Use of genomic information in mass-spawning fish

Cécile MASSAULT

Thesis committee

Thesis supervisor

Prof. Dr. Ir. Johan A.M van Arendonk
Professor of Animal Breeding and Genomics Centre
Wageningen University

Thesis co-supervisors

Dr. Ir Henk Bovenhuis
Assistant professor
Animal Breeding and Genomics Centre
Wageningen University

Dr Dirk-Jan de Koning
Principal Investigator
The Roslin Institute and Royal (Dick) School of Veterinary Studies
University of Edinburgh (UK)

Others members

Prof. dr. ir. E.W. Brascamp
Wageningen University

Dr. Anna Sonesson
Nofima, Ås (Norway)

Prof. Filip Volckaert
Katholik Universiteit Leuven (Belgium)

Dr. Ir. Geert Wiegertjes
Wagenigen University

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Science

Use of genomic information in mass-spawning fish

Cécile MASSAULT

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 Academicl Board
to be defended in public
on Wednesday 30 June 2010
at 11 AM in the Aula

Cécile Massault
Use of genomic information in natural mating mass-spwaning fish
172 pages.

Ph.D. thesis: Animal Breeding and Genomics Centre, Wageningen University, Wageningen, NL (2010)

With references and including summary in english and dutch

ISBN 978-90-8585-688-7

ABTRACT

The aim of this thesis was to detect quantitative trait loci (QTLs) in aquaculture species for various traits of economical importance. The results of this work can be integerated into breeding programs after validation and increase our biological knowledge on the complex traits studied.

The first step was to design experiment that will successfully detect QTLs in aquculture species. The designs depended largely on the family structure of the species, which varied from highly controlled (such as trout and oysters) to poorly controlled (sea bream and sea bass) and the actual size of the QTLs. Others factors such as number of progeny per family, heterozygosity and heritability of the trait were also found to impact on the success of QTL experiments.

The next part of the thesis focuses on the actual QTL mapping for stress response in sea bass and disease resistance in sea bream, using the experimental design previously analysed for complex family structure with skewed parental contribution. Significant QTLs were found for body weight and morphology in sea bass and for survival and body length in sea bream. Only suggestive QTLs were detected for stress response in sea bass.

Finally, the possibilities of using genomic selection for mass-spawning species, where natural mating is used, were investigated. Genomic selection gave significant higher genetic gain and accuracy of selection than traditional selection methods, but rate of inbreeding is dangerously high and need to be controlled.

TABLE OF CONTENTS

Chapter 1

General introduction

1.1 Current state of sea bream and sea bass aquaculture in Europe

European sea bass (*Dicentrarchus labrax*) and gilthead sea bream (*Sparus aurata*) are two mass-spawning species mostly cultivated in the Mediterranean sea. Aquaculture in both fish started in early 1980s and production has increased almost continuously over the years in Europe (Figure 1 – FAO 2006), except for 2002 and 2004 due to the food price crisis. The main producer for both species is Greece, which controls over 80% of sea bream and over 70% of sea bass production in Europe.

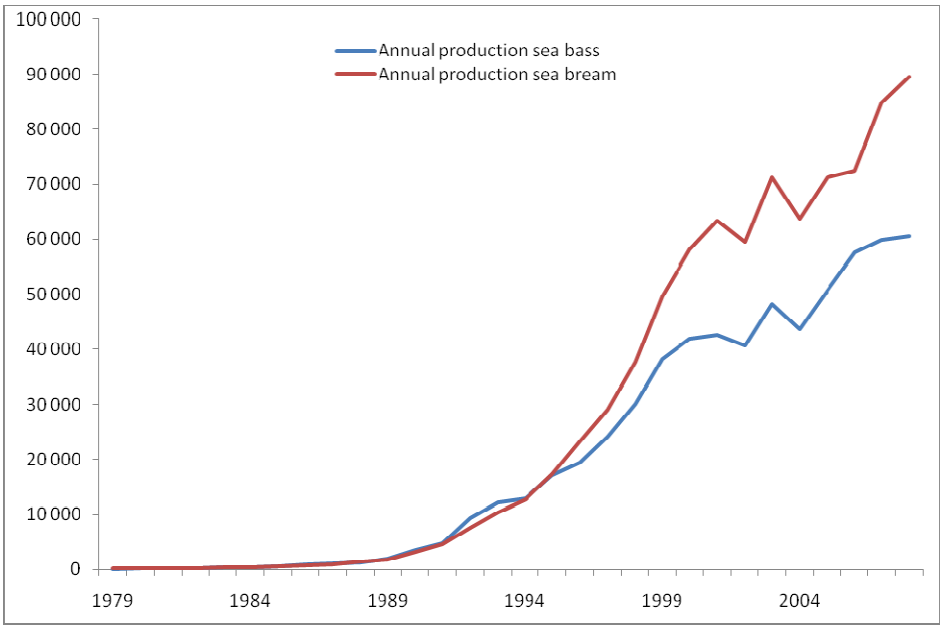

Figure1. Annual aquacultured production of European sea bass and gilthead sea bream in Europe (FAO 2006)

Sea bream and sea bass aquaculture are important for the Mediterranean economy. Although selective breeding is not yet widely used in sea bass and sea bream aquaculture, interest in the genetic improvement of farmed fish is motivated by the increase in consumer demand. European regulations for food

production emphasise general well-being of reared individuals. Therefore stress response and disease resistance are being considered as important breeding goals.

1.2 AQUAFIRST project

The economic importance of sea bream and sea bass aquaculture played an important part in the launch of the AQUAFIRST project, a sixth framework European project (contract no 513692), which ran between 2004 and 2008. It combined genetic and functional genomic approaches to stress and disease resistance in fish and shellfish. This project involved the collaborations of seventeen institutes located in ten European countries over four years. Fish are subject to increased stress in intensive production systems, which in turn have an impact on the quality, performance and the susceptibility to disease and results in loss of production.

The main objectives of the AQUAFIRST project were

- To characterize in sea bream, sea bass, trout and oyster, stress- and disease-responsive genes as potential candidate gene markers for desirable traits.

- To determine genetic parameters of stress in these four species and genetic parameters of disease in sea bream and trout by seeking associations with (i) variations in response to stress and resistance to pathogen and (ii) selected candidate genes and microsatellite markers by segregation analysis in appropriate families (QTL analysis)

- To provide specific protocols for selection, marker-assisted selection (MAS) and marker-assisted introgression

This thesis focuses on two species of the AQUAFIRST project, European sea bass (*Dicentrarchus labrax*) and gilthead sea bream (*Sparus aurata*), the detection of QTL for stress response and disease resistance in these species, and the use of genomic information in selection for natural mating mass-spawning populations.

1.3 Stress response and disease resistance

Stress response and disease resistance are two complex traits, which are influenced by many genes. At present, genes involved in stress and/or disease response are largely unknown and biological pathways that comprise those genes are not well-understood. Engelsma et al. (2002) reported the link between stress response and disease resistance pathways, where chronic stress is associated with a reduced immune response.

Various factors can cause stress such as low water quality, fish handling procedures, pathogen exposure and confinement, inflicted on animals during their rearing time in the fish farm. Those factors trigger chronic stress which activates the release of cortisol in an abnormally high quantity. The presence of cortisol in large quantity has a negative impact on macrophage production and reduces the immune response. Fish affected by chronic stress therefore are expected to have a reduced immune response.

A commonly accepted measure of stress response in fish is the quantification of cortisol levels in blood. Disease resistance is often recorded through survival and correlated traits such as immunological and physiological parameters (Fjalestad *et al*. 1993).

1.4 Selective breeding and QTL mapping in aquaculture

Selective breeding in aquaculture is mostly achieved to date by selection of parents for the next generation through mass selection or family-based selection. While mass selection is based on an animal's own phenotypic records, family based schemes select animals either on the performance of the family (between family selection) or relative performance of an animal in the family (within family selection). Despite progress in artificial reproduction (Morreti 1999, Saillant *et al*. 2001, Dupont-Nivet *et al*., 2006), natural mating is still widely used in sea bass and sea bream farms. Selection based on genomic information such as marker assisted selection (Fernando and Grossman, 1989) and genomic selection (Meuwissen *et al*., 2001) have not been implemented in sea bass and sea bream.

The quality of breeding programs is evaluated by two main parameters: the rate of genetic gain (how much the trait is improved) and increase of the rate of inbreeding (loss of genetic variation). A major issue in fish breeding is the high fecundity of fish species. While it is rather an advantage in animal breeding with the large number of offspring that can be obtained with limited number of parents, it can also cause a rapid build up of inbreeding in a closed population. Fessehaye et al. (2006) describes a rate of inbreeding in tilapia that is higher than an acceptable value of 1% (Bijma 2000) due to unequal parental contribution, which was also described in sole (Blonk *et al*. 2009) and sea bream (Brown *et al*. 2005). A solution adopted by fish farmers is to regularly introduce parents from the wild into the brood stock (Gjedrem *et al*. 2005). The side effect of this practice is the partial loss of genetic gain achieved in previous generations.

Detection of quantitative trait loci (QTL) provides genetic knowledge that helps to integrate MAS into a breeding scheme. QTLs are genes or genome regions associated with a trait of economical importance. Experimental design is very important for QTL detection as it maximizes the power to detect associations (Kolbedhari *et al*. 2005). The number of individuals genotyped and phenotyped, the population structure and the density of genetic maps has to be considered carefully. While a large number of QTLs have been discovered in terrestrial farm animals such as cattle, pig, chicken and sheep, aquaculture has not yet benefitted from such attention. Although domestication of some fish species started in ancient civilization, breeding programs are recent (early 1970s). The advances in genomics tools for aquaculture species (Canario *et al*. 2008), such as the development of large panels of markers, permit the use of high throughput technologies and the use of MAS and genomic selection. Genetic maps for both European sea bass (Chistiakov *et al*. 2008) and gilthead sea bream (Franch *et al*. 2006) have been established with a large number of microsatellites and AFLP markers. This thesis uses those resources to provide the foundation of dissection of economic traits in sea bass en sea bream.

1.5 Aim and outline of the thesis

The aims of this thesis are to (i) develop an experimental design for QTL mapping in aquaculture species, (ii) use the data provided by such experiments (performed by collaborators) to find QTLs associated with stress response and disease resistance in gilthead sea bream and European sea bass and (iii) examine the impact of genomic selection on genetic gain and inbreeding.

Chapter 2 describes three experimental designs to perform QTL analysis: “hierarchical design” where two divergent lines are crossed, “large full-sib

families design" where selective genotyping can be applied and "mass-spawning design" for species such as sea bream and sea bass, where natural mating is used to produce the F1 generation. The impact of number of families, family size, heritability, heterozygosity and size of QTL are investigated.

Chapter 3 and **Chapter 4** summarize the results of QTL mapping for stress response and disease resistance in sea bream and sea bass. **Chapter 3** focuses on linkage analysis performed for stress response, body weight and morphology using two methods: half-sib regression analysis and variance component using 570 European sea bass. **Chapter 4** uses the same methodology to detect associations with disease resistance after exposure to *Photobacterium damselae subsp. piscicida* in sea bream.

Chapter 5 investigates the genetic level and inbreeding for three selective breeding methods (phenotypic, BLUP and genomic selection) for natural mating population for 10 generations of selection. The natural mating population was simulated following a structure observed in the sea bream population in **Chapter 4.**

Chapter 6 discusses the four main issues presented in this thesis: (i) power of experimental designs and comparison with simulated and real mass-spawning data set, (ii) QTL mapping in aquaculture and genome wide association study (GWAS), (iii) comparison of selective breeding approaches and (iv) parentage assignment.

Reference

Bijma, P. (2000) Long term genetic contributions: prediction of rates of inbreeding and genetic gain on selected populations. Thesis (PhD) University of Wageningen

Blonk R.J.W, Komen J., Kamstra A., Crooijmans R.P.M.A., van Arendonk J.A.M. (2009) Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. *Aquaculture* **289**, 26-31

Brown R.C., Woolliams J.A., McAndrew B.J. (2005) Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture* **247**, 219-225
Canario A.V.M., Bargelloni L., Volckaert F., Houston R.D., Massault C., Guiguen Y. (2008) Genomics toolbox for farmed fish. *Review in Fisheries Science* **16**, 3-15

Chistiakov D.A., Tsigenopoulos C.S., Lagnel J., Yuanmei G., Hellemans B., Haley C.S., Volckaert F.A.M., Kotoulas G. (2008) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Animal Genetics* **39**, 623-634

Dupont-Nivet M., Vandeputte M., Haffray P., Chevassus B. (2006) Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture* **252**, 161-170

Engelsma M.Y., Huising M.O., van Muiswinkel W.B., Flik G., Kwang J., Savelkoul H.F.J., Verburg-van Kemenade B.M.L. (2002) Neuroendocrine-immune interactions in fish: a role for interleukin-1. *Veterinary Immunology and Immunopathology* **87,** 467-479

Fernando R.L., Grossman M. (1989) Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* **21**, 467-477

Fessehaye Y., El-Bialy Z., Rezk M.A., Crooijmans R., Bovenhuis H., Komen H. (2006) Mating systems and male reproductive success in Nile tilapia (*Oerochromis niloticus*) in breeding hapas: a microsatellite analysis. *Aquaculture* **256**, 148-158

Fjalestad K.T., Gjedrem T., Gjerde B. (1993) Genetic-Improvement of Disease Resistance in Fish - an Overview. *Aquaculture* **111**, 65-74.

Food Administration Organisation (2006) http://www.fao.org

Franch R., Louro B., Tsalavouta M., Chatziplis D., Tsigenopoulos C.S., Sarropoulou E., Antonello J., Magoulas A., Mylonas C.C., Babbucci M., Patarnello T., Power D.M., Kotoulas G., Bargelloni L. (2006) A genetic

linkage map of the hermaphrodite teleost fish *Sparus aurata L*. *Genetics* **174**, 851-61.

Gjedrem, T., Andersen, O., Bentsen, K.B., Fjalestad, K.T., Gjerde, B., Hayes, B.J., Kolstad, K.,Meuwissen, T.H.E., Olesen, I., Refstie, T., Rye, M., Sonesson, A., Thodesen, J.,Woolliams, J.A., (2005). Selection and Breeding Programs in Aquaculture. Springer,Dordrescht, The Netherlands, p. -364.

Kolbehdari D., Jansen G.B., Schaeffer L.R., Allen B.O. (2005) Power of QTL detection by either fixed or random models in half-sib designs. *Genetics Selection Evolution* **37**, 601-14.

Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics* **157**, 1819-1829

Morreti A. (1999) Manual on hatchery production of seabass and gilthead seabream. *Food and Agricultural Organisation*, Rome

Saillant E., Chatain B., Forstier A., Przybyla C., Fauvel C. (2001) Parental influence on early development in the European sea bass. *Journal of Fish Biology* **58**, 1585-1600

Chapter 2

QTL mapping designs for aquaculture

Cécile Massault
Henk Bovenhuis
Chris Haley
Dirk-Jan de Koning
Aquaculture (2008) 285:23-29

2.1 Abstract

Rapid development of genomics technology is providing new opportunities for genetic studies, including QTL mapping, in many aquaculture species. This paper investigates the strengths and limitations of QTL mapping designs for fish and shellfish under three different controlled breeding schemes. For each controlled breeding scheme, the potential and limitations are described for typical species and are illustrated by three different designs using interval mapping. The results show that, regardless of the species, the family structure is extremely important in experimental designs. The heritability of the QTL (controlled by its allele frequency and effect on the trait) also has an important impact on the power to detect QTL, while the overall polygenic heritability of the trait is less important. Marker density does not greatly affect the power when the distance between markers is less than 10 cM; but ideally spacing should not exceed 20 cM. For each of the systems studied, it is possible to design an experiment that would have an 80% power to detect a QTL of moderate effect (explaining between 1.5 and 5% of the trait variation) by genotyping 1000 or fewer individuals.

2.2. Introduction

Quantitative variation characterizes most traits of economic importance in livestock, including disease resistance, growth or meat quality.Variation in such "complex" traits often is controlled by a number of different genetic loci (quantitative trait loci or QTL) and environmental influences. QTL mapping studies have led to the identification of many genomic regions associated with QTLs in agricultural animals. Such studies are a prerequisite to the dissection and understanding of complex trait variation and the use of QTL in marker-assisted selection (Martinez 2007, Sonesson 2007a, Sonesson 2007b). QTL studies have been successfully applied to most farm animal species (reviewed by Andersson and Georges 2004) and more recently to aquaculture species such as Atlantic salmon, rainbow trout and tilapia (reviewed by Korol *et al*. 2007). However, for some mass-spawning species such as sea bream and sea bass, QTL mapping has barely been undertaken and linkage maps have only recently become available (Sonesson 2007a).

Several studies have examined statistical approaches to optimize the power of QTL detection experiments (Weller *et al*. 1990, van der Beek *et al*. 1995; Williams and Blangero 1999). Optimal designs for QTL detection depend on specific characteristics of a species and, therefore, optimal designs for terrestrial livestock species may be sub-optimal or impractical for aquaculture species.

Aquaculture species present both challenges and opportunities for the experimental design of QTL studies because of their high fecundity (which enables breeders to produce large families) and because species differ in the degree to which breeding can be controlled (Gjedrem *et al*. 2005). Both fecundity and breeding control have an impact on the family structure of a

species and therefore on the design of QTL studies. Some studies have considered utilizing specific aspects of aquaculture for QTL mapping designs such as gynogenesis/androgenesis (Martinez *et al*. 2002) or exploiting the difference in recombination fractions between males and females in some species (Hayes *et al*. 2006), and therefore require fewer markers to detect QTL. Breeding control is variable among aquacultural species, and the level of control imposes limits on an experimental design. The overall size of an experiment is limited by the total resources available for genotyping and phenotyping. This consideration includes the number of individuals used for mapping, as well as the number of markers that will be typed, which in turn determines the average distance between markers.

A complication with mass-spawning species is that these species generally have an effective number of parents that is much lower than the potential total number of parents due to the unequal contribution of parents to the next generation. Some potential parents do not contribute at all, and for the ones that contribute, the contribution is variable (Bekkevold *et al*. 2002, Porta *et al*. 2006, Brown *et al*. 2006). QTL mapping designs for those species require knowledge about the population structure, i.e. family type and size (Vandeputte *et al*. 2005).

The goal of this study is to explore experimental designs for successful QTL detection in aquaculture species. Three experimental designs, each corresponding to a different level of breeding control are chosen. The relative importance of various parameters of the design, such as family structure, heritability of a trait and the segregation of QTL alleles in the parental population are investigated.

2.3. Materials and methods

Experimental design

The experimental designs that were evaluated were designed to detect QTL with an effect between 1.5 and 20% on the phenotypic variation of a trait of interest, given experimental and financial limitations in terms of genotyping and phenotyping of outbred populations. We used the concept of experimental power (a statistic that describes how often a particular experiment would detect QTL of a given size) to compare experimental designs. In other words, the power of a QTL experiment is the success rate of discovering a QTL with a given effect.

The power of QTL experiments depends on different factors that fall into three categories: controlled, partially controlled and uncontrolled factors. The controlled factors comprise the numbers of individuals to be phenotyped and genotyped, the interval distance between markers and the false positive rate. Those factors are either fully determined by the experimenter (setting up the false positive rate) or limited by the availability of resources (number of markers used, maximum number of individuals in an experiment). The partially controlled factors are family structure and size of the experiment (number of families and number of progeny per family), heterozygosity of the parents for the QTL, and heritability of the trait of interest. Family structure, as well as the financial resources available, will determine the number of individuals to be genotyped and phenotyped. The marker contrast associated with the specific number of individuals genotyped and phenotyped will play an important role into success of QTL detection. Heterozygosity of the parents for the QTL corresponds to the fraction of parents that are heterozygous and therefore

informative for detecting QTL (Weller *et al*. 1990).

The major uncontrolled factors are the number of QTL and the magnitude of the QTL effects, which cannot be estimated prior the experiment. For a bi-allelic, additive QTL with allelic effect a (difference between alternative homozygotes is *2a*) and allele frequencies *p* and *q* (=1 − *p*) the variance of the QTL (σ_q^2) is $2pqa^2$. The proportion of phenotypic variance explained by the QTL, also referred to as the heritability of the QTL (h_q^2), is σ_q^2 / σ_p^2, where σ_p^2 is the phenotypic variance. For a single additive-effect QTL (h_q^2), increases as a function of the frequency of the rare allele, with the highest value at $p = q = 0.5$. In this study, results of power calculations will be presented as a function of the variance explained by the QTL.

It is assumed that the method used to map QTL is interval mapping, which uses information from two markers simultaneously and searches for the QTL in the bracketed interval. This method requires a known pedigree with phenotypic records on the last generation, as well as genotypes for parents and offspring (Lander and Botstein 1989).

Description of designs

The simulated experimental designs are relevant for a number of fish and shellfish species. The designs are associated with the level of breeding control for the species and hence the family structure. We simulated three different experimental designs representing each level of breeding control described above. The first design, named here the “hierarchical design”, is applicable to

species for which we have knowledge and control of reproductive behaviour and genetics. In species like Atlantic salmon, rainbow trout and common carp, full-sib families can be obtained with a relatively large number of progeny (up to thousands). The second experimental design corresponds to mass-spawning or batch-spawning species, like sea bream, European sea bass and tilapia, ("mass-spawning design"), where designed paired matings are hard to achieve. The third design is appropriate with species for which artificial reproduction is partially controlled, like oysters, and in which large full-sib family sizes permits use of selective genotyping ("large full-sib family design").

The "hierarchical design": The standard scenario used 1000 individuals, structured in five full-sib families of 200 progenies each. The heritability of the trait of interest was primarily set to 0.5 and would vary in other scenarios. The heterozygosity was set at 0.5 (50% of the parents are heterozygous for the QTL). A heterozygosity of 50% or higher is within reach when the parents were the result of a cross between two divergent (outbred) lines. The false positive rate α was fixed at 0.01 and the distance between markers at 20 centiMorgans (cM). Various aspects of this basic scenario were changed in order to evaluate the effect of family structure, heritability and heterozygosity on the power to detect QTL. The scenarios are described in Table 1. The effect of marker spacing was investigated for the basic scenario using a spacing of 5, 10, 20 and 50 cM. The power to detect QTL was calculated for QTL effect (in s.d.) from 0.134 to 0.387 (corresponding to a proportion of phenotypic variation explained by the QTL from 1 to 15%). The power of different scenarios was calculated using the deterministic method described by van der Beek et al. (1995). The QTL mapping method underlying these power calculations tests for the presence of a QTL by using the difference between offspring inheriting

alternative chromosome segments from their parents. This contrast will depend upon the probability that a parent is heterozygous for the QTL, and if it is, on the QTL effect and the recombination fractions between the markers and the QTL. The standard error of this estimated contrast depends upon the within-family variation, which is different for full- and half-sib families and is a function of the overall heritability of the trait and the family size. The method for prediction of power assumes that for all offspring, it can be determined which marker allele was inherited from the parents. This approach can handle a variety of two- and three-generation family structures (van der Beek *et al*. 1995). The computation of power is as follows:

$$power = \sum_{x=0}^{np} P(x) * P[\chi 2(NC(x), np) > T] \quad (1)$$

with x representing the number of heterozygous parents, np the total number of parents for which marker contrasts are computed, P(x) is the binomial probability that x out of np parents are heterozygous and $\chi 2$(NC(x), np) is the $\chi 2$ distribution with non-centrality parameter (NC) as a function of x and the relative QTL effect. T is the threshold for detecting a QTL at a given α. The whole power computation is detailed in van der Beek et al. (1995). We will refer to this method as the full- or half-sib regression method in the rest of the paper.

Table 1 - Parameters for hierarchical experimental designs

Scenario	*# full-sib families*	*# of offspring per family*	*Heritability*	*Heterozygosity*
Scenario 1	5	200	0.5	0.5
Scenario 2	5	200	0.2	0.5
Scenario 3	10	100	0.5	0.5
Scenario 4	5	200	0.5	0.2
Scenario 5	5	100	0.5	0.5

The “mass-spawning design”: Mass-spawning species present challenges for QTL mapping. The reproductive behaviour of some species is such that females only spawn in groups, making reproduction difficult to manipulate artificially. Natural spawning in groups of males and females is often practised in aquaculture and may be the only tractable way to produce offspring for further study. However, females may produce progeny sired by a number of different males and, similarly, males may sire progeny from a number of different females. Animals within a broodstock have uneven genetic contributions; some males and some females will produce more offspring than others, with a varying proportion of breeders having no offspring at all, and family sizes may vary widely. The traditional sexual reproduction is such that a small number of parents produce the majority of the offspring (Brown *et al*. 2005, Fessehaye *et al*. 2006). Those contributions determine the expected family structure for a

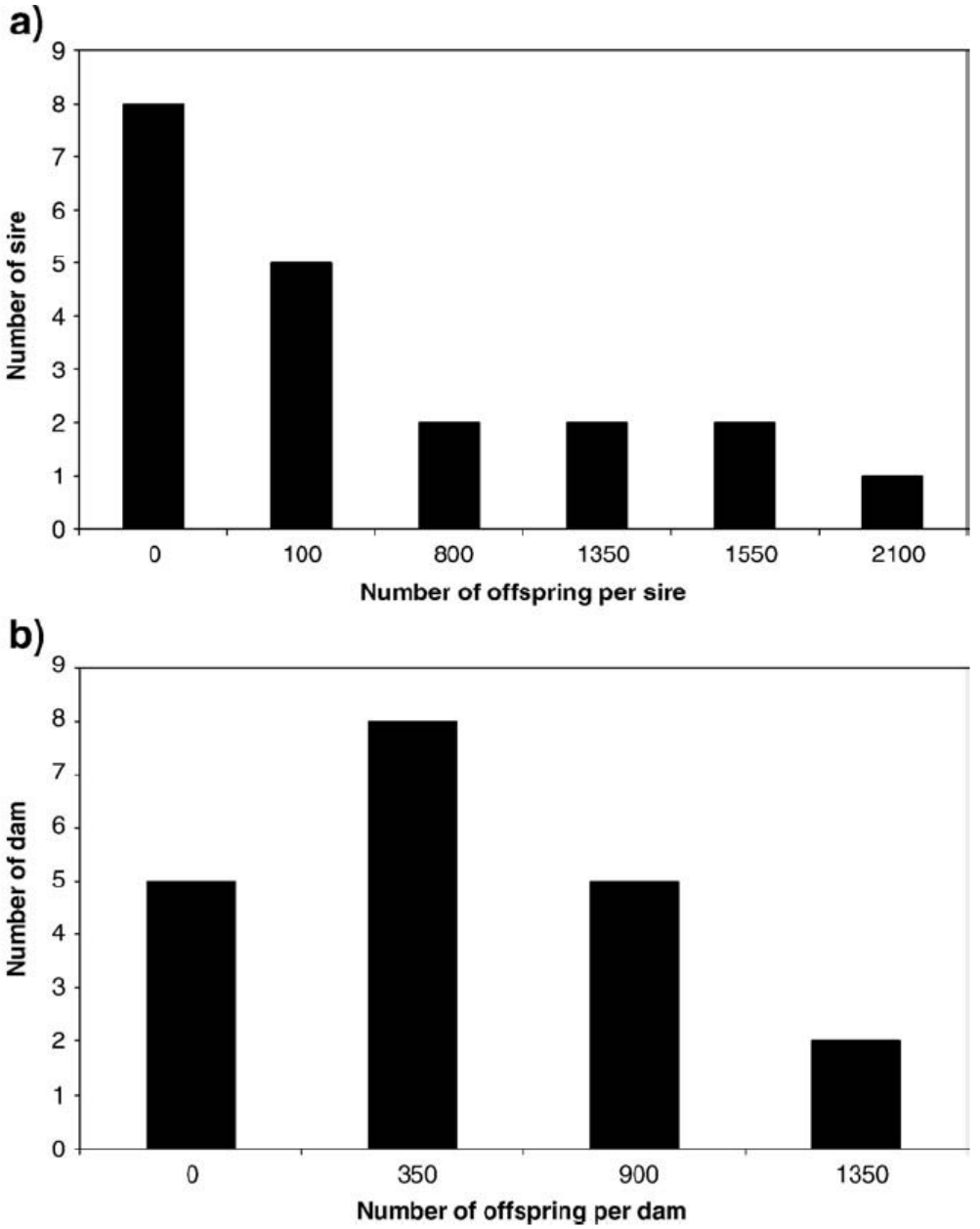

Figure 1 - Histograms a) and b) show the distribution of the family sizes for males and females, respectively, for a population of 10,000 offspring generated by a brood stock of 20 males and 20 females

species. Another complicating factor is that parentage needs to be shown in Figure 1, were used to simulate nine populations, each characterized by a different mass-spawning population size (10,000, 20,000 or 30,000) and a different brood stock structure (20 males and 20, 40 or 60 females). Each simulated population was evaluated for three different parentage-testing population sizes (1000, 1500 and 2000 individuals). From QTL mapping populations that contain at least seven families of 100 progenies, we drew two pedigrees: seven or five families of 100 offspring each, which were used for power calculations. In this experiment, only the case of selecting exactly 100 offspring was investigated. In reality, such an experiment will rarely have families of exactly the same size and is more likely to show a mix of family sizes.

Power calculations for sea bream were done with a power calculator for variance component analysis that is available through GridQTL (www.gridqtl.org.uk). This power calculator is derived from the simulation approach developed by Williams and Blangero (1999) and subsequently extended for large and complex animal pedigrees (Yu *et al*. 2004). These power calculations assume that the QTL is completely linked to a fully informative marker (no recombination between QTL and the marker) and heterozygosity for the QTL is based on an identical-by-descent inheritance pattern among on individuals in the pedigree. The heritability of the trait of interest was set at 0.5 with varying QTL effects. We also used the half-sib regression approach (van der Beek *et al*. 1995) to make a comparison between the two methods for power calculation and to evaluate the effect of incomplete marker-QTL linkage.

The “large full-sib family design”: This design is suitable for species for which very large full-sib families can be obtained, such as oysters. Selective

genotyping, meaning that among all animals phenotyped, only some will be genotyped (usually the extreme phenotypes), is potentially useful to reduce cost and increase experimental power (Lebowitz *et al*. 1987). The level of breeding control of these species allows a large experimental family size. In our scenario, we had 5000 individuals, divided into five families of 1000 progenies each, from which a subset of 200 full-sib animals would be selected from each family for genotyping (100 from each tail of the trait distribution). While the number of genotyped individuals was constant (200/family), we varied the total number of individuals per family (between 200 and 1000) that were phenotyped, thus providing different levels of selective genotyping.

Table 2 - Scenarios for large full-sib family design where number of individuals genotyped and phenotyped are different for different genotyped fractions.

Scenario	*Heritability*	*Heterozygosity*	*Genotyped fraction*	*Individuals genotyped*	*Individuals phenotyped*
1	0.5	0.5	0.1	1000	5000
2	0.1	0.5	0.1	1000	5000
3	0.5	0.5	0.2	1000	2500
4	0.1	0.5	0.5	1000	1000
5	0.5	0.1	0.1	1000	5000

The heritability of the trait was 0.50 and the probability that a parent was heterozygous for the QTL was 0.5. Table 2 shows the five different scenarios investigated, considering different genotyped fractions (scenarios 1, 3 and 4) with their corresponding number of genotyped and phenotyped individuals, and considering lower heritability (scenario 2) and lower heterozygosity (scenario 5).

For the power calculations, we used the full-sib regression method described by van der Beek et al. (1995), but adapted for selective genotyping in full-sib families based on methods described by Bovenhuis and Spelman (2000).

