Original article

Within-family marker-assisted selection for aquaculture species

Anna K. SONESSON*

AKVAFORSK (Institute of Aquaculture Research Ltd), P.O. Box 5010, 1432 Ås, Norway

Abstract – A within-family marker-assisted selection scheme was designed for typical aquaculture breeding schemes, where most traits are recorded on sibs of the candidates. Here, sibs of candidates were tested for the trait and genotyped to establish genetic marker effects on the trait. BLUP breeding values were calculated, including information of the markers (MAS) or not (NONMAS). These breeding values were identical for all family members in the NONMAS schemes, but differed between family members in the MAS schemes, making within-family selection possible. MAS had up to twice the total genetic gain of the corresponding NONMAS scheme. MAS was somewhat less effective when heritability increased from 0.06 to 0.12 or when the frequency of the positive allele was < 0.5. The relative efficiency of MAS was higher for schemes with more candidates, because of larger fullsib family sizes. MAS was also more efficient when male:female mating ratio changed from 1:1 to 1:5 or when the QTL explained more of the total genetic variation. Four instead of two markers linked to the QTL increased genetic gain somewhat. There was no significant difference in polygenic genetic gain between MAS and NONMAS for most schemes. The rates of inbreeding were lower for MAS than NON-MAS schemes, because fewer full-sibs were selected by MAS.

marker-assisted selection / aquaculture / breeding

1. INTRODUCTION

For aquaculture species, most traits are recorded on sibs of the candidates, because the timing of recording (fillet quality traits are measured after slaughter) or kind of recording (disease risk in case of a challenge test) make it impossible to measure these traits on the selection candidates. The heritability of these traits is often low to medium (see [9] for a review). These two features of selection schemes of aquaculture have been shown to favour within-family marker-assisted selection schemes [12,17] and they are the incentive to investigate the potential for within-family marker-assisted selection (MAS) schemes of aquaculture species in this paper.

---

* Corresponding author: Anna.Sonesson@akvaforsk.no

Kashi *et al.* [11] presented a grand-daughter design for MAS schemes, in which a sire and his progeny-tested sons are genotyped and phenotypic information is collected on the grand-progeny as daughter group means. Significantly different means of daughter-groups of sires that inherited the M or m allele of the grandsire that has genotype Mm indicates a marker that is linked to a QTL. Mackinnon and Georges [14] presented a two-generational design for MAS schemes, where progeny are phenotyped and genotyped and this information is used to calculate the effect of the marker allele on the trait. In the grand-daughter scheme, fewer genotypic records are needed than in the bottom-up scheme and at the same number of genotypes, the statistical power is 3-4 times larger for the grand-daughter scheme [30]. However, progeny tests are not performed for aquaculture species, which makes the grand-daughter design badly suited. In this paper, a two-generational design for a typical aquaculture breeding scheme will be presented.

Genetic marker maps are available for some of the aquaculture species. For most species, these genetic marker maps are made up of a large number of AFLP markers, which are anchored to fewer microsatellites. The marker density is rather low and the markers are unevenly spread over the genome on these maps. The most developed maps are the ones for rainbow trout [21] and tilapias [13].

Only few QTL for breeding goal traits have been identified in aquaculture species, mainly in salmonids, where QTL for resistance to different disease traits have been found [20, 22, 24, 26]. Also, QTL related to body weight and size have been found [15, 23, 25]. All of these QTL are candidates for selection in a MAS scheme.

Some genetic features of aquatic species have implications for QTL mapping studies. For example, recombination rates can differ between males and females, and therefore the marker map lengths between males and females differ considerably. In salmonids, females have the highest recombination rate. The ratio between female/male recombination rate was 3.25:1.00 for rainbow trout [27], 1.69:1.00 for Arctic char [31] and 8.25:1.00 for Atlantic salmon [19]. However, in other aquaculture species, males have the highest recombination rate. For example, the ratio between male and female recombination rate was 7.4:1.00 in Japanese flounder [3]. There are also differences in recombination rate over the length of the chromosomes, *i.e.* recombination rate was higher in telomeric regions than in proximal regions [27].

The aim of this study was to design a within-family MAS breeding scheme for typical aquaculture species and to investigate its superiority over traditional breeding schemes by computer simulation. In typical aquacultural breeding

schemes, many traits are recorded on sibs of the selection candidates and ge-
netic improvement of such traits is considered here. Schemes with different
mating ratio, heritability of trait, size of QTL and different numbers of QTL
affecting the trait, markers flanking the QTL, breeding candidates and sibs
tested for the trait, were evaluated.

2. MATERIALS AND METHODS

2.1. Design of aquaculture MAS breeding schemes

The MAS breeding scheme aims at genetic improvement of traits that are
measured on sibs of the breeding candidates. The sibs of the candidates will be
phenotyped and genotyped, in order to estimate the marker-trait associations
of the markers that are inherited from the parents. The parents transmit the
same markers to their progeny that are selection candidates, and these candi-
dates will be selected for the markers that they inherited. In this way, within
family differences in breeding values between the progeny were estimated and
selected for, whereas in NONMAS breeding schemes (breeding schemes that
do not use marker information) it is only possible to select between families,
which implies that less than half of the total genetic variance is used.

