*Proc. Assoc. Advmt. Anim. Breed. Genet. 22:537-540*

# DEVELOPING A GENOMIC SELECTION BREEDING PROGRAM FOR COMPLEX PEARL COLOUR TRAITS WITHIN THE SILVER-LIPPED PEARL OYSTER

**D.B. Jones[1], P.B. Toole[1, 2], M.S. Khatkar[3,4], H.W. Raadsma[3,4], D.R. Jerry[1,3] and K.R. Zenger[1,3]**

[1] Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Queensland, Australia
[2] Atlas South Sea Pearls Ltd., Western Australia, Australia
[3] ARC Hub for Advanced Prawn Breeding, James Cook University, Townsville QLD, Australia
[4] Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, Camden, NSW, Australia

## SUMMARY

Selective breeding within the pearling industry is heavily confounded by the complexity of production traits and the ability to accurately measure these phenotypes. Pearls are produced by implanting a nucleus along with a small piece of donor tissue from a sacrificed oyster into a host oyster. Unravelling trait architecture for these complex pearl quality traits is integral if genomic selection is to be implemented into established selective breeding programs. By combining simulated and exploratory datasets, this study uses genome-wide loci to better understand the genetic architecture of pearl colour which provides insights into the optimal design and implementation for a genomic selection breeding program within the pearling industry.

## INTRODUCTION

The silver-lipped pearl oyster is a globally important aquaculture species with pearl production from this species currently the second most valuable Australian aquaculture export (GLOBEFISH 2016). However, like most aquaculture industries, pearling has yet to establish and implement the advanced selective breeding programs required for industry progression. Although traditional animal improvement methods have had some success with simple traits (i.e. animal growth), they are inefficient for the complex pearl traits (size, colour, lustre and shape), which are polygenic, hard to measure and have a low heritability (Jerry *et al*. 2012, Jones *et al*. 2014). For the Australian pearl industry (and aquaculture in general) to maintain international competitiveness, industry must engage in a paradigm shift in breeding practices and implement pioneering technologies that circumnavigate current limitations associated with sole reliance on phenotypic selection.

Successful incorporation of genomic data into traditional selective breeding programs depends on many factors including the diversity within a farm, the species genome size and structure, and the architecture of traits of interest. Herein, we describe the development of an optimal genomic selection approach required for rapid-genetic improvement in pearl colour in pearl oysters. In doing so, we propose a breeding system which promises to not only improve efficiency of selection within the pearling industry, but will serve as a case study for many aquaculture species.

## MATERIALS AND METHODS

**Experimental animals, pearl seeding and phenotypic records.** To investigate the ideal design of a genomic selection breeding program for pearl colour, we utilised a dataset previously published in Jones *et al*. (2014) as a pilot dataset to explore the parameters required for robust application. Briefly, this dataset contains 2,306 individually traced commercial pearl grading phenotypes for 358 donor oysters from 6 families, as well as genotypic data for 1,146 SNPs across these individuals. Herein, we focus on the analysis of pearl colour, categorised into five sub-categories; SW.O.G: silver and white vs. gold vs. all remaining colours; G.O: gold vs. remaining colours; S.O: silver vs.

remaining colours; W.O: white vs. remaining colours; and finally SW.O: silver and white vs. remaining colours (Jones *et al*. 2014).

**Variance components, heritability and genetic parameters.** Previous estimates of variance components and heritability ($h^2$) for pearl colour have been based on pedigree information. To test the premise that genomic relationships are equally useful in calculating unbiased heritabilities (or variance explained by SNPs), we generated genomic relationship matrixes (GRMs) across the 358 oysters using GCTA v1.26.0 (Yang *et al*. 2011). Variance components and heritabilities were calculated using mixed linear models (MLM) whereby all SNPs were fit as random effects. Briefly,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{g} + \boldsymbol{\varepsilon} \text{ with } \boldsymbol{V} = \boldsymbol{A}\sigma_g^2 + \boldsymbol{I}\sigma_\varepsilon^2$$

where $\boldsymbol{g}$ in an $n$ x 1 vector of the total genetic effects of the individuals with $\boldsymbol{g} \sim N(0, \boldsymbol{A}\sigma_g^2)$, and **A** is interpreted as the GRM between individuals. An estimate of $\sigma_g^2$ can then be produced by the restricted maximum likelihood (REML) approach, relying on GRM estimates from all SNPs. Co-variables identified as significant in the previously published heritability estimates (i.e. seeding nucleus size and seeding technician) were also included within current the analysis described here. In addition to calculating GRMs, individual animal breeding values (best linear unbiased predictors; BLUPs) and SNP effects were calculated after incorporating in genomic relationship information. Genetic parameters generated using GRM were then compared to previously published pedigree derived results (Jones *et al*. 2014).

