New Method Of Trait Recording In Aquaculture Breeding Based On The Use Of Marker Technology

H.A.M van der Steen[*], J.W.M. Bastiaansen[&] and B.P. Kinghorn[$]

Introduction

In aquaculture breeding, it is often useful to collect information from two or more distinct subpopulations but to test them in a mixed population in order to avoid confounding with environmental effects. Each of the subpopulations has an at least partially specific genetic profile, and individuals from the subpopulations commingle and co-develop in the mixed population. It is difficult to conduct individual identification and trait recording when a large number of individuals are involved, when the individuals are small (before tagging weight) and identification of specific individuals is technically difficult. Examples are product evaluation and evaluation of families and inbred lines from hatching onwards. These studies may also be conducted to evaluate the subpopulation's performances under different environmental conditions or experimental treatments. The QTMP (Quantitative Tracing in Mixed Populations) marker technology provides a method and a system that avoids the need to conduct individual identification and individual trait recording in the mixed population. A wide range of applications includes evaluation in plant and aquaculture breeding, breeding by design and the optimization of microbial processes (Van der Steen, 2008). The method and markers were tested using DNA pools of two contributing individuals.

Material and methods

QTMP method. By measuring the allele frequencies for a set of genetic markers in a mixed population (MP) and combining that with knowledge of the genetic profiles of the contributing populations, various characteristics of the mixed population can be determined. MP is a mixed population which consists of individuals originated from *p* subpopulations with known, unknown or estimated contributions from each of the subpopulations. The genetic profile of the subpopulations is based on *m* markers, with $m \geq p\text{-}1$. The allele frequency for each of the genetic markers can be determined in the MP and in each of the contributing populations. The following set of equations for *p* populations using *m* markers can be defined and solved:

$$F^1_{MP} = C_1 \cdot F^1_1 + C_2 \cdot F^1_2 + \ldots + C_p \cdot F^1_p$$
$$F^2_{MP} = C_1 \cdot F^2_1 + C_2 \cdot F^2_2 + \ldots + C_p \cdot F^2_p$$
$$\vdots$$
$$F^m_{MP} = C_1 \cdot F^m_1 + C_2 \cdot F^m_2 + \ldots + C_p \cdot F^m_p$$

---

[*]Stonebridge Breeding, Gate House, Abbotswood, Evesham, WR11 4NS, UK. [&]Bastiaansen Genetics, Lange Weide 33, 5397 AG, Lith, The Netherlands. [$]University of New England, Armidale, NSW 2351, Australia

where, F^i_{MP} is the observed allele frequency for the i^{th} marker in the MP, with $i=1,m$
F^i_j, is the allele frequency for i^{th} marker in the j^{th} subpopulation, with $i=1,m$ and $j=1,p$
C_j is the fraction of individuals (or contribution), from the j^{th} subpopulation, with $j=1,p$
The set of equations with p unknowns can be solved with mixed model statistical technology to obtain the values of c_1 to c_p. We define M as a m x p matrix of marker allele frequencies for the m markers and p subpopulations. Y is the m x 1 vector of observed marker allele frequencies and $\hat{c}$ is a vector with the solution for the subpopulation contributions. The model is $y = Mc + e$, with solution $\hat{c} = (M'M)^{-1}M'y$. The fact that the values for $\hat{c}$ should add to 1 can be incorporated by using the Reduced Least Squares Model (Kinghorn *et al*, 2009). Estimates of c might include negative estimates of contributions. This can be solved by putting the largest negative estimate at zero and removing the subpopulation involved from the set of equations before re-estimating the contributions. An other approach is to use differential evolution to find values of $\hat{c}$ that minimize the sum of squares for error, under the constraint that no element of $\hat{c}$ should be less than zero, and that the elements of $\hat{c}$ should add to one (Kinghorn *et al*, 2009).

At the end of the experimental period or test, the MP can be split into bands based on phenotypic characteristics. For a quantitative trait such as shrimp body weight, the individuals may be approximately divided into two or more categories, ranking from low to high. Application of QTMP for each of the trait classes will result in estimates for the contributions of the subpopulations to each of the classes. With this information the means and standard deviations per subpopulation can be calculated for the trait (Kinghorn *et al*, 2009). The accuracy of these estimates will depend amongst others on how accurately the individuals are allocated to the trait classes and on the number of classes used. For qualitative traits the trait classes are obvious. This might for instance be the classes 'normal' versus 'defect'.

Material. The DNA of 49 pairs of individuals were pooled into 49 pools with 50/50 contributions of two individuals to one pool. The concentration of the 98 contributing DNA samples was measured with the NanoDrop method. The 49 pooled samples, the 98 individual samples and an additional set of 255 individual samples were genotyped in duplo by ServiceXS for 48 SNP with the Fluidigm BioMark™ system.

Statistical analyses. Although the Fluidigm system results in genotype calls, the signal intensity X and Y results are also available for analysis. The results of the 353 individual samples and the 98 pools were used to produce prediction equations for individual SNPs. The models used were:

$y = Plate + b_1 X + b_2 Y + b_3 XY + e$ (Model 1)

$y = Plate + b_1 X + b_2 Y + b_3 XY + b_4 R + b_5 X^2 + b_6 Y^2 + b_7 R^2 + e$ (Model 2)

where, y is the frequency of allele-1 (Fluidigm genotype call), b_i's are regression coefficients, X and Y are the signal intensities and R = X/(X+Y). Model 2 results were used to estimate the frequency of allele-1 in the pooled samples. The estimated allele frequencies were used to estimate the contributions from the two individuals to the pool.