2.2. Genetic model

A polygenic effect of the trait, g_i, was simulated. The polygenic effect of
each individual in the base population was sampled from $N(0, V_a)$, where $V_a =$
base generation additive genetic variance of polygenic effects, which was as-
sumed to be 0.06 or 0.12. The polygenic effect of later generations was ob-
tained by simulating genotypes from $g_i = \frac{1}{2} g_s + \frac{1}{2} g_d + m_i$, where m_i is the
Mendelian sampling component, which was sampled from $(0, \frac{1}{2}(1 - F)V_a)$,
where F is the average inbreeding coefficient of parents s and d.

It was assumed that there were either one or three QTL genetically affecting
the trait, nloc = 1 or 3, in addition to the polygenic effect. These QTL had
been identified in earlier QTL mapping experiments and were unlinked (*e.g.* on
different chromosomes). Each QTL had two alleles. Marker haplotypes were
formed, where the QTL was in the middle of the haplotype. The number of
markers flanking the QTL, nmarkers, was two or four. For the basic scheme,
starting values of the variance explained by the QTL, V_{QTL} was 1/6 of V_a,
i.e. the variance due to QTL was 0.01 and 0.02 for the schemes with $V_a =$
0.06 and 0.12, respectively. The environmental variance was chosen such that

the phenotypic variance in the base generation was 1.0, *i.e.* the environmental variance was 0.94 and 0.88, respectively. Thus, the heritability, h^2, in the base population was 0.06 and 0.12, respectively.

For the base generation, two alleles for each individual were sampled, where the favourable QTL allele was sampled with probability p and the other allele was sampled with probability $(1 - p)$. For later generations, individual genotypes were sampled using Mendel's rules. The value of p was 0.2 in all schemes. Hence, when one QTL explained all variance of the QTL effect, the genotypic value, a, was deduced from $V_{QTL} = 2p(1 - p)a^2$ [6] and was a = 0.1768 for schemes with $h^2 = 0.06$ and 0.25 for schemes with $h^2 = 0.12$. When three QTL explained the variance of the QTL effect, V_{QTL} was divided over the three QTL in a ratio of 6:3:1. For schemes with h^2 of 0.06, V_{QTL} of the three QTL was 0.006, 0.003 and 0.001 giving values of 0.1369, 0.0968 and 0.0559, respectively. Schemes were also run when the QTL explained 1/3 of the genetic variation at a heritability of 0.06, *i.e.* V_{QTL} was 0.02 and a was 0.25. The genetic model was the same for the MAS and NONMAS schemes.

It was assumed that the QTL was mapped to a region with a recombination rate of 0.1 for females and 0.033 for males, but the exact position of the QTL was unknown. Two or four marker haplotypes were used to mark this region, and the two outermost markers flanked the region. If the markers did not show a recombination, it was assumed that the entire region did not recombine and thus that the QTL followed the inheritance of the marker haplotype. If the markers showed a recombination, the inheritance of the QTL allele was assumed unknown, *i.e.* each of the paternal and maternal allele was inherited with 50% probability.

Let T denote the probability that the marker haplotypes surrounding the QTL can trace the inheritance of the QTL allele from parent to offspring, *i.e.* whether the paternal or maternal QTL was transmitted. T includes both the probability that a recombination occurred in the haplotype, which makes it impossible to trace the inheritance with certainty, and that the markers were informative. Double recombination probabilities between two markers within a haplotype were assumed negligibly small. This implies that if the markers do not recombine, the QTL allele can be traced with certainty. Following [17], the actual haplotypes were not simulated, only an indicator, S_{ij}, was simulated, where $S_{ij} = 1$ denotes that the inheritance of the jth allele of the ith animal could be traced or $S_{ij} = 0$ otherwise. The recombination rate, Θ, was assumed three times as high for females (0.10) as males (0.033). The probability that markers were able to trace the inheritance of the QTL allele is derived in the appendix, and was T = 1 − 0.375 = 0.625 for females and 0.709 for males for

the schemes with two markers flanking the QTL. For the situation with four markers flanking the QTL, T was 0.819 for females and 0.898 for males.

2.3. Breeding value estimation

Estimated breeding values (EBV) were obtained using the BLUP models of [7] and [10]. For the MAS schemes, the EBV estimation model was:

$$\mathbf{y} = \mu + \mathbf{Zu} + \Sigma_i \mathbf{Q}_i \mathbf{q}_i + \mathbf{e}_i,$$

where $\mathbf{y}$ = vector of records, $\mathbf{u}$ = vector of polygenic effects, $\mathbf{Z}$ = incidence matrix linking animals to records, $\mathbf{Q}$ = incidence matrix linking animals to QTL alleles, $\mathbf{q}_i$ = vector of allelic effects for the ith QTL and $\mathbf{e}$ = vector of environmental effects. For the basic scheme, where the QTL explained 1/6 of the genetic variation, in the BLUP estimation of MAS breeding values, the variance for error, polygenes and QTL were 0.94, 0.05 and 0.01 for the scheme with $h^2 = 0.06$, and 0.88, 0.10, and 0.02 for the scheme with $h^2 = 0.12$.

