</td>
<td>.07-.3</td>
<td>.19</td>
</tr>
<tr>
<td>Hind leg claws</td>
<td>.09-.13</td>
<td>.16</td>
</tr>
<tr>
<td>Back</td>
<td>.15-.22</td>
<td>.19</td>
</tr>
<tr>
<td>Locomotion</td>
<td>.08-.13</td>
<td>.11</td>
</tr>
<tr>
<td>Number of teat</td>
<td>.07-.42</td>
<td>.30</td>
</tr>
</tbody>
</table>

Table 4.1 Estimated direct additive heritability for different exterior traits in literatures. (Pramod K. Mathur, 2002).

### 4.3. Material and Methods

#### 4.3.1. Breeding Program and Station-Tested Traits

SUISAG is a part of Swiss Federation of Pig Breeders and Producers. SUISAG is a herd book breeding organization and conducts a national breeding program. In the breeding program, Swiss Large White (SLW) are used as sire line (SLW_{SL}) and dam line and Swiss Landrace (SLR) as sire line in the nucleus herds. In the multiplier herds the two (cross-bred) lines are crossed. The F1-sow is used in production herds. Swiss Large White (SLW) and Duroc are bred in nucleus herds and used as terminal sires in production herds. This breeding program with specialized sire and dam lines has been defined since 2000. Performance testing is for production, reproduction and type traits. Station tests are conducted for about 3000 animals per year, via sib / progeny testing. Animals are fed in pen feeders (group size of 10).
4.3.2. Data

Table 4.2 lists descriptive statistics of the performance records for traits evaluated in this study. Edits consisted of checking for missing values and descriptive tendency statistics for trait normality (it was only done for some traits which showed Gaussian-like distribution). Traits were centralized to have N (0,1) distribution before analysis. Records with missing values were eliminated in this study and this was done for both exterior and reproductive traits. In Table 4.2, traits with star sign indicate reproductive traits. The measured traits were collected in (SLW), (SLR) and (SLW DL) breeds. The number of observations within breeds was different. SLW contained 10614 observations and 1018 dam, SLR contained 1645 observations and 222 dam, and SLW SL contained 842 observations and 232 dams respectively. The pedigree contained the following information: 31267 animals, 3388 sires and 13582 dams. There were 1048 animals as base animal for which date of birth were not cited in original pedigree file. The birth date of individuals in pedigree scratched from 1990 until 2004. The main pedigree was quite large. We got seven generations back into main pedigree to come up with computationally feasible pedigree, yet preserving the required information. In doing so, we tried to create a pedigree in such way that half-sib families would have least enough progenies. In the process of dataset edging, removing missing values and putting reproductive and exterior records together for the same animal, reduced the dimension of dataset.
<table>
<thead>
<tr>
<th>Traits</th>
<th>Swiss-Large-White(SLW)</th>
<th>Swiss-Landrace(SLR)</th>
<th>Swiss-Large-White-Sire-Line(SLWSL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Pho-Var</td>
</tr>
<tr>
<td>(ZiRe)</td>
<td>7.42</td>
<td>0.53</td>
<td>0.28</td>
</tr>
<tr>
<td>(ZiLi)</td>
<td>7.31</td>
<td>0.49</td>
<td>0.24</td>
</tr>
<tr>
<td>(XoHi)</td>
<td>3.50</td>
<td>0.63</td>
<td>0.39</td>
</tr>
<tr>
<td>(SaHi)</td>
<td>3.75</td>
<td>0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>(FsHi)</td>
<td>4.08</td>
<td>0.65</td>
<td>0.42</td>
</tr>
<tr>
<td>(IkHi)</td>
<td>3.36</td>
<td>0.55</td>
<td>0.31</td>
</tr>
<tr>
<td>*(LGF)</td>
<td>11.31</td>
<td>2.76</td>
<td>7.62</td>
</tr>
<tr>
<td>*(AGF)</td>
<td>9.89</td>
<td>2.13</td>
<td>4.52</td>
</tr>
<tr>
<td>* (GSF)</td>
<td>11.36</td>
<td>2.16</td>
<td>4.65</td>
</tr>
<tr>
<td>*(FAR)</td>
<td>87.69</td>
<td>14.36</td>
<td>206.16</td>
</tr>
</tbody>
</table>

Table 4.2 Descriptive statistics of traits used in this study.
Count right (ZiRe); Count left (ZiLi); Rear Legs X-shaped-O-shaped (XoHi); Rear Legs steep –sickle hocked (SaHi); Rear Legs pastern low-steep angle (FsHi); Rear Legs inner claw small-large (IkHi); No. of piglets born alive (LGF), No. of weaned piglets (AGF); Number of piglets during the lactation (GSF); Piglets survival (FAR); * stands for reproductive traits. Pho-Var stands for Phenotypic Variance.
4.3.3. Statistical Genetic Analysis

In order to determine which fixed effects to be included in the model, a preliminary analysis was performed using the GLM procedure described by SAS Version 8.1. Those fixed effects thought to be important enough to be included in the genetic analysis then are determined. Fixed effects include in the Model were Farm-stable-period, Stable, and technician for linear description, technician for teat evaluations. Farm-stable-period, contained 307, 168, 172 levels for SLW, SLR and SLW_{SL} respectively. This fixed effect was treated as sparse in genetic analysis for all traits. Stable contained 16, 7, 6 levels, technician for linear description contained 16, 15, 15 levels and technician for teat evaluations contained 16, 18, 5 levels respectively for aforementioned breeds. Some effects were explicitly fitted for reproductive traits and letter size indirectly extracted from main pedigree. Variance covariance of genetics components were estimated using the ASREML program (Gilmour, et al., 2002). The method involved maximizing the likelihood function given the data. Across the breeds the same fixed effects and co-variables were used. Random effect of animals, dam and litter size were included in separate models to estimate genetic components as described in the following models.

Direct additive genetic model (model 1)

\[ y = X\beta + Z_1a + e \]

\[ E\left[ \begin{array}{c} a \\ e \end{array} \right] = \left[ \begin{array}{c} 0 \\ 0 \end{array} \right] \quad \text{and} \quad \text{var} \left[ \begin{array}{c} a \\ e \end{array} \right] = \left[ \begin{array}{cc} A\sigma_a^2 & 0 \\ 0 & 1\sigma_e^2 \end{array} \right] \]

\[ \lambda = \sigma_e^2 / \sigma_a^2 \]

Model 1 with litter size (model 2)

\[ y = X\beta + Z_1a + Z_2c + e \]

\[ E\left[ \begin{array}{c} a \\ c \\ e \end{array} \right] = \left[ \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right] \quad \text{and} \quad \text{var} \left[ \begin{array}{c} a \\ c \\ e \end{array} \right] = \left[ \begin{array}{ccc} A\sigma_a^2 & 0 & 0 \\ 0 & 1\sigma_c^2 & 0 \\ 0 & 0 & 1\sigma_e^2 \end{array} \right] \]

\[ \lambda = \sigma_e^2 / \sigma_a^2 \quad \lambda = \sigma_e^2 / \sigma_c^2 \]

Maternal genetic model (model 3)

\[ y = X\beta + Z_1a + Z_2m + e \]
In these models: $y$ is a vector of observations for a given trait; $\beta$ is a vector of fixed effects which influence the trait; $a$ is a vector of direct additive genetic effects; $m$ is a vector of maternal genetic effect; $e$ is a vector of residuals; $c$ is a vector of litter size effect; $\sigma_{am}$ is the covariance between additive genetic effects and maternal genetic effects; $X$ is an incidence matrix relating observations to their respective fixed effects; $Z_1$ and $Z_2$ are incidence matrices relating observations to their respective effects; $I$ is an identity matrix; $A$ is a numerator additive relationship matrix; $\lambda$ is the residual variance divided by the variance of the effect. $\lambda$ is always a function of the number of random effects in the model. The following genetic parameters were estimated: direct additive heritability: $h^2_a = \frac{\sigma^2_a}{\sigma^2_p}$; maternal genetic heritability: $h^2_m = \frac{\sigma^2_m}{\sigma^2_p}$; ratio of litter size: $h^2_c = \frac{\sigma^2_c}{\sigma^2_p}$ and total heritability defined as $h^2_t = \frac{\sigma^2_{am} + 0.5 \sigma^2_m + 1.5 \sigma^2_{am}}{\sigma^2_p}.$ (Robison, O.W., 1972) where $\sigma^2_a =$ additive genetic variance, $\sigma^2_m =$ maternal genetic variance, $\sigma^2_c =$ litter size variance, $\sigma^2_e =$ variance of error, $\sigma^2_p =$ phenotypic variance and $\sigma_{am} =$ covariance of additive genetic effect with maternal genetic effect. The phenotypic values of...
a given trait modeled in model 4 were as \( \sigma^2_p = \sigma^2_a + \sigma^2_m + \sigma_{am} + \sigma^2_e \). For estimation genetic correlation analysis, the following bi-model was used within each breed

\[
\begin{bmatrix}
    y_1 \\
    y_2
\end{bmatrix} =
\begin{bmatrix}
    X_1 & 0 \\
    0 & X_2
\end{bmatrix}
\begin{bmatrix}
    \beta_1 \\
    \beta_2
\end{bmatrix} +
\begin{bmatrix}
    Z_1 & 0 \\
    0 & Z_2
\end{bmatrix}
\begin{bmatrix}
    a_1 \\
    a_2
\end{bmatrix} +
\begin{bmatrix}
    e_1 \\
    e_2
\end{bmatrix}
\]

This bivariate genetic animal model was used to estimate three genetic components (two genetic additive variances and one single genetic covariance) per each run. \( y_1 \) and \( y_2 \) stand for two traits. The remaining terms were as defined for the univariate genetic animal models.

### 4.4. Results and Discussion

For univariate genetic animal model, we describe the results by looking at traits, however for bivariate genetic animal model analysis we describe the results by breeds, in other words, in former case, breed are categorized within traits, but in latter case trait are categorized within breeds. However one of first implication of this study was how to describe the results. Since many results trued out and managing the result in wise way to avoid falling into pleonasm was difficult task.

#### 4.4.1. Teat Traits

**4.4.1.1. Count right (ZiRe)**

Estimated direct additive heritability (\( h^2_a \)) for count right (ZiRe) in different models across breeds are given in Table 4.3. The estimated \( h^2_a \) for SLW was higher than estimated \( h^2_a \) for SLR and SLWSL breeds. This can indicate that genetic improvement of count right (ZiRe) would be more successful in SLW than two other breeds. However; this trend was not observed for estimated maternal heritability (\( h^2_m \)) of ZiRe as \( h^2_m \) was the highest (0.10 ± 0.03) for SLWSL using model 3. Estimated ratio of litter size (\( h^2_c \)) was quite small across breeds, however including litter size as a random effect, lowered the estimated additive effects (model 2). By comparing Models 3 and 4, an increase in estimated \( h^2_a \) was observed when the covariance between direct and maternal genetic effects was included and this was the case for all breeds. The estimated total heritability shows that this trait can be considered to be under more environmental than genetic.
Comparisons of predicted additive genetic heritability of pig’s teat number have been published (Borchers, N. et al., 2002). Their work indicates a wide range of heritability for teat number across and within breeds (0.15 to 0.59). Our estimations fall within this range across breeds and fitted models.

4.4.1.2. Count left (ZiLi)

Table 4.3 shows estimated genetic parameter for this trait. Biologically count left and count right should be similar, but the estimated genetic parameters were not the same. The estimated genetic components were higher across the breeds for count left. Like count right in different models across breeds the estimated $h^2$ values were higher for SLW breed. These traits can play an effective role in pig evaluation. By our knowledge there are no genetic parameters available for these traits in the literatures. However, it is shown that teat number (which here we treat it as left and right count) has bilateral symmetry, with only minor differences between the two sides. Besides, the current study has shown that the traits as described are heritable traits with moderate heritability. Number of teats with some deviations from our traits description, have been the subject of several studies. Fernandez, A. et al., (2004) used line-cross analysis of the number of teats and their asymmetry. Using Bayesian paradigm they estimated 0.25 posterior mean for additive heritability of teat number. The 95% Bayesian highest posterior density credible (HPD) for this posterior mean was (0.197 to 0.301). This is close to our estimation of heritability for count right (ZiRe) and count left (ZiLi) across breeds. The difference in estimated heritability between count right (ZiRe) and count left (ZiLi) in the current study is most likely because of structure of these data. Paying close attention to estimated maternal genetic heritabilities of teat-related traits is important as these traits are being investigated for major gene analysis and QTLs mapping e.g. as in (Kadarmideen, H. N. et al., 2004). Sanchez, M. P. et al., (2003) investigated the existence of major genes affecting false and good teat number and some other traits in the Tameslan line by applying both maximum likelihood via a Quasi Newton algorithm and Bayesian via a Gibbs sampling algorithm method. They showed clear evidence for the segregation of a fourth major gene influencing the numbers of false and good teats. Also, Sato, S. et al., (2006) analyzed the QTL affecting teat number by the nonparametric interval mapping method. They detected 5 QTL affecting number of teats in the region under study. Of these 5, only the QTL located at 29.7 cM on SSC8 QTL showed a significant imprinting
effect. In addition, some selection has been performed on teat number by culling animals carrying false teats (Rodriguez, C. et al., 2005; Zhang, S. Q. et al., 2000) investigated chromosomal regions affecting the number of teats in pigs and possible epistatic interactions between the identified QTL. The results showed three genome-wide significant QTL mapping on chromosomes 5, 10 and 12, whose joint action control up to 30% of the phenotypic variance of the trait. Also a positive-additive and additive-epistatic interaction was detected between QTL on chromosomes 10 and 12.
<table>
<thead>
<tr>
<th>Breeds</th>
<th>Model</th>
<th>Count right (ZiRe)</th>
<th>Count left (ZiLi)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$h_a^2$</td>
<td>$h_m^2$</td>
</tr>
<tr>
<td>SLW</td>
<td>1</td>
<td>0.21±0.06</td>
<td>0.21±0.06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.20±0.01</td>
<td>0.20±0.01</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23±0.04</td>
<td>0.23±0.04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.28±0.04</td>
<td>0.28±0.04</td>
</tr>
<tr>
<td>SLR</td>
<td>1</td>
<td>0.17±0.06</td>
<td>0.17±0.06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.16±0.06</td>
<td>0.16±0.06</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.21±0.04</td>
<td>0.21±0.04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.26±0.00</td>
<td>0.26±0.00</td>
</tr>
<tr>
<td>SLWSL</td>
<td>1</td>
<td>0.16±0.06</td>
<td>0.16±0.06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.13±0.01</td>
<td>0.13±0.01</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.18±0.06</td>
<td>0.18±0.06</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.22±0.09</td>
<td>0.22±0.09</td>
</tr>
</tbody>
</table>

Table 4.3 Estimation of teat traits genetic components across breeds.

