THE USE OF DNA FINGERPRINTING FOR HIGH-INTENSITY, WITHIN-FAMILY SELECTION IN FISH BREEDING

R.W. Doyle and C. Herbinger
Marine Gene Probe Laboratory, Dalhousie University
Halifax, NS. Canada B3H 4J1

SUMMARY

The Marine Gene Probe Laboratory (MGPL) is developing and testing selection protocols that use DNA fingerprinting to identify sibships in aquaculture populations. Superior individuals are selected from families that are grown together from birth onwards with a minimum of physical tagging and without interfering with commercial operations. A simple within-family selection protocol (called "walk-back") exploits the high fecundity of aquaculture organisms to achieve intense selection while minimizing inbreeding.

The relative response to selection of within-family *versus* combined selection is derived as a function of relative selection intensities and the intraclass correlation coefficient, t. For full sibs, the relative response to within *vs.* combined selection is $S_{w/c}t^{1/2}$, where $S_{w/c}$ is the ratio of selection intensities. Within-family is expected to be more effective than combined selection in many aquacultural circumstances. Extensive computer simulations suggest that the number of animals required to be DNA fingerprinted is a few hundred or less, to achieve selection intensities of 2 to 4 sigma, with an effective breeding number > 80. The current state of development of DNA microsatellite fingerprinting in the MGPL is briefly described.

INTRODUCTION

The greatest practical problem facing aquaculture genetics is that newborn aquatic animals are too small to tag by physical methods. In current breeding programmes, families must be reared for long periods in separate cages, tanks or ponds until they can be physically tagged at the family or individual level. The multiplicity of family tanks or cages adds a level of environmental variation to the experimental design, so statistically adequate experiments tend to be large and expensive.

The second difficulty (or more accurately, forgone opportunity) faced by current selective breeding procedures stems from the relatively small proportion of offspring that can be tagged -- when they have finally grown large enough. The limitations of physical tagging reduces selection intensities far below the levels theoretically permitted by the fecundity of many aquatic species, which may be in the hundreds for tilapia, tens of thousands for salmonids and millions in the case of carp. A large proportion of the potentially selectable population is simply discarded (from a geneticist's perspective, even if the fish are not actually thrown away). In this respect fish differ crucially from other domestic animals such as sheep or poultry, where low fecundity rather than our tagging capability ultimately limits the selection intensities that can be achieved. Fish have a closer

affinity to plants than to domestic animals in this regard. Unfortunately, whereas plant seeds will stay put when planted in numbered locations, fish seed will not.

To circumvent these problems, the Marine Gene Probe Laboratory (MGPL) is developing selection strategies that utilize DNA fingerprinting to simplify tagging and increase the intensity of selection. The essence of this strategy is to use individual- and family-specific DNA fingerprint probes for identifying broodstock animals and their offspring.

The DNA probes are not themselves used as criteria for selection (e.g. as QTL markers). Their role is to provide pedigree information. The actual on-farm breeding design follows the principles of classical quantitative genetics, with appropriate modifications to take advantage of the extraordinarily high individual fecundities of most fish and shellfish.

OPTIONS FOR SELECTIVE BREEDING USING DNA FINGERPRINT PEDIGREES

At any particular selection intensity, genetic progress is maximized by a combination of within-family and between-family selection (Falconer 1981). Ideally, the choice of breeders is determined from information on the performance of each individual fish and its relatives, possibly in several environments. Modern computer technology and statistical procedures allow good estimates of individual breeding values even with highly unbalanced data (e.g. Siitonen & Gall 1989). There appear to be a range of options for experimental designs using DNA fingerprint pedigrees in aquaculture. Of these, two basic types are considered here:

Full-information (combined) selection. Given sufficient performance information, optimal combined, mixed or pedigree selection can be carried out on a random sample of animals grown together until they have been identified by DNA fingerprinting. Animal breeding procedures utilizing the highest available level of sophistication can then be used for parameter estimation and selection of broodstock from this group of fish. Essentially, physical tagging is replaced by DNA tagging. This procedure does eliminate the costly separate housing of families but requires a large amount of genetic fingerprinting. There is a tradeoff between the theoretically optimal use of the animals that are fingerprinted, and loss of selection intensity through discarding most of the animals in the population before they can be measured or identified.