For the scheme, where the QTL explained 1/3 of the genetic variation, in the BLUP estimation of MAS breeding values, the variance for error, polygenes and QTL were 0.94, 0.04 and 0.02 for the scheme with $h^2 = 0.06$.

For the NONMAS schemes, the EBV estimation model was:

$$\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e}_i.$$

In the BLUP estimation of NONMAS breeding values, the variance for error, and polygenes were 0.94 and 0.06 for the scheme with $h^2 = 0.06$, and 0.88 and 0.12 for the scheme with $h^2 = 0.12$.

Following [7], mixed model equations were set up and polygenic and QTL effects were estimated. The total breeding values for MAS and NONMAS were $\mathbf{u} + \Sigma \mathbf{q}_i$ and $\mathbf{u}$, respectively.

2.4. The simulated breeding schemes

The simulated breeding scheme was that of a closed nucleus with a discrete generation structure. The number of selection candidates, N, was 5000 or 10 000. In each generation, 100 dams were selected by truncation on total breeding value. Similarly, either 100, 50 or 20 sires were selected, implying a male:female mating ratio of 1:1, 1:2 or 1:5. Mating among the selected sires and dams was at random. Each generation, either 10 or 20 progeny per fullsib-family were tested for the trait under selection (nprog), but these progeny were

**Table I.** Summary of scenarios of the simulated schemes.

	Scenario									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
N	5000	5000	10 000	10 000	5000	5000	5000	5000	5000	5000
h^2	0.06	0.06	0.06	0.06	0.12	0.06	0.06	0.06	0.06	0.06
nprog	10	20	10	20	10	10	10	10	10	10
male:female mating ratio	1:2	1:2	1:2	1:2	1:2	1:2	1:2	1:1	1:5	1:2
nQTL	1	1	1	1	1	1	3	1	1	1
QTL size	0.01	0.01	0.01	0.01	0.01	0.02	0.006 0.003 0.001	0.01	0.01	0.01
nmarkers	2	2	2	2	2	2	2	2	2	4

not selection candidates. These progeny were made in addition to the candidates, such that the number of candidates was either 50 (with N = 5000) or 100 (with N = 10 000) per fullsib family, irrespective of nprog. The phenotypic information from the progeny that were sibs of the candidates was only used to calculate the EBV.

Schemes were run for five generations and summary statistics of each scheme are based on 100 replicated simulations. The first generation fish are randomly selected, because all information about EBV came through the parents in the sib-testing schemes and the first generation fish do not have parents. Therefore, the presentation of the results will focus on selection response in generations two and three. The breeding schemes will be compared for the total genetic level (scaled to zero in generation 1), G_{tot}, total genetic gain, ΔG_{tot}, polygenic genetic level (also scaled to zero in generation 1), G_{pol}, and frequency of the positive QTL allele for QTL 1 (2 or 3), freq1, freq2 and freq3, and rate of inbreeding, ΔF. The different scenarios of the simulated schemes are summarised in Table I.

3. RESULTS

The results from the basic scheme of scenario 1 are shown in Figure 1. The total genetic gain (ΔG_{tot}) was 0.203 for MAS and 0.176 for NONMAS in generation two, *i.e.* the increase in total genetic gain (ΔG_{tot}) was 15% higher for the MAS schemes compared with the NONMAS schemes in the first generation of selection with marker information. In generation three, ΔG_{tot} was 68% higher for MAS than NONMAS. There was no significant difference in G_{pol} between MAS and NONMAS. This increase in ΔG_{tot} is therefore mainly due to
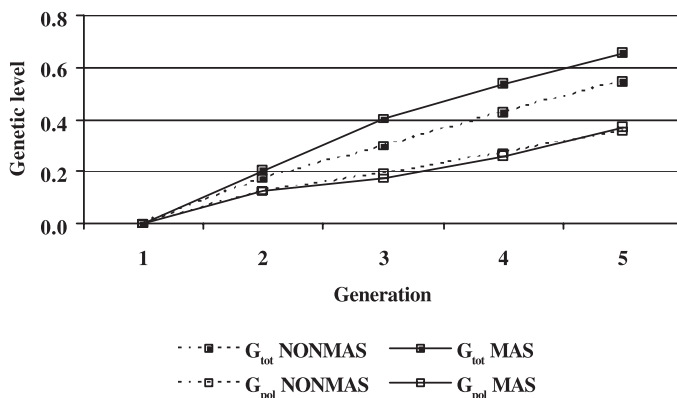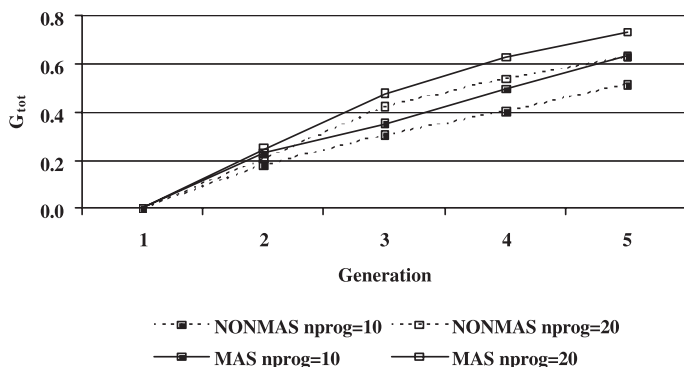
Figure 1. Polygenic (G_{pol}) and total (polygenic + QTL) genetic response (G_{tot}) for MAS and NONMAS schemes with scenario 1.