**Optimal number of markers.** To test the minimum number of SNPs necessary to produce robust estimates for GRM within the test farm data, we simulated a theoretical dataset to compare the potential benefit of including larger numbers of markers within GRM calculations. Firstly, a founder population was identified as the last generation of 1,000 historic simulated generations with 430 animals in each generation (equal to the effective population size of wild oyster populations as per Lind *et al*. 2007). Simulations were conducted for 4,200 SNPs using QMsim (Sargolzaei *et al*. 2009). From the founder population, 20 males and 20 females were used for breeding with each mating producing 50 offspring. The effective population size reflects the number of founders within the pilot dataset described above (N = 50). Simulations were run for 10 discrete generations with random selection of parents from the earlier generation. For positional information, the 4,200 SNPs were placed proportionally to the length of the 14 linkage groups of a *Pinctada maxima* linkage map published in Jones *et al*. (2013). In the last three generations of the simulations, only 2,000 SNPs remained polymorphic. The resulting dataset was utilised to run comparisons of GRMs between 1,000 and 2,000 SNPs.

**Optimal number of samples.** Power calculations for related vs unrelated individuals based on genetic parameters related to pearl quality were conducted to estimate the minimum number of samples required to accurately identify additive genetic variance. Using methods described in Visscher *et al*. (2014), we simulated the power, defined as 'the probability of detecting $h^2 > 0$ for a quantitative trait for the given type I error rate and the SNP-heritability assumed in the population'. Heritabilities of 0.05 - 0.30 were run to reflect previous estimates of $h^2$ for pearl colour (0.14 - 0.36; Jones *et al*. 2014). The type 1 error rate was set at 0.01 and the variance of the SNP-derived genetic relationships was 0.00002 (for unrelated individuals) and 0.025 (for related individuals within this study population, obtained from the genetic relatedness between individuals).

**RESULTS AND DISCUSSION**

**Pearl colour trait heritability and genetic parameters.** The average difference between heritability estimates of pearl colour using GRM instead of pedigree was minor (average $h^2$

difference of 0.02 ± SD 0.02) indicating that the substitution of the relationship matrix provides similar power and accuracy to pedigree data when separating variance components and calculating heritability (Figure 1). The calculation of animal BLUP values based on the GRM were highly correlated to those previously published with $r^2$ values ranging from 0.72 to 0.88 for the different colour categories except for S.O where the $r^2$ was 0.46. This lower $r^2$ may be due to the relatively low number of silver pearls observed within the data (N = 216).

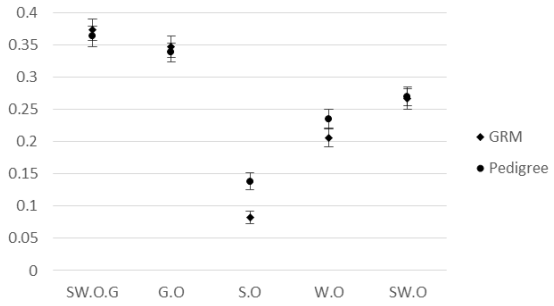


**Figure 1: Heritability estimates derived from GRM (♦) as compared to pedigree (●).**

The architecture of pearl colour has previously been reported to be polygenic and influenced by many genes of small effects (Jerry *et al*. 2012, Jones *et al*. 2014). One major region on linkage group 12 returned 13 significant genetic associations across the different categorisations of pearl colour which have SNP effect sizes ranging from 0.11 to 0.26 (Jones *et al*. 2014). For these SNPs, the SNP effects returned in GCTA using GRM were highly correlated to the previously published GWAS SNP effects (SW.O.G $r^2$ = 0.97; G.O $r^2$ = 0.96; S.O $r^2$ = 0.94; W.O $r^2$ = 0.98; SW.O $r^2$ = 0.98).

**Optimal number of markers.** To determine the potential effect of adding more SNPs into GRM calculation, we simulated a larger genotypic dataset containing 2,000 SNPs. The $r^2$ correlation between 1,000 and 2,000 SNPs was 0.98 indicating that increasing the number of SNPs with similar spacing throughout the genome yielded very little improvement to GRM accuracy in this test farm data. This indicates that 1,000 genome-wide markers is sufficient to give accurate GRM calculations for this closed farm population with limited founders (Ne previously estimated at 60). If however, if this is to be applied outside of this closed farm population, increasing the marker density would yield substantial benefit. The relative advantage of GBLUP models is at higher marker density and low heritability. In Atlantic salmon, GBLUP performed better with upwards of 4,000 SNPs (Ødegård *et al*. 2014).