2.4. Results

Hierarchical design

The power of QTL experiments was calculated for QTL heritabilities (h_q^2) between 0 and 0.50. Figure 2 shows the experimental power for the five scenarios described in Table 1. To obtain an 80% power, the proportion of variance explained by the QTL needed to be at least 4.1%, 5%, 5.2%, 6.6% and 8.1% for scenarios 1, 2, 3, 4 and 5, respectively. Scenario 1 (basic scenario) was the most powerful experimental design. In comparison with scenario 3, we can see that family structure (5×200 full-sib versus 10×100 full-sib) had an impact on the minimal QTL effect required (4.1% versus 5.2%, respectively) for an 80% power. Having a smaller number of families with a larger number of progeny per family increased the power (or decreased the QTL effect required). Scenario 2 showed that a decrease in heritability had a negative impact on the power to detect QTL. Reducing the heritability from 0.5 to 0.2 increased the QTL effect required from 4.1% in scenario 1 to 5% in scenario 2. A lower

heterozygosity (0.2) as described in scenario 4 required the QTL heritability to be at least 6.6%; a low heterozygosity means that fewer parents are heterozygous for the QTL and therefore less information is available for QTL detection. Scenario 2 and 3 had similar results because in scenario 2 the low heritability was compensated for by an improved family structure and vice-versa for scenario 3. Scenario 5 showed that the size of the experiment is a very important parameter; for an experiment of half the size, the QTL heritability required for a power of 80% increased to 8.3% (exactly twice as much as scenario 1). The full-sib regression method for power calculations (van der Beek *et al*. 1995) assumes QTL analysis via interval mapping and the QTL being located exactly in the middle of a marker bracket. With this method, it was possible to look at the effect of marker spacing (interval between 2 markers expressed in centiMorgans (cM)). Figure 3 shows the power for the basic scheme with different marker intervals. For a sparse marker map (distance between marker of 50 cM), the power is substantially lower than for marker intervals of 5,10 or 20 cM. But comparing the power with genetic marker maps that vary in density between 5 cM and 20 cM, the difference is small, but the genotyping load increases as the density of the genetic map increases. For our experimental design, a reasonable balance between power and genotyping load was reached using a 20 cM interval between markers, assuming that markers are fully informative (Darvasi *et al*. 1993, Xu *et al*. 2005). Using less informative markers will require higher density genetic maps for the same experimental power.

Mass-spawning design

Table 3 summarizes the characteristics of paternal and maternal half- sib families for the scenarios described in the Materials and methods section

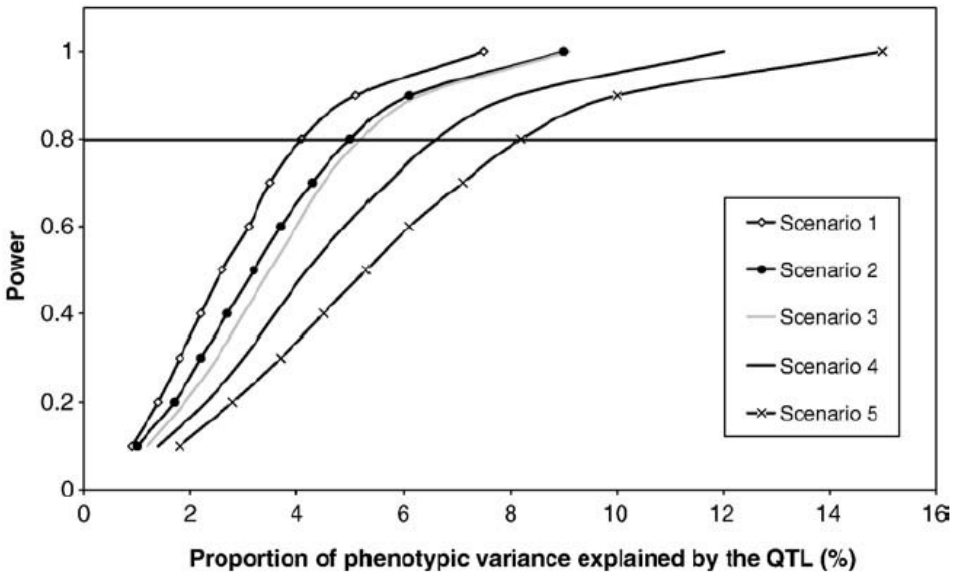

Figure 2 - Power to detect QTL as a function of the percentage of phenotypic variance explained by the QTL for the 5 scenarios of hierarchical breeding design (**Table 1**)

(population of 10,000, 20,000 or 30,000; sample size of 1000,1500 or 2000; mating schemes of 20 ♂×60 ♀, 20 ♂×40 ♀ or 20 ♂×20 ♀). The size of the total population does not seem to have an impact on the number of half-sib families obtained. However, the sample size has an effect, which is more pronounced in the female half-sib family sizes. Using a sample size of 2000 individuals results in at least 7 half-sib families >100; except for 2 cases out of a total of 18. When varying the number of females in the broodstock, the number of paternal families >100 did not change considerably, but as the number of females decreased, the number of maternal half-sib families >100 increased. For the ”mass-spawning” model, we simulated a population of 10,000 individuals produced by mating 20 males and 20 females. Pedigrees of five paternal half-

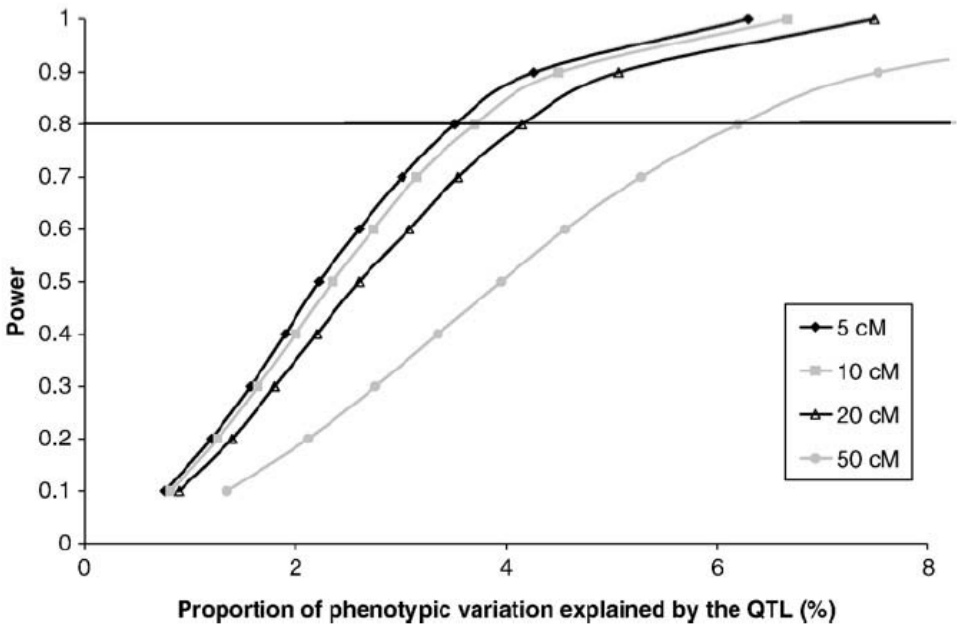

Figure 3 - Power of basic scheme(scenario 1, table 1) for hierarchical design as a function of percentage of phenotypic variation explained by the QTL with varying marker density

sib families of 100 and seven paternal half-sib families of 100 were selected from a sample of 2000 individuals that were randomly drawn from the simulated population. Figure 4 shows the power trends for the two pedigrees with the variance component method and the half-sib regression method. For a power of 80%, the variance component method required a QTL that explains 6.2% of the phenotypic variation for the pedigree of 700 individuals and 7.6% for the pedigree of 500.

The regression method required a larger QTL, explaining 12.2% of variance for the pedigree of 700 individuals and 15.5% for the pedigree of 500 individuals. The expected power to detect QTL varied considerably between methods; as the

power for the large (n = 700) pedigree was lower for the regression approach than the power for the small (n = 500) pedigree under the variance component approach (Figure 4).

Table 3. - Mass-spawning family distribution.

Sample size	*Average number of half-sib families > 100 offspring*								
	10,000 population			20,000 population			30,000 population		
	20♂ 60♀	20♂ 40♀	20♂ 20♀	20♂ 60♀	20♂ 40♀	20♂ 20♀	20♂ 60♀	20♂ 40♀	20♂ 20♀
Male family size									
1000	6	5	5	5	6	5	6	6	5
1500	7	6	6	6	7	7	6	7	6
2000	7	7	7	6	7	7	7	7	7
Female family size									
1000	0	1	1	0	3	5	0	2	5
1500	1	6	6	1	7	6	0	7	6
2000	7	7	7	8	7	6	7	7	7

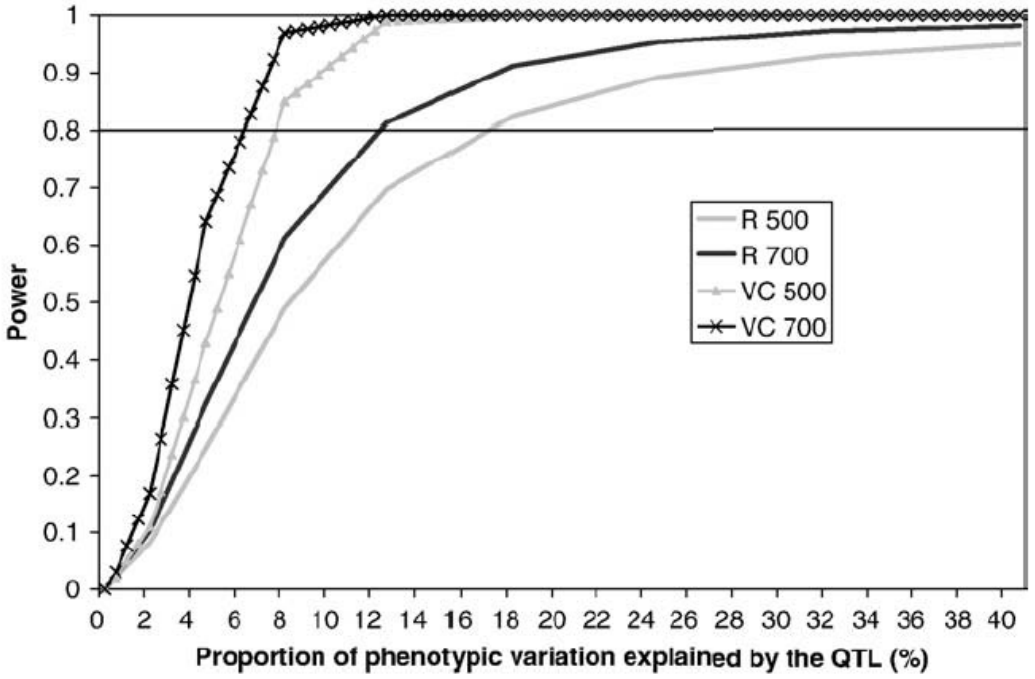

Figure 4 - Power to detect QTL for the mass-spawning breeding design. The letter R respresents the half-sib regression method and VC means variance component analysis. 500 or 700 indicate the number of individuals in the pedigree devided into paternal half-sib of 100 offspring each. The underlying full-sib strucure is taken into account by the VC method.

Large full-sib family design

Figure 5 shows that a QTL explaining 1.44% of the phenotypic variation using a genotyped fraction of 0.1 was expected to be detected with 80% power, while using a genotyped fraction of 0.2 required a QTL that explains 2.1% of the phenotypic variation for the same expected power. For a genotyped fraction of 0.5 we reverted back to scenario 1 of the salmonid design, where no selective

genotyping was applied. Therefore, for the same number of individuals genotyped, measuring a larger number of individuals and genotyping a smaller fraction of the total measured individuals was more powerful for the detection of QTL with small effect (explaining less than 5% of phenotypic variation). For a QTL of large effect (proportion of variance explained by the QTL >10%), the power was high, no matter what the genotyped fraction. With a genotyped fraction of 0.1 and a low overall heritability of 0.1 (scenario 2), a QTL that explains 1.8% of the phenotypic variation would be detected with an expected 80% power. When the heterozygosity was low (scenario 5), the power to detect QTL was much decreased; for a QTL effect as large as 12%, the maximum power was 65% for the parameters considered in this study.

2.5. Discussion

The commonality among the results of all the designs explored is that having fewer larger families is more powerful for QTL detection than using a large number of families with fewer offspring (Kolbehdari *et al*. 2005). The power to detect QTL for mass-spawning species when family distribution can be estimated is relatively high, when QTL heritability and QTL heterozygosity are estimated. When possible, selective genotyping is a useful method to increase power of QTL detection for a fixed genotyping cost.

Half-sib regression and variance component powers

We used two methods to calculate power according to the breeding design. The method developed by van der Beek et al. (1995) and revised by Bovenhuis and Spelman (2000) for selective genotyping is a deterministic method where

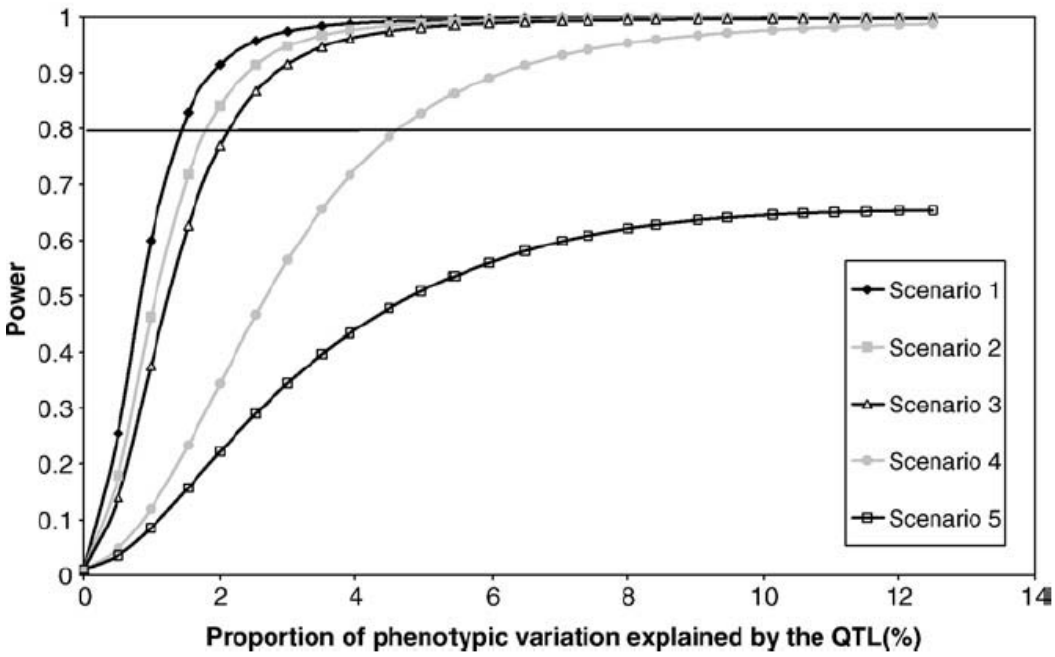

Figure 5 - Power of the 5 scenarios with various genotyped fractions and heritability and heterozygosity for the large full-sib design (see Table 2 for scenario description) as a function of the percentage of phenotypic variation explained by the QTL, with a fixed number of genotyped individuals (1000).

parameters such as heritability of the trait, marker spacing, heterozygosity and family structure are assumed known because we have some control over them. The values can be plugged directly into the formulae to obtain the power. However, this method does not give any standard error of the estimated power. The latter can be obtained only through extensive stochastic simulations. The method used for mass-spawners, derived from Williams and Blangero (1999), is slightly more stochastic in the sense that it uses actual pedigrees and simulates random QTL alleles in pedigree founders. The heritability of the trait and QTL

heritability have to be assumed, the QTL effect and the allele frequency remain unknown. This method does not require marker spacing information because it assumes that there is an infinite number of markers and therefore the QTL co-locates with a marker. Another assumption is that all founders are heterozygous for the QTL and therefore that all families contribute to the QTL variance, a consequence of the random model. These assumptions will lead to an overestimation of the power of the experiment. This is clearly illustrated by the results in Figure 4 where we compared the variance component and half-sib regression methods for the sea bream example. An explanation for this large difference between the two methods is that the regression methods assume only half-sib families, while the pedigree given for variance component shows a number of full-sib relationships that are intertwined within the half-sib relationships, and that taking account of these relationships increases the information and the chance to discover QTL. The power of half-sib regression and variance component approaches was studied through extensive stochastic simulation by Kolbehdari et al. (2005). Kolbehdari et al. (2005) also showed that variance component analyses were more powerful, but the differences were much smaller than in the present study. This can most likely be attributed to the fact that Kolbehdari et al. (2005) only studied half-sib populations, with no underlying maternal sib structures. We recommend using the full- or half-sib regression method as a starting point when there are for hierarchical and large full-sib families designs (e.g., Atlantic salmon, rainbow trout, oysters) and the variance component method for species such as sea bream or sea bass, which show a mix of full- and half- sib families. If possible, we recommend applying more than a single QTL mapping method in order to quantify the robustness of the given QTL results in the light of the different assumptions underlying the different QTL methods.

Effect of design on power

Clearly, family structure is an important variable in experimental design. Similar conclusions have been drawn in other studies on the effect of family structure on power to detect QTLs (van der Beek *et al*. 1995). The experiments carried out by Kolbehdari et al. (2005) showed that power is more affected by increasing the number of the progeny than it is by increasing the number of families.

Heterozygosity of the QTL has a major impact on QTL detection. The current power calculations do not allow a clear assessment of the risk of missing QTLs as a result of lack of heterozygosity; as a result, these approaches will always favour fewer larger families over a larger number of smaller families. The method of van der Beek et al. (1995) does model the heterozygosity of the parents, but the resulting power estimate did not reflect the difference between the expected heterozygosity and the real heterozygosity. For a design comprising a large number of families with only a few progenies each and also likewise a few large families, if the heterozygosity is low, the QTL will be segregating only for a limited number of individuals and will reduce the amount of information (Hayes *et al*. 2006, Verhoeven *et al*. 2006). However, selecting only a few large families carries the risk that none of the parents is heterozygous for the QTL and therefore the balance between family size and number of families needs careful consideration.

For species where there is some control in breeding, families can be generated from a cross between two divergently-selected lines, which increases the heterozygosity for QTL that has been under selection and therefore gives an advantage over species with low breeding control. Perry et al. (2001) show a

good example where divergently-selected lines were used to map QTL for upper thermal tolerance in rainbow trout. Populations with controlled breeding often also have better prior estimates for parameters such as heritability. For species with low breeding control, estimating heritability and heterozygosity is very difficult. When performing mapping experiments using fish derived from mass-spawning, it is necessary to characterize the population or use extensive simulations to determine how many fish will have to be sampled to provide sufficient family sizes for QTL mapping.

Marker spacing effect

The interval between two markers chosen here is rather large (20 cM). Kolbehdari et al. (2005) have shown that increasing the interval decreases the power. Some genetic maps in aquaculture have an average distance between markers much lower than 20 cM (about 5 cM for sea bass, (Chistiakov *et al*. 2005) and 9 cM for oysters, (Hubert and Hedgecock 2004)). The power calculations for rainbow trout showed that it is not really necessary to have intervals smaller than 20 cM. Because not all markers are expected to be informative in outbred populations; average spacing between informative markers of 20 cM may well require the genotyping of markers at 10 or 15 cM intervals.

Selective genotyping

Literature on selective genotyping predicts that selective genotyping does not affect the power to detect QTL when the genotyping is limited to 50% of the total population (genotyping 25% of each extreme phenotype) (Sen *et al*. 2005). For this proportion of genotyping, only one-eighth of the information provided by individual genotypes is lost if the marker density of the genetic map is

sufficient (Sen *et al*. 2005). The combination of interval mapping and selective genotyping can reduce up to 7-fold (depending on the marker density) the number of progeny to be genotyped (Lander and Botstein 1989). If the sample size is not large enough, selective genotyping will result in loss of power (Rabbee *et al*. 2004). As shown in the results of this study and other experiments (Darvasi *et al*. 1993, Chatziplis and Haley 2000), genotyping fewer individuals from large full-sib families (high breeding control) is very powerful.

Selective genotyping can reduce the overall cost of genotyping for a given power, but at the cost of increasing the number of animals measured for the phenotypic trait. For terrestrial farm animals, for which data have been collected over a period of years for routine breeding purposes, phenotyping should not be too expensive. Phenotyping, depending on the trait, can be expensive and time consuming (Zhao *et al*. 1997). If recording measurement for meat quality, number of individuals (relatives of the selection candidate) has to be significantly increased. Therefore, a balance between the cost of genotyping and phenotyping has to be established for each species studied. Selective genotyping is not an option if several traits are involved in the experiment, but is powerful when focussing on a single trait.

Mass-spawning family structures

One of the characteristics in reproductive behaviour of mass- spawning species is the unequal contribution of parents to the progeny. The model for the sea bream population simulation used in this study was a model based on sire and dam contributions found for a natural-spawning population studied by Brown et al. (2006).

In the present “mass-spawning” model used here, we assumed that all progenies

were assigned to both dam and sire with certainty. However, this is an idealistic situation. The various methods used for assigning parents to offspring do not provide a 100% confidence of correct assignment, but usually leave between 1 and 5% of wrong assignments and no assignments. Such pedigree mistakes will be identified when animals are genotyped for many more markers. However, this will result in a waste of resources as well as a very small loss of power. Erroneous or no parentage assignment can also introduce bias in quantitative genetic estimates like heritabilities and genetic correlations.

Both males and females have unequal contributions, but male contributions to progeny tend to be more variable. The population simulated represents a situation where eggs from females with variable contributions are collected over a long period of time (e.g. a month) and where dominant males have a higher probability to fertilize eggs, which corresponds to common hatchery method. Female spawning has been recorded as asynchronous in sea bream (Brown *et al.* 2006), cod (Bekkevold *et al*. 2002) and many other species. Egg collection frequency has a consequence on the power of QTL detection because this affects the pedigree structure. If eggs are collected on a single day, many may come from a single female. The predicted power remains high because the pedigree has a large maternal half-sib family with nested paternal full-sib families, as well as few small half-sib families. However, with a single large maternal half-sib family, the chance effects with regard to segregation of the QTL predominate; e.g. even if the QTL heterozygosity is 0.5, 50% of individual females will not be segregating for the QTL. For most QTL, we would expect the heterozygosity to be less than 0.5, with commensurate reduction in the chance of detection.

Family-based QTL mapping and linkage disequilibrium mapping

This paper describes QTL mapping designs using linkage analysis. Today, high-throughput SNP technologies allow the application of various other methods, such as linkage disequilibrium analysis (association studies) (Korol *et al*. 2007). However, the linkage analysis presented remains relevant for aquaculture because only a few species like salmonids, tilapias and channel catfish have the molecular resources required for such approaches. With new species being selected for aquaculture continuously, we expect linkage approaches to remain relevant for some time to come. The linkage disequilibrium analysis does present some advantages because they do not require known family relationships. One drawback of the family-based linkage analysis is that detected QTLs are specific for the families in which they are found and not for the entire population.

2.6. Conclusion

This study shows that QTL experiments with 80% power for QTL detection can be carried out for most aquaculture species where pedigree structure can be controlled or at least resolved. The goal was to provide some yardsticks for QTL experimental designs for fish and shellfish species that have not yet been studied extensively (unlike salmonids or tilapias). Strengths and limitations of designs were identified for different levels of breeding control. Family structure is one factor to consider with priority because of its major effect on power. Selective genotyping is recommended for species with high fecundity where control of breeding is possible, such as oysters, to improve power to discover QTL with small effects and cut the cost of the experiment where a single trait is of main interest. Classical QTL mapping designs (experimental crosses between

genetically divergent lines) are limited for mass-spawning species due to low breeding control and poor knowledge of those species. Stochastic population simulations using realistic parental contributions can be used to plan a QTL study for a mass-spawning species.

Acknowledgments

This study has been carried out for the AQUAFIRST project funded by the E.U. (E.U. contract number 513691). DJK & CSH acknowledge financial support from the Biotechnology and Biological Science Research Council (U.K.).

References

Andersson, L., Georges, M., (2004). Domestic-animal genomics: deciphering thegenetics of complex traits. Nat. Rev. Genet. 5, 202–212.

Bekkevold, D., Hansen, M.M., Loeschcke, V., (2002). Male reproductive competition in spawning aggregations of cod (Gadus morhua, L.). Mol. Ecol. 11, 91–102.

Bovenhuis, H., Spelman, R.J., (2000). Selective genotyping to detect quantitative trait loci for multiple traits in outbred populations. J. Dairy Sci. 83, 173–180

Brown, R.C., Woolliams, J.A., McAndrew, B.J., (2006). Factors influencing effective population size in commercial populations of gilthead sea bream, Sparus aurata. Aquaculture 247, 219–225.

Chatziplis, D.G., Haley, C.S., (2000). Selective genotyping for QTL detection using sib pair analysis in outbred populations with hierarchical structures. Genet. Sel. Evol. 32,547–560.

Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G.,Bertotto, D., Libertini, A., Volckaert, F.A., (2005). A microsatellite linkage map of the European sea bass Dicentrarchus labrax L. Genetics 170, 1821–1826.

Darvasi, A., Weinreb, A., Minke, V., Weller, J.I., Soller, M., (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134, 943–951.

Fessehaye, J., El-Bialy, Z., Resk, M., Crooijmans, R., Bovenhuis, H., Komen, H., (2006).Mating systems and male reproductive success in Nile tilapia (Oreochromis niloticus) in breeding hapas: a microsatellite analysis. Aquaculture 256, 148–158.

Gjedrem, T., Andersen, O., Bentsen, K.B., Fjalestad, K.T., Gjerde, B., Hayes, B.J., Kolstad, K.,Meuwissen, T.H.E., Olesen, I., Refstie, T., Rye, M., Sonesson, A., Thodesen, J.,Woolliams, J.A., (2005). Selection and Breeding Programs in Aquaculture. Springer,Dordrescht, The Netherlands, p. -364.

Hayes, B.J., Gjuvsland, A., Omholt, S., (2006). Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. Heredity 97, 4–18.

Hubert, S., Hedgecock, D., (2004). Linkage maps of microsatellite DNA markers for the Pacific oyster Crassostrea gigas. Genetics 168, 351–362.

Kolbehdari, D., Jansen, G.B., Schaeffer, L.R., Allen, B.O., (2005). Power of QTL detection by either fixed or random models in half-sib designs. Genet. Sel. Evol. 37, 601–614.

Korol, A., Shirak, A., Cnaani, A., Hallerman, E.M., (2007). Detection and analysis of quantitative trait loci (QTL) for economic traits in aquatic species. In: Liu, Z. (Ed.), Aquaculture Genome Technologies. Blackwell Publishing, Oxford, pp. 169–197.

Lander, E.S., Botstein, D., (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.

Lebowitz, R.J., Soller, M., Beckmann, J.S., (1987). Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor. Appl. Genet. 73, 556–562.

Martinez, V., (2007). Marker-assisted selection in fish and shellfish breeding schemes. Guimarães, E., Ruane, J., Scherf, B., Sonnino, A., Dargie, J. (Eds.), Marker- assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish. FAO, Rome, pp. 309–328.

Martinez, V.A., Hill, W.G., Knott, S.A., (2002). On the use of double haploids for detecting QTL in outbred populations. Heredity 88, 423–431.

Perry, G.M.L., Danzmann, R.G., Fergusson, M.M., Gibson, J.P., (2001). Quantitative trait loci or upper thermal tolerance in outbred strains of rainbow trout (Oncorhynchus mykiss). Heredity 86, 333–341

Porta, J., Porta, M., Martinez-Rodriguez, G., Alvarez, M.C., (2006). Genetic structure and genetic relatedness of a hatchery stock of Senegal sole (Solea senegalensis) inferred by microsatellites. Aquaculture 251, 46–55.

Rabbee, N., Speca, D., Armstrong, N.J., Speed, T.P., (2004). Power calculations for selective genotyping in QTL mapping in backcross mice. Genet. Res. Camb. 84, 103–108.

Sen, S., Satagopan, J.M., Churchill, G.A., (2005). Quantitative trait locus study design from an information perspective. Genetics 170, 447–464.

Sonesson, A.K., (2007a). Possibilities for marker-assisted selection in aquaculture breeding schemes. In: Guimarães, E., Ruane, J., Scherf, B., Sonnino, A., Dargie, J. (Eds.), Marker-assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish. FAO, Rome, pp. 309–328.

Sonesson, A.K., (2007b). Within-family marker-assisted selection for aquaculture species.Genet. Sel. Evol. 39, 301–317. van der Beek, S., van Arendonk, J.A.M., Groen, A.F., (1995).

Power of two- and three- generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. Theor. Appl. Genet. 91, 1115–1124.

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Gela, D., Vallod, D.,Linhart, O., Chevassus, B., (2005). Heritability estimates for growth- related traits using microsatellite parentage assignment in juvenile common carp (Cyprinus carpio L.).Aquaculture 247, 31–32.

Verhoeven, K.J., Jannink, J.L., McIntyre, L.M., (2006). Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96, 139–149. Weller, J.I., Kashi, Y., Soller, M., 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. 73, 2525–2537.

Weller, J.I., Kashi, Y., Soller, M., 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. 73, 2525–2537.

Williams, J.T., Blangero, J., (1999). Power of variance component linkage analysis to detect quantitative trait loci. Ann. Hum. Genet. 63, 545–563.

Xu, Z.L., Zou, F., Vision, T.J., (2005). Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. Genetics 170,401–408.

Yu, X.J., Knott, S.A., Visscher, P., (2004). Theoretical and empirical power of regression and maximum-likelihood methods to map quantitative trait loci in general pedigree.Am. J. Hum. Genet. 75, 17–26.

Zhao, H., Zhang, H., Rotter, J.I., (1997). Cost-effective sib pair designs in the mapping of quantitative trait loci. Am. J. Hum. Genet. 60, 1211–1221.

Chapter 3

QTLs for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*

Cécile Massault, Bart Hellemans,
Bruno Louro, Costas Batargias,
Jeroen Van Houdt, Adelino Canario,
Filip Volckaert, Henk Bovenhuis,
Chris Haley, Dirk-Jan de Koning
Animal Genetics (DOI: 10.1111/j.1365-2052.2009.02010.x)

3.1. Abstract

Natural mating and mass-spawning in the European sea bass (*Dicentrarchus labrax L., Moronidae, Teleostei*) complicate genetic studies and the implementation of selective breeding schemes. We utilised a two-step experimental design for detecting QTLs in mass spawning species: 2122 offspring from natural mating between 57 parents (22 males, 34 females and one missing), phenotyped for body weight, eight morphometric traits and cortisol levels, had been previously assigned to parents based on genotypes of 31 DNA microsatellite markers. Five large full-sib families (five sires and two dams) were selected from the offspring consisting of 570 animals, which were genotyped with 67 additional markers. A new genetic map was compiled, specific to our population, but based on the previously published map. QTL mapping was performed with two methods: half-sib regression analysis (paternal and maternal) and variance component analysis accounting for all family relationships. Two significant QTLs were found for body weight on linkage group 4 and 6, six significant QTLs for morphometric traits on linkage groups 1B, 4, 6, 7, 15 and 24 and three suggestive QTLs for stress response on linkage groups 3, 14 and 23. The QTLs explained between 8% and 38% of phenotypic variance. The results are a first step towards identifying genes involved in economically important traits like body weight and stress response in European sea bass.

3.2. Introduction

European sea bass (*Dicentrarchus labrax L., Moronidae, Teleostei*) is a marine fish, commonly distributed along the warm temperate coasts of the south-eastern Atlantic Ocean and Mediterranean Sea. While in 1980 only 10 tonnes were produced in semi-intensive aquaculture systems, production has stabilised since 2000 and the total worldwide production was an estimated 105,900 tonnes in 2008 (http://www.globefish.org). The value of sea bass aquaculture totalled 320 million Euros in 2007 at an average value of $5.1 per kilogramme. Greece is the main producer (44%), followed by Turkey (26%), Italy (12%) and Spain (8%), and sea bass represent 40% of Mediterranean aquaculture (www.feap.info).

Although sea bass aquaculture started 30 years ago, selective breeding is not widely used (Vandeputte *et al.* 2001). For a long time, natural mating and mass spawning were commonly used for reproduction. Artificial reproduction and mating are now fully controlled (Moretti 1999, Saillant *et al*. 2001, Dupont-Nivet *et al*. 2006) and have gradually become standard operational practice. In case of mass selection, the highly skewed unequal contribution of parents, so typical of mass-spawning fish (Jones and Hutchings 2002, Herlin *et al*. 2008), leads to inbreeding. Furthermore, the unknown pedigree of the individuals complicates selective breeding. However, the use of information from quantitative trait loci (QTL) is a first step towards selective breeding. Regions of the genome that are linked to a quantitative trait of interest are detected using QTL mapping.