the increase in frequency of the QTL for the MAS scheme, where a higher QTL frequency also implies more genetic variance (as long as freq < 0.5). The lower value of ΔG_{tot} in generation two compared to generation three can be explained by the relatively low number of heterozygous parents ($2 \times 0.2 \times 0.8 = 0.32$) and because the marker information is still building up. The higher frequency for MAS than NONMAS and little or no difference in G_{pol} between MAS and NONMAS was seen for most of the simulated schemes and is perhaps in contrast with experimental MAS schemes, where G_{pol} was decreased due to MAS [8]. Here, however, improvement of G_{pol} is due to between family selection, which is little or not reduced by MAS schemes, where the frequency of the positive QTL allele is mainly improved by within-family selection.

3.1. Effect of size of scheme

When the number of progeny per fullsib family tested for the trait under selection, nprog, was 20 as in scenario 2, ΔG_{tot} was 11 and 67% higher for MAS than NONMAS in generations two and three (results not shown). Hence, the superiority of MAS over NONMAS was about the same for schemes with nprog = 10 (scenario 1) or 20 (scenario 2).

With 10 000 candidates, ΔG_{tot} was 11 and 86% higher for MAS than NONMAS in generations two and three with nprog = 10 (scenario 3), and 7 and 87% higher in generations two and three with nprog = 20 (scenario 4; Fig. 2). Hence, MAS was more efficient with 10 000 than 5000 candidates, which can be explained by the larger fullsib family sizes being advantageous for within-family selection.

Figure 2. Total (polygenic + QTL) genetic response (G_{tot}) for MAS and NONMAS schemes, where the number of candidates was 10 000 and nprog 10 or 20 as in scenarios 3 and 4, respectively.

Rate of inbreeding (ΔF) remained zero until generation three for all schemes. In generation three, ΔF was 0.097 for MAS and 0.120 for NONMAS for the scheme in scenario 1 with ncand = 5000 and nprog = 10 (not shown). The lower ΔF for MAS is because the marker information allows for differentiation of breeding values within fullsib families, which reduces the correlation between EBV of fullsibs and thus the co-selection of sibs, which in turn reduces ΔF. For schemes with nprog = 20, ΔF increased especially for the MAS scheme with ncand = 10 000 as in scenario 4 ($\Delta F = 0.110$ in generation three) compared with ncand = 5000 as in scenario 2 ($\Delta F = 0.087$ in generation three), which may be due to higher selection intensity and higher accuracy of selection.

3.2. Effect of heritability

When h^2 was increased to 0.12 as in scenario 5, both G_{pol} and frequency increased more than in scenario 1 (Fig. 1), as expected, and ΔG_{tot} was 7% higher for the MAS schemes compared with the NONMAS schemes in generation two and 54% higher in generation three (Fig. 3). Hence, the extra response due to MAS was somewhat lower for schemes with a medium heritability than with low heritability, as was found by other authors, *e.g.* [17].

3.3. Effect of size of QTL

When the QTL explained a larger part of the genetic variation as in scenario 6, MAS increased frequency at a higher rate, such that frequency was
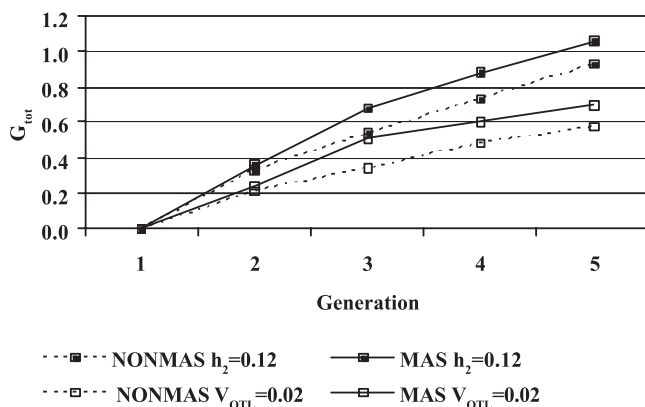
Figure 3. Total (polygenic + QTL) genetic response (G_{tot}) for MAS and NONMAS schemes, where the heritability was 0.12 (scenario 5) or the V_{QTL} was 0.02 (scenario 6).

0.97 already in generation three and the positive QTL allele thus was very near fixation (not shown). The ΔG_{tot} was 12% higher for the MAS schemes compared with the NONMAS schemes in generation two and ΔG_{tot} was twice as high for MAS compared to NONMAS in generation three (Fig. 3). Hence, MAS was, as expected, much more efficient for the scheme with a larger QTL effect.

