Genome-wide Association and Haplotype-based Association Mapping of Mastitis in Lacaune Sheep

Bingjie Li

Norwegian University of Life Sciences
Department of Animal and Aquacultural Sciences (IHA)
Master Thesis 30 credits 2013

European Master in Animal Breeding and Genetics (EM-ABG)

Genome-wide Association and Haplotype-based Association Mapping of Mastitis in Lacaune Sheep

Student: Bingjie Li

Supervisors: Rachel Rupp

Dag Inge V åge

JUNE 2013

Content

Abstract

The aim of this study was to perform a Genome-Wide Association Study (GWAS) and a Haplotype-based Association Mapping to identify potential Quantitative Trail Loci (QTLs) underlying four mastitis-related trails in Lacaune sheep genotyped by the Illumina Ovine 50K SNP chip. The four traits studied included two Somatic Cell Count (SCC)-based traits, and two traits based on mammary abscesses indicating chronic mastitis infection. A total of 393 genotyped animals from two lines of Lacaune sheep previously selected for high and low milk SCC were included in this study, within which 334 ewes were phenotyped based on clinical examinations of udders. The statistical analysis included a Genome-wide Association Study (GWAS) assuming independent markers and a Haplotype-based association study using phasing haplotype clusters. Mixed models accounting for population structure by including polygenic effect in the model were used in both studies. The results indicated five interesting regions on OAR 3, 5, 8, 16 and 20, which were associated with different mastitis-related traits. Among the five regions, OAR16 was assumed to have a potential QTL around 30Mb, since this region was associated with three mastitis-related traits. Region 49.1-49.5Mb on OAR20 was also very interesting since it was associated with one SCC-based trait and two mammary abscess traits. Regions on OAR5 and OAR8 were found to be specifically related to the presence of mammary abscesses, which may infer a specific immune pathway involved in forming mammary abscesses in response to chronic clinical mastitis. In conclusion, this study introduced the mammary abscess as a new phenotype for mastitis and detected two candidate QTLs strongly associated with this phenotype. Introducing new refined mastitis-related traits is important to determine the genetics underlying of mastitis infection.

Keywords: mastitis; mammary abscesses; somatic cell count; genome-wide association; haplotye-based association.

Genome-wide Association and Haplotype-based Association Mapping of Mastitis in Lacaune Sheep

Introduction

Mastitis refers to the inflammation in the mammary gland due to intra-mammary infection (IMI) caused by pathogenic organisms (Heringstad et al. 2000). It is considered as one of the most costly diseases in dairy animals (Davies et al. 2009; S J Wells et al. 1998; Bergonier et al. 2003), leading to a direct impact on milk production through increased culling rate, altered milk composition (Leitner et al. 2004), decreased milk quality (Coulon et al. 2002; Ma et al. 2000; Hogarth et al. 2004) and increased cost of veterinary treatment (Wells et al. 1998).

Staphylococci are the most prevalent pathogens inducing mastitis in dairy ruminants (Heringstad et al. 2000; Bergonier et al. 2003). In dairy cows, Staphylococcus aureus, Streptococcus agalactiae, Coliforms, Streptococci and Enterococci are major pathogens causing clinical mastitis, while Coagulase-negative Staphylococci and Corynebacterium bovis are minor pathogens causing subclinical mastitis (Heringstad et al. 2000; Harmon 1994). In small dairy ruminants, subclinical infection is much more frequent than clinical cases, which means Coagulase-Negative Staphylococci cannot be considered as minor pathogens any more (Bergonier et al. 2003). Coagulase-Negative Staphylococci are the most frequent pathogens in small dairy ruminants causing subclinical infection, while Staphylococcus aureus are found to be to be most associated with clinical mastitis in small ruminants (Bergonier et al. 2003).

The annual incidence of clinical mastitis in small dairy ruminants is generally lower

than 5% (Bergonier et al. 2003), compared with 20% to 40% in dairy cattle (Heringstad et al. 2000). During clinical mastitis, infected animals suffer from obvious inflammation and pain in the udder, declined milk yield and dramatic change in milk composition, in some cases, increased rectal temperature, lethargy, or anorexia (Harmon 1994; Heringstad et al. 2000). In subclinical mastitis, milk samples collected during early and mid-lactation contain coagulase-negative staphylococci, which may persist in many of the challenged glands and continue to influence the physiological conditions of the gland, leading to a chronic infection for long periods (Burriel 1997). In small ruminants, mastitis is generally a chronic and contagious infection (Bergonier et al. 2003). Acute clinical mastitis is very rare in small ruminants, but it generally causes severe clinical symptoms and leads to culling of the infected animals. By comparison, chronic mastitis is more common in small ruminants, causing pathological changes in mammary gland due to long-term infection. One of the key symptoms in chronic mastitis is mammary abscesses, referring to a collection of pus (full of neutrophils) accumulated in a cavity formed by the tissue as a result of an infection process (Marogna et al. 2010). Mammary abscesses can be detected through clinical examination of mammary glands, indicating a direct immune response to chronic mastitis.

Accumulating studies imply that animals' response to intra-mammary infection (IMI) is partly under genetic control (Rupp & Boichard 2003; Rupp 2007; Ogorevc et al. 2009). The occurrence of clinical mastitis (CM) can be diagnosed directly using milk bacteriological examination, or by indirect measures correlated with intra-mammary infection (Heringstad et al. 2000). In Scandinavian countries, clinical mastitis (CM) has been routinely recorded since 1980s, and mastitis has been included in breeding programs since 1990s based on a direct selection for mastitis resistance using CM records (Heringstad et al. 2000). However, due to the lack of regular records of CM in most other countries, and a generally low heritability of CM (h^2= 0.02-0.03) (Heringstad et al. 2000), Somatic Cell Count (SCC) has been developed as an indirect measurement of mastitis (Bergonier et al. 2003; Rupp et al. 2009). Milk SCC reflects

the number of neutrophils that migrate from blood to mammary gland, which increases dramatically in response to intra-mammary infection (Rupp et al. 2009). In practice, the measures of SCC are commonly log-transformed into SCS in order to achieve data normality (Ali & Shook 1980). Genetic parameters have been established for SCS with a heritability ranging from 0.08 to 0.19 for lactation mean SCS (LSCS) in dairy cows (Heringstad et al. 2000), while in dairy sheep the heritability is estimated from 0.12- 0.15 (El-Saied et al. 1999; Barillet et al. 2001; Rupp et al. 2003). Specifically for Churra and Lacaune sheep, most studies based on larger data sets reported consistent heritability between 0.11 and 0.15 for the lactation mean SCS (LSCS) (Barillet et al. 2001; El-Saied et al. 1999; Rupp et al. 2003; Rupp et al. 2002). SCC is found to be in highly correlated with clinical mastitis (CM), with a genetic correlation from 0.53 to 0.70 (Heringstad et al. 2000; Heringstad et al. 2006; Ødegård et al. 2004; Koivula et al. 2005; Carlén et al. 2004). This indicates a shared genetic basis between SCC and clinical mastitis, but also highlights that SCC is not exactly the same trait as clinical mastitis, since their genetic correlation is still far from 1. Therefore, SCC should be partly related to clinical mastitis, and partly related to other mechanisms unrelated to mastitis. Recent studies have reported several potential QTL regions underlying SCC. In dairy cattle, QTLs associated with SCC/SCS are positioned on all chromosomes except on BTAX, according to the available reported QTLs in Animal QTL database (QTLdb; http://www.animalgenome.org/QTLdb). In dairy sheep, the number of published QTL is very limited and restricted to the SCC phenotype: 7 QTLs associated with SCC/ SCS are published in Animal QTL database, separately distributed on OAR2, 6, 10, 14, 17, 20, and 22 (QTLdb; http://www.animalgenome.org/QTLdb). There are also a few unpublished QTLs available. Rupp et al. (2011; unpublished paper, 2013) reported several candidate QTLs for SCC based on a grand daughter design from commercial flocks of Lacaune sheep. These candidate QTLs are positioned on OAR3, 8, 10, 11, 14, 16, 20, and constitute good candidate QTL regions for further genetic studies in Lacaune sheep.