The steady increase in genetic and molecular biological studies of European sea bass (for a review, see Volckaert *et al*. 2008) provides a platform to implement

breeding programs that do not depend solely on the measurement of phenotypic traits, but rely also on information from genetic markers, i.e., marker assisted selection . Large scale profiling of paternity with microsatellite markers has facilitated the breeding of families on an experimental and commercial scale in a reliable and affordable fashion (García de Léon *et al*. 1995, Chatziplis *et al*. 2007). From such communally bred families, quantitative genetic parameters have been reported for heritabilities and phenotypic and genotypic correlations for some traits, such as sex (Vandeputte *et al*. 2007), body weight and length (Vandeputte *et al*. 2004, Saillant *et al.* 2006, Chatziplis *et al.* 2007) and carcass traits (Saillant *et al*. 2009). Heritabilities for growth are relatively high and vary from 0.29 to 0.60 in European sea bass (Saillant *et al*. 2006, Dupont-Nivet *et al*. 2008). These high heritabilities have allowed a doubling in growth rate in just four generations (B. Chatain, pers. comm.). A medium density genetic map is available and has been updated (Chistiakov *et al.* 2005, Chistiakov *et al.* 2008), based on more than 200 microsatellites and more than 200 AFLPs (amplified fragment length polymorphism). The ESTs of several tissue-specific cDNA banks have been sequenced, analysed (A. Canario, pers. comm.) and used as a source of microsatellite (Chistiakov *et al*. 2008) and SNP markers (Souche 2009). At the same time, a QTL analysis for body weight has been performed in European sea bass (Chatziplis *et al.*, 2007) and has identified a QTL for growth on linkage group 1.

An important economical trait is stress, either linked to pathogen infection or behaviour (such as confinement and handling). Cortisol, a widely accepted proxy for stress in fish, affects the immune system response (Engelsma *et al*. 2002). Therefore, if genomic regions are found that influence the stress response, this could have a beneficial effect on the management of fitness. All

these features represent multiple steps towards the use of genetic and molecular tools for improvement of production.

Massault et al. (2008) (**Chapter 2**) suggested a two-step experimental design for QTL mapping in mass-spawning species with natural mating. In a first step, individuals are assigned to parents using a limited number of markers and in a second step a subset of large families is selected for a whole genome scan. So far, such a design has not been applied in practice to fish. The strategy has been implemented in a single study on European sea bass. Volckaert et al. (submitted) describe how the families for QTL analysis were established by assigning parentage among the progeny of a one-day batch spawning. Heritabilities for growth, morphometry and stress response were calculated on these data and are reported by Volckaert et al. (submitted). The aim of this study is to perform the QTL analysis on stress response, body weight and morphological traits.

3.3 Materials and methods

Experiment

We followed a two-step procedure for the stress experiment that takes into account the specificity of natural mating population, as suggested by Massault et al. (2008): (i) take a random sample of 2122 offspring, originating from a single day natural mating by mass spawning and assign parents using a limited number of genetic markers, (ii) select large families for further genotyping and QTL mapping.

Fifty-seven parents originating from commercial and wild lines made up the base population. Offspring were chosen at random and genotyped for 31 loci at

50 days of age. Based on these genotypes offspring were assigned to parents. Confinement stress was induced on all 2122 fish, 8 months old approximately, by reducing the volume of water to 0.2 $m^3.kg^{-1}$ of fish for 4 hours (for details see Volckaert *et al*., submitted)

Phenotyping

Following the stress experiment, individuals were phenotyped for ten traits: body weight, eight various morphometric traits and cortisol level (for full details see Volckaert et al., submitted; see Figure 1 and Table 1). About half of the offspring were stunned in the morning using icy water while the other half was stunned in the afternoon. After being stunned, each fish was bled, weighted and digitally photographed. The plasma, after separation from the red blood cells, was conserved at -20ºC. When photographed, fish were placed as laterally as possible to avoid shape variation, landmarks were placed at various points of the body after determining the common scale using ruler and coordinated axes (x, y and z). From those landmarks, morphometric traits were observed. The cortisol level was determined using radio immunoassay technology (RIA) based on antiserum raised in rabbit against cortisol-3-CMO-BSA and tritriated cortisol (Fitzgerald Industries International). RIA cross-reacted with only one plasma component co-migrating with cortisol on a layer chromatography. Plasma samples were diluted in phosphate buffer containing 0.5 $g.l^{-1}$ gelatine (pH 7.6) and heated at 80°. We extracted the cortisol according to the protocol of Scott et al. (1982), as detailed in Volckaert et al. (submitted).

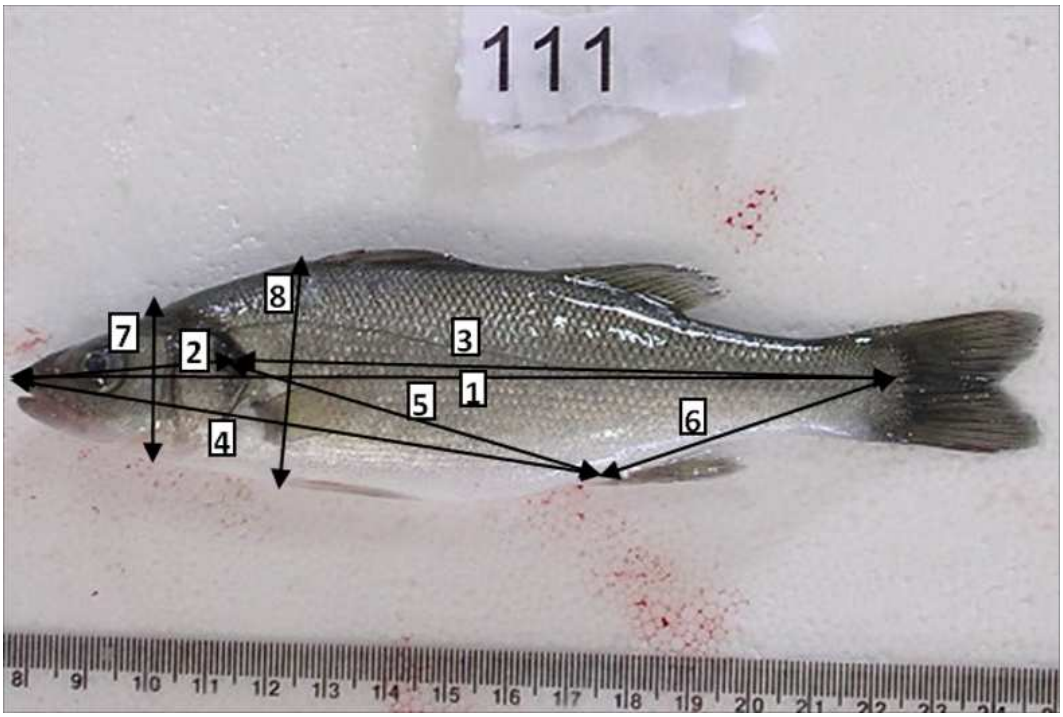

Figure 1 - Morphometric measurement on European sea bass. Numbers are related to the traits described in Table 1.

Genotyping and genetic map

Volckaert et al. (submitted) identified with a triple multiplex-PCR of 31 microsatellite markers 11 large full-sib families (n=922) from the 2122 offspring for heritability and correlation analysis (see above). Five of the eleven families (n=570 fish) and their seven parents were fully genotyped for 98 microsatellites markers (with 8 markers not fulfilling minimal quality requirements). The population structure for the QTL mapping includes the offspring from 5 males and 2 females; the number of offspring per family varies between 93 and 143 (Table 2). We generated a genetic map specific to our

Table 1 - Number of observations, mean, standard deviation, coefficient of variation and heritabilities (standard error) for ten traits and 570 European sea bass.

Trait	**Abbreviation****	**Number of individuals**	**Mean**	**SD**	**CV**	**h^2 * (±SE)**
Body weight	BW (-)	566	40.58	14.21	35%	0.54 (±0.2)
Cortisol	CORT (-)	444	319.5	136.40	43%	0.08 (±0.06)
1. Standard length	SL (SL)	540	13.26	1.61	12%	0.65 (±0.22)
2. Head length	HL (SNOP)	540	3.77	0.46	12%	0.61 (±0.21)
3. Body length	BL (OPCA)	540	9.50	1.17	12%	0.64 (±0.22)
4. Pre anal length	PrAnl (SNAN)	540	9.20	1.16	13%	0.68 (±0.23)
5. Abdominal length	AL (OPAN)	540	5.57	0.72	13%	0.66 (±0.23)
6. Post anal length	PsAnl (ANCA)	540	4.46	0.54	12%	0.52 (±0.20)
7. Head depth	HD (-)	540	2.46	0.33	13%	0.64 (±0.22)
8. Body depth	BD (DOPV)	540	3.38	0.41	12%	0.56 (±0.21)

* heritabilities calculated from 922 fish (including QTL population) in Volckaert *et al.* (submitted)

** Abbreviation according to Chatziplis et al., 2007

population using the software CRI-MAP v.2.4. (http://compgen.rutgers.edu/multimap/crimap).

QTL mapping

We used two methods to detect QTLs: half-sib regression analysis as described by Knott et al. (1996), available on the web (http://www.gridqtl.org.uk, Hernandez and Knott 2009), and variance component analysis as described by George et al. (2000) (http://qtl.cap.ed.ac.uk/puccinoservlets/hkloaderLoki). The model used for half-sib analyses is as follows:

$$y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij}$$

where y_{ij} is the offspring phenotype, α_i the mean for family i, β_i the regression coefficient for family i, x_{ij} the probability to inherit parental allele 1 conditional on the marker information and e_{ij} the residual. Information content was calculated as described by Knott et al. (1996). The QTL effect is estimated within families as the allele substitution effect. Knott et al. (1996) provide a method to calculate the QTL effect across families in terms of variance due to the QTL by looking at the difference between the residuals of full and reduced models:

$$QTL\ effect = 4 * (1 - (RES\ full/RES\ reduced))$$

where RES full and RES reduced are the residuals of full and reduced models respectively.

Regression was performed at each centiMorgan and the test statistic for presence of a QTL was calculated as described by Knott et al. (1996). We used permutation tests with 500 iterations and bootstrap analysis with 2000 iterations

to evaluate the significance and confidence intervals of detected QTLs, respectively. A QTL is considered significant if it exceeds the 5% genome-wide threshold and a QTL is considered suggestive if it exceeds the 5% chromosome-wide threshold.

For variance component analysis we used the genotypes and the linkage map to first calculate the identity-by-descent (IBD) matrix at each position of the genome with Loki (Heath 1997). In the analysis only linkage information on the transmission of alleles from parent to offspring was considered and IBD relationships between parental alleles were not considered (i.e., no linkage disequilibrium). The resulting IBD matrix is used to model a QTL in a linear mixed model using ASREML:

$$y = Xb + Zu + Zv + e$$

where y is the phenotype, X is the incidence matrix relating phenotypes to systematic environmental effects, b is the vector with solutions for systematic environmental effects, Z the incidence matrix relating animals to phenotype, u vector of additive polygenic effect, v vector of additive QTL effect and e vector of environmental effect ($\text{var}(u) = A\sigma^2 u$ and $\text{var}(v) = G\sigma^2 v$, with A the additive relationship matrix and G the IBD matrix). For CORT the environmental effect of sampling time was included. The sampling time corresponds to a half-day, resulting in three classes: first day morning, first day afternoon and second day morning. For the other traits the only other systematic effect was the mean. The QTL effect is given by the variance component analysis as the proportion of phenotypic variance due to the QTL.

Table 2 - The structure of the QTL population of European sea bass. Five males were mated with two females, providing full-sib families and paternal and maternal half-sib families.

Half-sib ID	***Full-sib ID***	***Dam ID***	***Sire ID***	***# offspring***
1	1	2	1	98
1	2	2	3	93
2	3	6	4	92
2	4	6	5	143
2	5	6	7	142

The null hypothesis model is:

$$y = Xb + Zu + e$$

With the same variable definitions as for the QTL model. The presence of a QTL, treated as a random effect, was tested at every centiMorgan using the likelihood ratio test:

$$LRT = 2*(lf - lr)$$

where LRT is the log likelihood ratio test, *lf* the log likelihood of the full model i.e. the model including the QTL and *lr* the log likelihood from the reduced model i.e. the model without a QTL. Under the null hypothesis the likelihood ratio test follows a chi-square distribution with one degree of freedom. For

VCA, we computed 1% chromosome-wide thresholds according to the method described by Piepho (2001). The thresholds vary as a function of the chromosome length and the trait studied.

3.4 Results

Summary statistics

Heritabilities for each trait (listed in Table 1 and including abbreviations) were calculated for 922 animals, including the QTL population (Volckaert *et al*., submitted), while all other statistics were derived from the actual QTL population (n=570). CORT heritability is 0.08, while heritabilities of the others traits ranged from 0.52 to 0.68. We note that all morphometric traits (SL, HL, BL, PRAL, AL, POAL, HD and BD) have a coefficient of variation (CV) within the same range (12-13%), while BW and CORT are more variable (CV of 35% and 43%, respectively). Morphometric traits are highly positively correlated among themselves and with BW (see also Volckaert *et al*., submitted).

Genetic map

Based on the marker genotypes collected on 570 offspring and their parents we built a genetic map containing 20 linkage groups and covering 639 cM (centiMorgans; Figure 2). Out of the 90 markers, 87 were located in linkage groups consisting of two or more markers. Three markers were unlinked to any of the other markers (linkage group LG9, LG18 and LG25).

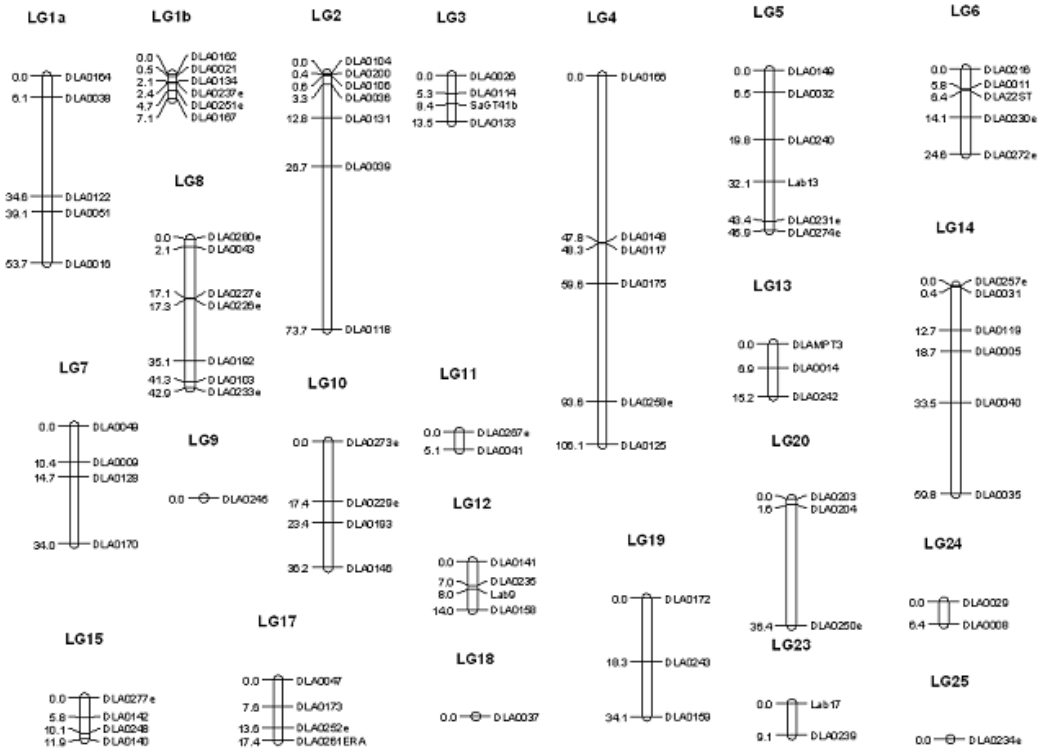

Figure 2 - Genetic map of QTL mapping sea bass population. Three linkage group LG16, LG21 and LG22 from Chistiakov et al. (2005) are missing as no markers from those groups are present.

The average marker spacing is 7.70 cM. The LG1 was cut into two parts (1a and 1b) because of a gap >100 cM. LG16, LG21 and LG22 are not represented in this genetic map as no markers were located on them.

Information content

Table 3 summarizes the average information content (IC) of each linkage group for both paternal (PHS) and maternal (MHS) half-sib analysis, as well as their

number of markers, average number of alleles and their length. The average information content of the genome amounts to 0.77 for PHS and 0.81 for MHS: it varied from 0.42 to 0.96 for PHS and 0.64 to 0.97 for MHS. We observed an average IC lower than 0.5 on LG2 and LG4 for both PHS and MHS. Another region on linkage group LG20, had low IC for PHS only.

QTL detected

Variance component analysis and regression analysis were performed for each trait for the whole genome (on the 20 linkage groups). Regression analysis was divided into PHS and MHS analysis (Table 4 – Figure 3). Significant QTLs for regression analysis were found when the F-statistics is above the 5% genome-wide threshold and suggestive when above 5% chromosome-wide thresholds, given by GridQTL and significant when log likelihood (LR) is above a 1% chromosome-wide thresholds.

There were two BW QTLs on LG4 and on LG6. The one on LG4 is present and highly significant in all analyses, although at various positions on the chromosome (F = 5.65 for PHS with significance thresholds of 5% CW of 3.10 and 5% GW of 4.48, F = 7.09 for MHS with significance thresholds of 5% CW of 3.76 and 5% GW of 7.02 and LR = 5.78 for VCA with a significance threshold of 1%CW of 13.5). It was not possible to estimate the variation due to the BW QTL on LG4 detected with VCA, as the likelihood converged at the boundary of the parameter space, i.e. at a heritability of that QTL of 0.99, but the other analyses indicate a large QTL effect on this trait. Another large BW QTL was detected on LG6 with the MHS method (F = 8.59 with 5% CW of 3.82 and 5% GW of 7.02).

Table 3- Number of markers, average number of alleles per marker, length of the linkage group, average information content for paternal and maternal half-sib regression analysis (PHS and MHS) of European sea bass

LG	# markers	Average # alleles	length	Avg IC PHS	Avg IC MHS
LG1a	5	5.2	53	0.80	0.76
LG1b	6	4.3	7	0.96	0.97
LG2	7	3.9	73	0.70 [1]	0.68 [4]
LG3	4	3.5	13	0.83	0.81
LG4	6	3.7	106	0.59 [2]	0.64 [5]
LG5	6	4	45	0.88	0.91
LG6	5	3.8	24	0.81	0.80
LG7	4	6	34	0.89	0.84
LG8	7	4.3	42	0.85	0.88
LG10	4	4.3	36	0.83	0.80
LG11	2	3.5	5	0.86	0.67
LG12	4	5.5	14	0.92	0.90
LG13	3	5	15	0.92	0.87
LG14	6	3.8	59	0.79	0.77
LG15	4	3	11	0.82	0.81

[1] average of 0.49 between 50 and 46 cM

[2] average of 0.27 between 0 and 34 cM

3 average of 0.32 between 12 and 36 cM

[4] average of 0.46 between 44 and 61 cM

[5] average of 0.37 between 0 and 33 cM

Table 3 – (cont.) Number of markers, average number of alleles per marker, length of the linkage group, average information content for paternal and maternal half-sib regression analysis (PHS and MHS) of European sea bass.

LG	# markers	Average # alleles	length	Avg IC PHS	Avg IC MHS
LG17	3	4.8	17	0.89	0.92
LG19	3	4	34	0.82	0.67
LG20	3	3	36	0.42 [3]	0.62
LG23	2	4.5	9	0.88	0.92
LG24	2	7	6	0.93	0.97

[1] average of 0.49 between 50 and 46 cM

[2] average of 0.27 between 0 and 34 cM

3 average of 0.32 between 12 and 36 cM

[4] average of 0.46 between 44 and 61 cM

[5] average of 0.37 between 0 and 33 cM

There was a large number of highly significant morphology QTLs, in particular for the PHS analysis. A common QTL is found in LG4 by all analyses, but again at a different position (F = 4.76 for PHS with 5% CW of 2.90 and 5% GW of 4.18, F = 10.31 with 5% CW of 4.37 and 5% GW of 6.98 and LR = 15.46 for VCA with 1%CW of 13.03). VCA showed a large proportion of phenotypic variation explained by the QTL (38%), while this proportion is less for PHS and MHS (13% and 14% respectively). Another morphology QTL was detected with VCA on LG6 with a QTL effect of about 9.4% (LR = 22.79 with 1% CW of 12.07). The other morphology QTLs found were located on four

different chromosomes and discovered with PHS analysis (F between 4.36 and 4.96 with 5% CW of 2.41-2.69 and 5% GW of 4.18). The QTL effects explain between 12% and 16% of phenotypic variation.

Three QTLs for CORT were detected with PHS (LG3 and LG14, F = 2.69 with 5% CW of 2.70 and F = 3.01 with 5% CW of 2.74 respectively) and MHS (LG23, F = 6.59 with 5% CW of 3.25), but were 5% significant chromosome-wide, therefore considered as suggestive QTLs. None were detected with VCA. The proportion of phenotypic variation explained by the QTLs varies between 8% and 10%.

Morphology and BW QTLs on LG4 were detected in the PHS as well as in the MHS analysis. The most likely location of the QTL was different in the two analyses: 55 cM in the HS analysis compared to 0 cM in the MHS analysis. However, the confidence intervals for QTL position are very large and overlap indicating that this might represent the same QTL. We can observe an interesting phenomenon between BW and morphological traits QTLs; there are three BW-morphology QTLs, one for PHS on LG4 and two for MHS on LG4 and LG6. For each pair of QTL, we notice that both BW and morphology QTLs are located at the same position of the linkage group. Body weight and all morphometric traits in this study are highly correlated, which might explain the observed link between both QTLs (Volckaert *et al.*, submitted).

Table 4. Body weight, morphological and cortisol QTL for paternal half sib (PHS), maternal half-sib (MHS) and variance component analysis (VCA) of European sea bass

Trait	Method	Position in cM	# segregating parents	Linkage Group	F	5%CW (HS)	5%GW (HS)	LR	1%CW (VCA)	QTL effect [1] (%)	Confidence Interval
BW	PHS	54	3	4	5.65	3.10	4.48	-	-	16%	1-88 cM
BW	MHS	0	2	4	7.09	3.76	7.02	-	-	8%	0 - 76 cM
BW	VCA	3	-	4	-	-	-	5.78	2.23	-	NA[2]
BW	MHS	12	2	6	8.59	3.82	7.02	-	-	10%	5 – 22 cM
MORPH	PHS	55	3	4	4.76	2.90	4.18	-	-	13%	1 – 105 cM
MORPH	MHS	2	2	4	10.31	4.37	6.98	-	-	14%	0 – 4 cM
MORPH	VCA	2	-	4	-	-	-	15.46	13.03	38 %	NA[2]
MORPH	VCA	19	-	6	-	-	-	22.79	12.07	9.4 %	NA[2]
MORPH	PHS	5	2	1B	4.36	2.41	4.18	-	-	14%	1 – 7 cM
MORPH	PHS	30	3	7	5.71	2.60	4.18	-	-	16%	0- 34 cM
MORPH	PHS	0	1	15	4.21	2.56	4.18	-	-	12%	0 – 11 cM
MORPH	PHS	0	2	24	4.93	2.50	4.18	-	-	13%	0 – 6 cM
CORT	PHS	3	2	3	2.69	2.70	4.59	-	-	8%	0 – 13 cM
CORT	PHS	1	2	14	3.01	2.74	4.59	-	-	9%	0 – 59 cM
CORT	MHS	3	1	23	6.59	3.25	7.61	-	-	10%	0 – 9 cM

BW Body weight, CORT cortisol level; MORPH combination of all morphometric traits HD, HD, BD, BL, AL, PRAL, POAL and SL
PHS paternal half-sib analysis; MHS maternal half-sib analysis; VCA variance component analysis
[1] QTL effect: proportion of the phenotypic variation explained by the QTL
[2] Not applicable

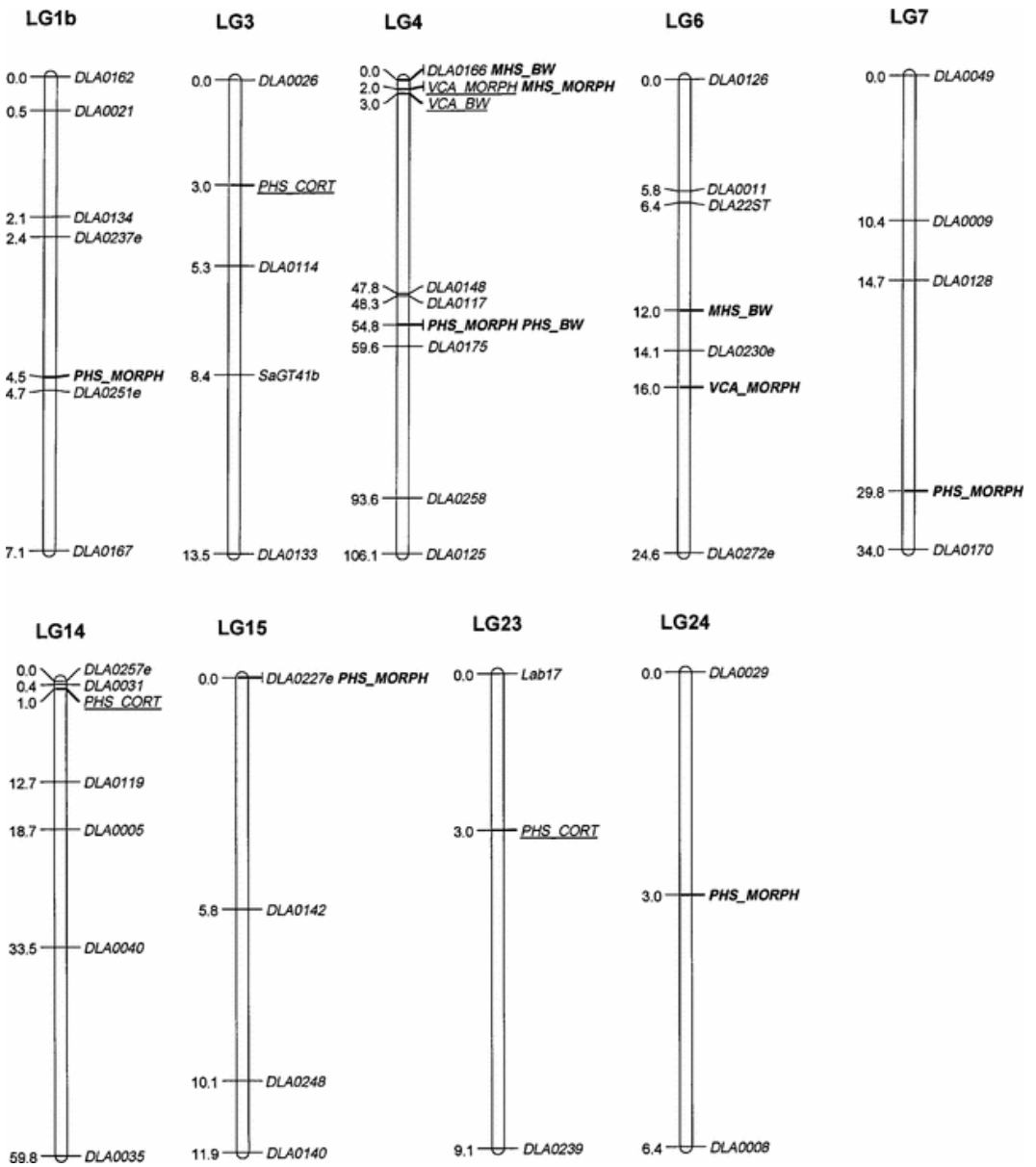

Figure 3 - QTLs in bold are 5% genome wide while the underlined one are 5% chromosome-wide. MHS stands for maternal half-sib regression, PHS for paternal half-sib regression and VCA for variance component analysis. The traits are MORPH for morphology and BW for body weight.

Morphology and BW QTLs on LG4 were detected in the PHS as well as in the MHS analysis. The most likely location of the QTL was different in the two analyses: 55 cM in the HS analysis compared to 0 cM in the MHS analysis. However, the confidence intervals for QTL position are very large and overlap indicating that this might represent the same QTL. We can observe an interesting phenomenon between BW and morphological traits QTLs; there are three BW-morphology QTLs, one for PHS on LG4 and two for MHS on LG4 and LG6. For each pair of QTL, we notice that both BW and morphology QTLs are located at the same position of the linkage group. Body weight and all morphometric traits in this study are highly correlated, which might explain the observed link between both QTLs (Volckaert *et al.*, submitted).

3.5 Discussion

Linkage map and information content

The genetic map that we built resembles closely the previous genetic maps published by Chistiakov et al**.** (2005, 2008), with the difference that 23 new markers were added to our map. Although it is specific to our QTL population, the similarities between both of them show consistency and relevance of our results. The map is of medium density with an average gap of about 8 cM, which is sufficient for QTL mapping according to Chistiakov et al**.** (2005).

The low IC detected in some regions on the linkage groups may impact the results, in particular for the QTLs detected on LG4, which are numerous. Results become less reliable as the extremity of the chromosome is not very informative.

Identification of QTLs

We found a total of 15 different QTLs using the three analyses summarized in Table 4. The results found with VCA are similar to those found with MHS regression analysis. BW and morphology QTLs were discovered at the same linkage groups and at similar positions; those traits were found to be highly positively correlated. Morphology QTL on LG4 explains 38% of phenotypic variation, but some of the effect might be due to the actual BW QTL on the same linkage group. PHS detected more QTLs, especially for the morphological trait than MHS and VCA. PHS and MHS differ in the number of detected QTLs and in their position on the genome. More QTLs were discovered for PHS. There are only two female parents included in the analysis and this might have an impact on QTL discovery in MHS analysis.

Potential QTLs might not have been detected, because the only two females in the parental population did not segregate for it. The 95% confidence intervals are very large for nearly all QTLs, covering the entire length of chromosomes. Fine mapping, which is the next logical step in this process, has to use a denser genetic map to be able to pinpoint interesting genes, but our results are precise enough for a first QTL analysis. For fine mapping, the use of more markers may help reduce the confidence intervals and narrow the interesting regions that we found. At the moment, only microsatellites are available, but development of AFLPs (Chistiakov *et al*. 2008) and soon SNPs may accelerate the process of fine mapping. Although the increase in the number of markers influences the precision of the results, population structure has an impact on the power to detect QTLs. In this study, we have a mixed population of half-sib and full-sib. To have the best results, we selected the largest families. But due to the unequal contribution of parents, the largest families will represent only very few parents.

The number of QTLs found with the variance component analysis is lower than with half-sib regression. This might be due to the approximate 1% chromosome-wide thresholds calculated with Piepho's method (Piepho 2001). Those thresholds are adjusted for the marker density of the genetic map.

Chatziplis et al**.** (2007) mention a QTL associated with morphometric traits located at the beginning of LG1; we also detected a morphology QTL on LG1. The two mapping populations differ, although they belonged to the same company. We split this linkage group into LG1a and LG1b because of a large region in the middle of LG1 devoid of markers. Our QTL is located at the end of the linkage group, in the LG1b section. The high confidence intervals suggest that the location of the QTL could be anywhere on the chromosome, and therefore this result potentially confirms the results of Chatziplis et al**.** (2007). They also mention a possible BW QTL on LG1, but we could not support this finding. No other QTLs were reported in European sea bass, suggesting that we found novel significant and suggestive QTLs for this species.