Results and discussion

The results of models 1 and 2 were compared with the base model of using the adjustment factor *k* to account for the deviation of the heterozygotes from the midpoint between the 2 homozygotes. The Prediction Error Variation (PEV), using model-1 was 0.073. Figure 1 gives the PEV of the 48 SNPs. Performance of one of the markers was particularly poor and this marker was removed from the results in Table 1. Model-1 performed better than the base model with PEV reducing from 0.119 to 0.070. Model-2 resulted in a further small reduction to 0.061.

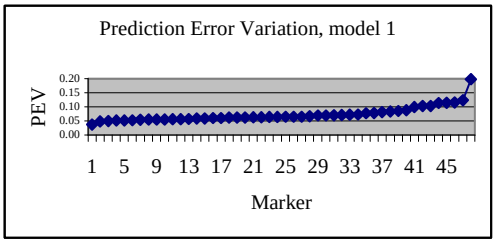

Figure 1: The prediction error variation for each of the 48 markers, predicting the frequency of allele-1

Table 1. Model statistics for the prediction of allele frequencies from signal intensities

Model	Correlation[a]	PEV[b]	Regression[c]
Base	0.893	0.119	0.900
Model-1	0.970	0.070	0.952
Model-2	0.981	0.061	0.963

[a]Correlation between predicted and true frequency

[b]Prediction error variation

[c]Regression of predicted on true frequency

The QTMP method was used to estimate the contributions of the two animals that made up the pool based on the allele frequencies of the 48 markers in the 49 DNA pools. The use of one marker already results in an estimate as we only deal with two contributing individuals.

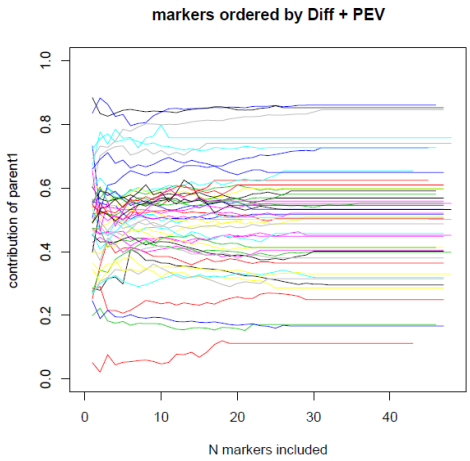

Figure 2: Estimates of the contribution of individual 1 using one marker (the best one) up to all 48 markers.

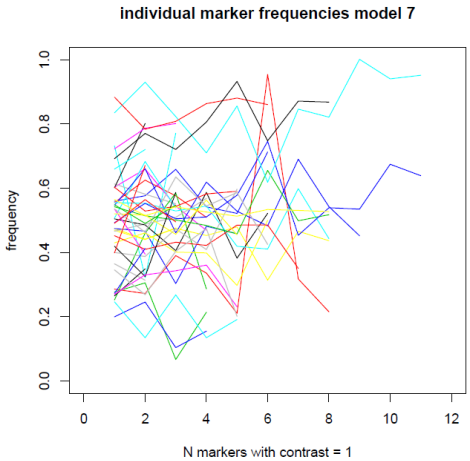

Figure 3: Est. of the contribution of ind. 1 using only one marker for the markers with maximum contrast between contributing individuals

For each pool we have two contributing individuals. A specific marker has no power to estimate contributions if both individuals have the same genotype. The power is at a maximum if the two individuals have the 11 and 22 genotype. The power is higher if the PEV of the marker is low (see Fig.1). Fig. 2 shows the contribution estimates as we increase the number of markers from one to using all 48. For each pool we start with all markers with allele frequency contrast of 1.0 and within that group starting with the marker with lowest PEV. Fig.3 shows the contribution estimates based on using only one marker, for those markers per pool with maximum contrast. We see considerable variation in estimates based on using only one marker, but as the number of markers used in the QTMP procedure increases, the estimates of individual contributions stabilize. Individual SNP results can be used to eliminate the poorest performing ones from the panel. In several pools, the contribution estimates deviate significantly from the 0.50 that could be expected. Given the consistent results across a large number of independent markers, the most likely explanation is that the NanoDrop measurement of DNA concentration was not accurate enough to guarantee 50/50 contributions. The DNA concentrations are now being re-estimates using the more accurate PicoGreen procedure. Results will be presented at the conference.

The QTMP technology can be used in several areas for several traits, including: a) mating design, b) fertilization rate, hatch% and early survival, c) performance testing, d) challenge experiments, e) estimation of competitive/ social effects and f) end product sampling. The technology allows mixing genetic groups without tagging but with collection of information on the individual genetic groups. There are no age/weight/size restrictions. The earliest opportunity to mix is at the gamete stage if the male and female gametes can be collected as is for instance the case with Salmon. The issues with performance testing have to do with tagging and rearing environment. With physical tagging, tags are inserted for individual or family identification. Tank effects can be avoided by testing all tagged individuals in one big tank. The problem is that a lot of animals need to be tagged (or small scale testing) and that testing can only start after reaching the age/weight at which the animals can be tagged. With 'traditional' DNA tagging (walk back selection) all individuals can also be tested in one big tank. The heaviest individuals are selected at the end of the test and the family they belong to is determined by genotyping for a number of markers. This system allows selection for growth but not for survival, unless we sort out the parentage for all surviving individuals. If families are tested in individual family tanks, no tagging is required. The main problem then is the confounding of true family effects and the environmental tank effects. QTMP provides an alternative to physical tagging. Several families are tested in one tank without tagging. For disease challenge tests, the same approach can be used as for performance testing.

Conclusion

The quality and number of markers can be managed in order to increase the power of the QTMP technology.

References

Kinghorn, B.P., Bastiaansen, J.W.M., Ciobanu, D.C. and Van der Steen, H.A.M. (2010). ACTA, submitted.
Van der Steen, H.A.M. (2008). *International patent application,* No. PCT/IB2008/002317.