- $h_a^2$ - Direct additive heritability.
- $h_m^2$ - Maternal additive heritability.
- $h_c^2$ - Ratio of litter size on phenotypic variance.
- $h_t^2$ - Total heritability,

$$h_t^2 = \left[ (\sigma_a^2 + 0.5\sigma_m^2 + 1.5\sigma_{am}) / \sigma_p^2 \right]$$.
4.4.2. Legs Traits

4.4.2.1. Rear Legs X-shaped-O-shaped (XoHi)

The estimated genetic parameters for this trait are shown in Table 4. 4, Table 4. 5 and Table 4. 6. Comparing the breeds based on estimation by fitting model 1 shows that greatest estimated $h^2_a$ was for SLW (0.23 ± 0.02) but comparing breeds based on the result of fitted model 4 shows SLWSL were superior in terms of estimated $h^2_a$ (0.32 ± 0.04). Also estimated $h^2_m$ for this trait were highest for SLWSL breed using either model 4 or 3. Estimation of a total genetic heritability had a unique trend across breeds. But estimated ratio of litter size to total phenotypic variance was substantial for SLR and SLWSL breeds. As in the previous traits, going from model 1 to model 4 showed an increase in the estimated genetic heritability observed. Including maternal effects in models for genetic analysis of Rear Legs X-shaped-O-shaped (XoHi) was important for all three breeds with large estimated maternal heritability. Probably more importantly, when maternal effects exist, maternal heritability may be underestimated and correlations between direct and maternal effects may be biased downward with models typically used to estimate maternal effects for Rear Legs X-shaped-O-shaped (XoHi). Because of the relatively large standard errors in the present analyses, larger data sets need to be analyzed in order to estimate the magnitude of biases in different populations more conclusively. Looking at Table 4. 1, it indicates that our results are lower than the predicted parameters in the literature. Comparison with other studies is difficult due to variation in trait definition.

4.4.2.2. Rear Legs steep – sickle hocked (SaHi)

Table 4. 4, Table 4. 5 and Table 4. 6 shows genetic parameter estimated for this trait. Specifying covariance between direct additive genetic effect and maternal genetic effect was important for this trait. Model failed to converge and hence $h^2_a$ in SLW and SLR was not estimated. This may reflect the importance of data structures in estimating genetic components. Applying more complex variance covariance structures often meant that model did not converge, but changing starting values help to reach convergence. Serenius, T. et al., (2004) used linear models to estimate the direct additive heritability of overall leg score as (0.06 ± 0.01). Also, they estimated additive direct heritability for overall leg
action (0.10) and for buck-kneed forelegs (0.11). These results are lower than our estimated parameters, though the traits were defined differently. Fukawa, K. et al., (2001) estimated the heritability of leg weakness and joint lesion scores in a range from 0.01 to 0.42. The ratio of litter size effect to phenotypic variance \((h^2)\) in SLW and SLW\(_{SL}\) in the current study was \((0.12 \pm 0.02)\) and \((0.11 \pm 0.02)\) respectively. The total estimated heritability was less than literature estimates (Table 4.1).

4.4.2.3. Rear Legs pastern low-steep angle (FsHi)

Table 4.4, Table 4.5 and Table 4.6 shows estimated genetic parameters across all breeds. The largest estimated direct additive heritability \((h^2)\) using model 1 for SLW\(_{SL}\) was 0.33 and SLR was 0.33. Fluctuation of estimated \(h^2\) was observed by applying more random effects across all breeds. This may indicate the complexities of underlying genetic effect which governs this trait i.e. most likely genes with pleiotropy effects are involved for this trait. This can be even further more justified with substantial \(h^2\) across breeds. The result indicates that our estimation of additive genetic heritability of this trait is over than that reported in the literatures (Table 4.1). However in comparing the result of this study with some other studies we must take into account the methods which were applied by other researcher to estimate the genetic components. In most sire effect model, the estimation of parameters is less than the animal model effects. However, it can be shown that maternal genetic effects contribute to substantial genetic variance for this trait which must be taken into account. Leg-related traits in most studies have been related to affect incidence of osteochodrosis (Jorgensen, B. et al., 2005; Kadarmideen, H. N. et al., 2004). These studies suggest that leg-related traits can be used as a marker to select against osteochondrosis. However leg-related traits must also be considered in different perspectives, for instance, say indirect effect of these traits on longevity and animal comfortable gaits. Lee, G. J. et al., (2003) identified a QTL significant at the genome-wide level affecting gait on chromosome 1 and QTL significant at the suggestive level were detected on several other chromosomes with clusters of effects on chromosomes 1, 13, 14 and 15. This may have been caused by the increased number of animals per maternal half-sib group with their covariance being a quarter of the additive genetic variance (i.e., only the dams as common parent), and unity for direct-maternal covariance. Similar pattern of results was also observed by (Gerstmayr, S. 1992).
4.4.2.4. Rear Legs inner claw small-large(IkHi)

Table 4.4, Table 4.5 and Table 4.6 shows estimated genetic parameters across all breeds. Estimated total heritability across breeds using model 1 and model 4 were almost the same. Also almost estimated $h^2_a$ were low and close to zero for SLW and SLW_{SL}. Some studies used non-animal models to estimate genetic parameters. In most sire effect model, the estimation of parameters is less than the animal model effects. However, this trait may likely affect by environmental effects a lot. Despite the low magnitude of $h^2_m$ across breeds, and based on the likelihood ratio test (Table 4.13) the inclusion of the maternal effect was significant for the trait considered here. About the results, all we can say for certain is that maternal genetic effects were greater than additive and should be included and genetic improvement will be slow due to low heritability.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Model</th>
<th>$h^2_a$</th>
<th>$h^2_m$</th>
<th>$h^2_c$</th>
<th>$h^2_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rear Legs X-shaped-O-shaped (XoHi)</td>
<td>1</td>
<td>0.23±0.02</td>
<td></td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.17±0.02</td>
<td></td>
<td>0.07±0.03</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.18±0.10</td>
<td>0.10±0.10</td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.30±0.04</td>
<td>0.10±0.16</td>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td>Rear Legs steep –sickle hocked (SaHi)</td>
<td>1</td>
<td>0.18±0.02</td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.20±0.08</td>
<td></td>
<td>0.00±0.01</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.18±0.01</td>
<td>0.12±0.02</td>
<td></td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.09±0.01</td>
<td>0.07±0.01</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Rear Legs pastern low-steep angle (FsHi)</td>
<td>1</td>
<td>0.29±0.09</td>
<td></td>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.22±0.10</td>
<td></td>
<td>0.07±0.00</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23±0.10</td>
<td>0.08±0.15</td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.36±0.04</td>
<td>0.18±0.15</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>Rear Legs inner claw small-large(IkHi)</td>
<td>1</td>
<td>0.31±0.10</td>
<td></td>
<td></td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.17±0.10</td>
<td></td>
<td>0.08±0.09</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23±0.09</td>
<td>0.02±0.00</td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.29±0.02</td>
<td>0.02±0.11</td>
<td></td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 4.4 Estimation of leg traits genetic components for Swiss-Large-White breed. $h^2_a$- Direct additive heritability; $h^2_m$ - Maternal additive heritability $h^2_c$ - Ratio of litter size on phenotypic variance; $h^2_t$-Total heritability = $[(\sigma^2_a + .5\sigma^2_m + 1.5\sigma_{am}) / \sigma^2_p]$.  

88
<table>
<thead>
<tr>
<th>Trait Model</th>
<th>Model</th>
<th>$h_a^2$</th>
<th>$h_m^2$</th>
<th>$h_c^2$</th>
<th>$h_t^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rear Legs X-shaped-O-shaped (XoHi)</td>
<td>1</td>
<td>0.22 ± 0.02</td>
<td></td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.21 ± 0.10</td>
<td></td>
<td></td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.26 ± 0.10</td>
<td>0.11 ± 0.01</td>
<td></td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.25 ± 0.04</td>
<td>0.10 ± 0.16</td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>Rear Legs steep –sickle hocked (SaHi)</td>
<td>1</td>
<td>0.20 ± 0.02</td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.10 ± 0.02</td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.18 ± 0.09</td>
<td>0.12 ± 0.01</td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.09 ± 0.01</td>
<td>0.12 ± 0.01</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Rear Legs pastern low-steep angle (FsHi)</td>
<td>1</td>
<td>0.33 ± 0.09</td>
<td></td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.24 ± 0.10</td>
<td>0.05 ± 0.01</td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.24 ± 0.09</td>
<td>0.18 ± 0.15</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.37 ± 0.10</td>
<td>0.28 ± 0.15</td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>Rear Legs inner claw small-large(IkHi)</td>
<td>1</td>
<td>0.32 ± 0.20</td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.24 ± 0.05</td>
<td></td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23 ± 0.07</td>
<td>0.1 ± 0.09</td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.27 ± 0.04</td>
<td>0.1 ± 0.11</td>
<td></td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 4.5 Estimation of leg traits genetic components for Swiss-Landrace breed.

$h_a^2$ - Direct additive heritability; $h_m^2$ - Maternal additive heritability; $h_c^2$ - Ratio of litter size on phenotypic variance; $h_t^2$ - Total heritability = \( \left[ \frac{\sigma_a^2 + 0.5\sigma_m^2 + 1.5\sigma_{am}}{\sigma_p^2} \right] \);

<table>
<thead>
<tr>
<th>Trait Model</th>
<th>Model</th>
<th>$h_a^2$</th>
<th>$h_m^2$</th>
<th>$h_c^2$</th>
<th>$h_t^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rear Legs X-shaped-O-shaped (XoHi)</td>
<td>1</td>
<td>0.20 ± 0.02</td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.18 ± 0.09</td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.18 ± 0.12</td>
<td>0.19 ± 0.12</td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.32 ± 0.04</td>
<td>0.10 ± 0.16</td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td>Rear Legs steep –sickle hocked (SaHi)</td>
<td>1</td>
<td>0.2 ± 0.02</td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.2 ± 0.12</td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.24 ± 0.11</td>
<td>0.10 ± 0.12</td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.19 ± 0.02</td>
<td>0.11 ± 0.02</td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>Rear Legs pastern low-steep angle (FsHi)</td>
<td>1</td>
<td>0.33 ± 0.09</td>
<td></td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.14 ± 0.07</td>
<td></td>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23 ± 0.00</td>
<td>0.16 ± 0.15</td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.40 ± 0.10</td>
<td>0.20 ± 0.15</td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td>Rear Legs inner claw small-large(IkHi)</td>
<td>1</td>
<td>0.33 ± 0.10</td>
<td></td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.14 ± 0.20</td>
<td></td>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23 ± 0.07</td>
<td>0.09 ± 0.04</td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.29 ± 0.04</td>
<td>0.09 ± 0.11</td>
<td></td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 4.6 Estimation of leg traits genetics components for Swiss-Large-White-Sire-Line breed.

$h_a^2$ - Direct additive heritability; $h_m^2$ - Maternal additive heritability; $h_c^2$ - Ratio of litter size on phenotypic variance; $h_t^2$ - Total heritability = \( \left[ \frac{\sigma_a^2 + 0.5\sigma_m^2 + 1.5\sigma_{am}}{\sigma_p^2} \right] \);
4.4.3. Reproductive Traits

4.4.3.1. Number of piglets born alive (LGF)

Table 4. 7, Table 4. 8 and Table 4. 9 shows predicted results. Importantly, $h^2_a$ was estimated downwardly for almost all of reproductive traits by applying (model 3 and model 4). Litter size did not have low estimated $h^2_a$ as it was the case for reproductive traits. Estimation $h^2_m$ applying model 3 was small across breeds. Torres, R. A. et al., (2005) and Torres, R. D. A. et al., (2005) using restricted maximum likelihood method (REML) estimated direct additive heritability ranged from 0.17 to 0.34 for number of piglets born alive and other related traits. They concluded that because of favorable genetic correlations between this trait and age at first farrowing and total number of piglets born selection strategies should be sought to increase genetic gain and that response for total number of piglets born may be obtained by indirect selection on number of piglets born alive. Their heritability estimates were consistent with ours. Also, they reported low estimates of maternal genetic heritability which matches our estimation. Chu, M. X., (2005) estimated direct additive heritability for percentage born alive between 0.11 and 0.13. This estimation is less than our estimation across breeds and Torres, R. A. et al., (2005) and Torres, R. D. A. et al., (2005) but is close to result of Tholen, E. et al., (1996) whose estimates ranged between (0.09-0.16). Canario, L. et al., (2006) argued that in contrary to selection on the number of piglet born in total, selection on the number piglet born alive appears to be a good way to limit the negative side effects on stillbirth. However their estimate of (0.08 ± 0.02) for number piglet born alive was also less than our estimates. Holm, B. et al., (2004) estimated 0.12 additive heritability for number of live-born piglets in the second litter which this estimation is less than that our estimation across breeds.

4.4.3.2. Number of weaned piglets (AGF)

Table 4. 7, Table 4. 8 and Table 4. 9 show estimated genetic parameters for AGF across breeds. Estimated $h^2_a$ across breeds and models was similar to LGF. Estimates of $h^2_m$ (model 3 and 4) were generally low. Estimates of total heritability across breeds for number of weaned piglets was higher than that of Serenius, T. et al., (2004). Torres, R. A.
et al., (2005) reported low estimates of maternal genetic effects on reproductive traits in pigs. They argued that one factor that could explain low, null or negative estimates of genetic trends for maternal effects is the antagonism between direct and maternal effects. If selection is based only on direct additive genetic effects maternal effects tend to decrease. The additive maternal effect of a sow is inherited from both her sire and dam, and expressed in her offspring's reproductive performance, i.e. one generation later than the additive direct effect (Skorupski, M. T. et al., 1996).

4.4.3.3. Number of piglets during the lactation (GSF)

Table 4.7, Table 4.8 and Table 4.9 show estimated genetic parameters for this GSF. Estimated total heritability was low across breeds. However, including maternal genetic effect sharply affected estimated genetic components. In this way SLR shows more vulnerability. This indicates the importance of maternal genetic effect in genetic analysis. In particular, maternal effects may be caused by genes carried by the dam as well as by the environment she experiences. Such maternal genetic effects will represent a heritable source of phenotypic variance. Effective modeling and prediction of response to selection is therefore likely to require the estimation of maternal genetic effects. Quantitative genetic models have shown that, as a result, phenotypic responses to selection may be accelerated by maternal genetic effects, but can also be dampened (or even occur in counterintuitive directions) if there is negative covariance between direct and maternal genetic effects (Wilson, A.J. et al., 2005). At a mechanistic level such negative covariance might occur through pleiotropy, for example if a gene has a positive effect on an offspring trait but a negative effect on maternal performance for that trait.

4.4.3.4. Piglets survival (FAR)

Table 4.7, Table 4.8 and Table 4.9 show estimated genetic parameters for FAR. General estimation of $h^2_c$ obtained across the breeds. Estimates of $h^2_a$ across breeds and models for FAR were similar to estimates of total heritability ($h^2_t$). Total heritability has the intuitive interpretation of describing the potential response to selection, replacing direct heritability in the classical breeders equation ($R = \frac{1}{2} h^2 S$) when maternal genetic effects are present (Wilson, A. J. et al., 2005). White, I. M. S. et al., (2006) used different approaches and reported smaller values for additive genetic variance and maternal genetic
variance than our estimates. Estimates of genetic parameters for FAR in the current study using different animal models showed that including the covariance between direct and maternal effects did not result in a significantly better fit in comparison to other models when judged by the Log likelihood (Log L). As for other reproductive traits, the inclusion of litter size component generally reduced both additive genetic variance and maternal variance. Based on results in the present study, it appears that the maternal genetic effects are important across the three breeds studied. This is somewhat surprising, because the animals no longer depend on their dam and should reflect only direct effects of the genes for the traits investigated. However, maternal effects on other species have also been reported (Boipuso, A. 2004). In their study they argued the fact that for animals raised on pasture with little or no supplementary feeding, the length of time between weaning and yearling may not be enough to buffer maternal effects existing at weaning.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Model</th>
<th>$h_a^2$</th>
<th>$h_m^2$</th>
<th>$h_c^2$</th>
<th>$h_t^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of piglets born alive (LGF)</td>
<td>1</td>
<td>0.19 ± 0.05</td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.18 ± 0.02</td>
<td>0.10 ± 0.09</td>
<td>0.12 ± 0.07</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.18 ± 0.02</td>
<td>0.10 ± 0.09</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.16 ± 0.04</td>
<td>0.03 ± 0.01</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Number of weaned piglets (AGF)</td>
<td>1</td>
<td>0.15 ± 0.08</td>
<td></td>
<td>0.08 ± 0.20</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.19 ± 0.10</td>
<td>0.10 ± 0.09</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23 ± 0.10</td>
<td>0.08 ± 0.01</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.16 ± 0.04</td>
<td>0.03 ± 0.01</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Number of piglets during the lactation (GSF)</td>
<td>1</td>
<td>0.15 ± 0.02</td>
<td>0.05 ± 0.06</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.14 ± 0.05</td>
<td>0.06 ± 0.00</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.17 ± 0.05</td>
<td>0.03 ± 0.01</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.17 ± 0.01</td>
<td>0.03 ± 0.01</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Piglets survival (FAR)</td>
<td>1</td>
<td>0.13 ± 0.05</td>
<td></td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.10 ± 0.03</td>
<td>0.10 ± 0.05</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.10 ± 0.01</td>
<td>0.08 ± 0.01</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.11 ± 0.01</td>
<td>0.07 ± 0.00</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7 Estimation of reproductive traits genetics components for Swiss-Large-White breed.