This potential genetic basis of mastitis offers a possibility to select dairy animals with

better resistance to mastitis. A recent study has proven the effectiveness of SCC-based selection against mastitis in dairy small ruminants. Rupp et al. (2009) showed that SCS-based selection has improved resistance to natural IMI by using Lacaune dairy ewes divergently selected for milk SCS. They found significant difference in resistance to mastitis between ewes from the high SCC line (SCC+) and the low SCC line (SCC-), where all clinical cases of mastitis came from ewes in high SCC line. Additionally, the frequency of chronic clinical mastitis, diagnosed by the presence of mammary abscesses, was much higher in the high SCC line compared to the low SCS line (Rupp et al. 2009). These results indicate a potential of using "mammary abscesses" as a new phenotype to diagnose and record mastitis.

So far, no major genes underlying SCC are known. SCC has been used as a black box in selection whose nature might not fully explain the complexity of mastitis (Rupp et al. 2009). The effectiveness of using SCC as an indirect measure of mastitis may be lower than expected since the genetic correlation between SCC and CM is far from 1. In addition, animals with and without clinical mastitis were reported with different heritability for SCS, indicating a heterogeneous character of SCS in explaining mastitis (Heringstad et al. 2006). Therefore, it is highly recommended to include additional mastitis-related traits to give a more reliable phenotype of mastitis. Phenotypes that correlate stronger with mastitis will perform better both in selection programs against mastitis and in studies aimed at identifying genetic variation explaining different resistance to mastitis. However, except for SCC, no other phenotypes are currently available to measure mastitis on a large scale in small ruminants. One potential phenotype is mammary abscesses, a typical symptom of chronic mastitis. Although the mammary abscess has not been much used as a chronic mastitis measure yet, it offers valuable new information about the mastitis phenotype. In addition to including new mastitis-related phenotypes, fine-phenotyping of available mastitis-related traits can be also a way to increase the accuracy of QTL detection by providing more detailed and comprehensive information of the phenotype. A good example can be developing SCC-based traits with higher

resolution. These SCC-based traits can be records from different lactation periods, or from different methods in analyzing SCC phenotypes.

The aim of this study was to perform both a genome-wide association study (GWAS) and a haplotype-based association mapping to identify potential QTLs underlying four mastitis-related traits by using Illumina Ovine 50K SNP chip in Lacaune sheep from an experimental farm. The four studied traits include two SCC-based traits, and two traits based on mammary abscess. In the end, candidate QTLs from all traits were compared and contrasted to detect if they share some interesting regions.

Material and Methods

Animals

All animals used in this study were from the farm of La Fage, an experimental farm at INRA (UE 321, Roquefort, France). Two lines of Lacaune sheep, selected for high and low milk SCC, were included in this study, designated as "*Low SCC Line (SCC-)"* and "*High SCC Line (SCC+)*". Animals in the high SCC line were found to have significantly higher susceptibility to mastitis, while animals in low SCC line were relatively resistance to mastitis (Rupp et al. 2009). In total, 393 sheep in this population were genotyped, including 59 sires and 334 daughters (152 SCC+ and 182 SCC-). The mammary of the 334 ewes were clinically examined by experienced technicians on the farm after previous training by a veterinarian.

Phenotypes

Four mastitis-related traits were included in this study. Table 1 define and describes the data type of each trait: The first two traits are based on mammary abscesses and the other two traits are based on SCC (Somatic Cell Count). They are treated as four different traits but are in high phenotypic correlations with each other: the phenotypic correlation between two mammary abscess traits is 0.96, while the correlation

between the SCC-based traits is 0.92, and the correlation between mammary abscess traits and SCC-based traits is 0.50-0.60.

Table 1. Definitions of the four mastitis-related traits.

Trait name	Definition of the trait	Data type
Abscess (case-control)	The occurrence of a mammary abscess of an individual during the first lactation.	Case/control (0: no abscess, 1: abscess)
Abscess (category)	The number of mammary abscesses found in the first lactation. (Animals with five or more mammary abscesses were all recorded as 5)	Categorical (0,1,2,3,4,5)
Infection Status	Infection status of individuals is determined by SCC in first lactation: Animals with SCC measures higher than 300,000 cells/mL were regarded as infected. (method described in Bergonier et al. (2003))	Case/control (0: healthy, 1: infected)
LSCS	Lactation Mean SCS during first lactation.	Continuous

The phenotype data recorded from clinical examinations were filtered before being used in the data analysis: (1) For trait Abscess (case-control), animals with at least three clinical examinations were included in the dataset. Animals were recorded as "1" (case) when examinations detected at least twice the presence of an abscess, while animals were recorded as "0" (control) when they were examined as healthy (without any abscesses) at least three times. (2) For trait Abscess (category), animals with at least three clinical examinations were included. The number of mammary abscesses was used as phenotype for this trait. (3) For the trait infection status, the infection status of an individual was determined by repeated SCC measures over the lactation (6 measures). Animals in any SCC measure higher than 300,000 cells/mL were regarded as "infected" animals. The infection status from the second lactation (L2) was introduced to check the result from the first lactation (L1): animals were recorded as "1" (infected) when they were detected "infected" in L1 and "not healthy" in L2; Animals were recorded as "0" (healthy) when they were recorded as "healthy" in L1 and "not infected" in L2; Animals with other phenotypes were recorded as "doubtful" and not included in the data analysis. (3) For trait LSCS, all animals with lactation

mean SCC in first lactation (LSCS during L1) were included in the dataset. The number of animals included in the analysis after data filtering is described in Table 2, as well as the number of cases and controls, and the animals' distribution in SCC+/SCC- lines (Table 2).

Table 2. The number of individuals and the number of cases/controls for each trait. The distributions of individuals within the two SCC lines are given in brackets.

Trait	No. of animals (SCC-/SCC+)	No. of cases (SCC-/SCC+)	No. of controls (SCC-/SCC+)
Abscess (case-control)	137 (74/63)	20 (2/18)	117 (72/45)
Abscess (category) [1]	158 (80/78)	41 (8/33)	117 (72/45)
Infection Status	172 (109/63)	40 (8/32)	132 (101/31)
LSCS [2]	281 (152/129)	---	---

1: Abscess (category) is not a case-control trait. Here the "No. of cases" indicates the numbers of animals with none-zero numbers of abscesses, and "No. of controls" indicates the number of animals with no abscess.

2: LSCS contains continuous data indicating the mean value of SCC in L1, which is not a case-control trait.

Genotypes and Quality Control

Genotyping was performed using the Illumina Ovine SNP50 BeadChip on the Labogena platform according to the manufacturer's standard procedures (Illumina, San Diego, CA, USA). Samples were included in the analysis if call rates > 95% (percentage of SNPs genotyped for an individual > 95%). Genotypes were included if the SNP marker passed a threshold of call frequency > 97%, minor allele frequency (MAF) > 0.01, Mendelian error < 0.025, and markers' distribution followed Hardy-Weinberg equilibrium. After quality control, 43,445 SNPs (80% of the total SNPs) passed the quality filtering from a total 54,241 markers. Markers were distributed on all 26 chromosomes, and the average distance between two SNPs was 0.0629 Mb. In total, 334 ewes genotyped in this divergent population were included in the analysis of mastitis-related traits.

Statistical Analysis

Statistical analysis was performed separately for the 4 traits for both GWAS and haplotype-based association.

Genome-wide association study (GWAS)

A GWAS was performed based on a linear mixed model including the additive SNP effect as a fixed effect, and the polygenic effect as a random effect.

