Proc. Assoc. Advmt. Anim. Breed. Genet. 22:105-112

THE NEXT WAVE IN SELECTIVE BREEDING: IMPLEMENTING GENOMIC SELECTION IN AQUACULTURE

K.R. Zenger[1,2], M.S. Khatkar[2,3], D.R. Jerry[1,2] and H.W. Raadsma[2,3]

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

SUMMARY

Advanced animal breeding in aquaculture has reached a tipping point where the commercial implementation of genomic selection to improve productivity and disease resistance is becoming reality. However, the success of practical implementation of genomic selection depends on the specific aquaculture species, production system and available phenotyping and genetic resources. Using the experience learned from commercial programs for pearl oysters and marine shrimp, we highlight current benefits and options in cost-effective high-throughput genotyping and phenotyping technologies for genomic selection applications relevant to aquaculture species, followed by discussion of some of the lessons learnt when dealing with its practical implementation, including what is needed to build adequate genotype resources for non-model species; confounded breeding objective verse trait measurements; complex traits and unknown interactions; multi-family breeding schemes; multi-stage selection schemes, and transition to a genomic selection breeding program incorporating minimisation of inbreeding.

INTRODUCTION

Classical breeding programs for farmed plant and animal species are based on phenotypic selection of individuals in conjunction with knowledge on genetic relationships and quantitative genetic principles. Breeders have enhanced production traits of farmed species by selecting superior individuals as parents for succeeding generations. However, the efficiency of this method is limited when traits are difficult-to-measure, can only be measured late in life, are sex limited, or have low heritability. Over the past two decades, rapid developments in genomics have resulted in breeders incorporating genetic marker technology in the form of Marker Assisted Selection (MAS) to aid in the animal selection process. Although this technique can be useful for some simple traits, application of MAS to improve complex traits controlled by many genes of small effect is limited. Genetic improvement in these traits can only be achieved through more advanced genomic methods (Eggen 2014).

With recent advances in molecular biotechnology and quantitative analysis methods, it is now possible to accurately predict and use genome-wide molecular breeding values for improved animal selection. This approach is termed Genomic Selection (GS) and was first proposed by Meuwissen *et al.* (2001), and has gained significant application within the animal genetics community. In this approach, animal selection decisions are based on genomic breeding values (GBVs) predicted from genome-wide loci. GS is based on the theory that with sufficiently high numbers of loci across the genome, most quantitative trait loci will be in strong linkage disequilibrium with at least one marker. GS simultaneously estimates the combined genetic effects of all relevant genes and provides accurate predictions of genetic merit for a trait. Furthermore, genome wide markers can be directly use to compute the genomic relationship matrix (GRM), which can then be used to compute genomic breeding values using standard mixed model equations. GRM, even based on a smaller subset of

markers, can provide an accurate estimate of the proportion of the genome shared by related individuals and hence provides higher accuracy of estimation of breeding values as compared to estimates based on pedigree information alone (Forni *et al.* 2011).

Integration of GS methods into aquaculture breeding programs promise to rapidly increase genetic gains through improved accuracy of breeding value estimation. GS has the highest potential for traits that cannot be directly measured on the selection candidates and can be used to capture both within- and between-family genetic variances (Nielsen *et al.* 2009). This makes genomic selection a powerful approach in aquaculture, since many traits (eg., disease resistance, carcass quality and pearl quality traits) must be measured on the siblings of the actual selection candidates, rather than the selected candidates themselves. Furthermore, GS can minimise inbreeding while maximising genetic gain beyond that of current practices (Daetwyler *et al.* 2007). This is of particular benefit to aquaculture where species are often highly fecund and the number of contributing families reared in closed farms is low, resulting in rapid inbreeding if pedigree is not tracked (Gjedrem 2005). Despite all of these advantages, a limited number of aquaculture breeding companies and associated research programs are attempting to implement GS into commercial operations for long-term genetic gain (eg., Tsai *et al.* 2017; Khatkar *et al.* 2017a; Jones *et al.* 2017).

The success of the practical implementation of GS in aquaculture production systems depends on the breeding objectives, selection criteria, infrastructure, genomic resources and phenotypic recording / analysis systems. Each of these aspects can have different challenges depending on the specific aquaculture species and production system. Here we aim to provide an overview of the opportunities for the adoption of genomic selection within aquaculture, with particular focus on the challenges of implementation and long-term use in aquaculture commercial systems.

VALUE OF GENOMIC SELECTION IN AQUACULTURE

The breeding design of aquaculture species is primarily governed by the biology of the animal and available farm resources. Commercial selective breeding programs have recently expanded to a diverse range of species (eg., crustaceans such as shrimp, oysters and finfish). Primarily, most aquaculture selection programs have focused on growth, which can be selected easily based on either simple individual, or pedigree family-based selection approaches (eg., between, within and combined). For disease traits or other traits that require destructive sampling, family-based sib-selection is more commonly practised. Sib-selection, whilst allowing family average breeding values to be calculated, only exploits half of the available additive genetic variance (ie., exploits the between family variance), which limits genetic gains, and can also lead to increased inbreeding as not all families are selected to contribute to the next generation stocks.