$h_a^2$ - Direct additive heritability; $h_m^2$ - Maternal additive heritability $h_c^2$ - Ratio of litter size on phenotypic variance; $h_t^2$ - Total heritability = \([\sigma_a^2 + .5\sigma_m^2 + 1.5\sigma_m^2]/\sigma^2_p\).
<table>
<thead>
<tr>
<th>Trait Model</th>
<th>Trait Description</th>
<th>Model</th>
<th>$h^2_a$</th>
<th>$h^2_m$</th>
<th>$h^2_c$</th>
<th>$h^2_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of piglets born alive (LGF)</td>
<td>0.18±0.05</td>
<td>0.24±0.02</td>
<td>0.11±0.02</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.23±0.10</td>
<td>0.13±0.01</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.16±0.04</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Number of weaned piglets (AGF)</td>
<td>0.18±0.08</td>
<td>0.09±0.20</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.23±0.07</td>
<td>0.03±0.01</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.16±0.04</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.19±0.15</td>
<td>0.10±0.01</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Number of piglets during the lactation (GSF)</td>
<td>0.17±0.02</td>
<td>0.09±0.05</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.13±0.05</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.14±0.01</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.19±0.15</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Piglets survival (FAR)</td>
<td>0.13±0.05</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.12±0.10</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.10±0.01</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.09±0.07</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.8 Estimation of Reproductive traits genetics components for Swiss-Landrace breed.

$h^2_a$ - Direct additive heritability; $h^2_m$ - Maternal additive heritability; $h^2_c$ - Ratio of litter size on phenotypic variance; $h^2_t$ - Total heritability = \left(\frac{\sigma^2_a + 0.5 \sigma^2_m + 1.5 \sigma_{am}}{\sigma^2_p}\right).

<table>
<thead>
<tr>
<th>Trait Model</th>
<th>Trait Description</th>
<th>Model</th>
<th>$h^2_a$</th>
<th>$h^2_m$</th>
<th>$h^2_c$</th>
<th>$h^2_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of piglets born alive (LGF)</td>
<td>0.18±0.05</td>
<td>0.24±0.02</td>
<td>0.10±0.11</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.28±0.02</td>
<td>0.09±0.01</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.16±0.03</td>
<td>0.10±0.11</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.11±0.08</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Number of weaned piglets (AGF)</td>
<td>0.20±0.09</td>
<td>0.11±0.10</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.23±0.10</td>
<td>0.10±0.01</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.16±0.10</td>
<td>0.09±0.01</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.19±0.02</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Number of piglets during the lactation (GSF)</td>
<td>0.20±0.05</td>
<td>0.09±0.05</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.19±0.05</td>
<td>0.10±0.00</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.18±0.01</td>
<td>0.11±0.01</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.13±0.00</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Piglets survival (FAR)</td>
<td>0.14±0.05</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.11±0.05</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.10±0.00</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.09±0.02</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9 Estimation of reproductive traits genetics components for Swiss-Large-White-Sire-Line breed.

$h^2_a$ - Direct additive heritability; $h^2_m$ - Maternal additive heritability $h^2_c$ - Ratio of litter size on phenotypic variance; $h^2_t$ - Total heritability = \left(\frac{\sigma^2_a + 0.5 \sigma^2_m + 1.5 \sigma_{am}}{\sigma^2_p}\right).
4.4.4. Genetic Correlations

First and foremost it must be mentioned that estimated genetic correlations were sensible to starting parameters values. Estimation of genetic correlations among traits could be of substantial interest because, depending on their sign (positive or negative), they can either facilitate or impede the joint genetic progress of the traits involved in long or short-term genetic program. A conflict can immediately arises when two negatively genetically correlated traits are both selected in the same direction, as the selective advance of each trait tends to pull the other trait in the opposite direction. However, genetic correlation between traits could be less straightforward, as both the magnitude and the sign of this parameter depend on the general pleiotropic effects of genes and, in some cases, on the pattern of gametic-phase disequilibrium (Lynch, M. 1999). At the follows, we turn to explain genetic correlation among traits within breed and then general words are given with respect to applied data and results of genetic correlations. Course of many estimated genetic parameters; we restrict to address some outlines of results. In other words, we don’t delve too much into explaining every single aspect of results. However, the primary idea of having such enormous genetic correlation estimation was to provide enough estimated genetic parameters for breeds of under study in a way that efficient and secure breeding program can be set up. Having all said, we turn to explain only the “outline of results” and intensive discussion of results is skipped.

4.4.4.1. Swiss-Large-White Breed

The estimates of the correlations (genetic and phenotypic) from the bivariate genetic animal model analysis among all traits within SLW breed are given in Table 4. 10. Heritabilities from the bivariate genetic model were generally higher than those from the univariate analyses. Genetic correlations between the traits studied were favorable among reproductive traits, indicating that selection for one trait will improve others in a desired direction, helping the breeding process as a whole. Teat count right showed highest positive genetic correlation with teat count left across breeds (0.91 ± 0.03). This trait showed negative genetic correlation with rear legs steep–sickle hocked across breeds (Table 4. 10 to Table 4. 12). However the magnitude of negative genetic correlations weren’t similar across breeds. The largest negative genetic correlation was observed for SLW (-0.13 ± 0.01). The phenotypic correlations of this trait with other traits were low.
Generally reproductive and exteriors traits were genetically related to each other. The genetic correlation between ZiRe and LGF was low across breeds (average of -0.04). This indicates that genetic selection for one won’t affect the other across breeds. SaHi showed highest negative genetic correlations with FsHi (-0.33 ± 0.30). In total, this trait showed more negative genetic correlation with other traits than other considered traits. SaHi showed highest positive genetic correlation with IkHi in SLR but not for other breeds. Draper, D. D. et al., (1988) reported that selection for different degrees of leg weakness resulted in accompanying alterations in angularity of joints. 

Table 4.10 shows, estimated genetic correlations among exterior and reproductive traits, were moderate. In other words, genetic selection for reproductive traits won’t affect exterior traits and vice versa based on our study. Serenius, T. et al., (2004) showed no clear genetic associations between prolificacy and leg conformation which comply with the current study. Linear exterior and reproductive traits can be useful predictors of survival in SLW breed. Teat traits especially could have a profound effect on pig survival. Inter-genetic negative correlation among these traits especially with reproductive traits can cause failure of genetic progress. Estimated genetic correlations of LGF with AGF and GSF in the current study were moderate to high (0.30 and 0.50 respectively). Serenius, T. et al., (2004) estimated the genetic correlation between number of piglets born alive and number of weaned piglets at 0.54 ± 0.12, which is to somewhat higher than our estimation. Genetic correlation between AGF and GSF in the current study was low (0.04± 0.13). Estimation of phenotypic correlation among the traits under study was moderate to low.
<table>
<thead>
<tr>
<th>Traits</th>
<th>ZiRe</th>
<th>ZiLi</th>
<th>xoHi</th>
<th>SaHi</th>
<th>FsHi</th>
<th>IkHi</th>
<th>LGF</th>
<th>AGF</th>
<th>GSF</th>
<th>Far</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZiRe</td>
<td>0.23 ± 0.15</td>
<td>0.91 ± 0.03</td>
<td>0.01 ± 0.7</td>
<td>-0.13 ± 0.01</td>
<td>-0.15 ± 0.06</td>
<td>0.08 ± 0.07</td>
<td>-0.05 ± 0.03</td>
<td>0.04 ± 0.02</td>
<td>0.05 ± 0.01</td>
<td>-0.07 ± 0.02</td>
</tr>
<tr>
<td>ZiLi</td>
<td>0.39 ± 0.01</td>
<td>0.32 ± 0.06</td>
<td>0.01 ± 0.07</td>
<td>-0.12 ± 0.02</td>
<td>-0.17 ± 0.06</td>
<td>0.12 ± 0.01</td>
<td>0.07 ± 0.02</td>
<td>-0.02 ± 0.02</td>
<td>0.14 ± 0.05</td>
<td>0.10 ± 0.03</td>
</tr>
<tr>
<td>xoHi</td>
<td>-0.02 ± 0.09</td>
<td>0.01 ± 0.02</td>
<td>0.31 ± 0.30</td>
<td>-0.11 ± 0.00</td>
<td>-0.33 ± 0.06</td>
<td>-0.12 ± 0.07</td>
<td>0.05 ± 0.08</td>
<td>0.07 ± 0.09</td>
<td>0.09 ± 0.02</td>
<td>0.09 ± 0.01</td>
</tr>
<tr>
<td>SaHi</td>
<td>-0.02 ± 0.08</td>
<td>-0.01 ± 0.02</td>
<td>-0.07 ± 0.01</td>
<td>0.28 ± 0.09</td>
<td>0.17 ± 0.07</td>
<td>0.26 ± 0.07</td>
<td>-0.05 ± 0.03</td>
<td>0.09 ± 0.04</td>
<td>-0.01 ± 0.09</td>
<td>-0.10 ± 0.01</td>
</tr>
<tr>
<td>FsHi</td>
<td>-0.09 ± 0.09</td>
<td>-0.06 ± 0.02</td>
<td>-0.22 ± 0.01</td>
<td>0.07 ± 0.02</td>
<td>0.21 ± 0.20</td>
<td>-0.08 ± 0.06</td>
<td>0.02 ± 0.05</td>
<td>-0.03 ± 0.02</td>
<td>0.05 ± 0.02</td>
<td>-0.07 ± 0.09</td>
</tr>
<tr>
<td>IkHi</td>
<td>0.03 ± 0.02</td>
<td>0.01 ± 0.00</td>
<td>-0.03 ± 0.01</td>
<td>0.11 ± 0.01</td>
<td>-0.02 ± 0.01</td>
<td>0.30 ± 0.02</td>
<td>-0.09 ± 0.00</td>
<td>-0.08 ± 0.00</td>
<td>-0.06 ± 0.01</td>
<td>-0.08 ± 0.00</td>
</tr>
<tr>
<td>LGF</td>
<td>0.04 ± 0.00</td>
<td>0.16 ± 0.02</td>
<td>0.04 ± 0.01</td>
<td>-0.12 ± 0.02</td>
<td>0.04 ± 0.01</td>
<td>-0.09 ± 0.04</td>
<td>0.22 ± 0.20</td>
<td>0.50 ± 0.12</td>
<td>0.30 ± 0.07</td>
<td>0.11 ± 0.03</td>
</tr>
<tr>
<td>AGF</td>
<td>-0.016 ± 0.017</td>
<td>-0.04 ± 0.01</td>
<td>-0.16 ± 0.09</td>
<td>-0.16 ± 0.07</td>
<td>-0.25 ± 0.07</td>
<td>-0.04 ± 0.00</td>
<td>-0.04 ± 0.06</td>
<td>0.28 ± 0.15</td>
<td>0.04 ± 0.13</td>
<td>0.20 ± 0.09</td>
</tr>
<tr>
<td>GSF</td>
<td>0.05 ± 0.01</td>
<td>-0.15 ± 0.13</td>
<td>0.14 ± 0.01</td>
<td>0.01 ± 0.02</td>
<td>0.07 ± 0.00</td>
<td>-0.03 ± 0.07</td>
<td>0.41 ± 0.06</td>
<td>-0.22 ± 0.05</td>
<td>0.23 ± 0.20</td>
<td>0.14 ± 0.03</td>
</tr>
<tr>
<td>Far</td>
<td>0.12 ± 0.04</td>
<td>0.23 ± 0.19</td>
<td>0.11 ± 0.01</td>
<td>-0.13 ± 0.07</td>
<td>0.07 ± 0.05</td>
<td>-0.34 ± 0.07</td>
<td>0.77 ± 0.02</td>
<td>0.11 ± 0.08</td>
<td>0.41 ± 0.04</td>
<td>0.31 ± 0.02</td>
</tr>
</tbody>
</table>

Table 4.10 Estimates of heritabilities (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among traits in Swiss-Large-Whit breed.