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e}$$

Where **y** is the vector of observations; $\boldsymbol{\mu}$ is the overall mean; $\boldsymbol{b}$ is the vector of fixed SNP effect (allele effect of SNP); $\boldsymbol{X}$ is the incidence matrix of $\boldsymbol{b}$ corresponding SNP effects to individuals. The random polygenic effect $\boldsymbol{u}$ was included to account for the family structure in the data. $\boldsymbol{u}$ was assumed following $\boldsymbol{u} \sim N(0, A\sigma_a^2)$ where A is the relationship matrix based on pedigree information. $\boldsymbol{Z}$ is the incidence matrix of $\boldsymbol{u}$, accounting for pedigree relationship structure among individuals. $\boldsymbol{e}$ is the random residual effect following $\boldsymbol{e} \sim N(0, I\sigma_e^2)$ where I is the identity matrix.

Restricted maximum likelihood (REML) was applied to solve the model by using ASReml (Gilmour et al. 2009). SNP effects of all 43,445 SNPs were calculated, and T-test was performed to calculate the $-\log_{10}$ (p-value) of each SNP effect. In significance test, Bonferroni correction (significance threshold = $-\log_{10}(\alpha/Nbsnp)$) was applied to α=5% for both genome-wise and chromosome-wise thresholds in order to correct the bias from multiple testing.

In addition, a Quantile-Quantile plot (QQ plot) was performed to check the distributions of test-statistics, in order to detect the "goodness of fitting" of this model. Our dataset includes several sire families in the population, which may produce population stratification leading to unexpected false positives in the association test. A

polygenic effect was therefore included in the mixed model to correct for the population structure. QQ plot was used here to detect the fitting of the model, especially for the goodness of fitting the population structure into the model.

Haplotype-based Association Study

Haplotype reconstruction and clustering was based on the method from Druet & Georges (2010) using PHASEBOOK Software. PHASEBOOK is a package of four haplotyping programs: LinkPHASE, HiddenPHASE, DualPHASE and DAGPHASE (Druet & Georges 2010). Haplotype reconstruction and clustering were performed following the procedures below:

1. Genotype data from 334 individuals was partially phased by LinkPHASE based on familial information (Mendelian segregation rules and linkage information).
2. The unphased missing alleles from Step 1 were randomly sampled by DAGPHASE. The sampling output was used by BEAGLE program to create a directed acyclic graph (DAG) which indicates a summary of localized haplotype clusters (Browning & Browning 2007).
3. A Hidden Markov Model (HMM) (Scheet & Stephens 2006) was used in DAGPHASE to sample missing alleles in base haplotypes from the previous DAG. The sampling result was output into BEAGLE to generate an improved DAG. This process was performed iteratively between DAGPHASE and BEAGLE to generate improved DAG, until the last DAG was constructed and output as the most-likely haplotype clusters.

Specifically for the third step, sampling parameters are required to set by users for running BEAGLE. There was controversy in setting the value of parameter "scale" according to the previous studies (Browning 2006; Browning & Browning 2007; Druet & Georges 2010). Parameter "scale" was reported to be related to "the number of samplings per individual" in BEAGLE sampling (Browning & Browning 2007). In

this study, two values of “scale” (scale=4.0 and scale=2.0) were tried according to previous studies, in order to detect a better setting of “scale” and the influence of parameter setting on haplotype reconstruction in a small data set like this study.

This phasing algorithm involved iteratively sampling haplotypes and assigning reconstructed haplotypes to a limited number of “localized haplotype clusters” called “hidden clusters”. In this case, the “hidden cluster” value of a marker represents the haplotype cluster this marker belongs to, and markers originating from the same haplotype were denoted by the same hidden cluster value. For association analysis, these hidden clusters were seen as the haplotype type at each marker that directly fit the mixed model for REML calculation.

The mixed model used for this haplotype-based association test was identical to that in GWAS, except that a haplotype effect (or hidden cluster effect) was fitted instead of a SNP effect in the model.

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{W\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

Where **y** is the vector of observations; $\boldsymbol{\mu}$ is the overall mean; $\boldsymbol{\beta}$ is the vector of haplotype effect (hidden cluster effect); $\boldsymbol{W}$ is the incidence matrix of $\boldsymbol{\beta}$ corresponding haplotype effects to individuals. The random polygenic effect $\boldsymbol{u}$ was included to account for the family structure as it was in GWAS. $\boldsymbol{Z}$ is the incidence matrix of $\boldsymbol{u}$, accounting for pedigree relationship structure among individuals. $\boldsymbol{e}$ is the random residual effect following $\boldsymbol{e} \sim N(0, I\delta_e^2)$ where I is the identity matrix.

Restricted maximum likelihood (REML) was applied to solve the model using ASReml (Gilmour et al. 2009). The haplotype effects of all haplotyopes at all markers were calculated. An F-test was performed for significance test, with Bonferroni correction applied to α=5% to correct for the bias from multiple testing. As previously described in GWAS, QQ plot was performed to check the fitness of the model.

Results

Validation of test-statistics: QQ-plot

Figure 1 gives the QQ plots for four mastitis-related traits obtained in both GWAS and Haplotype-based association. The four plots from haplotype-based association were all based on haplotype reconstruction with phasing parameter "scale" =4.0. The QQ plot of phasing with scale=2.0 showed severe false positives in markers' distribution (Figure 6 in discussion), whose mapping result was discarded in this study in order to maintain the accuracy of association. Therefore, "scale"=4.0 was regarded as a better setting of parameter in this study, and the followed results from haplotype-based association were all based on scale=4.0.

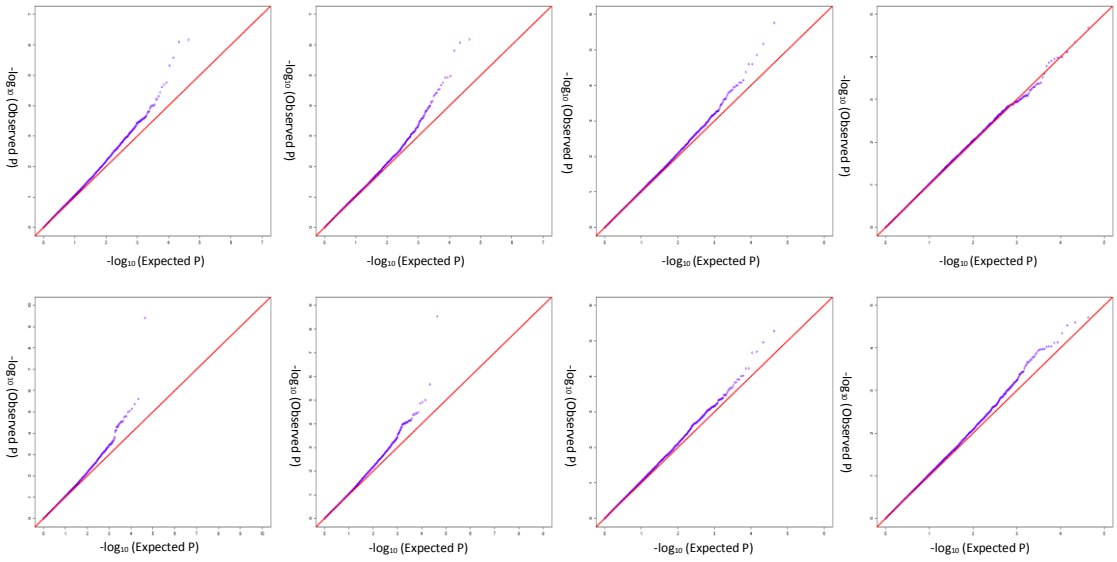

Figure 1. QQ plots of $-\log_{10}$ P resulting from GWAS (first line) and Haplotype-based Association (second line) for traits: Abscess (case-control), Abscess (category), Infect status, LSCS, from left to right within each line. The red diagonal line is the standard y=x, representing Observed p-values = Expected p-values, as expected according to the null hypothesis. Dots on this line or very close to it represent SNPs that follow the null hypothesis with no significant association with the trait. Dots obviously deviated from this line correspond to SNPs that disobey null hypothesis and were significant SNPs.

