Combining different pedigrees to fine-map QTL in the pig.

Thesis committee

Thesis supervisors
Prof. dr. M. A.M. Groenen
Personal chair at Animal Breeding and Genomics Centre
Wageningen University

Dr. D. Milan
Researcher
Institut National de la Recherche Agronomique, Toulouse , France

Thesis co-supervisors
Dr. J. Riquet
Researcher
Institut National de la Recherche Agronomique, Toulouse, France

Dr. R.P.M.A. Crooijmans
Assistant professor at Animal Breeding and Genomics Centre
Wageningen University

Other members
Prof. dr. N. Buys
Katholieke Universiteit Leuven, Leuven, Belgium

Dr. B. Harlizius
The Institute for Pig Genetics, Beuningen , The Netherlands

Dr. P. Le Roy
Institut National de la Recherche Agronomique, Rennes, France

Prof. dr. R.G.F. Visser
Wageningen University

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

Combining different pedigrees to fine-map QTL in the pig.

Flavie Tortereau

Thesis

submitted in fulfillment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof.dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday May 15, 2012
at 1.30 p.m. in the Aula

Tortereau, F.
Combining different pedigrees to fine-map QTL in the pig.

PhD thesis, Wageningen University, the Netherlands (2012)
With references, with summaries in English and Dutch

ISBN 978-94-6173-141-8

Contents

1

General Introduction

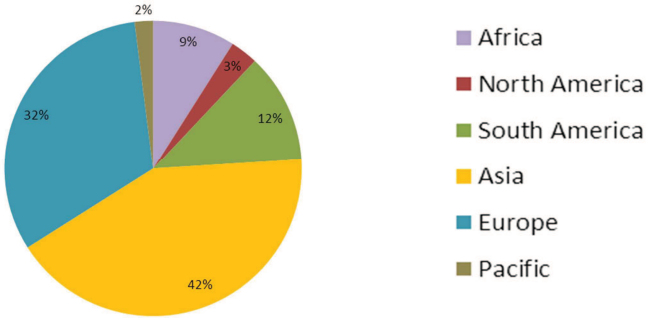
1.1 INTRODUCTION

Pig domestication and breed formation

The major sources of animal proteins, besides fish, come from the few livestock species (cattle, sheep, goat, pig, and poultry) that were domesticated around 10,000 years ago, during the Neolithic age. For most of them, domestication occurred mainly in Asian areas. In swine, two major domestication events took place independently in Europe and China from different sub-species of wild boars (GIUFFRA *et al.* 2000; LARSON *et al.* 2005), and pig breeds resulting from these domestication regions amount to more than 70% of the worldwide pig breed diversity (SCHERF 2000). Most current pig breeds originate from Europe and Asia, and both areas also represent 82% of worldwide pig populations (62% in Asia and 20% in Europe) (FAO 2007).

For more than 7,000 years, domestication processes and more recently pig breeders have relied on naturally occurring mutations to select animals exhibiting favourable traits related to growth, fatness, reproduction, behaviour, and resistance to diseases. As a result, in 2007, the Food and Agriculture Organization (FAO) estimated at 560 the number of existing pig breeds worldwide, most originating from Europe and Asia (figure 1.1). As a comparison, about 990 cattle breeds, 559 goat breeds, 1,129 sheep breeds and 1,132 chicken breeds have been recorded (FAO 2007).

Figure 1.1 – Worldwide distribution of pig breeds. Adapted from (FAO 2007).



Pig selection

About 180 pig breeds have been recorded in Europe (FAO 2007), but only a few of them have been intensively selected for production, including Large White, Landrace, Piétrain, and Duroc. These breeds are genetically very distinct from each other (figure 1.2), although despite strong selection, there is still a high common genetic background within these major European pig breeds (MEGENS *et al.* 2008; SANCRISTOBAL *et al.* 2006).

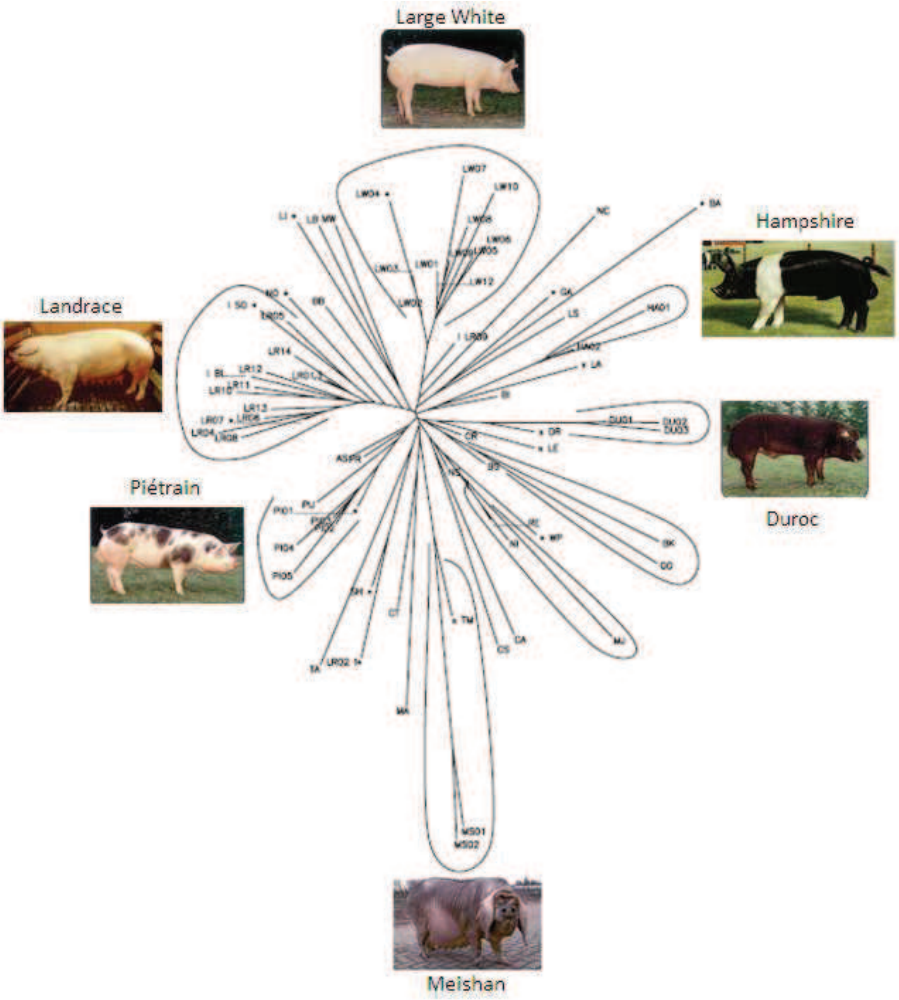


Figure 1.2 – Major European pig breed diversity based on the Reynolds' distance. The Meishan breeds were used as an outgroup. Adapted from SanCristobal *et al*., 2006.

Slaughtered pigs usually originate from multi-stage crosses involving different parental and grandparental breeds to benefit from heterosis. For a given trait, heterosis is defined as the superiority of hybrid animals as compared to the mean of the two parental breeds. Within the Large White, Landrace, and Duroc breeds, different lines have been independently selected as exhibiting more pronounced paternal or maternal features. Dam lines are usually selected for prolificacy, maternal abilities, hardiness, and ease of adaptation whereas sire lines are selected for growth rate, fatness, feed conversion ratio, carcass muscle content, and meat quality. In addition to these purebred lines, various synthetic lines have been created (as dam or sire lines) to benefit from the different

phenotypic potential of original breeds. Most of the synthetic lines produced as male lines are based on the main pig breeds mentioned above. As an example, the Laconie is a synthetic line created from Large White, Piétrain, and Hampshire. Other synthetic lines have been created to be used as dam lines, and some of them were obtained by crossing main breeds with Chinese breeds such as Meishan and Erhualian. Chinese breeds are well known for their high prolificacy and good maternal abilities. Because of their low growth, high fatness, and poor carcass conformation, they cannot be used as purebred lines. However, to benefit from their favourable maternal characteristics, they have been used to create synthetic lines. As examples, we can mention two synthetic French lines created to be used as parents of dam lines: the Taizumu synthetic line, produced by crossing a hyperprolific Large White line with Meishan, and the Tiameslan line, which is based on crosses between Meishan, Jia Xing, and Laconie (IANNUCCELLI *et al.* 2006).

In the last 40 years, pig breeds have been improved by selection for different traits such as reproduction, growth, and fatness. For example, the number of piglets born alive and number of weaned piglets per sow per year increased by almost 3 and 1.3 respectively, from 1992 until 2004 (figure 1.3).

Most traits of interest, such as growth and body composition, are quantitative traits characterized by a normal distribution of phenotypic values and are usually influenced by multiple genes and environmental factors. Trait improvements are thus the result of advances in management and environmental conditions and in genetics. Until now, traditional selection based on performance of animals and their relatives, together with the use of the animal model BLUP (best linear unbiased predictor) methodology to estimate breeding values, has given significant results as evaluated through a research protocol (TRIBOUT *et al.* 2010).

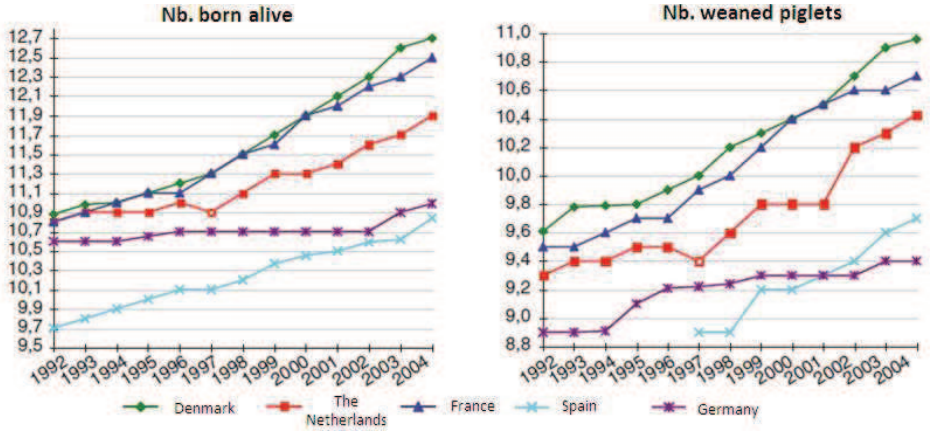


Figure 1.3 - Technical results from five European countries between 1992 and 2004. Data from IFIP (France), Landudvalg for svin (Denmark), Agrovision (the Netherlands), IRTA (Spain), and Verden (Germany).

By inseminating Large White sows with frozen semen from Large White boars born in 1977 or 1998, the genetic progress could be estimated in the French collective Large White breed for different traits. When significant genetic gains were obtained for lean meat content and growth and fatness traits, meat quality turned out to be slightly degraded (TRIBOUT *et al.* 2004). The number of born-alive piglets was also improved, in contrast to the number of born dead piglets, which increased by 0.03 ± 0.01 per year in the same period (TRIBOUT *et al.* 2003).

Because the ability to genetically improve a trait depends on its heritability, genetic correlations with other traits, and ease of measurement, all traits are not equally selected or improved. Genetic correlations can be illustrated from the previous example: growth rate and meat quality traits are negatively correlated, so improving one of the two traits degrades the other. In addition, moderate to highly heritable traits (such as average daily gain) will be selected more rapidly than traits having low heritabilities (such as reproduction traits). Heritability represents the proportion of the phenotypic variance arising from additive genetic effects. Therefore, for each heritable trait, specific regions in the genome control its variations. Different models have been proposed to describe the genetic control of phenotypic variability. The infinitesimal polygenic model assumes that an infinite number of genes influences the trait, each of them having a very small effect. Even if this model is not true because the number of genes is finite, this polygenic model has been successfully used to estimate genetic values because it underlies BLUP methodology, which has been used since the 1990s to genetically improve all livestock species. Another model assumes that genetic values arise from a limited number of loci (quantitative trait loci (QTLs)), each of them having a medium to high effect. This model is supported by results from a meta-analysis showing that up to 35% of QTLs with the highest effects explain 90% of the genetic variance (HAYES and GODDARD 2001). Under this last hypothesis of few QTLs having moderate to large effects, many studies set up in the last 20 years have aimed at detecting such genomic regions in all livestock species, including swine. These studies became possible with advances in molecular biology, and the results obtained enabled breeders to consider new selection strategies.

Molecular resources in swine

Since the 1990s, pig geneticists have benefited from major advances in molecular genetics to better understand the mode of inheritance of complex traits and to identify the causes of phenotypic variations. Among the available genomic resources developed in swine are linkage maps, physical maps (from somatic cell hybrid or radiation hybrid panels analyses), and bacterial artificial chromosomes (BAC) libraries.

Two major linkage maps were published in 1995 (ARCHIBALD *et al.* 1995) and 1996 (ROHRER *et al.* 1996), comprising 239 and 1042 loci, respectively, by the time of their publication. Both maps are based on different genetic markers, the major ones being restriction fragment length polymorphisms and microsatellites. This latter family of markers has been mainly used in livestock genetics because microsatellites are abundant and evenly distributed in the genome, highly polymorphic, and easy to genotype. In pig, between 65,000 and 100,000 microsatellites are supposed to be scattered over the 18 autosomes and the X and Y chromosomes (WINTERO *et al.* 1992). In 2004, Rothschild finally estimated that about 5,000 genetic markers were included in the published linkage maps (ROTHSCHILD 2004).

While linkage maps based on recombination events have been computed, different tools were set up to establish physical maps of the porcine genome. Several BAC libraries have been generated (ANDERSON *et al.* 2000; ROGEL-GAILLARD *et al.* 1999), and a somatic cell hybrid panel (YERLE *et al.* 1996) as well as two radiation hybrid panels (YERLE *et al.* 2002; YERLE *et al.* 1998) were developed. They were genotyped for several markers, including those in the linkage maps and also the expressed sequence tags that enable comparative mapping. The integration of both linkage and physical maps has helped in identifying candidate genes within the genetic regions detected as influencing the variation of quantitative traits. A significant result from such a combination of information from different genetic resources was the discovery of the RN mutation (MILAN *et al.* 2000).

The pig genome is currently being sequenced through shotgun sequencing of BAC clones and whole-genome shotgun sequencing, DNA sampling, and BAC clones originating from the same Duroc female (ARCHIBALD *et al.* 2010). Until now, the sequences of only two livestock species have been published: chicken (HILLIER *et al.* 2004) and cattle (ELSIK *et al.* 2009).

In the last few years, major technological advances have occurred particularly involving the ability to generate millions of sequences through next-generation sequencing (with technologies such as the Illumina Genome Analyzer, Roche's 454 FLX, or Applied Biosystems' SOLiD). By sequencing 19 reduced-representation libraries from five pools of DNA, each pool gathering about 30 individuals from the same breed, up to 372,000 single nucleotide polymorphisms (SNPs) have been identified (RAMOS *et al.* 2009). From this set of SNPs added to the previously detected ones, a final set of 549,282 was available for designing a high-density SNP chip for swine. After filtering on minor allele frequency, distance between SNPs (with build7 as a reference), number of SNPs per chromosome, and Illumina design score, a high-density SNP chip comprising more than 64,232 SNPs distributed all over the pig genome was designed and became available for the whole scientific community. With this new tool, thousands of genotypes can be obtained for one individual in one experiment, a tremendous increase in the number of genotypes that can be used to detect

genomic regions involved in the variation of phenotypic traits. Optimal use of high-density SNP chips requires knowing the exact position of SNPs in the genome. These locations can be obtained in silico when the genomic sequence is available, through radiation hybrids (RH) mapping, or through linkage analysis as has been done for other genetic markers. In pig, these three approaches are possible because resource families as well as radiation hybrid panels already exist.

Genetic variations and phenotypic variability

All these genetic resources have been widely used to detect genomic regions in which variations explain part of the observed phenotypic variability. In pig as well as in other livestock species, major genes were the first loci influencing economically important traits to be identified, and they have been more successfully investigated as compared to QTLs. Usually, a locus has been considered as a major gene when the difference between both homozygote groups reaches around three phenotypic standard deviations or more. Most of the genes that are currently used in pig breeding are major genes: They explain a large part of the phenotypic variation (if not all). When the described polymorphism explains a large proportion of the phenotypic variation, a genotype probability can be determined for each offspring with respect to its phenotypic value. Then, only individuals with the highest probabilities are kept in the studies, which facilitates the mapping of the causative genes. In addition to the genes listed in table 1.2 for pig (most of which are major genes), successful mapping of causative genes has been achieved, for example, in chicken for plumage colour (KERJE *et al.* 2003; KERJE *et al.* 2004; TOBITA-TERAMOTO *et al.* 2000) and in cattle and sheep for muscular hypertrophy (CHARLIER *et al.* 2001; GROBET *et al.* 1997).

When all the phenotypic variation cannot be explained by only one mutation, the identification of the causative genes becomes more difficult because there is no direct relationship between the genotype and the phenotype. The phenotypic difference between both homozygote groups at each locus remains low. Different methodologies have been used to detect such QTL regions.

The first QTL study carried out in a livestock species was published in 1994 in pig (ANDERSSON *et al.* 1994), and many QTL analyses have followed in pig as well as in other livestock species. Since 1994, many QTL detection programs have been performed, resulting in the detection of more than 6,300 QTL regions in pig, 4,600 in cattle, 2,450 in chicken, and 450 in sheep (QTL database on July 18, 2011) (HU and REECY 2007). In pig, a large majority of QTL studies have aimed at detecting regions underlying economically important traits such as growth, fatness, carcass characteristics, reproduction, immune system, and meat quality. Because pig is not only a meat resource but also a major biomedical

model for humans (LUNNEY 2007), some studies have also been dedicated to the detection of regions that are involved in the development of some diseases such as anal atresia (CASSINI *et al.* 2005) or cutaneous melanoma (DU *et al.* 2007).
Whatever the species studied, the general idea of QTL detection is to observe the segregation of alleles from a parent to its offspring and to verify if this segregation is similar to the segregation of the trait of interest (figure 1.4). When the analysed genetic marker is close to a QTL, the phenotypic difference between the offspring that inherited one allele or the other will be higher than when the genetic marker is far from the QTL.
In most livestock species, QTLs have been mainly mapped using backcross (BC) or F2 experimental designs with crosses between pure breeds or commercial lines as divergent as possible to maximize the chance that different QTL alleles segregate in the F1 parents. The production of a BC pedigree is more straightforward because it requires only the production of some F1 sires whereas enough F1 females also need to be produced to generate an F2 pedigree. If BC designs are much more efficient to detect dominance effects, F2 pedigrees return a more general overview of the QTLs in segregation (DARVASI 1998).
In swine, most QTL designs were thus based on F2 or BC (reciprocal or not) pedigrees with European White breeds (Landrace, Large White) and Chinese breeds or Wild Boar used as parental breeds. These breeds display large phenotypic differences in traits such as reproduction, behaviour, growth, fatness, and meat quality to maximize the chance that genes affecting those traits are heterozygous in the F1 generation. In addition, crossing wild boars with modern breeds should result in the detection of genes that have been under selection during pig domestication.

A major exception to these F2 or BC experimental strategies is the detection of QTLs that are already available from the industry for dairy cattle, which is carried out within breeds through a daughter or granddaughter design (GEORGES *et al.* 1995). In a granddaughter design, sires and their sons are genotyped, and the phenotype of each son is its breeding value for the trait of interest estimated through the performances of its daughters. The daughter design is a restriction of the granddaughter design to sons (for which the QTL status is investigated) and their daughters. To minimize the genotyping costs, only a portion of the daughters, usually with extreme phenotypic values, can be genotyped, in a strategy known as selective genotyping (COPPIETERS *et al.* 2009).
More than 30 QTL experimental designs have been set up in pig, and table 1.1 lists the most prominent. Most were used to detect QTLs in a large variety of traits (e.g., growth, fatness, cut weights, meat quality, reproductive traits), resulting in several publications for each design. Some designs not reported in

this table were, however, dedicated to the analysis of particular phenotypes such as ovulation rate and thus involved lines divergently selected for the traits of interest (RATHJE *et al.* 1997).

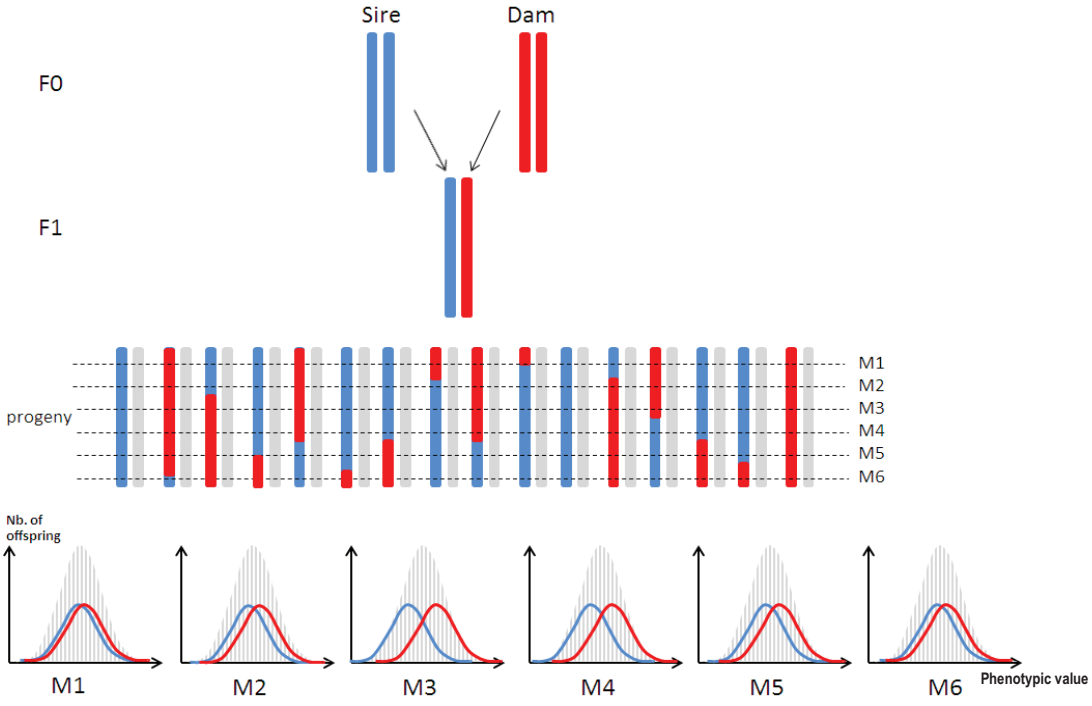


Figure 1.4 - General methodology for QTL detection. Different breeds are used as F0 grandparents, which are crossed to produce F1 animals that have high probabilities to be heterozygous at most loci and particularly at QTLs. F1 are then intercrossed to produce F2 animals, or as presented here, backcrossed to one of the two parental lines. Informative sets of microsatellites (i.e., for which F1 sires are heterozygous) are genotyped on F0, F1, and F2 animals, and phenotypes are recorded on F2 animals. For each microsatellite, offspring are grouped according to the allele (blue or red) they inherited from their sire, and phenotype distributions are compared between both groups. In this example, the highest differences are observed when the offspring are grouped according to the allele they inherited on M3 and M4.

Table 1.1 - Description of major QTL detection experimental set-ups in swine

References	Breeds	Cross	Nb. Markers	Nb. F1	Nb. offspring
(ANDERSSON *et al.* 1994) (ANDERSSON-EKLUND *et al.* 1998)	Large White Wild Boar	F2	236	4 ♂ x 22♀	200
(Rohrer and Keele 1998)	Meishan White-composite	Reciprocal Backcross	154	41♀	540
(de Koning et al. 1999)	Large White Landrace Meishan	F2	127	39♂ x 264♀	844
(Perez-Enciso et al. 2000)	Iberian Landrace	F2	90	6♂ x 73♀	577
(Wada et al. 2000)	Meishan Gottingen miniature pig	F2	318	2♂ x 19♀	265
(Bidanel et al. 2001)	Large White Meishan	F2	123	6♂ x 20♀	1103
(Malek et al. 2001)	Berkshire Yorkshire	F2	125	8♂ x 28♀	525
(NEZER *et al.* 2002)	Pietrain Large White	F2	137	20♂ x 45♀	525
(Lee et al. 2005)	Large White Meishan	F2	117	11♂ x 86♀	600
(Ren et al. 2009)	White Duroc Erhualian	F3	194	F1: 9♂ x 51♀ F2: 62♂ x 149♀	F2:1912 F3: 1530

In pig QTL detection experiments, as in other livestock species, the average density of markers is about one every 20 to 30 cM. Because much information remains between markers, interval mapping was developed (LANDER and BOTSTEIN 1989): Genotypes at positions between two markers are given as probabilities conditional on flanking marker genotypes.

Different statistical methods have been developed to verify whether a QTL is segregating in a given chromosomal region. QTLs have been usually detected with linear regression (HALEY and KNOTT 1992) or maximum-likelihood approaches (LANDER and BOTSTEIN 1989). Both methods give similar results when simple situations are considered (HALEY and KNOTT 1992) or when inbred lines are studied (REBAI *et al.* 1995). However, even if the regression interval mapping is faster in computation, the maximum-likelihood method tends to give better

parameter estimates, particularly when QTL effects are high, when QTLs are located in large marker intervals, or in situations of interaction between or among loci (KAO 2000). Different assumptions are also considered related to the QTL allele frequencies. Under a line-cross model, alternate QTL alleles are supposed to be fixed in the parental breeds. With a half-sib model, no assumption about QTL allele frequencies is made (GEORGES *et al.* 1995; KNOTT *et al.* 1996). Permutation tests usually are performed to estimate the statistical significance of the detected QTLs (CHURCHILL and DOERGE 1994). In 1995, Lander and Kruglyak proposed three different significance levels: suggestive (statistical evidence expected to randomly occur once per genome scan), significant (statistical evidence expected to randomly occur 0.05 times per genome scan), or highly significant (statistical evidence expected to randomly occur 0.001 times per genome scan) linkages (LANDER and KRUGLYAK 1995).
Most of the genetic markers currently used in pig breeding are the results of linkage analyses carried out within families, followed by further fine-mapping.

Genetic markers in pig selection

Until now, in pig, only a few causative polymorphisms have been described: in the *PRKAG3* (known as RN) and *RYR1* (known as halothane) genes for meat quality (FUJII *et al.* 1991; MILAN *et al.* 2000), in the *MC4R* gene for growth, fatness, and feed intake (KIM *et al.* 2000), and in the *IGF2* gene for muscling (VAN LAERE *et al.* 2003). Other polymorphisms are highly associated with different traits, such as in the *CAST* gene for tenderness (CIOBANU *et al.* 2004) and in the *ACACA* gene for some fatty acid concentrations (MUNOZ *et al.* 2007), but until now, no functional proofs of their causality have been provided. Pig breeding companies have already used some of these genes (table 1.2). The halothane gene was the first gene to be used in the pig industry, followed by the *PRKAG3* gene and some others that influence major economically important traits such as reproduction and resistance to disease. In addition to these well-described polymorphisms, other markers can be used for marker-assisted selection. Those markers are usually used to trace QTL regions in the selected lines. For these regions, the causative polymorphisms are not known and the underlying genes can also be unknown, but specific sets of SNPs can be designed to allow distinction of the different alleles at the QTLs.
From the genotypes at this causative polymorphism or at the targeted QTL regions, the molecular information can be computed differently based on the application (DEKKERS and HOSPITAL 2002). The molecular information can be considered as a complement to traditional selection, with genotypes being used to pre-select animals that will go through performance testing (i.e., exclusion of part of the candidate for selection based on this first diagnosis). Molecular information can also be applied in combination with phenotypes or estimated

breeding values to calculate an index based on which animals will be selected. Moreover, the molecular information is accessible from early stages of life, so early diagnostics can be done. Using genetic markers is also highly useful for traits that are difficult or expensive to measure, such as traits that can be recorded only in one of the two sexes or for carcass traits.

Table 1.2 - Main genes in which polymorphisms are used in commercial pig breeding

Traits	Gene	References
Stress susceptibility	*RYR*	(FUJII *et al.* 1991)
Coat color	*cKIT*	(MARKLUND *et al.* 1998)
Meat Quality	*RYR*	(FUJII *et al.* 1991)
	PRKAG3	(MILAN *et al.* 2000)
Feed Intake	*MC4R*	(KIM *et al.* 2000)
Growth and Composition	*MC4R*	(KIM *et al.* 2000)
	IGF2	(VAN LAERE *et al.* 2003)
	CAST	(CIOBANU *et al.* 2004)
Disease resistance	*F18*	(MEIJERINK *et al.* 2000)
	K88	(JORGENSEN *et al.* 2003)
Reproduction	*ESR*	(ROTHSCHILD *et al.* 1996)
	PRLR	(VINCENT *et al.* 1998)
	RBP4	(ROTHSCHILD *et al.* 2000)

Adapted from (DEKKERS 2004).

Fine-mapping of QTLs

All the detected QTLs are valuable information that could be used in selection schemes provided their localization interval is shortened. Most of them are still characterized by large intervals of 20 to 40 cM, where hundreds of genes can be found. Many studies have thus been dedicated to the refinement of QTL regions. In pig, until now only the SSC2 QTL that the *IGF2* gene underlies has been fine-mapped up to the causal mutation. In other livestock species, the number of completely fine-mapped QTLs is also very low. In cattle, two polymorphisms have been described to influence milk production, in the *DGAT1*

(GRISART *et al.* 2004) and *ABCG2* (COHEN-ZINDER *et al.* 2005) genes. In sheep, two mutations underlying muscle development have been described: in the 3'-UTR region of the *MSTN* gene (CLOP *et al.* 2006) and in the 90-kb *DLK1-GTL2* intergenic region (FREKING *et al.* 2002; SMIT *et al.* 2003).

This low number of fine-mapped QTLs highlights how difficult it is to detect causative polymorphisms underlying QTLs. Limitations to fine-mapping are directly linked to the different elements involved in QTL detection through linkage analyses:

- the number of offspring in which the different alleles of the QTLs are segregating and in which the phenotypes are recorded;
- the density of informative genetic markers; and
- appropriate methodologies to identify correlations between the segregation of marker alleles with the distribution of the phenotype.

The first limitation in the reduction of QTL confidence intervals is the **number of recombinant chromosomes** that can be studied. This number is directly linked to the number of animals included in the analysis, as suggested by the inverse relationship between confidence interval and sample size in BC and F2 designs (DARVASI and SOLLER 1997). To achieve a power of 0.8 to detect a QTL with an effect of 0.3, 0.5, or 0.7 phenotypic standard deviation and considering a type I error of 0.05, each sire should have 351, 128, or 66 informative offspring, respectively (VAN DER WERF *et al.* 2007). In the approximately 40 QTL detection experiments that have been set up in pig since 1994, two to 30 F1 sires were usually studied, each of them having 50 to 100 offspring. Finally, in those experiments, phenotypes were recorded on hundreds of F2 or BC animals, commonly ranging from 200 to 1,000 and more exceptionally around 2,000. Therefore, the power to detect a QTL is limited to QTLs with medium to high effects. The increase in the number of animals included in the study is thus a way to increase the accuracy of QTL detection and to better identify QTLs with smaller effects.

The **number of markers** used in QTL analyses is another key parameter. In swine, QTL detection studies have been mainly conducted with microsatellites from the two major published linkage maps (ARCHIBALD *et al.* 1995; ROHRER *et al.* 1996). The first genome scans in pig were carried out with 100 to 250 genetic markers with an average spacing of 20 to 30 cM. Simulation studies concluded that in linkage analyses done with F2 or BC pedigrees, the reduction of marker density beyond 10 cM had little effect on the power of QTL detection and the standard errors of the estimates of gene effects (DARVASI *et al.* 1993; PIEPHO 2000).

To overcome these first two limitations, fine-mapping strategies that have been applied until now aimed at increasing the number of markers in the QTL regions so that the recombination events were better characterized, and at studying

additional animals that were recombinant in the QTL interval of interest. One way to guarantee the study of recombinant chromosomes within the targeted QTL region is to use marker-assisted backcrossing to produce animals carrying these desired chromosomes (figure 1.5). This method has already yielded significant results for the fine-mapping of a QTL underlying fatness traits segregating on pig chromosome 4, with a decrease of the interval from 70 to 3.3 cM (BERG *et al.* 2006; MARKLUND *et al.* 1999).

The development of BC pedigrees is, however, expensive and time consuming, particularly when the size of the interval becomes smaller, because the probability that recombination occurs in an interval decreases with the size of the interval.

Therefore, most fine-mapping strategies have been based on the analysis of existing populations. These studies rely on linkage disequilibrium (LD). LD is defined as a non-random association of alleles at different loci and is estimated at the level of a population (FALCONER and MACKAY 1996). Markers are said to be in LD when the frequency of the combination of alleles is higher than the frequency that is estimated based on the individual allele frequencies.

Linkage disequilibrium was first used to fine-map causative polymorphisms responsible for recessive diseases in inbred children through homozygosity mapping (LANDER and BOTSTEIN 1987). The underlying idea is that the disease results from a single polymorphism that arose in a unique ancestor, so affected individuals are characterized by homozygous identical by descent (IBD) haplotypes in which the mutated allele is embedded. These IBD homozygous regions are associated with a very high level of LD, and their length depends on the recombinations that occurred through the generations since the mutation appeared.

For fine-mapping of QTLs, the same hypothesis as that underlying homozygosity mapping is considered: The causative polymorphism underlying a QTL appeared in a particular haplotype and has been subsequently transmitted to the next generations. All the carriers of this mutation will thus have a common IBD chromosomal segment surrounding the mutation, which will be gradually shortened because of recombination events (figure 1.6). Therefore, the QTL interval can be fine-mapped to the region characterized by a high level of LD between the markers and the causative polymorphism.

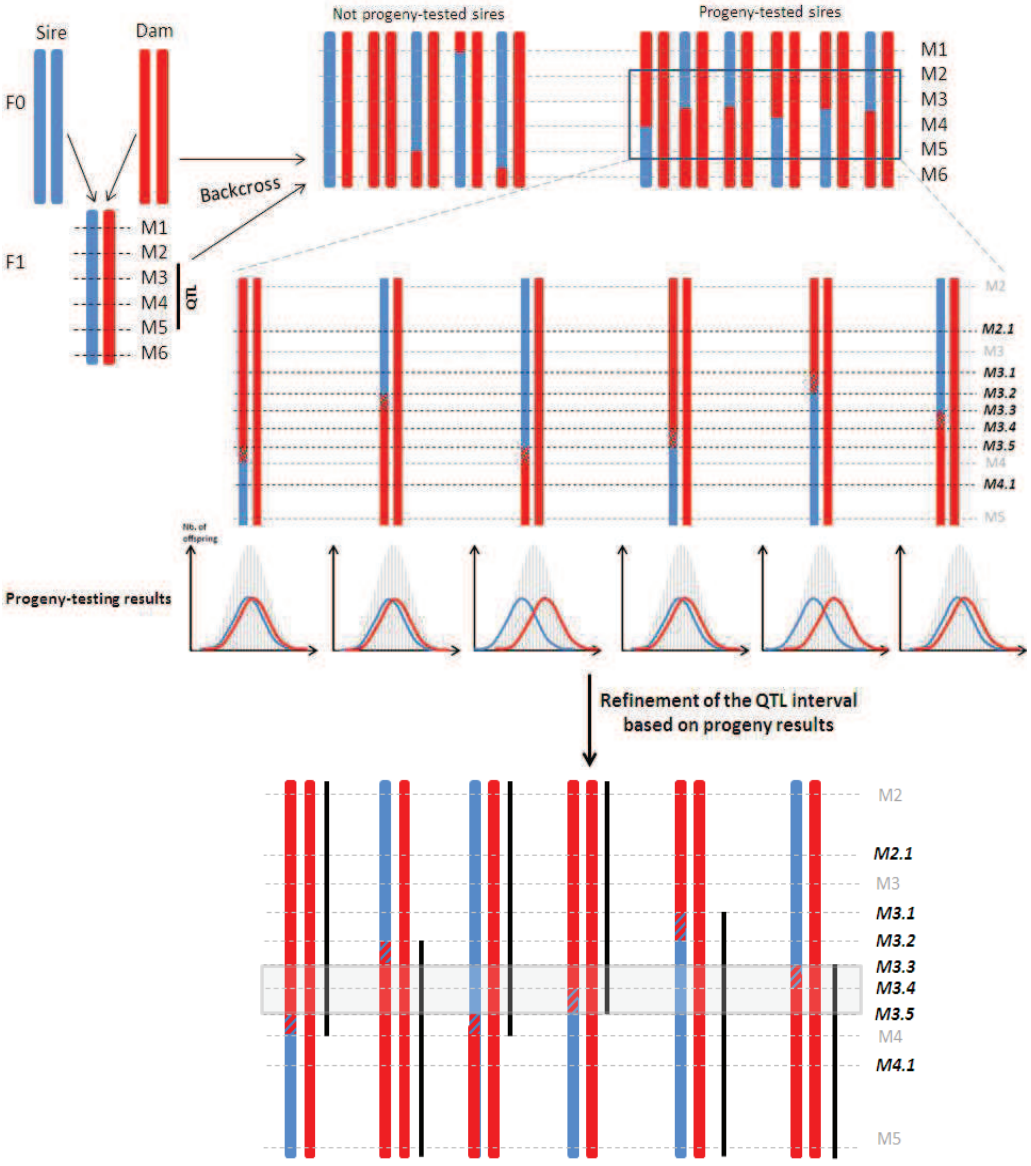


Figure 1.5 - Fine-mapping of a QTL by marker-assisted backcrossing. Following the example of figure 1.2, a QTL was detected around the interval M3–M4. To fine-map it, only backcross sires carrying recombination events within the QTL interval are progeny tested. Additional markers (in italic) are genotyped to precisely identify the recombination points; the hatched areas remaining are of undetermined origin. Blue (red) distribution is obtained from the progeny that inherited the recombinant red/blue (non-recombinant red) chromosome. When a QTL is detected, the sire is heterozygous at the QTL, and this QTL should thus be in the heterozygous region. On the other hand, when no QTL is detected, the sire is homozygous at the QTL, which should therefore segregate in the homozygous region. The region defined by the overlap of each sire interval finally gives a reduced QTL interval. The new, shorter QTL interval is thus between *M3.3* and *M3.5*.

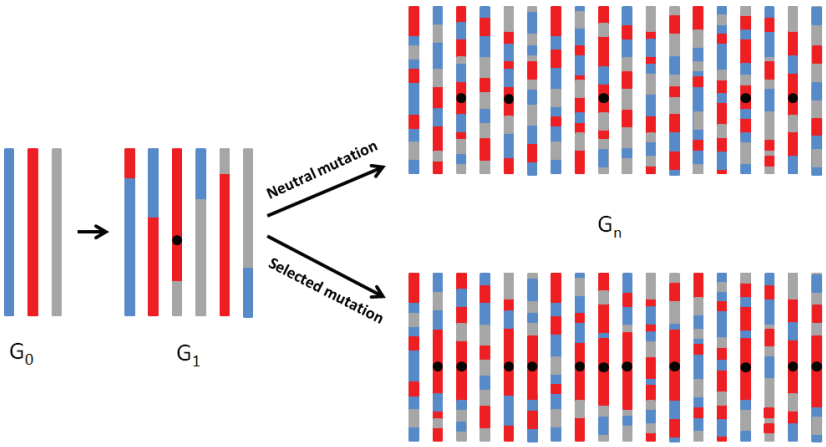


Figure 1.6 – Linkage disequilibrium around a neutral or selected mutation. The mutation (black circle) appears in the G1 generation in a red haplotype. If this mutation is neutral, its frequency will vary with genetic drift. If this mutation is selected, its frequency will increase in the G_n generation. In addition, the mutation will always be found within a red haplotype, which has different length due to recombination events.

A condition to exploit LD is that the density of markers must fit with the extent of LD in the studied population. The extent of LD is an evaluation of the distance over which there is LD. It mainly depends on mutation and recombination and, to a broader extent, on the structure of the population and its growth, on genetic drift, inbreeding, selection, and hitchhiking (ARDLIE *et al.* 2002). In European pig breeds, LD can extend up to 400 kb whereas it usually extends over 10 kb in Chinese breeds. This information implies that the marker densities needed to exploit LD must be around one marker every 0.1 cM (AMARAL *et al.* 2008).

Until recently, high densities of markers were achieved by developing new markers within QTL regions. However, these regions were usually too large (10–40 cM) to be considered for complete sequencing, so the strategy was to identify candidate genes that will be subsequently sequenced. Candidate genes were first retained because they were mapped within QTL intervals and then because of their known function in other mammals (human or mouse). Nevertheless, successful candidate gene analyses have been carried out directly, bypassing the mapping step (ROTHSCHILD *et al.* 1996). Identified candidate genes were then sequenced (completely or partly, with coding sequences being preferentially sequenced), and the identified polymorphisms were subsequently genotyped on different individuals originating from one or several breeds. These polymorphisms were tested either through association analyses, with a significant association between the alleles and the phenotypic variations indicating that the tested marker is in strong LD with the causative polymorphism, or they were used to refine QTL intervals through haplotype sharing. The principle of haplotype sharing is to identify the IBD haplotype

surrounding the causative polymorphisms. This method was successfully used to fine-map the QTL underlying muscle mass segregating on SSC2 to a 250 kb interval (NEZER *et al.* 2003). Subsequently, the 250-kb region identified on SSC2 was completely sequenced, and only one mutation was identified and confirmed to be the causative polymorphism (VAN LAERE *et al.* 2003). Similarly, in dairy cattle, haplotype sharing was used to fine-map to 5 cM a QTL underlying milk production in dairy cattle (RIQUET *et al.* 1999).

Since the availability of genome-wide high-density SNP chips, LD can be considered at the level of the entire genome, and not only at the level of candidate regions or genes.

As shown on figure 1.6, the frequency of a selected mutation can rapidly increase in a population, causing at the same time an increase in the frequency of alleles that are in LD with this mutation. Such variations in allele frequencies can help in mapping and fine-mapping QTLs and are called signatures of selection. In humans, regions under selection since about 6,500 years ago have been identified by applying the extended haplotype homozygosity test (SABETI *et al.* 2002). This test is applied within a population to detect signatures of selection by analysing haplotypes surrounding a small region characterized by a strong LD. The underlying idea is that regions surrounding selected mutations are characterized by a high frequency and long-conserved haplotypes (figure 1.6). This test was also used to detect signatures of selection in dairy cattle (HAYES *et al.* 2008; QANBARI *et al.* 2010). Signatures of selection were also detected among different pig breeds at the KIT locus, which underlies coat colour variation (FONTANESI *et al.* 2010). Very recent selection is also detectable at the level of the genome. In a Sino-European synthetic line obtained by repeated intercrosses between Large White and Meishan breeds and selected on BFT at 100 kg, a signature of selection was detected on SSC7 by comparing allele frequencies obtained in the second and the fifth generations. In this region, the frequency of Chinese alleles in the animals from the fifth generation reached 80%, which is significantly higher than what is expected under genetic drift (RIQUET *et al.* 2006).

Through the analysis of LD, genomic regions that have been under selection are thus detectable and do not require phenotypic data. However, in a fine-mapping context, it is interesting to identify regions that underlie a specific trait of interest. To do so, phenotypic variations are analysed with respect to marker variations through association analyses. As presented earlier, association studies used to be restricted to candidate genes, but now, with genome-wide high-density SNP chips, they can be extended to the whole genome and are known as genome-wide association studies (GWAS). In domestic animals, spectacular results have been obtained by using this method to fine-map and detect the causative polymorphisms of recessive genetic defects (CHARLIER *et al.* 2008) and to dramatically reduce the localization interval of causative variants underlying white spotting and hair ridge in dogs (KARLSSON *et al.* 2007). These

two experiments were based on the genotyping of case (affected) and control (non-affected) animals. For complex traits, a GWAS is usually carried out through the use of linear models including an SNP effect, fixed effects, and a polygenic value (GODDARD and HAYES 2009). As compared to the results obtained with monogenic traits, GWAS applied to quantitative traits returned less-striking results, although significant associations were detected, for example, for milk production traits in dairy cattle (DAETWYLER *et al.* 2008) and for resistance to salmonella in chicken (HASENSTEIN *et al.* 2008).
Based on the extent of LD estimated in different pig breeds, Amaral *et al.* projected that around 30,000 informative markers would be required for a whole-genome approach within European populations (AMARAL *et al.* 2008). With the availability of the porcine 60k SNP chip, the analysis of LD at the level of the genome is thus possible in swine. First GWAS carried out in pig have thus been recently published for boar taint (DUIJVESTEIJN *et al.* 2010; GRINDFLEK *et al.* 2011; RAMOS *et al.* 2011), body composition (FAN *et al.* 2011), and lifetime reproductive traits (ONTERU *et al.* 2011).
New developments in molecular genetics in this way have enabled consideration of new methodologies to map and fine-map QTLs, based on LD. With a high-density SNP chip, the number of markers no longer seems to be a limitation in the fine-mapping process.

Finally, the last important limitation to the fine-mapping of QTLs is the **ability to detect "complex" QTLs**. Most QTLs have been detected using models fitting additive and dominant effects of the QTLs. The additive and dominance QTL effects are given respectively as half the phenotypic difference between the two groups of homozygous animals and as the deviation of the heterozygous animals from the mean estimated with the two groups of homozygotes. Both effects can be estimated under the line-cross model, whereas only the additive effect is estimated under the half-sib model. By applying these models, thousands of QTLs with medium to large effects have been detected in pig, as well as in other species, underlying a large variety of traits. However, in swine and chicken, few QTLs, if any, have been detected for some traits whereas breeds involved in the studied crosses displayed large phenotypic differences for the trait of interest (GEORGES 2007). Similarly, in humans, most of the GWAS analyses returned significant results, but considered together, they explained only a small percentage of the studied trait's heritability (MAHER 2008). Epigenetics and epistatic interactions have been proposed to partly explain why some QTLs cannot be detected in experimental designs.
Epigenetics covers the study of changes in gene activity and expression (e.g., from histone modifications, methylation of the cytosine residue in DNA, and RNA interference) that do not depend on modifications of the genomic sequence (SLATKIN 2009). In livestock, the importance of epigenetic effects on phenotypic variations has also been reported. In pig, a major QTL has been

explained by a substitution in the third intron of the *IGF2* gene, which is maternally imprinted; only the allele transmitted by the father is expressed (VAN LAERE *et al.* 2003). In sheep, the callipyge phenotype results from a mutation in the 90-kb *DLK1-GTL2* intergenic region and is characterized by polar overdominance effects: The phenotype is expressed only in heterozygous individuals having inherited the mutation from their sire (COCKETT *et al.* 1996; FREKING *et al.* 2002; GEORGES and COCKETT 1996). Statistical models used for QTL detection have been adapted to consider parental imprinting (DE KONING *et al.* 2000; KNOTT *et al.* 1998) and polar overdominance (BOYSEN *et al.* 2010; KIM *et al.* 2004). Examples mentioned above have been detected because mutations occurred in this particular genes or regions, but the question of broader epigenetic variation effects on phenotypic variations remains (JOHANNES *et al.* 2008).