Count right (ZiRe); Count left (ZiLi); Rear Legs X-shaped-O-shaped (XoHi); Rear Legs steep–sickle hocked (SaHi); Rear Legs pastern low–steep angle (FsHi); Rear Legs inner claw small-large (IkHi); No. of piglets born alive (LGF), No. of weaned piglets (AGF); Number of piglets during the lactation (GSF); Piglets survival (FAR).
Like SLW breed, in this breed small portion of genetic correlation was observed among exterior traits (leg and teat) with reproductive traits. Bottom principle is that genetic correlations between two traits are caused by two mechanisms, linkage, and pleiotropy. Linkage can occur when the underlying controlling genes of two traits are located near each other on the same chromosome and therefore are transmitted from parent to offspring together. Pleiotropy is a phenomenon where one gene, or group of genes, controls more than one trait. Having this introduction, what we can say is that the genes that are responsible for penetrance of exterior and reproductive traits are either located in different chromosomes or are located far away in the same chromosome (small linkage). Therefore, genetic selection for one trait won’t affect the other traits by much. Genome mapping could be good scientific tool to decipher this phenomenon in DNA level. Generally speaking, multiple traits genome mapping more likely makes sense for highly correlated traits not for not non-correlated trait. If the idea in this population would be to increase the performance of all traits, genetic correlations can show which trait likely share in common the same genomic distance at DNA level. Du, F. X. et al., (2007) with aid of average population-wise LD in six commercial pig lines, recommended a DNA interval sapping from 0.1 to 1 cM for a whole genome association study in pig populations. Also highly correlated genetic traits can help for running multiple traits QTL mapping too. Rothschild, M. F. et al., (2007) issued that QTL discovery in the pig has advanced rapidly but funding remains a limiting factor as does the extension , development and maintenance of specialized pig families and populations for certain traits and disorders. Table 4. 11 shows almost the same profile for genetic correlation among traits with SLW breed. Genetic correlations from about 0.3 to 0.55 within reproductive traits for this breed mean that those traits will progress strongly in the same direction if selection is practiced for one trait of the pair. Negative genetic correlation between IkHi and GSF (-0.55 ± 0.10), mean that selection to increase IkHi trait will result in a reduction in genetic merit for the second GSF and vise versa. Genetic correlations in the middle ground which is exist among leg traits mean that slow change will occur due to selection for one pair. Sometimes such kind of correlations indicate serious, long term economic problems or opportunities to improve traits that are difficult or expensive to record and measure. Justifying it is based on researcher knowledge. Sometimes such correlations can be ignored in a breeding program. Comparing to SLW breed,
reproductive traits in SLR showed larger genetic correlations. However the leg trait correlation remain to be almost the same as SLW though some discrepancies are observed.
<table>
<thead>
<tr>
<th>Traits</th>
<th>ZiRe</th>
<th>ZiLi</th>
<th>xoHi</th>
<th>SaHi</th>
<th>FsHi</th>
<th>IkHi</th>
<th>LGF</th>
<th>AGF</th>
<th>GSF</th>
<th>Far</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZiRe</td>
<td>0.22±0.15</td>
<td>0.89±0.19</td>
<td>0.02±0.07</td>
<td>-0.12±0.07</td>
<td>-0.14±0.06</td>
<td>0.08±0.07</td>
<td>-0.03±0.00</td>
<td>0.07±0.02</td>
<td>0.06±0.01</td>
<td>-0.05±0.02</td>
</tr>
<tr>
<td>ZiLi</td>
<td>0.33±0.01</td>
<td>0.22±0.06</td>
<td>0.03±0.07</td>
<td>-0.11±0.05</td>
<td>-0.18±0.06</td>
<td>0.15±0.05</td>
<td>0.07±0.08</td>
<td>-0.02±0.04</td>
<td>0.16±0.04</td>
<td>-0.07±0.03</td>
</tr>
<tr>
<td>xoHi</td>
<td>-0.02±0.01</td>
<td>0.01±0.00</td>
<td>0.41±0.30</td>
<td>-0.10±0.06</td>
<td>-0.34±0.06</td>
<td>-0.11±0.07</td>
<td>0.04±0.07</td>
<td>0.09±0.00</td>
<td>0.07±0.02</td>
<td>0.07±0.00</td>
</tr>
<tr>
<td>SaHi</td>
<td>-0.06±0.08</td>
<td>-0.01±0.02</td>
<td>-0.07±0.10</td>
<td>0.32±0.09</td>
<td>0.15±0.07</td>
<td>0.28±0.07</td>
<td>-0.09±0.03</td>
<td>0.09±0.04</td>
<td>-0.01±0.09</td>
<td>-0.05±0.01</td>
</tr>
<tr>
<td>FsHi</td>
<td>-0.07±0.02</td>
<td>-0.04±0.01</td>
<td>-0.22±0.01</td>
<td>0.07±0.02</td>
<td>0.32±0.20</td>
<td>-0.07±0.06</td>
<td>0.08±0.05</td>
<td>-0.01±0.02</td>
<td>0.04±0.02</td>
<td>-0.09±0.09</td>
</tr>
<tr>
<td>IkHi</td>
<td>0.06±0.02</td>
<td>0.01±0.01</td>
<td>-0.03±0.10</td>
<td>0.11±0.05</td>
<td>-0.02±0.00</td>
<td>0.37±0.02</td>
<td>-0.08±0.00</td>
<td>-0.07±0.00</td>
<td>-0.5±0.10</td>
<td>-0.08±0.00</td>
</tr>
<tr>
<td>LGF</td>
<td>0.04±0.007</td>
<td>0.15±0.02</td>
<td>0.04±0.01</td>
<td>-0.02±0.02</td>
<td>0.04±0.01</td>
<td>-0.09±0.04</td>
<td>0.18±0.20</td>
<td>0.53±0.50</td>
<td>0.38±0.07</td>
<td>0.31±0.03</td>
</tr>
<tr>
<td>AGF</td>
<td>-0.12±0.07</td>
<td>-0.19±0.10</td>
<td>-0.18±0.09</td>
<td>-0.26±0.07</td>
<td>-0.15±0.07</td>
<td>-0.04±0.00</td>
<td>-0.04±0.06</td>
<td>0.18±0.15</td>
<td>0.55±0.13</td>
<td>0.21±0.09</td>
</tr>
<tr>
<td>GSF</td>
<td>0.05±0.01</td>
<td>-0.19±0.07</td>
<td>0.18±0.06</td>
<td>0.01±0.02</td>
<td>0.01±0.01</td>
<td>-0.03±0.07</td>
<td>0.41±0.06</td>
<td>-0.22±0.05</td>
<td>0.23±0.20</td>
<td>0.16±0.03</td>
</tr>
<tr>
<td>Far</td>
<td>0.03±0.04</td>
<td>0.23±0.09</td>
<td>0.08±0.01</td>
<td>-0.10±0.07</td>
<td>0.07±0.05</td>
<td>-0.26±0.07</td>
<td>0.77±0.02</td>
<td>0.11±0.08</td>
<td>0.41±0.04</td>
<td>0.27±0.02</td>
</tr>
</tbody>
</table>

Table 4.11 Estimates of heritabilities (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among traits in Swiss-Landrace breed.

Count right (ZiRe); Count left (ZiLi); Rear Legs X-shaped-O-shaped (XoHi); Rear Legs steep –sickle hocked (SaHi); Rear Legs pastern low-steep angle (FsHi); Rear Legs inner claw small-large(IkHi); No. of piglets born alive (LGF), No. of weaned piglets (AGF); Number of piglets during the lactation (GSF); Piglets survival (FAR).
4.4.4.3. Swiss-Large-White-Sire-Line

Table 4.12 shows the estimated genetic correlation among traits within this breed. Generally speaking, the result more or less fall in the middle of estimated genetic components for two aforementioned breeds. A significant matter which must be issued is this reality that teat traits (ZiRe and ZiLi) genetic correlations obtained substantial (bigger than 90%), however this matter almost had the same trend across whole breeds. General principle is that these traits definitely are controlled by same genes or genes which show large pleiotropy or linkage effect. However their corresponding estimated heritabilities were not the same across the traits. Also across breeds the genetic correlation between teat traits obtained positive which means as we select to change one trait, the second trait moves in the same direction (both increase or both decrease). However, genetic correlations may be negative for example between leg traits and reproductive traits, which means the traits respond in opposite directions (one increases as the other decreases). These correlations among traits must be considered in assessing the total impact of selection for a trait. As a trait is altered, performance for all correlated traits change as well. The net effect of changing a trait is the summation of the changes in the trait itself and all correlated traits. It means when many traits selected jointly, genetic correlation and its mode (negative or positive) plays central role. Correlations among traits may be exploited to reduce testing costs. Not of this breed but also for aforementioned breeds, estimated standard error due to each genetic parameter was high for some traits. This matter can inflict on weakness on the estimated genetic parameters. Our justification is that maybe we didn’t use linear genetic threshold model to estimate genetic parameters, since most of studied traits belonged to categorical traits. However, using this kind of modeling instead of usual genetic animal models which used here is controversial. Reader is requested to look at implication part of this study for more arguments. Estimated standard error also pushed us to look at some blind learning statistical model (see appendix) to explore the structures of data and it was somehow useful. For example in this study the genetic correlation among teat traits across breeds obtained high. As will be seeing at the appendix, these two traits always made a cohesively single cluster, which itself can likely support why estimated genetic correlation in this study obtained high. Table 4.10, Table 4.11 and Table 4.12 bear up a reality which must be taken into account, that is, even though a genetic correlation exists among leg traits, but in almost most cases this genetic correlation is negative. This matter must be taken in to account if
the idea is to improved genetic background of breeds of under study. To understand genetic characteristics of SLW\textsubscript{SL}, SLW and SLR a DNA based on study using say microsatellite loci are recommended. This can not only help to find the sate of inbreeding in aforementioned breeds but also the true genetic difference among breeds in term of their genetic performance. Kim, T. H. \textit{et al.}, (2005) across number of breeds indicated that the Korean native pig has been experiencing progressive interbreeding with Western pig breeds after originating from a North China pig breed with a black coat color. To compare the breeds, one other option would be to provide the same economic values for each trait studied for each breeds and compare either long to short-term genetic progress or economic profit across breeds. The animals are ranked on the index value and those with the greatest values are selected. It is obvious by keeping all other factors constant, number of selected of animal from a breed, can provide a general idea about their genetic superiority too. This genetic superiority to some extent depends upon potential of genetic correlation among traits. In continues this story teat traits showed sharp negative genetic correlation with leg traits. What emerges from this study to some extent is that estimating of genetic correlation within breed can be extrapolated for other breeds. Since more or less the estimated genetic correlation ranges among trait were almost similar across breeds. However, this study does not support this idea that a significant correlation under one breed would exist under different breeds. For example evolutionary pathways predicted from genetic correlations likely shall depend on the conditions experienced by breeds across generations and environmental situations. Estimation of genetic parameters is a demanding task and enterprise even with nicely balanced designs in the most controlled environments, a matter which can hardly be achieved in the practical scenarios. However, compared with univariate genetic model, genetic correlations are particularly difficult to assess because they require accurate estimates of three parameters i.e. genetic variances of the two traits, and the genetic covariance between them. Because all three estimates are generally obtained from the phenotypic covariances of relatives, they can take on any value. Thus, it is possible that estimates of the genetic correlation can fall outside of the parametric limits. Lynch, M., (1999) tried to explore new method for estimating the genetic correlations between traits expressed in individuals in completely undisturbed natural populations. In this analysis, we used general product moment correlation which is actually the classical approach to estimating genetic components of variance and covariance- based upon the phenotypic resemblance between individuals of known relatedness- since we were deal with
controlled breeding programs in which individuals are raised in artificial environments. However, it must be mentioned that even in well-controlled genetic populations explicitly when non-Gaussian traits are involved, it seems that similar approach like (Lynch, M., 1999) must be developed. One reason for this development would be that some time estimation of genetic correlation fallen out of the scope of parameter space in this study. However, most reasons which sound to be compelling were investigated to tackle with this problem but some time problems remain unsolved. Discrepancy among estimated genetic correlations for the same traits across breed could be likely due to environment or nutrition too. While comparisons of pig lines could be useful for investigating correlations across a range of environments, the interpretation of results from these experiments can be complicated by maternal and non-additive genetic effects. Sample sizes of a few thousand pairs of relatives in the pedigree are often necessary to achieve estimates that can confidently be interpreted at the level of even single significant digits (Lynch, M. 1999) although a simple knowledge of the sign of the genetic correlation can be achieved with less, but still substantial, effort. Of great importance regarding to genetic correlation estimation within breeds was to provide an overview in terms of applied model. We found it is difficult to compare the breeds based on trait-wise genetic correlation. As we can see some consistency can be seen among some traits (teat traits) in terms of estimated genetic correlations, but discrepancies are exist as well. Having nature of data, it is strongly recommended to run pair-wise genetic correlation with aid of linear threshold models. Across whole breeds, very often, estimated genetic correlation between FAR trait and other traits obtained far from the scope of parameter. This mostly forced us to conduct the analysis which is discussed at the appendix. However, still we hope that applied model has brought about some general insights for investigated traits taking into account that many literatures have been issued regarding to short-goal traits or economic traits (Hoque, M. A. et al., 2007; Zumbach, B. et al., 2007; Gilbert, H. et al., 2007) but scant literatures exist that only covered analyzing of exterior traits with reproduce traits genetically jointly.
**Table 4.12** Estimates of heritabilities (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among traits in Swiss-Large-White-Sire-Line breed.

<table>
<thead>
<tr>
<th>Traits</th>
<th>ZiRe</th>
<th>ZiLi</th>
<th>xoHi</th>
<th>SaHi</th>
<th>FsHi</th>
<th>IkHi</th>
<th>LGF</th>
<th>AGF</th>
<th>GSF</th>
<th>Far</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZiRe</td>
<td>0.28 ± 0.10</td>
<td>0.97 ± 0.37</td>
<td>0.01 ± 0.07</td>
<td>-0.02 ± 0.07</td>
<td>-0.11 ± 0.09</td>
<td>0.09 ± 0.07</td>
<td>-0.03 ± 0.03</td>
<td>0.06 ± 0.02</td>
<td>0.03 ± 0.01</td>
<td>0.10 ± 0.09</td>
</tr>
<tr>
<td>ZiLi</td>
<td>0.32 ± 0.31</td>
<td>0.20 ± 0.10</td>
<td>0.06 ± 0.10</td>
<td>-0.12 ± 0.07</td>
<td>-0.19 ± 0.06</td>
<td>0.14 ± 0.07</td>
<td>0.07 ± 0.02</td>
<td>-0.02 ± 0.02</td>
<td>0.16 ± 0.05</td>
<td>-0.07 ± 0.03</td>
</tr>
<tr>
<td>xoHi</td>
<td>-0.03 ± 0.01</td>
<td>0.01 ± 0.02</td>
<td>0.28 ± 0.30</td>
<td>-0.10 ± 0.07</td>
<td>-0.34 ± 0.06</td>
<td>-0.11 ± 0.07</td>
<td>0.04 ± 0.07</td>
<td>0.07 ± 0.09</td>
<td>0.06 ± 0.06</td>
<td>-0.09 ± 0.01</td>
</tr>
<tr>
<td>SaHi</td>
<td>-0.09 ± 0.08</td>
<td>-0.01 ± 0.02</td>
<td>-0.07 ± 0.00</td>
<td>0.23 ± 0.09</td>
<td>0.15 ± 0.07</td>
<td>0.28 ± 0.07</td>
<td>-0.03 ± 0.03</td>
<td>0.02 ± 0.04</td>
<td>-0.01 ± 0.09</td>
<td>-0.05 ± 0.01</td>
</tr>
<tr>
<td>FsHi</td>
<td>-0.06 ± 0.06</td>
<td>-0.02 ± 0.02</td>
<td>-0.22 ± 0.01</td>
<td>0.07 ± 0.02</td>
<td>0.23 ± 0.20</td>
<td>-0.07 ± 0.06</td>
<td>0.01 ± 0.05</td>
<td>-0.01 ± 0.02</td>
<td>0.04 ± 0.02</td>
<td>-0.07 ± 0.09</td>
</tr>
<tr>
<td>IkHi</td>
<td>0.00 ± 0.00</td>
<td>0.03 ± 0.01</td>
<td>-0.08 ± 0.01</td>
<td>0.13 ± 0.03</td>
<td>-0.02 ± 0.01</td>
<td>0.25 ± 0.02</td>
<td>0.00 ± 0.00</td>
<td>-0.07 ± 0.00</td>
<td>-0.6 ± 0.01</td>
<td>-0.08 ± 0.00</td>
</tr>
<tr>
<td>LGF</td>
<td>0.04 ± 0.06</td>
<td>0.15 ± 0.02</td>
<td>0.04 ± 0.01</td>
<td>-0.02 ± 0.02</td>
<td>0.04 ± 0.01</td>
<td>-0.09 ± 0.04</td>
<td>0.16 ± 0.10</td>
<td>0.43 ± 0.10</td>
<td>0.33 ± 0.12</td>
<td>-0.9 ± 0.03</td>
</tr>
<tr>
<td>AGF</td>
<td>-0.13 ± 0.07</td>
<td>-0.14 ± 0.07</td>
<td>-0.16 ± 0.09</td>
<td>-0.16 ± 0.17</td>
<td>-0.15 ± 0.17</td>
<td>-0.04 ± 0.00</td>
<td>-0.04 ± 0.06</td>
<td>0.15 ± 0.09</td>
<td>0.35 ± 0.33</td>
<td>0.12 ± 0.09</td>
</tr>
<tr>
<td>GSF</td>
<td>0.05 ± 0.08</td>
<td>-0.05 ± 0.00</td>
<td>0.04 ± 0.01</td>
<td>0.01 ± 0.07</td>
<td>0.01 ± 0.01</td>
<td>-0.13 ± 0.07</td>
<td>0.41 ± 0.06</td>
<td>-0.22 ± 0.05</td>
<td>0.17 ± 0.20</td>
<td>0.32 ± 0.13</td>
</tr>
<tr>
<td>Far</td>
<td>0.18 ± 0.14</td>
<td>0.33 ± 0.11</td>
<td>0.08 ± 0.08</td>
<td>-0.38 ± 0.07</td>
<td>0.17 ± 0.05</td>
<td>-0.36 ± 0.07</td>
<td>0.67 ± 0.02</td>
<td>0.51 ± 0.08</td>
<td>0.51 ± 0.04</td>
<td>0.25 ± 0.02</td>
</tr>
</tbody>
</table>