We found three suggestive QTLs for stress response. Although they are not highly significant, it is an important discovery because QTLs for stress response have not been reported before. A candidate gene for this trait could be the glucocorticoid receptor, which has been sequenced and whose position remains unknown (Terova *et al***.** 2005). Several genes related to carbohydrate metabolism, genes involved in Na^+/K^+ transport, polyamine biosynthesis, and iron homeostasis (Sarropoulou *et al.* 2005) could be also taken as candidate gene as they are involved into the mechanism of stress response. Our identified regions could be of high interest for candidate gene approach studies.

Methodological comparison

We compared half-sib regression and variance component analysis, each with its advantages and drawbacks. We performed separately a paternal half-sib regression analysis (PHS) and maternal half-sib regression analysis (MHS). Therefore, in the regression analysis the full-sib relationships are not taken into account. We additionally performed a variance component analysis (VCA) that deals with all family relations in our experimental population. Half-sib regression does not require large computing resources and time, and therefore it allows to perform a permutation test and calculate confidence intervals using bootstrapping.

Variance component analysis has been implemented to simultaneously account for all relations in our data. The QTL is modelled as a random effect in the model; therefore the QTL may have an infinite number of alleles. But VCA is computationally demanding, especially the calculation of IBD and thresholds for 5% chromosome-wide significance were computed according to an approximation given by Piepho (2001). Those calculations are time-consuming.

Power of the experimental design

This study has implemented the two-step procedure developed in Massault et al**.** (2008). When applying simulated parameters from our specific population (rounded to 600 individuals), we found that, for an 80% power, we were theoretically able to detect QTLs explaining at least 14.5% of phenotypic variation using half-sib regression analysis and at least 6.5% of phenotypic variation using variance component analysis. Morphology QTL on LG4 explains a large proportion of phenotypic variation (38%) and will have been detected with a less powerful experimental design. In case of paternal half-sib

regression, this QTL could be associated with marker DLA0175, while for MHS and VCA, it will be more linked to a marker at the beginning of the linkage group DLA0166. On the contrary, morphology QTL on LG6 might not have been detected as the QTL effect was 9.4% of the phenotypic variation (against 14.5% expected theoretically). It is a QTL of medium size that our design was powerful enough to detect; it a proof that the design is perfectly adapted for natural mating population. The proportion of phenotypic variation explained by QTLs detected with half-sib regression was of medium size effect (8%-16%). It shows that our method is able to detect not only QTLs with large effects but also medium effects for sea bass, and could be generalized to mass-spawning species.

The fact that there were only two female parents involved in the experiment may also have an impact on the success to detect QTLs. The heterozygosity for each allele is therefore none, half or one. As explained before, existing QTLs might be ignored because none of the females are segregating for it. The power therefore might be lower, as in theory we assumed heterozygosity of 0.5, which in reality is impossible to predict.

The theory was based on a sparse genetic map, with marker spacing of 20 cM. Our genetic map had average marker intervals close to 8 cM. A denser map could have improved the power, and therefore our experiment with a sparser map might not have detected some QTLs.

The power of the experiment depended on the heritability of the trait. We found three suggestive CORT QTLs for stress response. The heritability of this trait shown in Table 1 is 0.08. Massault et al**.** (2008) have set a high heritability of 0.5 for the two-step experimental design. Trait heritability is known to play a

role in QTL discovery (Kolbehdari *et al.* 2005) and in the case of low heritability such as cortisol, there is less chance of detecting important QTLs. Our study shows that QTL mapping is possible for natural mating mass-spawning species, using a specific experimental design and tools, such as parentage assignment. The results of this experiment are preliminary, but they are promising. QTLs were detected and could be located at specific chromosomes, so that complementary fine mapping can be undertaken.

Acknowledgments

This study has been funded by the European Commission (project AQUAFIRST, contract number FP6-STREP-2004-513692). DJK and CSH acknowledge financial support from the BBSRC. E. Couto, S. Darivianaku, J. Fuentes, P. Guerreiro, M. Kampakli, G. Kotoulas and D. Power helped with phenotyping on site.

References

Chatziplis, D., Bartagias, C., Tsigenopoulos, C.S., Magoulas, A., Kollias, S., Kotoulas, G., Volckaert, F.A.M., Haley, C.S. (2007) Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): the BASSMAP pilot study. *Aquaculture* **272**, S172-S182

Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A, Volckaert, F.A.M. (2005) A microsatellite linkage map of the European sea bass *Dicentrarchus labrax L*. *Genetics* **170**, 1821-1826

Chistiakov, D.A., Tsigenopoulos, C.S., Lagnel, J., Yuanmei, G., Hellemans, B., Haley, C.S., Volckaert, F.A.M., Kotoulas, G. (2008) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Animal Genetics* **39**, 623-634

Dupont-Nivet, M., Vandeputte, M., Haffray, P., Chevassus, B. (2006) Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture* **252**, 161-170

Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., Chatain, B. (2008) Heritabilities and GXE interactions for growth in European sea bass (*Dicentrarchus labrax* L.) using a marker-based pedigree. *Aquaculture* **275**, 81-87

García de León, F.J., Dallas, D.J., Chatain, B., Canonne, M., Versini, J.J., Bonhomme, F. (1995) Development and use of microsatellite markers in seabass, *Dicentrarchus labrax* (Linnaeus, 1758) (Perciformes: Serranidae). *Molecular Marine Biology and Biotechnology* **4**, 62 – 68

Engelsma, M.Y., Huising, M.O., van Muiswinkel, W.B., Flik, G., Kwang, J., Savelkoul, H.F.J., Verburg-van Kemenade B.M.L. (2002) Neuroendocrine-immune interactions in fish: a role for interleukin-1. *Veterinary Immunology and Immunopathology* **87,** 467-479

George, A., Visscher, P., Haley, C.S. (2000) Mapping quantitative trait loci for complex pedigree: a two-step variance component approach. *Genetics* **156**, 2081-2092

Heath, S.C. (1997) Markov chain Monte Carlo methods for radiation hybrid mapping. *Journal of Computational Biology* **4**, 505-515

Herlin, M., Delghandi, M., Wesmajervi, M., Taggart, J.B., McAndrew, B.J., Penman, D.J. (2008) Analysis of the parental contribution to a group of fry from a single day of spawning from a commercial Atlantic cod (*Gadus morhua*) breeding tank *Aquaculture* **274**, 218-224

Hernandez J., Knott S. (2009) Haplotyping via minimum recombinant paradigm. *BMC Proceedings/2009 **3**(Suppl 1), S7*

Jones M.W., Hutchings J.A. (2002) Individual variation in Atlantic salmon fertilization success: implications for effective population size. *Ecological applications* **12**, 184-193

Knott, S., Elsen, J., Haley C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71-80

Kolbehdari, D., Jansen G.B., Schaeffer, L.R, Allen, B.O. (2005) Power of QTL detection by either fixed or random models in half-sib designs. *Genetics Selection Evolution* **37**, 601-614

Massault, C., Bovenhuis, H., Haley, C.S., de Koning D.J. (2008) QTL mapping designs for aquaculture. *Aquaculture* **285**, 23-29

Moretti, A. (1999) Manual on hatchery production of seabass and gilthead seabream. Food and Agricultural Organisation, Rome

Piepho, H.P. (2001) A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics* **157**, 425-432

Saillant, E., Chatain, B., Forstier, A., Przybyla, C., Fauvel, C. (2001) Parental influence on early development in the European sea bass. *Journal of Fish Biology* **58**, 1585-1600

Saillant, E., Dupont-Nivet, M., Haffray, P., Chatain, B. (2006) Estimates of heritability and genotype-environment interactions for body weight in sea bass (*Dicentrarchus labrax* L.) raised under communal rearing conditions. *Aquaculture* **254**, 139-147

Saillant, E., Dupont-Nivet, M., Sabourault, M., Haffray, P., Laureau, S., Vidal, M.0., Chatain, B. (2009) Genetic variation for carcass quality traits in cultured sea bass (Dicentrarchus labrax). Aquatic Living Resources 22, 105-112

Sarropoulou, E., Kotoulas, G., Power, D.M., Geisler R. (2005) Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by application of cDNA microarray technology. *Physiological Genomics* **23**, 182-191

Scott, A., Sheldrick, E., Flint, A. (1982) Measurement of 17[alpha], 20[beta]-dihydroxy-4-pregen-3-one in plasma in trout (*Salmo gairdneri* Richardson): Seasonal changes and response to salmon pituitary extract. *General and Comparative Endocrinology* **46**, 444-451

Souche, E. (2009) Genomic variation in European sea bass *Dicentrarchus labrax* : from SNP discovery within ESTs to genome scan. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium

Terova, G., Gornati, R., Rimoldi, S. Bernardini, G., Saroglia, M. (2005) Quantification of a glucocorticoid receptor in sea bass (*Dicentrarchus labrax*, L.) reared at high stocking density. *Gene* **357**, 144-151

Vandeputte, M., Dupont-Nivet, M., Chatain, B., Chevassus B. (2001) Setting up a strain-testing design for the seabass, *Dicentrarchus labrax*: a simulation study. *Aquaculture* **202**, 329-342

Vandeputte, M., Launey, S. (2004) The genetic management of fish domestication. *Productions animales* **17**, 237-242

Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Chatain, B.A (2007) Polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax. Genetics* **176,** 1049-1057

Volckaert, F.A.M., Batargias, C., Canario, A., Chatziplis, D., Chistiakov, D., Haley, C., Libertini, A., Tsigenopoulos C. (2008) European sea bass. In: Genome mapping and genomics in animals. Eds. T.D. Kocher, C. Kole. Volume 2: 117-133. Springer- Heidelberg, Berlin

Volckaert, F.A.M., Hellemans, B., Batargias, C., Louro, B., Massault, C., Van Houdt, J.K.J., Haley, C., de Koning, D.J, Canario AV.M. (subm.) Family assignment and heritability of body weight, morphometric traits and stress response in aquacultured European sea bass *Dicentrarchus labrax* (submitted)

Chapter 4

QTLs for resistance to fish pasteurellosis in gilthead sea bream *Sparus aurata*

Cécile Massault, Rafaella Franch,
Chris Haley, Dirk Jan de Koning,
Henk Bovenhuis, Catrina Pellizzari,
Tomaso Partanello, Luca Bargelloni
Animal Genetics (submitted)

4.1 Abstract

Fish pasteurellosis is a bacterial disease causing important losses in farmed fish, including gilthead sea bream, a teleost fish of great relevance in marine. We report in this study a QTL analysis for resistance to fish pasteurellosis in this species. An experimental population of 500 offspring originating from eight sires and six dams in a single mass-spawning event, was subjected to a disease challenge with *Photobacterium damselae piscicida*, the causative agent of fish pasteurellosis. A total of 151 microsatellite loci were genotyped in all experimental fish, and half-sib regression QTL analysis was carried out on two continuous traits, body length at time of death and survival, and for two binary traits, survival at day 7 and survival at day 15, when the highest peaks of mortality were observed.

Two significant QTLs were detected for disease resistance. The first one was located on linkage group (LG) 3 affecting late survival (survival at day 15). The second one, for overall survival, was located on LG 21, which allowed us to narrow the region to a potential marker (Id13) linked to pasteurellosis disease resistance. A large significant QTL was also found for body length at death on LG 6 explaining 5-8% of the phenotypic variation.

4.2 Introduction

Gilthead sea bream (*Sparus aurata* L.) is a marine teleost fish of great importance for aquaculture, with an average cultured production of 80-100 thousands metric tonnes per year. This species is relatively well-adapted to aquaculture conditions, but it is susceptible to fish pasteurellosis, which often results in high mortality, up to 90% (Hawke *et al*. 2003). Fish pasteurellosis is a bacterial disease (caused by *Photobacterium damselae piscicida*) that spreads rapidly and is fatal in most cases after a few days. The presence of white tubercules in some internal organs is typical of chronic pasteurellosis. The outbreak may be associated with high stock density or a prolonged period of high water temperatures, which favours the development of the bacteria.

Disease outbreaks are a tremendous challenge for aquaculture production. Large economic losses can be generated by the destruction of entire production batches, while the potential infection of juveniles might pose a risk for the entire fish farm. Antibiotics and vaccines have been developed for a number of fish species and diseases. However, there is a general demand to minimize the use of antibiotics because bacteria can develop resistance and, drug residues in fish products could have a negative impact on consumers' health (Grewal and Tiwari 1990, Teuber *et al*. 2001). Antibiotic treatment does not reduce mortality to an acceptable level as it still results in a loss of 40% of fish over a four week period (Toranzo *et al*. 1991). Several vaccines have been designed to limit the effects of pasteurellosis, but few are actually efficient (Magarinos *et a*l. 1996, Romalde and Magarinos 1997), in particular in juvenile fish (Magarinos *et al*. 1999). In addition, vaccination presents safety issues for the personnel when carried out by manual injection, which is still widely used in parallel with

automatic vaccination devices (www.thefishsite.com) and leads to handling stress in the fish.

A potential strategy to reduce the probability of disease outbreak and thus to avoid the dramatic consequences of high mortality in fish farms is to implement selective breeding for disease resistance (e.g Fjalestad *et al*. 1993). Various modes of selection can be applied to perform genetic improvement of disease resistance: direct selection or indirect selection. Direct selection occurs when fish are selected among survivors following a disease challenge or natural disease outbreaks. Estimated heritabilities of survival on specific diseases have been obtained in several species (rainbow trout, Rye *et al*. 1990; Atlantic salmon for vibriosis, furonculosis and sea lice, Gjedrem and Gjøen 1995, Kolstad *et al*. 2004) As already observed in other species for various infectious diseases, resistance to fish pasteurellosis in sea bream has a low heritability (Antonello et al. 2009).. It is difficult and expensive to estimate, since the disease occurs as sporadic epizootic outbreaks rather than as endemic infections. When a trait has low heritability and is difficult to measure, an option can be to resort to marker assisted selection (MAS) (Dekkers and Hospital, 2002). MAS and genomic selection consist of selecting animals based on markers information. Application in aquaculture was investigated by Sonesson (2007) for MAS and by Sonesson and Meuwissen (2009) for genomic selection. In case of MAS, this requires initial QTL mapping to detect the relevant loci. This method detects regions of the genome that are associated with the trait of interest, in this case disease resistance. The information can be used for selective breeding and also to track genes that underlie variation in disease resistance in order to understand the biology of this trait. A better

understanding of the immune response could also assist in designing better drugs or other treatments.

QTL mapping in gilthead sea bream is made possible by the rapid development of genomic tools for this species and more in general in aquacultured fish (Canario *et al*. 2008). A first generation linkage map has been developed (Franch *et al*. 2006) for *S. aurata*, providing a large panel of microsatellite markers. In this study, existing as well as newly developed microsatellite markers are used in a genome-wide scan for QTLs involved in resistance to fish pasteurellosis.

4.3 Materials and Methods

Fish

The experimental population used in the present experiment is part of a larger group of juvenile fish that were subjected to an experimental challenge with *P. damselae piscicida* to estimate heritability of disease resistance as reported in Antonello *et al*. (2009), where a detailed description of the challenge experiment, on the family structure of the challenged fish, the methods used for parentage assignment, and the composition of the contributing brood stocks can be found. An experimental population, approximately 3,500 animals originating from mass-spawning of four broodstocks, was experimentally infected with a highly virulent strain of *P. damselae piscicida*. Mortality was monitored daily for 19 days. Upon completion of challenge experiment, genotypes at seven microsatellite loci were obtained for 1753 animals and for all (256) broodstock fish, for parentage assignment.

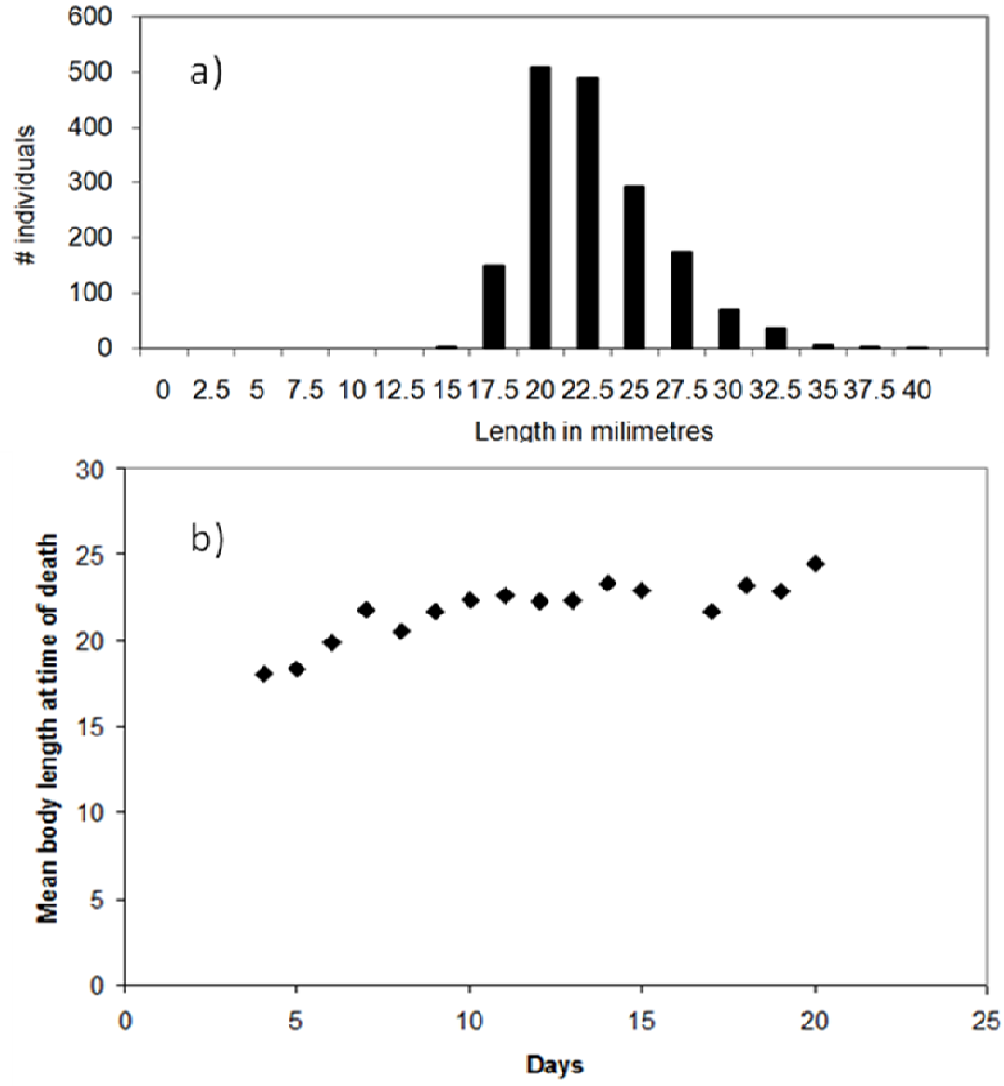

Figure 1 – Distribution of body length at death (a) and mean of body length at death where fish did no die before Day 4 and death were not reported at Day 16 (b).

Trait measurements

Dead fish were removed and day of death was recorded for each individual. Digital photos were used to determine the body length at death (BLD), which represents the body length on the day where the fish was found dead or at day 20 for surviving fish. Since body length was recorded only after death, it cannot be considered to be independent from survival time. Fish that were alive at the end of the experiment were attributed a value of 20 for survival. Figure 1 a) shows the overall distribution of log-transformed BLD of dead fish and Figure 1 b) the mean body length of collected dead fish for each sampling day for a subset of 500 fish.

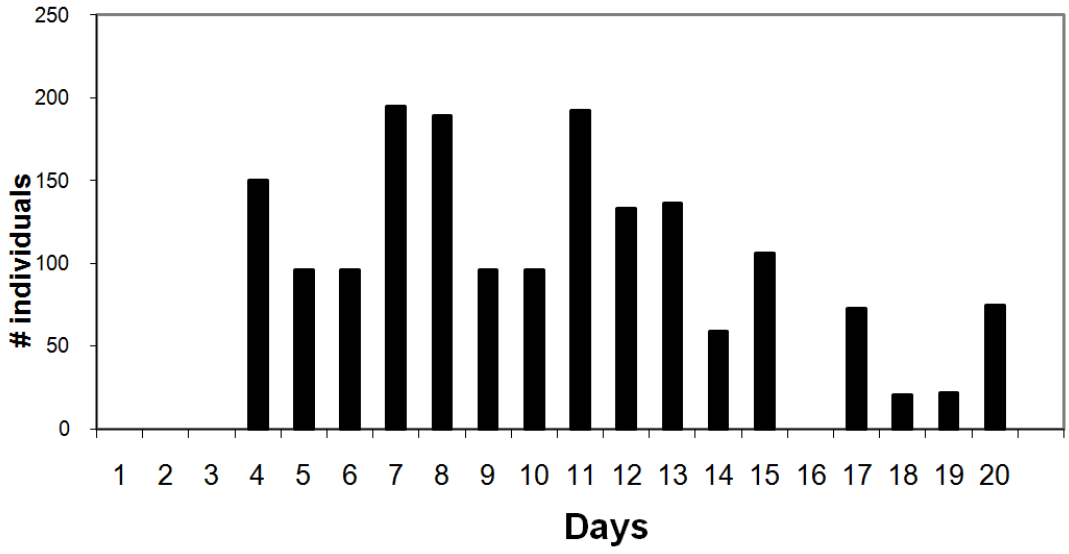

Figure 2 - Death frequencies per day during 19 days of challenge. The value 20 was allocated to survivors.

Antonello et al. (2009) observed a peak of mortality on day 7 (11.4 % of loss) and a smaller peak on day 15 of the challenge experiment. To take into account

the peculiar dynamics of the infection, two additional (binary) traits were considered, survival at day 7 and survival at day 15, where animals received a score of 0 if they are dead at that time or 1 if they are still alive. Figure 2 shows the mortality peaks for the same subset as Figure 1 (500 animals).

Two-step QTL design

To perform QTL mapping, a two-step experimental design was implemented as described by Massault et al. (2008). The first step consists of collecting DNA samples and phenotypic data for a larger set of individuals. From the 3,577 fish that entered the disease challenge test, DNA was collected on a random subset of 1753 animals. Parentage of this subset fish was assigned using a set of 7 microsatellite markers. A total 1257 fish could be uniquely assigned to parents using the software PAPA (Duchesne *et al*. 2002). More details can be found in Antonello *et al*. 2009). From these 1257 fish, 500 individuals were selected for QTL mapping, in order to analyse the largest full-sib families available (Massault *et al.* 2008). Selected fish originated from eight sires and six dams and consisted of 17 full-sib families (Table 1). The largest full-sib family comprises 151 offspring, three full sib families have medium sizes (between 35 and 59), while the other full sib families have a low number of offspring (between 4 and 27). A total of eight paternal half-sib and six maternal half-sib families were selected for the QTL analysis. Two large paternal half-sib families from sire 147 with more than the half of the offspring (264) and sire 178 with 107 offspring were present, while the six other sire families comprise less than 30 individuals per family. Maternal half-sib families were split into two large families (dam 151 with 186 offspring and dam 172 with 160 offspring), two medium-size families (of 71 and 52 offspring) and two small families (totalling 31 individuals). These 500 fish (referred to as the QTL

Table 1 - Family structure with number of offspring per half-sib families and per full-sib families

Sire ↓ *Dam →*	*147*	*153*	*159*	*165*	*169*	*174*	*178*	*180*	***Total***
136	-	-	-	-	-	-	7	-	**7**
151	151	-	-	-	-	-	35	-	**186**
160	59	-	-	-	-	-	8	4	**71**
168	-	-	-	-	-	-	18	-	**18**
170	-	7	6	-	19	-	20	-	**52**
172	54	-	21	20	-	27	19	25	**166**
Total	**264**	**7**	**27**	**20**	**19**	**27**	**107**	**29**	**500**

population) were fully genotyped for 151 microsatellite markers. Of these, 111 loci had been already mapped onto the first generation linkage map (Franch *et al.* 2006), 39 were developed for the present study as described in Franch et al. (2006).

Genetic map

The genetic map for the QTL population was constructed based on the genotypes of all animals for the 151 markers using CRI-MAP v.2.4. (http://compgen.rutgers.edu/multimap/crimap). First, the *twopoint* option was

used to determine which markers were significantly linked (LOD score > 3). Linkage groups were formed using the *build* option. Using information from a published map (Franch *et al*. 2006) the CRI-MAP *all* option was used to position new markers into a pre-determined map order of published markers. Subsequently, the resulting orders of all markers were evaluated using the *flipsn* option (flipping up to 4 markers). When multiple alternative orders were not significantly different (LOD< 3.0), the order with the highest likelihood was used.

QTL mapping

To detect QTL for body length and survival, survival at day 7 and survival at day 15, half-sib regression was implemented as described by Knott et al. (1996), both for maternal and paternal half-sib families. The half-sib regression tool is available online on GridQTL web page (http://gridqtl.org.uk). The model used is as follows:

$$y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij}$$

where y_{ij} is the offspring phenotype, α_i the mean of family i, β_i the regression coefficient for family i, x_{ij} the probability to inherit parental haplotype 1 conditional on the marker information and e_{ij} the residual. Information content and the phenotypic variance explained by the QTL were calculated as described by Knott et al. (1996). The F statistic is given for each linkage group for the most likely position as well as the 5% chromosome-wide (putative) and the 5% genome-wide (significant) thresholds to determine the significance of the detected QTL. The thresholds were calculated using permutation analysis and confidence interval using bootstrap analysis.

Variance component analysis (George *et al*. 2000) was also performed for body length at death and survival, which are continuously distributed. The model used here (including the QTL) is as follows:

$$y = Xb + Zu + Zv + e$$

where y is the phenotype of the offspring, X the incidence matrix relating phenotypes to systematic environmental effects, b is the vector with solutions for systematic effects, Z the incidence matrix relating animals to phenotype, u vector of additive polygenic effect, v vector of additive QTL effect and e vector of environmental effect (var(u) = A σ_u^2 and var(v) = G, with A the additive relationship matrix and G the IBD matrix). The log likelihood is then compared to the model without QTL, by performing a log likelihood ratio test (LRT), which follows a χ^2 distribution with one degree of freedom.

LRT = 2 (loglikelihood(model with QTL) – loglikelihood (model without QTL)

4.4 Results and discussion

Genetic map

Out of 151 markers genotyped on 500 fish, 140 could be mapped to 24 linkage groups (LGs) Figure 3), the same number of chromosomes described for gilthead sea bream haploid set (Cataudella *et al*. 1980). Eleven markers remain unlinked. The coverage from our average sex map was 1041.3 centiMorgans and the average space between markers is 5.8 centiMorgans. A comparison between the map presented here and the first generation map (Franch *et al*. 2006) showed concordance between most of the LGs of the two maps. The only exceptions are i) marker Ad75, which in Franch et al. (2006) was included in

LG 9 (instead of LG12'), ii) LG22 is not present in the map from this study, iii) LG18 now includes the original LG18 plus LG26. These differences are easily explained especially if evidence from the sea bream Radiation Hybrid (RH) map (Senger *et al.* 2006, Sarropoulou *et al.* 2007) is taken into consideration. In fact, 99 loci from the present map have been mapped in the RH map, allowing anchoring of the new linkage map to the physical map (Table 2). First, the original position of locus Ad75 was uncertainly linked to LG9 (Franch *et al.* 2006), and it mapped on a different RH group (RH25, Sarropoulou *et al.* 2007) compared to the remaining markers on LG9 (RH23), therefore it seems likely that the present position is the correct one. Second, LG22 in Franch et al. (2006) was a small LG with only two loci, which are not informative in the present population study. However, based on the comparison between the RH map and the first generation linkage map (Sarropoulou *et al.* 2007) LG22 should join LG8, since markers from both LGs are found on the same RH group (RH2). Third, fusion of LG18 and LG26 into a single group (LG18) in the new map was predicted by the RH map, where both LGs pointed to a single RH group (RH12). Regarding to the latter point, although the present map consists of a smaller number of mapped markers (140) compared to the first generation map (198), the much larger size of the present genetic map (500 individuals compared to 50) evidently provides a better representation of the sea bream genome. Finally, the presence in the present map of 99 anchoring loci to the RH map allows a comparative genomic approach to be implemented, since the RH map is easily linked to fish genome sequences (Sarropoulou *et al.* 2007).

Disease resistance QTL

Paternal half-sib (PHS) and maternal half-sib (MHS) regression analysis were performed for body length and survival traits (survival, survival at day 7 and

survival at day 15). Two genome-wide significant and 18 putative chromosome-wide significant QTLs for survival, survival at day 7, and survival at day 15 were found as detailed in Table 3. These 20 QTLs are distributed on 10 different linkage groups. The first genome-wide significant QTL was found for resistance against pasteurellosis on LG21 with PHS. This QTL explains 4% of the phenotypic variation (Table 3). The confidence interval covers the whole linkage group (4 cM). A second significant QTL was detected, using PHS, on LG3 for survival at day 15, which explains 6% of the phenotypic variation and for which the confidence interval again extends to the whole chromosome. Variance component analysis also detected a QTL for survival with a LRT of 15.2 on LG21 (Table 4), confirming the results found with half-sib regression analysis. Putative QTLs for survival were also reported on LG1, LG3, LG10, and LG19 with PHS and LG21 with MHS (Table 3). Putative QTLs for survival at day 7 were reported on LG1, LG4, LG5, LG14, LG19 and LG21. In one case, on LG1, a QTL for overall survival was also detected on this linkage group. Most of the QTLs were found with PHS, except for those on LG5 and LG21. The phenotypic variation explained by the QTLs varies between 3.5% and 5% and confidence intervals generally cover the entire linkage groups. Putative QTLs for survival at day 15 were found on LG10, LG12', LG16, and LG19 with PHS, while MHS analysis detected a suggestive QTL on LG19 and LG21. All confidence intervals are large and QTLs explain between 4% and 6% of the phenotypic variation.