3.4. Effect of QTL number

In Table II, the results are shown for schemes where the number of QTL loci affecting the trait, nloc, increased from one to three and V_{QTL} was 0.006, 0.003 and 0.001 for the three QTL, respectively, as in scenario 7. The total genetic gain (ΔG_{tot}) was 0.209 for MAS and 0.192 for NONMAS, *i.e.* the increase in total genetic gain (ΔG_{tot}) was 9 % higher for the MAS schemes compared with the NONMAS schemes in generation two. In generation three, ΔG_{tot} was 54% higher for MAS than NONMAS. MAS was thus somewhat less efficient for schemes with three loci instead of one. This somewhat lower genetic response can be explained by the low start frequency, which makes it difficult to select individuals with the positive QTL alleles of all three loci simultaneously with the fullsib family sizes that were used here. Also, the variance due to each QTL is smaller, which makes it more difficult to estimate the QTL effect, which reduces estimation accuracy. Note that freq3 < freq2 < freq1, which is expected, because the genetic value was the largest for the first locus and smallest for the third locus.

Table II. Polygenic (G_{pol}) and total (polygenic + QTL) genetic level (G_{tot}) and frequency of positive QTL allele (freq1–3) for schemes where three QTL affected the trait as in scenario 7.

		Generation				
		1	**2**	**3**	**4**	**5**
G_{tot}	**NONMAS**	0.000	0.192	0.315	0.424	0.540
	MAS	0.000	0.209	0.399	0.567	0.716
G_{pol}	**NONMAS**	0.000	0.136	0.218	0.287	0.364
	MAS	0.000	0.137	0.199	0.261	0.353
freq1	**NONMAS**	0.195	0.311	0.398	0.488	0.561
	MAS	0.197	0.365	0.708	0.933	0.984
freq2	**NONMAS**	0.198	0.283	0.349	0.411	0.486
	MAS	0.198	0.296	0.454	0.650	0.833
freq3	**NONMAS**	0.200	0.272	0.310	0.332	0.373
	MAS	0.199	0.262	0.296	0.351	0.417

### 3.5. Effect of mating ratio

In Figure 4, the effect of different male:female mating ratios was tested for MAS and NONMAS schemes, *i.e.* the mating ratio was 1:1, 1:2 or 1:5, as in scenarios 8, 1, and 9, respectively. Mating ratios of 1:2 or 1:5 result in an increased selection differential of the males, of which the MAS schemes can make extra good use due to the availability of within family information (*i.e.* intense within family selection is only possible in the MAS schemes). MAS had higher G_{tot} for all three mating ratios, and the relative increase was somewhat greater when the mating ratio went from 1:1 to 1:2 than from 1:2 to 1:5. ΔF was as expected the lowest for mating ratio 1:1, because the number of selected sires was the highest for that mating ratio (Fig. 5).

3.6. Effect of number of flanking markers

When the numbers of markers flanking the QTL, nmarkers, increased from two to four as in scenario 10, ΔG_{tot} was 2% higher for the MAS schemes compared with the NONMAS schemes in generation two and 11% higher in generation three (results not shown). Hence, the two extra markers did increase the informativeness of the markers, and thus the accuracy of selection and genetic gain somewhat.
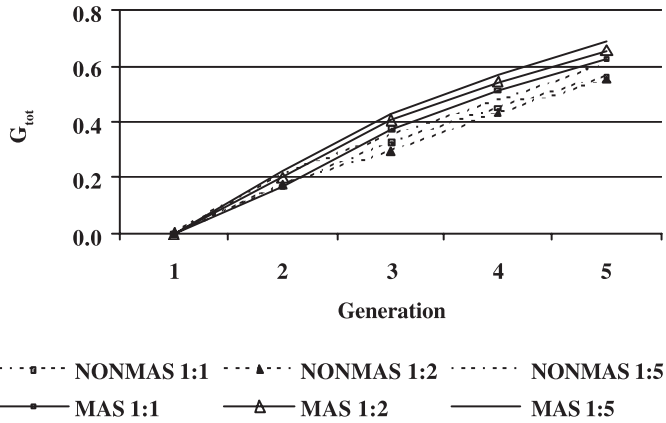
Figure 4. Total (polygenic + QTL) genetic response (G_{tot}) for MAS and NONMAS schemes where the male:female mating ratio was 1:1, 1:2, or 1:5 as in scenarios 8, 1, and 9, respectively.
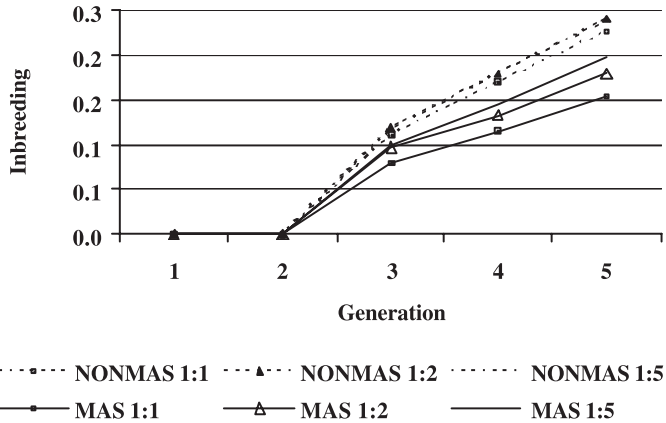
Figure 5. Inbreeding for MAS and NONMAS schemes, where the male:female mating ratio was 1:1, 1:2, or 1:5, as in scenarios 8, 1 and 9, respectively.