**Optimal number of samples.** Simulations of the power to detect the unbiased heritable component of a trait from related individuals (variance of SNP-derived genetic relationships of 0.025) at sample sizes ranging from 100 – 400 and $h^2$ of 0.05 – 0.30 reveal that 99% power is obtained at 300 samples for a $h^2$ of 0.2 (Figure 2). For the pilot dataset with 2,000 phenotypic records (from 358 unique related individuals), power to detect the heritable component of a trait is estimated to range from 0.78 – 1.00 (for $h^2$ of 0.10 – 0.30) indicating that the current number of individuals is sufficient for estimating trait heritability. However, if this was to be expanded to unrelated individuals (i.e. variance of SNP-derived genetic relationships of 0.00002), 99% power is only reached at 8,000 samples for $h^2 > 0.2$.

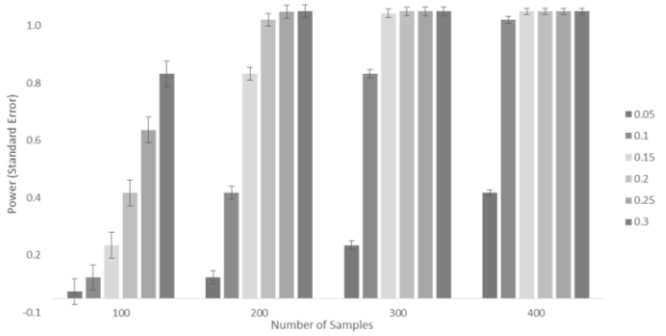


**Figure 2: Power simulations for varying heritabilities across increasing sample size.**

## CONCLUSIONS

The re-analysis of the farm test dataset within GCTA using GRM instead of pedigree returned comparable heritability, animal BLUP and SNP effects to previously published GWAS results. This suggests that the use of GRM alone is as effective as pedigree data within a closed breeding population and adds support to the notion that pearl colour is a highly complex polygenic trait. The ability to use GRM instead of pedigree to calculate BLUP enables the inclusion of individuals for which pedigree information is not known, but more importantly the relationship between relatives can be calculated more accurately (Veerkamp *et al*. 2011). Considering this, it is hypothesised that the variance components can be estimated more precisely with GRM. Furthermore, for a genomic selection breeding program to be implemented for complex traits such as pearl colour within a closed population of pearl oysters, a minimum number of 300-400 farm data records are required to estimate the variance explained by the genome-wide SNPs for the range of heritabilities evaluated. To extend this to a breeding population with a larger number of founders (N = 300), simulations indicate that the minimum number of markers required to achieve an equivalent GRM outcome is ~3,000 SNPs and that a minimum of 8,000 samples would be required to reliably detect heritable components of pearl production traits. Based on these recommendations, the current pearl oyster breeding program has collected phenotypic data (i.e. pearl quality and growth traits) and genotypic data (a minimum of ~3,500 genome-wide SNPs) from 10,000 farm production animals to achieve these outcomes. These data simulations described herein are integral to refining the direction of ongoing research into implementing advanced genomic selection into traditional breeding programs.

## REFERENCES

GLOBEFISH, F. (2016). FAO GLOBEFISH Highlights, 1/2016: 48

Jerry, D. R., R. Kvingedal, C. E. Lind, B. S. Evans, J. J. U. Taylor and A. E. Safari (2012) Aquaculture 338-341: 66-71.

Jones, D. B., D. Jerry, M. Khatkar, H. Raadsma and K. Zenger (2013) BMC Genomics 141: 810.

Jones, D. B., D. R. Jerry, M. S. Khatkar, G. Moser, H. W. Raadsma, J. J. Taylor and K. R. Zenger (2014) Aquaculture 4340: 476-485.

Lind, C. E., B. S. Evans, J. J. U. Taylor and D. R. Jerry (2007) Mol. Ecol. 1624: 5193-5203.

Ødegård, J., T. Moen, N. Santi, S. A. Korsvoll, S. Kjøglum and T. H. E. Meuwissen (2014) 5402.

Sargolzaei, M. and F. S. Schenkel (2009) Bioinformatics 255: 680-681.

Veerkamp, R. F., H. A. Mulder, R. Thompson and M. P. L. Calus (2011) 948: 4189-4197.

Visscher, P. M., G. Hemani, A. A. E. Vinkhuyzen, G.-B. Chen, S. H. Lee, N. R. Wray, M. E. Goddard and J. Yang (2014) PLoS Genet. 104: e1004269.

Yang, J., S. H. Lee, M. E. Goddard and P. M. Visscher (2011) Am. J. Hum. Genet. 881: 76-82.