According to Figure 1, all plots gave reasonable distributions of p-values of markers,

with major SNPs following the null hypothesis and only a few deviating SNPs indicate significant SNPs. This result indicated that our model succeeds in correcting for population stratification in the analysis.

However, false positives may still exist in our analysis: In both GWAS and Haplotype-based association, plots of abscess traits (Abscess (case-control), Abscess (category)) show slight over-estimation of p-values of a few markers, which may lead to presence of false positives in the test-statistics. This infers a certain degree of fitting problem of our model, which may come from the definition of the model itself, or due to the limited number of animals and cases in abscess traits.

QTL mapping and potential QTL regions

The result of QTL mapping was summarized in Table 3. In GWAS, 14 chromosome-wise significant SNPs over 11 chromosomes were detected, within which 3 are genome-wise significant SNPs, distributed on OAR5, OAR21, OAR24. In the haplotype-based association test, 24 chromosome-wise significant SNPs over 13 chromosomes were detected, with 1 genome-wise significant SNP on OAR25. Among all these regions, 4 regions were found significant in both association methods (Table 3). These 4 regions were thought to be candidate QTLs for their respective traits. The details of these 4 interesting regions were described below trait by trait. Manhattan plots of all traits for both association methods are shown in Figure 2.

Table 3. Summary of significant SNPs detected in GWAS and Haplotype-based association.

Mastitis-related traits	**Number of significant SNPs**				**Number of shared regions by both methods**
	GWAS		**Haplotype-based association**		
	Chromosome-wise	Genome-wise	Chromosome-wise	Genome-wise	
Abscess(case-control)	7	2	13	1	1
Abscess(category)	6	2	4	1	1
Infect status	3	0	4	0	2
LSCS	2	0	3	0	0

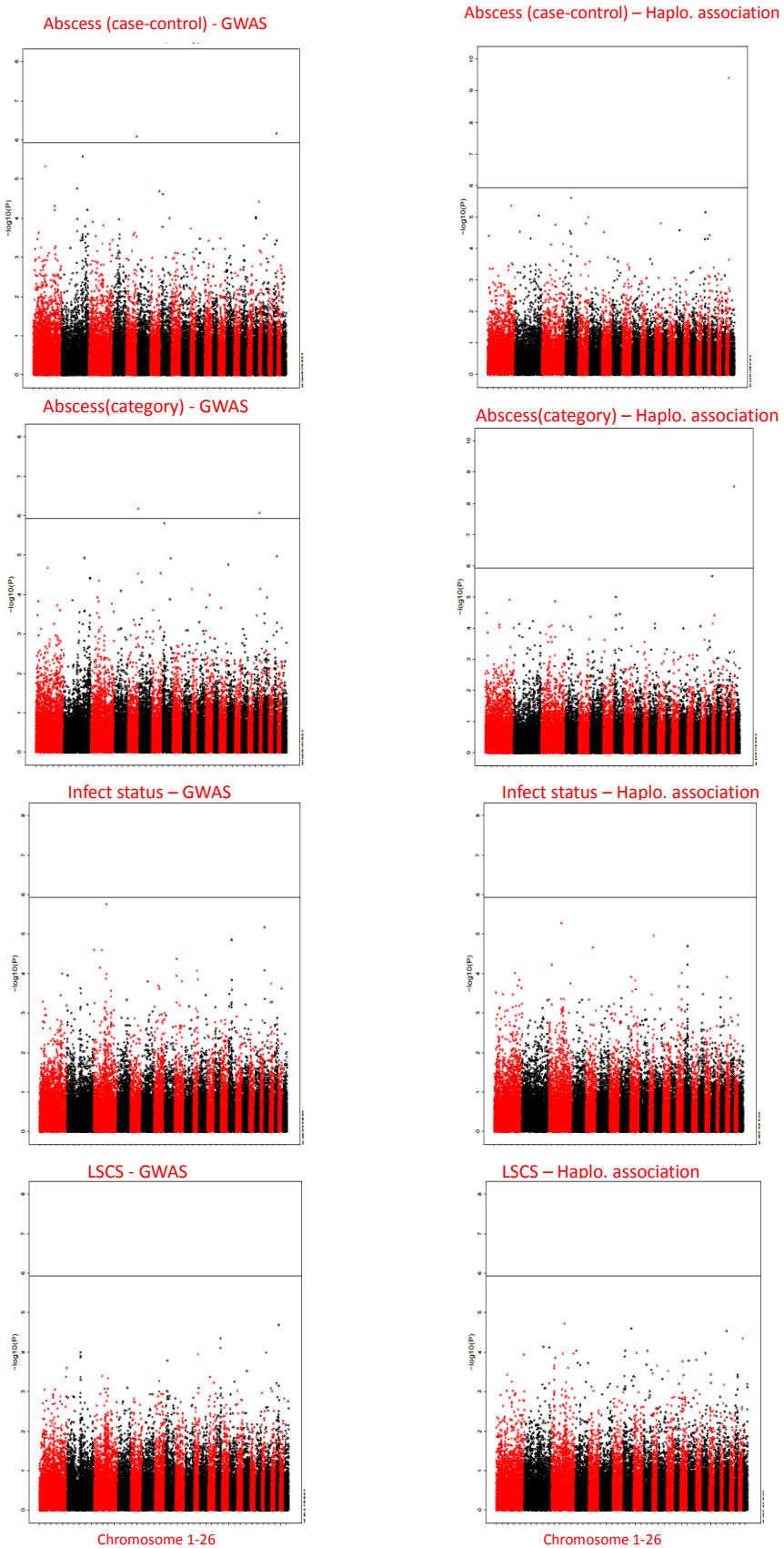

Figure 2. Manhattan plots for 26 chromosomes of four mastitis-related traits obtained from GWAS and Haplotype-based association. The horizontal line in each plot describes the genome-wise significance threshold $-\log_{10}$(p-value)=5.94 after Bonferroni correction.

Mammary abscess traits: Abscess (case-control) and Abscess (category)

For Abscess (case-control), region 95.9-96.7 Mb on OAR5 was detected by both methods to be significantly associated with mammary abscesses in Lacaune sheep (Table 4). In addition, there are also 5 regions on OAR2, 4, 8, 20 which showed significant association by one method and moderate p-values by the other method. Since Bonferroni correction is a conservative (severe) method in significance testing, these 5 regions may also be interesting in affecting mammary abscesses, but need further validation.

Table 4. Potential QTL regions affecting Abscess (case-control) from GWAS and Haplotype-based association testing. The line in bold shows the region significant according to both association methods.

Chr.	Potential QTL regions (Mb)[1]	$-\log_{10}$ P of the potential region		Chromosome Significance Threshold of $-\log_{10}$ P
		GWAS	Haplo. Association	
5	**95.9-96.7**	**6.10 *** **	**4.99 ***	**4.60**
2	192.1-192.7	5.58 *	3.48	4.97
2	220.6-221.9	3.53	5.04 *	4.97
4	54.0-54.7	3.97	5.61 *	4.65
8	14.5-16.3	4.62 *	3.66	4.54
20	21.5-23.2	4.03	4.29 *	4.26

***: genome-wise significance *: chromosome-wise significance

1: The intervals of the QTLs were defined based on the QTL regions found in two association methods.

For trait Abscess (category), region 14.5-16.6 Mb on OAR8 was detected by both methods significantly associated with mammary abscesses (Table 5). Two additional regions on OAR5 and OAR8 are also interesting since they were detected to be significant in GWAS and showed moderate p-values in Haplotype association tests.