Alternatively, combined or pedigree selection can be carried out on previously mass-selected animals, after fingerprinting only the large ones in the right-hand tail of the size distribution (supposing size is the criterion for selection). High selection intensities are obtainable because only the largest animals produced by the population are used. This procedure may prove to be optimal but appears to present formidable estimation problems.

Within-family selection. The approach discussed in the remainder of this paper is within-family selection. Selection is all done within families in this simple protocol, which maximizes the effective breeding number and achieves

very strong selection because of the characteristically high fecundities. The amount of DNA fingerprinting is minimized by identifying animals at the far right of the size distribution and working towards the mean only as far as is required to minimize inbreeding.

The simplest version of this procedure merely requires one to know whether any two selected offspring have the same parents, without necessarily knowing which parents. The parental generation need not be DNA fingerprinted to determine this. The probable sib relationships of selected individuals can be calculated by maximum likelihood, given a posteriori knowledge of allele frequencies at each locus in the population (Herbinger and Doyle, in prep).

THE RELATIVE RESPONSE TO COMBINED AND WITHIN-FAMILY SELECTION

Falconer's classic text discusses the relative response to the basic selection procedures when selection intensities are the same for each method (e.g. Fig. 13.1, page 216 of the 1981 edition). When selection intensities are not equal Falconer's analysis needs to be extended slightly. The response to within-family selection relative to combined selection, $R_{within}/R_{combined}$ or $R_{w/c}$, can be shown to depend on the relative selection intensities, $S_{w/c}$. The relationship is

$$S_{w/c}\ \{(1-r)\ [(n-1)/n(1-t)]^{1/2}\}\ /\ \{1+\ (n-1)\ (r-t)^2/(1-t)\ (1+t(n-1))\}^{1/2},$$

where r = genetic correlation, t = intraclass correlation, n = family size and $S_{w/c}$ is the ratio of standardized selection intensities $i_{within}/i_{combined}$. Note that the relative response is independent of the heritability and phenotypic variance of the trait.

Within-family selection would normally be carried out on full-sib groups, in which case r=1/2 and the relative response of within and combined selection becomes

$$S_{w/c}\ [(n-1)/n(1-t)]^{1/2}\ /\ 2\ [1\ +\ (n-1)(\tfrac{1}{2}-t)^2/(1-t)(1+(n-1)t]^{1/2}$$

These equations can be simplified when the family size, n becomes large. The limit at infinite family size is

$$\underset{n\infty}{\text{Lim}}\ [R_{w/c}]\ =\ S_{w/c}\ [1/(1-t)]^{1/2}\ /\ 2\ [1/(4t-4t^2)]^{1/2},$$

Which reduces to $S_{w/c}t^{1/2}$.

This simple expression describes the relative response to selection for full sibs and large family sizes when selection intensities are unequal.

The relative response $R_{w/c}$ is plotted in Fig. 1 for five values of $S_{w/c}$. Since we are interested in the relative response of within-family to combined selection when the latter uses physical tagging methods, $S_{w/c} = i_{within}/i_{combined}$ is assumed to be greater than 1. (Very large families permit strong selection.)