4. DISCUSSION

A within-family marker-assisted selection scheme for aquaculture species was presented. The design of normal family-based selection schemes, whereby sibs of the candidates are tested for the trait can be utilised also for MAS. Sibs of the candidates are then genotyped to calculate the association between phenotype and genotype. The candidates are only genotyped and breeding values are calculated using the marker information according to [7] and [10]. The need to record sibs of the candidates instead of the candidates themselves

and the large fullsib family sizes that are typical for aquaculture species make that MAS can differentiate between family members, whereas NONMAS cannot. This explains the high superiority of MAS over NONMAS. The total (QTL + polygenic) gain was very high, up to 100%, when marker information was used. The increase of genetic gain for MAS is because it increases accuracy of the breeding values and makes within-family selection possible. In the MAS schemes, marker information for the EBV from the sib-tests is not available until generation one and non-random selection therefore did not start until generation one. With a rather low starting frequency of 0.2 for the positive QTL allele, together with rather few progeny per fullsib family that were tested and the high recombination rate used (although the QTL explained 1/6 of the total genetic variance), accuracy of selection was not much higher for MAS than NONMAS in generation two. However, in generation three, a large increase in genetic gain could be seen for the MAS scheme, because of the increased accuracy that was possible when another generation of marker information became available.

The large selection intensities and the relative large QTL (explaining 1/6 of the total genetic variance) resulted in fixation of the positive allele already in generation three, after two generations of non-random selection, for many schemes. In practical schemes, selection will not only be for one trait or on one marker linked to the QTL and therefore a slower fixation of the positive allele is expected.

The rate of inbreeding was high in these schemes, because selection was based only on truncation selection on BLUP estimated breeding values. With the extra information on genetic markers within families, ΔF decreased for MAS in all schemes, because fewer family members were selected. With a total size of the schemes of 5000 and 100 families (and dams selected), full-sib family size was 50 (25 males and 25 females). Therefore, NONMAS selection could be from four families only. In practical selection schemes, one should, however, try to restrict ΔF either by selecting parents from more families or by using methods that restrict ΔF, *e.g.* optimum contribution selection for MAS [29] or for GAS [18]. It is expected that MAS would increase genetic gain relatively more compared to NONMAS when ΔF is restricted, because NONMAS will have to select from more families to reduce ΔF (and genetic gain). Selection methods with constrained rates of inbreeding are, however, not utilised by breeding schemes for aquaculture species today.

The trait that was simulated in these schemes was measured only on sibs of the candidates. For traits that could be measured on selection candidates, also NONMAS would be able to select for within-family information and therefore

the superiority of MAS is expected to be less for traits that are measured on the candidates [17].

If markers can be found that are very close to the QTL, direct selection for the markers has been suggested as linkage-disequilibrium MAS (LD-MAS; [4]). In this paper, MAS was based on analysis within families because of the following: (1) marker maps in aquaculture species are often not dense enough for LD-MAS; and (2) fullsib family sizes are very large for most aquaculture species, such that the re-estimation of marker-QTL associations for every family are not so problematic as in livestock populations, *i.e.* the use of LD-MAS is less of an advantage in aquaculture populations with large families.

Instead of simulating markers and working out whether they could be used to trace the QTL alleles, the probability that the QTL alleles could be traced by the markers was modelled directly here, *i.e.* was input for the simulation program. Computationally, this had the advantage that the linkage phases between the markers and whether the markers in a particular fullsib family were able to trace the inheritance of the QTL alleles or not were not computed. Furthermore, many situations with different information content of the markers and different recombination rates between markers and QTL can lead to similar probabilities of being able to trace the inheritance of the QTL, T. However, having a constant probability of being able to trace the inheritance of the QTL implies that the simulations did not depict the loss of information content of the markers when selecting for the QTL. This is especially a problem in later generations, when the marker allele frequencies approach 1. In the present paper, focus was, however, on early generation responses from MAS. Also, in practical MAS schemes, the markers that lose their information content may be replaced by markers that are still informative in the same region.

The breeding value estimation model does not account for changes of variance since the allele frequencies of the QTL changed in the simulations. This implies that more accurate breeding value estimation could be obtained by re-estimating QTL variance each generation. This would increase the accuracy of the MAS estimated breeding values and thus further increase the benefits of MAS.

Different recombination fractions were simulated between males and females. Assuming that one can distinguish between males and females, only males could be tested instead of a mix of males and females (at equal test capacity), because males will be more informative. A higher accuracy of the QTL effect will result. When comparing scenario 1 with a similar scheme, where only males were tested at an equal test capacity, G_{tot} was, however, very

similar (result not shown), probably because the difference in T values between males and females were quite small.