In addition to epigenetics, interactions between or among two or more loci (referred to as epistasis) are another complex phenomenon that can make QTL detection tough and that can contribute to a large part of phenotypic variation. Until now, all the described causative polymorphisms underlying QTLs result from a single polymorphism, but epistatic interactions studies in mice have indicated that interaction effects can contribute up to 33% of the total phenotypic variation (BROCKMANN *et al.* 2000). In livestock, evidence of epistatic interactions has already been reported in chicken (CARLBORG *et al.* 2004; CARLBORG *et al.* 2006) and pig (ESTELLE *et al.* 2008; NOGUERA *et al.* 2009). The failure to detect interacting loci has usually been imputed to designs that were not powerful enough (VARONA *et al.* 2002), given that only considering epistasis between two loci increases the number of genotypic classes from three to nine under a line-cross hypothesis and that this number would be even higher under a half-sib model. Even if epistasis can partly explain phenotypic variations, accurately detecting it in livestock species has remained challenging.

1.2 AIM AND OUTLINE OF THIS THESIS

In comparison with the thousands of QTLs that were detected, the number of causative genes/polymorphisms identified, also called QTNs (quantitative trait nucleotides), is very low, despite the high number of studies dedicated to the refinement of QTL regions. This large gap between the number of detected QTLs and the number of causative polymorphisms that have been described exists not only in pig but in all other livestock species. Even if no QTN has been detected, the detected QTLs can be used for marker-assisted selection, provided their localization interval is small enough to be traced by a set of markers.

The overall aim of this thesis was to fine-map QTLs segregating on pig chromosomes 2, 4, and 6 and to characterize their mode of inheritance. To achieve this aim, the different fine-mapping limitations mentioned in this introduction were considered: the number of animals, the number of markers, and different models used to detect these regions. All the analyses presented here rely on the combination of different experimental pedigrees. This thesis contains five research chapters, each focusing on different limitation points and in a general conclusion, the different QTL fine-mapping strategies and the relevance of working with experimental pedigrees is discussed.

This thesis was conducted under collaboration between the Laboratoire de Génétique Cellulaire at INRA Toulouse and the Animal Breeding and Genetics group of Wageningen University. Both labs developed pig F2 crosses to detect QTLs underlying growth, fatness, and reproductive traits (table 1.1). The analyses of these two pedigrees based on European White breeds and Meishan returned several QTLs, a portion of which were similar (BIDANEL *et al.* 2001; BIDANEL *et al.* 2008; DE KONING *et al.* 1999; DE KONING *et al.* 2000; DE KONING *et al.* 2001a; DE KONING *et al.* 2001b; MILAN *et al.* 2002). The main discordant result was the absence of an imprinting effect in the *IGF2* region in the French design whereas this effect was described in the Dutch design, with the *IGF2*-intron3-G3072A mutation segregating in both pedigrees.

These two pig pedigrees were first combined into a single study to increase the power of the linkage analysis. The increase in the crossover densities should provide greater significance and mapping precision of sheared QTLs as well as the opportunity to exclude false suggestive QTLs and detecting QTLs with lower effects. Results for three chromosomes (SSC2, SSC4, and SSC6) analysed for nine traits common to both pedigrees are reported in chapter 2.

As mentioned above, divergent results were previously obtained for the genetic determinism of the SSC2p chromosomal region. The two pig pedigrees were thus combined to further investigate this point. The *IGF2*-intron3-G3072A polymorphism segregates in both designs, but its imprinting status was not confirmed in one of the two pedigrees. Chapter 3 presents simulation studies that were carried out considering the segregation of this polymorphism as well as a second QTL genetically linked to the *IGF2* gene to understand why such different results were previously obtained.

The effect of SSC2p on fatness traits was confirmed in the two pedigrees, so the fine-mapping of this region was carried out through a marker-assisted backcrossing strategy. The analysis of this new design, presented in chapter 4, aimed at confirming, fine-mapping, and describing the mode of segregation of a QTL underlying fatness traits and segregating on SSC2 around 40–60 cM.

In 2009, the high-density SNP chip became available for swine geneticists. The first high-density recombination map of *Sus scrofa* using data from the 60K SNP chip is presented in chapter 5. It was calculated from the genotypes of four different pedigrees (developed by different members of the Swine Genome

Sequencing Consortium), and this new genetic tool will be of major interest to further investigate the pig genome, particularly in relation to phenotypic variations.

REFERENCES

AMARAL, A. J., H. J. MEGENS, R. P. CROOIJMANS, H. C. HEUVEN and M. A. GROENEN, 2008 Linkage disequilibrium decay and haplotype block structure in the pig. Genetics **179:** 569-579.

ANDERSON, S. I., N. L. LOPEZ-CORRALES, B. GORICK and A. L. ARCHIBALD, 2000 A large-fragment porcine genomic library resource in a BAC vector. Mammalian genome : official journal of the International Mammalian Genome Society **11:** 811-814.

ANDERSSON-EKLUND, L., L. MARKLUND, K. LUNDSTROM, C. S. HALEY, K. ANDERSSON *et al.*, 1998 Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross. J Anim Sci **76:** 694-700.

ANDERSSON, L., C. S. HALEY, H. ELLEGREN, S. A. KNOTT, M. JOHANSSON *et al.*, 1994 Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science **263:** 1771-1774.

ARCHIBALD, A. L., L. BOLUND, C. CHURCHER, M. FREDHOLM, M. A. GROENEN *et al.*, 2010 Pig genome sequence--analysis and publication strategy. BMC Genomics **11:** 438.

ARCHIBALD, A. L., C. S. HALEY, J. F. BROWN, S. COUPERWHITE, H. A. MCQUEEN *et al.*, 1995 The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm Genome **6:** 157-175.

ARDLIE, K. G., L. KRUGLYAK and M. SEIELSTAD, 2002 Patterns of linkage disequilibrium in the human genome. Nature reviews. Genetics **3:** 299-309.

BERG, F., S. STERN, K. ANDERSSON, L. ANDERSSON and M. MOLLER, 2006 Refined localization of the FAT1 quantitative trait locus on pig chromosome 4 by marker-assisted backcrossing. BMC Genet **7:** 17.

BIDANEL, J. P., D. MILAN, N. IANNUCCELLI, Y. AMIGUES, M. Y. BOSCHER *et al.*, 2001 Detection of quantitative trait loci for growth and fatness in pigs. Genet Sel Evol **33:** 289-309.

BIDANEL, J. P., A. ROSENDO, N. IANNUCCELLI, J. RIQUET, H. GILBERT *et al.*, 2008 Detection of quantitative trait loci for teat number and female reproductive traits in Meishan X Large White F2 pigs. Animal **2:** 813-820.

BOYSEN, T. J., J. TETENS and G. THALLER, 2010 Detection of a quantitative trait locus for ham weight with polar overdominance near the ortholog of the callipyge locus in an experimental pig F2 population. Journal of Animal Science **88:** 3167-3172.

BROCKMANN, G. A., J. KRATZSCH, C. S. HALEY, U. RENNE, M. SCHWERIN *et al.*, 2000 Single QTL effects, epistasis, and pleiotropy account for two-thirds of the

phenotypic F(2) variance of growth and obesity in DU6i x DBA/2 mice. Genome Res **10:** 1941-1957.

CARLBORG, O., P. M. HOCKING, D. W. BURT and C. S. HALEY, 2004 Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. Genet Res **83:** 197-209.

CARLBORG, O., L. JACOBSSON, P. AHGREN, P. SIEGEL and L. ANDERSSON, 2006 Epistasis and the release of genetic variation during long-term selection. Nat Genet **38:** 418-420.

CASSINI, P., A. MONTIRONI, S. BOTTI, T. HORI, H. OKHAWA *et al.*, 2005 Genetic analysis of anal atresia in pigs: evidence for segregation at two main loci. Mamm Genome **16:** 164-170.

CHARLIER, C., W. COPPIETERS, F. ROLLIN, D. DESMECHT, J. S. AGERHOLM *et al.*, 2008 Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature genetics **40:** 449-454.

CHARLIER, C., K. SEGERS, L. KARIM, T. SHAY, G. GYAPAY *et al.*, 2001 The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. Nat Genet **27:** 367-369.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971.

CIOBANU, D. C., J. W. BASTIAANSEN, S. M. LONERGAN, H. THOMSEN, J. C. DEKKERS *et al.*, 2004 New alleles in calpastatin gene are associated with meat quality traits in pigs. J Anim Sci **82:** 2829-2839.

CLOP, A., F. MARCQ, H. TAKEDA, D. PIROTTIN, X. TORDOIR *et al.*, 2006 A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature genetics **38:** 813-818.

COCKETT, N. E., S. P. JACKSON, T. L. SHAY, F. FARNIR, S. BERGHMANS *et al.*, 1996 Polar overdominance at the ovine callipyge locus. Science **273:** 236-238.

COHEN-ZINDER, M., E. SEROUSSI, D. M. LARKIN, J. J. LOOR, A. EVERTS-VAN DER WIND *et al.*, 2005 Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome research **15:** 936-944.

COPPIETERS, W., T. H. MES, T. DRUET, F. FARNIR, N. TAMMA *et al.*, 2009 Mapping QTL influencing gastrointestinal nematode burden in Dutch Holstein-Friesian dairy cattle. BMC Genomics **10:** 96.

DAETWYLER, H. D., F. S. SCHENKEL, M. SARGOLZAEI and J. A. ROBINSON, 2008 A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. Journal of dairy science **91:** 3225-3236.

DARVASI, A., 1998 Experimental strategies for the genetic dissection of complex traits in animal models. Nat Genet **18:** 19-24.

DARVASI, A., and M. SOLLER, 1997 A simple method to calculate resolving power and confidence interval of QTL map location. Behav Genet **27:** 125-132.

DARVASI, A., A. WEINREB, V. MINKE, J. I. WELLER and M. SOLLER, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943-951.

DE KONING, D. J., B. HARLIZIUS, A. P. RATTINK, M. A. GROENEN, E. W. BRASCAMP *et al.*, 2001a Detection and characterization of quantitative trait loci for meat quality traits in pigs. J Anim Sci **79:** 2812-2819.

DE KONING, D. J., L. L. JANSS, A. P. RATTINK, P. A. VAN OERS, B. J. DE VRIES *et al.*, 1999 Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics **152:** 1679-1690.

DE KONING, D. J., A. P. RATTINK, B. HARLIZIUS, M. A. GROENEN, E. W. BRASCAMP *et al.*, 2001b Detection of characterization of quantitative trait loci for growth and reproduction traits in pigs. Livestock Production Science **72:** 185-198.

DE KONING, D. J., A. P. RATTINK, B. HARLIZIUS, J. A. VAN ARENDONK, E. W. BRASCAMP *et al.*, 2000 Genome-wide scan for body composition in pigs reveals important role of imprinting. Proc Natl Acad Sci U S A **97:** 7947-7950.

DEKKERS, J. C., 2004 Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci **82 E-Suppl:** E313-328.

DEKKERS, J. C., and F. HOSPITAL, 2002 The use of molecular genetics in the improvement of agricultural populations. Nature reviews. Genetics **3:** 22-32.

DU, Z. Q., S. VINCENT-NAULLEAU, H. GILBERT, F. VIGNOLES, F. CRECHET *et al.*, 2007 Detection of novel quantitative trait loci for cutaneous melanoma by genome-wide scan in the MeLiM swine model. Int J Cancer **120:** 303-320.

DUIJVESTEIJN, N., E. F. KNOL, J. W. MERKS, R. P. CROOIJMANS, M. A. GROENEN *et al.*, 2010 A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC genetics **11:** 42.

ELSIK, C. G., R. L. TELLAM, K. C. WORLEY, R. A. GIBBS, D. M. MUZNY *et al.*, 2009 The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science **324:** 522-528.

ESTELLE, J., F. GIL, J. M. VAZQUEZ, R. LATORRE, G. RAMIREZ *et al.*, 2008 A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. J Anim Sci **86:** 3290-3299.

FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to quantitative genetics*. Longman.

FAN, B., S. K. ONTERU, Z. Q. DU, D. J. GARRICK, K. J. STALDER *et al.*, 2011 Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. PLoS ONE **6:** e14726.

FAO, F. A. A. O. O. T. U. N., 2007 The state of the world's animal genetic resources for food and agriculture, pp. Commission on Genetic Resources for Food and Agriculture Food and Agriculture Organization of the United Nations Rome, 2007

FONTANESI, L., E. D'ALESSANDRO, E. SCOTTI, L. LIOTTA, A. CROVETTI *et al.*, 2010 Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. Anim Genet **41:** 478-492.

FREKING, B. A., S. K. MURPHY, A. A. WYLIE, S. J. RHODES, J. W. KEELE *et al.*, 2002 Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Res **12:** 1496-1506.

FUJII, J., K. OTSU, F. ZORZATO, S. DE LEON, V. K. KHANNA *et al.*, 1991 Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science **253:** 448-451.

GEORGES, M., 2007 Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. Annu Rev Genomics Hum Genet **8:** 131-162.

GEORGES, M., and N. COCKETT, 1996 The ovine callipyge locus: a paradigm illustrating the importance of non-Mendelian genetics in livestock. Reproduction, nutrition, development **36:** 651-657.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO *et al.*, 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

GIUFFRA, E., J. M. KIJAS, V. AMARGER, O. CARLBORG, J. T. JEON *et al.*, 2000 The origin of the domestic pig: independent domestication and subsequent introgression. Genetics **154:** 1785-1791.

GODDARD, M. E., and B. J. HAYES, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet **10:** 381-391.

GRINDFLEK, E., S. LIEN, H. HAMLAND, M. H. HANSEN, M. KENT *et al.*, 2011 Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. BMC Genomics **12:** 362.

GRISART, B., F. FARNIR, L. KARIM, N. CAMBISANO, J. J. KIM *et al.*, 2004 Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc Natl Acad Sci U S A **101:** 2398-2403.

GROBET, L., L. J. MARTIN, D. PONCELET, D. PIROTTIN, B. BROUWERS *et al.*, 1997 A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nature genetics **17:** 71-74.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315-324.

HASENSTEIN, J. R., A. T. HASSEN, J. C. DEKKERS and S. J. LAMONT, 2008 High resolution, advanced intercross mapping of host resistance to Salmonella colonization. Developments in biologicals **132:** 213-218.

HAYES, B., and M. E. GODDARD, 2001 The distribution of the effects of genes affecting quantitative traits in livestock. Genetics, selection, evolution : GSE **33:** 209-229.

HAYES, B. J., S. LIEN, H. NILSEN, H. G. OLSEN, P. BERG *et al.*, 2008 The origin of selection signatures on bovine chromosome 6. animal genetics **39:** 105-111.

HILLIER, L., W. MILLER, E. BIRNEY, W. WARREN, R. C. HARDISON *et al.*, 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature **432:** 695-716.

HU, Z. L., and J. M. REECY, 2007 Animal QTLdb: beyond a repository. A public platform for QTL comparisons and integration with diverse types of structural genomic information. Mamm Genome **18:** 1-4.

IANNUCCELLI, N., J. RIQUET, M. J. MERCAT, H. LEGROS, M. SANCRISTOBAL *et al.*, 2006 Diversité génétique des populations porcines françaises dans les régions chromosomiques soumises à la sélection : projet DIVQTL. Les Actes du BRG **6:** 111-128.

JOHANNES, F., V. COLOT and R. C. JANSEN, 2008 Epigenome dynamics: a quantitative genetics perspective. Nat Rev Genet **9:** 883-890.

JORGENSEN, C. B., S. CIRERA, S. I. ANDERSON, A. L. ARCHIBALD, T. RAUDSEPP *et al.*, 2003 Linkage and comparative mapping of the locus controlling susceptibility towards E. COLI F4ab/ac diarrhoea in pigs. Cytogenetic and genome research **102:** 157-162.

KAO, C. H., 2000 On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. Genetics **156:** 855-865.

KARLSSON, E. K., I. BARANOWSKA, C. M. WADE, N. H. SALMON HILLBERTZ, M. C. ZODY *et al.*, 2007 Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet **39:** 1321-1328.

KERJE, S., J. LIND, K. SCHUTZ, P. JENSEN and L. ANDERSSON, 2003 Melanocortin 1-receptor (MC1R) mutations are associated with plumage colour in chicken. animal genetics **34:** 241-248.

KERJE, S., P. SHARMA, U. GUNNARSSON, H. KIM, S. BAGCHI *et al.*, 2004 The Dominant white, Dun and Smoky color variants in chicken are associated with insertion/deletion polymorphisms in the PMEL17 gene. Genetics **168:** 1507-1518.

KIM, K. S., J. J. KIM, J. C. DEKKERS and M. F. ROTHSCHILD, 2004 Polar overdominant inheritance of a DLK1 polymorphism is associated with growth and fatness in pigs. Mamm Genome **15:** 552-559.

KIM, K. S., N. LARSEN, T. SHORT, G. PLASTOW and M. F. ROTHSCHILD, 2000 A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. Mamm Genome **11:** 131-135.

KNOTT, S. A., J. M. ELSEN and C. S. HALEY, 1996 Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theoretical and Applied Genetics **93:** 71-80.

KNOTT, S. A., L. MARKLUND, C. S. HALEY, K. ANDERSSON, W. DAVIES *et al.*, 1998 Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics **149:** 1069-1080.

LANDER, E., and L. KRUGLYAK, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet **11:** 241-247.

LANDER, E. S., and D. BOTSTEIN, 1987 Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science **236:** 1567-1570.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185-199.

LARSON, G., K. DOBNEY, U. ALBARELLA, M. FANG, E. MATISOO-SMITH *et al.*, 2005 Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science **307:** 1618-1621.

LEE, G. J., A. L. ARCHIBALD, A. S. LAW, S. LLOYD, J. WOOD *et al.*, 2005 Detection of quantitative trait loci for androstenone, skatole and boar taint in a cross between Large White and Meishan pigs. animal genetics **36:** 14-22.

LUNNEY, J. K., 2007 Advances in swine biomedical model genomics. International journal of biological sciences **3:** 179-184.

MAHER, B., 2008 Personal genomes: The case of the missing heritability. Nature **456:** 18-21.

MALEK, M., J. C. DEKKERS, H. K. LEE, T. J. BAAS and M. F. ROTHSCHILD, 2001 A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mamm Genome **12:** 630-636.

MARKLUND, L., P. E. NYSTROM, S. STERN, L. ANDERSSON-EKLUND and L. ANDERSSON, 1999 Confirmed quantitative trait loci for fatness and growth on pig chromosome 4. Heredity **82 (Pt 2):** 134-141.

MARKLUND, S., J. KIJAS, H. RODRIGUEZ-MARTINEZ, L. RONNSTRAND, K. FUNA *et al.*, 1998 Molecular basis for the dominant white phenotype in the domestic pig. Genome research **8:** 826-833.

MEGENS, H. J., R. P. CROOIJMANS, M. SAN CRISTOBAL, X. HUI, N. LI *et al.*, 2008 Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. Genetics, selection, evolution : GSE **40:** 103-128.

MEIJERINK, E., S. NEUENSCHWANDER, R. FRIES, A. DINTER, H. U. BERTSCHINGER *et al.*, 2000 A DNA polymorphism influencing alpha(1,2)fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to Escherichia coli F18 adhesion. Immunogenetics **52:** 129-136.

MILAN, D., J. P. BIDANEL, N. IANNUCCELLI, J. RIQUET, Y. AMIGUES *et al.*, 2002 Detection of quantitative trait loci for carcass composition traits in pigs. Genet Sel Evol **34:** 705-728.

MILAN, D., J. T. JEON, C. LOOFT, V. AMARGER, A. ROBIC *et al.*, 2000 A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science **288:** 1248-1251.

MUNOZ, G., E. ALVES, A. FERNANDEZ, C. OVILO, C. BARRAGAN *et al.*, 2007 QTL detection on porcine chromosome 12 for fatty-acid composition and association

analyses of the fatty acid synthase, gastric inhibitory polypeptide and acetyl-coenzyme A carboxylase alpha genes. Anim Genet **38:** 639-646.

NEZER, C., C. COLLETTE, L. MOREAU, B. BROUWERS, J. J. KIM *et al.*, 2003 Haplotype sharing refines the location of an imprinted quantitative trait locus with major effect on muscle mass to a 250-kb chromosome segment containing the porcine IGF2 gene. Genetics **165:** 277-285.

NEZER, C., L. MOREAU, D. WAGENAAR and M. GEORGES, 2002 Results of a whole genome scan targeting QTL for growth and carcass traits in a Pietrain x Large White intercross. Genet Sel Evol **34:** 371-387.

NOGUERA, J. L., C. RODRIGUEZ, L. VARONA, A. TOMAS, G. MUNOZ *et al.*, 2009 A bi-dimensional genome scan for prolificacy traits in pigs shows the existence of multiple epistatic QTL. BMC Genomics **10:** 636.

ONTERU, S. K., B. FAN, M. T. NIKKILA, D. J. GARRICK, K. J. STALDER *et al.*, 2011 Whole-genome association analyses for lifetime reproductive traits in the pig. Journal of Animal Science **89:** 988-995.

PEREZ-ENCISO, M., A. CLOP, J. L. NOGUERA, C. OVILO, A. COLL *et al.*, 2000 A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross. J Anim Sci **78:** 2525-2531.

PIEPHO, H. P., 2000 Optimal marker density for interval mapping in a backcross population. Heredity **84 (Pt 4):** 437-440.

QANBARI, S., E. C. PIMENTEL, J. TETENS, G. THALLER, P. LICHTNER *et al.*, 2010 A genome-wide scan for signatures of recent selection in Holstein cattle. animal genetics **41:** 377-389.

RAMOS, A. M., R. P. CROOIJMANS, N. A. AFFARA, A. J. AMARAL, A. L. ARCHIBALD *et al.*, 2009 Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One **4:** e6524.

RAMOS, A. M., N. DUIJVESTEIJN, E. F. KNOL, J. W. MERKS, H. BOVENHUIS *et al.*, 2011 The distal end of porcine chromosome 6p is involved in the regulation of skatole levels in boars. BMC genetics **12:** 35.

RATHJE, T. A., G. A. ROHRER and R. K. JOHNSON, 1997 Evidence for quantitative trait loci affecting ovulation rate in pigs. Journal of Animal Science **75:** 1486-1494.

REBAI, A., B. GOFFINET and B. MANGIN, 1995 Comparing power of different methods for QTL detection. Biometrics **51:** 87-99.

REN, D. R., J. REN, Y. Y. XING, Y. M. GUO, Y. B. WU *et al.*, 2009 A genome scan for quantitative trait loci affecting male reproductive traits in a White Duroc x Chinese Erhualian resource population. Journal of Animal Science **87:** 17-23.

RIQUET, J., W. COPPIETERS, N. CAMBISANO, J. J. ARRANZ, P. BERZI *et al.*, 1999 Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc Natl Acad Sci U S A **96:** 9252-9257.

RIQUET, J., G. LAVAL, N. IANNUCCELLI, J. DEMARS, O. DEMEURE *et al.*, 2006 Identification of pig chromosomal regions under selection by a molecular whole genome scan analysis of the Taizumu sino-european synthetic line. 8th World Congress on Genetics Applied to Livestock Production, August 13-18, 2006, Belo Horizonte, MG, Brasil.

ROGEL-GAILLARD, C., N. BOURGEAUX, A. BILLAULT, M. VAIMAN and P. CHARDON, 1999 Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. Cytogenetics and cell genetics **85:** 205-211.

ROHRER, G. A., L. J. ALEXANDER, Z. HU, T. P. SMITH, J. W. KEELE *et al.*, 1996 A comprehensive map of the porcine genome. Genome Res **6:** 371-391.

ROHRER, G. A., and J. W. KEELE, 1998 Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits. J Anim Sci **76:** 2247-2254.

ROTHSCHILD, M., C. JACOBSON, D. VASKE, C. TUGGLE, L. WANG *et al.*, 1996 The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proceedings of the National Academy of Sciences of the United States of America **93:** 201-205.

ROTHSCHILD, M. F., 2004 Porcine genomics delivers new tools and results: this little piggy did more than just go to market. Genetical research **83:** 1-6.

ROTHSCHILD, M. F., L. MESSER, A. DAY, R. WALES, T. SHORT *et al.*, 2000 Investigation of the retinol-binding protein 4 (RBP4) gene as a candidate gene for increased litter size in pigs. Mammalian genome : official journal of the International Mammalian Genome Society **11:** 75-77.

SABETI, P. C., D. E. REICH, J. M. HIGGINS, H. Z. LEVINE, D. J. RICHTER *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832-837.

SANCRISTOBAL, M., C. CHEVALET, C. S. HALEY, R. JOOSTEN, A. P. RATTINK *et al.*, 2006 Genetic diversity within and between European pig breeds using microsatellite markers. Anim Genet **37:** 189-198.

SCHERF, B. D., 2000 World Watch List for Domestic Animal Diversity, 3rd edition, pp. FAO, Rome, Italy.

SLATKIN, M., 2009 Epigenetic inheritance and the missing heritability problem. Genetics **182:** 845-850.

SMIT, M., K. SEGERS, L. G. CARRASCOSA, T. SHAY, F. BARALDI *et al.*, 2003 Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. Genetics **163:** 453-456.

TOBITA-TERAMOTO, T., G. Y. JANG, K. KINO, D. W. SALTER, J. BRUMBAUGH *et al.*, 2000 Autosomal albino chicken mutation (ca/ca) deletes hexanucleotide (-deltaGACTGG817) at a copper-binding site of the tyrosinase gene. Poultry science **79:** 46-50.

TRIBOUT, T., J. C. CARITEZ, J. GOGUE, J. GRUAND, Y. BILLON *et al.*, 2003 Estimation, par utilisation de semence congelée, du progrès génétique réalisé en France

entre 1977 et 1998 dans la race porcine Large White : résultats pour quelques caractères de reproduction femelle. Journées Recherche Porcine **35:** 285-292.

TRIBOUT, T., J. C. CARITEZ, J. GOGUE, J. GRUAND, M. BOUFFAUD *et al.*, 2004 Estimation, par utilisation de semence congelée, du progrès génétique réalisé en France entre 1977 et 1998 dans la race porcine Large White: résultats pour quelques caractères de production et de qualité des tissus gras et maigres. Journées Recherche Porcine **36:** 275-282.

TRIBOUT, T., J. C. CARITEZ, J. GRUAND, M. BOUFFAUD, P. GUILLOUET *et al.*, 2010 Estimation of genetic trends in French Large White pigs from 1977 to 1998 for growth and carcass traits using frozen semen. Journal of Animal Science **88:** 2856-2867.

VAN DER WERF, J. H. J., M. K. and L. S., 2007 Methods and experimental designs for detection of QTL in sheep and goats. Small Ruminant Research **70:** 21-31.

VAN LAERE, A. S., M. NGUYEN, M. BRAUNSCHWEIG, C. NEZER, C. COLLETTE *et al.*, 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature **425:** 832-836.

VARONA, L., C. OVILO, A. CLOP, J. L. NOGUERA, M. PEREZ-ENCISO *et al.*, 2002 QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects. Genet Res **80:** 145-154.

VINCENT, A. L., G. EVANS, T. H. SHORT, O. I. SOUTHWOOD, G. S. PLASTOW *et al.*, 1998 The prolactin receptor gene is associated with increased litter size in pigs. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia**:** 15-18.

WADA, Y., T. AKITA, T. AWATA, T. FURUKAWA, N. SUGAI *et al.*, 2000 Quantitative trait loci (QTL) analysis in a Meishan x Gottingen cross population. Anim Genet **31:** 376-384.

WINTERO, A. K., M. FREDHOLM and P. D. THOMSEN, 1992 Variable (dG-dT)n.(dC-dA)n sequences in the porcine genome. Genomics **12:** 281-288.

YERLE, M., G. ECHARD, A. ROBIC, A. MAIRAL, C. DUBUT-FONTANA *et al.*, 1996 A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. Cytogenetics and cell genetics **73:** 194-202.

YERLE, M., P. PINTON, C. DELCROS, N. ARNAL, D. MILAN *et al.*, 2002 Generation and characterization of a 12,000-rad radiation hybrid panel for fine mapping in pig. Cytogenetic and genome research **97:** 219-228.

YERLE, M., P. PINTON, A. ROBIC, A. ALFONSO, Y. PALVADEAU *et al.*, 1998 Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. Cytogenetics and cell genetics **82:** 182-188.

2

Combining two Meishan F2 crosses improves the detection of QTL on pig chromosomes 2, 4 and 6

Flavie Tortereau[1,3], Hélène Gilbert[2], Henri C.M. Heuven[3], Jean-Pierre Bidanel[2], Martien A.M. Groenen[3], Juliette Riquet[1]

[1] INRA, Laboratoire de Génétique Cellulaire, 31320 Auzeville-Tolosane, France ; [2] INRA, Génétique Animale et Biologie Intégrative, 78352 Jouy-en-Josas Cedex, France ; [3] Wageningen University, Animal Breeding and Genetics Group, 6700 AH Wageningen, The Netherlands

Genetics Selection Evolution (2010), 42:42

Abstract

Background: In pig, a number of experiments have been set up to identify QTL and a multitude of chromosomal regions harbouring genes influencing traits of interest have been identified. However, the mapping resolution remains limited in most cases and the detected QTL are rather inaccurately located. Mapping accuracy can be improved by increasing the number of phenotyped and genotyped individuals and/or the number of informative markers. An alternative approach to overcome the limited power of individual studies is to combine data from two or more independent designs.

Methods: In the present study we report a combined analysis of two independent designs (a French and a Dutch F2 experimental designs), with 2000 F2 individuals. The purpose was to further map QTL for growth and fatness on pig chromosomes 2, 4 and 6. Using QTL-map software, uni- and multiple-QTL detection analyses were applied separately on the two pedigrees and then on the combination of the two pedigrees.

Results: Joint analyses of the combined pedigree provided (1) greater significance of shared QTL, (2) exclusion of false suggestive QTL and (3) greater mapping precision for shared QTL.

Conclusions: Combining two Meishan x European breeds F2 pedigrees improved the mapping of QTL compared to analysing pedigrees separately. Our work was facilitated by the access to raw phenotypic data and DNA of animals from both pedigrees and the combination of the two designs with the addition of new markers allowed us to fine map QTL without phenotyping additional animals.

2.1 INTRODUCTION

Over the past fifteen years, the construction of genetic maps in livestock species has enhanced efforts to dissect the molecular basis of the genetic variation of agriculturally important traits. In pig, a number of experiments have been set up to identify QTL and many chromosomal regions harbouring genes influencing traits of interest have been identified (BIDANEL and ROTHSCHILD 2002) and reported in QTLdb (http://www.genome.iastate.edu/cgi-bin/QTLdb/index) (HU and REECY 2007). However, in most cases mapping resolution remains limited and the QTL detected are rather inaccurately located. Mapping accuracy can be improved by increasing the number of phenotyped and genotyped individuals and/or the number of informative markers. However, collecting this additional information is often time-consuming and/or expensive. An alternative approach to overcome the limited power of individual studies is to combine data from two or more independent designs. Combining several pedigrees together increases the number of animals without additional phenotyping or genotyping costs. Without access to raw data, meta-analysis of published results can be an informative approach to increase precision. Allison and Heo (ALLISON and HEO 1998) have proposed meta-analytical techniques that can be used under difficult conditions. However, these analyses are complicated by the differences among testing methods and experimental designs and finally, the gain in accuracy of QTL mapping is limited. Availability of the raw data to analyse jointly independent data sets is probably a better way to combine different QTL mapping designs. In pig, some studies aiming at combining pedigrees in order to increase the power of QTL detection have already been carried out. Walling et al. (WALLING *et al.* 2000) have combined French, British, Dutch, American, Swedish and German studies to detect QTL on pig chromosome 4 or SSC4 (for *Sus scrofa* chromosome 4) and Perez-Enciso et al. (PEREZ-ENCISO *et al.* 2005) have combined pedigrees from Spanish, French and German designs to refine the location of a QTL for growth and fatness traits on SSCX. However, these analyses are complicated by the lack of common markers and often by slight differences in trait definitions and measurements. In addition, parental populations are usually different, and the QTL segregating in the various designs are not necessarily the same. Under these conditions, combining data sets from different origins may not be optimal to improve estimations of QTL localisations and effects.

Here we report an analysis of QTL located on SSC2, SSC4 and SSC6, that combines French and Dutch F2 pedigrees involving Meishan (MS), Large White (LW) and Landrace (LR) breeds. The analysis was focused on these three pig chromosomes because previous detection analyses (BIDANEL *et al.* 2001; DE KONING *et al.* 2001a; DE KONING *et al.* 1999; DE KONING *et al.* 2001b; HIROOKA *et al.* 2001; LARZUL *et al.* 2008; MILAN *et al.* 2002) have shown that the QTL identified on these chromosomes contribute less to the global variance of the traits than QTL detected for example on SSC7 or SSCX. To optimize this joint analysis and re-construct a unique genetic

map from the 2000 F2 offspring of this combined design, additional microsatellite markers were included in either one or both pedigrees. Using single- and multiple-QTL mapping analyses on each pedigree and on the combined pedigree, we investigated the benefits of combining pedigrees (i.e. doubling the pedigree size) to refine the location of QTL for growth and fatness on SSC2, 4 and 6.

2.2 MATERIALS AND METHODS

Pedigrees and phenotypic data

QTL mapping data from two experimental F2 crosses between European pig breeds x Meishan were used: (1) the French PORQTL pedigree produced at INRA (BIDANEL *et al.* 2001), by mating six Large White sires and six Meishan dams and then six F1 sires and 20 F1 dams to produce 1052 F2 animals; all pigs were born and raised at the INRA GEPA experimental unit (Poitou-Charentes) and (2) a Dutch pedigree, obtained at the University of Wageningen (WU) (JANSS *et al.* 1997a; JANSS *et al.* 1997b) by mating 19 Meishan sires and 126 Large White or Landrace dams and then 39 F1 sires and 265 F1 dams to produce 1212 F2 offspring; this Dutch design was conducted in five different breeding companies. Among the 39 Dutch half-sib families, we selected the 24 largest families, amounting finally to 1919 French and Dutch F2 animals.

Details on the phenotypic data have been reported respectively for the French pedigree in [6] for growth and fatness traits, (BIDANEL *et al.* 2008) for teat number, (MILAN *et al.* 2002) for carcass composition traits and (LARZUL *et al.* 2008) for IntraMuscular Fat (IMF) and for the Dutch pedigree in [10-12] for growth, fatness and meat quality traits (DE KONING *et al.* 2001a; DE KONING *et al.* 1999; DE KONING *et al.* 2001b) and in [9] for teat number(HIROOKA *et al.* 2001). Nine traits related to growth, fatness and teat numbers (Table 2.1) were included in a joint analysis of the pedigrees. Seven of these nine traits i.e. birth weight, weaning weight, carcass weight, teat number, IMF, Back-Fat Thickness (BFT) between the 3rd and 4th ribs of a carcass at 6 cm from the spine and *Longissimus Dorsi* (LD) depth were chosen because they had been recorded in both designs with the same conditions. The two remaining common traits i.e. Life Weight Gain (LWG) and meat percentage were already available for the Dutch pedigree and had to be computed for the French pedigree. Meat percentage was computed in the Dutch pedigree with the Hennessy Grading Probe formula taking BFT and muscle depth into account. For the French pedigree, we applied a similar formula as that used in France at the time of the experiment (DAUMAS and DHORNE 1994) which is also based on back fat thickness and muscle depth (meat percentage = 55.698 – 0.710 x BFT + 0.198 x LD). LWG is an average daily gain estimated throughout the entire animals' life and is calculated as weight / age. Some additional economically important traits that had

been recorded only in one of the two pedigrees but for which a significant QTL had been previously detected, were analysed only in the corresponding pedigree. These traits were related to additional fatness (X2, measured between the 3rd and the 4th lumbar vertebrae at 8 cm from the spine), cut weights (shoulder, midriff, ham, loin, leaf fat, foot, belly, kidney and head) for the French pedigree and meat quality (pH in m. *Longissimus Dorsi* and in m. *Semimembranosus* taken 24 h after slaughter, L*, a* and b* colour values of m. *Longissimus Dorsi*, driploss, cookloss and shear force) for the Dutch pedigree (Table 2.1). These traits are not shared by both designs, but are economically important and thus were re-analysed in this study with additional microsatellite markers.

Genotyping

In order to compare QTL detection results among the French, Dutch and combined pedigrees, a consensus linkage map based on genotyping data from the two Meishan x European breeds F2 populations was generated. The aim was to have a density of one marker every 10 cM within the previously described QTL regions and every 20 cM on the rest of the chromosomes. QTL regions extended from the telomere of the p arm to microsatellite SW240 on SSC2, between microsatellites S0301 and S0214 on SSC4 and along two regions on SSC6 (between SW2535 and SW1057 and between S0059 and SW607). Initially, French and Dutch pedigree were genotyped over these three chromosomes with 22 and 29 microsatellite markers respectively (BIDANEL *et al.* 2001; DE KONING *et al.* 1999). Five microsatellite markers on SSC2, five on SSC4 and six on SSC6 were common to both designs. Additional informative microsatellite markers were included for one or both pedigree(s), to obtain a unique set of common markers. Among the markers genotyped on both pedigrees on SSC6, microsatellite 261M17 was specifically designed from the BAC end-sequence bT261M17SP6 with primers 5'-CTCTTCCATTCCCTGATTGC-3' and 5'-CCCCTTCCTCACCTCTTTCT-3' to fill the gap between S0121 (122 cM) and SW322 (152 cM). On the common map, this new microsatellite is located 12.8 cM from S0121 and 17.4 cM from SW322. Finally, for SSC2 four additional microsatellites were analysed in the French pedigree and two in the Dutch population, for SSC4, two in the French pedigree and one in the Dutch population and for SSC6, four in the French pedigree and three in the Dutch population. New genotyping data were obtained at INRA as previously described (BIDANEL *et al.* 2001). All the genotypes were validated and stored in the Gemma database (IANNUCCELLI *et al.* 1996). Only common markers were kept in the analysis, except for S0217 and SW2466 that were used only on the Dutch animals and SW1089 and SW607 only on the French animals. Microsatellites S0217 and SW1089 and microsatellites SW2466 and SW607 which mapped to the same position respectively on SSC4 and SSC6 were considered as unique markers in the combined analysis. Marker maps were ordered using CriMap package (GREEN *et al.* 1990), considering all the F2 animals of the two designs. The sex-averaged maps are presented in Figure 2.1.

Table 2.1 - Studied traits in the French and Dutch pedigrees

Pedigree	Trait	N	mean	SD
French	**birth weight (kg)**	**1052**	**1.25**	**0.27**
French	**weaning weight (kg)**	**1052**	**5.46**	**1.13**
French	**teat number**	**1046**	**14.82**	**1.56**
French	**carcass weight (kg)**	**529**	**59.11**	**10.48**
French	**BFT (mm)**	**521**	**17.00**	**5.00**
French	**LD depth (mm)**	**521**	**35.00**	**9.00**
French	**meat percentage (%)**	**521**	**50.44**	**3.66**
French	**LWG (kg/day)**	**960**	**0.49**	**0.10**
French	**IMF (%)**	**248**	**1.69**	**0.54**
French	X2 (mm)	521	17.80	5.22
French	shoulder weight (kg)	489	4.71	0.84
French	midriff weight (kg)	489	1.16	0.33
French	ham weight (kg)	489	5.79	0.92
French	loin weight (kg)	489	8.07	1.46
French	leaf fat weight (kg)	489	0.42	0.23
French	foot weight (kg)	489	1.03	0.22
French	belly weight (kg)	489	2.96	0.65
French	kidney weight (kg)	487	0.11	0.02
French	head weight (kg)	481	4.81	0.82
Dutch	**birth weight (kg)**	**867**	**1.22**	**0.22**
Dutch	**weaning weight (kg)**	**864**	**8.23**	**2.00**
Dutch	**teat number**	**869**	**15.42**	**1.20**
Dutch	**carcass weight (kg)**	**548**	**70.24**	**9.59**
Dutch	**BFT (mm)**	**565**	**22.19**	**5.70**
Dutch	**LD depth (mm)**	**563**	**40.82**	**6.72**
Dutch	**meat percentage (%)**	**565**	**48.53**	**4.25**
Dutch	**LWG (kg/day)**	**551**	**0.53**	**0.08**
Dutch	**IMF (%)**	**557**	**1.87**	**0.88**
Dutch	pH_24 (LD)	565	5.67	0.27
Dutch	L*	563	53.85	4.73
Dutch	a*	565	17.13	1.83
Dutch	b*	565	9.52	1.91
Dutch	pH_24 (ham)	565	5.83	0.32
Dutch	driploss (%)	563	2.68	1.52
Dutch	cookloss (%)	564	26.35	3.46
Dutch	shear force (N)	564	39.20	10.75

Traits indicated in bold are common to the two independent pedigrees; N = number of analysed F2 for the corresponding trait; SD = Standard Deviation; BFT = Back-Fat Thickness (referred as X4 in Milan *et al.* 2002); LD = Longissimus Dorsi; LWG = Life Weight Gain; IMF = IntraMuscular Fat content; X2 is another measurement of back-fat thickness; pH_24 = pH measured 24 hours after slaughter; a* = redness of meat; b* = yellowness of meat; L* = lightness of meat

Statistical Analyses

Before QTL detection, phenotypes were corrected for the usual fixed effects using a linear model (GLM procedure, SAS® 9.1, SAS Institute, Inc.). For traits previously analysed, published fixed effects and covariates were used and for the other traits, fixed effects and covariates that were significant at the 5% level in the variance analysis were kept in the final model. Corrected data showed similar variances for the traits common in both designs but recorded independently so that no standardisation was applied.

QTL analysis using these corrected data was performed with the QTLMap software developed at INRA (ELSEN *et al.* 1999; MANGIN *et al.* 1999) based on interval mapping without any hypothesis on the number of QTL alleles present in the Meishan and European breeds.

The test statistic was computed as the ratio of likelihoods under the hypothesis of one (H1) vs. no (H0) QTL linked to the set of markers considered. Under hypothesis H1, a QTL effect for each sire and each dam (only dams with more than 10 offspring were taken into account) was fitted to the data. All sufficiently probable (above 0.10) dam phases were considered, so that the likelihood Λ could not be entirely linearised. For every cM along a linkage group, the likelihood Λ could be written as:

$$\Lambda^x = \prod_{i,j} \sum_{hd_{ij}} p(hd_{ij} | M_i, \hat{hs}_i) \prod_{ijk} f(\tilde{y}_{ijk} | \hat{hs}_i, hd_{ij}, M_i)$$

where: $\prod_{i,j}$ is a product over full-sib families of sire *i* and dam *ij*, $\sum_{hd_{ij}}$ is a summation over dam phases hd_{ij} with a probability greater than 0.10, $\hat{hs}_i = \arg\max_{hs_i} p(hs_i | M_i)$, $p(hs_i | M_i)$ = linkage phase hs_i probability for sire *i* given marker information M_i, $p(hd_{ij} | M_i, \hat{hs}_i)$ = linkage phase for dam *ij* given marker information M_i and sire linkage phase, $f(\tilde{y}_{ijk} | \hat{hs}_i, hd_{ij}, M_i)$ = density function of the adjusted phenotype $\tilde{y}_{ijk}$ of the offspring *ijk* of the *ij*th dam and the *i*th sire, conditional on the chromosome segments transmitted by the sire (q_s) and the dam (q_d). $\tilde{y}_{ijk}$ is supposed to be normally distributed with a mean $\sum_{q_s=1}^{2} \sum_{q_d=1}^{2} p(d_{ijk}^x = (q_s, q_d) | hs_i, hd_{ij}, M_i)(\mu_i^{xq_s} + \mu_{ij}^{xq_d})$ and a variance σ^2_i estimated within each sire family, where $p(d_{ijk}^x = (q_s, q_d) | hs_i, hd_{ij}, M_i)$ is the

transmission probability from parents *i* and *ij* to offspring *ijk*, and μ_i^{xqs} and μ_{ij}^{xqd} can be parameterised as $\mu_{i(j)}^{x1} = \mu_{i(j)} + \alpha_{i(j)}^{x}/2$ and $\mu_{i(j)}^{x2} = \mu_{i(j)} - \alpha_{i(j)}^{x}/2$, α_i^x and α_{ij}^x being the within-half-sib and within-full-sib average QTL substitution effects and $\mu_{i(j)}$ being the family mean for parent *i(j)*. Average substitution effects, which in the present case are equivalent to additive values (*a*), were hence estimated within each sire family as $\mu_i^{x1} - \mu_i^{x2}$ and within each dam family as $\mu_{ij}^{x1} - \mu_{ij}^{x2}$, and averaged over families to estimate the average QTL effect in the population.

To guarantee an accurate estimation of the sire QTL effects, only sire families with more than 30 progeny were retained in the analysis, thus 15 sire families were omitted from the Dutch pedigree. Due to number of progeny per dam, dam effects were estimated for all dams in the French pedigree, whereas none was estimated in the Dutch pedigree.

The maximum LRT along the linkage group indicated the most likely position for a QTL. Significance thresholds were empirically computed using 1000 simulations under the null hypothesis, assuming an infinitesimal polygenic model (i.e. the trait is controlled by an infinite number of additive loci, each with infinitesimal effect and is thus not influenced by a major QTL) and a normal distribution of performance traits (LE ROY *et al.* 1998). In practice, for progeny p, simulated phenotypes y_p were sampled as the sum of a polygenic part u_p and an environmental part e_p with normal distributions of mean = 0 and variances depending on the heritability of the trait, the total phenotypic variance being 1. The polygenic parts were sampled on the F1 sires and dams (u_s and u_d, respectively) and the transmission to the progeny was simulated as $u_p = 0.5\ u_s + 0.5\ u_d$, resulting in $y_p = 0.5\ u_s + 0.5\ u_d + e_p$ for progeny p. Simulations were preferred to permutations because of the family structures (CHURCHILL and DOERGE 2008). With these structures, permutations have to be performed within full-sib families to respect data variability in the different families. In our case, data number within each family was not sufficient to achieve an extensive description of the distribution of the test statistic under the null hypothesis. QTL were considered significant if the chromosome-wide significance threshold exceeded 5% and suggestive if it exceeded 10%. Chromosome-wide significance thresholds were preferred to genome-wide significance thresholds since only three chromosomes were included in the analyses. Estimated significant thresholds (at the 5% chromosome-wide level) varied with traits and pedigree ranging between 45 and 50 for each independent pedigree and 80 and 85 for the combined pedigree. Confidence intervals around a QTL position were empirically determined by the "2-LOD drop-off" method (LANDER and KRUGLYAK 1995).

For each chromosomal region, QTL detection analyses were applied separately on the French and Dutch pedigrees, and then on the combination of both pedigrees thereafter named “combined pedigree”.