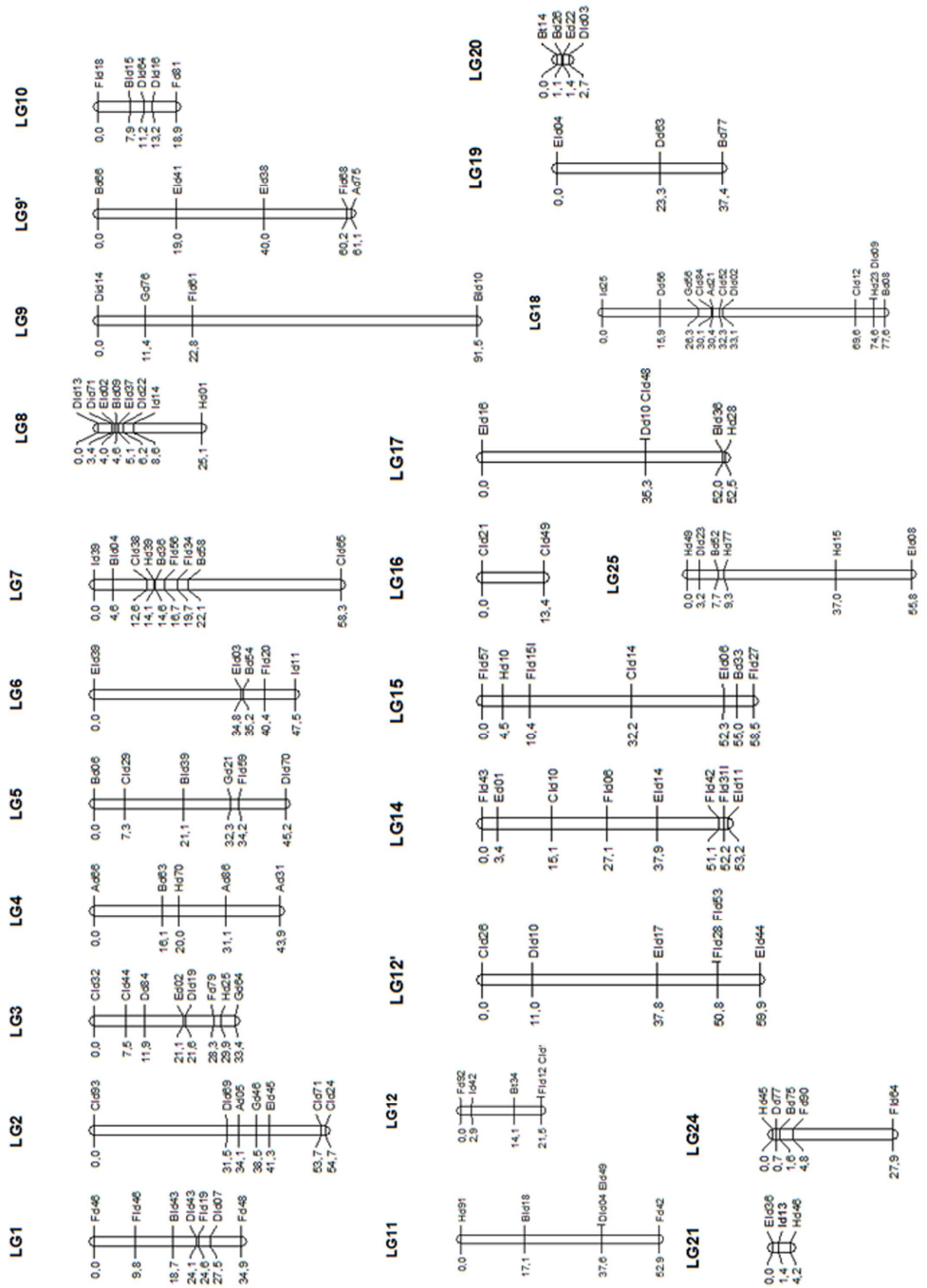

Figure 3 - Sex-averaged genetic map on centiMorgans with 24 linkage groups

Comparing the results of detected QTLs at different survival times (day 7, day 15, overall) suggests that the effects of some loci might be stronger or even exclusive during specific phases of the infection. For instance, QTLs on LG4, LG5, and LG14 are observed only for survival at day 7, whereas QTLs on LG16 and LG12' are found exclusively for survival at day 15. This evidence should be taken with caution since these QTLs are suggestive and might not be confirmed after further analysis and/or could have an effect also on later stages of the infection, but these effects might have gone undetected due to limited power of the analysis. However, it is not unexpected to find phase-specific QTLs, since mechanisms of disease resistance can be biphasic or even multiphasic (*e.g.* resistance to trypanosomosis in mice, Koudandé *et al.* 2008). A first line of defence might involve the innate immune response, while later on other mechanisms (e.g. adaptive immune response) could play a stronger role. A bimodal pattern has been found in the present challenge experiment as well as in other studies (Antonello *et al.* 2009), with a primary infection, after experimental exposure to the pathogen, and a secondary infection with bacteria released from moribund and dead fish. The observed infection/re-infection dynamics appears therefore to further support the hypothesis of a biphasic defence response.

On the other hand, the effect of some QTLs seem to extend over the entire duration of the infection, being found either for at least two survival traits (QTLs on LG1, LG3, LG10), or for all three time points, as in the case of the QTLs on LG21, which reaches genome-wide significance only for overall survival, but is detected also on day 7 and day 15, and on LG19. The latter one explains a percentage of phenotypic variation between 7% and 9%, which is large for a putative QTL. LG19 is one of the linkage groups where sire '147'

Table 3 - QTL for survival significant at the chromosome-wide level.

Trait[1]	*Method*[2]	*LG*[3]	*Pos*[4]	*F*[5]	*CW*[6]	*GW*[7]	*Effect (%)*[8]	*CI*[9]
Significant								
Survival	PHS	LG21	2	3.4	2.1	3.4	4	0-4
Day15	PHS	LG3	0	3.8	2.3	3.6	6	0-33
BLD	PHS	LG6	11	5.0	1.8	2.3	8	0-40
BLD	MHS	LG6	19	3.9	1.9	3.3	5	2-47
Putative								
Survival	PHS	LG1	19	2.5	2.3	3.4	4	0-36
	PHS	LG3	0	2.3	2.1	3.4	4	0-33
	PHS	LG10	8	2.3	1.9	3.4	4	0-18
	PHS	LG19	3	2.5	2.5	3.4	7	0-37
	MHS	LG21	1	2.7	2.3	3.6	3	0-4
Day 7	PHS	LG1	32	2.6	2.2	3.3	4	0-36
	PHS	LG4	7	2.8	2.2	3.3	5	5-38
	MHS	LG5	27	2.9	2.7	4.2	3.5	0-44
	PHS	LG14	24	3.2	2.4	3.3	5	5-38
	PHS	LG19	0	3.3	2.0	4.2	7	0-37
	PHS	LG21	2	2.7	2.2	4.3	4	0-4
	MHS	LG21	1	3.6	2.6	4.2	4	0-4

Table 3 - QTL for survival significant at the chromosome-wide level. (cont.)

Trait[1]	*Method*[2]	*LG*[3]	*Pos*[4]	*F*[5]	*CW*[6]	*GW*[7]	*Effect (%)*[8]	*CI*[9]
Day15	PHS	LG10	8	2.3	2.1	3.6	6	0-18
	PHS	LG12'	2	3.2	2.3	3.6	5	0-59
	PHS	LG16	13	2.5	2.3	3.6	4	0-13
	PHS	LG19	18	3.1	3.1	3.6	9	0-37
	MHS	LG19	37	1.9	1.9	4	2	0-37
	MHS	LG21	2	3.9	2.2	4	4	0-4
BLD	MHS	LG4	4	3.0	2.0	3.4	4	0-43
	MHS	LG5	38	2.1	1.9	3.4	2.5	21-45
	PHS	LG8	7	2.2	1.8	2.7	5	0-25
	PHS	LG15	3	2.4	1.8	2.7	4	5-55

[1] Trait of interest: survival, survival at Day7, survival at Day15, and body length at time of death

[2] Method: paternal half-sib regression (PHS) or maternal half-sib regression (MHS)

[3] LG: linkage group

[4] Position: position of the QTL in centiMorgans

[5] F: F statistic for the QTL

[6] 5%CW: value of 5% significance chromosome-wide threshold

[7] 5%GW: value of 5% significance genome-wide thresholds

[8] Effect: the proportion of phenotypic variation in % explained by the QTL

[9] CI: 95% confidence interval

did not contribute to the analyses because he was not informative for any of the markers.

Body length at death QTL

QTLs found for BLD are also described in Table 3. A genome-wide significant QTL for BLD was detected on linkage group 6 for both PHS and MHS explaining 8% and 5% of the phenotypic variation, respectively. In total, 299 offspring out of 500 are from PHS families segregating for this QTL. The QTL is positioned at 11 centiMorgans according to PHS and at 19 according to MHS. But in both cases the confidence interval is very large, therefore no conclusion can be drawn for the exact position of the QTL. The QTL was also detected using the variance component analysis on LG6 with a LRT of 32.9 exceeding 1% chromosome-wide thresholds (Table 4). Two putative QTLs were found with PHS on LG8 and LG15 which explain respectively 5% and 4% of the phenotypic variation. With MHS, two putative QTLs were located on LG4 and LG5 which explain respectively 4% and 2.5% of the phenotypic variation. For all QTLs, confidence intervals are large and cover the entire linkage group.

Table 4 - Variance component results for body length at death and survival

Linkage Group	*Trait*	*LRT*	*P=0.005*	*P = 0.001*
6	BLD	32.9	3.8	6.6
21	Survival	15.2	3.8	6.6

Id13 marker

A significant QTL was found on LG21, which is a small linkage group of 4 centiMorgans and 3 markers. Using a contingency table and pearson chi-square test, we found that locus Id13 was a possible marker linked to disease resistance (Table 5). We had a more close look at the genotype of marker Id13. In total, 192 animals were actually typed for Id13 and among the 28 fish that survived to the challenge; only 15 were genotyped, 13 from the largest family (147 x 151) and two from a smaller family of 20 offspring (165 x 172). The fact that some parents have not been genotyped for the marker reduces the number of offspring genotypes available. In both cases, about 10% of the family survived. The QTL segregates in the largest maternal half-sib family of dam 151. Figure 4 shows the mortality curve in percentage of animals that inherited the 177 microsatellite allele (63 animals) and the 188 allele (89 animals). Looking at Figure 4, we notice that individuals that inherited the 177 allele register a loss of 35% for day 7 and day 8 alone, while the ones that inherited allele 188 record for those two days a loss of 13.5%. The pattern was reversed on day 10 - day 11

Table 5 - Contingency table for individuals inheriting alternative genotypes for marker Id13 at the end of the experiment (Day 20)

Id13 genotype	*Dead*	*Alive*	*Total*
177-188	62	0	62
188-188	75	13	88
Total	137	13	150

* p-value with Pearson chi-square test of 0.004

where individuals inheriting allele 188 lost a total of 26% individuals, but individuals inheriting allele 177 were less affected (9.5% loss). A sudden loss at day 13 affected both groups equally (11% for allele 188 and 9% for allele 177). Only the group that inherited allele 188 has animals that survived (13 animals, 14.5% of the group). Those data indicate that that animals that inherited allele 188 resist pasteurellosis longer, dying in large numbers from day 10 and not at day 7 like animals inheriting allele 177, which did not survive after day 17, showing a more severe course of infection.

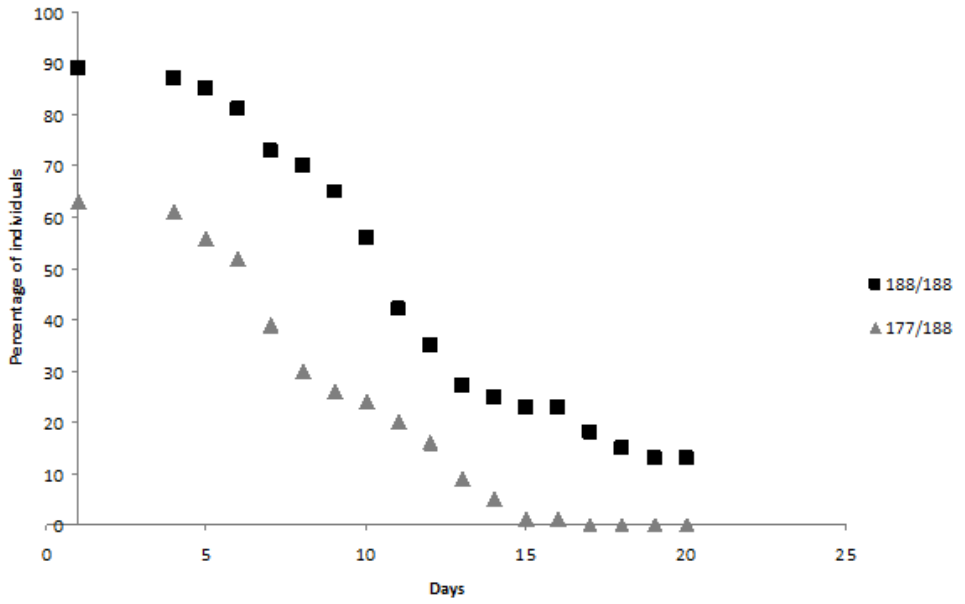

Figure 4 - Survival curves for animals of a large full-sib family according to the segregation of the maternal allele 177 or 188 at the Id13 locus.

As mentioned above, locus Id13 maps onto LG21. This small linkage group corresponds to RH18 in the RH map (Sarropoulou *et al.* 2007). The length of RH18 is 822.9 $cRAD_{3000}$, which translates into an estimated size of approximately 35 Mega base pairs (Mbp) (1 Mbp = 23.48 $cRAD_{3000}$, following

the approach of Senger et al. 2006 based on a total map length of 18,787.14 $cRAD_{3000}$ and an estimated genome size of 800 Mbp for *S. aurata*). This estimate is in agreement with the size range of orthologous chromosomes in other fish species (*Tetraodon nigroviridis* chromosome 5, 13.3 Mbp; *Gasterosteous aculeatus* group II, 23.3 Mbp; *Oryzias latipes* chromosome 3, 36.6 Mbp;. *Danio rerio* chromosome 776.9 Mbp) and most similar to the one of the species with a comparable genome size (*O. latipes,* 700 Mbp). Loci Eid36 and Hd46 on LG21 (Figure 3) are found at opposite ends of RH18. Therefore LG21 corresponds nearly to the entire chromosome, and its small length cannot be explained as a consequence of incomplete marker coverage. It is therefore quite likely that the chromosome corresponding to LG21/RH18 has a reduced recombination rate. In fact, in the first generation map, which was based on a completely unrelated mapping panel, LG21 consisted of four loci in complete linkage and with length zero. While the association of allele 188 of Id13 locus with longer survival to fish pasteurellosis is certainly suggestive, it is quite likely that this evidence is due to reduced recombination in the entire linkage group. Low recombination rates on LG21 represent an advantage when using genetic markers to assist breeding programs (discussed below) because linkage between the genetic marker and the QTL is stronger. On the other hand, to refine mapping of disease resistance QTL on LG21 will probably require to shift from family-based to population-based genome scan. To this end, the "mass-spawning" scheme for QTL identification (Massault *et al.* 2008) produced a larger experimental population, with over 250 full-sib families, generated by a substantial number of different sires and dams, all of them genetically unrelated, because brood stock fish originated directly from wild populations.

Disease resistance QTLs in aquaculture

The discovery of two significant QTLs for resistance to pasteurellosis resistance and several suggestive ones can be added to the list of detected QTL associated with resistance to various diseases for a number of aquaculture species. QTLs of major effect, explaining a large proportion of phenotypic variation between 25% and 50%, have been reported for salmon, trout and flounder (Ozaki *et al*. 2000, Houston *et al*. 2006, Fuji *et al.* 2006) with between 50 and 80 markers. We found a lower percentage of phenotypic variance explained by QTL, varying between 2% and 9%. Rodriguez et al. (2004) obtained a smaller percentage of phenotypic variation explained by the QTL in trout (11% on average) with a marker density similar to ours (139 markers). The differences observed could be biological differences between species. In some species, resistance to disease can be influenced by a major gene, and in some other, it can ne influenced by many genes with smaller effects. The marker density does play a role in the power to detect those effects (Kolbedhari *et al*. 2005).

One genomic region that has been frequently found to be associated with disease resistance/susceptibility is the MHC I complex, also in fish species (*e.g.* Grimholt *et al.* 2003, Miller *et al.* 2004, Johnson *et al.* 2008, Evans and Neff 2009). The MHC I locus position in the teleost genome appears to be conserved as it has been mapped onto homologous chromosomes in *G. aculeatus* (Group X), *D. rerio* (chromosome 19), and *O. latipes* (chromosome 11), which corresponds to sea bream RH19/20 and LG7. Although MHC I has not mapped in *S. aurata*, it should be noticed that no QTL was found on LG7.

MAS and genomic selection

The possibility of genetic improvement for disease resistance in fish has been demonstrated by Fjalestad et al. (1993), although MAS can be a better solution for traits such as disease resistance, where no measurement can be observed on the selection candidate. Results of QTL mapping, even if not integrated into a MAS scheme, could still be exploited to unravel the mechanisms of disease resistance by identifying regions of the genome that explain complex traits such as survival.

Sonesson and Meuwissen (2009) explored the possibility of genomic selection as a selective breeding method in aquaculture, which could be particularly powerful for traits such as disease resistance, as it tends to eliminate requirement for observations on relatives. Their basic conclusions were higher genetic gain and lower rate of inbreeding. The main disadvantage with genomic selection is the need to re-estimate breeding values every few generations, as selection is based on linkage disequilibrium, which declines fast. Although this could be feasible in Atlantic salmon, high throughput genomic information is not currently available for a large number of aquatic species, including the gilthead sea bream.

4.5 Conclusion

This study shows the results of the first QTL mapping experiment done in gilthead sea bream for disease resistance to Pasteurella. The QTL found for body length at death, after validation, can be integrated into a breeding program, while an interesting marker Id13 has been associated with survival. With advancing genomics tools, comparison mapping with other species and

full genome sequencing will help to understand the mechanisms of innate immune response to bacterial infection.

Acknowledgment

This study has been funded by the European Commission (project AQUAFIRST, contract number FP6-STREP-2004-513692). DJK and CSH acknowledge financial support from the BBSRC.

References

Antonello, J., Massault, C., Franch, R., Haley, C., Pellizzari, C., Bovo, G., Patarnello, T., de Koning, D., Bargelloni, L. Estimates of heritability and genetic correlation for body length and resistance to fish pasteurellosis in the gilthead sea bream (*Sparus aurata* L.). *Aquaculture* **298,** 29-35

Canario, A.V.M., Bargelloni, L., Volckaert, F., Houston, R.D., Massault, C., Guiguen Y. (2008) Genomics Toolbox for Farmed Fish. *Reviews in Fisheries Science* **16**, 3-15.

Cataudella, S., P., Perin, R., Sola, L. (1980) A chromosome study of eight Mediterranean species of Sparidae (Pisces, Perciformes). *Genetica* **54**, 155–159.

Dekkers, J.C.M., Hospital, F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet*. **3**, 22-32

Duchesne, P., Godbout, M.H., Bernatchez, L. (2002) PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. *Molecular Ecology Notes* **2**, 191-193.

Evans, M.L., Neff, B.D. (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, in press, corrected proof

Fjalestad, K.T., Gjedrem, T., Gjerde, B. (1993) Genetic improvement of disease resistance in fish - an Overview. *Aquaculture* **111**, 65-74.

Franch, R., Louro, B., Tsalavouta, M., Chatziplis, D., Tsigenopoulos, C.S., Sarropoulou, E., Antonello, J., Magoulas, A., Mylonas, C.C., Babbucci, M., Patarnello, T., Power, D.M., Kotoulas, G., Bargelloni, L. (2006) A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata L*. *Genetics* **174**, 851-861.

Fuji, K., Hasegawa, O., Honda, K., Sakamoto, T., Okamoto, N. (2007) Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **272**, 291-295.

George, A.W., Visscher, P.M., Haley C.S. (2000) Mapping quantitative trait loci in complex pedigrees: A two-step variance component approach. *Genetics* **156**, 2081-2092.

Gjedrem, T., Gjoen H.M. (1995) Genetic-Variation in Susceptibility of Atlantic salmon (*salmo salar*) to furunculosis, BKD and cold water temperature. *Aquaculture Research* **26**, 129-134.

Grewal, J.S., Tiwari, R.P. (1990) Resistance to metal ion and antibiotics in *Escherichia coli* isolated from foods. *Journal Medical Microbiology* **32**, 223-226.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., Stet, R.J. (2003) MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* **55**, 210-219.

Hawke, J.P., Thune, R.L., Cooper, R.K., Judice, E., Kelly-Smith, M. (2003) Molecular and phenotypic characterization of strains of Photobacterium damselae subsp piscicida isolated from hybrid striped bass cultured in Louisiana, USA. *Journal of Aquatic Animal Health* **15**, 189-201.

Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J., Bishop, S.C. (2008) Major Quantitative Trait Loci Affect Resistance to Infectious Pancreatic Necrosis in Atlantic Salmon (*Salmo salar) Genetics* **178**, 1109-1115.

Johnson, N.A., Vallejo, R.L., Silverstein, J.T., Welch, T.J., Wiens, G.D., Hallerman, E.M., Palti, Y. (2008) Suggestive association of major histocompatibility IB genetic markers with resistance to bacterial coldwater disease in rainbow trout (*Oncorhynchus mykiss*). *Marine Biotechnology* **10**, 429–437.

Knott, S.A., Elsen, J.M., Haley C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71-80.

Kolbehdari, D., Jansen, G.B., Schaeffer, L.R., Allen B.O. (2005) Power of QTL detection by either fixed or random models in half-sib designs. *Genetics Selection Evolution* **37**, 601-614.

Kolstad, K., Heuch, P.A., Gjerde, B., Gjedrem, T., Salte R. (2005) Genetic variation in resistance of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*. *Aquaculture* **247,** 145-151

Koudandé, OD., Thomson, PC., Bovenhuis, H., Iraqi, F., Gibson, JP, van Arendonk, JAM (2008) Biphasic survival analysis of trypanotolerance QTL in mice. *Heredity* **100**, 407-414.

Magariños, B., Toranzo, A.E., Romalde, J.L. (1996) Phenotypic and pathobiological characteristics of Pasteurella piscicida. *Annual Review of Fish Diseases* **6**, 41-64.

Magarinos, B., Romalde, J.L., Barja, J.L., Nunez, S., Toranzo, A.E. (1999) Protection of gilthead seabream against pasteurellosis at the larval stages. *Bulletin of the European Association of Fish Pathologists* **19**, 159-161.

Massault, C., Bovenhuis, H., Haley, C., de Koning, D.J. (2008) QTL mapping designs for aquaculture. *Aquaculture* **285**, 23-29.

Miller, K.M., Winton, J.R., Schulze, A.D., Purcell, M.K., Mling, T.J. (2004) Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. *Environmental Biology of Fishes* **69**, 307–316.

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M.R.M., Akutu T., Okamoto, N. (2001) Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPN) in rainbow trout (oncorhynchus mykiss) *Molecular Genetics and Genomics* **265**, 23-31.

Romalde, J.L., Magarinos, B. (1997) Immunization with bacterial antigens: Pasteurellosis. *Fish Vaccinology* **90**, 167-177.

Rye, M., Lillevik, K.M., Gjerde, B. (1990) Survival in early life of Atlantic salmon and rainbow trout: estimates of heritabilities and genetic correlations. *Aquaculture* **89**, 209–216.

Rodriguez, F., LaPatra, S., Williams, S., Famula, T., May, B. (2004) Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow trout and steelheads trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* **241,** 93-115.

Sarropoulou, E., Franch, R., Louro, B., Power, D.M., Bargelloni, L., Magoulas, A., Senger, F., Tsalavouta, M., Patarnello, T., Galibert, F., Kotoulas, G., Geisler, R. (2007) A gene-based radiation hybrid map of the gilthead sea bream sparus aurata refines and exploits conserved synteny with tetraodon nigroviridis. *BMC Genomics* **8**, 44.

Senger, F., Priat, C., Hitte, C., Sarropoulou, E., Franch, R., Geisler, R., Bargelloni, L., Power, D., Galibert, F. (2006). The first radiation hybrid map of a perch-like fish: The gilthead seabream (sparus aurata L). *Genomics* **87**, 793-800.

Sonesson, A.K., (2007). Possibilities for marker-assisted selection in aquaculture breeding schemes. In: Guimarães, E., Ruane, J., Scherf, B., Sonnino, A., Dargie, J. (Eds.), Marker-assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish. FAO, Rome, pp. 309–328.

Sonesson, A.K., Meuwissen, T. (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetics selection evolution* **41**, 37

Teuber, M. (2001) Veterinary use and antibiotic resistance. *Current opinion in Microbiology* **4,** 493-499.

Toranzo, A.E., Barreiro, S., Casal, J.F., Figueras, A., Magarinos, B., Barja J.L. (1991) Pasteurellosis in Cultured Gilthead Seabream (Sparus-Aurata) - 1st Report in Spain. *Aquaculture* **99**, 1-15.

Chapter 5

Comparing genomic and traditional selection in mass-spawning species

Cécile Massault, Robbert Blonk,
Chris Haley, Dirk Jan de Koning,
Hans Komen, Johan van Arendonk,
Henk Bovenhuis
In preparation

5.1 Abstract

Selective breeding is not widely used in populations that reproduce by natural mating and mass-spawning such as sea bream or sea bass. We simulated mass selection, best linear unbiased prediction (BLUP) and genomic selection (GBLUP) for natural mating population and controlled mating with heritabilities (h^2) 0.2, 0.5 and 0.8, various population sizes (200, 512 and 800) and various number of animals selected (40, 48 and 100). Values observed for genetic level after 10 generations ranged from 17.45 to 35.95. Average inbreeding rate over 10 generations (ΔF) ranges from 1.7% to 9.7%, with the highest values reached by the BLUP method and accuracy (r_{IH}) ranged from 0.32 to 0.93. With a high h^2, the genetic level is similar for the three methods. When the number of animals selected increases, genetic gain, r_{IH} and ΔF decrease. When using controlled mating instead of natural mating, genetic level and r_{IH} are similar with $h^2 = 0.2$, while ΔF is lower using controlled mating. Therefore GBLUP is advantageous for natural mating mass-spawning species but inbreeding rates remain very high.

5.2 Introduction

High fecundity of species is in general considered to be advantageous in selective breeding as large families are produced and high selection intensities realised (Gjerde and Rye 1997). For a number of species in aquaculture, especially those still in process the of domestication, natural mating and mass spawning are used for the production of the next generation. In this system, a number of males and females are placed together in a tank, which results in a large number of offspring of unknown paternity.

Parentage assignment has become routine using DNA markers (see Jones *et al*. (2010) for review) and pedigree reconstruction has permitted a closer look at the actual population structure of mass-spawning species. Skewed parental contributions are common in fish: sea bream (Brown *et al*. 2005), tilapia (Fessehaye *et al*. 2006), cod (Herlin *et al*. 2008), sea bass (Chatziplis *et al*. 2007) and in sole (Blonk *et al*. 2009). In general a small number of sires and dams contributed a large number of offspring, while other individuals are not contributing any offspring. The offspring are a mix of full-sib and half-sib families (both paternal and maternal) with few large families and many small families. This variation in family size may dramatically increase the rate of inbreeding per generation (Hedgecock 1994, Brown *et al*. 2005, Fessehaye *et al*. 2006).

Mass selection is currently the most-used method of selection in mass-spawning species (Brown *et al*. 2005, Gjedrem *et al.* 2005, Vandeputte *et al*. 2009). It is the simplest way of selective breeding, in which only the animals' own performance is used as criterion of selection. A major disadvantage with mass selection is that genetic progress will be low for traits with low heritabilities.

But access to parentage assignment has made it possible to do family selection in communanlly reared progeny (walk-back selection; e.g. Herbinger *et al*. 1995)

A mixed model accounting for all family relations (Best Linear Unbiased Prediction (BLUP)) can be used to increase the accuracy of selection. Sonesson et al. (2005) investigated truncation selection based on BLUP estimated breeding values for fish breeding schemes and concluded that this method was more efficient in terms of genetic gain, but resulted in higher rates of inbreeding without restriction inbreeding. The rate of inbreeding can be constrained in populations with controlled reproduction by using optimum contribution selection (Sonesson and Meuwissen 2000).

Genomic selection is a more recent method of selection that includes information on thousands of marker genotypes. Genomic selection can be used to increase the accuracy of selection, especially for sex-limited traits, traits with low heritability or traits recorded late in life. Advantages of genomic selection over marker assisted selection is that all genetic variance can be captured due to the large number of markers (see Goddard and Hayes 2009 for review). Various approaches have been developed to estimate breeding values using information of many marker such as partial least squares regression, BayesA and BayesB (Meuwissen *et al*. 2001, Hayes *et al*. 2009). Due to rapid development in the field of genomics, it has become feasible to genotype individuals for large SNP panels at relatively low costs. At present these are not yet commercially available for many fish species but it is expected that in the near future this will be the case. This will open up the opportunity to apply genomic selection in fish. A specific aspect on the use of genomic selection in mass-spawning fish is that specifically reconstructed pedigree is not needed.

The consequences of genomic selection for genetic improvement and rate of inbreeding in mass spawning fish have to our knowledge not been studied. In this study, we use simulation to determine the genetic gain and the rate of inbreeding in a mass pawning population. We compared the results of three selection methods, i.e. mass selection, BLUP selection and genomic selection, over a period of 10 generations.

5.3 Material and methods

Genetic model and selection process

We simulated a genome with a total size of 10 Morgans, divided over 10 chromosomes (1 Morgan length each) under a finite locus model. For a random mating population consisting of 50 males and 50 females, where one male and one female per mating was used as parents for the next generation and a mutation rate of 10^{-5} (Meuwissen *et al.* 2001), mutation-drift equilibrium was reached after 1700 generations, i.e. the number of polymorphic marker stabilizes on average at 4 400. We kept record of the pedigree for the last 5 generations of random mating in order to reach an equilibrium situation prior selection. The base population G_0 was the last generation of random mating after reaching mutation-drift equilibrium (generation 1700). The base population was identical for the three selection methods: mass selection, BLUP selection and genomic selection. Pedigrees were stored for each method of selection to calculate inbreeding coefficients, although pedigree records were only used for BLUP breeding value estimation. We simulated 5 different scenarios (for description, see Table 1) varying the number of selection candidates, the mating design and the number of phenotyped offspring. For each scenario, in which each of the three selection methods were applied for

the three different heritabilities (0.2, 05 and 0.8) over 10 generations (G_1 to G_{10}). For each alternative, results were averaged over 50 replicates.

Natural or controlled mating

After each generation of selection, selected fish reproduced in two ways: either natural mating or controlled mating. For controlled mating (scenario 5) the number of selection candidates and the number of phenotyped offspring was not varied: 16 males were mated to in total 32 . Females were mated to only one male and males were mated to two females. Sixteen offspring were kept per full-sib family, 8 males and 8 females and the total of number of offspring with phenotypes was therefore 512.

For the natural mating situation we tried to simulate a population structure similar to the population described by Antonello *et al.* (2009). For that purpose, we first allocated 245 mating pairs by randomly selecting randomly parents with replacement (males and females can be chosen for more than one mating) using a brood stock of 59 males and 68 females. We then allocated the number of offspring to the matings, by drawing those from a gamma distribution with parameters $\alpha = 0.17$ and $\beta = 0.75$. This process was replicated 1000 times, and for each replicate a sample of 1257 offspring was randomly selected to form the target population. This simulation procedure resulted in number of parents, family structure and variation in number of offspring per family that was similar to the gilthead sea bream population described by Antonello *et al.*(2009). The family variance of that population is showed in Figure 1a. Therefore, these parameters of the gamma distribution were chosen to simulate natural mating populations.

Table 1. - Characteristics of the 5 scenarios that were simulated

	# offspring	*cm[1]/nm[2]*	*nselmales*	*nselfemales*
Scenario 1				
$h^2 = 0.2$	512	nm	20	20
$h^2 = 0.5$	512	nm	20	20
$h^2 = 0.8$	512	nm	20	20
Scenario 2				
$h^2 = 0.2$	200	nm	20	20
$h^2 = 0.5$	200	nm	20	20
$h^2 = 0.8$	200	nm	20	20
Scenario 3				
$h^2 = 0.2$	800	nm	20	20
$h^2 = 0.5$	800	nm	20	20
$h^2 = 0.8$	800	nm	20	20
Scenario 4				
$h^2 = 0.2$	512	nm	50	50
$h^2 = 0.5$	512	nm	50	50
$h^2 = 0.8$	512	nm	50	50
Scenario 5				
$h^2 = 0.2$	512	cm	16	32
$h^2 = 0.5$	512	cm	16	32
$h^2 = 0.8$	512	cm	16	32

[1] natural mating [2] controlled mating

Figure 1a and 1b show the distribution of family sizes of the population described by Antonello *et al*. (2009)and the simulated population averaged over 1000 replicates for the same number of parents and number of offspring. The distributions are both skewed with few families contributing most of the offspring and a large number of families contributing fewer offspring. Variance of family size in simulated populations (148.33 ± 12.18) falls in the range of family size variance observed (138.61) by of Antonello et al. (2009).

The breeding schemes simulated in this study are smaller than those in Figure 1a Therefore, we adjusted the parameter α of the gamma distribution to 1 in order to obtain the family size variance showed in Figure 1a for a situation were 20 males and 20 females are randomly sampled with replacement to become parents. We randomly selected 512 fish that survived until reproductive age to form our population of selection candidates. We used an equal number of males and females in the brood stock for natural mating, while the number of females was twice the number of males for controlled mating.

Forty parents (20 males and 20 females) were selected for scenario 1, scenario 2 and scenario 3. One hundred parents were selected in scenario 4 (50 males and 50 females) and forty-eight parents (16 males and 32 females) for scenario 5, were we used controlled mating. Sex ratio in the selection candidates was assumed to be 1.