In practical schemes for dairy cattle, MAS has been applied as a pre-selection step when selecting bulls for progeny testing [1, 2]. In aquaculture breeding schemes, MAS could be implemented together with walk-back se-lection [5] as a pre-selection step. The advantage with walk-back selection schemes is that identification is done using genetic markers, which avoids the need for expensive fullsib family tanks. In the walk-back scheme by Doyle and Herbinger [5], the individual with the largest phenotypic value is first selected. Thereafter, the individual with the second largest phenotypic value becomes identified using genetic markers and selected if it is not a full- or half sib of the previously selected parents. Hence, the original walk-back selection proposed within-family selection. Sonesson [28] modified the walk-back schemes such that inbreeding was controlled using optimum contribution selection [16]. By sampling around 100 individuals, 76–92% of the genetic gain was achieved compared to genotyping all selection candidates (5000 or 10 000). One prob-lem with the walk-back selection scheme is that it is difficult to select for traits that are not measured on the candidates, *e.g.* disease traits. However, a MAS scheme similar to the one presented here could be applied, where MAS will be performed on individuals that are pre-selected (for *e.g.* growth) by some form of walk-back selection method. The marker-information could be used for both identification and within-family MAS.

REFERENCES

[1] Bennewitz J., Reinsch N., Reinhart F., Liu Z., Kalm E., Top down preselec-tion using marker assisted estimates of breeding values in dairy cattle, J. Anim. Breed. Genet. 121 (2004) 307–318.

[2] Boichard D., Fritz A., Rossignol M.N., Boscher M.Y., Malafosse A., Colleau J.J., Implementation of marker-assisted selection in French dairy cattle. Proceedings of 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France, CD communication no. 22-03.

[3] Coimbra M.R.M., Kabayashi K., Koretsugu S., Hasegawa O., Ohara E., Ozaki S., Sakamoto T., Naruse K., Okamoto N., A genetic map of the Japanese flounder *Paralichthys olivaceus*, Aquaculture 220 (2003) 203–218.

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

[5] Doyle R.W., Herbinger C.M., The use of DNA fingerprinting for high-intensity within-family selection in fish breeding, Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, August 7–12 1994, Vol. 19, Guelph, Canada, pp. 364–371.

[6] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, 4th edn., Longman Sci. and Tech., Harlow, UK, 1996.

[7] Fernando R.L., Grossman M., Marker-assisted selection using best linear biased prediction, Genet. Sel. Evol. 21 (1989) 467–477.

[8] Gibson J.P., Short term gain at the expense of long term response with selection of identified loci. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, August 7–12 1994, Vol. 21, Guelph, Canada, pp. 201–204.

[9] Gjedrem T., Olesen I., Basic statistical parameters, in: Gjedrem T. (Ed.), Selection and breeding programs in aquaculture, Springer, Doordrecht, the Netherlands, 2005, pp. 45–72.

[10] Goddard M.E., A mixed model analyses of data on multiple genetic markers, Theor. Appl. Genet. 83 (1992) 878–886.

[11] Kashi Y., Hallerman E., Soller M., Marker assisted selection of candidate bulls for progeny testing programs, Anim. Prod. 51 (1990) 63–74.

[12] Lande R., Thompson R., Efficiency of marker-assisted selection in the improvement of quantitative traits, Genetics 124 (1990) 743–756.

[13] Lee B.Y., Lee W.J., Streelman T., Carleton K.L., Howe A.E., Hulata G., Slettan A., Stern J.E., Terai Y., Kocher T.D., A second-generation genetic linkage map of tilapia *(Oreochromis* spp.*)*, Genetics 170 (2005) 237–244.

[14] Mackinnon M.J., Georges M.A.J., Marker-assisted preselection of young dairy sires prior to progeny-testing, Livest. Prod. Sci. 54 (1998) 229–250.

[15] Martyniuk C.J., Perry G.M.L., Mogahadam H.K., Ferguson M.M., Danzmann R.G., The genetic architecture of correlations among growth-related traits and male age at maturation in rainbow trout, J. Fish Biol. 63 (2003) 746–764.

[16] Meuwissen T.H.E., Maximizing the response of selection with a predefined rate of inbreeding, J. Anim. Sci. 75 (1997) 934–940.

[17] Meuwissen T.H.E., Goddard M.E., The use of marker haplotypes in animal breeding schemes, Genet. Sel. Evol. 28 (1996) 161–176.

[18] Meuwissen T.H.E., Sonesson A.K., Genotype-assisted optimum contribution selection to maximise selection response over a specified time period, Genet. Res. 84 (2004) 109–116.

[19] Moen T., Hoyheim B., Munck H., Gomez-Raya L., A linkage map of Atlantic salmon *(Salmo salar)* reveals an uncommonly large difference in recombination rate between the sexes, Anim. Genet. 35 (2004) 81–92.