Additional analyses were carried out with QTLMap to test the segregation of two linked QTL in a linkage group (GILBERT and LE ROY 2004) and revealed two situations : (1) when a significant QTL had been previously detected (H0 versus H1), the null hypothesis was “one QTL segregating at the maximum likelihood position estimated under H1” and (2) when no QTL had been previously detected (H0 versus H1), the null hypothesis was “no QTL”. In both cases, the alternative hypothesis (H2) was “two linked QTL segregating on the linkage group”. The LRT were computed following a grid-search strategy, using first 5 cM steps along the chromosome to identify significant regions in which then finer steps (1 cM) were applied. Significance thresholds were empirically estimated by a thousand simulations under the null hypothesis as described by Gilbert and Le Roy (GILBERT and LE ROY 2007). When H0 was “no QTL”, thresholds were the same as those computed previously for single QTL tests. When H0 was “one QTL segregating at the maximum likelihood position x_max estimated under H1”, simulations were done assuming that the trait was controlled by a QTL located at x_max and having the effect estimated for the maximum likelihood at position x_max in the single QTL analysis, all F1 being considered as heterozygous for the QTL. Estimated significant thresholds (at the 5% chromosome-wide level) varied with traits and pedigree, ranging between 85 and 90 for each independent pedigree and between 140 and 150 for the combined pedigree.

Finally, QTL detection was also carried out on the adjusted data using a Line-Cross model (LC) with the online version of QTLExpress (SEATON *et al.* 2002). In this report, the method is only briefly described since results are not shown in detail. The LC model assumed that Meishan and European breeds were fixed for alternate QTL alleles. With the LC model, the adjusted phenotypes were fitted to a linear model including additive and dominant components (HALEY *et al.* 1994) and the chromosome-wide significance thresholds were determined by permutations of data as described by Churchill and Doerge (CHURCHILL and DOERGE 1994).

2.3 RESULTS

Linkage maps

For chromosomes SCC2, SSC4 and SSC6, the estimated marker orders of their linkage maps were consistent with those of the published USDA-MARC linkage maps (http://www.marc.usda.gov/genome/swine/swine.html) and their sex-averaged lengths were 149 cM, 116 cM and 166 cM, respectively (Figure 2.1).

QTL detection

Table 2.2 shows the QTL detection results for each chromosome separately and for each independent pedigree and the combined pedigree. These results were obtained by using the half-full sib model with the QTLMap software. Additional analyses were done with the Line Cross model (which assumes that parental breeds are fixed for alternate QTL alleles) with the QTLExpress online Software (SEATON *et al.* 2002). The same QTL were detected with the HFS and LC models (data not shown), except for the QTL underlying birth weight that was only described with the HFS model in the combined pedigree. Moreover, it is worth noting that with LC model, most QTL were significant at a 1% chromosome-wide threshold whereas with HFS model significant QTL were detected at different thresholds (Table 2.2).

SSC2 Single-QTL detection analyses identified a chromosome-wide significant QTL affecting loin weight with the French design and a suggestive QTL affecting back fat thickness. These two QTL are located on the telomeric end of pig chromosome 2 (SCC2p) where the IGF2 gene is located and have already been described (BIDANEL *et al.* 2001),(MILAN *et al.* 2002). With the Dutch design, three chromosome-wide significant QTL affecting meat percentage, BFT and a* colour value were detected around position 25 cM, the latter two QTL already described in (DE KONING *et al.* 1999). An additional suggestive QTL affecting the weight at weaning was also detected in the same region with the Dutch design. Combining the French and Dutch pedigrees, only two QTL reached the 5% chromosome-wide significance threshold: significant LRT values were obtained for a QTL influencing back fat thickness at position 1 cM and for a QTL affecting meat percentage at position 32 cM. Using multiple-QTL analyses no additional QTL was detected on SSC2. Additional analyses related to parent-of-origin effect were computed but are not reported in the present study.

SSC4. With the French design on SSC4, two significant QTL affecting birth weight and belly weight around position 55 cM, and one QTL affecting life weight gain around position 66 cM were detected, all three QTL already described in (BIDANEL *et al.* 2001; MILAN *et al.* 2002). Two additional suggestive QTL also previously described in (LARZUL *et al.* 2008) were identified affecting intra-muscular fat content at position 0 cM and back fat thickness at 61 cM (MILAN *et al.* 2002). Two new suggestive QTL were detected, affecting X2 at 14 cM, and head weight at 42 cM. Using multiple-QTL detection analyses, a pair of QTL localised at positions 30 cM and 74 cM was detected for teat number. With the Dutch design, only one QTL, previously described in (DE KONING *et al.* 1999) and affecting intra-muscular fat content at 0 cM reached the 10% chromosome-wide suggestive. When combining the French and Dutch pedigrees this QTL reached the 5% chromosome-wide significance threshold (Figure 2.2). A QTL affecting birth weight around position 55 cM, and a QTL affecting life weight gain at 83 cM, detected only with the French design, were also confirmed in the combined analysis.

Using multiple-QTL tests, the hypothesis of two QTL affecting this trait was more likely than a single-QTL hypothesis: the test for loci at 59 and 90 cM reached the 5% chromosome-wide significance threshold (Figure 2.3). Additionally, a suggestive QTL influencing teat number was detected in the combined analysis at 46 cM using the single-QTL analysis. The two-QTL model retained in the analysis of the French pedigree for this trait was not significant with the combined analysis.

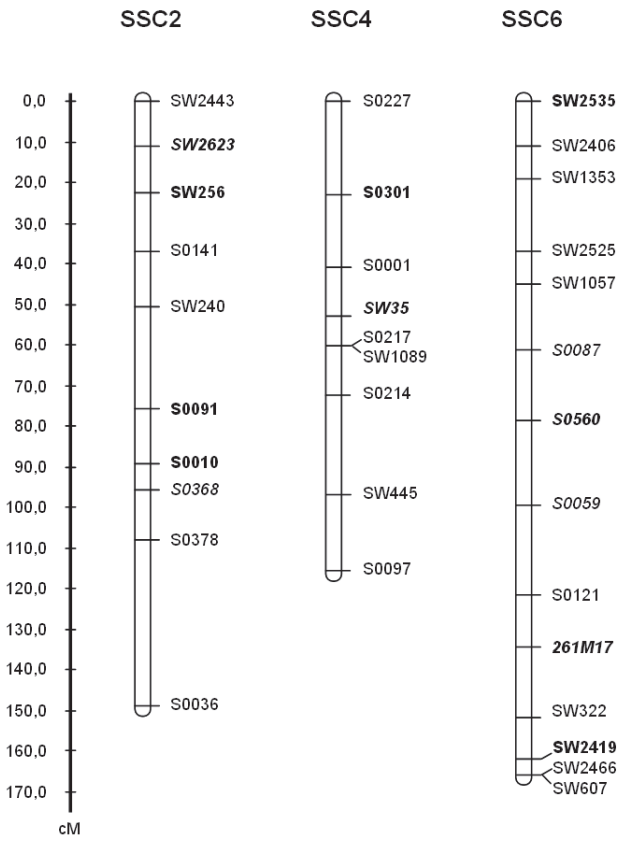


Figure 2.1 - Linkage maps of SSC2, SSC4 and SSC6 for the combined pedigree. Microsatellite markers in bold were added for the analysis of the French animals and markers in italics for the Dutch animals; * : on SSC4, microsatellite S0217 is genotyped only on Dutch animals and SW1089 on French animals; on SSC6, SW2466 is genotyped only on Dutch animals and SW607 on French animals

Table 2.2 - QTL detected in the two independent and the combined pedigrees by the single QTL detection analysis with significance level <10%

Pedigree	Trait	SSC	Position (cM)	LRT	Threshold	Number of segregating sires	QTL effect	References
French	**BFT**	**2**	**0 [0-8]**	**47.2**	**+**	**2**	**+ 0.25**	**[8]**
French	loin	2	0 [0-3]	58.4	**	3	- 0.30	[8]
French	**IMF**	**4**	**0 [0-12]**	**37.0**	**+**	**3**	**- 0.009**	**/**
French	**BFT**	**4**	**61 [55-66]**	**45.5**	**+**	**4**	**+ 0.19**	**[8]**
French	**LWG**	**4**	**66 [56-71]**	**69.6**	**	**2**	**- 0.18**	**[6]**
French	x2	4	14 [6-23]	47.0	+	2	- 0.006	/
French	head weight	4	42 [17-49]	44.6	+	3	- 0.25	/
French	birth weight	4	53 [45-58]	61.3	**	4	- 0.07	[6]
French	belly weight	4	55 [48-61]	56.9	**	2	- 0.04	[8]
French	**teat number**	**6**	**110 [98-130]**	**46.0**	**+**	**3**	**+ 0.03**	**[15]**
French	loin weight	6	99 [92-108]	46.6	+	3	+ 0.22	/
French	midriff weight	6	144 [137-150]	59.1	**	4	- 0.15	/
Dutch	**weaning weight**	**2**	**21 [17-24]**	**46.2**	**+**	**5**	**+ 0.03**	**/**
Dutch	**BFT**	**2**	**28 [18-46]**	**60.4**	**	**9**	**+ 0.22**	**[11]**
Dutch	**meat**	**2**	**30 [19-47]**	**63.8**	**	**8**	**- 0.31**	**/**
Dutch	a*	2	26 [19-36]	51.2	*	10	+ 0.02	[10]
Dutch	**IMF**	**4**	**0 [0-6]**	**47.0**	**+**	**9**	**+ 0.14**	**[11]**
Dutch	**birth weight**	**6**	**136 [130-144]**	**47.4**	**+**	**8**	**- 0.13**	**/**
Dutch	**teat number**	**6**	**154 [146-161]**	**47.5**	**+**	**8**	**- 0.05**	**/**
Dutch	L*	6	103 [91-118]	61.3	**	7	+ 0.20	/
Combined	BFT	2	1 [0-5]	95.6	*	12	+ 0.34	
Combined	meat	2	32 [25-46]	97.8	**	10	- 0.32	
Combined	IMF	4	0 [0-5]	84.0	*	12	+ 0.11	
Combined	teat number	4	46 [37-51]	75.7	+	10	- 0.04	
Combined	birth weight	4	53 [49-57]	82.1	*	9	- 0.10	
Combined	LWG	4	83 [59-90]	98.6	**	11	- 0.30	
Combined	teat number	6	104 [94-113]	84.8	*	11	- 0.01	

Traits indicated in bold are common to the two independent pedigrees. LRT is the maximum Likelihood Ratio Test; the QTL effect is expressed in standard deviation units; the effect is given as Meishan – European alleles; +, *, ** are the 10%, 5% and 1% chromosome-wide significance levels respectively; the references of the previously published data are given; / : when no previous result was available for this trait on the chromosome or when this QTL was not detected in the previous analyses

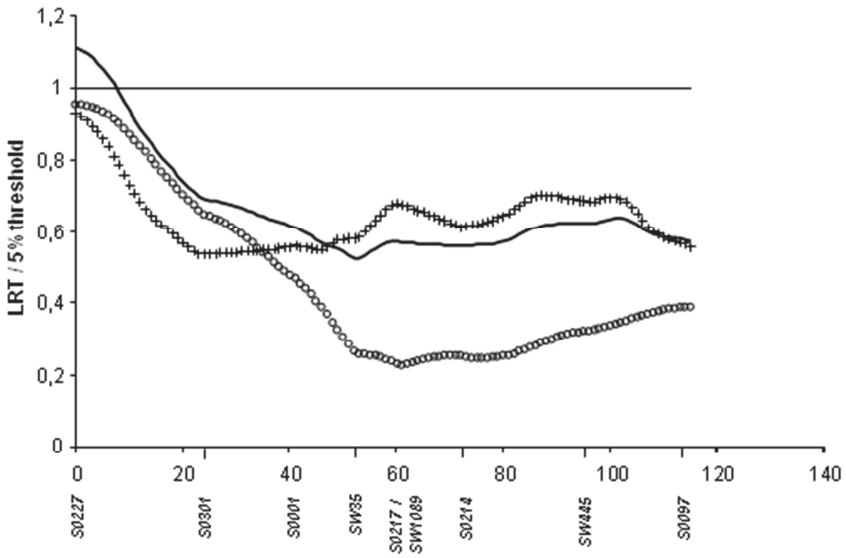


Figure 2.2 - QTL underlying Intra Muscular Fat content on SSC4 for the three studied pedigrees. The solid line gives the result for the combined pedigree, the circled line for the French pedigree and the crossed line for the Dutch pedigree. For each analysis, the LRT is presented in proportion to the 5% threshold on the chromosome.

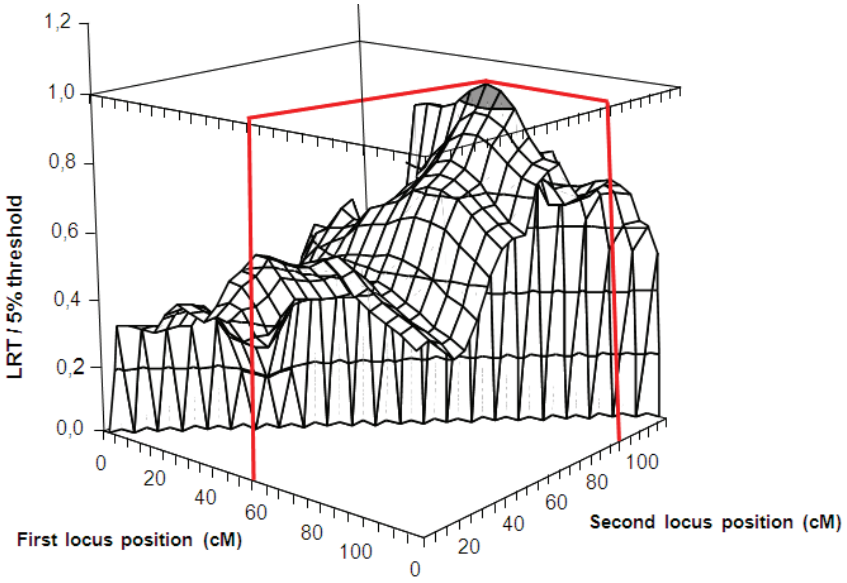


Figure 2.3 - Two-QTL analysis results for Life Weight Gain on SSC4 with the combined pedigree. The z axis gives the value of the LRT divided by the 5% threshold at the chromosome-wide level, the surface shown in gray corresponding to a ratio higher than 1.

SSC6. With the French design on SSC6, a significant QTL affecting midriff weight at 144 cM and two other suggestive QTL one influencing loin weight (99 cM) and one affecting teat number (110 cM) were detected. With the Dutch design, a significant QTL influencing L* parameter at 103 cM and two other suggestive QTL one affecting birth weight (136 cM) and one affecting teat number (154 cM) were identified. For this last trait, a two-QTL model was significant at the 5% chromosome-wide significance threshold for two loci localized at 50 and 155 cM. When combining the French and Dutch pedigrees only one significant QTL affecting teat number at 104 cM was detected at the 5% chromosome-wide threshold.

2.4 DISCUSSION

The aim of this study was to combine two F2 designs produced independently in France and in the Netherlands to detect QTL influencing economically important traits. These two designs were selected on the following criteria: (1) comparable founder breeds (Meishan and Large White and/or Landrace European breeds) and (2) the quasi-equal number of offspring produced in both protocols. Furthermore, although the European breeds were not identical, the Meishan sires used in the Dutch pedigree are related to the French Meishan dams. This supports the assumption that common Meishan QTL alleles segregated in both designs. European and Meishan breeds should have contrasted haplotypes (haplotypes being highly similar in both Meishan pedigrees) and QTL should segregate for the same loci. However, the two populations differ with respect to the number of families (six F1 sire families in the French design versus 39 F1 sires families in the Dutch design) and the reciprocal cross used to produce the F1 animals. To combine these designs the six French families and 24 of the 39 Dutch families, composed of more than 30 offspring, were retained. Our study focused on three pig chromosomes, SSC2, SSC4 and SSC6, for which QTL had already been detected. Despite the lack of overlap between some of the identified QTL from the pedigrees analysed separately, joint analyses of the combined pedigree should provide (1) greater significance of shared QTL, (2) exclusion of false suggestive QTL and (3) greater mapping precision for shared QTL. First, we investigated how the addition of new genotypes influenced the two designs. QTL detections were performed independently for each pedigree and for all the traits to be compared to the results previously published. We were then able to estimate the advantage between a combined analysis and independent analyses.

QTL detected with the French design

In the analysis with the French design, the results were consistent with those previously reported although some differences were observed. In 2001, Bidanel et al. (BIDANEL *et al.* 2001) have reported a highly significant QTL underlying BFT on

SSC4 between two markers located at positions 43 and 83 cM. Addition of microsatellite SW35 at 52.7 cM in the present study resulted in a loss of significance for this trait (10% chromosome-wide threshold). On the opposite, we detected in the same region a suggestive QTL affecting head weight which had not been previously identified in this pedigree (MILAN *et al.* 2002). This shows that the addition of a single highly informative marker in a region with a low marker density can change results to a great extent. Cepica et al. (CEPICA *et al.* 2003) have reported a QTL affecting head weight in the same region on this chromosome. On SSC6, Milan et al. (MILAN *et al.* 2002) have described a suggestive QTL for belly weight between positions 2 and 32 cM In our analysis, this QTL was neither significant, nor suggestive. In this case also, a microsatellite, SW2535, was also added above SW2406, which allowed a better coverage of the telomeric part of the chromosome. Due to the addition of this marker, we can conclude that the previously suggested QTL is probably a false-positive. Finally, with the French pedigree, three new QTL on SSC6: two QTL, one affecting loin weight (99 cM) and one affecting teat number (110 cM), and one significant QTL influencing midriff weight (144 cM). Previous studies had revealed QTL affecting loin weight on SSC6 at 83 cM in crosses involving Pietrain, Large-White and Leicoma animals (MOHRMANN *et al.* 2006), and between 122 and 149 cM with a Duroc x Pietrain design (EDWARDS *et al.* 2008).

QTL detected with the Dutch design

With the Dutch pedigree alone, it was possible that our results differed from those previously reported because of (1) the addition of new markers (as for the French pedigree), (2) the selection of 24 families among the 38 that were used by de Koning et al. (DE KONING *et al.* 1999) and (3) the model used (mixture of full and half-sib families vs. line cross model). Initially, de Koning et al. have detected many QTL regions using a line-cross model and our results are closer to those obtained using a half-sib model (DE KONING *et al.* 2001a; DE KONING *et al.* 1999; DE KONING *et al.* 2001b). Among the previously QTL detected, we did not confirm two suggestive QTL influencing the b* colour value (DE KONING *et al.* 2001a) and LWG on SSC6 (DE KONING *et al.* 2001b). These differences could be due to the addition of new markers showing that these QTL are false-positive ones. However, we cannot exclude that they were lost because of the reduction of the number of families. It is possible that these QTL were segregating in some of the 14 excluded families, which results in loss of power. However, with the addition of new markers in the Dutch pedigree, we detected a new QTL underlying the L* colour value on SSC6 at 103 cM. No QTL influencing this meat quality trait has been reported in this region. In a Landrace x Iberian design, Ovilo et al. (OVILO *et al.* 2002) have described one QTL at 40 cM on SSC6 influencing meat lightness in interaction with another locus on SSC18. We also detected a QTL underlying birth weight at 136 cM while Yue et al. (YUE *et al.* 2003) had reported a QTL for the same trait in a cross between Meishan and Pietrain breeds at 98.2 cM.

Our results illustrate the influence of marker density to exclude false-positive QTL and to detect new ones. In 2008, Hu and Xu (HU and XU 2008) have demonstrated using simulations that when the interval between two adjacent markers decreases, the power of QTL detection increases. Here we confirm this statement with real datasets.

QTL detected combining the French and Dutch designs

The size of a QTL design has a major influence on the power of QTL detection. A QTL can be detected thanks to the recombination events occurring during the gamete production in F1 animals. In an F2 protocol, the number of recombination events is limited. Therefore, to confirm or refine the position of QTL initially detected with an F2 protocol, the number of crossing-over events cannot be increased without additional animals. Producing additional animals is expensive and time-consuming. Thus, a joint analysis of independent pedigrees is an easy alternative to enhance the number of F2 animals in the design. One suggestive QTL affecting intramuscular fat content on SSC4 had been detected in both the French and Dutch pedigrees independently. By combining the two designs, this QTL reached a 5% chromosome-wide significant threshold. As shown on Figure 2.2, the two pedigrees contributed equally to the LRT. The proportion of segregating fathers per pedigree and/or the QTL effect were probably too small to detect this QTL in each pedigree independently but the combined pedigree resulted in more power and the QTL reached the significance threshold. For SSC2, independent analyses of each pedigree indicated that a single QTL influencing BFT and meat percentage was segregating on the telomeric end of the short arm of this chromosome but not at the same position. With the French pedigree, LRT value was maximal in the IGF2 region (0 cM) whereas with the Dutch design it was around 28 cM. Analysis of the combined pedigree confirmed that in fact two linked loci are segregating in this region, i.e. the IGF2 gene and a locus located around 30 cM (Figure 2.4). QTL influencing meat percentage have been previously detected between 40 and 60 cM with crosses involving either Meishan, Pietrain and Wild Boar breeds (LEE *et al.* 2003), or Large White and Meishan animals (QU *et al.* 2002).

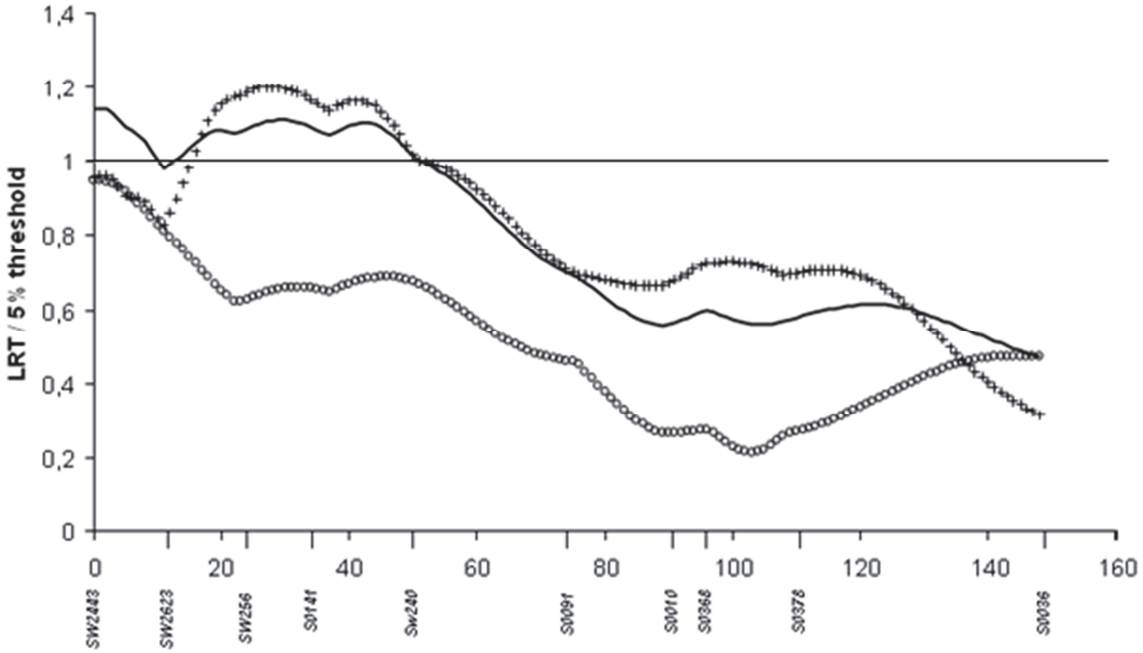


Figure 2.4 - QTL underlying BFT on SSC2 for the three studied pedigrees. The solid line gives the result for the combined pedigree, the circled line for the French pedigree and the crossed line for the Dutch pedigree; for each analysis, the LRT is presented in proportion to the 5% threshold on the chromosome.

Concerning the two-QTL detection analysis, our study provides evidence that combining two pedigrees and adding new markers increases the power of QTL detection. With a single-QTL analysis, we can conclude that there is a single QTL at 80 cM on SSC4 influencing life weight gain whereas in fact this maximum statistic is probably due to the combined actions of two QTL. The combination of two designs indicates that the presence of two different QTL located at 60 and 90 cM and common to both populations is more likely than the segregation of a unique QTL at 80 cM (Figure 2.3).

Limits of the combined analysis

The benefit of combined analyses is essentially obtained when the QTL is segregating in both pedigrees. If QTL were segregating in only one of the two designs, detection power and estimated QTL effect could be reduced. We observed this situation for some QTL initially detected independently in one of both pedigrees. For example, a significant QTL affecting birth weight was detected only

with the French design at 53 cM on SSC4 and was confirmed with the combined pedigree but it was not a major segregating QTL in the Dutch pedigree, the significance of this LRT was reduced in the combined analysis. Thus, this QTL is specific to the French pedigree. At the extreme, a QTL previously detected in one pedigree may be undetectable in the combined analysis. This situation was observed only for suggestive QTL such as that affecting birth weight on SSC6 detected in the Dutch pedigree only. Thus these QTL are also specific to the pedigree in which they were detected. The same situation has been reported by Walling et al. (WALLING *et al.* 2000) who have shown that among seven different pedigrees a QTL affecting birth weight detected on SSC4 segregated in the French pedigree only but after adding the six other pedigrees, although the QTL was still detected, its significance was lower and its effect was divided by two.

Influence of the breeds used

QTL segregating in several independent designs will be largely influenced by the breeds used. On the one hand, by combining two pedigrees (involving only three breeds), Kim et al. in 2005 (KIM *et al.* 2005) have detected 10 new QTL undetected in either of the two pedigrees. In this case, the combination of pedigrees increased the number of interesting regions. On the other hand, Pérez-Enciso et al. (PEREZ-ENCISO *et al.* 2005) in 2005 have demonstrated that by combining pedigrees, the possibility of detecting new QTL is sometimes reduced. This analysis was performed using five independent crosses involving six pig breeds (Iberian, Landrace, Large White, Meishan, Pietrain and Wild Boar). If a QTL allele is fixed in one breed involved in one cross only, the addition of other pedigrees that do not involve these breeds can reduce the effect of this QTL and therefore make it less detectable. This is also supported by the report of Guo et al. (GUO *et al.* 2008) in 2008 who have shown that if a QTL is population-dependant it is highly probable that combining pedigrees will provide no benefit. To avoid this drawback of a joint analysis, Guo et al. have combined two Meishan x Large White pedigrees with F0 animals coming from the same populations. Combining pedigrees seems more advantageous if pedigrees involve similar breeds like in our combined study. In our case, the breeds involved in both pedigrees were genetically very similar, which led to the detection of common QTL (BFT on SSC2 or IMF on SSC4). In some cases, the QTL detected with the combined pedigree had only been described in one of the two designs like birth weight on SSC4. This result can be due to either a population specific QTL, or to differences in the proportion of families contributing to the statistic signal.

Reduction of QTL interval

An important advantage of a combined analysis is to estimate more precisely mapping intervals. This parameter may be largely influenced by the total number of markers used and the number of common markers analysed on the different designs. In our study, new data were generated by genotyping animals of both

pedigrees with additional microsatellite markers to maximise the power and the precision. Previous studies that combined two or more pedigrees only analysed pre-existing datasets. The originality of our study is the large number of common markers, i.e. 29 common markers evenly spaced along the three chromosomes. In 2003, Bennewitz et al. (BENNEWITZ *et al.* 2003) have combined two bovine granddaughter designs and surprisingly, obtained larger confidence intervals than with single designs. Previously, using simulations they had estimated that an increase of population size led to a reduction of confidence intervals (BENNEWITZ *et al.* 2002; VISSCHER *et al.* 1996). Thus, they have proposed that the increase in confidence intervals is probably due to the fact that the families were genotyped with different sets of markers and only a few belonging to both sets. Using a dataset of highly common markers makes it possible to increase the confidence interval of a QTL. In our analysis, we avoided this drawback by analysing a unique set of microsatellite markers on both pedigrees and on the combined one. With a single set of markers, we obtained narrower confidence intervals than previously reported estimated intervals calculated with the same drop-off method (BIDANEL *et al.* 2001; MILAN *et al.* 2002). In the present study, confidence intervals varied from 3 to 30 cM, with an average size of 15 cM. In the initial analyses, the confidence intervals spanned on average 40 cM. Unfortunately, since in the first analyses reported by de Koning et al. [11,12] no confidence intervals were calculated, we could not carry out a comparison.

2.5 CONCLUSION

Combining two Meishan x European breeds F2 pedigrees improved the mapping of QTL compared to separate analyses of the pedigrees. We detected new QTL, confirmed some QTL which were previously described and excluded false positive QTL. Our work was facilitated by the access to raw phenotypic data and DNA of animals from both pedigrees. Analysis of a single set of markers proved more efficient to obtain smaller QTL intervals. The combination of two designs involving very similar breeds with the addition of new markers allowed us to fine-map QTL without phenotyping additional animals.

Acknowledgements

The authors would like to thank Hélène Hayes for her precious help for the English revision of the manuscript and the reviewers for their useful comments.

REFERENCES

ALLISON, D. B., and M. HEO, 1998 Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. Genetics 148: 859-865.

BENNEWITZ, J., N. REINSCH, C. GROHS, H. LEVEZIEL, A. MALAFOSSE *et al.*, 2003 Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. Genet Sel Evol 35: 319-338.

BENNEWITZ, J., N. REINSCH and E. KALM, 2002 Improved confidence intervals in quantitative trait loci mapping by permutation bootstrapping. Genetics 160: 1673-1686.

BIDANEL, J. P., D. MILAN, N. IANNUCCELLI, Y. AMIGUES, M. Y. BOSCHER *et al.*, 2001 Detection of quantitative trait loci for growth and fatness in pigs. Genet Sel Evol 33: 289-309.

BIDANEL, J. P., A. ROSENDO, N. IANNUCCELLI, J. RIQUET, H. GILBERT *et al.*, 2008 Detection of quantitative trait loci for teat number and female reproductive traits in Meishan X Large White F2 pigs. Animal 2: 813-820.

BIDANEL, J. P., and M. F. ROTHSCHILD, 2002 Current status of quantitative trait locus mapping in pigs. Pig News Inf. 23: 39N-53N.

CEPICA, S., A. STRATIL, M. KOPECNY, P. BLAZKOVA, J. SCHRÖFFEL JR *et al.*, 2003 Linkage and QTL mapping for Sus scrofa chromosome 4. J Anim Breed Genet 120 (suppl.1): 28-37.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

CHURCHILL, G. A., and R. W. DOERGE, 2008 Naive application of permutation testing leads to inflated type I error rates. Genetics 178: 609-610.

DAUMAS, G., and T. DHORNE, 1994 Nouvelles équations françaises de prédiction du taux de muscle des carcasses de porc. Journées Recherche Porcine 26: 151-156.

DE KONING, D. J., B. HARLIZIUS, A. P. RATTINK, M. A. GROENEN, E. W. BRASCAMP *et al.*, 2001a Detection and characterization of quantitative trait loci for meat quality traits in pigs. J Anim Sci 79: 2812-2819.

DE KONING, D. J., L. L. JANSS, A. P. RATTINK, P. A. VAN OERS, B. J. DE VRIES *et al.*, 1999 Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics 152: 1679-1690.

DE KONING, D. J., A. P. RATTINK, B. HARLIZIUS, M. A. GROENEN, E. W. BRASCAMP *et al.*, 2001b Detection of characterization of quantitative trait loci for growth and reproduction traits in pigs. Livestock Production Science 72: 185-198.

EDWARDS, D. B., C. W. ERNST, N. E. RANEY, M. E. DOUMIT, M. D. HOGE *et al.*, 2008 Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. J Anim Sci 86: 254-266.

ELSEN, J. M., B. MANGIN, B. GOFFINET, D. BOICHARD and P. LE ROY, 1999 Alternative models for QTL detection in livestock. I. General introduction. Genet Sel Evol 31: 213-224.

GILBERT, H., and P. LE ROY, 2004 Power of three multitrait methods for QTL detection in crossbred populations. Genet Sel Evol 36: 347-361.

GILBERT, H., and P. LE ROY, 2007 Methods for the detection of multiple linked QTL applied to a mixture of full and half sib families. Genet Sel Evol 39: 139-158.

GREEN, P., K. FALLS and S. CROOKS, 1990 Documentation for CRIMAP version 2.4. Washington Uniersity School of Medicine, St. Louis.

GUO, Y. M., G. J. LEE, A. L. ARCHIBALD and C. S. HALEY, 2008 Quantitative trait loci for production traits in pigs: a combined analysis of two Meishan x Large White populations. Anim Genet 39: 486-495.

HALEY, C. S., S. A. KNOTT and J. M. ELSEN, 1994 Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136: 1195-1207.

HIROOKA, H., D. J. DE KONING, B. HARLIZIUS, J. A. VAN ARENDONK, A. P. RATTINK *et al.*, 2001 A whole-genome scan for quantitative trait loci affecting teat number in pigs. J Anim Sci 79: 2320-2326.

HU, Z., and S. XU, 2008 A simple method for calculating the statistical power for detecting a QTL located in a marker interval. Heredity 101: 48-52.

HU, Z. L., and J. M. REECY, 2007 Animal QTLdb: beyond a repository. A public platform for QTL comparisons and integration with diverse types of structural genomic information. Mamm Genome 18: 1-4.

IANNUCCELLI, E., N. WOLOSZYN, J. ARHAINX, J. GELLIN and D. MILAN, 1996 GEMMA: a database to manage and automate microsatellite genotyping. Anim Genet 27 Suppl. 2: 55.

JANSS, L. L., J. A. VAN ARENDONK and E. W. BRASCAMP, 1997a Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population. Genetics 145: 395-408.

JANSS, L. L., J. A. VAN ARENDONK and E. W. BRASCAMP, 1997b Segregation analyses for presence of major genes affecting growth, backfat, and litter size in Dutch Meishan crossbreds. J Anim Sci 75: 2864-2876.

KIM, J. J., M. F. ROTHSCHILD, J. BEEVER, S. RODRIGUEZ-ZAS and J. C. DEKKERS, 2005 Joint analysis of two breed cross populations in pigs to improve detection and characterization of quantitative trait loci. J Anim Sci 83: 1229-1240.

LANDER, E., and L. KRUGLYAK, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11: 241-247.

LARZUL, C., J. P. BIDANEL, N. IANNUCCELLI, J. GRUAND, J. MOUROT *et al.*, 2008 Détection de QTL pour la teneur en lipides intramusculaires et les activités enzymatiques lipogéniques dans un croisement entre porcs Meishan et Large White. Journées Recherche Porcine 40: 127-130.

LE ROY, P., J. M. ELSEN, D. BOICHARD, B. MANGIN, J. P. BIDANEL *et al.*, 1998 An algorithm for QTL detection in mixture of full and half sib families. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production 26: 257-260.

LEE, S. S., Y. CHEN, C. MORAN, S. CEPICA, G. REINER *et al.*, 2003 Linkage and QTL mapping for Sus scrofa chromosome 2. J Anim Breed Genet 120 (suppl 1.): 11-19.

MANGIN, B., B. GOFFINET, P. LE ROY, D. BOICHARD and J. M. ELSEN, 1999 Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations. Genet Sel Evol 31: 225-237.

MILAN, D., J. P. BIDANEL, N. IANNUCCELLI, J. RIQUET, Y. AMIGUES *et al.*, 2002 Detection of quantitative trait loci for carcass composition traits in pigs. Genet Sel Evol 34: 705-728.

MOHRMANN, M., R. ROEHE, P. W. KNAP, H. LOOFT, G. S. PLASTOW *et al.*, 2006 Quantitative trait loci associated with AutoFOM grading characteristics, carcass cuts and chemical body composition during growth of Sus scrofa. Anim Genet 37: 435-443.

OVILO, C., A. CLOP, J. L. NOGUERA, M. A. OLIVER, C. BARRAGAN *et al.*, 2002 Quantitative trait locus mapping for meat quality traits in an Iberian x Landrace F2 pig population. J Anim Sci 80: 2801-2808.

PEREZ-ENCISO, M., A. MERCADE, J. P. BIDANEL, H. GELDERMANN, S. CEPICA *et al.*, 2005 Large-scale, multibreed, multitrait analyses of quantitative trait loci experiments: the case of porcine X chromosome. J Anim Sci 83: 2289-2296.

QU, Y. C., C. Y. DENG, Y. Z. XIONG, R. ZHENG, L. YU *et al.*, 2002 [The construction of the genetic map and QTL locating analysis on chromosome 2 in swine]. Yi Chuan Xue Bao 29: 972-976.

SEATON, G., C. S. HALEY, S. A. KNOTT, M. KEARSEY and P. M. VISSCHER, 2002 QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics 18: 339-340.

VISSCHER, P. M., R. THOMPSON and C. S. HALEY, 1996 Confidence intervals in QTL mapping by bootstrapping. Genetics 143: 1013-1020.

WALLING, G. A., P. M. VISSCHER, L. ANDERSSON, M. F. ROTHSCHILD, L. WANG *et al.*, 2000 Combined analyses of data from quantitative trait loci mapping studies. Chromosome 4 effects on porcine growth and fatness. Genetics 155: 1369-1378.

YUE, G., A. STRATIL, M. KOPECNY, D. SCHRÖFFELOVA, J. SCHRÖFFEL JR *et al.*, 2003 Linkage and QTL mapping for Sus Scrofa chromosome 6. J Anim Breed Genet 120 (suppl. 1): 45-55.

3

Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in Sino-European pig pedigrees

Flavie Tortereau[1,3], Hélène Gilbert[2], Henri C.M. Heuven[3], Jean-Pierre Bidanel[2], Martien A.M. Groenen[3], Juliette Riquet[1]

[1] INRA, Laboratoire de Génétique Cellulaire, 31320 Auzeville-Tolosane, France ; [2] INRA, Génétique Animale et Biologie Intégrative, 78352 Jouy-en-Josas Cedex, France ; [3] Wageningen University, Animal Breeding and Genetics Group, 6700 AH Wageningen, The Netherlands

Genetics Selection Evolution (2010), 43:11

Abstract

Background: In the pig, multiple QTL associated with growth and fatness traits have been mapped to chromosome 2 (SSC2) and among these, at least one shows paternal expression due to the IGF2-intron3-G3072A substitution. Previously published results on the position and imprinting status of this QTL disagree between analyses from French and Dutch F2 crossbred pig populations obtained with the same breeds (Meishan crossed with Large White or Landrace).

Methods: To study the role of paternal and maternal alleles at the IGF2 locus and to test the hypothesis of a second QTL affecting backfat thickness on the short arm of SSC2 (SSC2p), a QTL mapping analysis was carried out on a combined pedigree including both the French and Dutch F2 populations, on the progeny of F1 males that were heterozygous (A/G) and homozygous (G/G) at the IGF2 locus. Simulations were performed to clarify the relations between the two QTL and to understand to what extent they can explain the discrepancies previously reported.

Results: The QTL analyses showed the segregation of at least two QTL on chromosome 2 in both pedigrees, i.e. the IGF2 locus and a second QTL segregating at least in the G/G F1 males and located between positions 30 and 51 cM. Statistical analyses highlighted that the maternally inherited allele at the IGF2 locus had a significant effect but simulation studies showed that this is probably a spurious effect due to the segregation of the second QTL.

Conclusions: Our results show that two QTL on SSC2p affect backfat thickness. Differences in the pedigree structures and in the number of heterozygous females at the IGF2 locus result in different imprinting statuses in the two pedigrees studied. The spurious effect observed when a maternally allele is present at the IGF2 locus, is in fact due to the presence of a second closely located QTL. This work confirms that pig chromosome 2 is a major region associated with fattening traits.

3.1 INTRODUCTION

Many QTL associated with economically important traits like growth, fatness and meat quality have been detected since the 2000s, as reviewed by Bidanel and Rotschild in 2002 (BIDANEL and ROTHSCHILD 2002). However, even for those that have been fine-mapped, successful identification of the causal mutation is rare. In 1999, a paternally expressed QTL affecting backfat thickness (BFT) and muscle mass was identified on the short arm of SSC2 close to the IGF2 gene in crosses between Large White (LW) and European Wild Boar (JEON *et al.* 1999) and between LW and Pietrain (NEZER *et al.* 1999). In 2003, Van Laere et al. [4] reported that the IGF2-intron3-G3072A substitution is the causal mutation. This mutation affects the binding site of a repressor and up-regulates IGF2 expression in skeletal muscles and heart, inducing major maternally imprinted effects on muscle growth, heart size and fat deposition. Therefore, selection for animals carrying allele A at this locus is a major issue in pig production. Analysis of the frequency and effects of this mutation in pig populations of different genetic origins showed that both wild (G) and mutant (A) alleles still segregate in modern populations (LW x Pietrain cross) (CARRODEGUAS *et al.* 2005; JUNGERIUS *et al.* 2004), and that allele A is very rare or even nonexistent in local breeds and Wild Boars (OJEDA *et al.* 2008). The strong favourable effect of allele A was confirmed in both Spanish (ESTELLE *et al.* 2005) and Polish (OCZKOWICZ *et al.* 2009) LW and Landrace (LR) breeds. In 2004, Jungerius et al. [5] demonstrated that the mutation also explains the major imprinted QTL for backfat thickness in a cross between Meishan (MS) and European White pigs (LW and LR). Yet, although significant effects of the IGF2 mutation were revealed both by ultrasonic and carcass BFT measurements, the presence of a second QTL at a position near 40 cM, as previously described in this population by de Koning (DE KONING *et al.* 1999), cannot be excluded (JUNGERIUS *et al.* 2004). In the French LW x MS cross, QTL affecting loin weight and BFT on carcass have also been detected near the IGF2 locus (MILAN *et al.* 2002). However, surprisingly, no imprinting effect could be detected (QUINTANILLA *et al.* 2002), although the breeds involved are similar (European White breeds and MS) in the Dutch and French studies, and the MS animals in both crosses are related. It has been shown that spurious imprinting effects can exist because of maternal effects (HAGER *et al.* 2008) or because of linkage disequilibrium (SANDOR and GEORGES 2008). The aims of the present work were to estimate more precisely the IGF2 substitution effect by combining the two MS x European intercrosses, and to investigate further the genetic determinism of the SSC2p chromosomal region by testing the hypothesis of an additional QTL segregating on SSC2 in these populations. In addition, simulation studies were conducted to investigate how the presence of two QTL could affect the apparent mode of inheritance of IGF2 alleles.

3.2 MATERIALS AND METHODS

Animals and phenotypic data

The French and Dutch F2 MS x European breeds crosses and the recorded phenotypes have been described previously (BIDANEL *et al.* 2001; JANSS *et al.* 1997a; JANSS *et al.* 1997b). Briefly, the French INRA-PORQTL pedigree consisted of 12 F0 (six LW sires and six MS dams), 26 F1 (six sires and 20 dams) and 521 castrated male F2 pigs. All animals were born and raised at the INRA GEPA experimental research unit (Surgères, Charentes). The Dutch pedigree, obtained from the University of Wageningen (WU), was initiated by mating 19 MS sires to 126 LW and LR dams, resulting in an F1 population of 39 sires and 265 F1 dams, which produced a total of 1212 F2 offspring. The Dutch pedigree was bred in five different breeding companies. Among the 39 Dutch half-sib families, only the 24 largest (more than 30 progeny) were retained in the present analysis in order to homogenize the family structure of the two pedigrees.

Among the traits recorded in the two populations, BFT measured between the third and the fourth rib of carcass at 6 cm from the spine (DE KONING *et al.* 1999; MILAN *et al.* 2002) was considered here as the main common trait shared in both designs affected by the QTL under study. This trait was recorded on 565 Dutch pigs (castrated males and females) and on 521 French pigs (castrated males only).

Phenotypic data were first adjusted for fixed effects and covariates with the GLM procedure in SAS® (SAS® 9.1, SAS® Institute, Inc.). The models used to adjust the data included the effects of batch, slaughter day and carcass weight for the INRA pedigree and breeding company, sex, slaughter day and carcass weight for the Dutch pedigree.

Genetic data

Animals from both pedigrees were genotyped for 11 microsatellites evenly spaced on chromosome 2 (SW2443 (0 cM); SWC9 (2 cM); SW2623 (11 cM); SW256 (23 cM); S0141 (37 cM); SW240 (51 cM); S0091 (76 cM); S0010 (90 cM); S0368 (96 cM); S0378 (108 cM) and S0036 (149 cM)), as previously reported (TORTEREAU *et al.* 2010).

Genotyping of the IGF2-intron3-G3072A substitution was performed on some of the F0 and F1 animals of both pedigrees. Previously, F1 boars and their parents (JUNGERIUS *et al.* 2004) from the Dutch pedigree had been genotyped by the pyrosequencing technique (Pyrosequencing AB) described in (VAN LAERE *et al.* 2003). In the French design, all F0 and F1 animals were genotyped by PCR-RFLP using

primers 5'-GGACCGAGCCAGGGACGAGCCT-3' and 5'-GGGAGGTCCCAGAAAAAGTC-3'. The polymerase chain reaction was carried out at 57°C using the GC-RICH PCR system (ROCHE), in presence of 1 M GC-RICH Resolution solution, and 1.5 mM of MgCl2. PCR-RFLP with the restriction enzyme ApeK1 was used to detect the mutation according to the manufacturer's recommendations for time, temperature and buffer conditions. Then, genotypes of all F2 animals at the mutation were inferred for non-recombinant haplotypes inherited from F1 individuals, using information from the pedigree and from the transmission of parental haplotypes for surrounding markers (SW2443 and SWC9). No genotype was assigned for recombinant F2 piglets with a heterozygous A/G parent or if the mother had not been genotyped for the mutation. The parental origin of the allele inherited at the A/G substitution was also inferred when possible according to the phase they inherited from their parents.

QTL analyses

QTL detection was performed on the adjusted data using the QTLMap software (ELSEN *et al.* 1999; MANGIN *et al.* 1999) as explained in (TORTEREAU *et al.* 2010). Parameter estimates were obtained by maximization of the likelihood with a Newton-Raphson algorithm, and a Likelihood Ratio Test (LRT) was computed at each cM along SSC2. The maximum LRT along the linkage group indicated the most likely position for a QTL. For each sire, the substitution effect corresponds to the difference between the Meishan and the European alleles, a positive effect indicating an increased value of the trait due to Meishan alleles. The average QTL substitution effect was computed as the mean of the absolute values of the sire substitution effects. QTL significance thresholds were empirically computed using 1000 simulations under the null hypothesis, assuming an infinitesimal polygenic model for the trait, as described by Gilbert and Le Roy (GILBERT and LE ROY 2007).

QTL detection analyses were carried out first for the French and Dutch pedigrees separately, and then for the combined pedigree. A potential second QTL segregating within these pedigrees was investigated with two different methods. First, the multi-QTL option of QTLMAP was used to detect two linked QTL on SSC2 for BFT. The alternative hypothesis (H1) of two QTL segregating was compared to the null hypothesis of one QTL segregating at the IGF2 locus. The LRT were computed following a grid-search strategy, using 5 cM steps along the chromosome. Significance thresholds were empirically estimated by 1000 simulations under the null hypothesis, as described by Gilbert and Le Roy (GILBERT and LE ROY 2007). In a second analysis, the segregation of a potential additional QTL was investigated: (1) by analysing the data from the progeny of sires homozygous

at the IGF2 locus (G/G) and (2) by performing a QTL detection analysis on the full combined pedigree with a model that included IGF2 as a fixed effect.