Count right (ZiRe); Count left (ZiLi); Rear Legs X-shaped-O-shaped (XoHi); Rear Legs steep–sickle hocked (SaHi); Rear Legs pastern low-steeplangle (FsHi); Rear Legs inner claw small-large(IkHi); No. of piglets born alive (LGF), No. of weaned piglets (AGF); Number of piglets during the lactation (GSF); Piglets survival (FAR).
4.5. Model Comparison in Univariate Analyses

In this study we investigated and compared model 3 and model 4. The model 4 was tested against model 3 by likelihood ratio test. The statistics of the likelihood ratio (LR\textsubscript{Model4,3}) test for sequentially reduced models (Rao, 1973) was computed as follows:

\[ LR_{Model4,3} = -2 \log_e \left( \frac{L_{Model4}}{L_{Model3}} \right) = 2 \log_e \left( L_{Model4} - L_{Model3} \right) = X^2 \sim X^2_{0.05,1} \]

It was assumed that the difference is distributed as Chi-Square with degrees of freedom equal to the difference in number of variances and/or covariance in the two models (3 and 4). The models were compared at 05 percent significance with 1 DF, with a critical value of 3.841. Likelihood-ratio tests (Table 4.13) indicated that applying genetic correlation among additive and maternal effect was almost significant for exterior traits across breeds. It was not that much trivial to say whether these differences are statistically important or are little more than random deviations. However, even though the likelihood-ratio test indicated that maternal genetic effects were unimportant, for some exterior traits across breeds, it is not easily possible to say that these effects should be excluded in estimation of genetic parameters. The direction of the bias on the heritability estimate depends partly on the size of the additive maternal variance and the size and sign of the maternal genetic covariance. In summary, the importance of covariance between maternal and additive effects differed by breed and trait. Table 4.13 shows this discrepancy across breeds. Including covariance between maternal and additive effects make much sense for exteriors trait. Also it can be also argued that number of observations can also affect this ratio, as the number of observation in SLR and SLW\textsubscript{SL} were less than SLW. We believe that this is one of the reasons why most of estimated log likelihood ratio tests showed high level of significance in SLW breed.
<table>
<thead>
<tr>
<th>Traits</th>
<th>Swiss-Large-White</th>
<th>Swiss-Landrace</th>
<th>Swiss-Large-White-Sire-Line</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>logL</td>
<td>LRModel4,3</td>
<td>logeL</td>
</tr>
<tr>
<td></td>
<td>Model4</td>
<td>Model3</td>
<td></td>
</tr>
<tr>
<td>ZiRe</td>
<td>-173752.2</td>
<td>-173757.2</td>
<td>5.0*</td>
</tr>
<tr>
<td>ZiLi</td>
<td>-139931</td>
<td>-140066</td>
<td>135.0***</td>
</tr>
<tr>
<td>XoHi</td>
<td>-72325.4</td>
<td>-72401.8</td>
<td>76.4***</td>
</tr>
<tr>
<td>SaHi</td>
<td>-99995</td>
<td>-99994</td>
<td>1.0 NS</td>
</tr>
<tr>
<td>FsHi</td>
<td>-12238.1</td>
<td>-12249.2</td>
<td>11.1***</td>
</tr>
<tr>
<td>IkHi</td>
<td>-53781.1</td>
<td>-53792.5</td>
<td>11.4***</td>
</tr>
<tr>
<td>LGF</td>
<td>-91139.0</td>
<td>-91185.1</td>
<td>46.1***</td>
</tr>
<tr>
<td>AGF</td>
<td>-83752.2</td>
<td>-83756.2</td>
<td>4.0*</td>
</tr>
<tr>
<td>GSF</td>
<td>-139931</td>
<td>-140066</td>
<td>135.0***</td>
</tr>
<tr>
<td>FAR</td>
<td>-53780.1</td>
<td>-53781.2</td>
<td>1.1 NS</td>
</tr>
</tbody>
</table>

**Table 4.13** Model comparison based on log likelihood ratio test within traits across breeds.

logL: Logarithm of models, LRModel4,3: log likelihood ratio test which assumed to be asymptotically distributed as chi-square with 1 degree of freedom. Model 3 includes direct genetic effects, and maternal genetic effects; model 4 is model 3 with the correlation between additive direct and maternal genetic effects added. *P < 0.05; **P < 0.01; ***P < 0.001; NS: Not significant.
4.6. Implications

The general implication of this study is the lack of interchangeable exterior and even reproductive traits names across publications and literatures. This means that referencing predicted genetic component for given trait in literatures was cumbersome work. Perhaps it could be that the same trait name is served for trait(s) that are quite far away from other researcher’s purpose. It was observed that underlying genetic potential of reproductive traits was not so high. Results showed that the overall genetic prediction (either, additive effect and maternal) effect was almost low across breeds. On the other hand, it is being indicated that including litter size as random effect gave arise to that additive genetic effect to be estimated downwardly. This was almost always observed for exterior traits rather than reproductive traits. Estimated of all heritability was almost low across all breeds and traits. Compared the estimated additive effect, estimated maternal genetic effect was substantial across breeds and traits. In this way, SLW was superior to SLR and SLWSL with respect to exteriors traits. But for reproductive traits, the two last breeds showed superiority. The maternal effects are may be important for some traits in some breeds and should be examined in large commercial herds. If important, they need to be included in genetic evaluation systems to get unbiased estimates of direct breeding values. Improvement of maternal response in addition to direct response can lead to greater overall response in a trait. Model comparison ended up with significant results for exterior trait and this was mostly observed for SLW, though other breeds showed some predisposition for reproductive traits too. However, though so far no maternal genetic variance or heritability have addressed for most investigated traits, a substantial genetic variance can be attributed to this component for some traits. As a result, it is recommended to run another study to check in if this component within other pig breeds exists to similar levels. We analyzed Gaussian-like distribution trait and categorical traits with standard mixed linear models (MLM). However some bias may also result from the use of standard mixed linear models (MLM) to analyze categorical traits such as ZiRe or ZiLi. To deal with categorical traits, threshold models have been shown to be superior to MLM when the number of categories or level of a given trait is small and at the same time corresponding frequencies for these categories are unequal. Canario, L. et al., (2006) argued that use of threshold models are still difficult, particularly to estimate genetic correlations between categorical and normally distributed traits. Software based on approximate methods often gives unexpected results, and exact methods based on stochastic inference such as Gibbs sampling e.g (Holm, B. et
al., 2004) are still computationally demanding. The use of standard MLM, which is known
to be robust to strong departures from normality assumptions, hence considered as a simple
and satisfactory method to obtain genetic parameters in this study. Estimated genetic
correlations- irrespective of kind of applied method- have always been an important part of
carefully constructed, sustainable, secure breeding programs, but the concept of a genetic
correlation is perhaps even less well understood than the concept of heritability. Partly since
the genetic correlations are difficult to estimate. Genetic correlation was substantial among
exterior traits (leg and teat traits), but small genetic correlation was observed among
exterior traits with reproductive traits. In model 2, we used litter size as random effect.
Researchers have different views in terms of treating litter size as random effect for
aforementioned traits. Across whole breeds, these results tell us that improving genetic
progress for exterior traits won’t imply any significant change on reproductive traits and
vice versa. Genetic correlations tell us how pairs of traits co-vary or change together. This
study indicated that reproductive traits show more complex behavior genetically than
exterior and teat traits. The estimated genetic correlation in this study can be sued to set up
secure breeding programs.

4.7. Acknowledgements

Authors thank SUISAG Company for providing data, Professor Wolfgang Langhans for his
support, Professor L. Dale Van Vleck for his instant help and Dr. Juan Pablo Gutierrez for
his insightful comments. Also thanks for Dr. Andreas Hofer for his comments on very early
estimation on this data.

4.8. Appendix A

Estimated additive genetic trend has delineated for traits over different year of birth of
animals (Figure 4. 1 to Figure 4. 3). We included base animal as well to get clue of the
magnitude of estimated breeding values (EBVs) of animal for different traits from base
population. For all EBVs, a simple linear model with time (year of birth) as regressor was
fitted. This is being indicated one the top of left hand side of each figure. R2 indicates how
fit this regression is for each trait. This is very important to be mentioned that these trends
were made by lumping over the whole breeds into one dataset and also different fixed
effects were used to calculate them. In general EBVs is function of breed, that is, it must be
estimated within breed not across breeds. However, monotonic decreasing patterns of EBVs
over time can be seen for some reproductive traits (i.e. GSF). Therefore, reproductive traits require to be more investigated.

Figure 4.1 Estimated breeding values for teat traits.

Figure 4.2 Estimated breeding values for leg traits.
In the above we pinpointed that across all breeds, estimated genetic correlation among FAR trait and other traits frequently obtained out of expected parameter space. This led us to explore data structure using clustering algorithms and principal component analysis. In general clustering is built up due to result of some distance functions. There are many different distance functions. However, given two p-dimensional data objects \( i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \) and \( j = (x_{j1}, x_{j2}, \ldots, x_{jp}) \) the following common distance functions was used:

**Euclidian Distance Function:**

\[
d(i, j) = \sqrt{|X_{i1} - X_{j1}|^2 + |X_{i2} - X_{j2}|^2 + \ldots + |X_{ip} - X_{jp}|^2}
\]

**Manhattan Distance Function:**

\[
d(i, j) = |X_{i1} - X_{j1}| + |X_{i2} - X_{j2}| + \ldots + |X_{ip} - X_{jp}|
\]

In our analysis, \( P=10 \) (number of traits). The traits were normalized to lie between 0 and 1. Before proceeding along, we should mention that from statistical point of view, this is the likelihood of data which impose kind of algorithm which should be learned over a given data. When using the Euclidian distance function to compare distances, it is not necessary to calculate the square root because distances are always positive numbers. Clustering was done using cluster package in R. In learning hierarchical clustering, we constructed 10 kinds of clusters due to combination of clustering algorithms and distance functions. Only
results of two of them are given at Figure 4.4. Across whole clustering algorithms, the results indicated that, irrespective kind of clustering algorithms and applied distance function (Euclidian, Manhattan), piglets survival (Far) trait was clustered into one singleton cluster. It suggests that FAR cannot be clustered with any other traits. This could be maybe because of this fact that this trait is a percentage-measured trait. Results of average agglomerative clustering algorithm with either Euclidian and Manhattan distance function turned out to be highly similar. In other words, applying Euclidian or Manhattan distance function did not sharply affect the final results of clustering on mixed data. For some algorithms the topology of clustering slightly changed with applying Manhattan instead of Euclidian distance function. Teat-related traits across all clustering learning processes were cohesively tightly clustered together. This likely can be interpreted in this way that estimated breeding values (EBVs) for one teat trait can be used for other one. Therefore, there is no need to separately estimate EBVs for each teat trait. In other words, the results can pinpoint to this reality that maybe teat traits are governed almost by the same genetic background (already we saw that they had genetic correlation close to one). From genetic analysis point of view the results indicated that simultaneously running genetics analysis for these traits (exterior, teat and reproductive traits) will end up with some problems. As a result, it would be better to run pair-base genetic correlation analysis across traits. Our experience with analysis this data support above argument. In most genetic correlation analysis of other traits with FAR trait, the results failed to be meaningful and fallen within the parameter space. One of the drawbacks of learning clustering algorithms for our dataset could be the small size of dimensionality (ten) of dataset which used in this study. In general, clustering is used for high-dimensional dataset. However still we assume that the results can reveal the structure among traits and can show some general idea about the hidden structure pattern among pig exterior and reproductive traits. It seems that agglomerative approach has more promising application in breeding setting. In breeding setting we are going to see how traits are related. Clustering approach can provide prior knowledge for researcher before running genetic correlation in high-dimensional data analysis, since genetic analysis can be run based on more phonetically cohesively clustered traits. Likely, in cases that traits are invariant to algorithms and distance functions, prior knowledge about one cluster can be used for other one. In this way we reduce the burden of computations. Based on our idea about data, we suggest that complete method algorithm (Figure 4.4) can likely visualize the hidden topology structure among traits.
We applied two kinds of principal component analyses (PCA) to data as well. The PCA were based on correlation and covariance among observations. Table 4.14 and Table 4.15 show, there is no straightforward relationship between the PCs obtained from a correlation matrix and those based on and the corresponding covariance matrix.

<table>
<thead>
<tr>
<th>Traits</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZiRe</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.36</td>
<td>0.72</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td>ZiLi</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.57</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>xoHi</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.00</td>
<td>-0.86</td>
<td>-0.63</td>
<td>0.39</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SaHi</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.13</td>
<td>-0.21</td>
<td>0.15</td>
<td>-0.02</td>
<td>-0.22</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>FsHi</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>1.08</td>
<td>-0.46</td>
<td>0.30</td>
<td>0.04</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>IkHi</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.16</td>
<td>-0.11</td>
<td>0.00</td>
<td>-0.20</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LGF</td>
<td>1.11</td>
<td>-5.15</td>
<td>-1.43</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AGF</td>
<td>-2.14</td>
<td>-3.42</td>
<td>0.89</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.02</td>
</tr>
<tr>
<td>GSF</td>
<td>0.98</td>
<td>-4.04</td>
<td>1.07</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>FAR</td>
<td>-26.95</td>
<td>-0.09</td>
<td>-0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>STDE</td>
<td>8.56</td>
<td>2.07</td>
<td>0.66</td>
<td>0.46</td>
<td>0.30</td>
<td>0.27</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>PROV</td>
<td>0.93</td>
<td>0.05</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CUMP</td>
<td>0.93</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4.14 Estimation of principal components based on variance covariance matrix among traits.
Table 4.15 Estimation of principal components based on correlation matrix among traits. In the above tables, STDE means Standard deviation; PROV means Proportion of Variance and CUMP means Cumulative Proportion.

We can see that each of the first four PCs for the correlation matrix has moderate-sized coefficients for several of the traits, whereas the first four PCs for the covariance matrix are each dominated by a single trait. The first component is a slight perturbation of the single trait FAR, which has the largest variance; the second component is almost the same as the trait LGF with the second highest variance; and so on. In fact, this pattern continues for the fifth and sixth components, which are shown in (Table 4.15). Also, the relative percentages of total variation accounted for by each component closely mirror the variances of the corresponding traits. In PCA due to covariance matrix, the five first principal components explained about 99% of the total variation (Table 4.14) whereas for correlation matrix this value was 73% (Table 4.15). In other words sharp deduction in dataset dimensionality (from ten to five) can be seen retaining most information. Because trait FAR has a variance several times larger than any other variable, the first PC accounts for over 93% of the total variation. Thus the first six components for the covariance matrix tell us almost nothing apart from the order of sizes of variances of the original traits. By contrast, the first few PCs for the correlation matrix show that certain non-trivial linear functions of the (standardized) original traits account for substantial, though not enormous, proportions of the total variation in the standardized traits. In particular, a weighted contrast
between the first four and the last four traits is the linear function with the largest variance. This data illustrates the dangers in using a covariance matrix to find PCs when the variables have widely differing variances; the first few PCs will usually contain little information apart from the relative sizes of variances, information which is available without a PCA. To interpret the PCs in terms of the original traits each coefficient must be divided by the standard deviation of the corresponding trait. In our view, results of PCA supported clustering results and vice versa. Results of clustering indicated that FAR trait didn’t cohesively cluster with other traits. The bottom line is here that there are small portions of information in data studied here. Therefore, if we see it in large scale genetic evaluation, without proof of principle, multivariate genetic evaluation for this data will be likely failed. In multivariate genetic analysis, we are dealing with big design matrices in which when there are dependencies among rows or columns (small number of independent information), rank of matrix will be reduced. In this case, not explicitly, researcher must apply useful intelligent algorithm to solve system of linear equations. However, soundly using results of PCA based on correlation matrix in genetic evaluations can reduce the computational burden of genetic evaluation.
Chapter 5

Bayesian Major Gene Segregation Analysis of Carcass Length Trait across Three Pigs Breeds

5.1. Abstract

The objective of this study was to establish and to investigate the mode of inheritance of carcass length in three pig breeds. These breeds were Swiss-Large-White (SLW); Swiss-Landrace (SLR) and Swiss-Large-White-Sire-Line (SLWSL). Mean and standard deviation of carcass length were 96.56 ± 2.65, 99.33 ± 2.22 and 97.00 ± 2.67 for SLW, SLR and SLWSL respectively. Gibbs sampling algorithm was applied to sample marginal posterior densities of the model parameters in a mixed inheritance model. Segregation analysis showed evidence of a major gene in SLW and SLR breeds but not for SLWSL breed. The magnitude of dominance effect of major gene across breeds was substantial. Estimated non-parametric posterior additive major gene effect and posterior standard deviation were 0.26 ± 0.23 and 0.66 ± 0.31 for SLW and SLR respectively. Also, the estimated non-parametric posterior dominance major gene effect and posterior standard deviation were 1.04 ± 0.49 and 0.37 ± 0.44 and for SLW and SLR respectively. These results indicate the need for molecular detection of major genes for carcass length traits in SLW and SLR breeds for possible inclusion in the pig breeding programs.