Table 5. Potential QTL regions affecting Abscess (category) from GWAS and Haplotype association tests. The line in bold shows the region significant according to the two association methods.

Chr.	Potential QTL regions (Mb)[1]	$-\log_{10}$ P		Chromosome Significance Threshold of $-\log_{10}$ P
		GWAS	Haplo. Association	
8	**14.5-16.6**	**5.81 ***	**5.00 ***	**4.54**
5	96.1-96.7	6.18 ***	3.63	4.60
8	73.5-74.5	4.92 *	4.00	4.54

***: genome-wise significance *: chromosome-wise significance

1: Intervals of the QTLs were defined based on the QTL regions found by two association methods.

By comparing the results from Abscess (case-control) and Abscess (category) horizontally (Table 4 & Table 5), it is encouraging that the two regions, 95.9-96.7 Mb on OAR5 and 14.5-16.6 Mb on OAR8, were shared by both these two abscess traits (as significant QTL regions). This may infer a functional relationship between these two regions and mammary abscesses in Lacaune sheep. Figure 3 and Figure 4 show the Manhattan plots of these two candidate regions on their chromosomes, respectively.

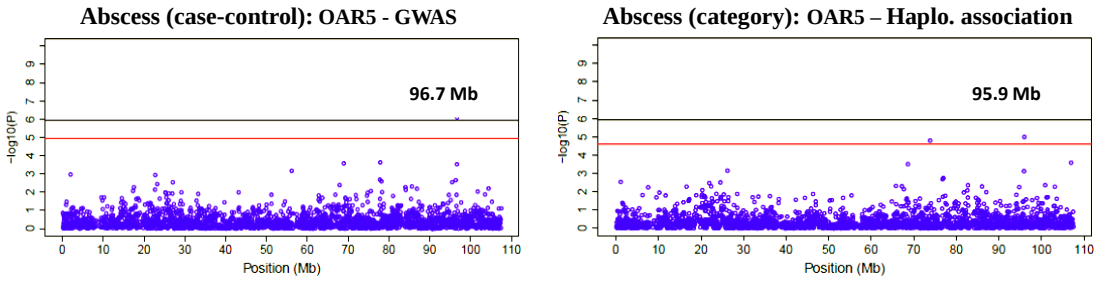

Figure 3. Manhattan plots on OAR5 for Trait Abscess (case-control): Region 95.9-96.7Mb showed significant association with Abscess (case-control) in both GWAS and Haplotype association. The black line indicates the genome-wise significance threshold, and the red line indicates the chromosome-wise significance threshold.

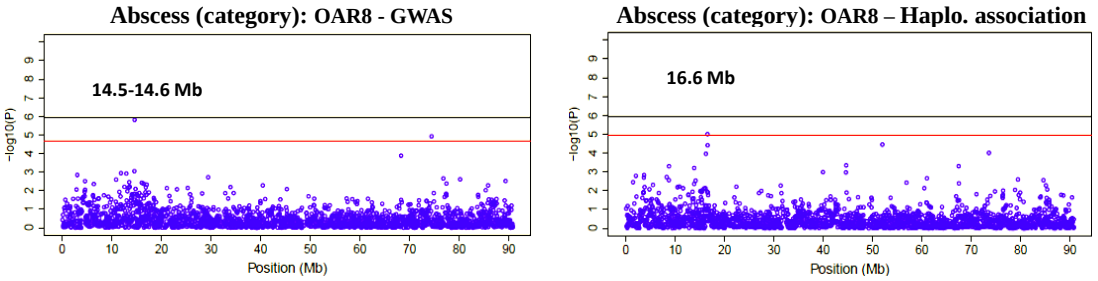

Figure 4. Manhattan plots on OAR8 for Trait Abscess (category): Region 14.5-16.6 Mb showed potential association with Abscess (category) in both GWAS and Haplotype-based association. The black line indicates the genome-wise significance threshold, and the red line indicates the chromosome-wise significance threshold.

Apart from the regions shared by the two association methods, the haplotype-based association test detected two other regions that had been validated by previous QTL projects on LSCS trait in Lacaune sheep (Rupp et al., 2011; unpublished paper, 2013). These two regions were associated with two mammary abscess traits (from

haplotype-based association) and LSCS trait, which infer themselves as general QTLs underlying mastitis infection. Table 6 gives the information of these two QTLs detected in this study and in previous study.

Table 6. Validated QTLs detected from haplotype-based association in this study.

Chr.	Haplotype- based association in this study		Previous study [a]	
	Trait	Potential QTL regions (Mb)	Trait	Potential QTL regions (Mb)
16	Abscess(case-control) Abscess(category)	**31.1-31.8 ***	**LSCS**	**29-36**
20	Abscess(case-control) Abscess(category)	**49.1-49.5 ***	LSCS	**49.1-50.6**

*: chromosome-wise significance.

a: GWAS analysis from 1000 Lacaune AI sires for LSCS trait, SheepSNPQTL Project (ANR; final report) (Rupp et al. 2012).

Somatic Cell Count traits: Infection status and LSCS

For the trait Infection status, two regions on OAR3 and OAR16 were detected to be significantly associated with mastitis infection by both association methods (Table 7). The two candidate regions were shown in Figure 5. Specifically, the candidate region on OAR16 was very close to the region previously detected in mammary abscess on OAR16 which is also inside the validated QTL region reported by (Rupp et al. 2012). This region, around 30Mb on OAR6, has been detected three times significantly associated with different mastitis-related traits (mammary abscess, infection status, LSCS), inferring a big possibility of existence of QTL associated with mastitis around 30Mb on OAR16. Figure 5 shows the Manhattan plots of these two candidate regions.

Table 7. Potential QTL regions for Infection status from GWAS and Haplotype association mapping.

Chr.	Potential QTL regions (Mb)	$-\log_{10}$ P		Chromosome Significance Threshold of $-\log_{10}$ P
		GWAS	Haplo. Association	
3	**117.2-117.5**	**5.76 ***	**5.28 ***	**4.92**
16	**27.3-29.7** [a]	**4.86 ***	**4.70 ***	**4.41**

*: chromosome-wise significance

a: The region lies very close to the validated QTL found in GWAS analysis from 1000 Lacaune AI sires for LSCS trait, SheepSNPQTL Project (ANR; final report) (Rupp et al. 2012).

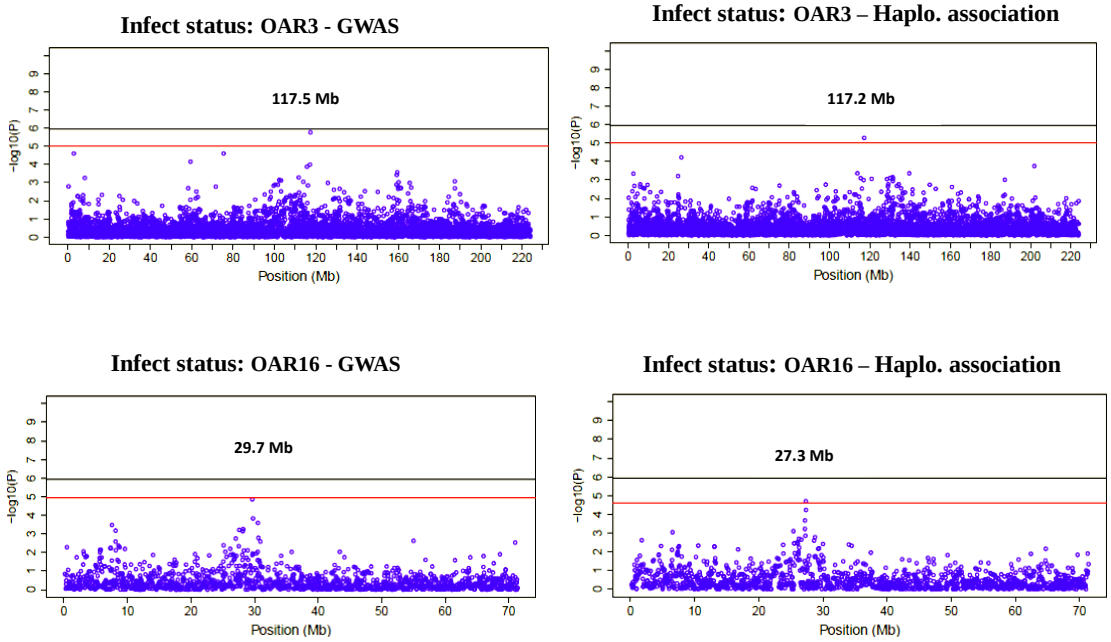

Figure 5. Manhattan plots on OAR3 and OAR16 for Trait Infect status: Region 117.2-117.5Mb on OAR3, 27.3-29.7Mb on OAR16 showed significant association with infection status in both GWAS and Haplotype-based association. The black line indicates the genome-wise significance threshold, and the red line indicates the chromosome-wise significance threshold.