Mode of inheritance of the QTL

Analyses of variance (ANOVA) were carried out to infer the inheritance pattern of the SSC2 QTL, using data adjusted for the previously described fixed effects. Tests were applied to compare different effects α_i in the model $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where Y_{ij} is the adjusted performance of individual j of genotype i (see below), μ is the population mean, ε_{ij} is the residual error of individual j of genotype i, and α_i is the tested effect. Three different effects for α_i were built based on the following inheritance patterns:

- Only the paternally inherited allele at the mutation has an effect (model IGF2pat, i = {A,G})
- Only the maternally inherited allele at the mutation has an effect (model IGF2mat, i = {A,G})
- Both the paternally inherited allele and the maternally inherited allele at the mutation have effects (model IGF2patmat, i = {AA,AG,GA,GG}, the paternal allele being written first).

These analyses of variance were applied to all F2 individuals of the combined pedigree, of both pedigrees separately, and to sub-groups of animals defined according to the genotype of the parents at the IGF2 mutation:

- F2 having A/G sires
- F2 having A/G dams
- F2 having A/G sires and G/G dams
- F2 having G/G sires and A/G dams
- F2 having A/G sires and A/G dams.

Detection of spurious effects of the maternally inherited IGF2-allele

Simulation studies were performed with the QTLMap software to evaluate the power of the inheritance pattern analyses and of the additional QTL studies proposed in this paper, in the presence of a major imprinted gene in the

chromosomal region investigated. A QTL segregating at 44 cM affecting the trait was simulated while assuming a paternal effect of the IGF2 locus. Only the phenotypes were simulated; family structures and genotypes were obtained from real data from the two pedigrees. The effect of the IGF2pat model was set to 0.48 phenotypic standard deviations of the trait (as estimated in the data set, see below). The QTL was assumed to be bi-allelic, with the Q allele decreasing backfat level as compared to the q allele. The simulated QTL effect represented the substitution effect of allele q by allele Q. Simulations were then performed with the following parameter values:

- Frequency of the QTL alleles:
    - in the F0 European breeds (French F0 males and Dutch F0 females) for allele Q: 0.25, 0.50, 0.75 or 1.00
    - in the Meishan populations (French F0 females and Dutch F0 males) for allele q: 1.00. Fixation was assumed based on the small size of the original population and based on the fact that MS individuals were also homozygous for IGF2.
- Effect of the simulated QTL: 0.22 or 0.32 or 0.42 phenotypic standard deviations of the trait.

For each simulation, a QTL analysis was performed as described above and the value of the maximum LRT (LRTmax) and its position were recorded. Simulated phenotypes were exported to the SAS® software and analyses of variance were performed as previously described to determine which inheritance pattern was detected depending on the simulated parameters, applying either the IGF2pat or the IGF2mat models. For the analyses of the two pedigrees separately, 2000 replicate simulations were performed for each combination of frequency x effect parameters. For the combined pedigree, 2000 replicates were performed with an effect of 0.32 and a frequency of 0.50 for the QTL in both grandparental populations, as the two pedigrees were reciprocal. The percentage of replicates returning significant results for each pattern of inheritance of IGF2 and detection of the QTL were then computed from the 2000 replicates for each situation.

3.3 RESULTS

Genotyping results for the IGF2 mutation

The IGF2-intron3-G3072A mutation was genotyped for most of the F0 and F1 founders of both pedigrees (Table 3.1). Presence of IGF2 wild type and mutant alleles in the Dutch pedigree was reported previously (JUNGERIUS *et al.* 2004). To

summarize, all MS F0 sires were homozygous (G/G) for the wild allele, and allelic heterogeneity was identified for the LW F0 dams: in the two Dutch LR lines, all the dams were homozygous (G/G), whereas in the three other LW lines allele A was found with frequencies over 80%. Among the 24 sire families selected for our study, 12 F1 sires were homozygous (G/G), and 12 F1 sires were heterozygous (A/G). These 24 F1 sire families involved 65 heterozygous females (A/G) and 71 homozygous females (G/G), while the genotype of 38 F1 dams remained unknown.

All F0 and F1 animals were genotyped in the French pedigree (Table 3.1). All MS F0 dams were homozygous G/G. Among the six LW sires, five were heterozygous (A/G) and one was homozygous for the mutant allele (A/A). Among the six F1 sires, four were homozygous (G/G) and two were heterozygous (A/G). These six F1 sire families involved 15 heterozygous females (A/G) and five homozygous females (G/G).

Table 3.1 – Distribution of genotypes at IGF2-intron3-G3072-A substitution.

	Dutch pedigree					French pedigree				
	A/A	**A/G**	**G/G**	**unknown**	**total**	**A/A**	**A/G**	**G/G**	**unknown**	**total**
F0 males	0	0	19	0	19	1	5	0	0	6
F0 females	30	21	22	27	100	0	0	6	0	6
F1 males	0	12	12	0	24	0	2	4	0	6
F1 females	0	65	71	38	174	0	15	5	0	20
F2	20	180	342	23	565	22	188	226	85	521

Numbers of heterozygous F2 are given regardless parental origin of alleles.

The genotypes of the F2 pigs at the IGF2-intron3-G3072A were inferred from the genotypes of their parents at the mutation and the haplotypes inherited at the surrounding SW2443 and SWC9 microsatellite markers. A complete genotype at the IGF2 mutation could be obtained for 90% of the F2 pigs. Analyses of variance to test the inheritance pattern of the IGF2 mutation were thus performed on 980 F2 animals (543 Dutch F2 and 437 French F2). For ANOVA studies with the IGF2patmat model, the heterozygous (A/G) F2 pigs were split into two groups depending on the parental origin of the two alleles. For the combined pedigree, the total numbers of animals of each genotype at the mutation were 42 A/A, 568 G/G, 155 A/G and 213 G/A, with the first allele identifying the paternal allele.

QTL detection

First, each pedigree was analysed independently. In the French pedigree, the maximum of the test statistic was obtained in the IGF2 region (0 cM) but was only significant at the 10% threshold. Analysis of the Dutch pedigree gave a significant result at the 5% threshold, but the maximum of the test statistics was reached at 28 cM.

The QTL detection analysis was then performed on the combined pedigree (Figure 3.1). The maximum LRT value was obtained in the region surrounding the IGF2 position. However, between 13 and 51 cM, the values of the test statistics were also higher than the 5% threshold.

A multi-QTL analysis was then performed with the combined pedigree but neither significant nor suggestive results were obtained for the hypothesis of two QTL segregating within both pedigrees.

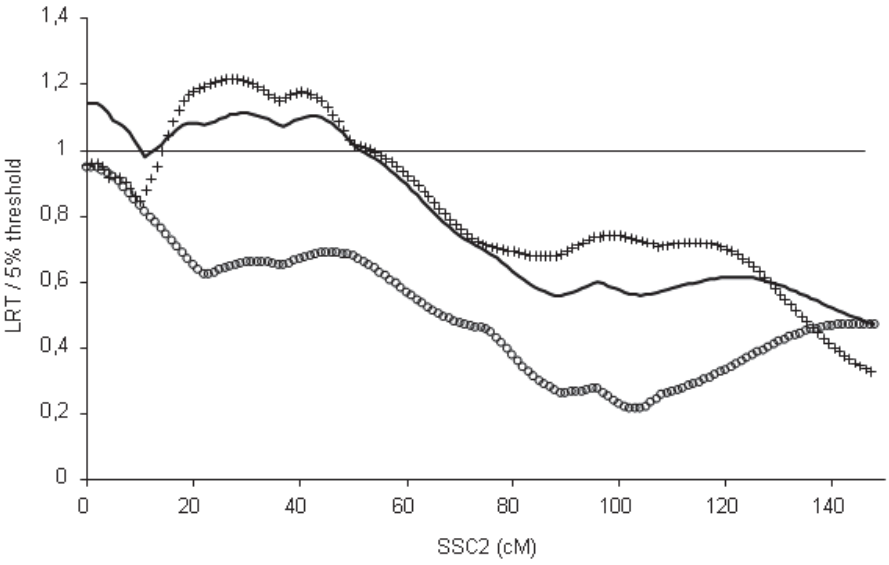


Figure 3.1 - QTL detection on SSC2 in the three studied pedigrees. Solid, circled and crossed lines represent respectively the combined, French and Dutch pedigrees; for each analysis, the LRT is presented as a proportion of the 5% threshold on the chromosome.

The QTL detection analysis performed on the 14 families from sires heterozygous at the mutation revealed a significant QTL close to the IGF2 locus (Figure 3.2). The decrease of the test statistic values downstream from the IGF2 gene was abrupt and no other region reached the 5% threshold. A complementary analysis was

performed on the 16 families originating from homozygous F1 sires (G/G) and detected a significant QTL at 44 cM. The substitution effects estimated at this second QTL position showed that, among the 16 sires analysed, three F1 sires could not be validated as heterozygous for the QTL. The 13 remaining sires were heterozygous with MS alleles associated with high BFT values in nine families and with low BFT values in two families. For the two remaining sires, the breed origin of the favourable allele could not be determined. On average, the QTL effect was estimated to be 0.32 s.d. of the trait. A similar result was obtained with the combined pedigree using phenotypic data corrected for the effect of the IGF2-intron3-G3072A genotype (data not shown). These results clearly indicate that a significant QTL affecting BFT is segregating around 40 cM on SSC2.

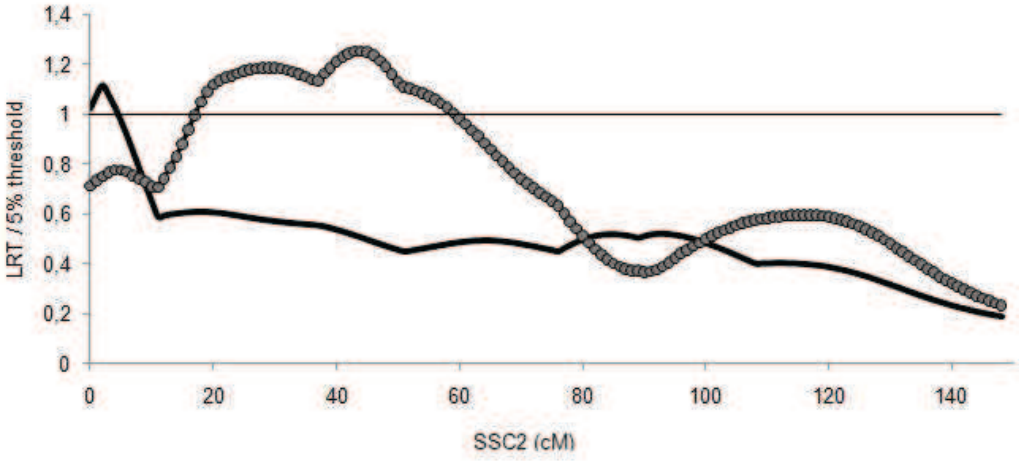


Figure 3.2 - QTL analyses on SSC2 on sub-groups of the combined pedigree. The solid line represents the QTL detection results from the segregating sire families (A/G sires) and the circles-marked line results from the no-segregating sire families (G/G sires) at the IGF2-intron3-G3072A mutation; for each analysis, the LRT is presented as a proportion of the 5% threshold on the chromosome.

Mode of inheritance of the QTL

Analyses of variance were performed on different sub-groups of animals to investigate the effect of the IGF2 mutation. Results obtained using the IGF2pat model confirmed the strong effect of the paternally inherited allele at the mutation (Table 3.2), since significant p-values were observed in all studied sub-groups of F2 pigs. The p-values obtained with the IGF2patmat model were always very similar to those obtained with the IGF2pat model (data not shown). To investigate the potential effect of the maternally inherited IGF2 allele, the IGF2mat model was also

tested. When the analysis was performed on the F2 progeny of heterozygous dams, a significant p-value was obtained with the combined pedigree (p=0.04). When the analysis was carried out on the progeny of heterozygous dams mated to homozygous sires, a significant p-value was also observed (p=0.01). Analysing each pedigree independently, results tended to be significant (p<0.10) for these two progeny sub-groups in the French pedigree and for the F2 produced from A/G dams and G/G sires in the Dutch pedigree (Table 3.2).

Table 3.2 – Statistical analyses of inheritance of the IGF2 mutation effect.

F2 studied	Model	Combined pedigree	French pedigree	Dutch pedigree
All	IGF2 pat	<0.0001 ***	<0.0001 ***	<0.0001 ***
All	IGF2mat	0.30	0.15	0.95
From A/G F1 sires	IGF2pat	<0.0001 ***	0.0007 ***	<0.0001 ***
From A/G F1 sires	IGF2mat	0.94	0.91	0.69
From A/G F1 dams	IGF2pat	0.0001 ***	0.005 **	0.005 **
From A/G F1 dams	IGF2mat	**0.04 ***	**0.08 †**	0.26
From A/G sires and G/G dams	IGF2pat	<0.0001 ***	0.01 *	0.001 ***
From A/G sires and G/G dams	IGF2mat	-	-	-
From G/G sires and A/G dams	IGF2pat	-	-	-
From G/G sires and A/G dams	IGF2mat	**0.01 ***	**0.07 †**	**0.08 †**
From A/G sires and A/G dams	IGF2pat	0.0004 ***	0.02 *	0.008 **
From A/G sires and A/G dams	IGF2mat	0.88	0.93	0.97

ANOVA were done on BFT standardized residuals with either the allele inherited at IGF2 from the sire (pat) or from the dam (mat); *** p-value < 0.001, ** p-value <0.01, * p-value < 0.05, † p-value < 0.10; in bold are indicated the p-values < 0.10 obtained with the maternal allelic effect model.

Detection of spurious effects of the maternally inherited IGF2-allele

The simulated QTL was detected in about 80% of replicates when its effect was at least 0.32 s.d. regardless of the frequency of allele Q in the European grand-parental population (Figure 3.3). When the simulated QTL had a small effect (0.22), the French pedigree tended to be more powerful than the Dutch pedigree to detect the QTL. With the Dutch design, the simulated QTL was detected in fewer than 50% of replicates. For the simulations performed with the combined pedigree, the QTL was detected in 88% of replicates.

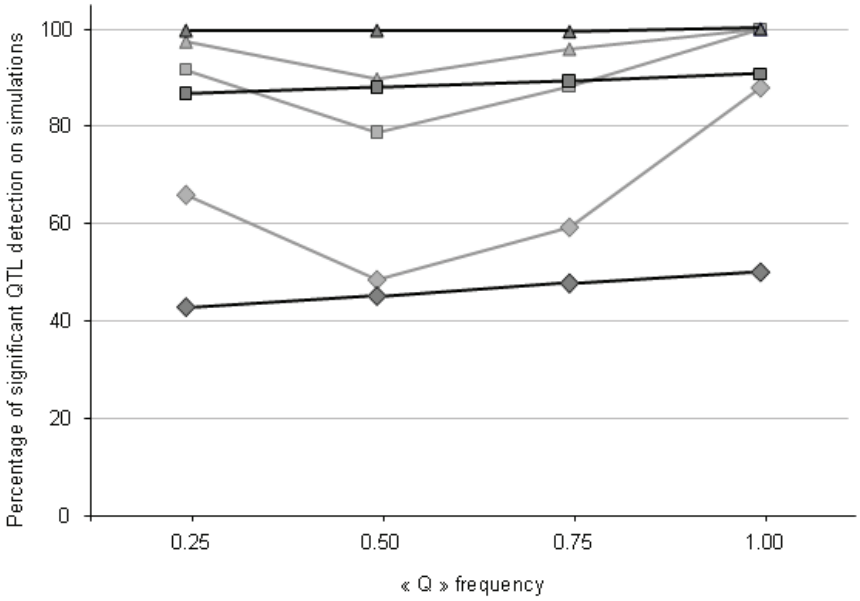


Figure 3.3 - Percentages of replicates for which the simulated QTL was detected, depending on the frequency of the Q allele simulated in the European grand-parental populations. The QTL was considered detected when the 5% threshold was reached; analyses were performed for each pedigree independently (in grey for the French design, in black for the Dutch design) and for three values for the QTL effect (diamonds 0.22 s.d., squares 0.32 s.d. and triangles 0.42 s.d.).

ANOVA was first carried out with the IGF2pat model, using all families. For both pedigrees, the simulated effect of the paternally inherited allele at IGF2 was detected in most replicates (Figure 3.4). The Dutch pedigree gave more significant results than the French pedigree. When the frequency of the simulated Q allele increased in the European populations, the percentage of replicates resulting in a significant effect for the paternally inherited allele decreased. With the combined pedigree, 83% of replicates showed a significant effect of IGF2 on backfat thickness.

Using the model of maternal inheritance on simulated imprinted paternally expressed IGF2 effects, the proportion of results reaching significance for an effect of the maternal allele at IGF2 (IGF2mat) was expected to be low or null. Variance analyses were performed on the sub-group of progeny produced by heterozygous dams regardless of the genotypes of the sires.

With the Dutch pedigree, few replicates led to validation of the maternal expression. In contrast, with the French pedigree, more significant results were

obtained (Figure 3.4). When the simulated QTL had a large effect (0.42 s.d.) and a low frequency of the Q allele was simulated in the European F0 (0.25), up to 75% of the replicates gave a significant result for the IGF2mat model in the French pedigree.

With the combined pedigree, 6.6% of the simulations detected a significant maternally inherited allele effect. When only progeny from the G/G sires among the heterozygous dams' families were considered, the effect of the allele inherited from the mother at the IGF2 mutation was significant in 23% of replicates.

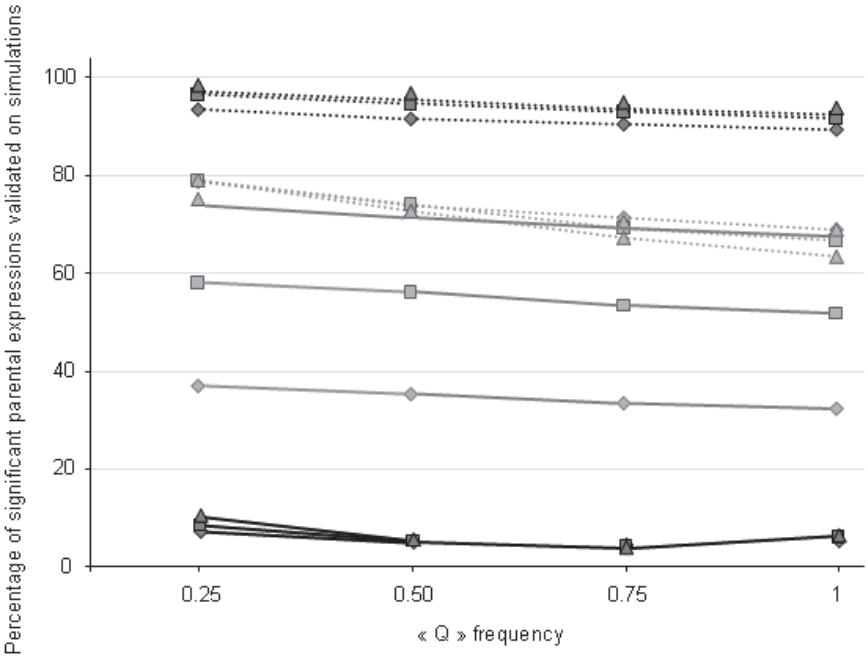


Figure 3.4 - Percentage of replicates that resulted in a statistically significant effect for the allele inherited from the sire or from the dam at the IGF2 mutation. The effects were considered significant when the p-value was lower than 5%; the dotted and solid lines represent, respectively, the results obtained for the allele inherited from the sire and from the dam at the IGF2 mutation; different frequencies of the Q allele in the European grand-parental populations were tested; analyses were performed for each pedigree independently (in grey for the French design, in black for the Dutch design) and for three values of the QTL effect (diamonds 0.22 s.d., squares 0.32 s.d. and triangles 0.42 s.d.).

3.4 DISCUSSION

The aim of this study was to confirm the existence of two QTL associated with BFT on SSC2p and to further dissect the imprinting effect of the IGF2 gene, in order to resolve contradictory results published for the French and Dutch MS x European pedigrees (JUNGERIUS *et al.* 2004; QUINTANILLA *et al.* 2002). The two designs were based on similar founder breeds (MS were crossed with LW and/or LR) and they contributed equally (considering the number of F2) to the so-called "combined pedigree". However, two major differences should be noted. First, the two pedigrees are reciprocal: the MS breed was used as the sire breed in the Dutch pedigree and as the dam breed in the French one. The second discrepancy lies in the pedigree structure: a limited number of large F1 sire families were produced in the French pedigree, whereas the Dutch pedigree consisted in many small F1 sire families.

Our objectives were to (1) show that another QTL on SSC2p is associated with BFT in addition to the IGF2 gene and (2) determine the most likely hypothesis to explain the discrepancies regarding observed imprinting effects in the two pedigrees. Two hypotheses were proposed: a) the imprinting of IGF2 gene is not complete and b) the mendelian effect detected at 0 cM (at the position of the IGF2 locus) in the French pedigree [12] is in fact an artefact due to genetic linkage between IGF2 and a second QTL.

Detection of QTL underlying BFT on the short arm of SSC2

Analysis of the combined pedigree with the single-QTL model gave a most likely position of the QTL at 0 cM, but LRT values over the 5% threshold were obtained for all positions in the first 50 cM of SSC2 (Figure 3.1). When the two pedigrees were analysed independently, two different most likely QTL positions were obtained, in accordance with previously published results on the two pedigrees, i.e. at 0 cM in the French pedigree (BIDANEL *et al.* 2001; MILAN *et al.* 2002) and in an interval between positions 20 and 50 cM in the Dutch pedigree (DE KONING *et al.* 2000; RATTINK *et al.* 2000). After genotyping the Dutch pedigree for the IGF2-intron3-G3072A mutation, it was concluded that the IGF2 QTN, localised at 0 cM, explained most of the observed paternally expressed QTL for BFT on SSC2 (JUNGERIUS *et al.* 2004). However, the presence of an additional QTL around 30 cM could not be excluded. In the present work, nearly every position on the short arm of SSC2 was significant, which is consistent with the large variability of positions found in the Dutch pedigree, and tends to confirm the hypothesis that more than one QTL associated with BFT is segregating in this region. However, this hypothesis had not been validated before for these two pedigrees.

With the combined pedigree, the multiple-QTL model gave no significant result, so that the hypothesis of two co-segregating QTL could not be validated. These results indicate that the F1 sires in general are not heterozygous for both loci and/or that the dataset does not provide enough information to validate the alternative hypothesis, which might be due to the proportion of homozygous sires for the IGF2 mutation in the combined pedigree (0.5).

The separate analyses of the A/G and G/G F1 fathers provided evidence for the segregation of two QTL, the IGF2 mutation and a second one most likely positioned at 44 cM, which is close to the position initially reported in the Dutch pedigree (DE KONING *et al.* 1999). Segregation of a QTL affecting BFT around 40 cM was also reported by Lee et al. (LEE *et al.* 2003) in a Wild Boar x MS pedigree in which all founders were A/A for the IGF2-intron3-G3072A mutation. Therefore, our results confirm the hypothesis of a second QTL segregating in the Dutch LW x MS cross, as suggested by Jungerius et al. (JUNGERIUS *et al.* 2004), and extend this observation to the French cross.

Is the IGF2 gene only paternally expressed?

The IGF2 gene has been studied in detail in several species and, in most cases, paternal expression has been described. Nevertheless, modifications of the imprinted status of genes have been reported in humans (most often associated with diseases) (FALLS *et al.* 1999). For the IGF2 gene, such pathological modifications have already been described (CUI 2007; JIRTLE 2004). However, a study on the loss of imprinting of the IGF2 gene in colorectal cancers has also shown a loss of imprinting in normal mucosa and peripheral blood leukocytes (CUI *et al.* 1998). Moreover, Sakatani et al. (SAKATANI *et al.* 2001) have reported maternal expression of IGF2 in a healthy human population. Li et al. (LI *et al.* 2008) have demonstrated that IGF2 P1 transcripts are bi-allelically expressed in all studied organs from adult healthy pigs. In addition, several studies have reported that the imprinting pattern of a locus can be variable over ontogenetic time (WOLF *et al.* 2008) or under different environmental effects (HAGER *et al.* 2009).

One present objective was to evaluate whether the differences of the imprinting status previously reported for the French (QUINTANILLA *et al.* 2002) and the Dutch (JUNGERIUS *et al.* 2004) pedigrees could be clarified. Significant effects of the allele inherited from the dam at the IGF2 mutation were obtained when analysing the progeny of A/G mothers, which is not in accordance with exclusive paternal expression of the IGF2 gene. Simulations showed that segregation of a second QTL at 40 cM from IGF2 can lead to the false detection of expression of the maternally inherited IGF2-allele in the French pedigree, whereas this was almost never observed in simulations of the Dutch or the combined pedigree. The discrepancies

between simulation results can be explained by differences in the pedigree structures. The Dutch pedigree is based on 104 matings between F0 animals that produced many small half- and full-sib families (24 F1 sires and 174 dams), whereas the French pedigree is based on six matings only between the F0 animals that produced six large half- and full-sib families (six sires and 20 dams). Another difference between the two designs was the allele frequencies at the IGF2-intron3-G3072A mutation: in the French design, 75% of the F1 dams were heterozygous at the mutation whereas only 47% dams where heterozygous in the Dutch design. Since the simulations were performed using the real genotype data at the IGF2-intron3-G3072A mutation, we could not estimate the influence of the percentage of heterogeneous dams. The relative impact of the different family designs and allele frequencies at the mutation could not be differentiated with these simulations. Nevertheless, the ANOVA and simulation analyses suggest that the differences in the effect of the maternal allele at the IGF2 locus can be caused by the segregation of an additional QTL at 44 cM. The genetic linkage between IGF2 and a second QTL with an effect of 0.32 s.d. is high enough to create this artefactual maternal effect, even if the two loci are relatively distant.

3.5 CONCLUSION

Since 2003, several studies have reported the effect of the IGF2-intron3-G3072A mutation on BFT. Besides this QTN, several studies tend to show that additional loci in the surrounding chromosomal region could influence the same trait. By combining two F2 designs, this study demonstrates that a second significant QTL affecting pig fatness is localised around 44 cM and that segregation of this second locus can explain the maternal effect that was observed in the French pedigree at the IGF2 locus. Thus, selection schemes against BFT should not only take the status at the IGF2 mutation into account but also genotypes at other QTL in the region.

QTL for other economically important traits have been described on SSC2p, including QTL affecting daily feed intake (HOUSTON *et al.* 2005) and teat number (HIROOKA *et al.* 2001). For these traits, the influence of the IGF2-intron3-G3072A mutation via pleiotropic effects has been excluded (HOUSTON *et al.* 2005; JUNGERIUS *et al.* 2004). Therefore, the short arm of SSC2 seems to be an important chromosomal region for pig production. Thus, fine-mapping the other QTL on this chromosome will be of major interest. This task will require a cautious design of fine-mapping experiments since the pedigree structures and the variety of loci in the region can lead to false conclusions.

Acknowledgements

The authors would like to thank Hélène Hayes and Jack Dekkers for their precious help for the English revision of the manuscript and their useful comments.

REFERENCES

BIDANEL, J. P., D. MILAN, N. IANNUCCELLI, Y. AMIGUES, M. Y. BOSCHER *et al.*, 2001 Detection of quantitative trait loci for growth and fatness in pigs. Genet Sel Evol 33: 289-309.

BIDANEL, J. P., and M. F. ROTHSCHILD, 2002 Current status of quantitative trait locus mapping in pigs. Pig News Inf. 23: 39N-53N.

CARRODEGUAS, J. A., C. BURGOS, C. MORENO, A. C. SANCHEZ, S. VENTANAS *et al.*, 2005 Incidence in diverse pig populations of an IGF2 mutation with potential influence on meat quality and quantity : An assay based on real time PCR (RT-PCR). Meat Science 71: 577-582.

CUI, H., 2007 Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. Dis Markers 23: 105-112.

CUI, H., I. L. HORON, R. OHLSSON, S. R. HAMILTON and A. P. FEINBERG, 1998 Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability. Nat Med 4: 1276-1280.

DE KONING, D. J., L. L. JANSS, A. P. RATTINK, P. A. VAN OERS, B. J. DE VRIES *et al.*, 1999 Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics 152: 1679-1690.

DE KONING, D. J., A. P. RATTINK, B. HARLIZIUS, J. A. VAN ARENDONK, E. W. BRASCAMP *et al.*, 2000 Genome-wide scan for body composition in pigs reveals important role of imprinting. Proc Natl Acad Sci U S A 97: 7947-7950.

ELSEN, J. M., B. MANGIN, B. GOFFINET, D. BOICHARD and P. LE ROY, 1999 Alternative models for QTL detection in livestock. I. General introduction. Genet Sel Evol 31: 213-224.

ESTELLE, J., A. MERCADE, J. L. NOGUERA, M. PEREZ-ENCISO, C. OVILO *et al.*, 2005 Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian x Landrace cross. J Anim Sci 83: 2723-2728.

FALLS, J. G., D. J. PULFORD, A. A. WYLIE and R. L. JIRTLE, 1999 Genomic imprinting: implications for human disease. Am J Pathol 154: 635-647.

GILBERT, H., and P. LE ROY, 2007 Methods for the detection of multiple linked QTL applied to a mixture of full and half sib families. Genet Sel Evol 39: 139-158.

HAGER, R., J. M. CHEVERUD and J. B. WOLF, 2008 Maternal effects as the cause of parent-of-origin effects that mimic genomic imprinting. Genetics 178: 1755-1762.

HAGER, R., J. M. CHEVERUD and J. B. WOLF, 2009 Change in maternal environment induced by cross-fostering alters genetic and epigenetic effects on complex traits in mice. Proc Biol Sci 276: 2949-2954.

HIROOKA, H., D. J. DE KONING, B. HARLIZIUS, J. A. VAN ARENDONK, A. P. RATTINK *et al.*, 2001 A whole-genome scan for quantitative trait loci affecting teat number in pigs. J Anim Sci 79: 2320-2326.

HOUSTON, R. D., C. S. HALEY, A. L. ARCHIBALD and K. A. RANCE, 2005 A QTL affecting daily feed intake maps to Chromosome 2 in pigs. Mamm Genome 16: 464-470.

JANSS, L. L., J. A. VAN ARENDONK and E. W. BRASCAMP, 1997a Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population. Genetics 145: 395-408.

JANSS, L. L., J. A. VAN ARENDONK and E. W. BRASCAMP, 1997b Segregation analyses for presence of major genes affecting growth, backfat, and litter size in Dutch Meishan crossbreds. J Anim Sci 75: 2864-2876.

JEON, J. T., O. CARLBORG, A. TORNSTEN, E. GIUFFRA, V. AMARGER *et al.*, 1999 A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nat Genet 21: 157-158.

JIRTLE, R. L., 2004 IGF2 loss of imprinting: a potential heritable risk factor for colorectal cancer. Gastroenterology 126: 1190-1193.

JUNGERIUS, B. J., A. S. VAN LAERE, M. F. TE PAS, B. A. VAN OOST, L. ANDERSSON *et al.*, 2004 The IGF2-intron3-G3072A substitution explains a major imprinted QTL effect on backfat thickness in a Meishan x European white pig intercross. Genet Res 84: 95-101.

LEE, S. S., Y. CHEN, C. MORAN, S. CEPICA, G. REINER *et al.*, 2003 Linkage and QTL mapping for Sus scrofa chromosome 2. J Anim Breed Genet 120 (suppl 1.): 11-19.

LI, C., Y. BIN, C. CURCHOE, L. YANG, D. FENG *et al.*, 2008 Genetic imprinting of H19 and IGF2 in domestic pigs (Sus scrofa). Anim Biotechnol 19: 22-27.

MANGIN, B., B. GOFFINET, P. LE ROY, D. BOICHARD and J. M. ELSEN, 1999 Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations. Genet Sel Evol 31: 225-237.

MILAN, D., J. P. BIDANEL, N. IANNUCCELLI, J. RIQUET, Y. AMIGUES *et al.*, 2002 Detection of quantitative trait loci for carcass composition traits in pigs. Genet Sel Evol 34: 705-728.

NEZER, C., L. MOREAU, B. BROUWERS, W. COPPIETERS, J. DETILLEUX *et al.*, 1999 An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. Nat Genet 21: 155-156.

OCZKOWICZ, M., M. TYRA, K. WALINOWICZ, M. ROZYCKI and B. REJDUCH, 2009 Known mutation (A3072G) in intron 3 of the IGF2 gene is associated with growth and carcass composition in Polish pig breeds. J Appl Genet 50: 257-259.

OJEDA, A., L. S. HUANG, J. REN, A. ANGIOLILLO, I. C. CHO *et al.*, 2008 Selection in the making: a worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. Genetics 178: 1639-1652.

QUINTANILLA, R., D. MILAN and J. P. BIDANEL, 2002 A further look at quantitative trait loci affecting growth and fatness in a cross between Meishan and Large White pig populations. Genet Sel Evol 34: 193-210.

RATTINK, A. P., D. J. DE KONING, M. FAIVRE, B. HARLIZIUS, J. A. VAN ARENDONK *et al.*, 2000 Fine mapping and imprinting analysis for fatness trait QTLs in pigs. Mamm Genome 11: 656-661.

SAKATANI, T., M. WEI, M. KATOH, C. OKITA, D. WADA *et al.*, 2001 Epigenetic heterogeneity at imprinted loci in normal populations. Biochem Biophys Res Commun 283: 1124-1130.

SANDOR, C., and M. GEORGES, 2008 On the detection of imprinted quantitative trait loci in line crosses: effect of linkage disequilibrium. Genetics 180: 1167-1175.

TORTEREAU, F., H. GILBERT, H. C. HEUVEN, J. P. BIDANEL, M. A. GROENEN *et al.*, 2010 Combining two Meishan F2 crosses improves the detection of QTL on pig chromosomes 2, 4 and 6. Genet Sel Evol 42: 42.

VAN LAERE, A. S., M. NGUYEN, M. BRAUNSCHWEIG, C. NEZER, C. COLLETTE *et al.*, 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425: 832-836.

WOLF, J. B., J. M. CHEVERUD, C. ROSEMAN and R. HAGER, 2008 Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. PLoS Genet 4: e1000091.

4

Progeny-testing of full-sibs IBD in a SSC2 QTL region highlights epistatic interactions for fatness traits in pigs

Tortereau F.[1], Sanchez M.P.[2], Fève K.[1], Gilbert H.[1,2], Iannuccelli N.[1], Billon Y.[3], Milan D.[1], Bidanel JP.[2], Riquet J.[1]

[1] INRA, UMR444 Laboratoire de Génétique Cellulaire, 31326 Castanet-Tolosan, France ; [2] INRA, UMR1313 Génétique Animale et Biologie Intégrative, 78350 Jouy-en-Josas, France ; [3] INRA, UE967 Génétique Expérimentale en Productions Animales, 17700 Surgères, France

BMC Genetics (2011), 12:92

Abstract

Background: Many QTL have been detected in pigs, but very few of them have been fine-mapped up to the causal mutation. On SSC2, the IGF2-intron3-G3072A mutation has been described as the causative polymorphism for a QTL underlying muscle mass and backfat deposition, but further studies have demonstrated that at least one additional QTL should segregate downstream of this mutation. A marker-assisted backcrossing design was set up in order to confirm the segregation of this second locus, reduce its confidence interval and better understand its mode of segregation.

Results: Five recombinant full-sibs, with genotype G/G at the *IGF2* mutation, were progeny-tested. Only two of them displayed significant QTL for fatness traits although four inherited the same paternal and maternal chromosomes, thus exhibiting the same haplotypic contrast in the QTL region. The hypothesis of an interaction with another region in the genome was proposed to explain these discrepancies and after a genome scan, four different regions were retained as potential interacting regions with the SSC2 QTL. A candidate interacting region on SSC13 was confirmed by the analysis of an F2 pedigree, and in the backcross pedigree one haplotype in this region was found to mask the SSC2 QTL effect.

Conclusions: Assuming the hypothesis of interactions with other chromosomal regions, the QTL could be unambiguously mapped to a 30 cM region delimited by recombination points. The marker-assisted backcrossing design was successfully used to confirm the segregation of a QTL on SSC2 and, because full-sibs that inherited the same alleles from their two parents were analysed, the detection of epistatic interactions could be performed between alleles and not between breeds as usually done with the traditional Line-Cross model. Additional analyses of other recombinant sires should provide more information to further improve the fine-mapping of this locus, and confirm or deny the interaction identified between chromosomes 2 and 13.

4.1 INTRODUCTION

Many QTL underlying economically important traits have been detected in pigs over the last fifteen years (BIDANEL and ROTHSCHILD 2002). These QTL have usually been mapped in large intervals (10 – 30 cM) using experimental crosses between distant populations. Consequently, their use in pig selection schemes has been very limited so far. The only QTL fine-mapped up to the causal mutation in pigs is an A-G substitution in the third intron of the *IGF2* gene (position 3072) (VAN LAERE *et al.* 2003). This causative mutation influences muscle mass and backfat deposition in crosses between Large White (LW) and European Wild Boar (JEON *et al.* 1999), between LW and Pietrain (NEZER *et al.* 1999) and between European breeds and Meishan (MS) (JUNGERIUS *et al.* 2004). For fat-related traits like backfat thickness, many QTL have been mapped on pig chromosome 2 (SSC2) within various experimental populations: their most likely positions ranging from 0 to 50 cM (BIDANEL *et al.* 2001; DE KONING *et al.* 1999; KNOTT *et al.* 1998; RATTINK *et al.* 2000). For crosses involving European breeds such as Piétrain, LW and Landrace, the *IGF2* mutation effect was large enough to refrain from investigating further QTL affecting these traits. However, it has been demonstrated that this *IGF2* mutation does not explain the whole genetic variation of the SSC2 QTL in a LW x MS cross (SANCHEZ *et al.* 2006), and studies of pedigrees where the *IGF2*-intron3-G3072A mutation was not segregating, indicated the presence of a fatness-related QTL between 30 and 60 cM (LEE *et al.* 2003).

In this study, we present QTL analyses in advanced backcross families produced after multiple directed crosses from a F2 LW x MS cross. These were produced from sires carrying recombinant LW x MS chromosomes with recombination points evenly distributed on SSC2p. This strategy, known as marker-assisted backcrossing, is usually performed to refine QTL mapping intervals (MARKLUND *et al.* 1999; RIQUET *et al.* 2011; SANCHEZ *et al.* 2006). The aims of this study were: 1) to confirm and fine-map the fatness-related QTL segregating between 30 and 70 cM on SSC2 and 2) to determine the mode of inheritance of this QTL using the F2 and advanced backcross populations.

4.2 MATERIALS AND METHODS

Animals

Data analysed in this paper came from an advanced backcross population obtained by marker-assisted backcrossing and deriving from the French PorQTL F2 design described by Bidanel et al. (BIDANEL *et al.* 2001). The care and use of animals were performed in compliance with the guidelines of the French Ministry of Agriculture

and Fisheries. Phenotypes of all the animals were recorded at the experimental farm and in a commercial abattoir in standard conditions.

The PorQTL pedigree was created by mating six Large White (LW) sires with six Meishan (MS) dams. Six F1 sires and twenty F1 dams were then mated to produce 1052 F2 animals. All pigs were born and raised at the INRA GEPA experimental unit (Poitou-Charentes). Semen from the F1 sires was frozen. Bidanel et al. (BIDANEL *et al.* 2001) described QTL detection results obtained for major production traits from which QTL genotypes of the F1 LW x MS sires have been assessed.

The advanced backcross population was generated in the INRA GEPA experimental unit (Poitou-Charentes). To produce the first generation of males and females backcrosses (BC), LW dams were inseminated with previously frozen semen of a F1 LW/MS sire. The following five generations were dedicated to the introgression of the F1 sire MS haplotype in *IGF2* region (from 0 to 7 cM) and to the production of various recombinant LW/MS segments in the SSC2 region located between 7 and 70 cM. After five generations of successive backcross, a panel of recombinant sires in the region 7 to 70 cM, homozygous MS/MS for the *IGF2*-intron3-G3072A mutation, was finally mated to LW sows and progeny-tested using on average 100 offspring per sire. In the present study, five half-sib families, obtained from five full-sib sires, were selected for fatness trait QTL analyses (figure 4.1).

Phenotypic Data

F2 animals from the PorQTL pedigree were assessed for average backfat thickness at 120 and 154 days of age. Details about those measurements can be found in (BIDANEL *et al.* 2001). In the present study, we assessed average backfat thickness at 120 days of age because this trait was measured in all F2 animals (n=1071) in contrast to average backfat thickness at 154 days of age that was measured in females only (n=542).

In the backcross population, all piglets were weaned at 28 days of age and were placed in postweaning collective pens until 10 weeks of age. They were then transferred to a fattening unit until 140 days of age. At 120 and 140 days of age, backfat thickness was measured using real-time ultrasound (Aloka SSD-500, Ecotron Aloka), on each side of the spine at 4 cm of the mid-dorsal line and 10 cm of the shoulder (neck), between the 3rd and 4th last ribs (back) and at the level of the last lumbar vertebra (rump). The pigs were finally slaughtered at a mean age of 175±10 days in a commercial slaughterhouse (Saint-Maixent, Deux-Sèvres). Shortly after slaughter, carcass weights and lengths were recorded and carcass fat depths were measured at the shoulder, the last rib and the hip joint. Additional fat (G2) and lean (M2 and M6) depths were recorded between the 3rd and 4th last ribs at 6

cm off the mid-dorsal line using a Fat-o-Meat'er (SFK Technology A/S, Herlev, Denmark) probe. Sixteen traits related to fatness were defined from the above mentioned measurements and analysed: the mean of the two ultrasonic backfat thickness measurements at the level of the neck (**UBFn**), the back (**UBFb**) and the rump (**UBFr**) as well as the mean of the six measurements (**UBFm**) at 120 and 140 days of age; carcass fat depths at the level of the neck (**BFneck**), the back (**BFback**) and the rump (**BFrump**) and the mean of the three measurements (**BFmean**); fat depth **G2** and lean depths **M2** and **M6**; lean meat content (**LMC**) estimated using G2 and M2 measurements (LMC = 62.19 – 0.729 * G2 + 0.144 * M2).

Genetic Data

Different sets of microsatellite markers were used for genotyping depending on the families. All amplifications were performed on ABI 9700 PCR machines (Applied Biosystems, Foster City, CA), and genotyping was carried out on an ABI 3730 automatic sequencer (Applied Biosystems). Genotypes were then determined using the Genemapper software (Applied Biosystems) and results of genotyping were checked, validated, and stored in the GEMMA database (IANNUCCELLI *et al.* 1996).

The F2 PorQTL population had already been genotyped for 123 microsatellites evenly spaced across the genome (BIDANEL *et al.* 2001). F0 and F1 animals were further genotyped for the *IGF2*-intron3-G3072A mutation, and the genotypes of the F2 at this mutation were inferred as explained in (TORTEREAU *et al.* 2011).

Each of the five progeny-tested families (sire and progeny) from the backcross pedigree was genotyped for a set of informative microsatellites covering the 7-70 cM region on SSC2 (marker names and positions are reported in figure 4.2). An additional set of 578 markers covering all the autosomes and SSCX were genotyped on the four full-sib sires, from which eight markers on SSC1, SSC3, SSC6, SSC8, SSC13, SSC14 and SC16 were selected and genotyped on their progeny to test epistatic interactions.

Statistical analysis

Phenotypic data of the backcross families were adjusted for fixed effects (sex and batch) and covariates (weight at measurement or carcass weight for ultrasonic backfat thickness and carcass composition traits, respectively) using the GLM procedure of SAS (SAS 9.1, SAS Institute, Inc.). Average backfat thickness measured at 120 days of age on PorQTL F2 animals was corrected for sex, batch, weight at measurement and their genotype at the IGF2 mutation.

- *QTL analyses*

QTL detection was performed for each backcross family on the adjusted data using the QTLMap software (ELSEN *et al.* 1999; MANGIN *et al.* 1999) as explained by Tortereau et al. (TORTEREAU *et al.* 2010), from 7 to 70 cM (or 7 to 90 cM for the sire 084026). Parameter estimates were obtained by likelihood maximization using a Newton-Raphson algorithm, and a Likelihood Ratio Test (LRT) was computed at each centimorgan along the linkage group. The maximum LRT indicated the most likely position for the QTL. For each sire, the substitution effect corresponds to the difference between its maternal and paternal chromosomes. In our case a positive effect indicates an increase of the trait value attributable to the maternal LW chromosome. Conversely, a negative significant effect indicates an increased trait value due to the paternally inherited MS chromosomal segment of the recombinant chromosome. QTL significance thresholds were empirically computed using 1,000 simulations under the null hypothesis, assuming an infinitesimal polygenic model for the trait, as described by Gilbert and Le Roy (GILBERT and LE ROY 2007).

- *Epistatic interactions analyses*

Epistatic interactions were assessed on the backcross and F2 animals by testing whether the detection of the QTL on SSC2 was conditional on the segregation of another locus in the genome. Four backcross families (from sires 072866, 072868, 084026, 085830) were retained for this analysis, with sires being full-sibs that inherited the same duo of parental chromosomes in the tested region of SSC2 (37-67 cM). In the SSC2 region the sires thus shared pairs of identical by descent chromosomes. The four backcross families were first clustered into two groups of two families (k=1,2) depending on the SSC2 QTL genotype estimate (q/q or Q/q). Epistatic interactions were then tested within each group, between the SSC2 chromosomal region and every other candidate position on the genome marked with microsatellites. First, each group of progeny was divided in sub-groups considering the inherited paternal allele at the tested microsatellite (from two to three sub-groups g with 36 to 93 progeny). Then, within each group k, the QTL detection in the SSC2 region (from 30 to 80 cM) was assessed with the following model:

$$perf_i = \mu_g + p_{SSC2i} * \alpha_{SSC2(kg)} + \varepsilon_i \qquad (1)$$

where, for all progeny *i* belonging to sub-group g of group k, μ_g is the trait mean for the group, $\alpha_{SSC2(kg)}$ is the substitution effect of the putative QTL for the sub-group g, p_{SSC2i} is the probability for individual *i* to inherit its sire maternal allele at the tested position on SSC2 and ε_i is the residual error.

Epistatic interactions in the F2 pedigree were tested with an adaptation of model (1). Because QTL analyses had been previously performed on this design (BIDANEL *et al.* 2001), breed origin (LW or MS) of the paternally transmitted allele in the candidate regions was known for each progeny, and could be used to define two groups of offspring. Within each of these two groups, model (1) was used, with p_{SSC2i} being computed as the probability that the progeny inherited a MS allele from its sire at the tested position *i* on SSC2.

When necessary, epistatic interactions were also tested in the backcross families between the SSC2 chromosomal region and a candidate region traced using transmitted paternal haplotypes in the progeny. Model (1) was jointly applied on the four backcross families with the following modifications:

$perf_i = \mu + p_{SSC2i} * \alpha_{SSC2(kg)} + \varepsilon_i$ (2)

where k was the breed origin of the paternal SSC2 segment inherited by progeny *i* and g the paternal haplotype transmitted in the candidate region (g=1,2,3).

For both backcross and F2 pedigrees, significant $\alpha_{SSC2(kg)}$ effects (p-value < 0.05) were retained as indicating a QTL detection on SSC2 in the g sub-group of the k group considered. The interaction between the QTL on SSC2 and a second locus in the genome was validated when the effect of the QTL was significant (test p-value < 0.05) in at least one of the groups of sires and not consistent across the (sub-) groups.