Phenotypes

We simulated a trait that is affected by 100 QTLs, with 10 QTLs per chromosome. QTLs were evenly spaced along the chromosome. The QTL effects were sample from a gamma distribution with shape parameter 0.4 (Meuwissen *et al.* 2001) and scale parameter 0.12 in order to obtain a average

genetic variance σ^2_A of 15 over 50 replicates. QTL effects were additive and had a probability of 0.5 to be either positive or negative. The true breeding value was calculated as the sum of the allelic effects accross all QTLs:

$$g_i = \sum_{j=1}^{n} (a_{1ij} + a_{2ij})$$

where g_i is the true breeding value of animal *i*, *n* is the total number of QTLs, a_{1ij} the effects of the paternal QTL allele of animal *i* and at locus *j* and a_{2ij} the effect of the maternal QTL allele of animal *i* and at locus *j* (Hayes and Goddard 2008). In order to simulate a trait with a given heritability h^2, the environmental values e_i were drawn from a normal distribution $N(0,\sigma_e)$, with $\sigma_e^2 = \frac{(1-h^2)}{h^2}\sigma_A^2$.

The phenotypic value for animal *i* was the sum of genetic value g_i and environmental value e_i. Genetic variance (σ^2_A) in a specific f generation was calculated as the variance of the true breeding values of the animals in that generation.

Selection methods

Parents for the next generation are selected using three different selection methods: mass selection, GBLUP (Hayes *et al.* 2009) and BLUP. No restriction on inbreeding was applied for any of the methods. We will refer the number of selected males as *nselmales* and the number of selected females as *nselfemales*. For all three methods, all animals were phenotyped and for genomic selection, all animals were genotyped.

Mass selection

The parents for the future generations were selected according to their own performance. Selection candidates were ranked from the highest phenotypes to the lowest. *Nselmales* and *nselfemales* with the highest phenotypes were therefore kept for reproduction.

Best Linear Unbiased Prediction (BLUP) selection

The estimated breeding values (EBVs) from the selection candidates were calculated using the following mixed model:

$$y = \mu + Zu + e$$

where y was the vector of phenotypes, μ the mean phenotype, Z the incidence matrix relating phenotypes to breeding values, u vector of estimated breeding values, with var(u) = A σ^2_A and e vector of environmental effects, with var(e) = $I\ \sigma^2_e$. A is the genetic relationship matrix of animals from 5 generations prior to selection up to the current generation and I the identity matrix. In our case, no fixed effects other than the mean were added to the model. Phenotypes from animals from the last 5 generations prior base generation up to the current generation were included. For estimated breeding values, heritabilities were assumed known and fixed either to 0.2, 0.5 or 0.8. The new brood stock was therefore composed of the *nselmales* males and *nselfemales* females with the highest EBVs.

Genomic BLUP selection

The genomic breeding values (GEBVs) of selection candidates were estimated with the following model (Hayes et al. 2009) and using a genomic relationship matrix to based on marker information:

$$y = \mu + Zu + e$$

where y was the vector of phenotypes, μ the mean phenotype, Z the incidence matrix relating phenotypes to breeding values, u vector of estimated breeding values, with var(u) = G σ^2_A and e vector of environmental effects, with var(e) = I σ^2_e . G, here, is the genomic relationship matrix between the animals of the current generation and I the identity matrix. Animals from the current generation were phenotyped and genotyped. For estimated breeding values heritabilities were fixed to 0.2, 0.5 or 0.8. In our case, no fixed effects other than the mean were added to the model. We constructed the genomic relationship matrix G as described by Hayes and Goddard (2008). On average information was available of 4 400 polymorphic markers. A similarity value between two individuals was calculated between each locus, which can be either 0, 0.5 or 1. The genomic relationship between 2 individuals is the sum of the similarity values across all loci. The new brood stock was composed of *nselmales* males and *nselfemales* females with the highest GEBVs.

Genetic level, inbreeding rate and accuracy of selection

Inbreeding level per generation was calculated as the mean of the inbreeding coefficients of the animals in a generation. Rate of inbreeding ΔF was calculated as :

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$$

where F_t is the average inbreeding level of generation t and F_{t-1} the average inbreeding level of generation *t-1*.

The genetic level in a generation was calculated as the average of the true breeding values of individuals in that generation and the accuracy of selection for a generation (r_{IH}) was the correlation between true breeding values and the estimated breeding values of animals of that generation.

5 .4 Results

Comparison of selection methods

Figure 2 shows the genetic level, rate of inbreeding, accuracy and genetic variance for scenario 1 with low heritability (0.2). After 10 generations of selection, GBLUP results in the high of genetic level, while mass selection realised the lowest genetic level. The genetic level obtained using BLUP selection falls in between GBLUP and mass selection. Rates of inbreeding are in general very high (up to 12% increase per generation). For the three methods of selection, inbreeding rate increases a lot between G_1 and G_2 after first selection, where sibs are more likely to be mated due to high intensity of selection. The generation G_0 being the last generation of random mating with equal family sizes and no selection taking place between G_0 and G_1, explains the fact the hige increase of rate of inbreeding between G_1 and G_2. While the BLUP and GBLUP selection method reach a 12% rate of inbreeding in G_2, the inbreeding rate with mass selection climbed to 6% and remains about 6% during the 10 generations of selection. The inbreeding rate using BLUP

stabilizes around 12%, while the inbreeding rate with GBLUP stabilizes around 8%. We can observed fluctuations in rate of inbreeding from one generation to the next. This could be explained by the presence of dominant families. In some generations, dominant families will be genetically superior and therefore a large amount of offspring of those families will be selected, increqsing the risk of sib

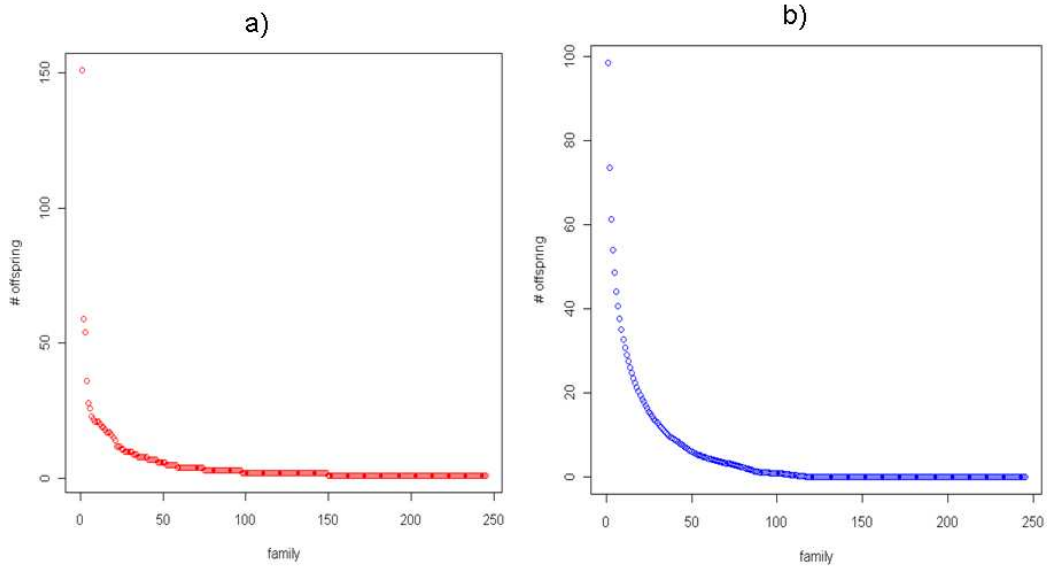

Figure 1 - a) Distribution of full-sib family size observed in a gilthead sea bream population (Antonello *et al*. 2009). b) Distribution of simulated full-sib family size averaged over 1000 replicates

mating, while in other generations, dominant families will not be genertically superior. Therefore fish from a lot of different families will be selected, decreasing the risk of sib mating. Accuracy of selection in generation G_1 (after

first selection) is the highest with GBLUP (0.70) followed by BLUP (0.63) and accuracy for mass selection method was 0.44. Accuracy of selection for all three selection methods decreased over time to 0.46 value of for GBLUP, 0.33 for BLUP and 0.32 for mass selection in generation G_{10}. Genetic variance decreases considerably for all three selection methods. This is partly due to the Bulmer effect (first generations) and the high rates of inbreeding. Further, allele fixation will reduce genetic variance. Genetic variance for mass selection decreases slower for BLUP or GBLUP selection, which have a similar reduction in genetic variance. Table 3 summarizes the results for the 5 scenarios. The genetic level at G_{10}, the average inbreeding level over10 generations, and the accuracy at G_1 and G_{10} is given. In all the studied alternatives GBLUP performs best or similar than the other two selection methods in respect of genetic level. Average inbreeding rate over 10 generations is very high (between 1.7% and 9.7%) for all selection methods and for any scenarios. This is mainly due to the nature of mass-spawning species and due to the fact that no attempts were made to the restrict inbreeding in our selection procedure. Accuracy of selection varies between 0.36 and 0.93 for GBLUP, between 0.33 and 0.90 for BLUP and 0.32 and 0.89 for mass selection. Accuracy is always higher when using GBLUP selection compared to BLUP and mass selection.

Effect of heritability

Figure 3a shows the genetic level for scenario 1 (512 selection candidates, 20 males and 20 females selected) with high heritability (e.g. 0.8) for the three selection methods. Genetic level is similar for GBLUPand mass selection. We observe from Table 3 the similar genetic level between the GBLUP and mass selection with high heritability for all the scenarios,, while the BLUP genetic

level is much lower. Figure 3b shows the genetic level for GBLUP selection in scenario 1 when heritabilities differ; here the genetic level is higher when the heritability is higher. This also holds for other scenarios (see Table 3). The rate of inbreeding for GBLUP and BLUP selection decrease as the heritability increases while the rate of inbreeding for mass selection is more stable.

Effect of populationsize

To study the effect of population size on the results of GBLUP we compared scenario 2, scenario 1 and scenario 3, where the population size varies (200, 512 and 800 respectively) for a heritability of 0.2 and 40 selected parents (Figure 3c). At heritability of 0.2, the genetic level is similar in the case of 512 animals and 800 animals, except for the last 2 generations, where the genetic level when using 512 animals is higher than when using 800 animals. However, for medium and high heritability, the genetic level is higher when using 800 animals (Table 3). The rate of inbreeding are higher at lower heritability and this could explains that for h2 = 0.2, the genetic level is higher with 512 animals than with 800. The advantage of having more animals is counterbalanced by the higher rate of inbreeding, which causes a larger loss of genetic variation. The rate of inbreeding level increases as the population size increases. A higher number of selection candidates produced by the same number of parents results in a larger size of the dominant families and potentially more sibs will be selected to produce the next generation.

Effect of intensity of selection and mating design

Figure 4a, Figure 4b and Figure 4c show the genetic level, inbreeding rate and accuracy of selection, with 512 selection candidates and heritability of 0.2 for GBLUP selection for a natural mating population with 20 selected males and

20 selected females (scenario 1), a controlled mating population with 16 selected males and 32 selected females (scenario 5) and a natural mating population with 50 selected males and 50 selected females (scenario 4), respectively. No major differences are observed between the genetic levels when selecting 40 animals with natural mating and 48 animals with controlled mating are very similar, while the genetic level obtained with 100 animals selected using natural mating is not as high as the two other scenarios (Figure 4 a). Accuracy of selection is similar for the three cases starting between 0.6 and 0.7 after the first selection and decrease in the same pattern and values for the three scenarios. The inbreeding rate is the highest when using natural mating with 40 selected animals and the lowest when using natural mating with 100 selected animals. The inbreeding rate of the controlled mating schemes is lower than natural mating with 40 animals selected, due to the population structure and the slight higher number of selected animals.

5.5 Discussion

We showed that in most cases, GBLUP performed better than or as good as BLUP and mass selection. However, while we included phenotypes for all animals of the pedigree (5 last generations of random mating up to the current generation) for BLUP estimation of breeding values, we only used genotypes and phenotypes of the selection candidates of the current generation for practical purposes. Therefore, we did not explore the full potential of GBLUP and we could expect a higher genetic level when including animals from previous generations. The inbreeding rate is very high in general (between 1.7% and 9.7%), with BLUP selection being the highest and mass selection the lowest. Accuracy at G_{10} of selection is always the highest with GBLUP.

Table 3 – Genetic level at generation 10, average rate of inbreeding and average of accuracy over 10 generations for the three methods

	Genetic level			Rate of inbreeding			Accuracy at G_1 and at G_{10}		
	GBLUP	BLUP	MASS	GBLUP	BLUP	MASS	GBLUP	BLUP	MASS
Scenario 1									
$h^2 = 0.2$	25.34	20.87	17.45	6.5%	9.7%	3.8%	0.70 – 0.46	0.63 – 0.33	0.44 – 0.32
$h^2 = 0.5$	29.53	25.51	26.77	5.3%	8.1%	3.9%	0.85 – 0.60	0.77 – 0.47	0.70 – 0.46
$h^2 = 0.8$	33.14	29.35	32.59	4.9%	7.5%	3.8%	0.93 – 0.72	0.90 – 0.60	0.89 – 0.61
Scenario 2									
$h^2 = 0.2$	18.97	16.35	14.50	5.5%	6.6%	3.7%	0.62 – 0.48	0.57 – 0.34	0.45 – 0.34
$h^2 = 0.5$	23.65	21.56	20.96	4.8%	6.0%	3.7%	0.82 – 0.57	0.74 – 0.50	0.68 – 0.53
$h^2 = 0.8$	26.62	24.01	26.20	4.1%	5.6%	3.8%	0.91 – 0.75	0.87 – 0.67	0.89 – 0.69
Scenario 3									
$h^2 = 0.2$	24.65	22.42	20.63	7.5%	9.1%	3.6%	0.66 – 0.36	0.65 – 0.34	0.45 – 0.30
$h^2 = 0.5$	29.87	26.39	27.00	6.3%	8.5%	4.0%	0.81 – 0.52	0.77 – 0.46	0.68 – 0.43
$h^2 = 0.8$	34.19	31.23	34.00	5.0%	7.4%	4.2%	0.91 - 0.71	0.90 -0.41	0.90 – 0.58
Scenario 4									
h2 = 0.2	22.16	19.46	16.82	3.3%	4.0%	2.0%	0.65 – 0.53	0.60 – 0.43	0.44 – 0.37
h2 = 0.5	26.01	23.98	23.55	2.9%	3.6%	2.1%	0.81 – 0.66	0.77 – 0.59	0.70 – 0.54
h2 = 0.8	29.16	27.01	28.45	2.5%	3.0%	2.0%	0.91 – 0.72	0.90 – 0.70	0.89 – 0.71
Scenario 5									
$h^2 = 0.2$	26.01	22.48	21.64	4.9%	6.2%	1.7%	0.69 0.53	0.63 0.41	0.45 0.37
$h^2 = 0.5$	31.03	28.52	28.95	3.7%	5.1%	1.9%	0.82 – 0.64	0.78 – 0.52	0.69 – 0.0.49
$h^2 = 0.8$	35.95	32.91	35.62	2.6%	4.1%	1.8%	0.92 – 0.77	0.90 – 0.68	0.89 – 0.67

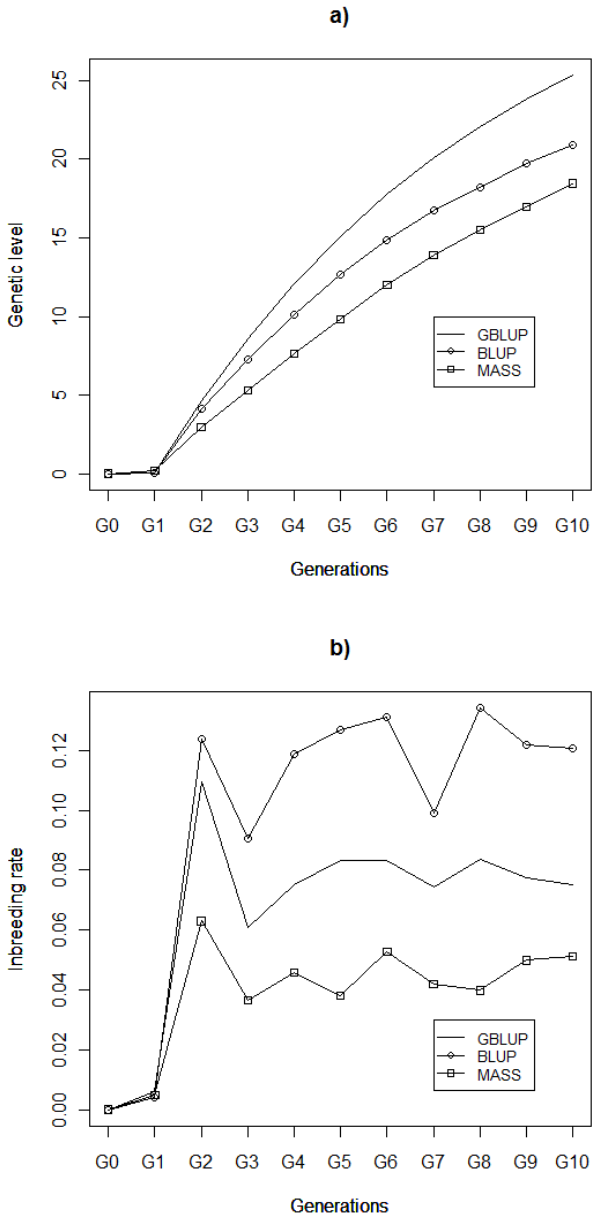

Figure 2 - Results of different selection methods for scenario 1, i.e. 20 selected males, 20 selected females, 512 candidate selection for a h^2=0.2 – averaged over 50 replicates - Genetic level (a), inbreeding rate (b)

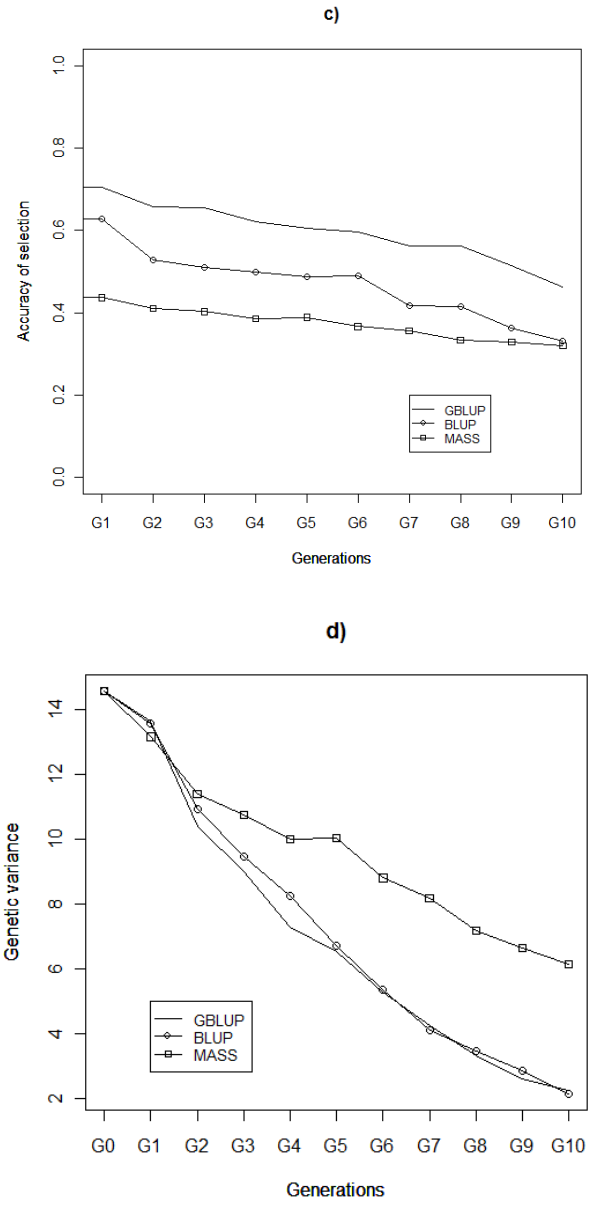

Figure 2 (cont.) - Results of different selection methods for scenario 1, i.e. 20 selected males, 20 selected females, 512 candidate selection for a h^2=0.2 – averaged over 50 replicates - accuracy of selection (c) and genetic variance (d).

Stochastic simulation versus deterministic simulation

We also predicted genetic level, rate of inbreeding and accuracy of selection for mass and BLUP selection for scenario 5 (controlled mating) using deterministic simulation (SelAction, Rutten *et al.* 2004) to compare with simulation results. Genetic level and accuracy of selection increase as the heritability increases for the 2 selection methods. While ΔF in mass selection remains constant when heritability increases (1.8%) , ΔF decreases in BLUP selection when heritability increases (from 5.7% at h^2=0.2 to 2.1% at h^2 = 0.8). The BLUP selection method performs better iwith a genetic level of 28.00 and accuracy of selection of 0.55 than mass selection with genetic level of 27.40 and accuracy of selection of 0.42, while inbreeding rate for BLUP (5.7%) is higher than in mass selection (1.8%) for low heritabity. For an heritability of 0.8, the 2 methods result in the same genetic level and accuracy (genetic level of 46.40 and accuracy of selection of 0.85).

In both stochastic and deterministic simulations for the mass and BLUP selection method, we observed similar trends. Accuracies found with stochastic simulation with mass selection are close to those expected by deterministic simulation ($\sqrt{h^2}$). While the predicted inbreeding rate from deterministic simulation are similar to stochastic simulation, inbreeding rate in BLUP is higher with stochastic simulation. Genetic level at G_{10} is lower than those predicted deterministically for the three heritabilities. The stochastic genetic level with low heritability (scenario 5) are 22.48 and 21.64, respecively, while the genetic level of BLUP and mass selection using deterministic simulation are 28.00 and 27.40. This can be explained by three factors. First, the deterministic simulation as implemented in selAction assumes an infinitesimal model, where traits are affected by an infinite number of genes each with small effects. In the

stochastic simulation, we used a finite locus model, with 100 genes with effects distributed according to a gamma distribution. This means that in our stochastic simulation of allele frequencies will change and some alleles might become fixed and no further improvement will be possible, while improvement is continuous and infinite with deterministic simulation. Secondly, the deterministic model does not account for reduction of genetic variance due to inbreeding. Therefore, the genetic variance remains higher than in our stochastic simulation. Finally, results given by the deterministic simulation correspond to an equilibrium situation, where genetic variance and heritability converged after few generations, where they do not change due to Bulmer effect (i.e genetic variance).

Inbreeding rate

As the results showed, the rate of inbreeding with GBLUP selection is very high (7% in scenario 1 with low heritability). The level of inbreeding of scenario 1 with heritability of 0.5 started from 0.002 in G_0 and reached 0.47 at G_{10} while the level of inbreeding that reported Sonesson and Meuwissen (2009) started from 0 in G_1 and reaches 0.06 in G_{10} for heritability of 0.4 in a scenario where QTLs effects are only estimated once. The large difference observed between the two results is mostly due to the population strcuture and the number of selection candidates. Sonesson and Meuwissen (2009) use a control on mating with equal family size and 3 000 selection candidates, while we used natural mating with 512 selection candidates. The inbreeding rate using the BLUP selection method without restriction on inbreeding was larger (9.6% in scenario 1 with low heritability). Sonesson (2007) found that the rates of inbreeding were higher using BLUP schemes than marker-assisted schemes, because less individuals per family were selected due to the extra information

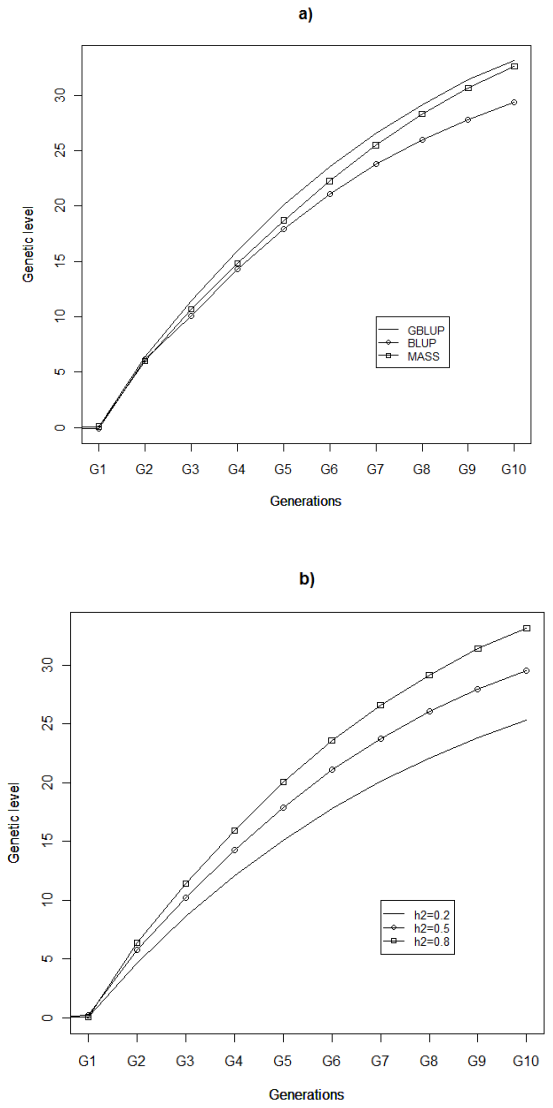

Figure 3 - a) Genetic level for GBLUP, BLUP and mass selection with 512 selection candidates, 20 selected males and 20 selected females for an heritability of 0.8 (scenario 1) – b) Genetic level of GBLUP with 512 selection candidates, 20 selected males, 20 selected females for heritability of 0.2, 0.5 and 0.8 (scenario 1)

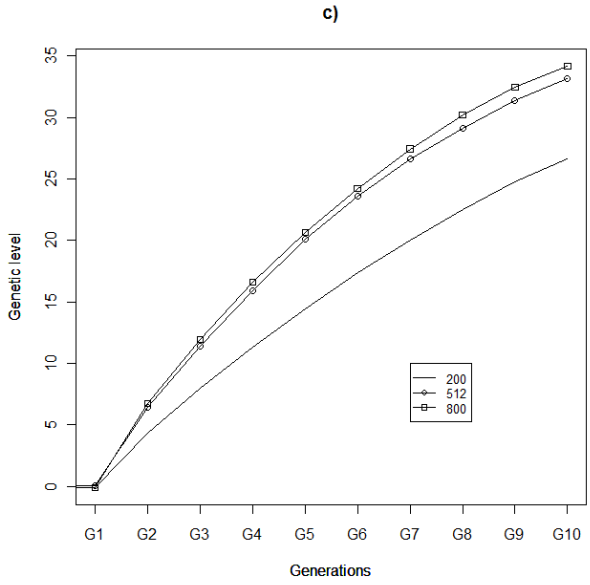

Figure 3 (cont.) - c) Genetic level of GBLUP with 20 selected males, 20 selected females, heritability of 0.2 and for 200, 512 and 800 selection candidates (scenario 2, scenario 1 and scenario 3)

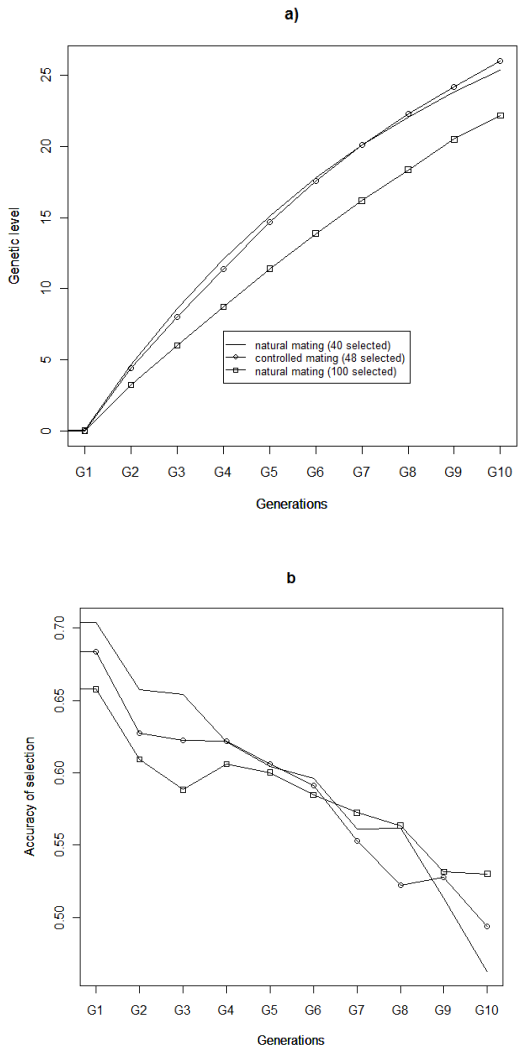

Figure 4 - Comparison between natural mating with 20 males selected, 20 females selected, 512 selection candidates and heritability of 0.2 (scenario 1), controlled mating with 16 selected males, 32 selected females, 512 selection candidates and heritability of 0.2 (scenario 5) and natural mating with 50 males selected, 50 females selected, 512 selection candidates and heritability of 0.2. - a) genetic level, - b) inbreeding rateand

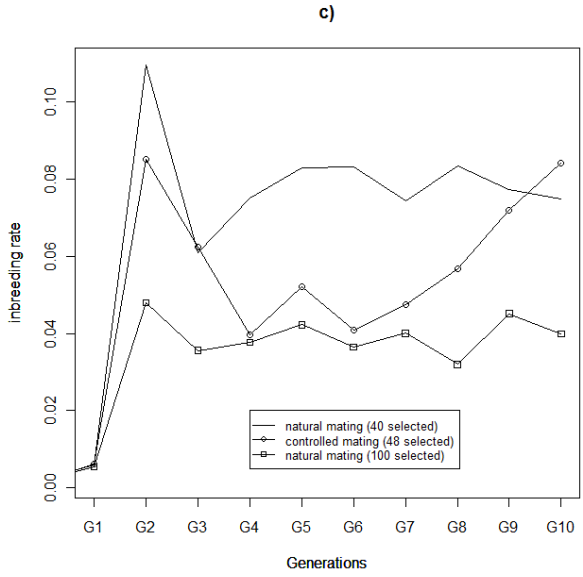

Figure 4 (cont.) - Comparison between natural mating with 20 males selected, 20 females selected, 512 selection candidates and heritability of 0.2 (scenario 1), controlled mating with 16 selected males, 32 selected females, 512 selection candidates and heritability of 0.2 (scenario 5) and natural mating with 50 males selected, 50 females selected, 512 selection candidates and heritability of 0.2. - c) accuracy of selection .

from markers. The same principle applies here if we consider that marker assisted selection corresponds to GBLUP with one QTL.

The high rate of inbreeding observed in our simulation can be explained by variability in family sizes. Brood stocks that were kept, were fairly small in our simulation. Furthermore , not all brood stock in mass spawning actually contributes to the next generation. This will not be the case when reproduction can be controlled. The low number of effective parents in combination with the

variation in family size has a dramatic impact on the rate of inbreeding Finally, we did not apply any restriction on inbreeding. The rate of inbreeding that we recorded in our simulation, are not acceptable for sustainable breeding schemes. Our study confirms the reported high inbreeding rate after one generation between 2.5% and 3.5% in sea bream (Brown *et al*. 2005) and of 3-6% in tilapia (Fesshaye *et al.* 2006). Those characteristics are specific from mass-spawning species. Our results show that, while the effect on the genetic level is rather minor compared to breeding programs with reproductive control, it is essential to control the rate of inbreeding for breeding schemes that use mass-spawning.

However, we can see in scenario 4 that increasing the number of selected animals does reduce inbreeding rates. Although inbreeding rates in scenario 4 are still too high for selective breeding purposes, increasing the number of selected animals remains an option to restrict inbreeding. The number of selected parents should be increased further more along with the number of selection candidates to keep a high selection intensity. Another option would be to use several mass-spawning units. While rates of inbreeding will remain high within each unit, the exchange of genetic material between units would restrict the overall rate of inbreeding to an acceptable level (Blonk *et al*. 2009). Finally, in genomic selection, the genomic relationship matrix could be used to restrict inbreeding. Further investigations should be undertaken in this area. Control over inbreeding rates for mass-spawning species are discussed in greater detail in the general discussion.