For trait LSCS, no significant region was found to be shared by the two association methods. There are 2 SNPs respectively on OAR14 and OAR24 that were found significant in GWAS and showed moderate p-values in haplotype association, which may be interesting but needs to be further checked by other studies (Table 8).

Table 8. Potential QTL regions in LSCS from both GWAS and Haplotype association mapping.

Chr.	Potential QTL regions (Mb)	$-\log_{10}$ P		Chromosome Significance Threshold of $-\log_{10}$ P
		GWAS	Haplo. Association	
14	47.8-48.7	4.35 *	3.19	4.28
24	30.1-30.9	4.69 *	3.37	4.08

*: chromosome-wise significance

Discussion

Genetic complexity of mastitis and fine-phenotyping of mastitis-related traits

In summary, five regions were found to be significantly associated with mastitis-related traits in the present study (Table 9).

Table 9. A summary of interesting regions detected in this study as candidate QTLs for mastitis.

Chr.	This study		Previous study [a]	
	Trait	Potential QTL regions (Mb)	Trait	Potential QTL regions (Mb)
3	Infection status	**117.2-117.5**	**---**	---
5	Mammary abscess	**95.9-96.7**	---	---
8	Mammary abscess	**14.5-16.6**	---	---
16	Mammary abscess Infection status	**31.1-31.8** **27.3-29.7**	LSCS	29-36
20	Mammary abscess	**49.1-49.5**	LSCS	**49.1-50.6**

a: GWAS analysis from 1000 Lacaune AI sires for LSCS trait, SheepSNPQTL Project (ANR; final report) (Rupp et al. 2012).

Among these five regions, the region on OAR16 (around 30Mb) is shared by three mastitis-related traits detected in this study and also in a previous study in Lacaune sheep. This result indicated a possible candidate QTL around 30Mb on OAR16, highly related to general mastitis infection. In addition, the region on OAR20 is also interesting since it has been related to one SCC-based trait and two mammary abscess traits, which also indicate a general association with the mastitis. Moreover, sheep MHC Class I and Class II genes are located around 25Mb - 27Mb on OAR20 (Sheep Genome Browser; http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv3.1/). Since loci on the same chromosome generally show different levels of linkage, further investigations should be carried out to explore the genetic linkage between MHC loci and the QTL regions detected in this study on OAR20 (49.1 – 49.5Mb). In addition, two regions on OAR5 and OAR8 were specifically related to mammary abscesses. This may infer a specific immune pathway involved in forming mammary abscesses

in response to chronic mastitis. However, due to the limited number of cases with abscess traits in our dataset (20 cases in Abscess (case-control) and 41 cases in abscess (category)), the power of QTL mapping may be not high enough to detect medium and small QTLs for abscesses in this study. More animals and cases are needed in order to detect more QTLs for abscess and at the same time validate the two regions on OAR5 and OAR8.

Mastitis is a complex disease with different pathogens involved, diverse clinical symptoms, and dynamic immune interactions. This means that that identifying genes influencing mastitis susceptibility is very challenging, especially in studies with small population size and non-dense markers. From this study, 5 chromosomes (OAR3, 5, 8, 16, 20) contained candidate QTL regions associated with mastitis. According to previous QTL studies on mastitis in sheep, putative QTLs of SCC were also present on OAR2, 6, 10, 14, 17, 22 (Árnyasi et al. 2009; Jonas et al. 2011; Raadsma et al. 2009), which means that at least 10 chromosomes could be involved in mastitis in sheep. The choice of phenotypes has obvious influence on the result of QTL mapping, and it is very difficult to give a whole picture of genetic basis of mastitis simply by analyzing one phenotype.

As the key symptom of chronic mastitis, mammary abscesses provide valuable phenotypic material in exploring the genetic basis for mastitis, in addition to SCC traits. Two candidate QTLs, one at OAR5 and one at OAR8 were found in this study. More studies, based on different methods and from different sheep breeds, should be carried out to validate these two regions and to find additional QTLs underlying mammary abscesses. Also, genetic parameters of mammary abscesses should be studied to estimate its heritability and genetic correlation with mastitis.

Fine-phenotyping of SCC traits offers another way to better phenotyping mastitis. Generally, traits based on SCC are defined by SCC observations during different lactation period, e.g., test-day SCC, average SCC during 1st lactation or during first

two lactations. In this study, no SNPs were found to be shared by LSCS and Infection status although they are both based on SCC. The reason is probably their big difference in defining SCC traits: Infection status is defined as case-control trait, derived from extreme values of SCC, while LSCS is defined as a continuous variable, estimated by averaging all SCC observations obtained over the first lactation. These differences will have an impact on which QTLs that are identified when different SCC-phenotypes are used. The variety of SCC-based traits can be seen as fine-phenotyping of SCC phenotype, which opens a way to explore the best method in defining SCC in association with mastitis. Different SCC-based traits may give different complexity and accuracy in profiling genetic basis of mastitis.

GWAS and Haplotype-based Association

Genome-wide association mapping is now widely used for QTL detection. Genome wide association study (GWAS) is considered to be more powerful than linkage analysis (LA) for good detection and localization of effects of common alleles with small effects (Grindflek et al. 2011). In GWAS, linkage disequilibrium between markers and causative QTLs is used to detect potential association between markers and phenotypes. Significant markers indicate the existence of QTLs that are in high LD with these markers. GWAS assumes independent markers, ignoring the genetic linkage between SNPs, which may cause false positives because markers are truly linked with each other and their linkage disequilibria are very variable (Grindflek et al. 2011). In comparison, haplotype-based association uses phased haplotype clusters in the association analysis, which takes into account the linkage between markers and this may reduce some false positives in association mapping. Therefore, in this study genotyping data was analyzed simultaneously with GWAS and haplotype-based association.

According to the results from the two association methods, a few significant regions were found shared by both association methods. It is very difficult to say which method

gave better results without further validation. But when comparing the results obtained with both methods with known putative QTLs found in previous studies, haplotype-based association tends to detect more putative QTLs. This may indicate a stronger power and robustness of haplotype-based association test, but it doesn't mean this method is powerful enough to give a whole picture of the associated QTLs. Since haplotype-based association is based on clustered haplotypes, the robustness of haplotype-based association is also highly affected by the accuracy of the haplotype reconstruction. By contrast, GWAS ignores marker linkage and will generally detect more common alleles with small effects, but may contain some false positives. Therefore, it is recommended that GWAS and haplotype-based association are used simultaneously in genome-wide association studies as a confirmation and supplement to each other.