4.3 RESULTS

Sire chromosomes

The backcross population was designed to dissect the region between 7 and 70 cM on SSC2. The five progeny-tested sires were all homozygous by descent for the MS haplotype in the first six centimorgans of SSC2 (figure 4.1), the haplotype originating from one founder MS chromosome of the F2 design.

Depending on the sire, different contrasts between LW and MS haplotypes were obtained between 7 and 70 cM. It is worthwhile to notice that the Meishan haplotypes were Identical By Descent (IBD) also in this chromosomal region. The LW haplotype origins were more diverse, because different LW animals were introduced during the construction of the pedigree. Among the five progeny-tested sires, three different LW origins were identified. Sires 072866, 072868, 084026 and 085830 inherited the same maternal haplotype, and sire 084024 inherited the alternative maternal chromosome (figure 4.1). All of these sires inherited a

paternal recombinant chromosome with recombination points between 20 and 35 cM.

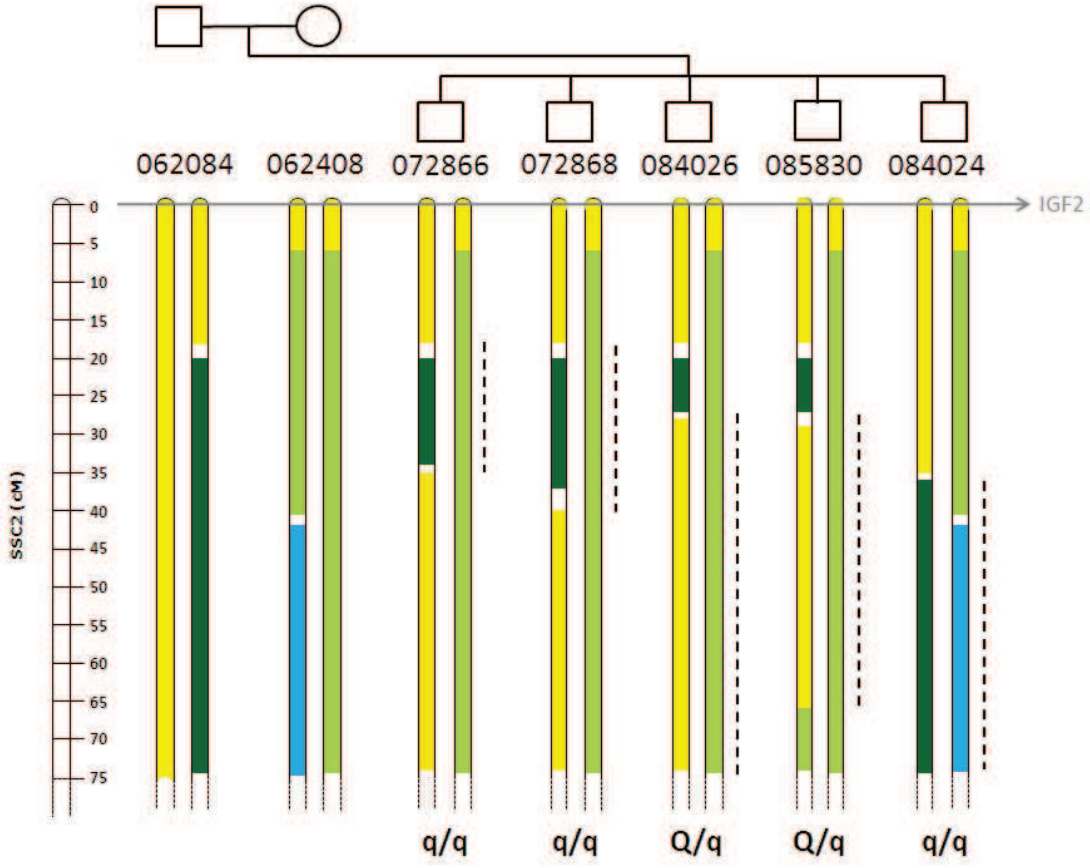


Figure 4.1 – SSC2 haplotypes of the five progeny-tested sires. The pairs of haplotypes (0-75 cM) inherited by the five progeny tested sires from their two parents (062084 and 062408).Yellow refer to the IBD Meishan haplotype (associated with the backfat increasing QTL allele), all other colors refer to different Large White haplotypes (associated with the backfat decreasing QTL allele). In the white areas, the origin could not be defined. The progeny testing results obtained for each sire is indicated by a genotype at the SSC2 QTL (q/q for no segregating families, Q/q for segregating families). The dashed lines represent the position of the studied QTL based on the progeny-testing results of these sires.

QTL detection on SSC2

Table 1 gives the QTL mapping results for the progeny-tested sires. Two groups of sires were distinguishable based on these results. Three sires (072866, 072868 and 084024) did not segregate for any fatness QTL in the region of interest whereas the two others (084026 and 085830) exhibited consistently significant QTL underlying backfat thickness. QTL effects indicated an increase in fatness attributable to the MS haplotype as compared to the LW haplotypes. The results indicate significant QTL segregation in the regions of heterozygosity LW/MS for these sires (figure 4.1). Based on the significant results of sires 084026 and 085830, the QTL underlying fatness traits would be expected to segregate between 27 and 67 cM. Alternatively, the lack of QTL detected within 072866, 072868 and 084024 sires would indicate that the QTL should be localised in the homozygous LW/LW region of these sires. As shown on figure 4.1, there was no overlap between the different intervals deduced from the progeny testing results.

Detection of epistatic interactions

The QTL analyses gave discordant results regarding the localization interval. As shown in table 4.1 and figure 4.2, the most likely position of the SSC2 QTL detected in sires 084026 and 085830 is around 60-70 cM. In this region, sires 072866, 072868, 084026 and 085830 exhibited exactly the same pair of identical by descent chromosomes, and thus carried the same QTL alleles. However, only two of them were validated for the segregation of a fatness QTL, with an effect size of about 0.5 mm on ultrasonic backfat. Family sizes were large enough (N>75 for carcass traits, N>85 for in vivo traits) for all sires to detect this QTL effect with certainty when it segregated. The hypothesis of epistasis was thus considered to explain these discrepancies. For a region to be considered to be candidate as the interacting region, at least two haplotypes should segregate in the four sire families, one enabling the detection of the QTL within sires 084026 and 085830 families, and another masking the segregation of the SSC2 QTL in sires 072866 and 072868 families.

Table 4.1 – Significant QTL results for fatness traits on SSC2

	Sire 072866		Sire 072868		Sire 084024		Sire 084026					Sire 085830				
								QTL results					QTL results			
	Nb off.	QTL results[a]	Nb off.	QTL results	Nb off.	QTL results	Nb off.	Max LRT	Pos (cM)	QTL effect	Significance.	Nb off.	Max LRT	Pos (cM)	QTL effect	Significance.
UBFr 120 (mm)	106	-	119	-	98	-	101	8.4	70	-0.44	*	85	9.2	62	-0.47	*
UBFb 120 (mm)	106	-	119	-	98	-	101	9.8	61	-0.45	*	85	9.5	30	-0.38	*
UBFn 120 (mm)	106	-	119	-	98	-	101	8.3	71	-0.58	*	85	9.8	62	-0.65	*
UBFm 120 (mm)	106	-	119	-	98	-	101	11.2	70	-0.49	**	85	9.8	62	-0.46	*
UBFr 140 (mm)	105	-	119	-	98	-	98	-	-	-	-	84	6.9	62	-0.47	*
UBFb 140	105	-	119	-	98	-	98	-	-	-	-	84	7.7	62	-0.57	*
UBFn 140 (mm)	105	-	119	-	98	-	98	10.6	66	-0.67	**	84	9.3	62	-0.75	*
UBFm 140 (mm)	105	-	119	-	98	-	98	6.7	67	-0.44	+	84	9.6	62	-0.59	*
G2 (mm)	74	-	100	-	97	-	94	17.4	67	-1.11	**	72	7.4	31	-0.88	+
M2 (mm)	74	-	100	-	97	-	94	7.1	58	1.39	+	72	-	-	-	-
M6 (mm)	74	-	100	-	97	-	94	7.6	77	1.10	*	72	-	-	-	-
LMC (%)	74	-	100	-	97	-	94	22.7	68	0.98	**	72	-	-	-	-
BFrump (mm)	84	-	102	-	97	-	94	-	-	-	-	72	-	-	-	-
BFback (mm)	84	-	102	-	97	-	94	-	-	-	-	72	-	-	-	-
BFneck (mm)	84	-	102	-	97	-	93	-	-	-	-	72	12	32	-1.67	**
BFmean (mm)	84	-	102	-	97	-	93	-	-	-	-	72	-	-	-	-

QTL effects are given as (maternal) – (paternal) allele effects. +, *, ** indicate the 10%, 5% and 1% significance levels respectively. UBFr = live ultrasonic backfat thickness measured at the level of the rump; UBFb = live ultrasonic backfat thickness measured at the level of the back ; UBFn = live ultrasonic backfat thickness measured at the level of the neck ; UBFm = live ultrasonic backfat thickness mean at 120 or 140 days of age. LMC = carcass lean meat content ; BFrump = carcass backfat thickness measured at the level of the rump ; BFback = carcass backfat thickness measured at the level of the back, ; BFneck = carcass backfat thickness measured at the level of the neck ; BFmean = average carcass backfat thickness.

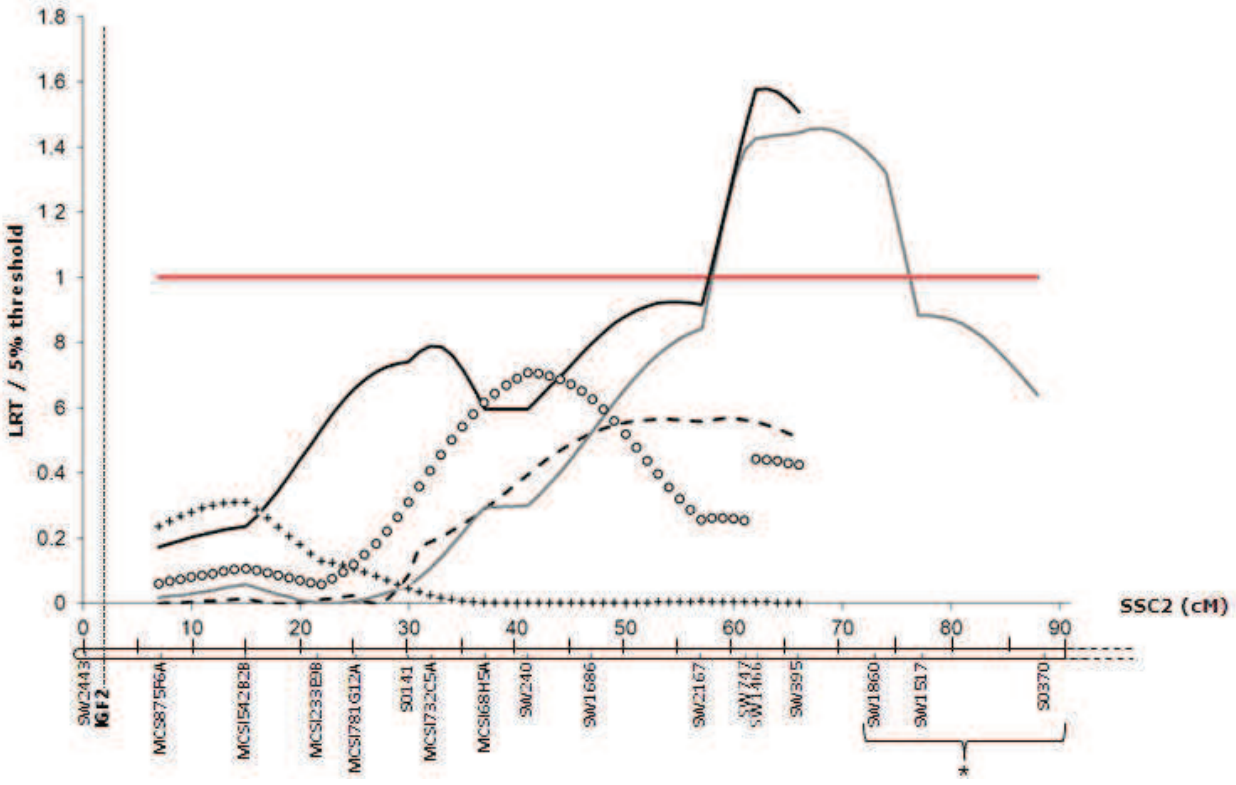


Figure 4.2 – LRT profile on SSC2 for live backfat thickness measured at the level of the neck at 140 days of age in the five sire families. The crossed, circled, grey, black and dashed lines give the results for sires 072866, 072868, 084026, 085830, and 084024, respectively. The LRT is presented as a ratio to the 5% threshold on the linkage group. *SW1860, SW1517 and S0370 were not informative for the 085830, 084024, 084026, 072868 and 072866 sire families.

To identify these candidate regions, a genome scan was performed on the four sires using a panel of 578 microsatellites distributed throughout the genome (figure 4.3). The criteria chosen to select candidate regions were: (1) When two fathers did not belong to the same family group and had the same genotypes then the marker was excluded.(2) A region was selected as candidate when several adjacent markers were retained. The genotypes in the two groups of sires (072866/072868 *versus* 084026/085830) were thus compared. Finally, seven different candidate regions, on SSC1 (140-145 cM), SSC3 (55-75 cM), SSC6 (90-120 cM), SSC8 (50-60 cM), SSC13 (55-75 cM), SSC14 (20-35 cM) and SSC16 (45-90 cM) were selected (figure 4.3).

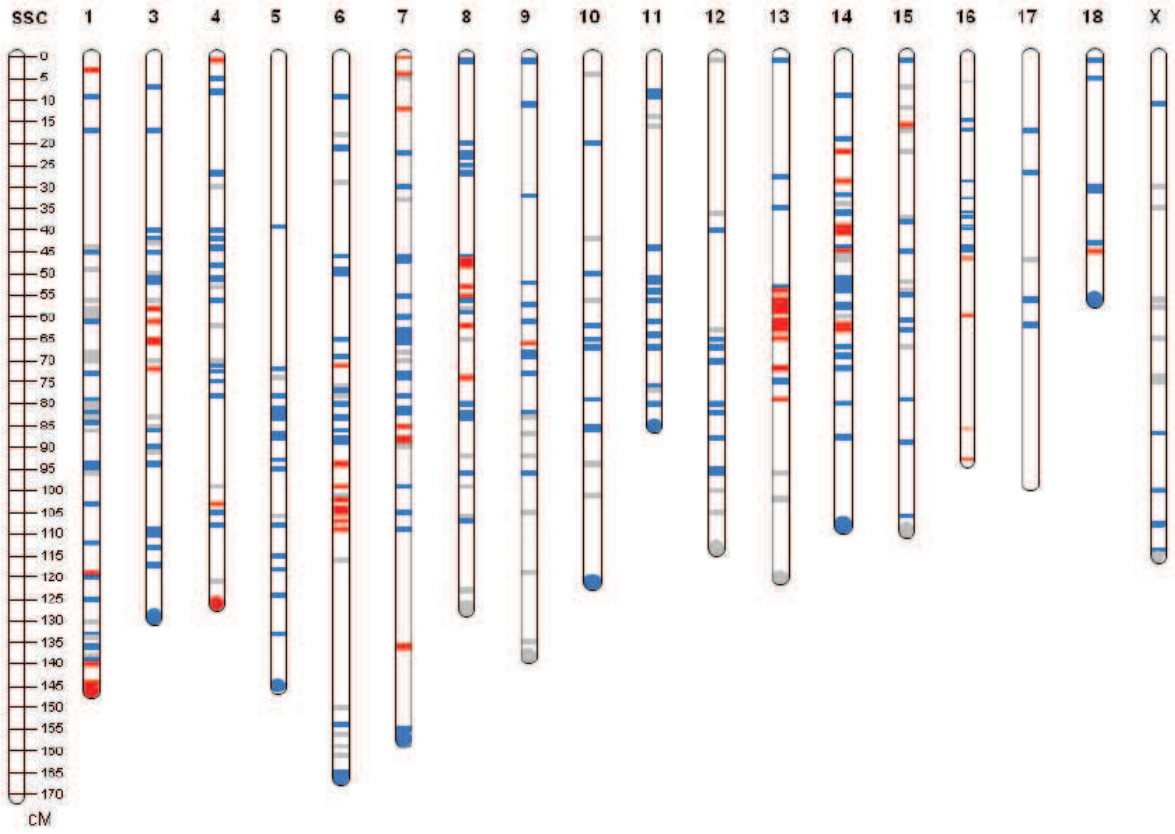


Figure 4.3 – Genome scan for regions candidate for epistatic interactions with SSC2 in the 072866, 072866, 084026 and 085830 sires. Each line represents a microsatellite. When two sires did not belong to the same familial group and had the same genotypes, the marker was excluded. In red are the markers retained, in blue are the excluded markers and in grey are the non-informative markers: all the sires were homozygous for the same allele.

One microsatellite (two for SSC13) was retained to mark each region to be genotyped on the four sire progeny. An interval mapping strategy was applied for the two groups of sires independently (072866/072868 or 084026/085830), using the model (1) where the interaction between the microsatellite alleles and the effect of every putative position of the SSC2 QTL within the 40-80 cM region was tested. This analysis was applied only to ultrasonic backfat thickness measured at the level of the neck at 140 days of age, which showed the highest significance of the SSC2 QTL in the 084026 and 085830 families (figure 4.2). Only the results obtained at the SW395 position (66 cM), microsatellite genotyped in all families closest to the most likely position of the SSC2 QTL, are presented (table 4.2).

Model (1) was applied to the two groups of sires (072866/072868 *versus* 084026/085830). The test p-values were always significant when the sires 084026 and 085830 were analysed (from 0.0034 to 0.0003), and were never significant (the smallest p-value being 0.10 when interaction was tested with SW1550 on SSC13) when sires 072866 and 072868 were analysed.

On SSC1, three different alleles were observed, allele 1 being present in the two groups of sires but the SSC2 QTL effect being detected only in sire 084026 and 085830 families. A similar pattern was observed on SSC14, SSC13 (SW207), and SSC16. As a result, regions on SSC1, around SW207 on SSC13, SSC14 and SSC16 were not retained as candidate regions because a common allele gave different results in the two groups of sires.

For the SSC3 candidate region, the SSC2 QTL effect (at 66 cM on SSC2) was detected in the two groups of sires, but only when the progeny inherited allele 1 at the tested microsatellite. For SSC8, the only allele enabling the detection of the SSC2 QTL segregated in the progeny of sires 084026 and 085830. These two regions on SSC3 and SSC8, were conserved as candidate epistatic regions, as significant interactions were obtained with only one allele of the tested microsatellite, segregating in one (SW205 on SSC8) or the two groups of sires (SW1436 on SSC3).

On SSC6 and for SW1550 on SSC13, all the alleles except for one were associated with significant detection of the SSC2 QTL. On each of these two microsatellites, the allele for which the SSC2 QTL could not be detected segregated only in the sires of the 072866 and 072868 families. Thus, SSC6 around SW1550 on SSC13 were considered as candidate regions for interaction with the SSC2 QTL.

Table 4.2 – Interaction analyses results

		Sires 072866 and 072868			Sires 084026 and 085830		
Tested marker	**Allele**	**Nb off.**	**Effect**	**P-val**	**Nb off.**	**Effect**	**P-val**
SSC1 (SW2512)	1	69	-0.51	0.308	58	-1.45	0.009
	2	81	-0.27	0.544	-	-	-
	3	-	-	-	67	-1.27	0.010
SSC3 (SW1436)	1	85	-1.49	**0.001**	87	-1.86	<0.0001
	2	93	0.37	0.372	-	-	-
	3	-	-	-	70	0.60	0.242
SSC6 (SW1055)	1	62	-1.07	**0.050**	67	-1.49	0.007
	2	55	-0.19	0.719	-	-	-
	3	-	-	-	91	-1.27	0.005
SSC8 (SW205)	1	64	-0.77	0.186	-	-	-
	2	77	-0.34	0.414	57	-0.74	0.147
	3	-	-	-	65	-1.57	0.003
SSC13 (SW207)	1	69	-0.54	0.286	73	-1.69	0.001
	2	66	0.32	0.540	-	-	-
	3	-	-	-	67	-1.01	0.046
SSC13 (SW1550)	1	93	-0.95	**0.025**	82	-1.60	0.001
	2	80	0.01	0.980	-	-	-
	3	-	-	-	67	-1.10	0.024
SSC14 (SW245)	1	83	0.09	0.833	45	-1.28	0.024
	2	85	-0.66	0.164	-	-	-
	3	-	-	-	76	-1.23	0.009
	4	-	-	-	36	-1.67	0.059
SSC16 (MCSeq5008)	1	74	-0.56	0.220	72	-1.39	0.009
	2	70	-0.48	0.350	-	-	-
	3	-	-	-	64	-1.31	0.012

Interactions were tested between eight microsatellites (representing eight candidate regions) at position 66 cM on SSC2. QTL effects $\alpha_{SSC2(kg)}$ given as (MS – LW) and the associated p-values are shown. In bold are the significant p-values obtained for sires 072866 and 072868 for which no QTL was detected on SSC2.

In order to reduce the number of candidate regions, a complementary genome scan was carried out using the F2 PorQTL design. The F2 animals were grouped based on the breed origin (LW or MS) of their paternal allele. All the microsatellites available from the original genome scan of the population were tested by interval mapping as previously described for interactions within the 40-70 cM region on SSC2. A region on SSC13 covered with three microsatellites (S0222, SW225 and SW38) showed the highest significance for an interaction with the SSC2 region (figure 4.4): the SSC2 QTL was detected only in F2 animals which inherited the LW haplotype from their sire on SSC13. The SSC13 region spanning 30 cM from SW225 to SW38, containing the SW1550 microsatellite used in the backcross population, was therefore considered to be the strongest candidate region for the interaction with the SSC2 QTL in both F2 and backcross pedigrees.

The three microsatellites of the SSC13 region that were significant in the F2 design were genotyped on the backcross animals. First, sire haplotypes were constructed from the familial segregation information. Three different haplotypes were segregating in 072866, 072868, 084024 and 085830 sires. Sire 084026 was heterozygous for haplotypes 1 and 2, sire 085830 was homozygous for haplotype 2, and sires 072866 and 072868 were heterozygous for haplotype 1 and haplotype 3. Second, the interaction of these haplotypes with the SSC2 region was tested. Progeny which inherited the paternal haplotype 1 or haplotype 2 showed significantly different phenotypic values according to the breed origin of their paternal haplotype at 66 cM on SSC2 (figure 4.5). By contrast, progeny that inherited the paternal SSC13 haplotype 3 showed no differences in phenotypic values based on their paternal haplotype on SSC2.

4.4 DISCUSSION

The aims of this study were to confirm and fine-map a QTL underlying fatness traits segregating around 30-70 cM downstream of the *IGF2* gene through the production of animals dedicated to this project. Progeny testing of 084026 and 085830 sires confirmed that a QTL affecting fatness was localized around 60-70 cM on SSC2. Surprisingly, the two additional full sibs 072866 and 072868, which shared the same identical-by-descent haplotypes in this region, showed no evidence for this QTL.

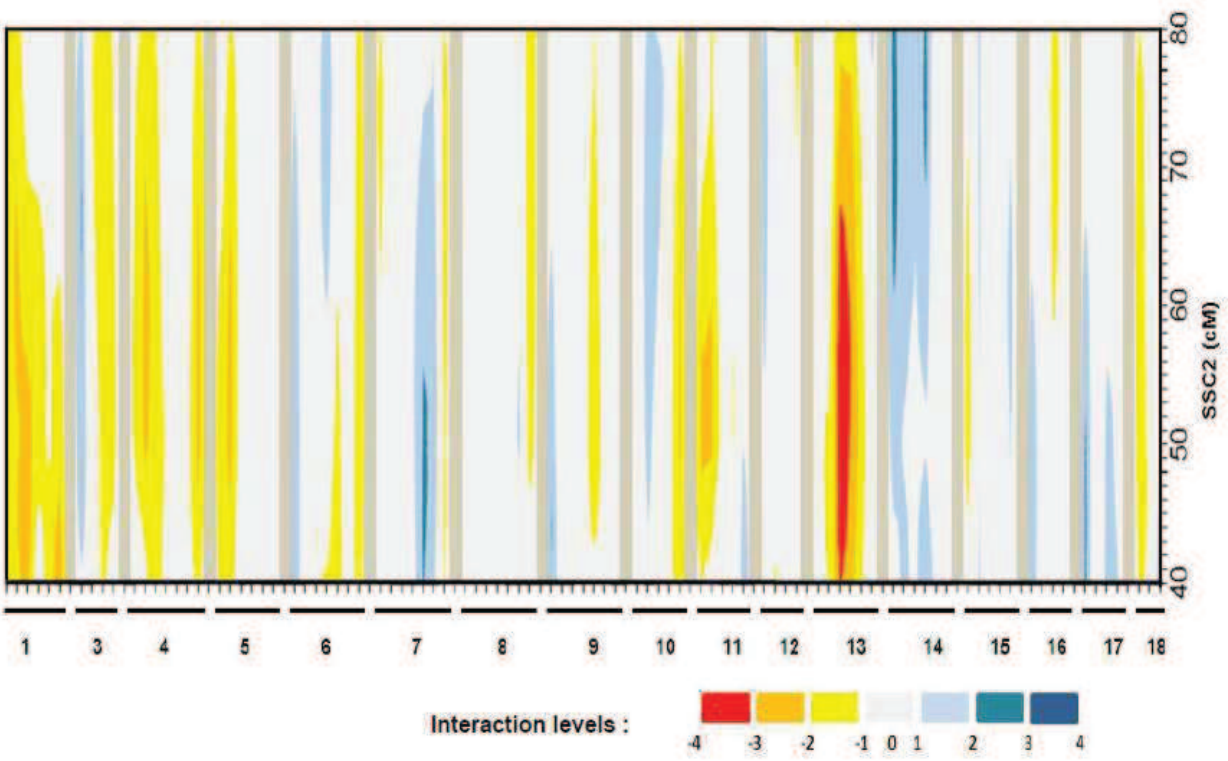


Figure 4.4 – Contrasts between p-values (log_{10} scale) obtained for interval mapping tests within progeny with LW paternal alleles and within progeny with MS paternal alleles on the PorQTL pedigree. Interval mapping tests were applied on SSC2 (y-axis, 40-70 cM region) with all available positions in the genome scan as candidate for interaction (x-axis, from SSC1 to SSC18 (SSC2 being excluded)). A blue (orange) spot indicates that the SSC2 QTL is more likely to be detected in progeny that inherited the Large White (Meishan) allele paternally at the candidate position for interaction. Extreme colors indicate the highest p-value differences, i.e. the strongest candidates for interaction with the SSC2 QTL.

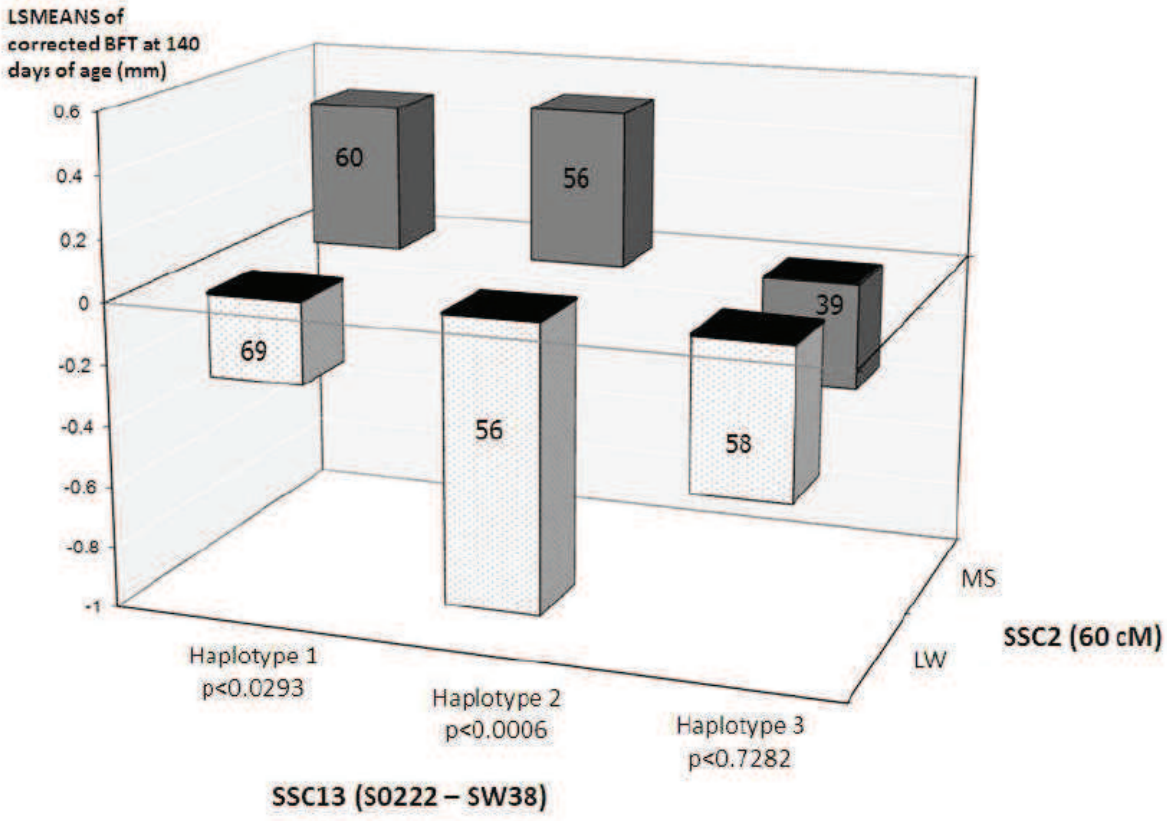


Figure 4.5 – Least-Squares MEANS of corrected values of live BFT measured at 140 days of age for progeny of sires 072866, 072868, 084026 and 085830. LSMEANS were computed for each combination of SSC13 paternal haplotype designed from S0222 to SW38 and paternal SSC2 allele breed origin at 66 cM. The number of animals per group is indicated on each bar. P-values, below the haplotype names, refer to the Student's tests applied to the LSMEANS differences obtained within each SSC13 haplotype, the p-value of the test being less than 0.0006.

To explain this discrepancy, eight regions (localised on seven different chromosomes) were retained and tested for interactions with SSC2. As concerns the two regions on SSC3 and SSC8, significant results were obtained with only one allele of the tested microsatellite. In these regions, interaction would imply that only one allele enabled the detection of a significant QTL effect on SSC2. In these cases, the power of the design to detect the SSC2 QTL would be influenced by the proportion of offspring carrying these alleles. Alternatively, for the regions selected on SSC6 and SSC13 (SW1550), the SSC2 QTL was detected with all alleles except one segregating within the 072866/072868 sires' families. This particular allele on SSC6 or SSC13, inhibited the segregation at SSC2 QTL. For these two candidate

regions, this suggests that one allele inhibits the MS QTL allele of SSC2. An additional argument in favour of the SSC13 region was provided by the analysis of the F2 PorQTL design, in which this region was the strongest candidate for interaction with SSC2 in a whole genome scan. Here, the QTL was detected on SSC2 when the progeny inherited a LW allele from their sire in the SSC13 region. SW207 is in this interacting region, and the non-significant result obtained (table 4.2) may be due to the low number of progeny to which a paternal allele could be attributed with certainty, as compared to SW1550.

Number of studies already aimed at identifying QTL interactions in pig (DUTHIE *et al.* 2010; DUTHIE *et al.* 2011a; DUTHIE *et al.* 2011b; ESTELLE *et al.* 2008; NOGUERA *et al.* 2009; UEMOTO *et al.* 2009; VARONA *et al.* 2002; WEI *et al.* 2010a), and many different pairs of interacting QTL were described underlying various traits including fatness. By comparing our results to the ones from these seven extant papers, no identical epistatic pair of regions may be found in common. This could be attributable to the different traits analysed, the different breeds involved in the pedigrees and to the power of the analyses. In all these studies, few pairs of interacting QTL were detected in contrast to studies carried out in mice where all detected QTL underlying adiposity were involved in epistatic interactions (CHEVERUD *et al.* 2001). A comparison of the regions detected in the pig in this study and orthologous regions in the mouse indicated no overlap in known epistatic genes. This indicates that the interaction described in this study between SSC2 and SCC13 is novel in both pigs and mice.

Most of the studies in search of epistasis were based on the Line-Cross model which assumes that founder lines are fixed for alternative QTL alleles for the trait of interest, so that each progeny-tested sire is heterozygous for each interacting QTL. This classifies the offspring in sixteen different classes, reduced to nine classes by merging the heterozygous individuals at each interacting locus, as illustrated in (WEI *et al.* 2010a). This assumption can be easily verified in mouse experiments using inbred lines but it is much more difficult to demonstrate in pigs or other livestock animals. For example, we already know, through the analysis of the *IGF2*-intron3-G3072A mutation, that the assumption of fixed alleles is not always confirmed (VAN LAERE *et al.* 2003)., Because the QTL alleles fixation cannot be considered as a rule, therefore it is very likely that animals carrying different QTL alleles can belong to the same "breed-based" class. This confusion between QTL alleles and breeds could partly explain why few interacting pairs of QTL are detected compared to mice. Additionally, it has been shown that departure from the Line Cross hypothesis decreases the power to detect epistasis (WEI *et al.* 2010b). The Line Cross model remains highly used for epistatic interaction as it allows keeping group sizes large enough to afford sufficient statistical power. For this reason, it was also used in this study for the analysis of the F2 animals, considering only the segregation of the paternal alleles. However, finer interaction

analyses were performed with the BC families, as it obviated any confusion between the breeds and the QTL alleles. The four sires analysed were full sibs and were IBD for both the paternal and maternal haplotypes in the 40-66 cM region of SSC2. Therefore, by construction, they all shared the same alleles (Q and q alleles) at the SSC2 QTL. This binary situation could however only be asserted in this SSC2 region, and not for the rest of the genome. Nevertheless, because the four sires analysed for epistasis were full-sibs, only a few different alleles were segregating and, for a pool of two sires, interactions have been tested with a maximum of three different alleles. This advantageous situation enabled us to work out of the Line Cross assumption and to determine that among the three possible haplotypes segregating in the SSC13 region only one inhibited the segregation at the SSC2 QTL.

QTL detection and localization

In this analysis, a QTL underlying fatness traits was detected around 60-70 cM, with Meishan alleles increasing fatness. Based on the recombination points, the QTL region spanned 30 cM and was delimited by the recombination point on the paternal allele of 072868 (at 37 cM) and by the one at the end of the Meishan haplotype on sire 085830 (at 67 cM). This study confirmed previous analyses of SSC2 based on Meishan and European White breeds such as Large White and Landrace, where Chinese alleles were always associated with a higher fatness as compared to European White breeds alleles (BIDANEL *et al.* 2001; DE KONING *et al.* 1999; MILAN *et al.* 2002). However, QTL intervals were different according to the trait with a region from 0 to 30 cM underlying backfat thickness measured on living animals (BIDANEL *et al.* 2001) and a 43 cM region from 40 to 83 cM underlying backfat thickness measured on carcass (DE KONING *et al.* 1999; MILAN *et al.* 2002). In a Meishan-White composite resource population, a QTL underlying backfat thickness measured at 14 weeks of age was detected around 74 cM (ROHRER 2000), which is consistent with the most likely positions we detected here. In crosses between Pietrain and Wild boar or Meishan, QTL underlying carcass fatness traits were mainly detected between 55 and 75 cM on SSC2 and not in the first centimorgans, even if the Pietrain breed is nearly fixed for the *IGF2*-intron3-3072A allele and Wild Boars and Meishan are supposed to be fixed for the alternate allele (GELDERMANN *et al.* 2010). When the *IGF2*-intron3-G3072A mutation was taken into account in the analysis of Meishan x European White breed crosses, the segregation of a QTL underlying BFT around 40-50 cM was observed (TORTEREAU *et al.* 2011), and the segregation of QTL affecting fatness traits between 40 cM and 60 cM was also reported by Lee et al in a Wild Boar x Meishan pedigree were all founders animals were G/G for the *IGF2*-intron3-G3072A mutation (LEE *et al.* 2003). In our study also, all the progeny-tested sires were homozygous G/G for the *IGF2*-intron3-G3072A mutation so variation of the studied traits would not be due to this polymorphism.

Fine-mapping through the production of additional animals.

The backcross design was set up to confirm and fine-map a QTL underlying fatness traits, segregating around 30-70 cM on SSC2. This marker-assisted backcrossing method already gave encouraging results in crosses involving European breeds mated with Meishan or Wild Boars (MARKLUND *et al.* 1999; RIQUET *et al.* 2011; SANCHEZ *et al.* 2006) and a highly significant result with the decrease from 70 to 3.3 cM of the FAT1 QTL interval on SSC4 (BERG *et al.* 2006). If in our case this design was not directly succesfull in decreasing the QTL confidence interval, it has allowed interactions between the SSC2 QTL and a locus mapped on SSC13 to be detected. Despite being a long and expensive method requiring a lot of animals, marker-assisted backcrossing remains a powerful method to finely dissect segregating QTLs in livestock.

4.5 CONCLUSION

The marker-assisted backcrossing method presented in this study ended first with the confirmation of a fatness QTL between 37 and 67 cM on SSC2 and then with the highlighting of interactions between this QTL and at least another region on SSC13. As far as we know, it is the first time that such a design has enabled the detection of epistasis. In addition, by considering only different haplotypes whatever the original breed, no assumption of line-cross model was used. This study reappraises the marker-assisted backcrossing strategy as not only an efficient method to fine-map QTL but also to understand the mode of their segregation. When such interactions are described, additional steps are needed to improve the mapping accuracy of the locus, which can delay the fine-mapping of the QTL. Therefore, the marker-assisted backcrossing design provides valuable information which, however, has to be balanced with the time that it requires for fine-mapping.

Acknowledgements

This work was funded by the animal genetics division of INRA. FT is supported by a PhD grant from Wageningen University. Genotyping were performed on the Genomic Platform of Toulouse (France): http://get.genotoul.fr/

The authors would like to thank Byron C. Jones for his precious help for the English revision of the manuscript and the reviewers for their helpful comments.

REFERENCES

BERG, F., S. STERN, K. ANDERSSON, L. ANDERSSON and M. MOLLER, 2006 Refined localization of the FAT1 quantitative trait locus on pig chromosome 4 by marker-assisted backcrossing. BMC Genet **7:** 17.

BIDANEL, J. P., D. MILAN, N. IANNUCCELLI, Y. AMIGUES, M. Y. BOSCHER *et al.*, 2001 Detection of quantitative trait loci for growth and fatness in pigs. Genet Sel Evol **33:** 289-309.

BIDANEL, J. P., and M. F. ROTHSCHILD, 2002 Current status of quantitative trait locus mapping in pigs. Pig News Inf. **23:** 39N-53N.

CHEVERUD, J. M., T. T. VAUGHN, L. S. PLETSCHER, A. C. PERIPATO, E. S. ADAMS *et al.*, 2001 Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. Mamm Genome **12:** 3-12.

DE KONING, D. J., L. L. JANSS, A. P. RATTINK, P. A. VAN OERS, B. J. DE VRIES *et al.*, 1999 Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics **152:** 1679-1690.

DUTHIE, C., G. SIMM, A. DOESCHL-WILSON, E. KALM, P. W. KNAP *et al.*, 2010 Epistatic analysis of carcass characteristics in pigs reveals genomic interactions between quantitative trait loci attributable to additive and dominance genetic effects. J Anim Sci **88:** 2219-2234.

DUTHIE, C. A., G. SIMM, A. DOESCHL-WILSON, E. KALM, P. W. KNAP *et al.*, 2011a Epistatic quantitative trait loci affecting chemical body composition and deposition as well as feed intake and feed efficiency throughout the entire growth period of pigs. Livestock Science **138:** 34-48.

DUTHIE, C. A., G. SIMM, A. DOESCHL-WILSON, E. KALM, P. W. KNAP *et al.*, 2011b Quantitative trait loci for meat quality traits in pigs considering imprinting and epistatic effects. Meat Sci **87:** 394-402.

ELSEN, J. M., B. MANGIN, B. GOFFINET, D. BOICHARD and P. LE ROY, 1999 Alternative models for QTL detection in livestock. I. General introduction. Genet Sel Evol **31:** 213-224.

ESTELLE, J., F. GIL, J. M. VAZQUEZ, R. LATORRE, G. RAMIREZ *et al.*, 2008 A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. J Anim Sci **86:** 3290-3299.

GELDERMANN, H., S. CEPICA, A. STRATIL, H. BARTENSCHLAGER and S. PREUSS, 2010 Genome-wide mapping of quantitative trait loci for fatness, fat cell characteristics and fat metabolism in three porcine F2 crosses. Genet Sel Evol **42:** 31.

GILBERT, H., and P. LE ROY, 2007 Methods for the detection of multiple linked QTL applied to a mixture of full and half sib families. Genet Sel Evol **39:** 139-158.

IANNUCCELLI, E., N. WOLOSZYN, J. ARHAINX, J. GELLIN and D. MILAN, 1996 GEMMA: a database to manage and automate microsatellite genotyping. Anim Genet **27 Suppl. 2:** 55.

JEON, J. T., O. CARLBORG, A. TORNSTEN, E. GIUFFRA, V. AMARGER *et al.*, 1999 A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nat Genet **21:** 157-158.

JUNGERIUS, B. J., A. S. VAN LAERE, M. F. TE PAS, B. A. VAN OOST, L. ANDERSSON *et al.*, 2004 The IGF2-intron3-G3072A substitution explains a major imprinted QTL effect on backfat thickness in a Meishan x European white pig intercross. Genet Res **84:** 95-101.

KNOTT, S. A., L. MARKLUND, C. S. HALEY, K. ANDERSSON, W. DAVIES *et al.*, 1998 Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics **149:** 1069-1080.

LEE, S. S., Y. CHEN, C. MORAN, S. CEPICA, G. REINER *et al.*, 2003 Linkage and QTL mapping for Sus scrofa chromosome 2. J Anim Breed Genet **120 (suppl 1.):** 11-19.

MANGIN, B., B. GOFFINET, P. LE ROY, D. BOICHARD and J. M. ELSEN, 1999 Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations. Genet Sel Evol **31:** 225-237.

MARKLUND, L., P. E. NYSTROM, S. STERN, L. ANDERSSON-EKLUND and L. ANDERSSON, 1999 Confirmed quantitative trait loci for fatness and growth on pig chromosome 4. Heredity **82 (Pt 2):** 134-141.

MILAN, D., J. P. BIDANEL, N. IANNUCCELLI, J. RIQUET, Y. AMIGUES *et al.*, 2002 Detection of quantitative trait loci for carcass composition traits in pigs. Genet Sel Evol **34:** 705-728.

NEZER, C., L. MOREAU, B. BROUWERS, W. COPPIETERS, J. DETILLEUX *et al.*, 1999 An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. Nat Genet **21:** 155-156.

NOGUERA, J. L., C. RODRIGUEZ, L. VARONA, A. TOMAS, G. MUNOZ *et al.*, 2009 A bi-dimensional genome scan for prolificacy traits in pigs shows the existence of multiple epistatic QTL. BMC Genomics **10:** 636.

RATTINK, A. P., D. J. DE KONING, M. FAIVRE, B. HARLIZIUS, J. A. VAN ARENDONK *et al.*, 2000 Fine mapping and imprinting analysis for fatness trait QTLs in pigs. Mamm Genome **11:** 656-661.

RIQUET, J., H. GILBERT, B. SERVIN, M. P. SANCHEZ, N. IANNUCCELLI *et al.*, 2011 A locally congenic backcross design in pig: a new regional fine QTL mapping approach miming congenic strains used in mouse. BMC Genet **12:** 6.

ROHRER, G. A., 2000 Identification of quantitative trait loci affecting birth characters and accumulation of backfat and weight in a Meishan-White Composite resource population. J Anim Sci **78:** 2547-2553.

SANCHEZ, M. P., J. RIQUET, N. IANNUCCELLI, J. GOGUE, Y. BILLON *et al.*, 2006 Effects of quantitative trait loci on chromosomes 1, 2, 4, and 7 on growth, carcass, and meat quality traits in backcross Meishan x Large White pigs. J Anim Sci **84:** 526-537.

TORTEREAU, F., H. GILBERT, H. C. HEUVEN, J. P. BIDANEL, M. A. GROENEN *et al.*, 2010 Combining two Meishan F2 crosses improves the detection of QTL on pig chromosomes 2, 4 and 6. Genet Sel Evol **42:** 42.

TORTEREAU, F., H. GILBERT, H. C. HEUVEN, J. P. BIDANEL, M. A. GROENEN *et al.*, 2011 Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in Sino-European pig pedigrees. Genet Sel Evol **43:** 11.

UEMOTO, Y., S. SATO, C. OHNISHI, S. TERAI, A. KOMATSUDA *et al.*, 2009 The effects of single and epistatic quantitative trait loci for fatty acid composition in a Meishan x Duroc crossbred population. J Anim Sci **87:** 3470-3476.

VAN LAERE, A. S., M. NGUYEN, M. BRAUNSCHWEIG, C. NEZER, C. COLLETTE *et al.*, 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature **425:** 832-836.

VARONA, L., C. OVILO, A. CLOP, J. L. NOGUERA, M. PEREZ-ENCISO *et al.*, 2002 QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects. Genet Res **80:** 145-154.

WEI, W. H., Y. DUAN, C. HALEY, J. REN, D. J. DE KONING *et al.*, 2010a High throughput analyses of epistasis for swine body dimensions and organ weights. Animal Genetics **42:** 15-21.

WEI, W. H., S. KNOTT, C. S. HALEY and D. J. DE KONING, 2010b Controlling false positives in the mapping of epistatic QTL. Heredity **104:** 401-409.

5

Sex specific differences in recombination rate in the pig are correlated with GC content

Flavie Tortereau[1,2], Bertrand Servin[2], Laurent Frantz[1], Hendrik-Jan Megens[1], Denis Milan[2], Gary Rohrer[3], Ralph Wiedmann[3], Jonathan Beever[4], Alan L. Archibald[5], Larry B. Schook[4] and Martien A.M. Groenen[1]

[1] Wageningen University, Animal Breeding and Genomics Centre, PO Box 338, 6700AH Wageningen, The Netherlands ; [2] INRA, Laboratoire de Génétique Cellulaire, 31320 Auzeville-Tolosane, France ; [3] USDA, ARS, US Meat Animal Research Center, PO Box 166, Spur 18D, Clay Center, NE 68933-0166, USA ; [4] University of Illinois, Institute for Genomic Biology, 382 ERML, 1201 W. Gregory Avenue, Urbana, IL 61801, USA ; [5] Division of Genetics and Genomics, Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK Cedex

Submitted

Abstract

Background: The availability of a high-density SNP chip and a reference genome sequence of the pig have enabled the construction of a high-density linkage map. A high density linkage map is an essential tool for the further fine-mapping of QTL for a variety of traits in the pig and for a better understanding of the mechanisms underlying genome evolution.