5.2. Introduction

Given the mixed inheritance multiplicative model, generally two different statistical paradigms (Bayesian and likelihood) can be used to predict the evidence of major gene for a given vector of phenotypic measurements (Sanchez, M.P. et al., 2003). However, in case of likelihood based methods, generally large sire families are required to estimate the genetic parameters than Bayesian analysis. In Bayesian Gibbs sampling, genetic trend either for polygene effect and major gene effect can be calculated readily over generation (Sanchez, M.P. et al., 2003). In the context of animal genetics, Bayesian complex segregation analysis was first introduced by (Hoeschele, I. 1988). Latter on, Janss, L.L.G et
al., (1995, 1997) applied Gibbs sampling within a Bayesian framework to investigate major
gene affecting meat quality traits in a crossed F2 population based on a mixed inheritance
model in animal breeding. The Bayesian Gibbs sampling methods of Janss, L.L.G. et al.,
(1995) have been used by researchers in detection of major genes for osteochondral
diseases in pigs (Kadarmideen, H. N. et al., 2005), milking speed in dairy cattle (Ilahi, H. et
al., 2004), for stress in pigs (Kadarmideen, H. N. et al., 2007), egg number and quality in
poultry (Hagger, C. et al., 2004), body composition, carcass, meat quality and the number
of false teats in pig (Sanchez, M.P. et al., 2003) and for carcass traits in dairy cattle
(Miyake, T. et al., 1999). Carcass traits in pig genetics evaluation programs have received
extensive attention over last decays (Olesen, I. et al., 2000; Kuhlers, D.L. et al., 1992;
Kuhlers, D.L. et al., 1993; Bereskin, D.L. et al., 1988; Sanchez M.P. et al., 2003). Here we
investigate an autosomal major gene which affects carcass traits with aim of Bayesian
Gibbs sampling methods of (Janss, L.L.G et al., 1995). As issued in relevant literatures
which some of them issued here, breeding goals in pigs are subject to change and are
directed much more toward retail carcass yield and meat quality because of the high
economic value of these traits. The genetic correlation (unfavorable) of pig carcass trait
with other economically motivated traits emphasis that detection of major gene is important
(Serenius, T. et al., 2004). The goal of this study were to detect a bi-allelic autosomal major
genes and estimate associated genetic parameters which affect carcass length across Swiss-
Large-White(SLW); Swiss-Landrace(SLR) and Swiss-Large-White-Sire-Lines(SLW SL)
breeds using a Bayesian complex segregation analysis and to compare these three breeds.

5.3. MATERIALS AND METHODS

5.3.1. Data

Data used in this study were obtained from the Suisag' company located in Switzerland.
General information is given in Table 5. 1. Relevant pedigree for each breed containing
seven generations was constructed. Within each breed, data were edited for missing values,
descriptive statistics and departures from normality. Since normality is central to run
segregation analysis, observations that showed fives standard deviation from the mean,
were not included in analysis.
### Table 5.1 General information about the data and records and number of fixed effects which used in this study.

<table>
<thead>
<tr>
<th>Breeds</th>
<th>Carcass length trait</th>
<th>Fixed effects used in the model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Swiss-Large-White(SLW)</td>
<td>96.59</td>
<td>2.65</td>
</tr>
<tr>
<td>Swiss-Landrace(SLR)</td>
<td>99.33</td>
<td>2.22</td>
</tr>
<tr>
<td>Swiss-Large-White-Sire-Line(SLWSL)</td>
<td>97.00</td>
<td>2.67</td>
</tr>
</tbody>
</table>

Trait distribution was examined for each breed before analyses (**Figure 5.1, Figure 5.2** and **Figure 5.3**). For all breeds, estimated empirical distribution for carcass length trait agreed with estimated theoretical normal distribution.

**Figure 5.1** Theoretical normal distribution and empirical normal distribution of carcass length distribution in SLW breed.

**Figure 5.2** Theoretical normal distribution and empirical normal distribution of carcass length distribution in SLR breed.
Figure 5.3 Theoretical normal distribution and empirical normal distribution of carcass length distribution in SLW<sub>SL</sub> breed.

5.3.2. Box-Cox Transformation

Major gene segregation analysis could be sensitive to deviations from Gaussian distribution of the trait, when such deviations are not caused by a single gene (Shizhong, Xu. et al., 2004). Data were one parameter Box-Cox transformed. This was done to remove noise (i.e. skewness) that may lead to the false inference of a major gene in the current study. Following one-parameter Box-Cox transformation was used (Charles, D. et al., 2007):

\[
y^{(\lambda)} = \begin{cases} 
  \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\
  \ln y & \lambda = 0
\end{cases}
\]

In the Box-Cox transform, \( \lambda \) controls the degree and direction of skew that can be removed from given distribution. Box-Cox transformation has been used in segregation analysis (Szydlowski, M. et al., 2001; Sanchez, M. et al., 2003). The optimal value of \( \lambda \), obtained by maximizing the log-likelihood function:

\[
l(y) = -\frac{n}{2} \ln \left[ \ln \sum_{i=1}^{n} \left( y^{(\lambda)}_i - \bar{y}^{(\lambda)} \right) \right] + (\lambda - 1) \sum_{i=1}^{n} (\ln y_i)
\]

Where \( i \) indexes the \( n \) observations and: \( \bar{y}^{(\lambda)} = \frac{1}{n} \sum_{i=1}^{n} y^\lambda_i \). To make \( y^{(\lambda)}_i \) scale-invariant,

\[
y^{(\lambda)}_k = \begin{cases} 
  \frac{y^\lambda - k^\lambda}{\lambda} & \lambda \neq 0 \\
  \ln(y/k) & \lambda = 0
\end{cases}
\]

used which \( k > 0 \) is an arbitrary constant in the same measurement units as \( y \). Major gene effects (dominance, additive and deviation b/w additive and dominance effects) presented later were back-transformed to the original scale using an inverse Box-Cox transformation.
(Charles, D. *et al.*, 2007). One reason for applying this Box-transformation was that the adjusting raw data for fixed effects slightly decreased absolute skewness but increased kurtosis which makes distribution more leptokurtic. This phenomenon was markedly observed for SLWSL breed’s records for carcass length.

### 5.3.3. Notation and Assumptions

In this study it was presumed that a bi-allelic autosomal major gene influencing on carcass length, had two alleles A1 and A2, with frequencies q and p (p = 1-q). Where p is the estimate of A1 allele frequency in the founder population. The Mendelian transmission probabilities for this autosomal gene’s allele was reserved (no relaxation was used). The genotypes A1A1, A1A2, and A2A2 were assumed to occur in Hardy-Weinberg frequencies, with q^2, 2q(1-q) and (1-q)^2 genotypes frequencies in the founder population. Also it assumed that corresponding above genotypes effects on carcass length trait with a, d, and –a respectively. Allele A1 is defined to decrease the phenotypic value, and allele A2 is defined to increase the phenotypic value.

### 5.3.4. Complex Segregation Analysis

A mixed inheritance model was used to analyze the presence of a major gene affecting the carcass length. The mixed inheritance model included non-genetic fixed and random effects, random polygenic effects and fixed effects of a major gene. The following mixed inheritance model was applied, as shown in (Kadarmideen, H. N. *et al.*, 2005; Ilahi, *et al.*, 2004).

\[
y = X\beta + Zu + ZWm + e
\]

where \(y\) is vector of observation, \(X\) and \(Z\) are the known incidences matrices, \(\beta\) is the vector of fixed effects, \(u\) is the vector of random polygenic effects, \(W\) is the matrix containing genotypes of each individual, \(m\) is a vector of mean genotypic values, and \(e\) is a vector of random residuals. \(Wm\) is a vector of random effects at the single locus. We assumed that the conditional distribution of the data vector \(y\) given relevant components is multivariate normal as:

\[
y | \beta, u, W, m \sim N(X\beta + Zu + ZWm, \sigma^2_e)
\]

Given the model, the marginal non-parametric density posterior distributions of the following components were estimated:
\[ \text{pr}(y | \mu, \sigma^2_c, \sigma^2_{po}, a, d, p, f, m, I) \sim \text{pr}(\mu, \sigma^2_c, \sigma^2_{po}, a, d, p, f, m, I | y) \]

Which \( \mu, \sigma^2_c, \sigma^2_{po}, a, d, p, f, m, I \) were general mean of model; error variance; variance of polygenic effect; additive major gene effect, dominance major gene effect; frequency of major gene of allele A1, fixed effects, genotypes of major gene and polygenic effects respectively. Residual effects were assumed to be distributed as \( e \sim N(0, \sigma^2_e) \) which \( I \) is identity matrix. Uniform prior distributions were assumed in the range \((-\infty, +\infty)\) for fixed effects and for additive and dominance major gene effect effects, also uniform prior distributions were assumed in range \((0, +\infty)\) for genetic variance components, and in the range \([0, 1]\) for allele frequencies. Using the above estimated marginal components, the marginal non-parametric density posterior distributions of following components were computed as follows:

\[
\begin{align*}
\sigma^2_a &= (2pq)(a + d(q - p))^2 \\
\sigma^2_d &= (2pqd)^2 \\
\sigma^2_g &= \sigma^2_a + \sigma^2_d \\
\sigma^2_{po} &= \sigma^2_{po} + \sigma^2_e \\
\sigma^2_h &= \sigma^2_{po} + \sigma^2_a + \sigma^2_d + \sigma^2_g \\
\end{align*}
\]

\( \sigma^2_a, \sigma^2_d, \sigma^2_g, \sigma^2_{po}, \sigma^2_h \) were variance of additive major gene effect, variance of dominance major gene effect; variance of major gene effect, heritability of polygenic effects and total heritability respectively. The vector of hyper-parameters (\( \theta \)) included the fixed effects (\( \beta \)), polygenic effects (\( u \)), residual variance, polygenic variance (\( \sigma^2_{po} \)), major gene effects (\( m \)), and major gene frequency (\( q \)) were reserved as it is indicated in (3).

5.3.5. Gibbs Sampling Algorithm

Markov Chain Monte Carlo (MCMC) sampling methods provide an efficient means to carry out complex integration numerically. In particular the Gibbs sampler, an MCMC method, is now widely used in genetic analyses (Navarro, P. et al., 2006; Kadarmideen, H. N. et al., 2005; Ilahi, H. et al., 2004). It is capable of generating samples from the joint distribution (usually complex) of several random variables by sampling from known and simple conditional distributions. From these samples, marginal distributions of each
variable can be obtained and provide estimates of the posterior distributions of the model parameters. In other words, Gibbs sampling is a Markov Chain Monte Carlo (MCMC) scheme for simulation from \( \pi(x) \) where transition kernel is formed by the full conditional distributions of \( \pi(x) \). In this study we directly sampled five parameters, in doing so it is being assumed that the full conditional distributions \( \pi_i(\theta_i | \theta_{-i}) \) \( \forall i = 1,\ldots,8 \) be known to the extent that it is possible to simulate from them. The Gibbs sampler algorithm is expanded as follows:

\[
\sigma^2_{i} \sim \pi(\sigma^2 | y, \sigma^2_{po}, a^t, d^t, p^t, f^t, m^t, I^t)
\]

\[
\sigma^2_{po} \sim \pi(\sigma^2_{po} | y, \sigma^2_{e}, a^t, d^t, p^t, f^t, m^t, I^t)
\]

\[
a^t \sim \pi(a | y, \sigma^2_{e}, \sigma^2_{po}, d^t, p^t, f^t, m^t, I^t)
\]

\[
d^t \sim \pi(d | y, \sigma^2_{e}, \sigma^2_{po}, a^t, p^t, f^t, m^t, I^t)
\]

\[
p^t \sim \pi(p | y, \sigma^2_{e}, \sigma^2_{po}, a^t, d^t, f^t, m^t, I^t)
\]

\[
f^t \sim \pi(f | y, \sigma^2_{e}, \sigma^2_{po}, a^t, d^t, m^t, I^t)
\]

\[
m^t \sim \pi(m | y, \sigma^2_{e}, \sigma^2_{po}, a^t, d^t, f^t, I^t)
\]

\[
I^t \sim \pi(I | y, \sigma^2_{e}, \sigma^2_{po}, a^t, d^t, f^t, m^t)
\]

Which, \( t \) indicates the number of iteration per chain. This algorithm with blocked sampling of genotypes was used for inference in the mixed inheritance model and implemented using the MAGGIC software package developed by (Janss, L.L.G et al., 1995). 20 replicates of Gibbs chains of 100000 cycles were run for analysis, using a spacing of 50 cycles, obtaining 2000 Gibbs samples per chain and 100000 samples in total for each trait. A burn-in period of 1000 cycles was used to allow the Gibbs chains to reach equilibrium. Marginal non-parametric density posterior distributions of aforementioned components were directly sampled from the model.