Natural mating and controlled mating

For the first 5 generations, natural mating and controlled mating have somewhat similar genetic level for heritability of 0.5 and 0.8, while natural mating has a

higher genetic gain than control mating for heritability of 0.2. But after generation 5, for all heritabilities, controlled mating results in high genetic level than natural mating. The high inbreeding rate observed for natural mating could partly explain the lower genetic level in the later generations. Many loci would become fixed at the unfavourable allele and reduce the genetic variance.

The rates of inbreeding in GBLUP are lower than in BLUP. The genomic relationship matrix distinguishes among full-sib individuals, while the traditional relationship matrix based on pedigree does not. Therefore, the genomic relationship matrix based on markers is more accurate than the genetic relationship matrix based on pedigree and become more accurate by increasing the number of markers (VanRaden 2007).

Candidate selections and training population

In this study, we estimated the breeding values of one set of phenotyped animals per generation. However, a strong argument in favour of genomic selection is that it is possible to estimate accurately breeding values of animals with no phenotypes (Goddard and Hayes, 2009). The genetic markers are used to link the selection candidates (with no phenotypes) to a reference population that has been genotyped and phenotyped. This can be especially advantageous for traits which are difficult to record. In our specific simulations, such application of genomic selection could not be achieved, due to the uneven family sizes. To estimate accurately breeding values of animals that where only genotyped, sibs of animals to be estimated have to be present in the reference population, where animals have been genotyped and phenotyped. However, in the case of mass-spawning species, the smallest families might not have enough

offspring to obtain accurate estimation of breeding values and therefore will reduce accuracy of selection.

Efforts to control reproduction in mass-spawning species (i.e. in sea bass Vandeputte et al. 2009) will increase the efficiency of genomic selection.

Acknowledgment

This study has been funded by the European Commission (project AQUAFIRST, contract number FP6-STREP-2004-513692). DJK and CSH acknowledge financial support from the BBSRC.

References

Antonello, J., Massault, C., Franch, R., Haley, C., Pellizzari, C., Bovo, G., Patarnello, T., de Koning, D. J., Bargelloni, L. (2009) Estimates of heritability and genetic correlation for body length and resistance to fish pasteurellosis in the gilthead sea bream (*Sparus aurata L.*). *Aquaculture* **298,** 29-35.

Blonk, R. J. W., Komen, J., Kamstra, A., Crooijmans, R., van Arendonk, J. A. M. (2009) Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. *Aquaculture* **289,** 26-31.

Brown, C. R., Woolliams, J. A., McAndrew, B. J. (2005) Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture* **247,** 219-225.

Chatziplis, D., Bartagias, C., Tsigenopoulos, C.S., Magoulas, A., Kollias, S., Kotoulas, G., Volckaert, F.A.M., Haley, C.S. (2007) Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): the BASSMAP pilot study. *Aquaculture* **272**, S172-S182

Fessehaye, Y., El-Bialy, Z., Rezk, M. A., Crooijmans, R., Bovenhuis, H., Komen, H. (2006) Mating systems and male reproductive success in Nile tilapia (*Oreochromis niloticus*) in breeding hapas: A microsatellite analysis. *Aquaculture* **256,** 148-158.

Gjedrem, T., Andersen, O., Bentsen, K.B., Fjalestad, K.T., Gjerde, B., Hayes, B.J., Kolstad, K.,Meuwissen, T.H.E., Olesen, I., Refstie, T., Rye, M.,

Sonesson, A., Thodesen, J.,Woolliams, J.A., (2005). Selection and Breeding Programs in Aquaculture. Springer,Dordrescht, The Netherlands, 364 pp.

Gjerde, B. , Rye, M (1997) Design of breeding programs in aquaculture species – Possibilities and constraints. Genetics and breeding of Mediterranean aquaculture species. Proc. Of the seminar of the CIHEAM Network on Technology of Aquaculture (TECAM) , jointly organized by CIHEAM and FAO, Zaragoza, Spain. Options Mediterr. 34, 181-193.

Goddard, M. E., Hayes, B. J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10,** 381-391.

Hayes, B. J., Goddard, M. E. (2008) Technical note: Prediction of breeding values using marker-derived relationship matrices. *Journal of Animal Science* **86,** 2089-2092.

Hayes B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K., Goddard, M.E. (2009) Accuracy of genomic selection in multi-breed dairy cattle populations. *Genetics Selection Evolution* **41**

Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? Pp. 122–134 in A. Beaumont, ed. Genetics and evolution of aquatic organisms. Chapman and Hall, London.

Herbinger, C., Doyle, R., Pitman, E., Paquet, D., Mesa, K., Morris, D., Wright, M., Cook, D. (1995) DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture*. **137** 245-256

Herlin, M., Delghandi. M, Wesmajervi. M., Taggart, J., McAndrew, B. Penman. D. (2008) Analysis of the paternal contribution to a group of fry from a single day of spawning from a commercial Atlantic cod (*Gadus morhua*) breeding tank. *Aquaculture*. **274**, 218-224

Jones, A., Small, C., Paczolt, K, Ratterman, N (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Ressources* **10**, 6-30

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157,** 1819-1829.

Rutten, M.M., Bijma, P., Wooliams, J.A, van Arendonk, J.A.M. (2002). SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. *The Journal of Heridity* **93**, 456-458

Sonesson, A.K. (2007). Within-family marker-assisted selection for aquaculture species. *Genetics Selection Evolution* **39**, 301-317

Sonesson, A. K., Gjerde, B., Meuwissen, T. H. E. (2005) Truncation selection for BLUP-EBV and phenotypic values in fish breeding schemes. *Aquaculture* **243,** 61-68.

Sonesson, A. K., Meuwissen, T. H. E. (2000) Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution* **32,** 231-248.

Sonesson, A. K., Meuwissen, T. H. E. (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution* **41**.

Vandeputte, M., Baroiller, J. F., Haffray, P, Quillet, E. (2009) Genetic improvement of fish: Achievements and challenges for tomorrow. *Cahiers Agricultures* **18,** 262-269.

VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. Interbull Bull. 37:33–36.

Chapter 6

General discussion

This thesis is part of a larger European project, designed to detect QTLs for disease resistance and stress in rainbow trout, Pacific oysters, gilthead sea bream and European sea bass. Information on detected chromosomal regions can subsequently be used in marker-assisted selection. The work presented here focuses on QTL mapping for traits of economic importance in mass-spawning species. The objectives were firstly to propose designs for QTL mapping experiments for various aquatic species, secondly to perform QTL analyses for stress response and disease resistance in sea bream and sea bass, using the designs previously introduced, and finally to integrate genomic information in breeding programs for mass-spawning fish species.

Although quantitative genetic theory can be relatively easily applied to genetically improve most species, natural mating mass-spawning fish have some specific characteristics that have to be taken into account when designing breeding schemes: (i) large number of offspring per parent, (ii) large variation in family sizes and unequal parental contribution (Brown *et al*. 2005, Fessehaye *et al.* 2006, Blonk *et al*. 2009). These characteristics carry with them the risk of high rates of inbreeding. Furthermore, natural mating in mass-spawning species leads to complex population structures, where males are mated to several females and females mated to several males within the same tank. An additional step, parentage assignment, must be undertaken to reconstruct pedigrees, which are necessary for QTL experiments and breeding programs. The main challenge of this thesis was to apply QTL detection and genomic selection of mass-spawning species (i.e. uncontrolled reproduction).

This discussion is divided into four parts: (i) power of experimental designs and comparison between expectations and real results, (ii) mass-spawning design for QTL experiments and the possibility of genome wide association studies

(GWAS), (iii) restriction of inbreeding rate in natural mating populations and marker assisted selection, and (iv) parentage assignment and pedigree reconstruction.

6.1. Power of experimental designs

Selective genotyping design

While the price of genotyping per marker decreases, the actual price of genotyping per animal does not necessarily decrease too. We presented in the power study (**Chapter 2**) a strategy for selective genotyping where 5 large full-sib families of 1000 offspring were phenotyped and for each family, 200 fish with extreme phenotypes were genotyped. The contrast between different alleles is increased, and therefore so is the power to detect QTLs. This method can easily be applied for fish and shellfish, i.e. Pacific oysters, where very large full-sib families can be obtained. The genotyping is significantly reduced for the experimental population (only 1000 animals are genotyped instead of whole population of 5 000 animals). Sauvage *et al*. (*in press*) used this strategy and detected 5 significant QTLS, which were associated to summer mortality in Pacific oysters.

However, selective genotyping presents several drawbacks. First, the number of individuals to phenotype remains very large (5000 in our study). The principle of selective genotyping is to phenotype the whole population, while only selected markers are genotyped to limit the costs. However, phenotyping can also be costly, depending on the trait of interest (i.e. carcass trait). Furthermore, it cannot be applied when multiple traits are measured. Increase of power is only achieved for one trait, i.e. the trait based on which individuals are selectively genotyped. It is possible to select extreme phenotypes for two or

multiple traits, however the gain in reducing genotyping cost will be lost as you quickly end up genotyping the whole population. Also, selective genotyping cannot be applied for non-continuous trait such as 0/1 trait. Therefore, in case of 0/1 trait, the whole population needs to be genotyped as only two different phenotypes exist.

Moen *et al*. (2004) proposed a multistage strategy using selective genotyping to detect QTLs for disease resistance in Atlantic salmon. After the disease challenge, only affected animals were genotyped and significant markers were identified based on the transmission disequilibrium test. In the next stage resistant animals were genotyped for the significant markers and a Mendelian segregation test was performed to ascertain if the markers were segregating in a Mendelian fashion. The last stage was to perform a survival analysis using a log-rank test on markers that were both significant in the transmission analysis disequilibrium and segregated in a Mendelian fashion. Although the Moen *et al*. (2004) method of selective genotyping is attractive because it works for binary traits such as survival and it reduces the amount of genotyping, it is not widely used in QTL mapping for disease resistance in fish. The only example is reported by Lallias *et al*. (2009) for European flat Oysters.

Moen et al. (2004) suggested a specific mating design (mating ratio of 1 male for 2 females), with maternal half-sib families of 40 offspring, where resistant and susceptible fish are compared. It is not easy to obtain such a structure in mass-spawning. The sample size should be 5000 fish (the smallest family contains 4 offspring in a sample of 500 animals in sea bream) to have at least 40 offspring per half-sib family. In addition, mass-spawning species reproduce in batches and the genotyping of a few microsatellites for all individuals is necessary to assign parents and reconstruct families. Therefore there is no

advantage anymore in performing a transmission disequilibrium test only on affected individuals. The method proposed by Moen et al. (2004) is not appropriate for mass-spawning species because of their population structure and the rapid development of markers. However, as all animals have to be genotyped for parentage assignment, effort could be made to equalize family sizes. In that way, the population structure will resemble more the one described by Moen *et al*. (2004).

Mass-spawning design

Chapter 2 investigates the power to detect QTL using a mix of full-sib and half-sib families for a mass-spawning species. The trait under study has a heritability of 0.5 and the heterozygosity of the QTL is 0.5. We used the variance component analysis to detect QTL for a design with 500 and 700 animals. Pedigrees were structured into 5 paternal families (VC500 pedigree) and 7 paternal families (VC700). In both cases, 15 dams contributed to the offspring. Those results were obtained by simulation using a natural mating population structure described by Brown et al. (2005). In the study by Brown *et al*. (2005) a brood stock of 54 parents was used: 13 males and 21 females contributing to 195 offspring. Proportions of parental contributions were calculated and simulated for a larger population (thousands of individuals) and various brood stock sizes (40, 60 and 80).

For an 80% power, VC500 and VC700 designs were able to detect QTLs that explain 7.6% and 6.2% of phenotypic variation, respectively. **Chapter 3** and **Chapter 4** describe QTL mapping experiments for European sea bass and gilthead sea bream. The characteristics of the simulated pedigrees for sea bass and sea bream are summarized in Table 1. The aim of the experiments in sea

bass and in sea bream was to obtain five sire families of 100 offspring and 15 dams contributing to the offspring, but the sea bass pedigree based on realized data contained 570 animals (70 additional animals). The number of dams in the realized pedigrees was lower than expected based on the simulation. Only two females contributed to offspring in the sea bass experiment and 8 in the sea bream experiment. The total number of full-sib families was therefore much lower than expected; two sire families had approximately 140 progeny and the three others about 95. The family sizes for sea bream were more uneven with one large full-sib family (≈ 150) and several small ones. The discrepancies between the simulated pedigrees based on the population described by Brown *et al*. (2005) and the realized pedigrees obtained in our experiments could be explained by several factors: (i) the population used for simulation was a gilthead sea bream population described by Brown *et al*. (2005) and might not be representative of sea bass reproductive behaviour, (ii) the brood stock used in the sea bream experiment was divided into actually 4 brood stocks taken from the wild at two different locations (Adriatic sea and Mediterranean sea) while only one brood stock was used for the population described by Brown *et al*. (2005) and came from a fish farm and (iii) the sample sizes were larger in our experiment (>1200 fish with assigned parents per experiment) while less than 200 animals composed the sample in Brown *et al*. (2005). Mass-spawning fish do not have the same family structure even within the same species due to variation in male success (Bekkevold *et al*. 2002, Brown *et al*. 2005). VC500 and VC700 were simulated according to the population structure described by Brown *et al*. (2005), while realized pedigrees were taken from one experiment in sea bass (**Chapter 3**) and one in sea bream (**Chapter 4**). Simulated and realized population structures were all different from each other, showing the lack of repeatability between experiments and between species. Difference

observed in population structure can be explained by some stochastic factors (i.e. water temperature or non-contributing parents) and by the random sampling of individuals to obtain the population. Therefore, it is difficult to predict a population structure for natural mating mass-spawning species.

Table 1 – Characteristics of simulated pedigrees (VC500, VC700) and real data set (SEABASS, SEABREAM).

Pedigree	**# offspring**	**#sires**	**#dams**	**# Full-sib families**
VC500	500	5	15	70
VC700	700	7	15	98
SEA BASS	570	5	2	5
SEA BREAM	500	6	8	17

Figure 1 shows the power for the simulated pedigrees and the two realized pedigrees (from experiment described in Chapter 4 and Chapter 5) for a heritability of 0.5 and heterozygosity for the QTL of 0.5. Power was calculated as described in Massault *et al.* (2008). This heritability corresponds to the heritability estimated for body weight in sea bass and body length in sea bream (Antonello *et al.* 2009). The simulated pedigree with 700 animals has higher power than the simulated pedigree with 500 as demonstrated in **Chapter 2**. Two results to point out are the better performance of the sea bream pedigree over VC500 and the better performance of the sea bass pedigree over VC700.

With equal trait heritability and equal heterozygosity, the only variable parameter here is the population structure. At 80% power, VC500 can detect QTLs explaining 7.6%, while the sea bream pedigree, with exactly the same number of animals, can detect QTLs explaining 6.5% of the phenotypic variation. Heterozygosity was set at 0.5, which is the best case scenario in a diallelic system. However, if heterozygosity is lower, realized pedigrees will perform worse than the simulated ones. The number of dams being lower in simulated pedigree, when heterozygosity is low, the chances of segregation for the QTL are lower.

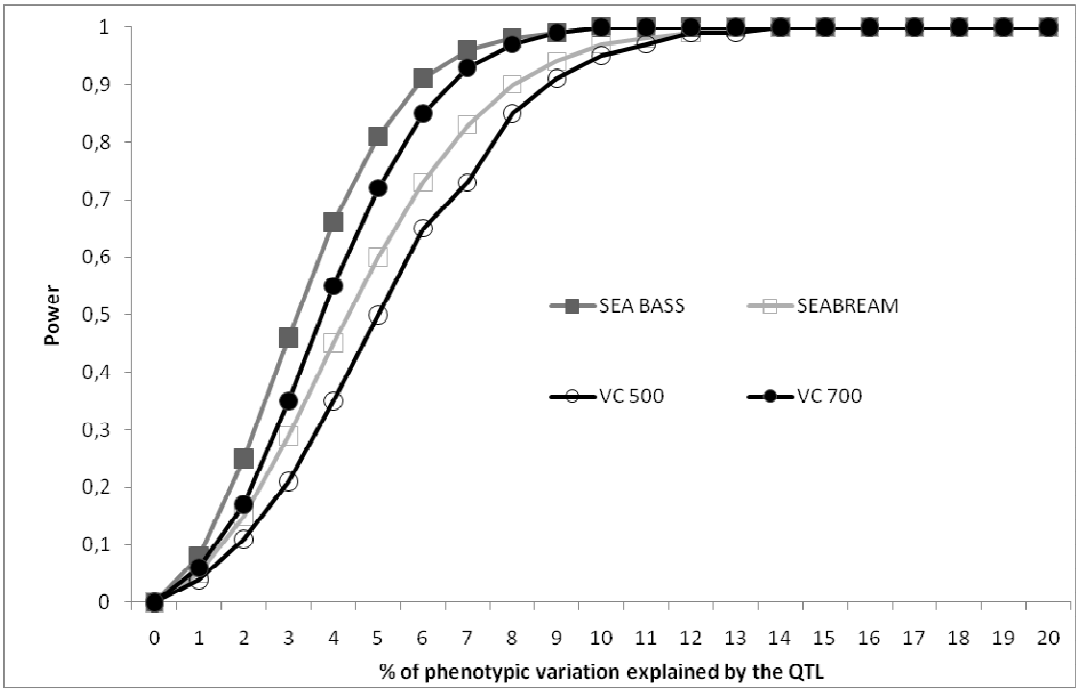

Figure 1 - Deterministic power for simulated and real data in this thesis

.

The realized number of dams contributing to progeny in the sea bream pedigree is about half the number of dams that contributed in the simulated scenario VC500. The same holds true for the sea bass pedigree and the VC700. At 80% power, VC700 pedigree allows the detection of QTLs that explain 6.2%, while the sea bass pedigree can detect QTLs that explain 5% of the phenotypic variation with 130 animals less. Here again, the number of dams contributing to the progeny seems to play an important role in the power to detect QTLs. Having few dams increases the power because more information can be captured from larger families, on condition that QTL are segregating in those families. Therefore, the number of dams in the sea bass pedigree associated with a higher number of progeny and more equal family sizes is more powerful.

Looking back at the results from QTL mapping, estimated effects of significant QTLs were between 8 and 38% in sea bass and between 4 and 8% in sea bream, while simulated pedigree VC500 predicted the detection of QTL sizes as low as 8% and the realized pedigrees predicted 6.5% in sea bass and 5% in sea bream. Although both experimental populations were generated through a natural mating mass-spawning event, population structures were different in sea bream and sea bass. While 6 males and 8 females contribute to the offspring, with uneven family sizes in sea bream, offspring in sea bass were generated by 5 males and 2 females with more equal family sizes. Uneven family size resulted in the detection of smaller QTLs (sea bream). The larger effects observed in sea bass QTL analysis can be explained by the presence of only two females that might not segregate for small QTLs by chance.

To conclude, we found a higher power in the realized pedigree than expected from the simulated pedigrees, due to the lower number of dams contributing to the offspring. The actual number of parents contributing to offspring and

family sizes in natural mass-spawning populations is difficult to predict because of limited knowledge on reproductive behaviour.

6.2 QTL mapping

Efficiency of two-step procedure

Chapter 3 and **Chapter 4** describe a two-step procedure, where first a large sample (i.e. 2000 individuals) is subjected to parentage assignment based on a small number of markers and then the largest full-sib families are fully genotyped and phenotyped for the QTL mapping experiment. Sea bass and sea bream studies identified significant QTLs: 2 for body weight and 6 for morphology in sea bass and, one for body length at time of death and one for survival at day 15 and one for survival to the disease for sea bream. A third experiment has been run with stress response, morphology and body weight on sea bream, where 9 QTLs were detected for morphology (unpublished results). The two step approach was successful in detecting QTLs for both continuous traits (i.e. morphology) and categorical traits (i.e. survival).

However, no significant QTLs were detected for stress response, neither in sea bream nor in sea bass. The heritability of cortisol level, a measure of stress response, was low (0.08), which may be pointing towards the lack of major QTLs. However, reported heritability of cortisol is variable, i.e. moderate to high in rainbow trout (Pottinger *et al.* 1999) and in common carp (Tanck *et al*. 2001) or not significant from zero in Atlantic Salmon (Fevolden *et al*. 1994). Tanck *et al*. (2001) discuss the fact that experimental set-up and model used to estimate heritability can be responsible for over-estimation of heritability. Experimental factors that can influence the level of cortisol are the species

themselves (trout have high heritability and salmon low heritability), fish handling during the blood collection, experimental set up (size of tank, length of stress inflicted and number of fish in the tank, replication, control) or accuracy of measurement. Using identical set up parameters across experiments and an identical procedure of cortisol sampling could decrease the variability of the stress response and allow the comparison between studies.

Results of sea bass stress and disease resistance of sea bream show that the two step experimental design described in **Chapter 2** is valuable using interval mapping. However, results for the third experiment (sea bream stress response) were not conclusive. While the number of animals was roughly the same for the three experiments, the number of microsatellites markers varies (Table 2). For both species, the number of chromosome (haploid) is 24. We may assume that most of the genome has been captured for the stress response in sea bass and the disease resistance in sea bream experiments, as the genetic maps have respectively 20 and 24 linkage groups. This is not the case for the stress response in the sea bream experiment, which covers only 13 linkage groups and 10 markers were not linked to any other marker. The results given by the analysis for this experiment are incomplete due to the low resolution of the genetic map. Some very strong associations were found, where the QTL size was equivalent to the genetic variance. The polygenic component was likely confounded with the QTL component, which captured all the effects.

The QTL mapping was done using interval mapping for the experiment on the stress response of sea bream with low resolution genetic mapping, where a large proportion of markers were unlinked, this method was not successful. A more appropriate approach would have been single marker analysis, which is as powerful as interval mapping (Haley *et al*. 1994) but less confident for the

location of the detected location. Another factor that explains the lack of success in detection of QTLs is the family structure, with 7 males and 74 females for 500 animals. Family sizes were much smaller than for the experiment on sea bream disease resistance stress, with the biggest full-sib family of 18 offspring.

Table 2 - The genetic map characteristics for the three experiments

Experiment	**# markers**	**Map length**	**Marker spacing**	**#linkage group**	**# unlinked markers**	**# fish**
Sea bream stress	51	329.1	12.2	13	10	549
Sea bass stress	90	639	7.7	20	3	570
Sea bream disease	151	1041.3	5.8	24	11	500

Future prospect of QTL mapping

Once a QTL has been detected and the position narrowed to a specific region, the next steps are to undertake fine mapping and confirm the QTL. Confidence intervals given by bootstrap analysis with half-sib regression were large, usually corresponding to an entire linkage group. In those regions, a large number of potential genes are present, but for the moment, very few genomes of fish species have been completely sequenced. Therefore employing a candidate gene approach to narrow the region to a single causative gene is difficult, as candidate genes remain unknown in most cases. However, fine mapping may help to reduce the length of the region. Fine mapping consists of typing a large number of markers in the specific region to increase the number of

recombinations between the markers and QTL and find the true association with a single marker (in the best case scenario) using a large population. A QTL confirmation experiment uses different animals than the original QTL study, and should consider a lower QTL size (usually over-estimated in the discovery experiment), an adequate type I error (α) and other parameters such as experimental designs and expected heterozygosity. A number of QTLs have been detected in various aquaculture species, but to date, very few have been fine-mapped and/or confirmed such as a QTL for IPN resistance, found first by Houston *et al*. (2008) and recently confirmed by Moen *et al.* (2009) or one of the body weight QTL in the sea bass experiment, reported in Chatziplis et al. (2007).

The aim of QTL mapping is to integrate the knowledge on markers associated to a quantitative trait into breeding schemes, known as marker-assisted selection (MAS). Sonesson (2007) demonstrated the possibilities of such a selection method in fish breeding. At the time of the preparation of the European project Aquafirst (2004), QTL mapping and development of breeding programs relying on MAS was an attractive solution. But MAS has been rarely applied in fish breeding to this date (Moen *et al*. 2009).

However, QTLs can be a useful source of biological knowledge. The detection of regions that influence a quantitative trait could lead to the discovery of specific genes responsible for the variation and to a better understanding of physiological pathways. It is particularly interesting concerning stress response and disease resistance pathways, where only few genes are known accurately. Moreover, those two pathways are linked in a way that prolonged stress negatively affects the immune response.

Recently, with the availability of large panels of markers (SNPs especially), genome wide association studies (GWAS) have been widely applied in studies

of humans. GWAS utilizes a very large number of markers to scan the whole genome and find associations with quantitative traits of interest. The main advantages of this method are: (i) it detects association between the trait and a SNP across the whole population (Goddard and Hayes 2009) and not within families like QTL mapping with half-sib regression and (ii) it is powerful for low effective population sizes (Goddard and Hayes 2009). The application of GWAS for mass-spawning species is achievable. The latest sex-averaged genetic maps for gilthead sea bream has a length of 1241.9 centiMorgans (cM) (Franch *et al*. 2006) and for sea bass of 1373.1 cM (Chistiakov *et al*. 2008). If it is assumed that the genome size of a mass-spawning species is 15 Morgans (or 1 500 000 kilobase pairs - kbp), the number of markers to be genotyped is fixed by the distance between two markers, assuming that they are evenly spaced. One cM is equivalent to 1 000 kbp, therefore for a genome size of 1 500 000 kbp and length segment between 2 markers of 50 kbp, 100 kbp and 200 kbp, the number of markers to be genotyped is respectively 7 500, 15 000 and 30 000. The measure of linkage disequilibrium r^2 is given in Table 3 for various effective population sizes and various lengths between two markers. The values were calculated according to Sved (1971).

Table 4 shows the deterministic power of GWAS experiment calculated according to Luo (1998) for two QTL size explaining 5% and 10% of the phenotypic variation. The phenotypic variation was arbitrarily set up at 100, the QTL allele frequency at 0.5 and marker allele frequency at 0.5. Power was calculated for three population sizes (100, 500, 1000) and for three lengths of segment between 2 markers (or three different numbers of markers genotyped) with a type I error of 0.01. However, linkage disequilibrium calculated in Table 3 is deterministic. The actual r^2 is unknown, as mass-spawning species have been recently domesticated (Duarte *et al*. 2007).

Table 3 – Expected linkage disequilibrium (r^2) for effective population size Ne of 50, 100 and 500 and for different lengths of segment between two markers – 50, 100 and 200 kilo base pair (kbp)

	Ne = 50	**Ne=100**	**Ne=500**
50 kbp (30 000 markers)	0.91	0.83	0.50
100 kbp (15 000 markers)	0.83	0.71	0.33
200 kbp (7 500 markers)	0.71	0.56	0.20

The power is high when the extent of LD is high (small effective population and large number of markers genotyped), as well as the sample size and the QTL effect. When only 100 individuals are genotyped, the power of the experimental design is lower, even if r^2 and QTL variance is high. When LD is low, determined by large effective population size (500) and large distance between genotyped markers (200 kbp), the explored experimental designs are not powerful, even with a large sample size and large QTL variance. Even if the number of markers remains limited today for mass-spawning species, GWAS can be achieved with at least an 80% chance to detect QTLs with an effective population size of 100, 7 500 genotyped markers (200 kbp) and 500 animals genotyped. Power can also be influenced by frequencies of marker and QTL sizes, which are determined for a specific trait. Here, we chose two QTL sizes of 5% and 10% of phenotypic variation. But in reality, some QTLs would be as large as those we used and some others will be much smaller. With smaller QTL sizes, even with large population sizes and high r^2, power will be lower and they might not be detected.

GWAS for mass-spawning seems to be powerful to detect association between trait and markers. The complex population structure should not have an important impact because the genome scan is done across the population and within families (Goddard and Hayes 2009). However, it is difficult to predict linkage disequilibrium pattern as no r^2 has been reported either for sea bream or sea bass.

Table 4 – Power of experimental design for GWAS with 3 sample population size (N), three lengths of segment between 2 markers (length) and two QTL size (h^2q)

Length	**N = 100**		**N = 500**		**N = 1 000**	
	h^2q = 0.05	**h^2q = 0.1**	**h^2q = 0.05**	**h^2q = .1**	**h^2q = 0.05**	**h^2q = 0.1**
Ne = 50, 50 kbp	0.23	0.58	0.97	1	1	1
Ne = 100, 50 kbp	0.21	0.52	0.96	1	1	1
Ne = 500, 50 kbp	0.10	0.27	0.75	0.99	0.98	1
Ne = 50, 100 kbp	0.21	0.52	0.96	1	1	1
Ne=100, 100 kbp	0.17	0.43	0.91	1	1	1
Ne=500, 100 kbp	0.06	0.15	0.51	0.89	0.88	0.99
Ne = 50, 200 kbp	0.17	0.43	0.91	1	1	1
Ne=100, 200 kbp	0.12	0.31	**0.81**	**0.99**	0.99	1
Ne=500, 200 kbp	0.03	0.08	0.27	0.62	0.62	0.95

6.3 Selective breeding

Chapter 5 reports genetic gain, inbreeding rate and accuracy of selection for genomic selection (GBLUP), BLUP selection and mass selection, without restriction on inbreeding for various heritabilities, various numbers of selection candidates, various intensities of selection and various mating designs.

Marker assisted selection

Chapter 5 mentions three main selective breeding methods: phenotypic selection, BLUP selection and genomic selection. Before the concept of genomic selection appeared (Meuwissen *et al.* 2001), studies on selective breeding were mainly focused on how to implement marker assisted selection (MAS).

In addition to the three methods of selection presented in **Chapter 5**, we simulated a breeding program including a known QTL (marker assisted selection) using a controlled mating design. Sixteen males were mated to thirty-two unrelated females, forming 32 full-sib families of 8 males and 8 females offspring (512 in total). At each new generation, intensity of selection was 0.0625 for males (16) and 0.125 for females (32). We simulated one QTL, which explained 25% of the genotypic variance, therefore explaining 5% of phenotypic variance. To include the effect of the QTL on the phenotype, the quantitative trait is correlated with both the polygenic component and the QTL as described by Schrooten *et a*l. (2005). Genetic variance was set at 15 to compare with the results from **Chapter 5** and genetic gain, rate of inbreeding and accuracy were calculated per generation deterministically using SelAction (Rutten *et al*. 2002). Table 5 reports genetic gain, rate of inbreeding and accuracy of selection for MAS, BLUP, GBLUP and mass selection.

From the results presented above, genetic response generated by MAS selection is higher than BLUP, but lower than GBLUP. Inbreeding rate in BLUP is higher than GBLUP, while inbreeding rate in Mass is lower than GBLUP. While MAS is expected to be efficient for a controlled mating design, this method is not widely used in commercial populations of fish. The attention has shifted to genomic selection for aquaculture breeding schemes in the past few years (Sonesson and Meuwissen 2009). The limited number of studies focused on QTL detection and validation could be an explanation for the lack of implementation of MAS in breeding schemes. GWAS offers the possibility to detect a large number of QTLs, without the need of a validation step. The sum of QTL effects can explain a large part of genetic variance. Therefore, more genetic gain can be achieved, while spending less time on experimentation to detect and validate QTLs. But implementation of marker-assisted selection might be problematic in the long term with large QTLs, because of the possibility of hitch-hiking due to strong selection in the first generations (Hospital 1997).

Inbreeding for natural mating mass spawning species

As shown in **Chapter 5**, the inbreeding rate is very high with any selection method used. Inbreeding rate decreases when we apply a controlled mating design. Sonesson and Meuwissen (2000) explored a number of options to constrain inbreeding rate to 1% per generation using the optimum contribution method (Meuwissen 1997), where numbers of selected individuals for the next generations varies from generation to generation to maintain inbreeding rate at a fixed value. The proposed strategies of Sonesson and Meuwissen (2000) to restrict inbreeding resulted in higher genetic improvement compared to random mating. But all solutions required the possibility to choose which males to mate

Table 5 - Genetic gain (ΔG) , increase of inbreeding rate (ΔF) and accuracy of mass (R_{gg}) , MAS, BLUP and GBLUP selection method for controlled mating (16 males X 32 females, 16 offspring per full-sib families), heritability of 0.2

	ΔG	**ΔF**	**Rgg**
Mass	1.72	1.70%	0.37
BLUP	2.23	6.30%	0.44
MAS	2.62	1.60%	0.77
GBLUP	2.92	4.70%	0.52

to which females, which is not possible in natural mating mass-spawning species. Therefore optimum contribution is one way to restrict inbreeding which can be applied to natural mating mass-spawning species. However, reducing the number of selected animals to constrain the inbreeding rate could be disastrous for mass-spawning species, because of the variability in family size and uneven contribution of parents. Dominant parents will contribute to more offspring and potential candidates from small families will be eliminated. A higher number of offspring from the largest families are therefore expected to be selected for the next generation and inbreeding will increase. The problem could be avoided by increasing the population size considerably. Another solution that could be effective in natural mating fish would be to restrict the number of offspring selected per sire (or dams) but keeping the number of animals selected per generation constant. This would lead to less genetic improvement compared to no restriction on inbreeding. Finally, keeping several

separate brood stock and select animals from all brood stock can help to control inbreeding (Blonk *et al.* 2009). However, in practice, the implementation of this solution will require a higher number of tanks and the handling of larger populations, which might not be always feasible in fish farms (cost and space to have several tanks).