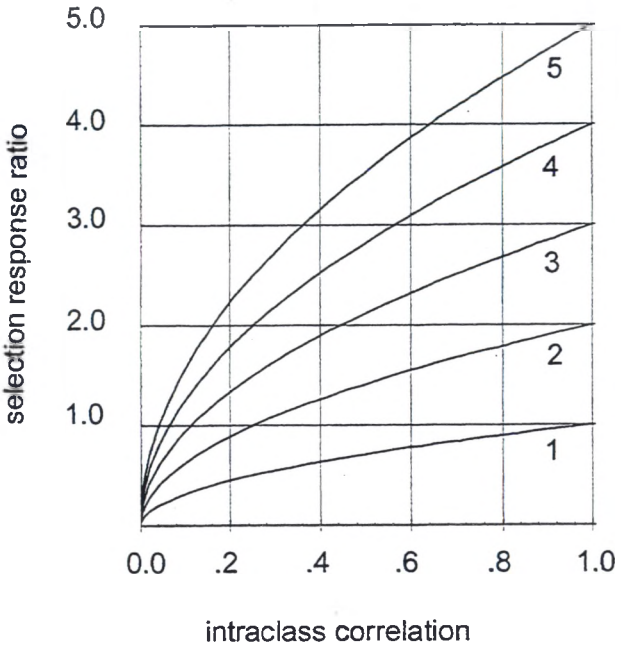

Figure 1. Relative response to within *vs.* combined selection, $R_{w/c}$.

Five relative selection intensities $S_{w/c}$ are indicated by numbers on the right of the graph.

The x-axis is t, the full-sib intraclass correlation.

The lower line in Fig. 1 corresponds to the relationship shown in Falconer's Fig. 13.1 for equal selection intensities. It is likely that within-family selection can be 2 or 3 times more intense than is achieved in the best current combined selection programmes based on physical tagging (e.g. the AKVAFORSK programme in Norway, Gjedrem et al. 1988).

The relative response to selection depends on the value of the intraclass correlation, which tends to be rather large in full sib groups reared together (Gall et al. 1993). When the intraclass correlation exceeds 0.1, approximately, within-family selection yields the larger response.

Fig. 2 is a computer simulation that illustrates within-family selection in which the whole population has been graded. The largest animal in the population is DNA fingerprinted and chosen to become one of the parents of the next generation. Then the second-largest animal is examined and also added to the broodstock if it comes from a different family than the first. Otherwise it is discarded. The third-largest individual is fingerprinted and accepted as broodstock if it is not a sibling of the first two. The process is repeated until a sufficient number of pairs of selected breeders, each from a different parental family, is obtained. A rotational mating scheme is assumed (Abella et al. 1992). For convenience we call this the *walk-back* selection procedure.

The walk-back may become very expensive if the number of animals that has to be identified is large. (The current fingerprinting cost in the MGPL is between $10 and $100 depending on technical requirements.) Obviously, the

number to be identified will depend critically on the intraclass correlation coefficient, since if most of the animals in the right-hand tail of the distribution belong to one family, many will be discarded before another family is encountered. Another potential source of difficulty is unequal family size, since if one family is much more numerous than the others many individuals may have to be fingerprinted. Only practical experience will tell whether these are serious problems, but computer simulation experiments suggest they probably are not.

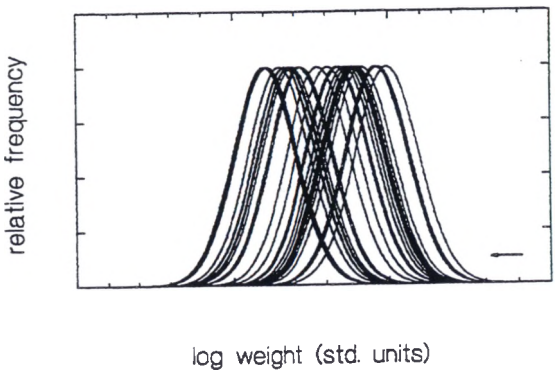

Figure 2. Simulated selection. Each curve is a different family. t = 0.58.

In "walk-back" selection The largest animals are fingerprinted sequentially from the right.

Relative frequencies on the Y-axis. Gaussian distributions simulating length or log weight.