[20] Moen T., Sonesson A.K., Hayes B.J., Lien S., Munck H., Meuwissen T.H.E., Mapping of a quantitative trait loci for resistance against infectious salmon anemia in Atlantic salmon (*Salmo salar*): comparing survival models with binary models, BMC (2006) Submitted.

[21] Nichols K.M., Young W.P., Danzmann R.G., Robison B.D., Rexroad C., Noakes M., Phillips R.B., Bentzen P., Spies I., Knudsen K., Allendorf F.W., Cunningham B.M., Brunelli J., Zhang H., Ristow S., Drew R., Brown K.H., Wheeler P.A., Thorgaard G.H., A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*), Anim. Genet. 34 (2003a) 102–115.

[22] Nichols K.M., Bartholomew J., Thorgaard G.H., Mapping multiple genetic loci associated with *Ceratomyxa shasta* resistance in *Oncorhynchus mykiss*, Diseas. Aq. Org. 56 (2003b) 145–154.

[23] O'Malley K.G., Sakamoto T., Danzmann R.G., Ferguson M.M., Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes, J. Hered. 94 (2003) 273–284.

[24] Ozaki A., Sakamoto T., Khoo S., Nakamura K., Coimbra M.R.M., Akutsu T., Okamoto N., Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*), Mol. Genet. Genom. 265 (2001) 23–31.

[25] Reid D.P., Szanto A., Glebe B., Danzmann R.G., Ferguson M.M., QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*), Heredity 94 (2005) 166–172.

[26] Rodriguez M.F., LaPatra S., Williams S., Famula T., May B., Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses, Aquaculture 241 (2005) 93–115.

[27] Sakamoto T., Danzmann R.G., Gharbi K., Howard P., Ozaki A., Khoo S.K., Woram R.A., Okamoto N., Ferguson M.M., Holm L.-E., Guyomard R., Hoyheim B., A microsatellite linkage map of rainbow trout *(Oncorhynchus mykiss)* characterized by large sex-specific differences in recombination rates, Genetics 155 (2000) 1331–1345.

[28] Sonesson A.K., A combination of walk-back and optimum contribution selection for fish, Genet. Sel. Evol. 37 (2005) 587–599.

[29] Villanueva B., Pong-Wong R., Woolliams J.A., Marker assisted selection with optimised contributions of the candidates to selection, Genet. Sel. Evol. 34 (2002) 679–703.

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

[31] Woram R.A., McGowan C., Stout J.A., Gharbi K., Ferguson M.M., Hoyheim B., Davidson E.A., Davidson W.S., Rexroad C., Danzmann R.G., A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregating distortion in hybrid versus pure strain mapping parents, Genome 47 (2004) 304–315.

APPENDIX

THE PROBABILITY THAT THE INHERITANCE OF THE QTL ALLELE FROM PARENT TO OFFSPRING COULD BE TRACED BY THE FLANKING MARKERS

The parent has two QTL alleles, which we will call the paternal and the maternal allele (depending on whether the parent inherited the allele from its sire or its dam). The question here is the following: what is the probability (T) that the markers indicate whether the offspring inherited the paternal or maternal allele of the parent. It is easier to calculate the probability that the markers cannot trace the inheritance of the QTL alleles of the parents $(1 - T)$. Let us assume there are n markers in a region of Θ Morgan, which is the region to which the QTL was mapped. The probabilities of having 0, 1, or ≥ 2 informative markers are $(1 - H)^n$, $n^*H^*(1 - H)^{n-1}$, and $1 - (1 - H)^n - n^*H^*(1 - H)^{n-1}$, respectively, where H is the probability that a marker is informative (*i.e.* the inheritance of the marker alleles of the parent can be traced from parent to offspring). If there are 0 informative markers, the QTL cannot be traced. If there are ≥ 2 informative markers and they do not recombine, we assume that the QTL is bracketed between those and follows the inheritance of the markers (*i.e.* the probability of a double recombination is assumed negligibly small). Thus, if there are ≥ 2 informative markers, the probability that the QTL cannot be traced is Θ, *i.e.* the probability that there is a recombination between the markers. If there is only one informative marker, the situation is different, because we cannot trace the QTL with near certainty. For instance, if the paternal allele was inherited at the marker locus, the probability of a paternal inheritance at the QTL locus is $(1 - \Theta)$, which is in between the extreme situations of 0.5 (no information) and 1 (full information). And, the probability of non-paternal inheritance is Θ, which is in between 0.5 (no information) and 0 (full information). If we interpolate the case of one informative marker between "no" and "full" information, we have "no information" with probability $\Theta/2 = 2\Theta$ (where $^1/_2$ is the difference between no (0.5) or full information (0)) and "full information" with probability $(1-2\Theta)$. The probability that the markers cannot trace the inheritance of the QTL alleles of the parents is (summation over 0, 1, ≥ 2 informative markers):

$$(1 - T) = (1 - H)^n$$
$$+ n^*H^*(1 - H)^{n-1}*2\Theta$$
$$+ [1 - (1 - H)^n - n^*H^*(1 - H)^{n-1}]^*\theta.$$

One minus this equation gives the value for T. The probability that the markers are informative is assumed H = 0.5.