Results: Four different pig pedigrees were genotyped with the Illumina PorcineSNP60 Beadchip. Recombination maps were computed for each individual pedigree using a common set of markers. The resulting genetic maps comprised 38,599 SNPs, including 508 SNPs not positioned on the current assembly of the pig genome (build 10.2). The total genetic length varied according to the pedigree, from 1797 to 2149 cM. Female maps were longer than male maps, with a notable exception for SSC1 where male maps are characterised by a higher recombination rate than females in the region between 91-250 Mb. The recombination rate varied among chromosomes and along individual chromosomes, regions with high recombination rates tending to cluster close to the chromosome ends, irrespective of the position of the centromere. Correlations between main sequence features and recombination rate were investigated and significant correlations were obtained for all the studied motifs. Regions characterised by a high recombination rate were enriched for specific GC-rich sequence motifs as compared to low recombinant regions. These correlations were higher in females than in males, and females were found to be more recombinant than males when the GC content of the region was higher than 0.4.

Conclusions: The analysis of the recombination rate along the pig genome highlighted that the more recombinant regions tend to cluster around the telomeres irrespective of the location of the centromere. Major sex-differences in recombination were observed with a higher recombination rate in the females only within GC-rich regions, with females exhibiting a much stronger correlation between recombination rate and specific sequence features.

5.1 INTRODUCTION

Linkage maps have been widely used to identify genomic regions that influence phenotypic traits. Besides the expected advances in fine-mapping of Quantitative Trait Loci (QTL) (DAW *et al.* 2000; FINGERLIN *et al.* 2006), high-density linkage maps provide a major framework in the assembly of genome sequences and for studies of the evolution of these genomes through the analysis of recombination. In fact, recombination lies at the heart of every genetic analysis, and whereas linkage maps in the past were constructed primarily to aid in the generation of a physical map, linkage maps are currently being recognized as indispensable tools to study virtually every aspect of genome biology. Genomic features that have been shown to correlate with recombination rate include GC content, gene density, gene expression, epigenetic modifications, nucleosome formation, repetitive element composition, isochore structure, but also patterns of genetic variation and differentiation within and between populations. For this reason, increasingly dense recombination maps have been constructed in the so called 'post-genomic era' for species such as human and mouse, focussing on identifying hotspots of recombination, and, recently, variation in the use of these hotspots between populations and between genders.

Despite the evident importance of accurate and comprehensive linkage maps in the post-genomic era, comprehensive maps are currently only available for a handful of vertebrate species (human, mouse, rat, cattle, dog, zebra finch and chicken). That limited scope on recombination landscape in genomes severely limits the possibility of drawing general conclusions regarding recombination rates in genomes, particularly now that it is becoming increasingly clear that various mechanisms can work together in creating a very dynamic use of recombination hotspots over time.

In swine, the first linkage map covering all the autosomes of the pig was established in 1995 (ARCHIBALD *et al.* 1995) and a more dense map comprising about 1,200 markers was published in 1996 (ROHRER *et al.* 1996). Two other linkage maps comprising around 240 loci were published in the late 1990s (MARKLUND *et al.* 1996; MIKAWA *et al.* 1999). These four maps were mainly based on microsatellites, Restriction Fragment Length Polymorphisms (RFLPs) and protein polymorphisms. More recently, SNPs were added to these maps (VINGBORG *et al.* 2009), but the resolution remained low with an average SNP distance of 3.94 cM. With the advent of genome-wide high-density SNP chips, genetic maps can comprise an increasing number of markers. Until now, such high-density genetic maps, based on microsatellites and SNPs, were computed for human (KONG *et al.* 2002), mouse (SHIFMAN *et al.* 2006), chicken (GROENEN *et al.* 2009; ELFERINK *et al.* 2010), cattle (ARIAS *et al.* 2009) and dog (WONG *et al.* 2010). With the release of Illumina's Porcine SNP60 Beadchip (RAMOS *et al.* 2009), it became possible to compute such a high-density genetic map of the porcine genome. In this work, we present four recombination maps based on the same SNPs for four different pedigrees, using an

a priori order of the SNPs. This a priori order was based on the position of the SNPs on the porcine RH map (Servin *et al.*, in preparation) and on the position of the SNPs on the genome sequence (build 10.2) of the pig. An average recombination rate (estimated over the four studied pedigrees) was calculated along the genome (sex-averaged, male and female maps) and particular attention was given to its variations between and along chromosomes with respect to some sequence features.

5.2 MATERIALS AND METHODS

Mapping populations

The animals used to compute the recombination maps belong to four independent pedigrees. Three were based on an F2 design (including one reciprocal cross) and one was based on multi-stage crosses. Details about the four pedigrees are presented in table 5.1.

To compute recombination maps, only families with more than four full-sibs were retained in the analysis. Therefore, recombination maps were calculated based on the information from 573 animals of the ILL pedigree, 247 from the UIUC pedigree, 204 from the ROS pedigree and 1298 from the USDA pedigree.

SNP genotyping

The four pig pedigrees were genotyped using the Illumina PorcineSNP60 Beadchip (San Diego, CA, USA). Each pedigree was genotyped independently, and a total of 664 samples from ILL, 337 from UIUC, 208 from ROS and 1337 from USDA were genotyped. To carry out the computation of recombination maps, only SNPs with a call rate higher than 97% were retained. In addition, all the genotypes were checked for Mendelian inheritance and erroneous genotypes were set as missing.

Table 5.1 – Description of the four pedigrees

Pedigree	Cross	F0 males	F0 females	F1 males	F1 females	F2
ILL	F2	3 Berkshire	17 Duroc	5	44	595
UIUC	F2	3 Meishan	7 Yorkshire	3	15	260
ROS	F2	5 Meishan	6 Large White	2	16	151
	(reciprocal)	5 Large White	5 Meishan	4	14	
USDA *	overlapping F2	13	35	12	27	97
		13	55	10	53	376
		12	66	8	62	547

*: the population from which were sampled the founders of the USDA pedigree is composed of ¼ Duroc, ¼ Large White, ¼ maternal Landrace and ¼ high growth Landrace.

Recombination map calculation

Recombination maps were computed for each pedigree independently with a unique set of SNPs. The a priori order of the SNPs was obtained by combining information from the RH mapping (Servin *et al.*, in prep) and in silico mapping of the SNPs. The genotyping of the two RH panels of the porcine genome on the PorcineSNP60 Beadchip enabled the computation of a physical map (Servin *et al.*, in prep). SNPs were positioned on the current pig genome sequence (build 10.2) by aligning the 200 bp sequence adjacent to the SNP against build 10.2 using BLAT (Kent 2002). In order to create the a priori order needed for the analysis, the RH order was considered as the basic order and when it was consistent with the sequence assembly, SNPs from the assembly were retrieved and included in the a priori order. Recombination rates along chromosomes were estimated using the method of Coop et al. (2008). Briefly, haplotypes transmitted by a parent to each of its offspring were inferred based on informative SNPs. Then, within a given nuclear family, one of the offspring (template) was successively compared to the others: at a marker, it was deduced whether both offspring were Identical By Descent (IBD) or not. Any switch from an IBD to a non-IBD status indicated a recombination. Regions where the majority of offspring showed a recombination were considered as indicative of a recombination in the template offspring. Finally, the parental phases were partially reconstructed, allowing identification of recombination events that occurred in each meiosis (COOP *et al.* 2008). Recombination rates were transformed into centimorgans (cM) using the Haldane genetic distance.

Finally, four recombination maps were computed and recombination rates in cM/Mb were calculated for each pedigree along the genome. These recombination

rates were estimated in non-overlapping bins of approximately 1 Mb considering the exact SNP positions as the delimiters of the bins. An average recombination rate was also estimated along the genome over the four pedigrees and was used to carry out further analyses in relation to correlation with sequence features. Similarly, female and male recombination rates were estimated along the genome.

Correlation of recombination with sequence parameters

The average recombination rate was compared to the distribution of various sequence motifs including repetitive elements (LINEs, SINEs, LTRs, simple repeats and low-complexity repeats), GC content, and GC rich motifs previously shown to be correlated with high recombination rates (CCTCCT, CCTCCCT, CTCTCCC, CCCCCCC, CCCCACCCC, the CTCF consensus sequence CCNCCNGGNGG and the PRDM9 consensus binding sequence CCNCCNTNNCCNC). The distribution of sequence motifs and GC content were calculated for bins of 1 Mb using the current assembly (build 10.2) and the correlations with recombination rates were tested using Pearson's correlation coefficient with the corr procedure in SAS (SAS® 9.1 , SAS Institute, Inc.). To further investigate the link between sequence features and recombination rate, the sequence composition of jungle and desert regions (SHIFMAN *et al.* 2006) were compared. Jungle regions were defined as the 1 Mb intervals with the 10% highest recombination rates, and conversely, desert regions were defined as the 1 Mb intervals with the 10% lowest recombination rates. These ratios were also estimated independently in males and females.

Finally, the correlation between recombination rate and the physical distance to the closest chromosome end was also estimated.

5.3 RESULTS

Genotyping quality

The Illumina PorcineSNP60 Beadchip comprising 64,232 SNPs was genotyped on the four studied pedigrees. The a priori order used to compute the genetic map comprised 44,760 SNPs: 35,098 from the RH order, and 9,662 derived from the sequence assembly. Of the 44,760 SNPs, 5,980 SNPs were discarded because of their low call-rate (<97 %), and an additional set of 181 SNPs was removed because there were Mendelian inconsistencies in several families. When Mendelian inconsistencies were only limited to one particular family per pedigree, genotypes were set to zero in this family. A total of 168 individuals were removed from the four pedigrees because of their high proportion of incorrect genotypes due to

either pedigree or genotyping errors. Finally, the average number of informative meiosis per marker was 432 for ILL, 200 for UIUC, 670 for USDA and 120 for ROS.

Recombination maps

The a priori order, on which the recombination analysis was based, comprised 44,760 SNPs, including 556 SNPs mapped to unplaced scaffolds and 480 SNPs with no match on the assembly. Finally, genetic maps comprised 38,599 SNPs among which 508 were positioned on the unknown chromosome and 420 did not have a match on the assembly. On average, there were 2,144 SNPs per chromosome, ranging from 1,011 (SSC18) to 5,293 (SSC1) (table 5.2). This set of SNPs was chosen as being valid for all four pedigrees; recombination maps being calculated separately for each of them. The rates of phase reconstruction differed for the four pedigrees. For the complete genome, the highest rate was obtained for the UIUC pedigree (99.0%) and the lowest rate was obtained for the ROS pedigree (87.0%). The ILL and USDA pedigrees were intermediate with phase reconstruction rates of 96.5% and 92.0% respectively.

The details of the genetic maps calculated for each of the four pedigrees are presented in table 5.2. The estimates of the total genetic length of the 18 autosomes were 2,012 cM for ILL, 2,149 cM for UIUC, 1,797 cM for USDA and 1,858 cM for ROS. The largest chromosome was SSC6 for ILL, UIUC and ROS pedigrees with 148, 151 and 148 cM respectively whereas it was SSC1 for the USDA pedigree with 130 cM. SSC18 was the smallest chromosome for all the pedigrees, its length varying from 44 cM for the ROS pedigree to 71 cM for the UIUC pedigree.

Because within the USDA and ROS pedigrees female recombination was not well taken into account (because of the low number of offspring per dam or because of missing genotypes), male and female recombination maps were described separately only for the ILL and UIUC pedigrees (table 5.3). As expected, the total lengths were longer for the female maps (2,244 and 2,545 cM for ILL and UIUC respectively) than for the male maps (1,782 and 1,747 cM for ILL and UIUC respectively). SSC1 stands out as an exception, with the male maps being longer than the female maps. This difference is due to a low recombination rate in the females in the region between 90 and 250 Mb (figure 5.2). In this 90-250 Mb region, the average recombination rate in females was 0.056 and 0.031 cM/Mb for ILL and UIUC respectively whereas it was 0.286 and 0.290 for males in ILL and UIUC pedigrees respectively.

Table 5.2 – Description of the linkage maps of the four pedigrees.

			ILL		UIUC		USDA		ROS	
SSC	Nb SNP	Physical length (Mb)	Linkage map (cM)	cM/Mb*	Linkage map (cM)	cM/Mb	Linkage map (cM)	cM/Mb	Linkage map (cM)	cM/Mb
1	5293	308	145	0.37	144	0.38	130	0.33	140	0.37
2	2492	158	122	0.64	137	0.68	110	0.57	113	0.60
3	2044	141	120	0.74	122	0.76	113	0.71	106	0.65
4	2789	143	125	0.70	129	0.73	111	0.64	115	0.61
5	1737	109	114	0.83	124	0.94	97	0.73	104	0.89
6	2156	157	148	0.78	151	0.87	122	0.68	148	0.85
7	2693	132	132	0.89	144	0.97	117	0.78	119	0.78
8	2008	147	112	0.63	124	0.70	110	0.62	110	0.62
9	2166	153	127	0.74	135	0.81	117	0.69	112	0.63
10	1173	77	109	1.22	116	1.29	99	1.10	89	0.96
11	1332	86	85	0.77	96	0.89	77	0.70	73	0.62
12	1038	63	99	1.33	99	1.30	86	1.11	94	1.24
13	2875	216	113	0.45	122	0.51	97	0.40	117	0.47
14	3142	153	124	0.75	138	0.85	110	0.64	111	0.64
15	2085	154	108	0.61	123	0.67	97	0.54	110	0.61
16	1337	85	83	0.80	91	0.86	78	0.75	77	0.73
17	1227	68	78	0.94	83	1.05	67	0.81	76	0.98
18	1011	60	68	0.91	71	0.95	59	0.80	44	0.56
TOTAL	***38599***	***2334***	***2012***	***0.78***	***2149***	***0.85***	***1797***	***0.70***	***1858***	***0.71***

* The ratio cM/Mb was calculated in bins of 1 Mb.

Table 5.3 – Description of sex-specific linkage maps of the ILL and UIUC pedigrees

	ILL ♀		ILL ♂		UIUC ♀		UIUC ♂	
SSC	**Linkage map (cM)**	**cM/Mb***	**Linkage map (cM)**	**cM/Mb**	**Linkage map (cM)**	**cM/Mb**	**Linkage map (cM)**	**cM/Mb**
1	138	0.34	151	0.40	129	0.33	159	0.42
2	133	0.70	111	0.58	167	0.83	107	0.53
3	130	0.80	111	0.69	135	0.85	109	0.67
4	137	0.76	112	0.63	150	0.87	108	0.60
5	128	0.93	100	0.74	147	1.05	100	0.82
6	170	0.89	125	0.68	181	1.05	120	0.69
7	146	0.96	118	0.82	176	1.17	112	0.77
8	120	0.66	105	0.60	140	0.78	108	0.63
9	144	0.85	110	0.64	167	1.00	103	0.63
10	126	1.41	93	1.03	140	1.53	91	1.05
11	107	0.95	63	0.59	118	1.07	74	0.72
12	117	1.58	81	1.08	124	1.70	74	0.90
13	115	0.46	112	0.44	133	0.51	110	0.44
14	135	0.84	112	1.01	163	0.66	112	0.69
15	115	0.63	102	0.58	151	0.81	95	0.53
16	96	0.90	70	0.70	119	1.13	62	0.59
17	100	1.19	56	0.70	114	1.40	52	0.70
18	87	1.17	50	0.65	91	1.23	51	0.68
TOTAL	**2244**	**0.89**	**1782**	**0.70**	**2545**	**1.00**	**1747**	**0.67**

* The ratio cM/Mb was calculated in bins of 1 Mb.

Recombination rates

Recombination rates were calculated for non-overlapping bins of 1 Mb with marker positions delimiting the intervals. At the level of the genome, the highest average recombination rate was obtained for the UIUC pedigree with 0.85 cM/Mb, the lowest being obtained for the USDA pedigree with 0.70 cM/Mb (table 5.2). This ratio highly varied depending on the physical length of the chromosomes, the shortest ones having higher ratios than the longest ones (figure 5.1).

For the four pedigrees, the highest recombination rate was observed for SSC12 with values of 1.33, 1.30, 1.11 and 1.24 cM/Mb for ILL, UIUC, USDA and ROS respectively. The lowest recombination rate was obtained on SSC1 with 0.37, 0.38, 0.33 and 0.37 cM/Mb for ILL, UIUC, USDA and ROS respectively (table 5.2). At the genome level, recombination rates were higher in females than in males. At the chromosome levels, only SSC1 displayed higher recombination rates in males than in females, for ILL and UIU pedigrees (table 5.3).

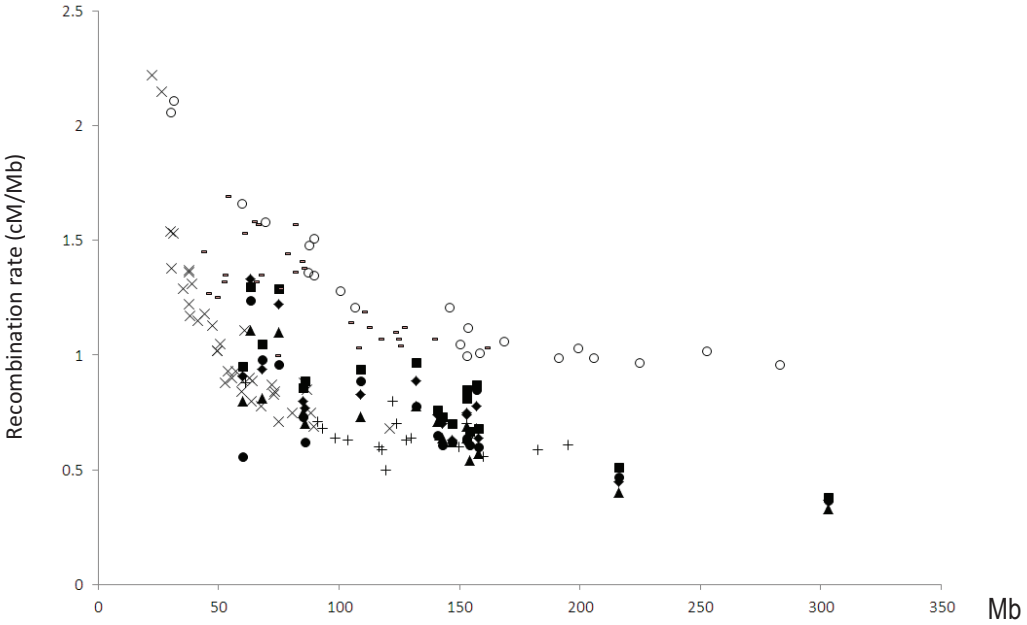


Figure 5.1 - Distribution of the recombination rate according to the physical chromosome size. Results are given for the four pig pedigrees in black (squares for ILL, diamond for UIUC, triangles for USDA and circles for ROS), and for other mammals in grey (cross for dog, circle for human, plus for mice and dash for cattle).

The distribution of recombination rates was not constant along the chromosomes with high recombination rates mostly concentrated around the telomeres (figure 5.2 and figure 5.3).

On SSC9, the large gap observed is due to the absence of SNPs that could be reliably included for the four pedigrees in the genetic maps. The distribution of the recombination rates plotted against the physical distance to the closest chromosome end confirmed that high recombination rates tend to cluster around the chromosome ends, irrespective of the position of the centromere (figure 5.4). For the sex-averaged map, the correlation between the recombination rate and the physical distance to the closest chromosome end was estimated to be -0.48 (p-value< 0.0001), and correlations for separate male and females maps were identical.

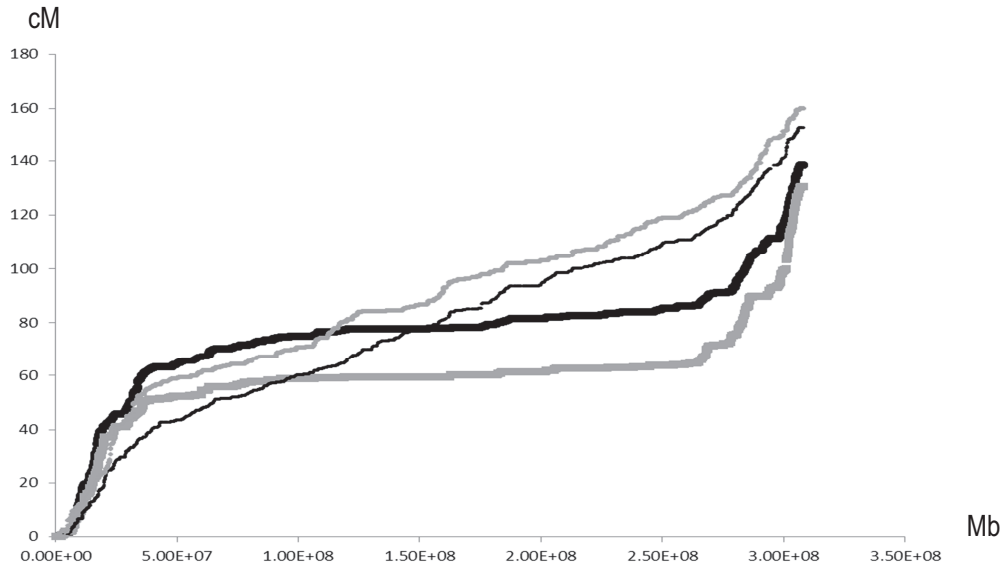


Figure 5.2 - Physical and genetic positions of the SNPs mapped on SSC1. The ILL and UIUC positions are given in black and grey respectively, thickest dots representing the female maps.

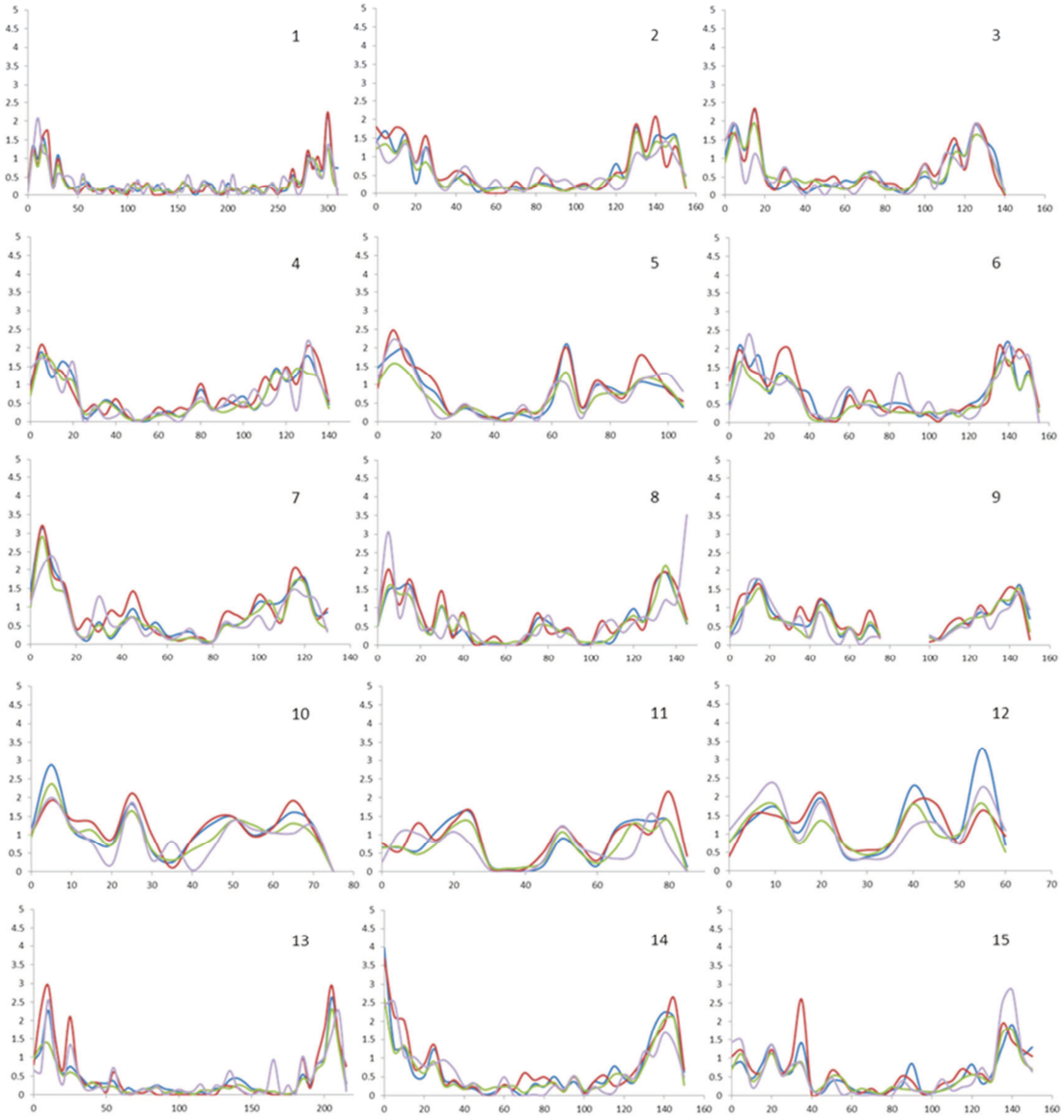


Figure 5.3 (part 1)

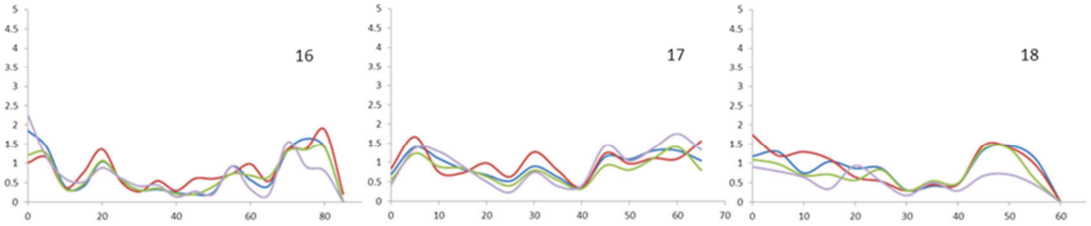


Figure 5.3 (part 2) - Recombination rate for the four pedigrees. Recombination rates were calculated for bins of 5 Mb. On the x-axis, the genomic position is given in million base pairs. On the y-axis, the recombination rate is given in cM/Mb. Results for ILL, UIUC, USDA and ROS pedigrees are given in blue, red, green and grey respectively.

Correlation of recombination with sequence parameters

Correlations between recombination rate and various sequence parameters (GC content, repetitive elements content and short sequences) have previously been observed in human (KONG *et al.* 2002), chicken (GROENEN *et al.* 2009), dog (WONG *et al.* 2010) and mouse (SHIFMAN *et al.* 2006). The occurrence of these sequence parameters was calculated within bins of 1 Mb and the correlations with the recombination rates were estimated. With the sex-average map, all sequence features were highly significantly correlated with the recombination rate (p-value <0.05). However, the level of the correlations was lower for LINEs and LTRs, with Pearson correlation coefficients of -0.05 and 0.06 respectively. The comparison of the sequence composition of recombination jungles and deserts also highlights this link between the occurrence of specific sequence features and recombination rate (table 5.4). Recombination jungles were enriched in specific GC rich motifs as compared to the deserts. The largest difference was observed for the CCCCACCCC sequence, this sequence being almost three times more frequent in recombination jungles than in deserts.

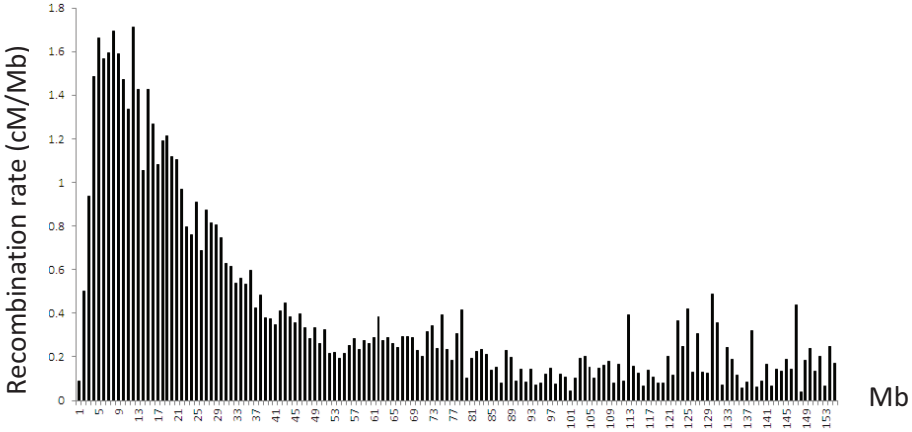


Figure 5.4 - Pig recombination rate distribution according to the distance (in Mb) to the closest chromosome end.

Male and female recombination rates were also analysed separately and large differences were observed. The correlation of the recombination rate with GC content was higher in females (0.44) than in males (0.15) (table 5.4). In agreement with this is the observation that in females recombination is higher only when the GC content of the region is higher than 0.40 whereas it is lower for regions where the GC ratio is smaller than 0.39 (figure 5.5).

Jungle / desert ratios were also highly different between both sexes for SINEs and short sequence motifs. In females, this ratio reached 3.41 for the CTCF consensus sequence (CCNCCNGGNGG), whereas it only reached 1.52 in males.

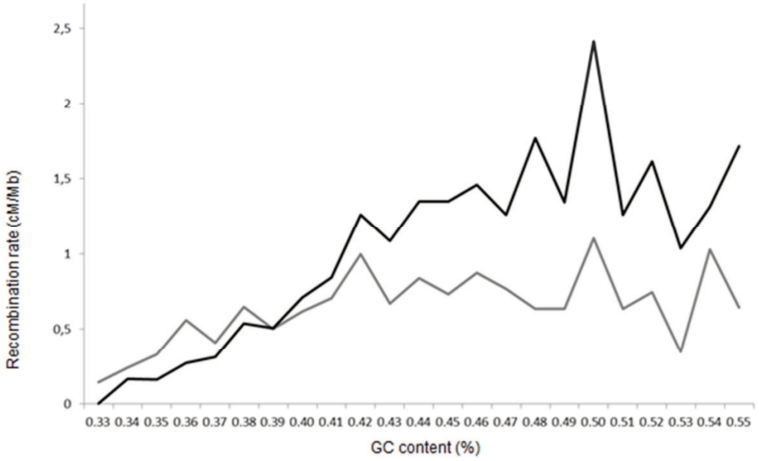


Figure5.5 - Distribution of recombination rate within males and females in relation to the GC content. Black and grey bars represent female and male recombination rates respectively.

5.4 DISCUSSION

Genetic maps

The reliability of a genetic map is of major importance for linkage and genome-wide association analyses (DAW *et al.* 2000). The presented genetic maps were computed for four different pedigrees, with a subset of SNPs being optimal for all of them, finally comprising 38,599 SNPs. The order of the SNPs was the best one that could be obtained for the four pedigrees based on the RH order, position on the sequence assembly and on the recombination analysis. Because only SNPs for which sequence and RH positions were in agreement were included in the analysis and because recombination maps confirmed the *a priori* order, the map presented

in this study is expected to be as accurate as possible with the currently available data.

The recombination map presented in this paper is the densest recombination map ever computed for the porcine genome. The first published genetic maps were mainly based on microsatellites, RFLPs and protein polymorphisms (ARCHIBALD *et al.* 1995 ; MARKLUND *et al.* 1996 ; MIKAWA *et al.* 1999 ; Rohrer et al., 1994). The first genetic map based on SNPs was published recently and comprised only 462 SNPs (VINGBORG *et al.* 2009). Until now, the shortest average marker intervals on a genetic map was reached by the USDA MARC map (ROHRER *et al.* 1996) and was on average 2.23 cM. The large number of SNPs as well as the high number of informative meiosis included in the present analysis enabled the computation of a high-density recombination map of the porcine genome with a major decrease in resolution (around 0.1 cM) compared to previous maps. The total length of the genetic map varied between the four pedigrees, from 1,797 cM to 2,149 cM, which is smaller than the previously published genetic maps. This decrease in the total length of the map can be explained by the lower rate of genotyping errors with SNP chip genotyping as compared to microsatellites or RFLP genotyping. Concerning the map computed with gene-associated SNPs (VINGBORG *et al.* 2009), the sex-averaged genetic maps presented in our study are 15 to 45% shorter, if we take into account only the regions covered in both studies (VINGBORG *et al.* 2009). The same is observed for the sex-specific maps. Female maps are 21 to 33 % shorter in our study, and the two male genetic maps are around 18-19% shorter than the one presented by Vingborg *et al.* Recently, two genetic maps based on the 60k SNP chip have been published for Landrace and Duroc, with similar chromosome lengths as in our study except for SSC1 where a length of 199.8 cM was obtained in Landrace, very different from all the others (GRINDFLEK *et al.* 2011).

Recombination rates

The recombination map of the porcine genome described in this paper, revealed major chromosomal as well as regional differences in recombination rates. The four pedigrees clustered into two different groups, ILL and UIUC having recombination rates close to 0.8 cM/Mb whereas the two other pedigrees had lower recombination rates close to 0.7 cM/Mb. All these values are in the range of previous findings in mammals (from 0.6 cM/Mb in mouse (SHIFMAN *et al.* 2006) to 1.25 cM/Mb in cattle (ARIAS *et al.* 2009). In birds, the observed recombination rate is higher with values of 1.5 cM/Mb in the zebra finch (BACKSTROM *et al.* 2010) up to 2.7 to 3.4 cM/Mb in chicken (ELFERINK *et al.* 2010). Differences in recombination rate within a species have already been described in mice (DUMONT *et al.* 2009) and chicken (ELFERINK *et al.* 2010; GROENEN *et al.* 2009).

Table 5.4 - Correlations between recombination rate and sequence composition in 1 Mb bins.

	Sex-average					Male					Female				
	Corr	p-value	J	D	J/D	Corr	p-value	J	D	J/D	Corr	p-value	J	D	J/D
GC %	0.34	<0.0001	0.45	0.38	1.19	0.15	<0.0001	0.43	0.40	1.09	0.44	<0.0001	0.46	0.38	1.21
Line	-0.05	0.0227	458	492	0.93	0.03	0.1643	477	485	0.98	-0.08	<0.0001	456	494	0.92
Low_complexity	-0.26	<0.0001	131	174	0.75	-0.15	<0.0001	139	165	0.84	-0.31	<0.0001	129	173	0.75
LTR	0.06	0.0037	136	132	1.03	0.07	0.0011	136	133	1.02	-0.01	0.5137	131	135	0.97
Simple_repeat	0.31	<0.0001	179	150	1.19	0.18	<0.0001	174	158	1.10	0.31	<0.0001	179	150	1.19
SINE	0.35	<0.0001	754	506	1.49	0.18	<0.0001	690	587	1.18	0.43	<0.0001	785	492	1.60
CCTCCT	0.34	<0.0001	463	294	1.57	0.17	<0.0001	424	339	1.25	0.41	<0.0001	474	291	1.63
CCTCCCT	0.37	<0.0001	159	82	1.95	0.19	<0.0001	143	103	1.38	0.45	<0.0001	166	81	2.05
CTCTCCC	0.38	<0.0001	137	79	1.72	0.21	<0.0001	125	93	1.34	0.44	<0.0001	141	80	1.76
CCCCCCC	0.41	<0.0001	184	64	2.88	0.21	<0.0001	155	89	1.73	0.49	<0.0001	193	62	3.10
CCCCACCCC	0.36	<0.0001	52	18	2.91	0.17	<0.0001	43	28	1.52	0.45	<0.0001	56	17	3.30
CCNCCNGGNGG	0.25	<0.0001	24	8	2.84	0.09	<0.0001	19	12	1.52	0.33	<0.0001	25	7	3.41
CCNCCNTNNCCNC	0.36	<0.0001	44	17	2.68	0.19	<0.0001	37	24	1.55	0.44	<0.0001	47	16	2.95

Differences in recombination rate observed in this study among the four pedigrees are partly explained by the percentage of phases that could be reconstructed. A lower number of phases could be reconstructed in the two pedigrees were family sizes were small (USDA) or where several mother genotypes were missing (ROS).

Besides these differences among the four pedigrees studied, the recombination rate also varied among chromosomes (table 5.2 and figure 5.1) and within chromosomes (figure 5.2). The distribution of the recombination rate according to the physical size of the chromosomes obtained with the pig was in agreement with the distributions observed in other mammalian species, as well as in birds: shortest chromosomes exhibiting higher recombination rates. This result is in line with the observation of at least one cross-over occurring per meiosis per chromosome (JONES and FRANKLIN 2006). It is noteworthy that for the longest chromosomes in pig, the overall recombination fraction (cM/Mbp) is much lower than for any other mammalian species known to date (figure 5.1).

The distribution of the recombination rate according to the distance to the closest telomere showed that higher recombination rates were mostly found towards the ends of the pig chromosomes. Moreover, the position of the centromere did not seem to influence this distribution: E.g. SSC13 is an acrocentric chromosome and the distribution of the recombination rate along this chromosome was very similar to the distribution along metacentric or submetacentric chromosomes (pig chromosomes 1 to 12 being meta- or submetacentric chromosomes, the others being acrocentric chromosomes (FORD *et al.* 1980)). Other species with acrocentric chromosomes, such as the dog, show a marked increase in recombination fraction for most of the chromosomes away from the start – i.e. centromeric part – of chromosomes (WONG *et al.* 2010). The general absence of this pattern in the acrocentric chromosomes in pig raises questions on how and particularly when the pig chromosomes became acrocentric. The evolution of centromere positions can be highly dynamic, and the current apparent disparity between centromere position and recombination rate may hint at a recent shift of the position of the centromere in several pig chromosomes. Similar to the pig, however, many of the dog chromosomes show a very stark increase in recombination fraction further away from the centromere.

In human and rat, recombination rates were also found higher in the telomeric regions and reduced close to the centre of the chromosomes (JENSEN-SEAMAN *et al.* 2004), but this pattern is not as pronounced as in the pig. This preferential distribution of crossing overs at the chromosomal ends is even more striking in zebra finch with long central regions where the recombination rate remains extremely low (BACKSTROM *et al.* 2010). However, in the zebra finch, and also in chicken, these telomeric regions of exceptionally high recombination compared to the other parts of the chromosomes seem to be much more confined to the extreme edges of the chromosomes, whereas in the pig these distal regions of high

recombination are less pronounced but much bigger in size. In some species, however, this particular distribution of recombination rate along a chromosome is not observed. In mouse, the correlation estimated between recombination rate and the distance to the centre of the chromosome did not differ from the one estimated with respect to the distance to the telomere (JENSEN-SEAMAN *et al.* 2004), which is in agreement with the distribution of the recombination rate estimated from the sex-averaged genetic map (SHIFMAN *et al.* 2006). Similarly, the plot of the genetic map against the physical map of the bovine genome does not show this sigmoid-like pattern that indicates higher recombination rates at the chromosome ends (ARIAS *et al.* 2009).

Recombination and sequence features

In this study, we show that the recombination rate varied with the distance to the closest chromosome end. In human, the GC content was shown to be negatively correlated with the distance to the telomere (DURET and ARNDT 2008), and the porcine genome exhibits the same negative correlation. The GC content has also been shown to be strongly positively correlated with recombination rates in human (BIRDSELL 2002; KONG *et al.* 2002; MARAIS 2003), mice (SHIFMAN *et al.* 2006), chicken (GROENEN *et al.* 2009) and zebra finch (BACKSTROM *et al.* 2010), and this was also confirmed in this study. This seemingly universal positive correlation between GC content and recombination is thought to signify a shared underlying mechanism determining recombination rate (PETES 2001; PETES and MERKER 2002), although it has been proposed that higher GC content can conversely be the result of high recombination rate ((GALTIER *et al.* 2001; Duret and Galtier 2009)).

Mechanisms explaining the direct relationship between GC content and recombination rate identify the presence of certain recognition motifs for DNA binding proteins that have a known function in meiosis or the recombination process directly, such as cohesin and PR domain-containing protein 9. In other mammalian and avian species, high-density linkage maps showed strong correlations between recombination rates and various sequences such as the consensus cohesion binding site; the 7-nucleotide oligomer CCTCCCT (MYERS *et al.* 2005; SHIFMAN *et al.* 2006) and a 13-nucleotide oligomer described in human CCNCCNTNNCCNC (MYERS *et al.* 2008). Recently, it was shown that this 13-nucleotide sequence is recognized in vitro by the human PR domain-containing protein 9, encoded by the *PRDM9* gene (BAUDAT *et al.* 2010). The PR domain-containing protein 9 is known to regulate recombination hotspot activity in human (BERG *et al.* 2010). These GC-rich motifs have been investigated in this study and all of them were overrepresented in recombination jungles and underrepresented in deserts. The sequences CCTCCCT and CCCCACCCC, overrepresented in about 10 % of human hotspots (MYERS *et al.* 2005) were also correlated with higher recombination rates in mouse and chicken, jungle/desert ratios being close to 2 or

higher. The same was observed in this study with a ratio close to 2 or higher (table 5.4).

Sex-differences

In our study, male and female maps were analysed separately for the ILL and UIUC pedigrees. In both designs, female meioses were better sampled than in the two other pedigrees for which dams were not always genotyped or had too few offspring. The ROS and USDA maps are thus closer to male maps which can explain their shorter lengths as compared to the sex-average maps of ILL and UIUC. It should also be noticed that the length of the females maps that are reported here are close to the original MARC map which was based primarily on female meiosis (ROHRER *et al.* 1996).

In most species, the heterogametic sex is expected to have a lower recombination rate than the homogametic sex (HALDANE 1922). This was confirmed in this study at the level of the genome with female maps being longer than male maps by 26% or 46% for ILL and UIUC pedigrees respectively. However, SSC1 stood out with more recombination events described within males than within females. As shown in figure 5.2, females displayed a 160 Mb region with a very low recombination frequency. Vingborg et *al.* found that SSC1 was longer in females than in males, but the 70-100 cM region of SSC1 also displayed higher recombination in males than in females (VINGBORG *et al.* 2009). The higher length of SSC1 in males as compared to females was already observed in previous pig genetic maps (ARCHIBALD *et al.* 1995; BEECKMANN *et al.* 2003; ELLEGREN *et al.* 1994; GUO *et al.* 2009). All these previous maps were based on different crosses involving either Wild Boars, different European breeds or different Asian breeds. The current study also included highly diverse pedigree origins, which makes breed effects therefore unlikely to be the major explanation to this locally low recombination rate. For the ILL pedigree, we observed a small difference between the male and female maps of SSC13 and this was also reported by Guo *et al.* who observed a ratio female to male of 0.98 for this chromosome (GUO *et al.* 2009). In the linkage map computed with gene-associated SNPs, SSC13 was also found rather similar in males and females (VINGBORG *et al.* 2009). For this chromosome, we did not observe such large sex-differences in the distribution of the recombination rates along the chromosome as for SSC1. To better understand this apparent discrepancy in recombination rates between male and female on different chromosomes, we plotted the recombination rate as a function of GC content for male and female separately (figure 5.5). Although in both sexes higher average recombination frequencies are observed for regions exhibiting a higher GC content, this correlation is much stronger in female than in male. In fact, in males a clear lower recombination rate at AT rich regions was observed. However, in females this reduced recombination at AT rich regions is

much stronger than in males, resulting in an overall lower recombination in females in AT rich regions compared to that in males. This explains why on SSC1, the recombination is higher in the males because the 90-250 Mb region is relatively AT rich (GC content of 0.39 compared to the genome average of 0.42). This effect is only clearly seen on SSC1 because the other chromosomes, lack such long regions of low GC content (figure 5.6). A positive correlation between recombination rate in female and GC content had already been reported in human (MEUNIER and DURET 2004), and this was confirmed in the present analysis (table 5.4). Recombination in males appeared to be less sensitive to the frequency of the GC rich motifs and the observed jungle/desert ratios are much higher in females.

The positive relationship between GC content and female recombination does not appear to be universal. Sex-specific GC related recombination rates for instance have been observed in dogs, but appears to be opposite in this species: higher GC content appears to be negatively correlated with female recombination rate (WONG *et al.* 2010). Since that study on dog recombination did not dissect the precise relationship of male and female recombination rate as a function of GC content as done in the present study it is difficult to compare results. However, this opposite relationship in dogs may hint at specific recombination mechanisms that apply to acrocentric vs. metacentric karyotypes, and demonstrates the importance of having detailed recombination maps for many different species for comparative genome biology purposes.

Even if the mechanisms underlying recombination are largely unknown, a number of mechanisms for sex-specific differences have been proposed: difference in time allotted for so called bouquet formation in meiosis (Paigen & Petkov, 2010), difference in the compactness of the chromosomes at pachytene phase of meiosis (Petkov et al., 2007), genomic imprinting (Paigen and Petkov, 2010), or differences in the use of specific recombination-hotspot specific motifs (Kong et al, 2008, Paigen & Petkov, 2010). For instance, it has been shown that different alleles of the *RNF212* gene can have opposite effects on male and female recombination rate (Kong et al. 2008). In mouse, a QTL analysis was carried out to detect regions of the genome underlying recombination rate and the most significant QTLs were observed on chromosome X (DUMONT and PAYSEUR 2011).