**5.4. Results and Discussion**

Identification of major genes affecting carcass length trait in pig populations could have a considerable impact on genetic improvement, for example, by increasing the accuracy of genetic selection. Many researcher addressed QTLs or chromosome locations which harbor of genetic variations carcass/meat quality trait across populations. In this study we referred to some of them though it could be that they have treated differently in their context (Malek, M. et al., 2001). There are no stringent rules to make distinguish between these traits. **Table 5.1** shows descriptive statistics of the carcass length trait across the three breeds. **Table 5.2** , **Table 5.3** and **Table 5.4** show corresponding estimated non-parametric
marginal posteriors distributions for each breed. For carcass trait within each breed twelve parameters were sampled. Estimated point estimate for polygenic heritabilities across breeds were substantial (Table 5.2, Table 5.3 and Table 5.4). In this way, SLWSL showed more superiority ($h^2_{po} = 0.63 \pm 0.14$). Sanchez, et al., (2003) estimated a range of 0.15 to 0.42 polygenic heritabilities for number of different meat and carcass traits in pig line. This is lower of our estimation for carcass length traits across almost all breeds. In genetic parameter estimation by REML approach to estimate polygenic components, similar estimates were observed. This can likely suggest that we have the same structure of genetic data and Bayesian and REML approaches work similarly in estimation of parameters of interest. Above tables shows marginal posterior estimates for additive affect (a), dominance effect (d), and the difference between additive and absolute values of dominance effects (a-abs (d)) at the single locus. For SLWSL breed (a-abs(d)) likely suggest the strong dominance effect of gene action in this breed. The (a-abs(d)) for carcass length trait was not that much high in both SLW and SLR breeds. However, it can be seen that the probability that carcass length in this breed (SLWSL) is under dominance control. It can be argued that the segregation of major genes in a SLWSL breed has undergone a long-term additive genetic selection that this could result from non-additive gene action or associated effects for carcass trait for this breed. However, we assumed that we didn’t have any genetic selection in the base population. The frequencies of unfavorable alleles were not relatively high across all breeds. This can per see show the complexity of gene effect mechanism for carcass trait in SLWSL breed. The reason for the lack of existence of segregating gene for carcass trait in SLWSL breed is difficult to assess. There was no extra data to assess the linkage and pleiotropic for this trait. The method applied gave no information about the linkage or pleiotropic effects, and only weak suggestions can be made here. A possible explanation for the existence of genes of large effects in these breeds could be due to pleiotropy or linkage with other genes. Associated effects may be involved in general fitness controlled by natural selection (Szydlowski, M. et al., 2001). Multiple traits major gene segregation analysis of other traits with carcass length trait can show the existence of genes affecting traits that are under the antagonistic pressures of artificial and natural selection. Exploring that kind of analysis maybe can shed a light on the differences among studied breeds in this investigation in terms of presence of major genes. Gene’s allele frequencies were calculated from the estimates of allele frequencies assuming the Hardy-Weinberg proportions. Almost in all breeds, all frequencies were similar this matter can be seen in Figure 5.5, Figure 5.8 and Figure 5.11. The results of this study confirm
that segregation is possible for genes with large effects in the SLW and SLR breeds since their HDP of variance of major gene is not included zero value. The other figures, that is, Figure 5.4, Figure 5.6, Figure 5.7, Figure 5.9, Figure 5.10, Figure 5.12 shed more light on computational task across breeds. It must be stressed that the mean of the marginal posterior distribution of the variance of major gene obtained at different level of the highest posterior density region in percent (HPD%). For SLW breed, we used HPD at 95% (HPD95%), for SLR breed we used 95% (HPD95%), and for SLWSL breed we used HPD at 93% (HPD93%). The idea by applying different HPD was to see if we can find any sign of major gene even with more prone to increase the level false positive rate. More concept pertaining to HDP can be found at (Kadarmideen, H. N. et al., 2005, 2007).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Posterior Means</th>
<th>Posterior Standard Deviation</th>
<th>95%HPD(left)</th>
<th>95%HPD(right)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_e^2$</td>
<td>1.28</td>
<td>0.68</td>
<td>0.05</td>
<td>2.47</td>
</tr>
<tr>
<td>$\sigma_{po}^2$</td>
<td>5.68</td>
<td>1.04</td>
<td>3.67</td>
<td>7.74</td>
</tr>
<tr>
<td>$\sigma_a^2$</td>
<td>0.15</td>
<td>0.04</td>
<td>0.11</td>
<td>0.23</td>
</tr>
<tr>
<td>$\sigma_d^2$</td>
<td>0.17</td>
<td>0.16</td>
<td>0.12</td>
<td>0.63</td>
</tr>
<tr>
<td>$\sigma_g^2$</td>
<td>0.23</td>
<td>0.22</td>
<td>0.10</td>
<td>0.74</td>
</tr>
<tr>
<td>$h_{po}^2$</td>
<td>0.61</td>
<td>0.11</td>
<td>0.62</td>
<td>0.91</td>
</tr>
<tr>
<td>$h_i^2$</td>
<td>0.65</td>
<td>0.11</td>
<td>0.60</td>
<td>0.99</td>
</tr>
<tr>
<td>$q$</td>
<td>0.46</td>
<td>0.04</td>
<td>0.39</td>
<td>0.54</td>
</tr>
<tr>
<td>$p$</td>
<td>0.54</td>
<td>0.04</td>
<td>0.46</td>
<td>0.61</td>
</tr>
<tr>
<td>$a$</td>
<td>0.26</td>
<td>0.23</td>
<td>-0.12</td>
<td>0.71</td>
</tr>
<tr>
<td>$d$</td>
<td>0.26</td>
<td>0.30</td>
<td>-1.25</td>
<td>1.79</td>
</tr>
<tr>
<td>$a - abs(d)$</td>
<td>-0.43</td>
<td>0.49</td>
<td>-1.45</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 5.2 Estimated non-Parametric posterior distributions of means, standard deviation, right and left bound of 95% highest posterior density (HPD) for carcass length in SWL breed.

$\sigma_e^2$: Error Variance; $\sigma_{po}^2$: Variance of polygenic effect; $\sigma_a^2$: Variance of additive major gene effect; $\sigma_d^2$: Variance of dominance major gene effect; $\sigma_g^2$: Variance of major gene effect $h_{po}^2$: heritability of polygenic effects; $h_i^2$: total heritability; $q$: frequency of major gene of allele A2; $p$: frequency of major gene of allele A1; $a$: additive effect of major gene; $d$: dominance effect of major gene; $a - abs(d)$: deviation b/w additive and dominance effects; HPD: Highest Non-Parametric Posterior Density

An advantage of the Bayesian approach through MCMC procedures is the possibility of easy construction of all kinds of confidence intervals (Bayesians called “credibility intervals”). The means of the marginal non-parametric posterior distribution of the additive effects (a) and corresponding credible interval have been given in related tables. The largest
additive effect for carcass length was found in SLR breed. The additive effect was lower for carcass length in SLW breed. This simply indicates that implementing more in-depth study in terms of genetic marker and expression analysis is sense in SLR and SLW breeds.

**Figure 5.4** Marginal non-parametric posterior distributions of error variance and polygenic variance for SLW breed.

**Figure 5.5** Marginal non-parametric posterior distributions of major gene allele frequencies and polygenic heritability for SLW breed.

**Figure 5.6** Marginal non-parametric posterior distributions of major additive and dominance effect for SLW breed.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Posterior Means</th>
<th>Posterior Standard Deviation</th>
<th>95%HPD(left)</th>
<th>95%HPD(right)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_e^2$</td>
<td>4.08</td>
<td>0.36</td>
<td>3.45</td>
<td>4.68</td>
</tr>
<tr>
<td>$\sigma_{po}^2$</td>
<td>1.15</td>
<td>0.26</td>
<td>0.79</td>
<td>1.58</td>
</tr>
<tr>
<td>$\sigma_a^2$</td>
<td>0.24</td>
<td>0.21</td>
<td>0.18</td>
<td>0.55</td>
</tr>
<tr>
<td>$\sigma_d^2$</td>
<td>0.33</td>
<td>0.25</td>
<td>0.14</td>
<td>0.68</td>
</tr>
<tr>
<td>$\sigma_g^2$</td>
<td>0.57</td>
<td>0.37</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>$h_{po}^2$</td>
<td>0.52</td>
<td>0.05</td>
<td>0.14</td>
<td>0.30</td>
</tr>
<tr>
<td>$h_t^2$</td>
<td>0.24</td>
<td>0.05</td>
<td>0.16</td>
<td>0.52</td>
</tr>
<tr>
<td>$q$</td>
<td>0.48</td>
<td>0.03</td>
<td>0.44</td>
<td>0.53</td>
</tr>
<tr>
<td>$p$</td>
<td>0.52</td>
<td>0.03</td>
<td>0.47</td>
<td>0.56</td>
</tr>
<tr>
<td>$a$</td>
<td>0.66</td>
<td>0.31</td>
<td>0.11</td>
<td>0.89</td>
</tr>
<tr>
<td>$d$</td>
<td>1.04</td>
<td>0.49</td>
<td>0.30</td>
<td>1.90</td>
</tr>
<tr>
<td>$a - abs(d)$</td>
<td>-0.40</td>
<td>0.44</td>
<td>-1.14</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table 5.3 Estimated non-parametric posterior distributions of means, standard deviation, right and left bonds of 95% highest posterior density (HPD) for carcass length in SLR breed.

$\sigma_e^2$: Error Variance; $\sigma_{po}^2$: Variance of polygenic effect; $\sigma_a^2$: Variance of additive major gene effect; $\sigma_d^2$: Variance of dominance major gene effect; $\sigma_g^2$: Variance of major gene effect; $h_{po}^2$: heritability of polygenic effects; $h_t^2$: total heritability; $q$: frequency of major gene of allele A2; $p$: frequency of major gene of allele A1; $a$: additive effect of major gene; $d$: dominance effect of major gene; $a - abs(d)$: deviation b/w additive and dominance effects, HPD: Highest Non-Parametric Posterior Density
Figure 5. 7 Marginal non-parametric posterior distributions of error variance and polygenic variance for SLR breed.

Figure 5. 8 Marginal non-parametric posterior distributions of major gene allele frequencies and polygenic heritability for SLR breed.

Figure 5. 9 Marginal non-parametric posterior distributions of major additive and dominance effect for SLR breed.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Posterior Means</th>
<th>Posterior Standard Deviation</th>
<th>93% HPD (left)</th>
<th>93%HPD(right)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{e}$</td>
<td>1.68</td>
<td>0.79</td>
<td>0.14</td>
<td>3.06</td>
</tr>
<tr>
<td>$\sigma^2_{po}$</td>
<td>4.69</td>
<td>1.20</td>
<td>2.53</td>
<td>7.05</td>
</tr>
<tr>
<td>$\sigma^2_{a}$</td>
<td>0.17</td>
<td>0.23</td>
<td>0.00</td>
<td>0.64</td>
</tr>
<tr>
<td>$\sigma^2_{d}$</td>
<td>0.25</td>
<td>0.31</td>
<td>0.00</td>
<td>0.90</td>
</tr>
<tr>
<td>$\sigma^2_{g}$</td>
<td>0.47</td>
<td>0.41</td>
<td>0.00</td>
<td>1.29</td>
</tr>
<tr>
<td>$h^2_{po}$</td>
<td>0.63</td>
<td>0.14</td>
<td>0.49</td>
<td>0.99</td>
</tr>
<tr>
<td>$h^2_{g}$</td>
<td>0.71</td>
<td>0.14</td>
<td>0.46</td>
<td>0.96</td>
</tr>
<tr>
<td>$q$</td>
<td>0.49</td>
<td>0.04</td>
<td>0.42</td>
<td>0.57</td>
</tr>
<tr>
<td>$p$</td>
<td>0.51</td>
<td>0.04</td>
<td>0.43</td>
<td>0.58</td>
</tr>
<tr>
<td>$a$</td>
<td>0.46</td>
<td>0.35</td>
<td>-0.02</td>
<td>1.13</td>
</tr>
<tr>
<td>$d$</td>
<td>0.37</td>
<td>0.44</td>
<td>-1.43</td>
<td>2.12</td>
</tr>
<tr>
<td>$a - abs(d)$</td>
<td>-0.37</td>
<td>0.72</td>
<td>-1.82</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 5.4 Estimated non-parametric posterior distributions of means, standard deviation, right and left bonds of 93% highest posterior density (HPD) for carcass length in SLW SL breed.

$\sigma^2_{e}$: Error Variance; $\sigma^2_{po}$: Variance of polygenic effect; $\sigma^2_{a}$: Variance of additive major gene effect; $\sigma^2_{d}$: Variance of dominance major gene effect; $\sigma^2_{g}$: Variance of major gene effect, $h^2_{po}$: heritability of polygenic effects; $h^2_{g}$: total heritability; $q$: frequency of major gene of allele A2; $p$: frequency of major gene of allele A1; $a$: additive effect of major gene; $d$: dominance effect of major gene; $a - abs(d)$: deviation b/w additive and dominance effects; HPD: Highest Non-Parametric Posterior Density.

In this study there was no evidence of segregating major gene for carcass length in SLW SL breed since HDP of major variance contained zero value. Box-Cox transformation and comparing theoretical and empirical distributions and also screened missing values were used to remove systematic and noise effect that sought to be central in addressing truly estimated major gene evidence. Having all of them, we argue most likely the small data size used in this study could give arise to non-evidence of major gene in SLW SL breed.
Figure 5. 10 Marginal non-parametric posterior distributions of error variance and polygenic variance for SLW$_{SL}$ breed.

Figure 5. 11 Marginal non-parametric posterior distributions of major gene allele frequencies and polygenic heritability for SLW$_{SL}$ breed.

Figure 5. 12 Marginal non-parametric posterior distributions of major additive and dominance effect for SLW$_{SL}$ breed.

The results also support differences between statistical methods: Bayesian Gibbs sampling approach generally gave lower values of additive and dominance effects than REML (Sanchez, M. P. et al., 2003). Failure to detect major gene for one of the breeds could be due to small family size, as most of the genetic information contained in the pedigree
concerned the segregation of genes within half-sib families. Yet, departures from other assumptions, such as the homogeneity of the distribution of residuals among families, may also lead to the false inference of a major gene. Having estimated different major genes, pleiotropic effect can be construed and in sound experiment. For SLW\textsubscript{SL} breed, it was indicated that dominance effect played likely a role in underlying genetic mechanism of carcass length trait. Kacser, H. \textit{et al.}, (1981) made the argument that dominance in metabolic pathways is an inevitable property of multi-enzyme systems and could not be significantly modified. As a consequence, they argued that dominance in metabolism does not require an evolutionary explanation (Bagheri, H. C. \textit{et al.}, 2004). The results from segregation analyses should therefore be considered as first indications of the presence of property of multi-enzyme systems due to dominance deviation which have then should be confirmed by analyses using molecular tools. To what extend argumentation of evolutionary genetics can be useful in genetic breeding/setting, is not quite clear, but we can be sure that dominance effect of segregating major gene can play major role and must be considered in genetic selection program. This effect must be wisely used in genetic program, because of its role in property of multi-enzyme systems.

5.5. Implications

In this study, we investigated an evidence for segregating major gene for carcass length in 3 Swiss pig breeds. We have found an evidence for an autosomal- bi-allelic gene influencing carcass length in SLW and SLR breeds but not SLW\textsubscript{SL} breed. It is well known that segregation analysis is not robust to violations of model assumptions, such as the normality of residuals. In this analysis, we minimized any impact of the non-normality of residuals by using a Box-Cox transformation. One possible reason for not having found a major gene for SLW\textsubscript{SL} breed could be that there was less number of records and departures from other assumptions, such as the homogeneity of the distribution of residuals among families within this breed. These results from segregation analyses should therefore be considered as first indication of the presence of the major gene, which have then to be confirmed by analyses using molecular tools i.e. running DNA-level genetic experiments for QTLs mapping and gene expression studies. However, because segregation analyses are less costly, it is suggested to search for a major gene using non-Bayesian Gibbs sampling approach like maximum likelihood. We also suggested looking for imprinting major gene effect for these breeds. In future, a model that tackle two major genes (either in Mendelian or imprinting
inheritance mode) must be developed. Overall, it is suggested for more future work, major gene to be tracked down within families, within breed and compare them.

5.6. Acknowledgments

The author thanks Suisag Company for providing data. Thanks a lot to Dr. Houcine Ilahi for his great help.
Chapter 6

General Discussion

In this thesis, we used data derived from a custom-designed cDNA microarray platform (Atlas® cDNA expression arrays, BD Biosciences, Allschwil, Switzerland) to explore the pattern of gene expression in skeletal muscle in rat species across different experiments. In addition, we used teat, exterior and reproductive traits across three pig breeds to explore genetic variations for each trait. An extensive stand-alone discussion for each chapter was provided. However, here we provide a general discussion for the results which were obtained in this thesis in critically summarized way.