Haplotype reconstruction

Generally, haplotypes can be reconstructed from unphased genotypes by using either familial information (Mendelian segregation and linkage) and/or population information (LD) (Windig & Meuwissen 2004; Scheet & Stephens 2006; Kong et al. 2008). From previous studies, phasing has been mainly based on familiar information from pedigree (Windig & Meuwissen 2004), which phasing leaves nonnegligible proportion of unphased genotypes, especially for the less connected individuals (Druet & Georges 2010). After phasing, IBD probabilities are calculated for each pair of base haplotypes on the identity-by-state (IBS) status of neighborhood markers using a window of a specific number of flanking markers. Haplotype clusters are further grouped based on these IBD probabilities (Meuwissen & Goddard 2001). Using this "standard" method for haplotype reconstruction, the accuracy of haplotyping is limited due to the partially unphased genotypes, and it can be very time-consuming because of the calculation of IBD probabilities, especially with high-dense SNP chips (Druet & Georges 2010).

In this study, a more efficient approach based on Hidden Markov Model (HMM) (Scheet & Stephens 2006) was used for haplotype reconstruction (Druet & Georges 2010). By this approach, phasing is first done by using familial information, as it is in the standard method, to generate partially phased genotypes due to the missing alleles. Afterwards, missing alleles are imputed, phased and clustered into haplotypes together with other phased data, which process is performed iteratively based on the Hidden Markov Model to continuously improve the estimation of haplotype clusters.

One important feature of this method is that it allows accurate imputation of missing genotypes, which increases the effective coverage of phasing data and improves the accuracy of haplotyping (Druet & Georges 2010). The other important advantage of this method is its substantial improvement in computation efficiency than the available standard method using IBD probabilities (Druet & Georges 2010).

In haplotyping reconstruction, one more attention should be paid to the parameter setting in using BEAGLE (Browning & Browning 2007). BEAGLE is a phasing program used together with DAGPHASE to iteratively sample phased haplotypes and construct improved DAG. Three parameters are needed to be set before running BEAGLE: “iteration”, “scale” and “shift”. The parameters of “iteration” and “shift” were set to 10 and 0.0 respectively, according to the instruction from Browning & Browning (2007). For parameter “scale”, it is said to be positively related to “the number of samples per individual performed in BEAGLE sampling”, i.e., larger value of scale indicates more samplings per individual (Browning & Browning 2007). Previous studies showed controversy on setting the parameter of scale: 1.0 (Browning 2006), 2.0 (Druet & Georges 2010), and 4.0 (Browning & Browning 2007). In order to maintain a high accuracy of phasing, scale=2.0 and scale=4.0 were tried in this study since generally more samplings per individual brings higher accuracy. Figure 6 gives the QQ plots generated from QTL mapping (for trait Infect status) based on scale=2.0 and scale=4.0, respectively.

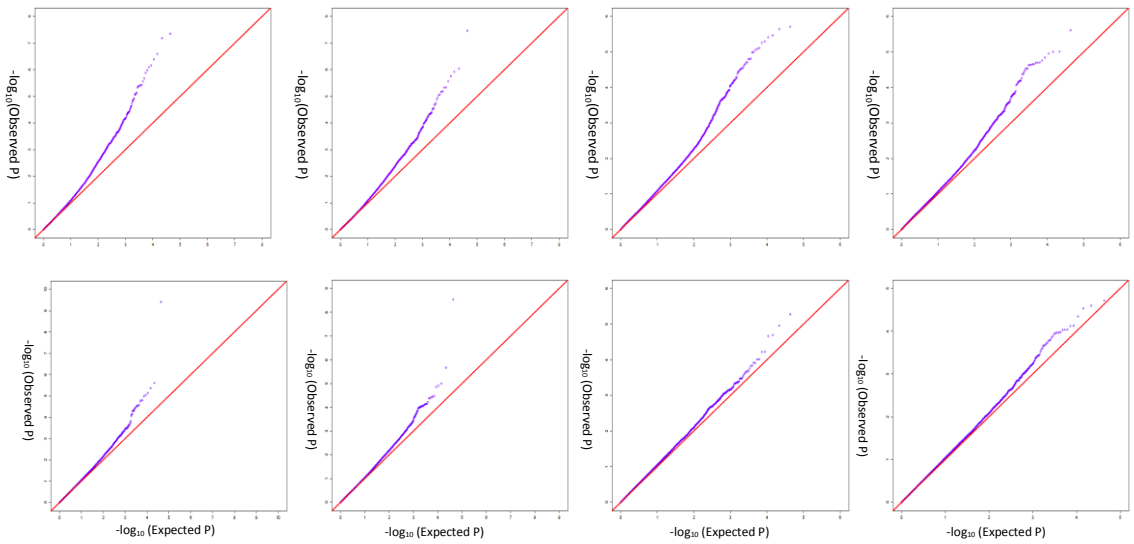

Figure 6. QQ plots of QTL mapping based on phasing with scale=2.0 (first line) and scale=4.0 (second line) for traits: Abscess(case-control), Abscess(category), Infect status, LSCS, from left to right within each line.

It is clear that more false positives appear from phasing with scale=2, which proves a significant influence of the parameter settings when using this method for haplotyping reconstruction. In conclusion, in small dataset like this study, "scale" should be set to at least 4 in order to ensure the accuracy of phasing. Larger "scale" value may increase phasing accuracy to some extent, but also increase computation time. A good balance must be found between computation time and the possible increase in accuracy.

Conclusion

This study introduced the concept of fine-phenotyping of mastitis by including two mammary abscess traits and two SCC-based traits. Association methods included both Genome-wide association mapping using independent markers, and Genome-wide Haplotype-based association mapping using phased haplotype clusters. The result indicated five interesting regions respectively, positioned on OAR3, 5, 8, 16, 20. Different QTLs were associated with different mastitis-related traits. Among the five QTL-regions, OAR16 was believed to contain a strong candidate region around 30Mb

since this region was associated with three different mastitis-related traits (Mammary abscess, Infection status, LSCS). Region 49.1- 49.5Mb on OAR20 was also very interesting for general presence of mastitis, since it was associated with one SCC-based trait and two mammary abscess traits. Regions on OAR5 and OAR8 were found to be specifically related to mammary abscesses, which may indicate a specific immune pathway related to forming mammary abscess in response to chronic clinical mastitis.

The widely-distributed QTLs and the variable genetic basis between mastitis-related traits demonstrate the genetic complexity of mastitis. One way to approach this complexity can be to record more mastitis-related phenotypes, refine the phenotyping of available mastitis-related traits, fine-mapping of candidate regions by more markers, and validate potential QTLs by different mapping approaches. This study introduced mammary abscess as a new phenotype for mastitis and two candidate QTLs highly associated with this phenotype were detected. Due to the complexity of the genetic basis for mastitis, the introduction of abscess and other mastitis-related phenotypes can be useful in mapping additional QTLs/genes involved in mastitis. Moreover, by combining QTL mapping results for different mastitis traits, more accurate QTL localization or ultimately causative mutations can found. This opens up ways for further studies on immune mechanisms in mastitis susceptibility or resistance in dairy sheep. In addition, validated QTLs related to mastitis infection can be applied into Marker Assisted Selection (MAS) for selecting animals with stronger resistance to mastitis.

Acknowledgement

This master thesis was under joint supervision of Rachel Rupp from INRA (France) and Dag Inge Våge from UMB (Norway). They gave me much help and great guidance during my study. I also received very kind help from Anne Ricard (INRA) when using the method of haplotype-based association mapping in this study. Here I give my sincere thanks to them for their precious help. I also thank the director and personnel of SAGA (INRA) to welcoming me in the lab and all EMABG teachers helping to manage all the things for my master courses and thesis.

References

Ali, A.K.A. & Shook, G.E. (1980) An Optimum Transformation for Somatic Cell Concentration in Milk1. *Journal of dairy science*, 63(3), pp.487–490.