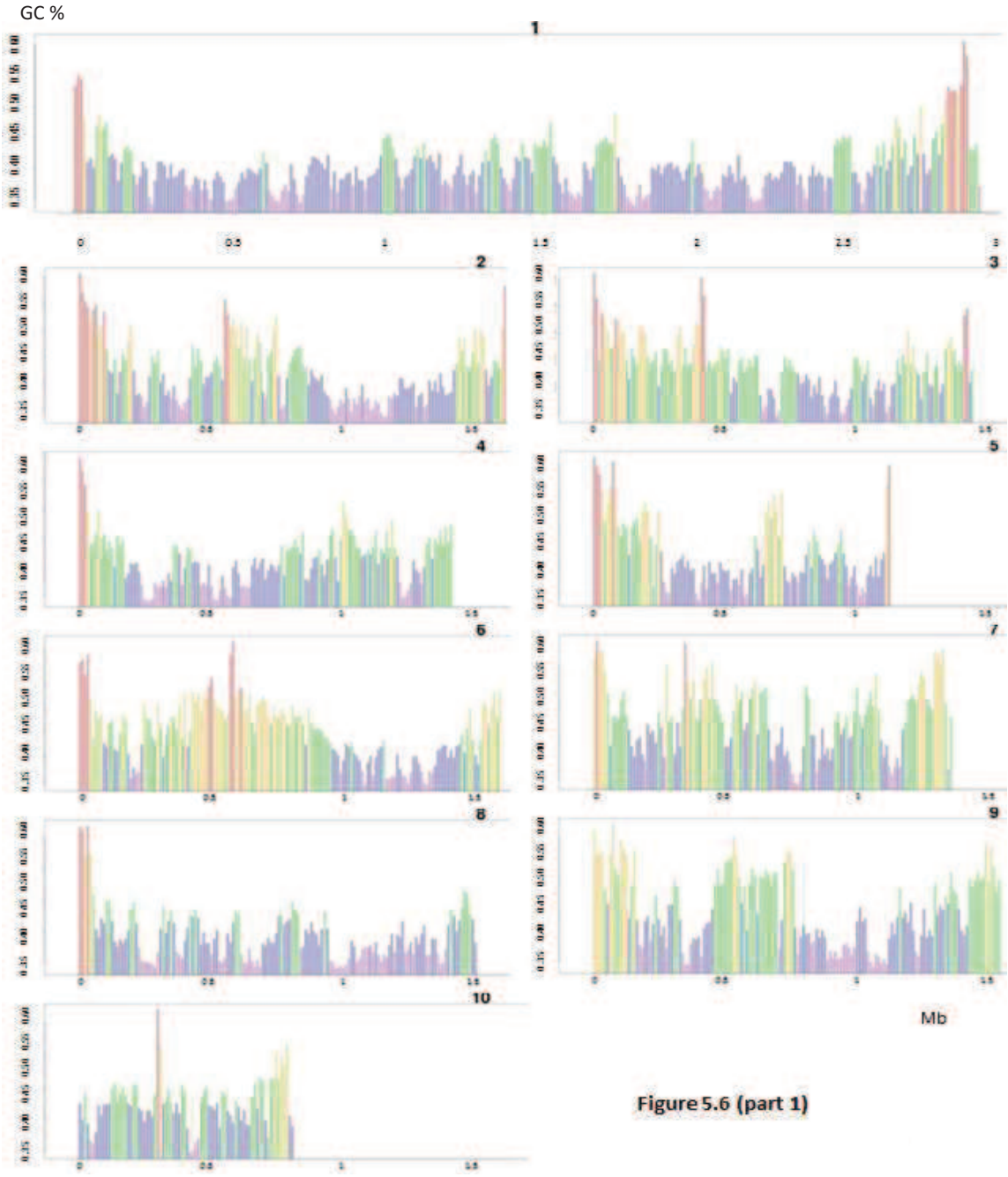


Figure 5.6 (part 1)

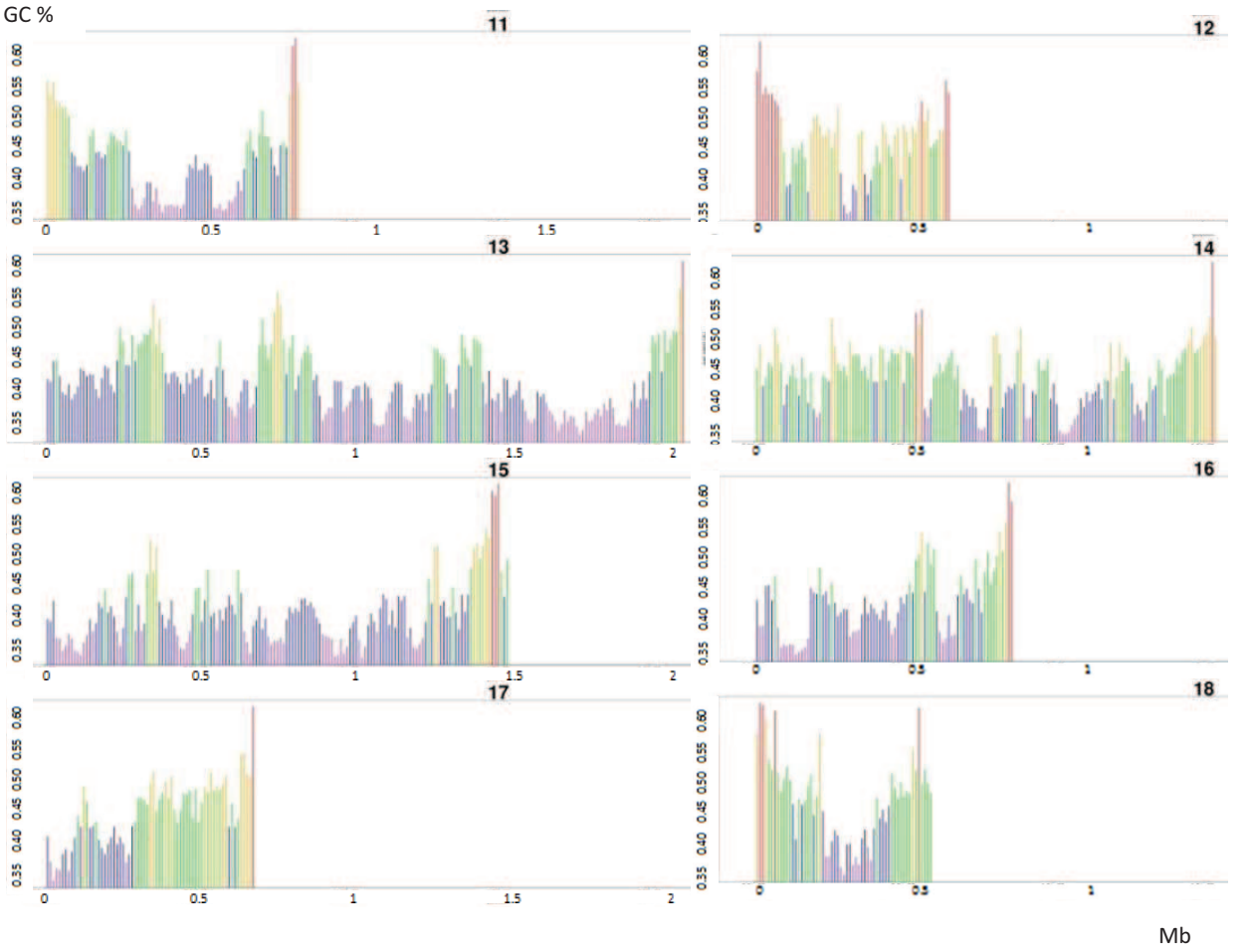


Figure 5.6 (part 2)– GC content distribution along the genome. The GC content is plotted against the physical positions of the bins (in Mb). Colours indicate different levels of GC content, the content increasing from purple, to blue, to green, to orange and finally to red (which colour represents the highest GC contents).

This raises the possibility that chromosomes X and/or Y may be involved in the observed striking difference of recombination rates between males and females. However, the analysis included only males, so no sex-specific QTL could be analysed. This study in mouse indicated that genomic variations on SSCX influenced the recombination rate, but it did not provide further explanation to why females recombine more than males. Finally, in mice, the analysis of meiocytes from XX females, XY males, XY sex-reversed and XO females indicated that recombination patterns depend more on being a male or a female than on the true genotype (LYNN *et al.* 2005). All of these mechanisms may be compatible with the patterns observed in the present paper. In fact, the evolution of recombination and

recombination hotspots seems highly dynamic, and may involve universal (e.g. chromosome compactness at the pachytene phase at meiosis) and species specific mechanisms (e.g. use of gender specific hotspots). The importance of each of these mechanisms will need to be tested using higher density linkage maps in the future.

5.5 CONCLUSION

In this study we presented the first high-density genetic map of the porcine genome, with a resolution substantially higher than previously published maps. This high resolution enabled us to focus on the differences between low and high-recombinant regions of the genome, and on the large differences that we observed between males and females. As expected, at the genome level, female maps were longer than male maps. The unexpected higher recombination rates in males observed on SSC1, could be explained by a large region of low GC content where females showed very low recombination rates. The higher correlation between recombination rate and GC content (as well as GC rich motifs) in females as compared to males was subsequently confirmed at the genome level. Until now, this high correlation between recombination rate in females and GC content has only been reported in human. Further analyses of the mechanisms underlying recombination are needed to identify the molecular mechanism underlying this observed difference. The increased insight in the porcine recombination landscape will help future studies aimed at understanding the evolution of the pig genome and at fine-mapping identified QTLs for economically important traits.

REFERENCES

ARCHIBALD, A. L., C. S. HALEY, J. F. BROWN, S. COUPERWHITE, H. A. MCQUEEN *et al.*, 1995 The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm Genome **6:** 157-175.

ARIAS, J. A., M. KEEHAN, P. FISHER, W. COPPIETERS and R. SPELMAN, 2009 A high density linkage map of the bovine genome. BMC Genet **10:** 18.

BACKSTROM, N., W. FORSTMEIER, H. SCHIELZETH, H. MELLENIUS, K. NAM *et al.*, 2010 The recombination landscape of the zebra finch Taeniopygia guttata genome. Genome research **20:** 485-495.

BAUDAT, F., J. BUARD, C. GREY, A. FLEDEL-ALON, C. OBER *et al.*, 2010 PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science **327:** 836-840.

BEECKMANN, P., J. SCHRÖFFEL JR, G. MOSER, H. BARTENSCHLAGER, G. REINER *et al.*, 2003 Linkage and QTL mapping for Sus scrofa chromosome 1. J Anim Breed Genet **120 (suppl.1):** 1-10.

BERG, I. L., R. NEUMANN, K. W. LAM, S. SARBAJNA, L. ODENTHAL-HESSE *et al.*, 2010 PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nature genetics **42:** 859-863.

BIRDSELL, J. A., 2002 Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Molecular biology and evolution **19:** 1181-1197.

COOP, G., X. WEN, C. OBER, J. K. PRITCHARD and M. PRZEWORSKI, 2008 High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science **319:** 1395-1398.

DAW, E. W., E. A. THOMPSON and E. M. WIJSMAN, 2000 Bias in multipoint linkage analysis arising from map misspecification. Genet Epidemiol **19:** 366-380.

DUMONT, B. L., K. W. BROMAN and B. A. PAYSEUR, 2009 Variation in genomic recombination rates among heterogeneous stock mice. Genetics **182:** 1345-1349.

DUMONT, B. L., and B. A. PAYSEUR, 2011 Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS genetics **7:** e1002116.

DURET, L., and P. F. ARNDT, 2008 The impact of recombination on nucleotide substitutions in the human genome. PLoS genetics **4:** e1000071.

DURET, L., N. GALTIER, 2009 Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genomics Hum. Genet.**10:** 285-311.

ELFERINK, M. G., P. VAN AS, T. VEENENDAAL, R. P. CROOIJMANS and M. A. GROENEN, 2010 Regional differences in recombination hotspots between two chicken populations. BMC Genet **11:** 11.

ELLEGREN, H., B. P. CHOWDHARY, M. FREDHOLM, B. HOYHEIM, M. JOHANSSON *et al.*, 1994 A physically anchored linkage map of pig chromosome 1 uncovers sex- and position-specific recombination rates. Genomics **24:** 342-350.

FINGERLIN, T. E., G. R. ABECASIS and M. BOEHNKE, 2006 Using sex-averaged genetic maps in multipoint linkage analysis when identity-by-descent status is incompletely known. Genet Epidemiol **30:** 384-396.

FORD, C. E., D. L. POLLOCK and I. GUSTAVSSON, 1980 Proceedings of the First International Conference for the Standardisation of Banded Karyotypes of Domestic Animals. University of Reading Reading, England. 2nd-6th August 1976. Hereditas **92:** 145-162.

GALTIER, N., G. PIGANEAU, D. MOUCHIROUD and L. DURET, 2001 GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics **159:** 907-911.

GRINDFLEK, E., S. LIEN, H. HAMLAND, M. H. HANSEN, M. KENT *et al.*, 2011 Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. BMC Genomics **12:** 362.

GROENEN, M. A., P. WAHLBERG, M. FOGLIO, H. H. CHENG, H. J. MEGENS *et al.*, 2009 A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res **19:** 510-519.

GUO, Y., H. MAO, J. REN, X. YAN, Y. DUAN *et al.*, 2009 A linkage map of the porcine genome from a large-scale White Duroc x Erhualian resource population and evaluation of factors affecting recombination rates. animal genetics **40:** 47-52.

HALDANE, J. B., 1922 Sex-ratio and unisexual sterility in hybrid animals. J. Genet. **12:** 101-109.

JENSEN-SEAMAN, M. I., T. S. FUREY, B. A. PAYSEUR, Y. LU, K. M. ROSKIN *et al.*, 2004 Comparative recombination rates in the rat, mouse, and human genomes. Genome research **14:** 528-538.

JONES, G. H., and F. C. FRANKLIN, 2006 Meiotic crossing-over: obligation and interference. Cell **126:** 246-248.

KONG, A., D. F. GUDBJARTSSON, J. SAINZ, G. M. JONSDOTTIR, S. A. GUDJONSSON *et al.*, 2002 A high-resolution recombination map of the human genome. Nat Genet **31:** 241-247.

KONG, A., G. THORLEIFSSON, H. STEFANSSON,G. MASSON, A. HELGASON *et al.*, 2008 Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science. **319:**1398-1401.

KONG, A., G. THORLEIFSSON, D.F. GUDBJARTSSON, G. MASSON, A. SIGURDSSON, *et al.*, 2010 Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010 **467:** 1099-103.

LYNN, A., S. SCHRUMP, J. CHERRY, T. HASSOLD and P. HUNT, 2005 Sex, not genotype, determines recombination levels in mice. American journal of human genetics **77:** 670-675.

MA, J.W., N. IANNUCCELLI, Y. DUAN, W. HUANG, B. GUO, J. RIQUET, L. HUANG, D. MILAN, 2010 Recombinational landscape of porcine X chromosome and individual variation in female meiotic recombination associated with haplotypes of Chinese pigs. BMC Genomics 11: 159.

MARAIS, G., 2003 Biased gene conversion: implications for genome and sex evolution. Trends in genetics : TIG **19:** 330-338.

MARKLUND, L., M. JOHANSSON MOLLER, B. HOYHEIM, W. DAVIES, M. FREDHOLM *et al.*, 1996 A comprehensive linkage map of the pig based on a wild pig-Large White intercross. animal genetics **27:** 255-269.

MEUNIER, J., and L. DURET, 2004 Recombination drives the evolution of GC-content in the human genome. Molecular biology and evolution **21:** 984-990.

MIKAWA, S., T. AKITA, N. HISAMATSU, Y. INAGE, Y. ITO *et al.*, 1999 A linkage map of 243 DNA markers in an intercross of Gottingen miniature and Meishan pigs. animal genetics **30:** 407-417.

MYERS, S., L. BOTTOLO, C. FREEMAN, G. MCVEAN and P. DONNELLY, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science **310:** 321-324.

MYERS, S., C. FREEMAN, A. AUTON, P. DONNELLY and G. MCVEAN, 2008 A common sequence motif associated with recombination hot spots and genome instability in humans. Nature genetics **40:** 1124-1129.

PAIGEN, K., P. PETKOV, 2010 Mammalian recombination hot spots: properties, control and evolution. Nat Rev Genet. **11:** 221-33.

PETES, T. D., 2001 Meiotic recombination hot spots and cold spots. Nature reviews. Genetics **2:** 360-369.

PETES, T. D., and J. D. MERKER, 2002 Context dependence of meiotic recombination hotspots in yeast: the relationship between recombination activity of a reporter construct and base composition. Genetics **162:** 2049-2052.

PETKOV, P. M., K.W. BROMAN, J.P. SZATKIEWICZ, K. PAIGEN, 2007 Crossover interference underlies sex differences in recombination rates. Trends Genet. **23:** 539-42.

RAMOS, A. M., R. P. CROOIJMANS, N. A. AFFARA, A. J. AMARAL, A. L. ARCHIBALD *et al.*, 2009 Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One **4:** e6524.

ROHRER, G.A., L.J. ALEXANDER, J.W. KEELE, T.P. SMITH and C.W. BEATTIE, 1994 A microsatellite linkage map of the porcine genome. Genetics **136:**231-245.

ROHRER, G. A., L. J. ALEXANDER, Z. HU, T. P. SMITH, J. W. KEELE *et al.*, 1996 A comprehensive map of the porcine genome. Genome Res **6:** 371-391.

SHIFMAN, S., J. T. BELL, R. R. COPLEY, M. S. TAYLOR, R. W. WILLIAMS *et al.*, 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol **4:** e395.

VINGBORG, R. K., V. R. GREGERSEN, B. ZHAN, F. PANITZ, A. HOJ *et al.*, 2009 A robust linkage map of the porcine autosomes based on gene-associated SNPs. BMC Genomics **10:** 134.

WONG, A. K., A. L. RUHE, B. L. DUMONT, K. R. ROBERTSON, G. GUERRERO *et al.*, 2010 A comprehensive linkage map of the dog genome. Genetics **184:** 595-605.

6

General Discussion

6.1 INTRODUCTION

The possibility of obtaining information directly from DNA has revolutionised genetics, leading not only to the detection of causal polymorphisms underlying human diseases, but also to improvements in livestock breeds and animal products. Many QTL detection programs have been set up in livestock since 1994. These studies aimed at identifying regions of the genome that underlie variations in economically important traits, such as reproduction, fatness, growth, and disease resistance in swine. Thousands of QTLs have been detected in pigs, but only one of them was fine-mapped to a causal polymorphism (VAN LAERE *et al.* 2003). Despite tremendous efforts over the last few years, the vast majority of QTLs are still characterised by large intervals and cannot be used in selection.

The aim of this thesis was to consider the different limitations of QTL fine-mapping: the number of individuals, marker density, and the ability to detect complex QTLs (chapter 1). To carry out this study, we took advantage of different intercrosses (F2 or marker-assisted backcross design) to confirm and fine-map QTLs underlying growth and fatness traits segregating on pig chromosomes 2, 4, and 6 and understand their mode of segregation. The number of animals was substantially increased by combining two F2 pedigrees based on European White breeds (Large White and Landrace) and Meishan. Consequently, the power of QTL detection analysis was higher than in previous independent studies, which resulted in a slight decrease in QTL intervals as well as the detection of new QTLs (chapter 2). This increase in power also let us consider more complex models, such as those accounting for two linked loci underlying the same trait but which were previously detected as a single locus (chapter 2), and detect a QTL genetically linked to an imprinted major gene (chapter 3). Finally, the development of a marker-assisted backcross design enabled the detection of epistatic interactions (chapter 4).

With the development of the commercial 60K SNP chip in 2009, a new milestone was achieved for the scientific community involved in pig genetics. In order to implement this SNP information in QTL mapping and fine-mapping, a high density linkage map was developed based on 60K SNP chip data (chapter 5). Until now, this high-density SNP chip has been used mainly to detect QTL through association studies. The number of such analyses is increasing in pigs, as well as other livestock species. The availability of these new genotyping tools makes the development of genomic selection possible when the selection strategy does not rely on QTL mapping.

In this last chapter, I first discuss the necessity to carry out QTL fine-mapping in pigs in the era of genomic selection, and then the different fine-mapping strategies that can be carried out with current knowledge and available

materials. Finally, I discuss which improvements are still needed to successfully fine-map QTLs.

6.2 IS QTL FINE-MAPPING STILL NECESSARY IN PIGS?

Until now, genetic progress in pigs and other livestock species was obtained through traditional selection based on pedigrees and phenotypic measurements using the BLUP method with an animal model. Over the last few years, the use of molecular genetics in animal breeding has increased significantly, and substantial improvements in breeding value estimates have been obtained in dairy cattle (GUILLAUME *et al.* 2008). Some breeding companies have already implemented marker-assisted selection (MAS) in their pig breeding programs, using either the identified markers (listed in table 1.2) or sets of markers developed for internal use (PLASTOW *et al.* 2003). In 2001, Meuwissen *et al.* set up the basis for genomic selection: the rate of genetic gain can be significantly increased if genomic values are predicted from a dense genome-wide set of markers (MEUWISSEN *et al.* 2001). With genome-wide high-density SNP chips, all QTLs underlying a trait are expected to be in LD with at least one marker of the chip. Therefore, genomic selection is theoretically insensitive to the identification of causal polymorphisms as all of the QTLs will be captured by the SNPs. When using genomic selection, further fine-mapping of the QTLs is not needed. In this first part, I will discuss whether genomic selection can play a major part in pig breeding and whether MAS is convenient.

The basic scheme of genomic selection is that SNP effects are estimated within a reference population that is phenotyped and genotyped, followed by a determination of the genetic estimated breeding values (GEBVs) in a validation population (with individuals genotyped but not phenotyped) as the sum of all SNP effects. Thus, genomic selection relies on two major assumptions: (1) the phase between the causal mutation and the SNP markers in the reference and validation populations is persistent and (2) the SNP effects are additive. With the current density of markers available in livestock species (50-60,000), the first assumption holds true given the extent of LD within commercial cattle breeds (MCKAY *et al.* 2007) and pig breeds (AMARAL *et al.* 2008). However, much higher SNP densities are required to guarantee that this assumption still holds true across breeds. However, the second assumption is still controversial. On one hand, epistatic interactions exist and influence phenotypic variation (chapter 4), but on the other hand, interactions between genes do not generate important variations at the variance level (HILL *et al.* 2008).

In order to be as accurate as possible, SNP effects need to be estimated in a large reference population, the size of which (usually thousands of individuals) directly influences the reliability of genomic selection (VANRADEN *et al.* 2009). Such large reference populations already exist for dairy cattle and comprise

thousands of bulls that were progeny-tested in the past. In contrast, commercial pigs originate from multi-stage crosses involving, on average, three or four different breeds in order to benefit from additive genetic effects and heterosis. A key point to investigate before starting genomic selection in pigs is how to set up the reference population: pure bred or crossbred individuals? The first results of genomic selection applied in pigs were presented recently, obtained with a reference population made up of 500 Piétrain boars for growth traits, with accuracies ranging from 0.45 to 0.50, which are lower than those usually obtained in dairy cattle (BENNEWITZ *et al.* 2011); for a reference population comprising 3,576 bulls, the average accuracy for production traits is around 0.7 (VANRADEN *et al.* 2009). This difference in accuracy illustrates a major limitation of the use of pure breeds in a pig reference population; each purebred line does not comprise enough individuals to guarantee the accuracy of the estimates of SNP effects. The second limitation is that a reference population would be required for each purebred line. One of the possibilities to overcome the limitation in population size is to combine reference populations. In dairy cattle, it was demonstrated recently that combining reference populations from related populations increases the reliability of genomic values, particularly for production traits (BRONDUM *et al.* 2011). The relatedness between dairy cattle populations is very high with few bulls that have been used widely across the countries. The relatedness between pig populations is different from that of dairy cattle, and combinations of reference populations will still need to be considered in this species. An alternative approach for setting up a reference population comprising enough individuals is the use of crossbred animals. Until now, this second strategy has only been considered through simulations, and results indicate that, assuming high marker density (40 markers per cM), estimating SNP effects in crossbred individuals and using them to estimate the genomic values of individuals belonging to all the pure breeds from which the crossbred population derives is possible. However, whatever the simulated crossbred population, accuracy was always lower than that of a reference population made up of pure breeds (TOOSI *et al.* 2010).

In addition to the question about the reference population, whether genomic selection will benefit in pig breeding is still unclear. In dairy cattle, benefits are clear, with shorter generation intervals and the end of bull progeny-testing. In contrast, gains are less evident in other livestock species for which the generation intervals are already short, such as pig and poultry (approximately 2.5 years and 1 year respectively), and the genotyping costs are still relatively high compared to the economic value of the animals and actual selection costs. Nevertheless, benefits from genomic selection are obtained at the level of the accuracy of the genetic values estimated for some traits. In pigs, meat quality traits are currently recorded based on relatives, and the accuracy is rather limited. Genomic selection will increase the efficiency of the selection of such traits. In addition, genomic selection will allow the selection of traits that are

not recorded routinely by breeding companies (e.g., resistance to disease and behavioural traits) or traits with low heritability, such as reproductive traits.
In the case that genomic selection cannot be applied, QTLs need to be fine-mapped in order to implement MAS. The expectations from MAS are similar to some from genomic selection: traits with low heritability and traits that are difficult to measure will be better selected for. The major difference with genomic selection is that not all of the QTLs are accounted for; GEBVs are computed from fewer loci for which the effects on the trait of interest have been detected and then validated in the population. Therefore, if QTLs with moderate to large effects (additive or not) underlie the trait, MAS will be enough to improve its selection. The efficiency of MAS will be even higher if true causal polymorphisms are used, because their effect is higher than the effect estimated from markers in high LD with the causal variant (OLSEN *et al.* 2007). Furthermore, non-additive effects are better taken into account by MAS. With genomic selection, GEBVs are determined as the sum of all SNP effects, even if we know that more non-additive effects exist. For example, the *IGF2* substitution only has an effect when the mutated allele is inherited from the father, which is information that can be implemented in MAS but not genomic selection. Nevertheless, if the trait of interest is due to a large number of loci, each one explaining a small proportion of the genetic variance, to be effective MAS will require that most of these loci be identified, which, for the time-being, is a difficult aim to achieve. In such a situation, genomic selection is more straightforward as it does not require causal loci to be identified.
Currently, no argument has been made in favour of the implementation of genomic selection instead of MAS in pig breeding, so there is still an interest in fine-mapping QTLs. For breeding purposes, the development of MAS was necessary because identifying the true causative variants and finding functional proof of their causality have been difficult. However, the selection on true causal variants (through gene-assisted selection) can be applied directly to all breeds and lines of a species provided that the frequency of the favourable allele is low, whereas selection based on haplotypes requires that haplotype effects are re-estimated regularly due to LD changing over time. For example, *Phi* and *Pgd* have been proposed as flanking markers used to select on the halothane gene, but depending on the breed, the alleles of these markers do not always show preferential association with alternate *RYR1* alleles (Houde *et al.*, 1993).
The identification of causal variants is also essential when dealing with the introgression of favourable haplotypes within a breed. An example of introgression in pigs was the creation of the Piétrain ReHal breed. Most of the Piétrain pigs are homozygous for the unfavourable allele (n) in the halothane gene (FUJII *et al.* 1991); thus, successive backcrosses were done to introgress the favourable halothane allele (N) from Large White into Piétrain in order to benefit from the good growth and leanness characteristics of the Piétrain breed

(HANSET *et al.* 1995). Similarly, in plants, many introgressions of advantageous alleles within different varieties have been done. However, the results did not always meet plant breeders' expectations because the introgressed QTLs' effects sometimes disappeared (HOSPITAL 2009). In the future, animal breeders may have some interest in introgressing particular favourable QTL alleles in breeds in which the favourable alleles do not segregate. To avoid the situation observed in plant breeding, a deep understanding of the QTL effect is necessary, particularly whether it is additive, whether any interactions occur with other loci, and whether the QTL affects only one trait or if it has any pleiotropic effects. In chapter 4, we showed that, similar to plants, epistatic interactions can mask QTL effects in pigs. From an economic point of view, the introgression of a QTL displaying undetected epistatic interactions would be disastrous for pig breeders given that, for example, the creation of the Piétrain ReHal line required more than seven backcross generations.

Finally, the pig has been commonly used as a model organism for human diseases, and for these particular experiments the identification of causal polymorphisms is an absolute necessity for understanding biological pathways underlying the diseases, which is an essential step in the process of drug elaboration.

Therefore, even though the interest in genomic selection is increasing in animal breeding, a need to fine-map QTLs still exists for breeding and research purposes. Over the last few years, major changes have occurred in molecular biology with new genotyping and sequencing tools, and this has modified QTL fine-mapping strategies.

6.3 FINE-MAPPING QTLs WITH NEW MOLECULAR TOOLS

6.3.1 Experimental designs and linkage analysis

When this thesis was initiated, various limitations to QTL fine-mapping had to be overcome, specifically the number of recombinant chromosomes to study, marker density, and the ability to consider QTLs displaying different modes of segregation. Experimental F2 or BC designs previously set up to detect QTLs have formed major resources for genetic analyses with hundreds to thousands of individuals with known ancestors, DNA samples, and several recorded phenotypes. Analyses of these designs resulted in the identification of thousands of QTLs. In this thesis, I described how to fine-map QTLs by studying these existing pedigrees as well as new designs. Because no high-density SNP chip was available at that time, this work was carried out with sets of microsatellite markers. Within such designs, LD extends over long distances as only one generation of recombination is available, abolishing the need to carry out linkage studies with marker densities higher than one every 10 cM. The

results of the studies presented in this thesis show that the benefits of combining large F2 pedigrees are rather limited for the purpose of fine-mapping, with final QTL intervals remaining around 10-20 cM (chapter 2). Nevertheless, this strategy proved to be efficient for de novo QTL detection and the investigation of more complex models (chapters 2, 3, and 4).

Meta-analysis of published QTL results was proposed as an alternative method for benefitting from already existing QTL experimental designs. The general methodology was proposed by Goffinet and Gerber (2000) and two applications in livestock have been published: in pigs for QTLs segregating on SSC4 and influencing various traits (Silva et al. 2011), and in dairy cattle for milk production traits and mainly focused on BTA6 (Khatkar et al. 2004). The meta-analysis conducted for SSC4 in swine was based on results from 25 independent studies. Confidence intervals of meta-QTL were generally reduced compared to single initial QTLs, but no meta-QTL effects were estimated because the diversity of breeds would have made their interpretation difficult and the authors estimated that the number of studies was too few to estimate them confidently. In addition, the number of studies that could be part of a meta-analysis was limited because of the low number of traits that have been recorded equally in different QTL experiments. This limit due to the definition and precision of phenotypes will be discussed later. The estimation of meta-QTL confidence intervals depends on the published results that are included in the meta-analysis (Goffinet and Gerber 2000). For example, in the SSC4 meta-analysis, Silva et al. (2011) accounted for a QTL underlying the ADG detected in the INRA pedigree. However, we demonstrated that what was previously detected as a unique QTL is more likely due to two closely linked QTLs (chapter 2), which probably modifies their results.

In addition to combining pedigrees (in a joint or meta-analysis), the re-analysis of existing designs is tempting in order to test different hypotheses related to, for example, the number of QTLs and their mode of segregation. However, we demonstrated that existing QTL designs are not always appropriate for testing some hypotheses. For example, the pedigree structure and frequency of the IGF2 mutation influence the ability to detect imprinting (chapter 3), whereas another QTL segregates 40 cM from the imprinted gene.

Although traditional QTL designs have been efficient for the identification of QTL, re-analysis of the data (either through combination or meta-analysis) does not significantly decrease the QTL confidence intervals. Nevertheless, one can wonder whether more interesting results would be obtained from these designs if new molecular developments are considered in the re-analysis.

6.3.2 LDLA mapping

In 2009, a high-density chip became available for pig geneticists (RAMOS et al. 2009). With more than 60,000 SNPs distributed across the genome, this new molecular tool enables LD to be taken into account. At first look, using 60,000 SNPs to analyse an F2 or BC experimental design seems like overkill because of the very long LD blocks segregating within these pedigrees. However, LD extends over shorter distances when only the founders of the pedigrees are considered. These individuals were usually sampled from the same original population, and most of them should share IBD haplotypes. With a high marker density, this LD between the genotyped markers and QTL can be implemented in LDLA mapping. This method benefits from the robustness of linkage analysis, as well as the high resolution of association analyses (Meuwissen et al. 2002). In 2009, Heuven et al. observed that LDLA mapping with an average inter-marker distance of 6 cM was efficient to improve the significance of QTLs, but not the precision to localise QTLs (Heuven et al. 2009). Until now, only two LDLA mapping studies have been carried out in pigs with dense marker maps (with either 3,000 or 60,000 SNPs) (Grindflek et al. 2011a, 2011b). In the analysis using the 60K chip, Grindflek et al. (2011b) were able to identify regions underlying androstenone in fat that extended over 3.6 Mb on average (from 0.1 to 28 Mb). The size of these intervals is considerably smaller than intervals obtained for the same population with a panel of 3,000 SNPs in which QTL intervals varied from 2 to 41 cM with an average size of 15 cM. The ability to decrease QTL confidence intervals with LDLA mapping depends on the level of LD existing among the animals that are analysed. Prior to LDLA mapping, the level of LD among the founders must be checked.

If a first analysis of the founders of an F2 or BC design, such as the ones analysed in this thesis, shows that the LD among founders extends over long distances, no benefit exists in applying LDLA mapping in these pedigrees because the QTL intervals will remain too large. However, if this first analysis shows that the extent of LD among the founders is small, I would advise carrying out such an analysis provided that the animals and their phenotypes already exist. Nevertheless, genotyping all of the hundreds of individuals of a pedigree would be too expensive. An alternative strategy is to restrict the genotyping to F0 and F1 individuals, and then to impute the genotypes of the F2, as performed in chapter 3 for the IGF2 mutation. From the genotypes of the F0 and F1 individuals, the identification of haplotypes segregating within the F1 animals is quite straightforward, and various software have already been developed to carry out genotype imputations (Marchini et al. 2006). To reliably impute genotypes, a well-established marker order is essential. The pig sequence is now available and the relative order of the microsatellites that were previously genotyped on the analysed pedigrees using the SNPs on the chip can be assessed by in silico mapping. However, to avoid any confusion due

to assembly errors, I recommend combining this in silico information with RH and recombination mapping when such data are available (chapter 5).

6.3.3 Populations

To benefit from LD when mapping and fine-mapping QTLs, analysing populations through GWAS or LDLA is more relevant. However, as presented above, carrying out LDLA mapping requires that all parents and offspring be known unambiguously and sampled, which is not always possible, and that they are all genotyped, which is more expensive (unless imputations can be implemented) than GWAS studies in which parents are not necessary. LDLA mapping also requires more computing time than GWAS (Goddard and Hayes 2009). These constraints can explain why GWAS have been more popular than LDLA.

When designing a GWAS, the choice of the animals to be genotyped is highly important; the number of individuals, which directly influences the power of the analysis, is less limiting than for linkage studies. Additional animals can be added in the experiment as long as they are genotyped and phenotyped, as no family structure is required for these analyses. However, the origin of the individuals is a major issue: which population(s) needs to be sampled and whether related individuals are better than unrelated ones. The choice of individuals is also driven by the characteristics of the available genome-wide SNP chips. In pigs, the current 60K SNP chip is more favourable for GWAS based on European breeds rather than the ones based on Asian breeds. This bias exists because of the different levels of LD within breeds and because the current chip was designed based on results obtained in European breeds and Wild Boar. In European breeds, LD extends over 400 kb haploblocks, whereas the size of haploblocks does not exceed 10 kb in Chinese breeds (AMARAL et al. 2008). With such an extent of LD, Amaral et al. estimated that approximately 500,000 SNPs would be necessary to carry out a whole genome analysis in Asian breeds, which explains why the current 60K SNP chip is not dense enough for analyses of Asian breeds. The bias due to the SNP chip design will be discussed later. Until now, five porcine GWAS have been published. These studies were carried out with European breeds and concerned boar taint (through the analysis of androstenone, skatole, or other sex steroids) (Duijvesteijn et al. 2010; Grindflek et al. 2011a; Ramos et al. 2011), body composition and structural soundness (Fan et al. 2011), and lifetime reproductive traits (Onteru et al. 2011).

Concerning the relationships between individuals, sampling related animals with known pedigrees (Grindflek et al. 2011a,b) is relevant because it is possible to take the known relationships into account in the statistical model, and knowing the relationships enables the combination of GWAS and LDLA mapping. If the structure of the population is not taken into account, GWAS can

end with false positive associations (Pritchard and Przeworski 2001). When applying both GWAS and LDLA, confidence intervals can be obtained from LDLA mapping, and the likelihood profiles obtained from LDLA mapping are smoother than those obtained from GWAS so QTLs are better localised (Grindflek et al. 2011a). Moreover, simulation studies carried out in humans indicate that including related individuals in GWAS only slightly reduces the power of the analysis, from about 1.00 to 0.93 when the coefficient of relationship increases from 1/8 to ½ (Visscher et al. 2008). Sampling related individuals is not only relevant for the reasons mentioned above, but also for further steps in fine-mapping QTLs, which I will discuss later.

Similar to linkage studies, the effects of the regions detected through GWAS must be confirmed in order to discard any false-positive results. As with LDLA mapping, a reliable SNP order must be established because GWAS results can be confusing if errors remain in the SNP order. For example, Sahana et al. could not determine exactly the number of QTLs underlying calving traits segregating on BTA25; they identified two regions separated by a 7 Mb gap and proposed that these two QTLs could actually be a single QTL with the two observed signals due to assembly errors (Sahana et al. 2011).
GWAS results can be validated by strategies similar to the ones presented for linkage analysis: by analysing independent populations, by combining several GWAS in a single analysis, or by conducting meta-analyses of the results. A major limitation of the combination of linkage analyses is the diversity of markers used in independent studies (chapter 2). Because only a few genome-wide high-density SNP chips exist for each species, the combinations of GWAS will not be hampered by this limitation. Combining several GWAS in a single analysis or meta-analysis is not only useful to validate results, but also increases the power to detect QTLs with smaller effects, similar to linkage studies. Such a meta-analysis has already been carried out in humans with a final analysis comprising up to 70,690 individuals. The analysis identified three loci that are significantly associated with adiposity, one of which was described for the first time (in MSRA); the other two were already described but at lower significance thresholds (Lindgren et al. 2009).
Another possibility for validating results from a GWAS is to re-analyse an experimental pedigree in which a similar QTL was detected. The idea is similar to what was done in chapter 3 with the IGF2 substitution. The most significant SNPs obtained by GWAS can be included as fixed effects in the linkage analysis of the pedigree. If the QTL is no longer detected, the GWAS results are confirmed. However, if the QTL is still segregating, it does not mean that the association was a false-positive.

If the validation of association results in an independent sample of individuals confirms a true effect of a genomic region, failure to reproduce the result does

not necessarily mean that the detected association was a false-positive. Difficulty replicating association results is well documented in human studies (Weitkunat et al. 2010). The validation of GWAS results appears to be affected by similar factors as highlighted in chapter 2 for linkage analyses. First, a QTL can be fixed in other populations, or the minor allele frequency can be low, so that the QTL effect is hardly detectable. Secondly, phenotypes can be considered similar among studies, whereas in reality, they are not. How to overcome this problem of heterogeneous phenotype definitions will be discussed later.

6.3.4 Sequencing and further fine-mapping

To further fine-map the QTL regions detected by GWAS or LDLA mapping, additional polymorphisms are needed so that haplotype-sharing can be applied (chapter 1). Next generation sequencing makes re-sequencing a complete genome possible, allowing the detection of millions of polymorphisms. For the purpose of fine-mapping, the choice of the individuals to sequence is, again, of major importance. To maximise the chance that the causal polymorphism(s) will be sequenced, individuals homozygous for the different QTL genotypes must be retained, which is essential to infer phases.

Before GWAS, the QTL genotype of the individuals was determined based on the results of progeny-testing. This strategy was successful for fine-mapping IGF2, but it would have failed if the sires studied in chapter 4 had been sequenced based on their progeny-testing results; two heterozygous sires would have been characterised as homozygous because of epistatic interactions.

With population analyses, the QTL genotype of individuals is even more difficult to assess. The choice of which individuals to sequence is usually based on the correlation between their haplotype and their phenotype. Individuals with extreme phenotypes and alternate haplotypes in the QTL region will be sequenced preferentially in order to maximise the probability that both QTL alleles are sequenced. If the QTL was validated previously in an F2 or BC design, then the sequencing of sires identified as heterozygous at the QTL is more relevant because their QTL genotypes are more reliable.

Sequencing the selected individuals returns a pool of polymorphisms that enables further fine-mapping of the QTL through haplotype-sharing; all polymorphisms with shared alleles can already be discarded from the list of putative causal variants. An alternative to sequencing complete genomes is to sequence target genomic regions. Recently, a QTL underlying tameness and aggression was re-sequenced for four tame rats and four aggressive rats. The QTL region was captured using custom arrays and next-generation sequencing applied to the retained fragments. Using this approach, the QTL was shown to be due to several variants rather than a unique variant (Albert et al. 2011).

Finally, once a causal variant is identified, or at least characterised by a small haplotype, re-analysis of the same population with this causal variant/haplotype included as a fixed effect in the statistical model will help detect new QTLs whose effects were previously masked by this specific causal variant (chapter 3). For example, in dairy cattle, taking into account the DGAT1 polymorphism in the QTL detection model enabled the detection of other QTLs that influence milk production traits (Bennewitz et al. 2004).

As highlighted in chapter 1, linkage analyses, LDLA, and GWAS returned a number of regions suspected to influence the variation of some traits, but the number of causal polymorphisms remains low. The fine-mapping strategies developed in this part of the thesis aim to speed up the process of identifying these variants. However, major limitations linked to phenotyping, genotyping, and the consideration of non-additive effects still need to be overcome for the optimal use of these strategies.

6.4 ACTUAL LIMITATIONS OF FINE-MAPPING QTLs

6.4.1 Traits

The basis of QTL detection is to compare genotypic and phenotypic data in order to identify genomic regions where the segregation of alleles is similar to the segregation of a given trait. Although major advances have been made in the last few years concerning genotypic data, the same cannot be said for phenotypic data. Major limitations in QTL fine-mapping due to phenotypes are the heterogeneity in the definition of phenotypic traits and the precision in their measurements. The heterogeneity in trait definitions makes combining datasets difficult, if not impossible, for both linkage analysis and GWAS. In chapter 2, we were only able to consider average daily gain from birth to slaughter because no identical intermediate weights were available. For GWAS, the number of different phenotypes recorded per individual is rather low because it is restricted to phenotypes that are routinely recordable and used for selection purposes. Furthermore, trait definitions are not necessarily the same across countries; for example, no homogeneity exists in the definition of die-cuts among countries. In addition to the lack of true common traits among QTL detection studies, the imprecision of phenotype measurements is another key problem. Recently, Barendse compared GWAS results obtained in beef cattle with two independent measurements of subcutaneous fat thickness; the ten and six significant SNPs obtained with alternate measurements had only one SNP in common (Barendse 2011). Therefore, to efficiently compare and

combine results of GWAS and/or linkage studies, standardised protocols must be proposed and applied. Projects dedicated to the standardisation of phenotypic data have been developed in humans (Pirmohamed et al. 2011) and mice (Wakana et al. 2009). For livestock, the International Committee for Animal Recording (ICAR) was created to improve performance recording, particularly through the establishment of definitions and standards for measuring economically important traits. Historically, ICAR has focused on the dairy cattle industry. The development of INTERBULL, which provides international evaluation services, attests that efforts in the development of standards are worthwhile (Philipsson 2011). Concerning trait recording in pig breeding, no standards have been proposed except for welfare traits for which standardisation has been studied within the Welfare Quality® European Project. Recently, the Animal Trait Ontology for Livestock (ATOL) project has been initiated by INRA and Iowa State University to precisely define phenotypic traits measured in all livestock species. If these projects manage to elaborate standards to which experimenters can refer, result and data sharing will become highly efficient.

Concerning the level of phenotype precision, until now most QTL detection programs (through linkage or association analyses) were limited to general traits, such as weight at different ages from which average daily gains are obtained, ultrasonic back fat measurements, and carcass traits (e.g., length, weight, cut weights). However, observed variations, such as back fat thickness, can be due to variations in a large number of underlying finer phenotypes, including enzymes involved in adipogenesis, adipocyte number and size, and lipid metabolism, with all of the different factors involved in these pathways. QTL fine-mapping can become a major gene study if the product of the mutation (its composition, quantity, or concentration) can be directly assessed. For example, to localise the RN gene, glycolytic potential has been preferred as a phenotype rather than the Napole yield, given that the heritability of the glycolytic potential (0.90) is higher than that of the Napole yield (0.78) (Le Roy et al., 1994). To further investigate the precise phenotypes involved in an SSC7 QTL underlying body composition, Demars et al. (2007) analysed homozygous and heterozygous individuals (based on their haplotypes in the region) for the SSC7 QTL for different cellular, histochemical, and metabolic traits, narrowing down the possible candidate genes to a small number. Several other QTL studies have analysed more specific quantitative traits, including different fatty acid and serum lipid concentrations (Sanchez et al. 2007; Uddin et al. 2011), and histological templates (Laenoi et al. 2011).

Specific “omics”, transcriptomics, metabolomics, and proteomics, have been developed to assess these more precise phenotypes. The transcriptome has

already been addressed directly in e-QTL studies. In pigs, e-QTL studies have highlighted genes and gene networks that are involved in meat quality (Ponsuksili et al. 2010) or various loin muscle phenotypes (Steibel et al. 2011). However, until now, no improvement has been obtained through these studies for fine-mapping. A new approach to study the transcriptome was recently developed that uses next-generation sequencing methodology (Wang et al. 2009). A major advantage of this approach, referred to as RNA-seq, does not require prior knowledge about the sequence or gene annotation. With e-QTL studies, only QTLs caused by a modification of the transcript level can be detected. To overcome these limitations, concentrations of metabolites and proteins have been used as new biomarkers for QTL detection. However, the results of all these "omics" analyses highly depend on the stage of development and on the tissues from which RNAs, metabolites, and proteins are sampled. Phenotypes change over time and are not identical from one organ to another, or even from one cell to another. Therefore, biological pathways involved in the phenotypic variations of complex traits in livestock species must be identified not only at the level of the factors involved, but also by taking into account when and where these factors are measured. After these levels of precision are targeted, highly precise phenotypes that improve the identification of the genomic regions underlying the variations will become available.

A variety of projects dedicated to high-throughput phenotyping with homogenous definitions of the different phenotypes have been initiated in mice, rats, dogs, and humans. For example, the Europhenome project set up more than 600 traits defined by standardised procedures in mouse targets, using inbred or knock-out lines (Morgan et al., 2010). The results from these projects will be highly useful for studies in livestock for which such projects are not developed. In order to transpose results from these species to livestock species, preliminary studies will be needed to develop standardised phenotyping procedures that can easily be set up outside of an experimental context.

6.4.2 Limitations of the current SNP chip

Even with the major advances in genomics over the last few years, some limitations still exist and need to be overcome to improve QTL fine-mapping. The limitations of high-density SNP chips are directly linked to their design. Genome-wide high-density chips comprise thousands of SNPs detected in a sample of individuals. For the 60K chip for pigs, SNPs were detected with 158 individuals originating from five breeds: Duroc, Piétrain, Large White, Landrace, and Wild Boar. Among the 549,000 SNPs that have been identified, the ones that were finally retained in the chip were chosen because of their favourable quality criteria and a minor allele frequency higher than 0.15. The average minor allele frequency of the porcine 60K SNP chip was estimated to be 0.274

for the breeds used for SNP identification (Ramos et al. 2009). Thus, the first major limitation of the current porcine 60K chip is its limited power to detect rare variants. By definition, these variants have a low frequency within a population and their frequency is not in the range of the minor allele frequencies of the SNPs selected for the chip. Therefore, the level of LD between these rare causal variants and the markers is too low for significant associations to be readily detectable. The second major limitation identified for the porcine 60K SNP chip is, as mentioned earlier, that the actual SNP density does not match the extent of LD in Asian breeds. Therefore, to improve the detection of rare variants and study QTLs segregating in Asian breeds, it is necessary to re-sequence some Asian pig genomes and select more relevant sets of polymorphisms that can be used in the studied populations.
These limitations due to minor allele frequencies and the adequacy between the density of SNPs and extent of LD of a given population are not only specific for pigs, but for all species. The number of chips available per species is rather low, and usually only one for livestock species, so these SNP chips do not fit with the genetic characteristics of all breeds. In humans, the number of genotyping chips is more important with, for example, seven chips with densities ranging from 300,000 to 4.3 million SNPs commercialised by Illumina (www.illumina.com) and five chips with densities from 10,000 to 906,600 SNPs commercialised by Affymetrix (www.affymetrix.com). The number of chips is likely to increase now that it is possible to customise classical Illumina chips by adding thousands of chosen SNPs, and this “customisation” of chips will constitute a major advance in future livestock studies.