The simulation in Fig. 2 is based on an assumed 75 pairs of spawned animals, of which 25 are shown in Fig. 2. The mean family weights and family sizes both vary randomly (even random distributions). The intraclass correlation is .58, rather high for fish spawned simultaneously and grown together in the same ponds and cages. The expected number of sibs to be encountered in each family during the walk-back was calculated by integrating the familial Gaussian distribution functions. In this particular simulation, walk-back requires the fingerprinting of only 141 animals before at least 2 sibs are identified from 23 out of the 75 families spawned.
This type of simulation has been carried out many times over a broad range of t values and family sizes. For the most part the number of identifications required for an effective breeding number of 80 or greater falls in the low hundreds. The within-family selection intensities, which depend on family size, range between 1% and .02% (i = 2.3 to 3.5).

The sizes of the sibships selected in such simulations are illustrated in Fig. 3 for the particular simulation shown in Fig. 2. The range of family sizes (number of sibs produced) is indicated on the X-axis. It can be seen that individuals from surprisingly small families (about 1/10 of the maximum) can be encountered during the walk-back.

The Y-axis indicates the length of the walk-back, in standard deviation units, from the largest fish to the 141st largest fish. The area of the symbols ("bubbles") is proportional to the number of each family identified during the walk-back. The smallest animal fingerprinted was about 1.4 std. deviations smaller than the largest. One abundant family that also had one of the largest mean (log) weights dominates the sample.

If mass selection were to be performed on this simulated population, inbreeding would be intense. Inbreeding may accumulate rapidly in small or artisanal operations where high fecundities permit farmers to use a few often related individuals as breeders and pedigree records are nonexistent (Doyle & Talbot, 1986). The MGPL has already identified (and, we hope, forestalled) potentially serious inbreeding problems in scallop and salmonid hatcheries through the use of DNA sibship identifications.

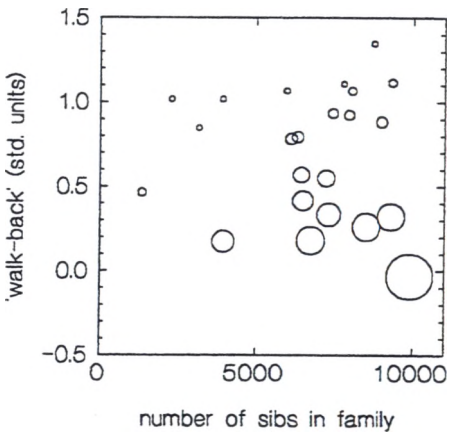

Figure 3. Properties of 23 families encountered during simulated "walk-back" selection in Fig. 2. (See text for explanation.)

THE BIOTECHNOLOGY OF DNA FINGERPRINTING AT THE MGPL

Protein polymorphisms (allozymes) have been used in the past to mark or identify genotypes in aquaculture experiments (Moav et al. 1976). The level of allozyme polymorphism available for marking is relatively low, however, so that even when special breeding arrangements are made, no more than 2 or 3 different genotypes (families) can be distinguished in pooled populations. This number is too small for selection or inbreeding avoidance, although it is useful for other purposes, e.g. for breed comparison. The extremely high levels of polymorphism at the loci used in DNA fingerprinting means that the sibship of any individual can be unambiguously identified even when the genotypes of the potential parents are unknown.

The laboratory practice and underlying molecular biology of DNA fingerprinting in fish has recently been comprehensively reviewed by Franck et al. (1991) and Wright (1993) and will not be presented in detail here. The technology has evolved rapidly since its first use in the MGPL (Harris et al. 1992). The current preference in the MGPL is to base its DNA fingerprinting on polymorphic DNA regions called microsatellites, which consist of tandemly repeated sequences of from 1 to 4 base pair repeats. These tandem arrays of core sequence repeats, which may exist at several locations in the same genome (i.e. at several loci), vary in length among individuals. Since the number of length variants (alleles) segregating at each locus may be very large, this type of polymorphism has been termed "hypervariable" variable number tandem repeat, or VNTR, polymorphism. The detection procedure involves the synthesis of pairs of oligonucleotides that are complementary to unique sequence DNA lying adjacent to each end of the polymorphic VNTR locus. The intervening VNTR sequence is amplified by PCR and electrophoresed on acrylamide sequencing gels (Jeffreys et al., 1988). High-resolution sequencing gels are required to avoid confounding alleles which have nearly the same length.