Árnyasi, M., Komlósi, I., Lien, S., Czeglédi, L., Nagy, S., Jávor, A. (2009) Searching for DNA markers for milk production and composition on chromosome 6 in sheep. *Journal of Animal Breeding and Genetics*, 126(2), pp.142–147.

Barillet, F., Rupp, R., Mignon-Grasteau, S., Astruc, J. M., Jacquin, M. (2001) Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. *Genetics, selection, evolution: GSE*, 33(4), pp.397–415.

Bergonier, D., Crémoux, Renée de., Rupp, R., Lagriffoul, Gilles., Berthelot, Xavier. (2003) Mastitis of dairy small ruminants. *Vet. Res.*, 34(5), pp.689–716.

Browning, S.R. (2006) Multilocus Association Mapping Using Variable-Length Markov Chains. *American journal of human genetics*, 78(6), pp.903–913.

Browning, S.R. & Browning, B.L. (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *American journal of human genetics*, 81(5), pp.1084–1097.

Burriel, A.R. (1997) Dynamics of intramammary infection in the sheep caused by coagulase-negative staphylococci and its influence on udder tissue and milk composition. *Vet Rec*, 140, pp.419–423.

Carlén, E., Strandberg, E. & Roth, A. (2004) Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish holstein cows. *Journal of dairy science*, 87(9), pp.3062–70. A

Coulon, Jean-B., Gasqui, Patrick., Barnouin, J., Ollier, A., Pradel, P., Pomiès,D. (2002) Effect of mastitis and related-germ on milk yield and composition during naturally-occurring udder infections in dairy cows. *Anim. Res.*, 51(05), pp.383–393.

Davies, G., Genini, S., Bishop, S. C., Giuffra, E. (2009) An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. *animal*, 3(03), pp.415–436.

Druet, T. & Georges, M. (2010) A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics*, 184(3), pp.789–98.

El-Saied, U. M., Carriedo, J. A., De La Fuente, L. F., San Primitivo, F. (1999) Genetic Parameters of Lactation Cell Counts and Milk and Protein Yields in Dairy Ewes. *Journal of Dairy Science*, 82(3), pp.639–644.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. (2009) ASReml User Guide Release 3.0. Available at: www.vsni.co.uk.

Grindflek, E., Lien, S., Hamland, H., Hansen, M., Kent, M., van Son, M., Meuwissen, T. (2011) Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. *BMC Genomics*, 12(1), p.362.

Harmon, R.J. (1994) Physiology of Mastitis and Factors Affecting Somatic Cell Counts. *Journal of Dairy Science*, 77(7), pp.2103–2112.

Heringstad, B., Gianola, D., Chang, Y. M., Ødegård, J., Klemetsdal, G. (2006) Genetic Associations Between Clinical Mastitis and Somatic Cell Score in Early First-Lactation Cows. *Journal of Dairy Science*, 89(6), pp.2236–2244.

Heringstad, B., Klemetsdal, G. & Ruane, J. (2000) Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. *Livestock Production Science*, 64(2-3), pp.95–106.

Hogarth, C. J., Fitzpatrick, J. L., Nolan, A. M., Young, F. J., Pitt, A., Eckersall, P. D. (2004) Differential protein composition of bovine whey: A comparison of whey from healthy animals and from those with clinical mastitis. *PROTEOMICS*, 4(7), pp.2094–2100.

Jonas, E., Thomson, P., Hall, E., McGill, D., Lam, M., Raadsma, H. (2011) Mapping quantitative trait loci (QTL) in sheep. IV. Analysis of lactation persistency and extended lactation traits in sheep. *Genetics Selection Evolution*, 43(1), p.22.

Koivula, M., Mäntysaari, E. A., Negussie, E., Serenius, T. (2005) Genetic and Phenotypic Relationships Among Milk Yield and Somatic Cell Count Before and After Clinical Mastitis. *Journal of Dairy Science*, 88(2), pp.827–833.

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson,G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F.,Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H.,Stefansson, K. (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet*, 40(9), pp.1068–1075.

Leitner, G., Chaffer, M., Shamay, A., Shapiro, F., Merin, U., Ezra, E., Saran, A., Silanikove, N. (2004) Changes in Milk Composition as Affected by Subclinical Mastitis in Sheep. *Journal of Dairy Science*, 87(1), pp.46–52.

Ma, Y., Ryan, C., Barbano, D M., Galton, D M., Rudan, M a., Boor, K J. (2000) Effects of somatic cell count on quality and shelf-life of pasteurized fluid milk. *Journal of dairy science*, 83(2), pp.264–74.

Marogna, G., Rolesu, S., Lollai, S., Tola, S., Leori, G. (2010) Clinical findings in sheep farms affected by recurrent bacterial mastitis. *Small Ruminant Research*, 88(2-3), pp.119–125.

Meuwissen, T.H. & Goddard, M.E., 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genetics, selection, evolution: GSE*, 33(6), pp.605–34.

Ødegård, J., Heringstad, B & Klemetsdal, G, 2004. Short communication: bivariate genetic analysis of clinical mastitis and somatic cell count in Norwegian dairy cattle. *Journal of dairy science*, 87(10), pp.3515–7.

Ogorevc, J., Kunej, T., Razpet, A., Dovc, P. (2009) Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal genetics*, 40(6), pp.832–51.

Raadsma, H., Jonas, E., McGill, D., Hobbs, M., Lam, M., Thomson, P. (2009) Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. *Genetics Selection Evolution*, 41(1), p.45.

Rupp, R., Lagriffoul, G., Astruc, J. M., Barillet, F. (2003) Genetic Parameters for Milk Somatic Cell Scores and Relationships with Production Traits in French Lacaune Dairy Sheep. *Journal of Dairy Science*, 86(4), pp.1476–1481.

Rupp, R., Bergonier, D., Dion, S., Hygonenq, M C.,Aurel, M R.,Robert-Granié, C., Foucras, G. (2009) Response to somatic cell count-based selection for mastitis resistance in a divergent selection experiment in sheep. *Journal of Dairy Science*, 92(3), pp.1203–1219.

Rupp, R.,Boichard, D.,Barbat, A.,Astruc, J M.,Lagriffoul, G.,Barillet, F. (2002) *Selection for mastitis resistance in French dairy sheep.* 2002. Session 9. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, ed., Montpellier: Institut National de la Recherche Agronomique (INRA) LA - English.

Rupp, R. (2012) *Détection de QTL pour des caractères de résistance aux mammites et de production laitière dans un dispositif petite fille de race ovine laitière Lacaune,*

Rupp R., Boichard, D. (2007) Genetics of udder health in dairy ruminants. In *EAAP meeting, Dublin, 26-29 August 2007- Session 34*. Dublin.

Rupp, R., & Boichard, D. (2003) Genetics of resistance to mastitis in dairy cattle. *Veterinary research*, 34(5), pp.671–688.

Rupp, R., Sall ę G.,Barillet, F., Larroque, H., Moreno-Romieux, C. (2011) Genome wide association study of somatic cell counts in a grand-daughter design of French Lacaune AI rams. In *EAAP - 62nd Annual Meeting, Stavanger*.

Scheet, P. & Stephens, M. (2006) A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *The American Journal of Human Genetics*, 78(4), pp.629–644.

Wells, S J, Ott, S L & Hillberg Seitzinger, A. (1998) Key Health Issues for Dairy Cattle—New and Old. *Journal of Dairy Science*, 81(11), pp.3029–3035.

Wells, S.J., Ott, S.L. & Hillberg Seitzinger, A. (1998) Key Health Issues for Dairy Cattle—New and Old. *Journal of Dairy Science*, 81(11), pp.3029–3035.

Windig, J.J. & Meuwissen, T. H. E. (2004) Rapid haplotype reconstruction in pedigrees with dense marker maps. *Journal of Animal Breeding and Genetics*, 121(1), pp.26–39.