6.1. Rat Microarray and System Genetic Analysis

Chapter 2 and 3 used gene expression dataset from custom-designed microarray experiments which were performed at the Institute of Anatomy, University of Bern, Switzerland (Flück, M. et al., 2005). Ensembles of different computational methods were used to check in different biologically motivated questions. In terms of finding differentially expressed genes, based on the changes in mRNA levels, our results supported the idea that loading, reloading for rat’s hindlimb muscle (third experiment), influenced some key metabolic adaptations to conserve energy and metabolic substrates, that is, sharply affected expression of genes encoding some key enzymes in the glycolytic pathway, and also affected expression of gene transcripts in the TCA cycle. Such phenomenon was observed for other experiments, thought its degree of magnitude was not definitively comparable with third experiment. As we discussed already, one of the main goals of the biological system-level modeling is to reveal the regulatory network(s) underlying cell function. In particular, transcriptional regulatory networks have received immense attention, mainly due to the availability of large amounts of relevant genomics data. For inferring regulatory networks, probabilistic graphical models i.e., Bayesian networks, dynamics Bayesian network, are in general considered as a natural mathematical framework (Friedman, N. et al., 2000). In the context of gene expression data, there are two kinds of caveats for using probabilistic graphical models i.e., Bayesian networks. The first caveat of this framework is that a vast number of features and distribution parameters need to be
learned given only a limited number of observations. That means, there are lots of numbers of genes but small number of observations per gene. This remains one of key computational limitation in this context. Module network, a very complex but an attractive way to remedy this limitation was first introduced by Segal, E. et al (2003). He took advantage of the inherent modularity which exists in biological networks (Taki, K. et al., 2004; Bundy, J. et al., 2007). In particular he identified that groups of genes which acting in concert way are often regulated by the same regulators. Module network consists of a set of modules (group of genes that share the same parents and conditional distributions) and connections with arcs among the modules. The probabilistic inference process of a module network is consisted of following two steps: first the modules are inferred by grouping together genes with similar statistical features; second, regulator genes for every module are inferred. These steps update modules and connections with arcs so that a score of correspondent module network is maximized and in repeated manner convergence of the score is met.

Second caveat (without generalizing) for learning probabilistic graphical model (Bayesian network, module network) in the context of gene expression is that they require rich area of biological and gene annotations information. For example, Segal, E. et al., (2003) first identified set of signaling molecules and regulator genes with overall and local effects, though he only used regulator genes with local effects in his synthetic regulation program. This is the point where the role of a pure biologist is essential to further expand and motivate theoretical biological models. Without understanding the transcriptome data in depth and width in terms of the implicit meaning of gene annotations, the proficient use of probabilistic graphical model will likely not be possible. The wealth of information has to be ordered by basic biological knowledge to assist the bioinformatician in the identification of relevant findings. In this study, such information was not immediately available.

Therefore, we looked at other module learning procedure (Zhang, B. et al., 2005) to address some questions for rat hindlimb muscle data. The former module learning process for our microarray study is different from the one set up by Segal, E. et al., (2003). In former one, we do have densely interconnected set of genes (module) which illuminate higher order of over-represented structures in gene expression data. From a computational point of view, Zhang, B. et al., (2005)’s approach is not as vulnerable as Segal, E. et al., (2003) idea in terms of number of observations per gene. Learned modules in Zhang, B. et al., (2005) are not explicitly regulators modules. This explains since the adjacency matrix in the module network by (Zhang, B. et al., 2005) is symmetric whereas this is not the case for its predecessor (Segal, E. et al., 2003). Also, Zhang, B. et al., (2005) used scale-free topology
criterion, therefore, singling out genes out are biologically justifiable. The marriage between aforementioned module learning processes appears justified. Zhang, B. et al., (2005) approach (which is non-probabilistic module learning method) lets extra source of information (i.e. set of biological interested in genes, sex of arrayed individual and etc) be associated with different modules in course of learning modules across microarray gene expression dataset. We suggest selection of highly enriched module with extra source of information and implementing (Segal, E. et al., 2003) approach for that module is biologically impressive. That is, from the derived modules, we predict and estimate tentative functions for un-annotated genes with high certainty therefore, this marriage method is a useful tool for identifying unknown functions of genes and prediction of hub or lethal-acting genes that highly connected with extra source of information, meanwhile, gene regulation for set of learned modules and conditions in which they function, can be inferred. For addressing many complex aspects of gene regulations, expression QTL mapping (Jansen, R. C. et al., 2001) or genetical genomics is itself a powerful method. Computationally, this method is daunting as well. However, to come up with biologically motivated results, having co-segregating of expression genes and genetics markers data is central. Having such data is not readily achievable. By using (Kadarmideen, H. N. et al., 2006) approach, we implemented interval eQTL mapping for six candidate genes. Table 6. 1 contains summary of interval eQTL mapping for these candidate genes.
<table>
<thead>
<tr>
<th>Gene Names</th>
<th>Physical Location</th>
<th>Log-Likelihood Peak</th>
<th>eQTL Act</th>
<th>eQTL Act</th>
<th>eQTL Act</th>
<th>Spurious</th>
<th>Inferred Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutathione S-transferase theta 1</td>
<td>Chromosome 20</td>
<td>Chromosomes 3 and 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>Chromosome 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>Chromosome 3</td>
<td>Chromosome 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td>Cytochrom c Oxidase Subunit VIII-H</td>
<td>Chromosome 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not Inferred</td>
</tr>
<tr>
<td>Solute Carrier Family 4, Member 3</td>
<td>Chromosome 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not Inferred</td>
</tr>
<tr>
<td>Carbonic Anhydrase 3</td>
<td>Chromosome 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td>ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 1</td>
<td>Chromosome 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td>Fatty Acid Binding Protein 3</td>
<td>Chromosome 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>Chromosome 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Complex</td>
</tr>
</tbody>
</table>

*Table 6.1* Summary of genetical genomic analysis of candidate genes.
Almost all candidate genes revealed trans-acting eQTL phase, which indicates complex regulation. We can see that chromosome 3 is a harbor for physical positions of most candidate genes. **Figure 3. 3** indicates densely located genes (or genetic markers) between 69 to 73M base pair for physical map of estimated trans-acting eQTL region of chromosome 10 for gene Glutathione S-transferase theta 1. It is not immediately obvious which of them are compelling regulatory element(s) for Glutathione S-transferase theta 1. It is likely possible that Zhang, B. *et al.*, (2005) approach can reduce the densely located genes in trans-acting eQTL region. For example, it is possible to find hub-acting gene(s) in this region. The new identified hub-acting gene(s) with help of other tools like metabolimic QTL mapping (mQTL) or gene annotations, will narrow down the space of uncertainty with respect to Glutathione S-transferase theta 1 regulation to some hub-acting genes identified in trans-acting eQTL region. In this sense, both Jansen, R. C. *et al.*, (2001) and Zhang, B. *et al.*, (2005) approaches support each other. Therefore, implementing such strategy is recommended. Comparing **Table 6. 1** and **Table 3. 3** tells us that there is small association between most connected genes (hub-acting genes) and differentially expressed genes. We implemented genetical genomics for most varying genes. We called those genes candidate genes. Generally speaking, candidate genes weren’t most connected genes. Assuming all postulations which were used in this study were correct (i.e running co-expression gene network for whole genes and using non-segregating population for eQTL mapping), this is the first time that we address the relationship between gene connectivity and gene differentially in the context of gene expression studies.

### 6.1. Pig Traits Genetics Evaluations

Chapter 4 and 5 were dealing with genetic parameter estimations for some functional (exterior, teat, reproductive) traits in pig breeds. In chapter 4 we used variance covariance genetic estimation for the aforementioned traits. The main aim in this chapter was to estimate additive and maternal genetics components across three pig breeds. Genetic correlations were estimated, too. Ten traits within three different pig breeds Swiss-Large-White (SWL), Swiss-Landrace (SLR) and Swiss-Large-White-Sire-Line (SLWSL) were analyzed. For each of the traits within each breed 4 univariate animal models were fitted. Litter size affected exterior traits more than reproductive traits. Estimated maternal genetic heritability in exterior traits was relatively substantial especially for SLR and SLWSL breeds. Also, by fitting a genetic covariance among additive and maternal genetic effects, estimated
additive heritability increased, though concurrent trend for whole traits were not observed. Heritability of reproductive traits was generally lower than that of the exteriors traits. Comparison models based on log likelihood ratio test indicated that fitting model 4 likely explains better the structure of data for reproductive traits especially in the SLW breed. Genetic correlations among traits across breeds were almost similar. The genetic correlations among teat traits were positive and substantial (>90%). Also, importantly, the genetic correlation among traits across breeds was not changed much in terms of their mode and magnitude, which likely indicates that the estimate of genetic correlation among traits used in this study for one breed can be used for other breeds. Low level genetic correlation among reproductive traits with teat and leg traits were observed. This indicates that genetic selection in favor of reproductive traits, won’t affect the performance of teat and exterior traits. However, significant discrepancies of estimated genetic correlations across breeds were not statistically investigated. This result indicated that for genetic selection, including maternal effect in the analytical models makes sense to be somewhat for exterior and reproductive traits not for teat traits. However, researchers have different views in terms of estimating maternal effect for the aforementioned traits. To our knowledge, this was the first time that we estimated large-scale genetic components for exteriors, teat and reproductive traits across three pig breeds. We used two kinds of statistical modeling: univariate genetic analysis and pair-wise genetic analysis. We used phenotypic records in which missing values were screened out. Also traits were centralized (0, 1). In light of multivariate analysis, it is well understood that there are limitations in handling high-dimensional dataset with missing values. Some treatment of missing values has provided in (Little, R. et al., 2004). Because of this reality, we skipped records with missing values which actually sharply reduced the data size. Likely it is possible to lose information due to removing records with missing values, centralizing traits and editing data, but our experience indicated that even with missing value removed records, implementing multivariate genetic analysis was difficult task to be done. This is why we resorted to pair-wise genetic analysis. In the context of pair-wise genetic analysis, finding genetic associations between pig survival trait (FAR) and other traits was not trivial. Many times, estimated genetics parameters fell outside of parameter scope. This matter happened across all three breeds. Since there were different number observations across the three breeds, therefore, it was hardly believable that a small number of observations have been culminated in this problem. One reason for this failure could be due to using general mixed model. Many traits in this study followed non-Gaussian distributions and we used general
mixed linear model to analysis them. Canario, L. et al., (2006) stipulated that the underlying theory of mixed model theory is quite robust to be used for non-Gaussian traits distributions too. But the main problem was why only for one trait (FAR), model parameter estimated outside the scope of model parameters. This led us to have closer look at data structure. In doing so, we used five agglomerative clustering algorithms and two kinds of principal component analysis (PCA): based on covariance among and on correlation among traits. Different clustering algorithms and PCA methods, revealed hidden structures in our data. Pig survival (Far) trait didn’t cluster cohesively with other traits. PCA showed that this trait was responsible for a lot of diversity in dataset. Small information content for our dataset was estimated by correlation-based PCA. Covariance-based PCA sharply reduced the dimensional of original dataset. Therefore, we don’t recommend using it in revealing data structure for genetic evaluation. For our dataset, we argue that there are limitations to the ability of general mixed model for obtaining precise estimates of heritability for exterior, teat and reproductive traits. Because current datasets had relatively small observations across breeds, confidence intervals for estimated genetic parameters, likely could be large and power is low. Larger datasets and further development of alternative analytic strategies i.e. multiple-threshold approach, will be necessary to provide more precise genetic estimates. Trait’s names definition provided a lot of difficulties to find corresponding literature for them. Finally we suggest, if the aim of breeding program is to improve prolific maternal line then teat-related traits and exterior traits must be recorded. This study indicates that direct genetic improvement of reproductive traits in pigs will be slower than exteriors traits. Published genetic parameters for the traits addressed in this study, particularly maternal genetic components were scant. Chapter 5 was dealing with major gene genetics parameter estimation. Complex Bayesian major gene segregation analysis indicated that an autosomal Mendelian major gene is segregated for carcass length trait in SLR and SWL breeds but not in SWL SL breed. A carcass length trait is economically motivated trait. Therefore, our results highlight the difficulties inherent with attempting to implement DNA-based studies for SWL SL breed in conjecture with carcass length trait. Generally speaking, for SLR and SWL breeds, fine mapping and gene expression study for deciphering black box of carcass trait is highly recommended. Carcass traits have been highly investigated (Malek, M. et al., 2001; Sanchez, M.P. et al., 2003; Sernius, T. et al., 2004) and high levels of correlation among them have been reported. The estimated genetic correlations are fundamentally based on infinitesimal theory (polygenic). We could argue that finding major gene across two traits and establishing genetic correlation based on
major gene effects can bring much biological insights. In this way, topology of metabolic network can be indirectly scratched. We argue that such a scenario makes sense to be done for carcass length trait. However, the polygenic heritability for carcass length trait across all breeds studied in this thesis obtained substantial. Therefore, genetic selection for this trait can boost genetic progress in eye-catching way. We recommend running simulations to see the effect of Box-Cox transformation on different likelihood probability of data, in which they should be used for finding major gene effects. Overall, this thesis has shed some light onto unseen parts of subjects which were covered up. My general impression in doing this thesis is that, quantitative genetic analyses were much time-consuming and intuitively demanding than genome based researches. In future, every feature of life will be translated into numbers. Numbers are not kind (in moral sense), but they make life better. However, this idea that accumulated biological data culminated into new era of logical interpretations requires daily updating of mathematical tools and imagination development.
References


Biography

Curriculum Vitae of
Mostafa Ghaderi-Zefrehei

Personal Information

Surname / First name     Mostafa
Family / Last name       Ghaderi-Zefrehei
Nationality              Iran
Date of birth            20, MARCH, 1976
Gender                   Male
Marital Status           Single
E-Mail: mostafa.ghaderi@inw.agrl.ethz.ch
E-Mail: mosms741@yahooo.com

Education and Training

1992-1995; Diploma in Science, Hafez Exemplary High School, Kuhpayeh       Iran
1995-1997; ASc in Animal Production Technology, Bou-Ali-Sina University       Iran
1998-2000; B.Sc in Animal Production Technology, University of Tehran        Iran
2000-2002; M.Sc in Breeding and Genetic, Isfahan University of Technology     Iran
2004-PhD Bioinformatics and Quantitative Genetics at ETH Zürich              Switzerland

Personal Skills and Competences

Language Skills

1. Farsi          Understanding
2. English        Understanding
3. German         Understanding
4. Italian        Aquatinted
5. Arabic         Aquatinted

Computer Skills

1. Programming Language
   • FORTRAN 90,
   • Shell Programming
2. Scripting Language
   • Perl
3. Markup Language
   • XML
4. DNA, RNA, Protein Sequence and Phylogenetic Tree Analysis
   • DARWIN
5. Latex Typesetting
General statistical Tools
1. SAS and programming (good especially for error structure)
2. Matlab (good for numerical analysis)
3. R (the best things in the world is free!)

Additional,
1. Experience working alone or as part of a team
2. Experience meeting tight deadlines
3. Experienced at multi-tasking
4. Excellent communications skills
5. Good knowledge of mysql, SQL
6. Feed formulation with NRC and CNCPS system

4 Operating System Skills
1. Multi-User, Multiprocessing, Multitasking and Multithreading GUI,
   (Windows Linux, UNIX, Mackintosh)

Genetic Evaluation Skills:
1. ASREML (Genetics Variance, Covariance Estimation),
2. DF-REML (Genetic Variance, Covariance Estimation),
3. MAGGIC (Bayesian Segregation Analysis),

Bioinformatics and Computational System Biology Skills:
1. Identifying Differentially Expressed Genes (Microarray Analysis)
2. Learning Co-Expression Gene Network (Based on Graph Theory)
3. Neural Network Algorithms in Expression Analysis
4. Modeling with Ensemble system of ordinary Differential Equations
5. Stochastic Metabolic network Analysis
6. Expression QTLs Mapping
7. Compendium (a great idea in Bioinformatics)

Master Thesis
   Covariance Estimations of dairy cattle data using DF-REML animal models.
2. PhD Papers
   Some papers ready for submitting