WITHIN-FAMILY SELECTION IN PRACTICE

Within-family selection has been shown to be a simple and effective selection procedure for tilapia, which is the only aquacultural species in which it has been tried thus far. Uraiwan (1990) reported gains of slightly less that 20% after three generations of within family selection in *Oreochromis niloticus* in Thailand. In the Philippines, Abella et al. (1990) reported similar gains after 3 generations of within family selection in a different strain of the same species. The Philippine tilapia project continues, and selection gains now approach 70% after 9 generations (Bolivar, pers. comm.) These selection programmes were done by conventional means, i.e. with each family maintained in separate tanks, rather than by DNA pedigree procedures. They required a relatively small commitment of resources but even so, special facilities were required, so selection was performed in experimental rather than in commercial environments.

The MGPL has pilot-scale projects employing DNA fingerprint pedigree and selection procedures under way on Nova Scotian commercial farms and governmental agencies that produce scallops (*Placopecten magellanicus*), Atlantic salmon and rainbow trout. The participating hatcheries do not do the DNA fingerprint analyses themselves, of course. They can, however, easily identify high-performing individuals after they have grown to market or breeding size. The DNA fingerprint technology is being implemented through a type of genetic counselling programme in which the MGPL advises the hatchery on which individuals to breed during routine commercial spawning and growout operations. The aim is to achieve cost-effective genetic progress in small programmes. This will potentially encourage the simultaneous improvement of several broodstocks in a single geographical region, including breeds that are optimized for contrasting and marginal environments.

REFERENCES

ABELLA, T.A., PALADA, M.S. and G.F. NEWKIRK. (1990) In Proc. Second Asian Fisheries Forum. Asian Fisheries Soc. Manila. eds. Hirano, R. and Hanyu, I. pp. 515-518.

DOYLE, R.W. and A.J. TALBOT (1986) Can. J. Fish. Aquat. Sci. 43:1059-1064.

FALCONER, D. S. (1981). Introduction to Quantitative Genetics, 2nd Ed. Longman, London. 340 p.

FRANCK, J.P.C., HARRIS, A.P., BENTZEN, E.M., DENOVEN-WRIGHT, E.M. and WRIGHT, J. (1991). Organization and evolution of satellite, minisatellite and microsatellite DNAs in teleost fishes. Oxford Surveys on Euaryotic Genes. Oxford University Press.

GALL, G.A.E., BAKAR, Y. and FAMULA, T. (1993) Aquaculture 111: 75-88.

GJEDREM, T., B. GJERDE, and REFSTIE, T. (1988) In Proc. 2nd Internat. Conf. on Quantitative Genetics. Eds. B. S. Weir, E. J. Eisen, M., Goodman and Namkoong, G. Sinauer, Sunderland, MA., 527-535.

HARRIS, A. S., BIEGER, S., DOYLE, R. W. and WRIGHT, J.M. (1992) Aquaculture 92:157-163.

JEFFREYS, A.J., WILSON, V., NEUMANN, R. and KEYTE J. (1988) Nucleic Acids Res. 16: 10953-10971.

MOAV, R., BRODY, T., WOHLFARTH, D., and HULATA, G. (1976) Aquaculture 9: 217-228.

SIITONEN, L. and GALL, G.A.E. (1989) Aquaculture 78:153-161.

URAIWAN, S. (1990) Artificial selection on growth and age of maturation of tilapia (*Oreochromis niloticus* Linn.) in Thailand. Ph.D. thesis, Dalhousie University. 259 pp.

WRIGHT, J., (1993) In Biochemistry and Molecular Biology of Fishes, Vol. 2. Eds. Hochachka, P. and Mommsen, T. Elsevier.