6.4.3 Sequencing and new polymorphisms

Re-sequencing genomes is not only useful for designing appropriate sets of SNPs, but also to consider different families of polymorphisms, particularly the ones that were difficult to consider before now because of technological limitations. In humans, the 1000 genome project (www.1000genomes.org) has been initiated in order to characterise human genetic variations. By re-sequencing thousands of human genomes, most variations will be described (MILLS et al., 2011), as well as the haplotypes in which they are identified. This new catalogue of haplotypes will help impute complete sequences from dense sets of markers and GWAS will include new variants. Until now, for livestock species, genomic variations were studied through HapMap projects, which are based on SNP data; these projects highlighted the relationships among breeds and helped describe breed history in cattle (Bovine HapMap Consortium, 2009), pigs (GROENEN et al., 2010), and sheep (KIJAS et al., 2009). The re-sequencing of thousands of human genomes to identify sequence variations is going to be mimicked in bovines. Re-sequencing projects have also been initiated recently in pigs and should finally gather more than 500 sequenced individuals from a

wide range of breeds and Wild Boars. Such re-sequencing projects are necessary to gain information on genomic variations, including SNPs and other variants. The major other polymorphisms are structural variants, among which copy number variants (CNVs, which are fragments of DNA usually longer than 1 kb and present at variable copy numbers), inversions (which is a modification in the orientation of a DNA segment relative to the rest of the chromosome), and translocations (which is a change in the position of a DNA segment without a change in the total DNA content) can be distinguished.

In chicken, two CNVs have been described as causal variants: a 180-kb duplication for the late-feathering phenotype (Elferink et al. 2008), and a 3.2-kb CNV within the Sox5 gene for the Pea comb phenotype (Wright et al. 2009). In the mouse, several CNVs affecting specific genes (Itlma, Trim12, and Trim34) have been shown to correlate with weight, adiposity, and triglyceride, glucose, or insulin levels (Orozco et al. 2009). Thus, traits of high economic importance in pig production, such as weight and adiposity, might be influenced by CNVs, which strengthens the need to consider these polymorphisms in pigs. By using comparative genomic hybridisation arrays (CGHa), CNVs have been detected on various pig chromosomes (Fadista et al. 2008; Tang et al. 2010). More recently, high-density SNP chips were used to assess CNVs in the cattle genome (Hou et al. 2011; Seroussi et al. 2010). Similarly, in pigs, 49 CNVs have been identified in 13 chromosomes using the 60K SNP chip. However, this number is considered an underestimation of the real number of CNVs in the porcine genome given that the current SNP density of the porcine SNP chip only enables the detection of CNVs longer than 200 kb (Ramayo-Caldas et al. 2010). With CGHa and SNP-based methodology, the number of variants is not determined exactly. With next-generation sequencing technologies, the detection and analysis of CNVs become more precise and a greater variety of structural variants can be detected. The basic idea is that read depth will vary depending on the number of copies (Alkan et al. 2009; Sudmant et al. 2010). Methodologies dedicated to the analysis of CNVs based on sequence data are currently being developed and, thus, within a few years, more will be known about these structural variants and their possible involvement in phenotypic variations.

In addition to re-sequencing, new genotyping tools are being developed that enable, in unique experiments, to obtain genotypes for SNPs and CNVs. In humans, an array comprising more than 906,600 SNPs and more than 946,000 probes for the detection of CNVs is already available, and its use has resulted in the identification of two CNVs associated with Addison's disease and one CNV associated with hip bone size (Bronstad et al. 2011; Liu et al. 2011). As long as the cost of next-generation sequencing remains higher than the cost of genotyping chips, the design combining SNPs and CNVs for livestock species will be a major improvement for investigating the genetic causes of complex traits.

6.4.4 Non-additive effects

In this thesis, we highlighted that non-additive effects can play a major role in QTL detection; the segregation of a major imprinted gene partly influences the detection of another QTL segregating in the same region (chapter 3), and epistatic interactions mask true QTL effects (chapter 4). These conclusions were obtained through traditional linkage analyses applied within experimental designs. As I already mentioned, additional experimental families are not likely to be developed for the purpose of QTL fine-mapping, so new developments are needed to identify non-additive genetic effects within populations. Among these non-additive effects I will discuss the ones that I considered in this thesis: epistatic interactions and epigenetics (through imprinting).

Starting with interactions between loci, I note the on-going debate about how important these effects must be considered in a breeding context. Nevertheless, the identification of these effects remains important for understanding the biological mechanisms underlying phenotypic variations and to set up appropriate MAS and/or introgression programs. The development of a marker-assisted backcross design enabled the identification of epistatic interactions (chapter 4). Although the strategy we set up was successful, it is unlikely to be applied again for other loci because setting up such designs is time-consuming and expensive given the number of generations that are needed. In addition, setting-up such an experimental design requires that at least one of the interacting loci has already been identified and that the alternate alleles involved in the interaction segregate in the new pedigree.
To directly investigate epistatic interactions within a given population, rather than within pedigrees, alternate methodologies must be considered. Without prior information about whether interacting loci influence a trait, no pre-selection of individuals included in a GWAS is required. In 2008, Ma et al. presented the epiSNP computer package that distinguishes between additive and dominance effects, and enables the detection of five different epistatic effects (MA et al. 2008). Using this model with human data, they successfully identified seven regions involved in epistatic interactions that influence total cholesterol or high-density lipoprotein cholesterol (MA et al., 2010). These epistatic effects only reached the significance of suggestive linkage, despite the large sample size (6,431 and 6,078 individuals for total cholesterol and high-density lipoprotein cholesterol, respectively). Studying interactions between loci requires that all possible combinations of alleles are available in the studied population, and that each one is represented by enough individuals so that statistical analysis can be performed, because rare combinations of loci can lead to false-positive epistatic effects (MA et al., 2010). Therefore, considering epistasis in GWAS requires that thousands of individuals are analysed, and this partly explains why, in livestock, no GWAS has yet considered interactions

between loci. To reduce the computation time needed, Kooperberg et al. (2009) applied a two-stage procedure through which only SNPs with significant associations are tested for interactions, and they identified 22 interactions that potentially underlie Crohn's disease. With this procedure, interacting loci with opposite effects can be missed by the selection step because they will not be identified individually. However, this strategy enables the ability to check whether a locus that influences a trait is involved in genetic interactions, which is major information prior to setting up selection strategies on this locus.

Epigenetic effects are other non-additive effects that are highly difficult to assess as, by definition, they are not due to sequence variations. The most studied epigenetic marker is methylation. In humans, a DNA methylation microarray for 450,000 CpG sites was recently developed and validated by comparing colorectal cancer cell lines and normal colon cells. Higher methylation was observed in cancer cells compared to normal cells, particularly in the regions known to influence cancer development (SANDOVAL et al. 2011). Such a tool does not exist in livestock species, but next-generation sequencing technologies make it possible to investigate DNA methylation and other epigenetic markers, such as histone modifications (Hirst and Marra 2010). Similar to transriptomic, metabolomic, and proteomic analyses, DNA methylation patterns vary according to the stage of development and tissue, so appropriate sampling requires prior knowledge of the determinism of the trait. Among the few epigenome-wide association studies that have been conducted in humans, one of them identified four variably methylated regions that are associated with variations in body mass index. These regions are within or close to the PM20D1, MMP9, PRKG1, and RFC5 genes, which are known to underlie diabetes and body weight variation (Feinberg et al., 2010). By combining both genomics and methylation analyses, Bell et al. (2010) found that the methylation level varied around the FTO locus with the genotype of the individual. This difference in methylation was due to three SNPs segregating in the risk haplotype (risk for obesity and type 2 diabetes), creating CpGs (BELL et al., 2010). This study confirms the major impact that epigenetics can have on phenotypic variations and that combining genetics and epigenetics can help identify genomic regions responsible for trait variations.

6.5 CONCLUSION

In this thesis, we took advantage of different pig experimental designs to overcome the different limitations of QTL fine-mapping. However, the different strategies considered in this thesis were not successful for fine-mapping the identified QTLs, though they were efficient for *de novo* QTL detection and the analysis QTLs displaying non-additive effects.

Since the development of genome-wide high-density SNP chips, interest in GWAS has increased because such studies do not require new developments of experimental designs and molecular tools are now adapted to population studies. However, the number of causal polymorphisms that have been identified through this new methodology still remains very low. Sequencing complete genomes is likely to help identify causal polymorphisms, but to make it truly effective, further developments in bioinformatics tools are needed. In addition, if causal polymorphisms are identified by sequencing, the functional proof of their effect still remains a weak point. Combining the different "omics" data that can now be obtained with next-generation sequencing technologies will help identify biological pathways involved in the phenotypic variations.

In this thesis, we analysed traditional "general" traits, but in order to improve QTL fine-mapping, efforts must now be made to standardise the definition of phenotypes and to propose standardised procedures of measurements, so reliable meta-analyses and combined GWAS become possible. Finally, combining precise phenotyping (identified in high-throughput phenotyping projects in other species) with "omics" data will result in the identification of causal regions responsible for variations in quantitative traits. This aim requires that additional methodologies be developed to integrate all of these data. High-throughput phenotyping projects need to be well documented so that new relevant phenotypes can be analysed in livestock species. Thus, QTL detection will be transformed progressively into major gene detection.

Finally, for all of the livestock species, use of the new molecular tools (high-density chips and next-generation sequencing) is still lagging behind the use in humans. In humans, tremendous results have already been obtained thanks to the use of these new molecular tools, and I am confident that major improvements can be expected in livestock species in the near future.

REFERENCES

ALBERT, F. W., E. HODGES, J. D. JENSEN, F. BESNIER, Z. XUAN *et al.*, 2011 Targeted resequencing of a genomic region influencing tameness and aggression reveals multiple signals of positive selection. Heredity **107:** 205-214.

ALKAN, C., J. M. KIDD, T. MARQUES-BONET, G. AKSAY, F. ANTONACCI *et al.*, 2009 Personalized copy number and segmental duplication maps using next-generation sequencing. Nature genetics **41:** 1061-1067.

AMARAL, A. J., H. J. MEGENS, R. P. CROOIJMANS, H. C. HEUVEN and M. A. GROENEN, 2008 Linkage disequilibrium decay and haplotype block structure in the pig. Genetics **179:** 569-579.

BARENDSE, W., 2011 The effect of measurement error of phenotypes on genome wide association studies. BMC Genomics **12:** 232.

BELL C.G., S. FINER, C.M. LINDGREN, G.A. WILSON, V.K. RAKYAN *et al.*, 2010 Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the *FTO* type 2 diabetes and obesity susceptibility locus. Plos One **5:** e14040.

BENNEWITZ, J., N. REINSCH, S. PAUL, C. LOOFT, B. KAUPE *et al.*, 2004 The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. J Dairy Sci **87:** 431-442.

BENNEWITZ, J., R. WELLMANN, N. NEUGEBAUER, E. THOLEN and K. WIMMERS, 2011 Results from genomic selection in Piétrain pig breeding., pp. 389 in *62nd EAAP*, Stavanger Norway.

BOVINE HAPMAP CONSORTIUM, 2009 Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science **324:** 528-532.

BRONDUM, R. F., E. RIUS-VILARRASA, I. STRANDEN, G. SU, B. GULDBRANDTSEN *et al.*, 2011 Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. Journal of dairy science **94:** 4700-4707.

BRONSTAD, I., A. S. WOLFF, K. LOVAS, P. M. KNAPPSKOG and E. S. HUSEBYE, 2011 Genome-wide copy number variation (CNV) in patients with autoimmune Addison's disease. BMC medical genetics **12:** 111.

DEMARS, J., J. RIQUET, M. P. SANCHEZ, Y. BILLON, J. F. HOCQUETTE *et al.*, 2007 Metabolic and histochemical characteristics of fat and muscle tissues in homozygous or heterozygous pigs for the body composition QTL located on chromosome 7. Physiol Genomics **30:** 232-241.

DUIJVESTEIJN, N., E. F. KNOL, J. W. MERKS, R. P. CROOIJMANS, M. A. GROENEN *et al.*, 2010 A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC genetics **11:** 42.

ELFERINK, M. G., P. VAN AS, T. VEENENDAAL, R. P. CROOIJMANS and M. A. GROENEN, 2010 Regional differences in recombination hotspots between two chicken populations. BMC Genet **11:** 11.

FADISTA, J., M. NYGAARD, L. E. HOLM, B. THOMSEN and C. BENDIXEN, 2008 A snapshot of CNVs in the pig genome. PLoS ONE **3:** e3916.

FAN, B., S. K. ONTERU, Z. Q. DU, D. J. GARRICK, K. J. STALDER *et al.*, 2011 Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. PLoS ONE **6:** e14726.

FEINBERG, A.P., R.A. IRIZARRY, D. FRADIN, M.J. ARYEE, P. MURAKAMI *et al.*, 2010 Personalizes epigenomic signatures that are stable over time and covary with body mass index. Science Translational Medicine **15:**49-67.

FUJII, J., K. OTSU, F. ZORZATO, S. DE LEON, V. K. KHANNA *et al.*, 1991 Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science **253:** 448-451.

GODDARD, M. E., and B. J. HAYES, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet **10:** 381-391.

GOFFINET, B., and S. GERBER, 2000 Quantitative trait loci: a meta-analysis. Genetics **155:** 463-473.

GROENEN, M.A.M., A. AMARAL, H-J. MEGENS, G. LARSON, A.L. ARCHIBALD *et al.*, 2010 the porcine HapMap project : genome-wide assessment of nucleotide diversity, haplotype diversity and footprints of selection in the pig. Plant and Animal Genoms XVIII Conference, San Diego, W609.

GRINDFLEK, E., S. LIEN, H. HAMLAND, M. H. HANSEN, M. KENT *et al.*, 2011a Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. BMC Genomics **12:** 362.

GRINDFLEK, E., T. H. MEUWISSEN, T. AASMUNDSTAD, H. HAMLAND, M. H. HANSEN *et al.*, 2011b Revealing genetic relationships between compounds affecting boar taint and reproduction in pigs. Journal of Animal Science **89:** 680-692.

GUILLAUME, F., S. FRITZ, D. BOICHARD and T. DRUET, 2008 Short communication: correlations of marker-assisted breeding values with progeny-test breeding values for eight hundred ninety-nine French Holstein bulls. Journal of dairy science **91:** 2520-2522.

HANSET, R., C. DASNOIS, S. SCALAIS, C. MICHAUX and L. GROBET, 1995 Effets de l'introgression dans le génome Piétrain de l'allèle normal au locus de sensibilité à l'halothane. Genet Sel Evol **27:** 77-88.

HEUVEN, H. C., R. H. VAN WIJK, B. DIBBITS, T. A. VAN KAMPEN, E. F. KNOL *et al.*, 2009 Mapping carcass and meat quality QTL on Sus Scrofa chromosome 2 in commercial finishing pigs. Genet Sel Evol **41:** 4.

HILL, W. G., M. E. GODDARD and P. M. VISSCHER, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS genetics **4:** e1000008.

HIRST, M., and M. A. MARRA, 2010 Next generation sequencing based approaches to epigenomics. Briefings in functional genomics **9:** 455-465.

HOSPITAL, F., 2009 Challenges for effective marker-assisted selection in plants. Genetica **136:** 303-310.

HOU, Y., G. E. LIU, D. M. BICKHART, M. F. CARDONE, K. WANG *et al.*, 2011 Genomic characteristics of cattle copy number variations. BMC Genomics **12:** 127.

HOUDE, A., S.A. POMMIER, R. ROY, 1993 Detection of the ryanodine receptor mutation associated with malignant hyperthermia in purebred swine populations. Journal of Animal Science **71:** 1414-1418.

KHATKAR, M. S., P. C. THOMSON, I. TAMMEN and H. W. RAADSMA, 2004 Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genetics, selection, evolution : GSE **36:** 163-190.

KIJAS, J.W., D. TOWNLEY, B.P. DALRYMPLE, M.P. HEATON, J.F. MADDOX et al., 2009 A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. Plos One **4:** e4668.

KOOPEBERG, C., M. LEBLANC, J.Y. DAI and I. RAJAPAKSE, 2009 Structures and assumptions: strategies to harness gene x gene and gene x environment interactions in GWAS. Statistical Science **24:** 472-488.

LAENOI, W., M. J. UDDIN, M. U. CINAR, C. GROSSE-BRINKHAUS, D. TESFAYE *et al.*, 2011 Quantitative trait loci analysis for leg weakness-related traits in a Duroc x Pietrain crossbred population. Genetics, selection, evolution : GSE **43:** 13.

LE ROY, P., W. PRZYBYLSKI, T. BURLOT, C. BAZIN, H. LAGANT and G. MONIN, 1994 Etude des relations entre le potentiel glycolytique du muscle et les caractères de production dans les lignées Laconie et Penshire. Journées de la Recherche Porcine en France **26:** 311-314.

LINDGREN, C. M., I. M. HEID, J. C. RANDALL, C. LAMINA, V. STEINTHORSDOTTIR *et al.*, 2009 Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. PLoS genetics **5:** e1000508.

LIU, S. L., S. F. LEI, F. YANG, X. LI, R. LIU *et al.*, 2011 Copy number variation in CNP267 region may be associated with hip bone size. PLoS ONE **6:** e22035.

MA, L., H.B. RUNESHA, D. DVORKIN, J.R. GARBE, Y. DA, 2008 Parallel and serial computing tools for testing single locus and epistatic SNP effects of quantitative traits in genome-wide association studies. BMC Bioinformatics **9:** 315.

MA,L., H.B. RUNESHA, T. TANAKA, L. FERRUCI, S. BANDINELLI and Y. DA, 2010 Genome-wide association analysis of total cholesterol and high-density lipoprotein cholesterol levels using the Framingham Heart Study data. BMC Medical Genetics **11:** 55.

MARCHINI, J., D. CUTLER, N. PATTERSON, M. STEPHENS, E. ESKIN *et al.*, 2006 A comparison of phasing algorithms for trios and unrelated individuals. American journal of human genetics **78:** 437-450.

MCKAY, S. D., R. D. SCHNABEL, B. M. MURDOCH, L. K. MATUKUMALLI, J. AERTS *et al.*, 2007 Whole genome linkage disequilibrium maps in cattle. BMC genetics **8:** 74.MEUWISSEN, T. H., B. J. HAYES and M. E. GODDARD, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics **157:** 1819-1829.

MEUWISSEN, T. H., A. KARLSEN, S. LIEN, I. OLSAKER and M. E. GODDARD, 2002 Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics **161:** 373-379.

MILLS, R.E., K. WALTER, C. STEWART, R.R. HANDSAKER, K. CHEN *et al.*, 2011 Mapping copy number variation by population scale genome sequencing. Nature **470:** 59-65.

MORGAN, H., T. BECK, A. BLAKE, H. GATES, N. ADAMS *et al.*, 2010 EuroPhenome: a repository for high-throughput mouse phenotyping data. Nucleic Acids Research **38:** D577-D585.

OLSEN, H. G., H. NILSEN, B. HAYES, P. R. BERG, M. SVENDSEN *et al.*, 2007 Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. BMC genetics **8:** 32.

ONTERU, S. K., B. FAN, M. T. NIKKILA, D. J. GARRICK, K. J. STALDER *et al.*, 2011 Whole-genome association analyses for lifetime reproductive traits in the pig. Journal of Animal Science **89:** 988-995.

OROZCO, L. D., S. J. COKUS, A. GHAZALPOUR, L. INGRAM-DRAKE, S. WANG *et al.*, 2009 Copy number variation influences gene expression and metabolic traits in mice. Human molecular genetics **18:** 4118-4129.

PHILIPSSON, J., 2011 Interbull's role in the era of genomics., pp. 78 in *62nd EAAP*, Stavanger, Norway.

PIRMOHAMED, M., G. P. AITHAL, E. BEHR, A. DALY and D. RODEN, 2011 The phenotype standardization project: improving pharmacogenetic studies of serious adverse drug reactions. Clinical pharmacology and therapeutics **89:** 784-785.

PLASTOW, G., S. SASAKI, T. P. YU, N. DEEB, G. PRALL *et al.*, 2003 Practical application of DNA markers for genetic improvement. Proc. 28th Annu. Mtg. Natl. Swine Improve. Fed., Iowa State Univ., Ames**:** 151-154.

PONSUKSILI, S., E. MURANI, M. SCHWERIN, K. SCHELLANDER and K. WIMMERS, 2010 Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. BMC Genomics **11:** 572.

PRITCHARD, J. K., and M. PRZEWORSKI, 2001 Linkage disequilibrium in humans: models and data. American journal of human genetics **69:** 1-14.

RAMAYO-CALDAS, Y., A. CASTELLO, R. N. PENA, E. ALVES, A. MERCADE *et al.*, 2010 Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genomics **11:** 593.

RAMOS, A. M., R. P. CROOIJMANS, N. A. AFFARA, A. J. AMARAL, A. L. ARCHIBALD *et al.*, 2009 Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One **4:** e6524.

RAMOS, A. M., N. DUIJVESTEIJN, E. F. KNOL, J. W. MERKS, H. BOVENHUIS *et al.*, 2011 The distal end of porcine chromosome 6p is involved in the regulation of skatole levels in boars. BMC genetics **12:** 35.

SAHANA, G., B. GULDBRANDTSEN and M. S. LUND, 2011 Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. Journal of dairy science **94:** 479-486.

SANCHEZ, M. P., N. IANNUCCELLI, B. BASSO, J. P. BIDANEL, Y. BILLON *et al.*, 2007 Identification of QTL with effects on intramuscular fat content and fatty acid composition in a Duroc x Large White cross. BMC Genet **8:** 55.

SANDOVAL, J., H. HEYN, S. MORAN, J. SERRA-MUSACH, M.A. PUJANA *et al.*, 2011 Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics **6:** 692-702.

SEROUSSI, E., G. GLICK, A. SHIRAK, E. YAKOBSON, J. I. WELLER *et al.*, 2010 Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. BMC Genomics **11:** 673.

SILVA, K. M., J. W. BASTIAANSEN, E. F. KNOL, J. W. MERKS, P. S. LOPES *et al.*, 2011 Meta-analysis of results from quantitative trait loci mapping studies on pig chromosome 4. Anim Genet **42:** 280-292.

STEIBEL, J. P., R. O. BATES, G. J. ROSA, R. J. TEMPELMAN, V. D. RILINGTON *et al.*, 2011 Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. PLoS ONE **6:** e16766.

SUDMANT, P. H., J. O. KITZMAN, F. ANTONACCI, C. ALKAN, M. MALIG *et al.*, 2010 Diversity of human copy number variation and multicopy genes. Science **330:** 641-646.

TANG, H., F. LI, H. FINLAYSON, S. SMITH, Z. LU *et al.*, 2010 Structural and copy number variation in the pig genome. Plant & Animal Genomes XVIII conference, San Diego, CA.

TOOSI, A., R. L. FERNANDO and J. C. DEKKERS, 2010 Genomic selection in admixed and crossbred populations. Journal of Animal Science **88:** 32-46.

UDDIN, M. J., N. DUY DO, M. U. CINAR, D. TESFAYE, E. THOLEN *et al.*, 2011 Detection of quantitative trait loci affecting serum cholesterol, LDL, HDL, and triglyceride in pigs. BMC genetics **12:** 62.

VAN LAERE, A. S., M. NGUYEN, M. BRAUNSCHWEIG, C. NEZER, C. COLLETTE *et al.*, 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature **425:** 832-836.

VANRADEN, P. M., C. P. VAN TASSELL, G. R. WIGGANS, T. S. SONSTEGARD, R. D. SCHNABEL *et al.*, 2009 Invited review: reliability of genomic predictions for North American Holstein bulls. Journal of dairy science **92:** 16-24.

VISSCHER, P. M., T. ANDREW and D. R. NYHOLT, 2008 Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. European journal of human genetics : EJHG **16:** 387-390.

WAKANA, S., T. SUZUKI, T. FURUSE, K. KOBAYASHI, I. MIURA *et al.*, 2009 Introduction to the Japan Mouse Clinic at the RIKEN BioResource Center. Experimental animals / Japanese Association for Laboratory Animal Science **58:** 443-450.

WANG, Z., M. GERSTEIN and M. SNYDER, 2009 RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics **10:** 57-63.

WEITKUNAT, R., E. KAELIN, G. VUILLAUME and G. KALLISCHNIGG, 2010 Effectiveness of strategies to increase the validity of findings from association studies: size vs. replication. BMC medical research methodology **10:** 47.

WRIGHT, D., H. BOIJE, J.R.S. MEADOWS, B. BED'HOM, D. GOURICHON *et al.*, 2009 Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chicken. PLoS Genet **5:** e1000512.

Summary

Pig domestication started around 10,000 years ago during the Neolithic age, independently in Europe and China, and most current pig breeds originate from these two areas. Among the 560 pig breeds that have been recorded over the world in 2007 by the FAO, only few of them have been intensively selected for production. Domestication and, more recently, pig breeders have relied on naturally occurring mutations to select individuals exhibiting favourable traits related to reproduction, growth, fatness, resistance to diseases and behaviour. In order to identify these mutations underlying the phenotypic variations of these traits, a number of QTL detection programs was set up, and thousands of QTLs have been detected in the 2000s. However, for only a few of them, fine-mapping has resulted in the identification of the causal polymorphism.

In **chapter 1**, the general introduction provides an overview of QTL detection in pig in relation to the molecular tools that are available for pig geneticists and to the different mapping strategies that can be used. Major limitations to QTL fine-mapping in pig (but also valid for other livestock species) concern the number of individuals, the number of informative genetic markers and the ability to detect non-additive QTLs.

To increase the statistical power by increasing the number of individuals, a combined linkage analysis is presented in **chapter 2**. To carry out this work, two pig F2 pedigrees comprising about a thousand individuals each and based on similar breeds (Large White and/or Landrace crossed with Meishan individuals) were combined. Both pedigrees had been developed in the late 1990s at INRA and WUR. Common QTLs segregating on SSC2, SSC4 and SSC6 were confirmed in the combined analysis, but QTLs that were specific to one pedigree disappeared or were detected at a lower significance threshold. Despite the limited benefits in term of the number of QTLs, the increase in the number of individuals, enabled us to separate two linked QTLs that were previously detected as a single one. False-positive QTLs were also detected as well as new QTLs characterised by a low frequency and/or a small effect. In addition, both pedigrees could be compared regarding the imprinting status at the IGF2-intron3-G3072A substitution, segregating on SSC2. The mutation was segregating within the European founders used in both pedigrees, Meishan individuals being all homozygous for the wild allele (G). This analysis, presented in **chapter 3,** shows that the structure of the pedigree (number of F1 individuals and size of half-sib families), the number of F1 heterozygous females at the IGF2 locus and the segregation of another QTL at a distance of 40 cM from the IGF2 locus influence the ability to detect imprinting at the IGF2 mutation. This spurious maternal effect can lead to incorrect conclusions regarding the imprinting status of the IGF2 mutation, with maternal effects being detected whereas they do not exist.

In order to fine-map the second QTL segregating on pig chromosome 2, a backcross design was set up. Sires that were finally progeny tested were all homozygous for a Meishan haplotype in the IGF2 region, so the phenotypic variation could not be due to the IGF2-intron3-G3072A mutation. Results from the progeny-testing presented in **chapter 4** confirmed that a QTL underlying fatness traits was segregating on the short arm of SSC2. However, the size of the QTL interval could not be reduced because of epistatic interactions. These epistatic effects could be detected because full-sibs with Identical-by-descent haplotypes in the QTL regions were progeny-tested. This particular design could be analysed without the strong assumptions of the line-cross models (according to which QTL alleles are fixed within breeds), so interactions could be detected. The re-analysis of one of the two F2 pedigrees confirmed that a region on SSC13 interacts with the QTL segregating in SSC2, but other candidate regions still need to be considered.

The combined analysis of different pedigrees finally gives few benefits regarding the number of new QTLs that were detected. However, these combined analyses enabled to successfully consider non-additive effects such as imprinting and epistasis.

During the work described in this thesis, a major technological advance occurred for pig geneticists, with the commercialisation of the Illumina PorcineSNP60 Beadchip. With this tool, the number of genotypes that can be included in a QTL analysis tremendously increased. In order to properly use this new type of information, the order of the SNPs along the genome must be reliable. In **chapter 5**, the first high-density genetic map of the pig is presented. This genetic map was computed using information from in silico and RH mapping of the SNPs in combination with recombination rates between them and finally comprised 38,599 SNPs. Four pig pedigrees based on different breeds were analysed separately, and the analysis of the recombination rate along the pig genome highlighted that the more recombinant regions tend to cluster around the telomeres irrespective of the location of the centromere. Two of the four analysed pedigrees comprised enough male and female meiosis to construct sex-specific maps. Major sex-differences in recombination were observed with a higher recombination rate in the females only within GC-rich regions, with females exhibiting a much stronger correlation between recombination rate and specific sequence features. This new information will be of major importance when dealing with QTL fine-mapping and pig genome evolution.

Finally, in the general discussion presented in **chapter 6**, arguments toward further fine-mapping of QTLs in pig are given despite the increasing interest in genomic selection. Despite major improvements that have been made in the development of high-density SNP chips, efforts are still needed to overcome the biases linked to the design of the chips. In parallel to the development of high-density genotyping tools, few improvements were made regarding

phenotyping. In this final chapter, various programs dedicated to the description of highly precise phenotypes and to the development of homogenous phenotyping practices are presented. Such programs, in combination with the development of appropriate genotyping tools, will facilitate the detection of causal variants. These efforts that are still necessary are not only required for pig but also to most livestock species for which QTL fine-mapping is still needed.

Samenvatting

De domesticatie van het varken vond plaats in het Neolithische tijdperk zo'n 10.000 jaar geleden en genetisch onderzoek heeft aangetoond dat dit onafhankelijk gebeurd is in Europa en China. De oorsprong van de huidige Europese en Aziatische rassen ligt dan ook in deze twee gebieden. De FAO registreerde in 2007 nog 560 varkensrassen waarvan er maar een klein aantal intensief gebruikt worden voor productie. In het proces van domesticatie en ras vorming heeft selectie plaatsgevonden op bestaande variatie zowel door natuurlijke selectie alsmede gerichte selectie door fokkers. Belangrijke kenmerken waarop geselecteerd is zijn reproductie, groei, vetaanzet, gedrag en resistentie tegen ziekten. Om de onderliggende mutaties op te sporen die de fenotypische variatie van deze kenmerken verklaart, zijn er de zogenaamde "quantitative trait loci (QTL)" studies opgezet. Vele duizenden QTL zijn de afgelopen jaren opgespoord maar de causale mutatie is maar voor een zeer beperkt aantal kenmerken opgespoord.

In de algemene introductie (**hoofdstuk1**) wordt een overzicht gegeven van QTL detectie in het varken. Hierin worden de meest belangrijke moleculaire technieken (tools) en karteringsstrategieën beschreven die gebruikt worden bij het varken. Belangrijke aspecten die van belang zijn bij het nauwkeurig in kaart brengen (fine mapping) van de opgespoorde QTL in het varken zijn o.a. het aantal individuen, het aantal informatieve genetische markers en de mogelijkheid voor het detecteren van niet-additieve QTL.

Een eerste aanzet voor het nauwkeuriger in kaart brengen van eerder gevonden QTL, is beschreven in **hoofdstuk 2.** Het principe van de in dit hoofdstuk beschreven gecombineerde analyse**,** was om de statische power te verhogen door het aantal individuen te verhogen. Hierbij zijn twee vergelijkbare F2-kruisingen (tussen Europese witte rassen en het Chinese Meishan ras), elk met ongeveer 1000 individuen, gecombineerd. Deze populaties zijn in de jaren 90 gemaakt in de onderzoeksfaciliteiten van INRA en WUR. De gemeenschappelijke QTL gedetecteerd op de chromosomen SSC2, SSC4 en SSC6 werden bevestigd in de gecombineerde analyse. Echter, de QTL specifiek voor één van de twee populaties, werden niet of met een veel lagere significantie gedetecteerd. Daarnaast werd een aantal nieuwe QTL met een lage frequentie en/of een klein effect gedetecteerd. Ondanks de geringe winst door het gezamenlijk analyseren van de twee verschillende populaties bleek het mogelijk door het gebruik van veel meer dieren om één QTL op te splitsen in twee nauw gekoppelde QTL. Bovendien was het mogelijk om het effect van de IGF2-intron3-G3072A substitutie op SSC2 binnen de twee populaties te vergelijken. Deze mutatie segregeert in de Europese voorouders van beide populaties terwijl de Meishan dieren homozygoot zijn voor het wild type allel (G). De analyse zoals beschreven in **hoofdstuk 3** laat zien dat de populatie structuur (aantal F1 dieren en de grootte van de half-sib families), het aantal heterozygote F1 zeugen voor het

IGF2 locus en het segregeren van een ander QTL, 40cM van het IGF2 locus, effect hebben in het opsporen van het QTL effect van de IGF2 mutatie. Dit schijnbare maternale effect kan leiden tot foute conclusies betreffende de "imprinting" status van de IGF2 locus, waarbij een maternaal effect geschat wordt terwijl dat niet aanwezig is.

Het tweede gedetecteerde QTL op chromosoom 2, op een afstand van 40CM van het IGF2 gen, is vervolgens met behulp van een terugkruisingspopulatie verder bestudeerd. De gebruikte beren in deze terugkruising zijn allemaal homozygoot voor het meishan haplotype rond het IGF2 gen. Hierdoor kan de fenotypische variatie niet het gevolg zijn van de IGF2-intron3-G3072A mutatie. Resultaten van de geteste nakomingen bevestigen dat er een QTL voor vetaanzet segregeert op de korte arm van chromosoom 2 (**Hoofdstuk 4**). Omdat er sprake was epistatische interacties kon het QTL interval niet verder verkleint worden. Deze epistatische effecten konden worden opgespoord omdat full-sibs met "Identity by decent (IBD)" haplotypen in het QTL gebied zijn gebruikt voor nakomelingen onderzoek, zonder gebruik te hoeven maken van lijn specifieke kruisingsmodellen. Het opnieuw analyseren van de INRA F2 kruising bevestigde bovendien dat er een interactie bestaat van een gebied op SSC13 en het QTL op SSC2. Andere kandidaat gebieden met mogelijk een interactie moeten nog verder worden onderzocht. Het onderzoek beschreven in dit hoofdstuk toont aan dat een gecombineerde analyse van verschillende populaties een aantal voordelen biedt ten aanzien van het aantal nieuwe gedetecteerde QTL. Bovendien maakt deze gecombineerde analyse het mogelijk om niet additieve effecten zoals imprinting en epistasie te besturen.

Tijdens het werk beschreven in dit proefschrift zijn er belangrijke technologische vooruitgangen geboekt zoals de ontwikkeling van de "Illumina PorcineSNP60 beadchip" door ABGC. Met deze SNP chip wordt het aantal genotyperingen dat voor QTL studies gebruikt kan worden enorm verhoogd. Om deze chip goed te gebruiken is het van belang de exacte volgorde alsmede de onderlinge genetische afstand van de SNPs op het genoom te kennen. In **hoofdstuk 5** wordt daarvoor de eerste recombinatie kaart met zeer hoge marker dichtheid gepresenteerd. Deze genetische kaart met in totaal 38.599 SNPs werd samengesteld door informatie te gebruiken van de SNPs welke geplaatst waren op de RH-kaart in combinatie met de recombinatie frequentie tussen markers in de verschillende karteringspopulaties. De 4 karterings-populaties zijn hiervoor apart geanalyseerd. Verdere analyse van de recombinatie frequentie binnen het varkensgenoom gaf aan dat de gebieden met hogere recombinatie clusteren aan de uiteinden van de chromosomen ongeacht de locatie van het centromeer. Twee van de vier populaties bevatte genoeg mannelijke en vrouwelijke meiosis om een sex-specifieke kaart te maken. De belangrijkste sex-gebonden verschillen in recombinatie werden gevonden in de vrouwelijk kaart waarbij een hoge recombinatie duidelijk gecorreleerd is met een hoger GC gehalte. Verder vertoont de vrouwelijk kaart

een duidelijke correlatie tussen de mate van recombinatie en specifieke sequenties in het genoom. Dit is belangrijke informatie voor QTL fijn kartering alsmede voor een beter inzicht en begrip van de evolutie van het varkens-genoom.
In de algemene discussie in **hoofdstuk 6** tenslotte, presenteer ik een aantal argumenten om door te gaan met het fijn karteren van QTL gebieden bij het varken ondanks een toenemende belangstelling voor genomische selectie. Ondanks de grote voortuitgang in de ontwikkeling van de mogelijkheden voor het met zeer hoge dichtheid genotyperen van dieren, is er relatief weinig vooruitgang geboekt betreffende fenotypering. In dit laatste hoofdstuk komen verschillende programma's aan de orde voor het zeer nauwkeurig beschrijven van fenotypes en nieuwe ontwikkelingen om te komen tot een uniformere en beter gedefinieerde fenotypering. Dit in combinatie met de geschikte genotypeer mogelijkheden zal het mogelijk maken om in toenemende mate de causale mutaties op te sporen. Deze inspanningen zijn niet alleen nodig bij het varken maar bij alle landbouwhuisdieren waar fijn kartering van QTL nodig is.

Voorwoord

This thesis has been a four-year journey during which I met a lot of people whom I would like to thank.

First of all I would like to thank Nadine Buys, Barbara Harlizius, Pascale Le Roy and Richard Visser who agreed to be members of my thesis committee.

I also would like to thank both directors of the labs I worked in, Johan Van Arendonk and Philippe Mulsant. Thank you for welcoming me in your labs where I enjoyed my stays.

This thesis would not have been possible without the support and confidence of Martien Groenen, Denis Milan and Juliette Riquet. I really thank you first to give me the chance to start this thesis and then to let me carry it out with all your support.

To carry on with my wonderful English, I thank all the people I met at Wageningen: Richard, Ole, Hendrik-Jan, Laurent, Yogesh, Mirte, good luck for your PhD.. Thank you to all of you, I spent a niece year with you in the Netherlands.

Parce qu'il m'est plus facile de m'exprimer dans la langue de Molière que dans celle de Shakespeare, je vais profiter de ces quelques lignes pour remercier toutes les personnes que j'ai croisées durant mes trois ans passés au LGC.

Je vais, bien sûr, commencer par remercier Juliette. Oui, je sais, je t'ai déjà remerciée quelques lignes plus haut, mais tu mérites bien quelques lignes de plus, en français. Merci beaucoup de m'avoir donné ma chance, j'ai énormément appris à tes côtés, tant scientifiquement qu'humainement. Je n'oublierai jamais les vendredi après-midi propices à l'élaboration d'hypothèses (voire même de théories) plus ou moins probables. Seulement voilà, le lundi matin, ces hypothèses penchaient plutôt du côté moins probable que plus probable. Et cela à cause de qui ? Ou devrais-je dire, grâce à qui ? Et oui, c'est grâce à toi Hélène. Merci beaucoup de m'avoir appris à apprivoiser QTLMAP (même si certains de ses aspects sont restés obscurs pour moi) et à prendre goût aux analyses statistiques. Merci beaucoup à toutes les deux, ça a été un immense plaisir pour moi de travailler avec vous, et merci également de m'avoir soutenue et entraînée pour mon concours.
Je tiens ici aussi à remercier les autres collègues de l'équipe porc. Nath, d'abord, comme ex-colocataire de bureau et comme docteur ès Gemma et docteur ès potin, Annie, ensuite également comme ex-colocataire de bureau. Je tiens maintenant remercier Katia. Merci pour ta bonne humeur, ta sympathie et

ta disponibilité malgré les nombreuses tâches que tu mènes. Merci aussi à Maguy, d'avoir réalisé une bonne partie des génotypages utilisés dans mon travail. Enfin, même si je n'ai pas travaillé directement travaillé avec vous, merci à Laurence et Laurianne, au moins pour m'avoir écoutée lors des réunions d'équipe !

Le LGC ne se réduisant pas seulement à l'équipe porc, je remercie toutes celles et ceux sans qui le LGC ne serait pas le monde des bisounours scientifiques qu'il est. Merci tout d'abord à Bertrand, sans qui une partie de cette thèse n'aurait pas été la même. J'ai beaucoup apprécié travailler avec toi, merci de t'être rendu disponible ! Maintenant, je remercie, l'équipe bonne humeur du labo, qui permet de passer des pauses café et déjeuner bien agréables : Pitou, Flo, Soso, Julien, Gwen, Simon, Magali, Eddie, Thomas, Patrice, Philippe, Kamila…
Je re-remercie aussi Philippe, de m'avoir aidé à me sentir moins seule en fin de repas!
J'ai aussi une pensée pour les thésards et CDD que j'ai eu la chance de rencontrer durant ces trois ans passés au LGC : Laurence, Marine, Romain, Natacha, Florian (courage, c'est bientôt la fin !), Man, Sophie, Jérôme, Claire…

J'en profite également pour remercier Maxime, mon colloc' Wageningenois et globe-trotter de première. Merci d'avoir été présent durant cette année néerlandaise. Je te souhaite beaucoup de bonheur que ce soit aux Pays-Bas, en France, ou là où tu auras décidé de poser tes valises !

Il est maintenant temps pour moi de remonter un peu dans le temps et de remercier celles et ceux que je connais depuis avant la thèse. Donc, juste avant la thèse, il y a eu l'ENSAT où j'ai rencontré des gens formidables, à commencer par Hervé Rémignon et Hélène Manse qui m'ont donné envie de poursuivre dans la recherche (et oui, j'avais commencé mon stage en étant déterminée à ne pas faire de thèse…). Et à l'ENSAT, j'ai eu la chance de rencontrer mes amies de maintenant : Katia et Mélanie, copines d'ENSAT et de thèse ! Il faut aussi que j'ajoute collègue maintenant pour toi Mélanie !
Je tiens aussi à remercier Maëlle. Merci d'être toujours là pour papoter de temps en temps, même si ces occasions se font de plus en plus rares, c'est toujours avec grand plaisir que je te retrouve !

Bon, je ne vais quand même pas remonter jusqu'à la maternelle… et bien si !! C'est là que je t'ai rencontrée Fanny ! Ma meilleure amie, avec qui j'ai grandi. Je suis très contente de t'avoir à mes côtés depuis ces quelques 23 ans qu'on se connaît.

Maintenant, et peut-être aurais-je dû commencer par vous, je remercie ma famille. Tout d'abord, merci à vous papa et maman d'avoir contribué à ce que je

suis devenue. Pourtant ce n'était pas gagné d'avance, moi qui comprends un mot sur deux et n'hésite pas à en inventer! … Je remercie aussi mes trois frères à moi (pas ceux du film), Raph, Ton et Soun. Merci pour tout: les fous-rires à table, les petits sourires en coin quand justement on veut rire mais qu'on ne peut pas ! Je remercie aussi mes beaux-parents, Annie et Rapha, ainsi qu'Arnaud, Béa et Martin, pour nous avoir soutenus, surtout la dernière année !

Enfin, je te remercie toi, Grégory, pour m'avoir soutenue durant cette thèse (même si les QTL sont devenus des QTP (?), les chromosomes des bâtons … !), durant la préparation de mon concours (tu y as toujours cru), et tout simplement pour être là, à mes côtés jour après jour.

About the author

Flavie Tortereau was born on the 23rd of July, 1984 in Chambray-lès-Tours (France) and was raised in Monts. In 2002, she obtained her baccalauréat at lycée Jean Monnet in Joué-lès-Tours. In September 2002, she started a two-years preparation class at lycée Pothier in Orléans and was accepted in September 2004 at the ENSAT in Toulouse to study agronomical sciences and particularly animal productions. In September 2007 she got her engineer's degree as well as a Master degree in animal productions and Quality and Food Safety, in Toulouse.
In September 2007, she started her PhD project "Combining different pedigrees to fine-map QTL in the pig". This project was carried out under the collaboration of the Laboratoire de Génétique Cellulaire at INRA, Toulouse and the Animal Breeding and Genetics Centre at Wageningen University. This project resulted in this thesis.
From the 1st of November 2011, Flavie has been employed by INRA as an engineer and is in charge of genetic evaluation of meat sheeps.

List of publications

Peer-reviewed publications

TORTEREAU F, Gilbert H, Heuven HC, Bidanel JP, Groenen MA, Riquet J: Combining two Meishan F2 crosses improves the detection of QTL on pig chromosomes 2, 4 and 6. Genet Sel Evol 2010, 42:42.

TORTEREAU F, Gilbert H, Heuven HC, Bidanel JP, Groenen MA, Riquet J: Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in Sino-European pig pedigrees. Genet Sel Evol 2011, 43:11.

TORTEREAU F, Sanchez MP, Feve K, Gilbert H, Iannuccelli N, Billon Y, Milan D, Bidanel JP, Riquet J: Progeny-testing of full-sibs IBD in a SSC2 QTL region highlights epistatic interactions for fatness traits in pigs. BMC Genetics 2011, 12:92.

Publications in review

TORTEREAU F, Servin B, Frantz L, Megens HJ, Milan D, Rohrer G, Wiedmann R, Beever J, Archibald AL, Schook LB, Groenen MA: Sex-specific recombination rate differences observed in the pig are correlated with GC content, *submitted to BMC genomics*

Conference proceedings

TORTEREAU F, Gilbert H, Heuven H, Bidanel JP, Groenen M, Riquet, J: Comparison of two pig pedigrees regarding the effect of the IGF2 mutation on backfat. Proceedings of the 9th World Congress on Genetic Applied to Livestock Production (WCGALP), Leipzig, Germany, 1-6 August 2010.

TORTEREAU F, Servin B, Beever J, Archibald A, Rohrer G, Milan D, Schook L, Groenen M: Strategies to build high-density linkage maps of the porcine 60k SNP chip. Proceedings of the 19th Plant and Animal Genome congress (PAG), San Diego, United States of Amercia, 15-19 January 2011.

TORTEREAU F, Sanchez MP, Billon Y, Burgaud G, Bonnet M, Bidanel JP, Milan D, Gilbert H, Riquet J: Fine-mapping of a QTL segregating on pig chromosome 2 highlighted epistasis, EAAP 2011, Stavanger, Norway, session 30.

Training and Supervision Plan

The Basic Package (3 ECTS)	
WIAS Introduction Course (mandatory, 1.5 credits)	2010
Course on philosophy of science and/or ethics (mandatory, 1.5 credits)	2011
Scientific Exposure (12 ECTS)	
ISAG meeting (20-24 july)	2008
WCGALP, Germany (1-6 august)	2010
PAG, San Diego (15-19 january)	2011
EAAP, norway (29 august - 2 september)	2011
Seminars Genomic Selection	2011
Ph D. students seminar 2008	2008, 2009, 2010, 2011
In-Depth Studies (9 ECTS)	
Multifactorial genetics	2007
DNA chip	2008
animal experimentation	2010
genomic selection course	2011
Professional Skills Support Courses (5 ECTS)	
fast reading	2009
preparing professionnal project	2009
how to communicate your knowledge	2010
TOEIC test, english course	2008
Didactic Skills Training (9 ECTS)	
Course to MSc minors on QTL mapping and positionnal cloning	2008, 2009, 2010
Course to MSc majors : DNA polymorphism, linkage maps, major genes and QTL detection	2008, 2009
Course to MSc majors and ingeneers students : Genetics and meat quality	2009
Supervising of one student	2010
Management Skills Training (3 ECTS)	year
organisation of the PhD students seminar of the genetics Department in France	2008
reviewing of an article for the scientific journal Genetics Selection Evolution	2010

Education and Training Total : 41

Colophon

The research described in this thesis was financially supported by the Laboratoire de Génétique Cellulaire (INRA) and Wageningen University.

This thesis was printed by: GVO printers & designers B.V., Ede, the Netherlands