6.4. Parentage assignment

Parentage assignment is a necessary step to be able to reconstruct pedigrees of mass-spawning species. Natural mating is still widely used in fish farms to produce the next generation. This implies that offspring are kept in the same tank, where it is impossible to distinguish families. For QTL mapping and selection purpose, it is essential to perform parentage assignment and retrieve the original pedigree. It requires first to genotype all offspring for a panel of markers and then applying statistical methods to allocate parents to offspring. The parentage assignment for the sea bream experiment described in **Chapter 4** used 7 microsatellites for 1753 fish (Antonello *et al.* 2009). Out of the 1753 genotyped fish, parents were allocated for 1257 (roughly 70%). The statistical method used (exclusion method) here did not allow for any genotyping errors and therefore decreases the number of allocations. Using less stringent thresholds (likelihood approach), the authors could assign parents for nearly 1500, but with less certainty. Using a higher number of microsatellites will increase the proportion of correct assignment (i.e. 15 markers). It is important to consider that, out of the microsatellites used; some will be excluded because they do not show polymorphism. Antonello et al. (2009) discarded markers from the 9 that were originally typed. As genotyping costs are

decreasing,adding more markers to increase the accuracy of parentage assignment is feasible.

The pedigree can also be reconstructed by measuring molecular relatedness (Toro *et al*. 2002), where relatedness is calculated as the similarity between marker alleles (Li *et al*. 1993). Blonk *et al.* (2010) showed, that pedigree estimation using the molecular relatedness method is more appropriate to natural mating mass-spawning species, due to skewed parental contribution to offspring and variable family sizes.

Errors in parentage assignment could have a large impact when using a selection method that relies on pedigree, such as BLUP as mentioned by Blonk *et al.* (2010). The breeding values will be over- or under-estimated for an animal, which is assigned to the wrong parents. But pedigree reconstruction is not necessary for mass selection and genomic selection. The practical and time limitations raise the question whether or not pedigree reconstruction is a necessary step in selection. Mass selection is solely based the record of its own performance, while genomic selection uses marker data to evaluate the relationship between individuals without a pedigree. However, the lack of pedigree information results in higher inbreeding in mass selection, because sibs will have a higher chance of being selected, especially for high heritability traits. The information provided by the pedigree is valuable to restrict inbreeding in mass selection. In the case of genomic selection, the information from the genomic relationship matrix essential for the breeding value estimation replaces the pedigree information and that information can also be used to minimize genomic relationship between selected individuals to restrict inbreeding. Therefore no pedigree in parallel is necessarily needed to constrain inbreeding.

Conclusion

Despite the inherent difficulties linked to skewed parental contribution to offspring and variable family size, selective breeding in natural mating mass-spawning is feasible through a common selection method (i.e. mass, BLUP and MAS selection). Genomic selection is also promising as it works at the population level, not at the family level for the detection of QTLs (GWAS). However, in any of the selection method studied in this thesis, inbreeding rate is high and efforts need to be made to constrain inbreeding rate for natural mating mass-spawning species.

References

Antonello, J., Massault, C., Franch, R., Haley, C., Pellizzari, C., Bovo, G., Patarnello, T., de Koning, D. J., Bargelloni, L. (2009) Estimates of heritability and genetic correlation for body length and resistance to fish pasteurellosis in the gilthead sea bream (*Sparus aurata L.*). *Aquaculture* **298,** 29-35.

Blonk, R. J. W., Komen, H., Kamstra, A., van Arendonk, J. A. M. (2010) Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, Solea *solea*. *Genetics* **184,** 213-219.

Blonk, R. J. W., Komen, J., Kamstra, A., Crooijmans, R., van Arendonk, J. A. M. (2009) Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. *Aquaculture* **289,** 26-31.

Brown, C. R., Woolliams, J. A., McAndrew, B. J. (2005) Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture* **247,** 219-225.

Chatziplis, D., Batargias, C., Tsigenopoulos, C. S., Magoulas, A., Kollias, S., Kotoulas, G., Volckaert, F. A. M., Haley, C. S. (2007) Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): The BASSMAP pilot study. *Aquaculture* **272,** S172-S182.

Chistiakov, D. A., Tsigenopoulos, C. S., Lagnel, J., Guo, Y. M., Hellemans, B., Haley, C. S., Volckaert, F. A. M., Kotoulas, G. (2008) A combined AFLP

and microsatellite linkage map and pilot comparative genomic analysis of European sea bass Dicentrarchus labrax L. *Animal Genetics* **39,** 623-634.

Duarte , C., Marbà, N. Holmer, M. (2007) Rapid domestication of marine species. *Science* **316**, 382-383

Fessehaye, Y., El-Bialy, Z., Rezk, M. A., Crooijmans, R., Bovenhuis, H., Komen, H. (2006) Mating systems and male reproductive success in Nile tilapia (Oreochromis niloticus) in breeding hapas: A microsatellite analysis. *Aquaculture* **256,** 148-158.

Fevolden, S. E., Roed, K. H., Gjerde, B. (1994) Genetic Components of Poststress Cortisol and Lysozyme Activity in Atlantic Salmon - Correlations to Disease Resistance. *Fish & Shellfish Immunology* **4,** 507-519.

Franch, R., Louro, B., Tsalavouta, M., Chatziplis, D., Tsigenopoulos, C. S., Sarropoulou, E., Antonello, J., Magoulas, A., Mylonas, C. C., Babbucci, M., Patarnello, T., Power, D. M., Kotoulas, G., Bargelloni, L. (2006) A genetic linkage map of the hermaphrodite teleost fish Sparus aurata L. *Genetics* **174,** 851-861.

Goddard, M. E., Hayes, B. J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10,** 381-391.

Haley, C. S., Knott, S. A., Elsen, J. M. (1994) Mapping quantitative trait loci in crosses between outbred lines using least-squares. *Genetics* **136,** 1195-1207.

Hospital, F., Moreau, L., Lacoudre, F., Charcosset, A, Gallais, A. (1997) More on the efficiency of marker assisted selection. *Theor. Appl. Genet*. **95**, 1181-1189

Houston, R. D., Gheyas, A., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., McAndrew, B. J., Haley, C. S., Bishop, S. C., 2008 Detection and cof a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (*Salmo salar*), pp. 199-204 in *Animal Genomics for Animal Health,* edited by Pinard, M. H., Gay, C., Pastoret, P. P. and Dodet, B. Karger, Basel.

Lallias, D., Gomez-Raya, L., Haley, C. S., Arzul, I., Heurtebise, S., Beaumont, A. R., Boudry, P., Lapegue, S. (2009) Combining two-stage testing and interval mapping strategies to detect QTL for resistance to bonamiosis in the european flat oyster *Ostrea edulis*. *Mar Biotechnol (NY)* **11,** 570-584.

Li, C. C., Weeks, D. E., Chakravarti, A. (1993) similarity of dna fingerprints due to chance and relatedness. *Human Heredity* **43,** 45-52.

Luo, Z. W. (1998) Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity* **80,** 198-208.

Massault, C., Bovenhuis, H., Haley, C., de Koning, D. J. (2008) QTL mapping designs for aquaculture. *Aquaculture* **285,** 23-29.

Meuwissen, T. H. E. (1997) Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* **75,** 934-940.

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157,** 1819-1829.

Moen, T., Fjalestad, K. T., Munck, H., Gomez-Raya, L. (2004) A multistage testing strategy for detection of quantitative trait loci affecting disease resistance in Atlantic salmon. *Genetics* **167,** 851-858.

Moen T., Baranski, M., Sonesson A.K. & Kjøglum S. (2009) Confirmation and fine-mapping to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 2009, **10:**368.

Pottinger, T. G., Carrick, T. R. (1999) Modification of the plasma cortisol response to stress in rainbow trout by selective breeding. *General and Comparative Endocrinology* **116,** 122-132.

Rutten, M. J. M., Bijma, P., Woolliams, J. A., van Arendonk, J. A. M. (2002) SelAction: Software to predict selection response and rate of inbreeding in livestock breeding programs. *Journal of Heredity* **93,** 456-458.

Sauvage, C.C., Bourdy, P., de Koning D.J., Haley C., Heurtebise, S., Lapuègue, S. Quantitative Trait Loci for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*) In press

Schrooten, C., Bovenhuis, H., van Arendonk, J. A. M., Bijma, P. (2005) Genetic progress in multistage dairy cattle breeding schemes using genetic markers. *Journal of Dairy Science* **88,** 1569-1581.

Sonesson, A. K. (2007) Marker-assisted selection in aquaculture breeding programs. *Aquaculture* **272,** S311-S311.

Sonesson, A. K., Meuwissen, T. H. E. (2000) Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution* **32,** 231-248.

Sonesson, A. K., Meuwissen, T. H. E. (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution* **41**, 1-8

Sved, J. A. (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* **2,** 125-141.

Tanck, M. W. T., Vermeulen, K. J., Bovenhuis, H., Komen, H. (2001) Heredity of stress-related cortisol response in androgenetic common carp (Cyprinus carpio L.). *Aquaculture* **199,** 283-294.

Toro, M., Barragan, C., Ovilo, C., Rodriganez, J., Rodriguez, C., Silio, L. (2002) Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetics* **3,** 309-320.

SUMMARY

Several aquaculture fish species reproduce through mass-spawning which allows little control over reproduction and generates a complex population structure with unequal family sizes and uneven parental contributions. As a result, mass selection is the most common method for selective breeding in these fish species at present. However, there is growing interest for genetic improvement of natural mating mass-spawning species and, following the availability of large panels of genetics markers, implementation of selection based on genomic information (marker-assisted selection or genomic selection) becomes a viable option.

This thesis investigates different aspects of utilizing genomic information in selective breeding of natural mating mass-spawning species, starting with the design of QTL mapping experiments (**Chapter 2**), followed by the actual results from QTL mapping experiments in European sea bass and gilthead sea bream for stress response and disease resistance (**Chapter 3 and Chapter 4**). Finally, we compared selection response and rate of inbreeding of various selection methods for natural mating mass-spawning species (**Chapter 5**).

Chapter 2 describes three experimental designs to perform QTL detection in aquaculture species: “hierarchical design” where two divergent lines are crossed, “large full-sib families design” where selective genotyping can be applied and “mass-spawning design” for species such as sea bream and sea bass, where natural mating is used. We concluded that family structure and QTL size have a large impact on the power to detect QTL, in all three designs, while the polygenic heritability has a small impact on the power. Marker

density does not greatly affect the power when the distance between markers is less than 10 cM; but ideally spacing should not exceed 20 cM. For each of the systems studied, it is possible to design an experiment that would have an 80% power to detect a QTL of moderate effect (explaining between 1.5 and 5% of the trait variation) by genotyping 1000 or fewer individuals.

Chapter 3 and **Chapter 4** summarize the results of QTL mapping for stress response and disease resistance in sea bream and sea bass. **Chapter 3** focuses on linkage analysis performed for stress response and two additional traits (body weight and morphology) in sea bass using the 2-step design for mass-spawning species presented in **Chapter 2**. 2122 fish were subjected to confinement stress followed by measures of cortisol, body weight and a range of body proportions. The pedigree was reconstructed on the basis of 31 genetic markers. 570 fish from the five largest families were selected for QTL mapping and genotyped for 67 additional markers. Two statistical methods to detect QTL were used: half-sib regression analysis (paternal and maternal) and a variance component analysis accounting for all family relationships. Two significant QTLs were detected for body weight, located on linkage group 4 and 6. Six significant QTLs were detected for morphology, located on linkage group 1B, 4, 6, 7, 15 and 24. Suggestive QTLs were detected for stress response located on linkage group 3, 14 and 23. The QTLs explained between 8% to 38% of the phenotypic variance. **Chapter 4** uses the same methodology as described in **Chapter 3** to detect QTL for disease resistance in gilthead sea bream. After exposure to *Photobacterium damselae subsp. piscicida*, day of death and body length at time of death were recorded. Parentage assignment was performed using a panel of 7 microsatellite markers. 500 fish from the largest families were selected for genome-wide genotyping using 151 microsatellite markers.

Besides ‘day of death’ and ‘body length at time of death’, two binary traits were defined to investigate whether ‘early’ and ‘late’ survival were under different genetic control: survival at day 7 and survival at day 15. We detected two significant QTL for disease resistance: one on linkage group 3 for survival at day 15 (‘late’ survival) and one on linkage group 21 for day of death (overall survival) explaining 6% and 4% of the phenotypic variation, respectively. Marker Id13 was identified as being linked to a QTL with an effect on survival after exposure to *Photobacterium damselae subsp. piscicida*. A large significant QTL was detected for body length on linkage group 6, explaining 5-8% of the phenotypic variation.

Chapter 5 compared selective breeding using three methods of selection in natural mating mass-spawning: selection based on own performance (mass selection), selection using own performance and performance of family members (BLUP, i.e. assuming family relations have been reconstructed) and selection based on own performance and performance of family members using a genomic relationship matrix (GBLUP). Natural mating populations were simulated to reproduce the population structure observed in gilthead sea bream and selection was for 10 generations without applying any restrictions on the increase of inbreeding. We simulated breeding schemes that differed with respect to the number of selection candidates, the number of selected parents, the heritability of the trait and the mating scheme (natural or controlled mating). The rate of inbreeding was high in all simulated scenarios. Lowest rates of inbreeding were found when mass selection was used. BLUP resulted in a higher rate of inbreeding than GBLUP. Average genetic merit in generation 10 was always highest for GBLUP whereas BLUP was superior to mass selection

only at lower heritabilities. We consider GBLUP a more efficient method, but steps must be taken to constrain inbreeding rate.

Chapter 6 discusses four main issues: experimental design for QTL mapping in fish species, future prospects of QTL mapping experiments, selective breeding using genomic information and finally parentage assignment. We focused on two main experimental designs: selective genotyping and the two-step design for natural mating mass-spawning species. Selective genotyping is advantageous if phenotypes are available on a large number of animals, but is not very efficient if interest is in multiple traits or binary traits (i.e. survival). The two-step design for natural mating mass-spawning species is complicated by the upredictability of family sizes. With our 2 QTL mapping studies, we demonstrated that the 2-step procedure is efficient for mass-spawning species. However, it is not very powerful when using low resolution genetics maps or if traits have low heritabilities. Genome wide association studies (GWAS) could be an alternative to the 2-step design: the efficiency of the method is less dependent on the family structure and it is more powerful than linkage studies. However, application of GWAS requires a large number of markers.

The comparison of MAS with BLUP and GBLUP shows that MAS using a QTL explaining 25% of the genetic variation results in a relatively high response to selection, the smallest inbreeding rate and the highest accuracy. However, few reports on QTLs of this size in fish limit the application of this method. Inbreeding is a major issue in selective breeding in mass-spawning species. Limiting the number of offspring per parent or using multiple spawning units could limit the rate of inbreeding.

Despite the inherent difficulties linked to unequal parental contributions, selective breeding in mass-spawning species while restricting the rate of inbreeding is challenging but feasible.

SAMENVATTING

Diverse vissoorten planten zich natuurlijk en in groepen voort. Er is weinig controle op deze vorm van reproductie en dit leidt tot een ingewikkelde familiestructuur met grote verschillen in familiegrootte en ongelijke bijdragen van ouders. Het gevolg is dat selectie op grond van eigen prestatie de meest gebruikte selectiemethode is bij deze vissoorten. Er is echter groeiende belangstelling voor de genetische verbetering van soorten die zich natuurlijk en in groepen voortplanten. De beschikbaarheid van grote hoeveelheden genetische merkers brengt ook alternatieve selectiestrategieën zoals merker ondersteunde selectie of genomische selectie binnen handbereik.

Dit proefschrift onderzoekt verschillende aspecten die gerelateerd zijn aan de implementatie van selectie met behulp van merkers voor vissoorten die zich natuurlijk en in groepen voortplanten. Hoofdstuk 2 handelt over het efficiënt opzetten van QTL detectie experimenten in vis en schaaldieren. Vervolgens worden in hoofdstuk 3 en 4 QTL detectie experimenten beschreven in de Europese zeebaars en de goudbrasem voor stress respons en ziekteresistentie. Ten slotte onderzochten we de selectierespons en de inteelttoename van verschillende selectiestrategieën voor soorten die zich natuurlijk en in groepen voortplanten (Hoofdstuk 5).

Hoofdstuk 2 beschrijft drie experimentele ontwerpen voor QTL detectie: "het hiërarchische ontwerp" waarin twee genetisch verschillende lijnen worden gekruist, "het ontwerp met grote full-sib families" waar selectieve genotypering kan worden toegepast binnen families en "het ontwerp voor soorten die zich natuurlijk en in groepen voortplanten". Dit laatste ontwerp is geschikt voor

soorten zoals de goudbrasem en de zeebaars waar natuurlijke voortplanting wordt gebruikt. We concluderen dat in alle drie ontwerpen familie structuur en de grootte van de QTL effecten een belangrijk effect hebben op het onderscheidend vermogen. De polygene erfelijkheidsgraad heeft een gering effect op het onderscheidend vermogen. De merkerdichtheid heeft een gering effect op het onderscheidend vermogen wanneer de merkerafstand kleiner is dan 10 cM. De optimale merkerafstand moet niet groter zijn dan 20 cM. Voor de onderzochte ontwerpen is het mogelijk om met 1000 of minder individuen een QTL dat 1.5 tot 5% van de variatie verklaart met 80% zekerheid opsporen.

In hoofdstuk 3 en hoofdstuk 4 worden de resultaten van QTL detectie voor stress respons en ziekteresistentie in goudbrasem en zeebaars beschreven. Hoofdstuk 3 richt zich op een koppelingsanalyse voor de reactie op insluitingstress en twee aanvullende kenmerken (lichaamsgewicht en morfologie) in zeebaars. Hierbij is gebruik gemaakt van het 2-staps ontwerp voor soorten die zich natuurlijk en in groepen voortplanten zoals gepresenteerd in hoofdstuk 2. 2212 vissen werden ingesloten, wat resulteert in stress, en dit werd gevolgd door een cortisolmeting en het vaststellen van het gewicht en enkele lichaamsmaten. De afstamming werd gereconstrueerd op grond van 31 microsatelliet merkers. Hierna werden 570 vissen van de vijf grootste families gegenotypeerd voor 67 extra microsatelliet merkers en deze vissen werden vervolgens gebruikt voor de QTL detectie. Twee statistische methoden voor QTL detectie werden gebruikt: de half-sib regressie analyse en de variantiecomponent analyse die gebruik maakt van alle familierelaties. Twee significante QTLs werden gevonden voor lichaamsgewicht op koppelingsgroep 4 en 6. Zes significante QTLs werden gevonden op koppelingsgroep 1b, 4, 6, 7, 15 en 24 voor morfologische kenmerken. Suggestieve QTLs werden gevonden

op koppelingsgroep 3, 14 en 23 voor stress respons. De QTLs verklaren tussen de 8% en de 38% van de fenotypische variatie.

Hoofdstuk 4 maakt gebruik van dezelfde methodiek als beschreven in hoofdstuk 3 om QTLs voor ziekteresistentie op te sporen in goudbrasem. Na blootstelling aan *Photobacterium damselae subsp. piscicida*, werden levensduur (in dagen, gemeten vanaf het moment van blootstelling) en de lichaamslengte ten tijde van overlijden geregistreerd. De ouders van de 1753 vissen werden toegewezen op grond van de informatie van 7 microsatelliet merkers. 500 vissen van de grootste families werden geselecteerd voor typering met 151 microsatelliet merkers. Naast levensduur en de lichaamslengte ten tijde van overlijden werden ook twee binaire eigenschappen gedefinieerd om te onderzoeken of overleving gedurende het begin van het experiment en overleving aan het eind van het experiment genetisch gezien dezelfde eigenschap is: overleving op dag 7 en overleving op dag 15. We detecteerden twee significante QTLs voor ziekteresistentie: één op koppelingsgroep 3 voor overleving op dag 15 en één op koppelingsgroep 21 voor levensduur. Het QTL op koppelingsgroep 3 verklaart 6% van de fenotypische variatie en het QTL op koppelingsgroep 21 verklaart 4% van de fenotypische variatie. Merker Id13 op koppelingsgroep 21 kon worden geïdentificeerd als zijnde gekoppeld aan een QTL met een effect op levensduur na blootstelling aan *Photobacterium damselae subsp. Piscicida*. Een QTL met een groot en significant effect op lichaamslengte kon worden gelokaliseerd op koppelingsgroep 6. Het QTL verklaart 6-8% van de fenotypische variatie.

In hoofdstuk 5 worden 3 verschillende selectiemethoden voor vissoorten die zich natuurlijk en in groepen voortplanten vergeleken: selectie op eigen prestatie, selectie op grond van eigen prestatie en op basis van de prestaties van

familieleden (BLUP, aannemende dat familierelaties worden gereconstrueerd) en selectie op grond van eigen prestatie en op basis van de prestaties van familieleden gebruikmakend van een genomische relatiematrix (GBLUP). Een zich natuurlijk voortplantende populatie werd gesimuleerd op basis van de informatie verkregen van de goudbrasem. De populatie werd geselecteerd gedurende 10 generaties zonder dat er enige beperkingen werden gesteld wat betreft de inteelttoename. Fokprogramma's werden gesimuleerd die verschilden wat betreft het aantal selectiekandidaten, het aantal geselecteerde ouders, de erfelijkheidsgraad van het kenmerk en het paringsschema (natuurlijke paring of gecontroleerde paring). De inteelttoename was hoog in alle gesimuleerde situaties. BLUP resulteert in een grotere inteelttoename dan GBLUP. Het genetisch niveau in generatie 10 was altijd het hoogst voor GBLUP. BLUP resulteert in een hoger genetisch niveau in generatie 10 dan selectie op grond van eigen prestatie wanneer de erfelijkheidsgraad van het kenmerk relatief laag is. We beschouwen GBLUP als een efficiënte selectie methode maar dan moeten er wel maatregelen worden genomen om de inteelttoename te beperken.

Hoofdstuk 6 bediscussieert vier belangrijke punten voor genetisch onderzoek in vissen: experimentele ontwerpen voor QTL detectie, de toekomst van QTL detectie experimenten, selectie met gebruik van genomische informatie en tenslotte de reconstructie van afstamming. We concentreerden ons op twee belangrijke experimentele ontwerpen: het selectief genotyperen en het twee-stappen ontwerp voor QTL detectie in soorten die zich natuurlijk en in groepen voortplanten. Het selectief genotyperen is voordelig wanneer fenotypes van een groot aantal dieren beschikbaar zijn, maar niet erg geschikt wanneer interesse bestaat in het detecteren van genen voor meerdere kenmerken of wanneer het een binair kenmerk (bijvoorbeeld overleving) betreft. Het twee-stappen ontwerp

voor QTL detectie in soorten die zich natuurlijk en in groepen voortplanten wordt bemoeilijkt door de onvoorspelbaarheid van de familiegrootte. We hebben met onze twee QTL detectie studies aangetoond dat het twee-stappen ontwerp toepasbaar is. Het onderscheidend vermogen is echter gering wanneer de merkerresolutie laag is of wanneer de erfelijkheidsgraad van het kenmerk laag is. Wanneer in de toekomst voldoende genetische merkers beschikbaar zijn dan is een associatiestudie een goed alternatief. Associatiestudies zijn minder afhankelijk van familiestructuur en hebben een beter onderscheidend vermogen.

Vergelijking van merker ondersteunde selectie (MAS) met BLUP en GBLUP laat zien dat MAS gebruikmakend van een QTL dat 25% van de genetische variatie verklaart resulteert in een hoge selectierespons, de laagste inteelttoename en de hoogste nauwkeurigheid van selectie. Er zijn echter weinig studies die QTLs van een dergelijke omvang melden en dit beperkt de toepassing van merker-ondersteunde selectie. De inteelttoename is een belangrijk probleem bij het selectief verbeteren van soorten die zich natuurlijk en in groepen voortplanten. Het beperken van het aantal nakomelingen per ouder of het gebruik maken van meerdere voortplantingsgroepen kan de inteelttoename beperken.

Het selectief verbeteren van zich natuurlijk en in groepen voortplantende soorten terwijl de inteelt wordt beperkt is een uitdaging maar is zeker niet onmogelijk.

CURRICULUM VITAE

Cécile was born in Lyon (France) the 14th of August 1982. After passing her final exams in science, she undertook a 2-years diploma at the Technological University of Lyon I in computer science. She moved to Edinburgh (Scotland) to finish her Bachelor of Engineering in Software Engineering at Napier University and then completed a master of science in Bioinformatics at Heriot-Watt University, also in Edinburgh. In January 2006, she started a joint Ph.d project in Animal Breeding and Genetics in aquaculture between the Roslin Institute (Edinburgh) and Wageningen University.

LIST OF PUBLICATIONS

Peered – reviewed article

Massault, C., Bovenhuis, H., Haley, C., de Koning, D.J. (2008) QTL mapping designs for aquaculture. *Aquaculture* **285**, 23-29.

Massault, C., Hellemans, B., Louro, B., Bartagias, C., Van Houdt, J., Canario, A., Volckaert,F., Bovenhuis, H., Haley, C., de Koning, D.J. QTLs for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. Animal Genetics (DOI: 10.1111/j.1365-2052.2009.02010.x)

Massault, C., Franch, R., Haley, C., de Koning, C., Bovenhuis, H., Pillizzari, C., Partanello, T., Bargelloni, T. QTLs for resistance to fish pasteurellosis in gilthead sea bream *Sparus aurata* Animal Genetics (*under eview*)

Canario, A.V.M., Bargelloni, L., Volckaert, F., Houston, R.D., Massault, C., Guiguen Y. (2008) Genomics Toolbox for Farmed Fish. *Reviews in Fisheries Science* **16**, 3-15.

Antonello, J., Massault, C., Franch, R., Haley, C., Pellizzari, C., Bovo, G., Patarnello, T., de Koning, D., Bargelloni, L. Estimates of heritability and genetic correlation for body length and resistance to fish pasteurellosis in the gilthead sea bream (*Sparus aurata* L.). *Aquaculture* **298,** 29-35

Conference proceedings/abstract

Massault, C., Bovenhuis, H., Haley, C. and De Koning, D. J. (2007) QTL mapping designs for species in aquaculture. *Aquaculture* **272,** S289-S289.

Massault, C., Bovenhuis, H., Haley, C., de Koning, D.J. Marker Assisted Selection in Aquaculture Breeding Programmes. *3rd International Conference of Quanatitative Genetics, August 2007, Hanghzhou, China*

Massault, C., Hellemans, B., Louro, B., Bartagias, C., Van Houdt, J., Canario, A., Volckaert, F., Bovenhuis, H., Haley, C., de Koning, D.J. QTLs for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. *IX International Symposium of Aquaculture Genetics, June 2009, Bangkok, Thailand*

ACKNOWLEDGEMENT

Four years have passed… Now it is my turn to thank all of you that gave me support in one way or another throughout the thesis.

I am sincerely grateful for the advices and support from “all” my supervisors, Chris, D.J., Henk and Johan, who have always been present, even when I was away. Chris, you have been very helpful with the project and the analysis.

I really appreciated the atmosphere in the North Wing Extension and the time that I spent there, even with the many sarcastic comments about arrogant french people during lunch! Thanks for the social interactions and for the answers to my numerous work related questions..

I would like to thank everybody for the constructive discussions from the Animal Breeding and Genomic Centre, where everything and anything is a good reason to share cake at coffee breaks.

Thanks to my family, who has been patient in the last months of the thesis writing, and to all my friends for their support.

Finally, special thanks to Chris and D.J. for giving me the opportunity to enter in the scientific world, by supervising my master thesis at the Roslin Institute. It triggered quite a change on my professional interests.

Training and Supervision Plan — **Graduate School WIAS**

Name PhD student	Cécile Massault
Groups	Animal Breeding and Genomics Centre / Roslin Institute
Daily supervisor(s)	Henk Bovenhuis and D.J. de Koning
Supervisor(s)	Johan Van Arendonk and Chris Haley
Project term	January 2006 to January 2010

The Basic Package	year	credits
WIAS Introduction Course	2007	1,5
Ethics course (University of Edinburgh)	2007	1,5
Subtotal		**3,00**
Scientific Exposure		
International conferences		
International Symposium Genetics in Aquaculture IX, Montpellier (France), 25/06/2006 - 30/06/2006	2006	1,5
3rd international conference on quantitative genetics, Hangzhou (China), 19/08/2007 - 24/08/2007	2007	1,5
3rd international symposium on Animal Functional Genomics, Edinburgh (UK), 07/04/2006 - 09/04/2008	2008	0,9
International Symposium Genetics in Aquaculture X, Bangkok (Thailand), 22/06/2009 - 26/06/2008	2009	1,5
Seminars and workshops		
Aquafirst Meeting Leuven, Belgium	2006	0,6
Conservation genetics, Wageningen, Netherlands	2006	0,3
Aquafirst Meeting La Rochelle, France	2008	0,6
Conservation genetics, Wageningen, Netherlands	2008	0,3
Aquafirst Meeting Paris,France	2009	0,6
Genetics in milk quality, Wageningen, Netherlands	2009	0,3
Presentations		
IX International Symposium Genetics in Aquaculture, Montpellier, France (oral)	2006	1,0
Genesis Faraday annual event, Cambridge ,UK (poster)	2007	1,0
3rd international symposium on quantitative genetics, Hangzhou,China (poster)	2008	1,0
WIAS science day, Wageningen, Netherlands (oral)	2009	1,0
X International Symposium Genetics in Aquaculture, Bangkok, Thailand (oral)	2009	1,0
Subtotal		**13,10**
In-Depth Studies		
Disciplinary and interdisciplinary courses		
Marine Genomics course, Roscoff, France	2007	3,0
QTL Mapping, MAS and Genomics selection, Lelystad, Netherlands	2008	1,5
Advanced statistics courses		
Experimental design, Wageningen, Netherlands	2007	1,0
MSc level courses		
Animal breeding and genetics, Wageningen, Netherlands	2006	6,0
Genetic improvement of livestock, Wageningen, Netherlands	2006	6,0
Bioinformatics and Genome analysis modules for Msc in Quantitative Genetics and Genome Analysis (University of Edinburgh, UK)	2006	1,0
Subtotal		**18,50**
Professional Skills Support Courses		
Public Speaking Training Course, Roslin, UK	2006	0,6
Course Techniques for Scientific Writing, Wageningen, Netherlands	2009	1,2
Genesis Faraday course - Group Dynamics, Edinburgh, UK	2008	0,6
Genesis Faraday course - Intellectual property, Edinburgh, UK	2009	0,6
Subtotal		**3,00**
Education and Training Total		**34,60**

COLOPHON

This thesis was conducted jointly at The Roslin Institute and R(D)SVS, University of Edinbugh, Roslin, UK, and at the Animal Breeding and Genomics Centre, Wageningen University, financed partly by the European Commission (project AQUAFIRST, contract number FP6-STREP-2004-513692) and the Animal Breeding and Genomics Centre, Wageningen University. KTN Biosciences (previously Genesis Faraday) contributed financially to the some of the trainings and conferences costs.

The printing of the thesis was funded by the Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands

Printed by: Wöhrmann Print Service, Zutphen The Netherlands