Aquaculture 285 (2008) 23–29

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aqua-online

QTL mapping designs for aquaculture

Cécile Massault [a,b,*], Henk Bovenhuis [b], Chris Haley [a], Dirk-Jan de Koning [a]

[a] Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, Roslin, United Kingdom
[b] Animal Breeding and Genetics Group, Wageningen University, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 12 November 2007
Received in revised form 27 June 2008
Accepted 29 June 2008

Keywords:
Experimental design
Fish
Statistical power
QTL mapping
Shellfish

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.

1. 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,b). 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 optimise 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.

* Corresponding author. Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, Roslin, United Kingdom. Tel.: +44 131 527 4449; fax: +44 131 440 0434.
E-mail address: Cecile.massault@roslin.ed.ac.uk (C. Massault).

0044-8486/$ – see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.aquaculture.2008.06.040

2. Materials and methods

2.1. 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).

2.2. 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:

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

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 $\chi^2(\text{NC}(x),np)$ is the χ^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.

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., 2006; Fessehaye et al., 2006). Those contributions determine the expected family structure for a species. Another complicating factor is that parentage needs to be

Table 1
Scenarios for hierarchical experimental designs

	# 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