In aquaculture, GS has been theoretically shown to simultaneously increase genetic gains, while decreasing inbreeding by up to 81% when compared with traditional selection programs (Sonesson and Meuwissen 2009). Although, the monetary value of individual animals of most aquaculture species is generally low (eg., compared to livestock), they are highly fecund and have a relatively short generation interval. This not only provides the ability for varied selective breeding strategies, but also for generating the thousands of phenotypic records required for accurate GS predictions. Furthermore, with a limited number of discrete broodstock capable of producing offspring for the entire production system, the farm effective population size is relatively small. This characteristic enables GS to be implemented on a family-based, or farm-wide basis, utilising a lower density of genome-wide loci compared to outbred populations (see genomic information section below). In aquaculture, GS improvement programs can have a rapid impact on genetic improvement particularly through the use of a structured nucleus breeding scheme. As with traditional selective breeding programs, the potential of GS will vary across different species depending on differences in life cycle, fecundity, effective population size and breeding objectives.

To date the successful application of GS in aquaculture has been limited to a handful of research

projects. For example, sea lice resistance in Atlantic salmon (Tsai *et al.* 2016), bacterial cold-water disease resistance in rainbow trout (Vallejo *et al.* 2017) pasteurellosis resistance in gilthead sea bream (Palaiokostas *et al.* 2016) and shell size in scallops (Dou *et al.* 2016). For commercial aquaculture applications of GS, there is limited public information available, and progress is reported here on optimisation and implementation within the authors own programs. Here, GS is being directly integrated into shrimp breeding programs for multiple production traits (eg., size, disease resistance, colour, survival, Khatkar *et al.* 2017a these proceedings), as well as pearl oyster breeding programs for both host oyster and donor oyster traits (eg., shell size and pearl quality traits, Jones *et al.* 2017 these proceedings). Within these programs, the feasibility of successfully applying GS has relied on the availability of high-quality genomic resources, comprehensive information on genetic parameters for all traits and extensive trait phenotype records in the reference population.

COST-EFFECTIVE GENOMIC INFORMATION

In aquaculture breeding, the number of individuals to genotype can be large (particularly for traits with low heritability). Apart from optimising the number of training or selection candidates for routine genotyping (ie., based on GS modelling and farm breeding scheme, eg., Sonesson and Meuwissen 2009), reducing the cost or number of genome-wide markers is a viable solution. Our own data show that derivation of genomic relationships can be achieved with relatively low-density SNP panels (Figure 1; 1,000-3,000 SNPs;) compared to those derived from medium-to high density SNP panels (eg., 50,000+ SNPs; see also Ødegård *et al.* 2014). However, such accuracies deteriorate rapidly if very low-density SNP panels are used (<1,000 SNPs).

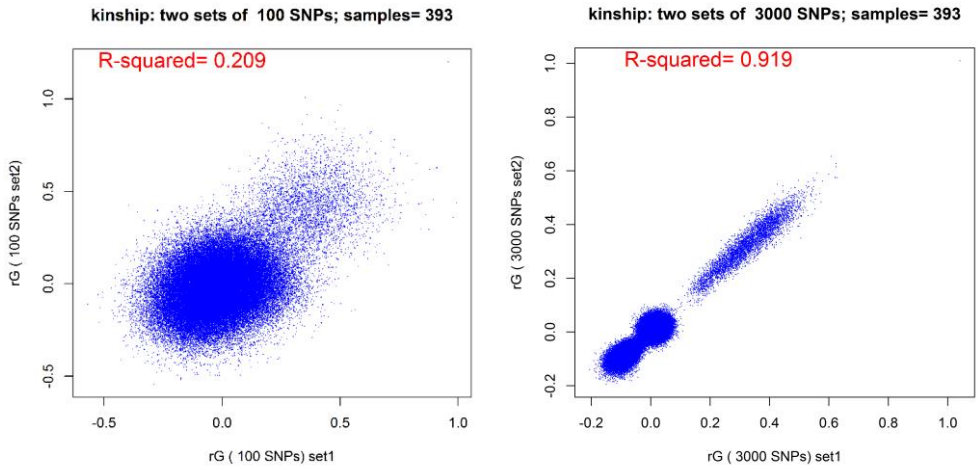

Figure 1. Comparison of SNP based kinship estimates (rG) computed using two independent sets of (a) 100 SNPs and (b) 3,000 SNPs, calculated on 393 shrimp samples.

To our knowledge, there are only a handful of aquaculture species that have commercially available SNP genotyping arrays available (ie., Affymetrix Axiom Salmon genotyping array, Affymetrix Axiom Trout genotyping array and Illumina Infinium ShrimpLD-24 genotyping array). The lack of commercially available genotyping SNP arrays for aquaculture adds significant additional cost to GS genotyping, as these resources need to be first development and tested. However, the recent development of high-throughput and cost-effective genotyping by sequencing (GBS) technologies has significantly reduced both the cost of developing and genotyping SNPs for

non-model species (eg., Lind *et al.* 2017). As such, GBS is rapidly becoming the methodology of choice for aquaculture species (Robledo *et al.* 2017). Compared to SNP array based genotyping platforms, GBS requires significantly more quality control (QC) measures to ensure robust genotype data is produced. This is primarily a result of the molecular technique itself, which can introduce spurious and missing data when proper control and data filtering methods are not put in place. Aquaculture species can be particularly sensitive to these anomalies given their sometimes highly polymorphic and repetitive genome structures, a problem particularly observed for crustaceans and oysters (eg., Yu *et al.* 2015; Lal *et al.* 2016).

Another method to reduce the cost of genotyping is through imputation of genotypes, where most of the animals can be genotyped with a low-cost, low-density SNP panel. The genotypes of these animals can be imputed up to high-density by using information on a smaller number of reference individuals (typically broodstock) genotyped with a larger high-density SNP panel that also captures the same SNP as represented on smaller arrays. Such imputed *in-silico* genotypes can then be used for GS and other genomic analyses. Such strategies have been shown to improve the accuracy of GS in livestock (Khatkar *et al.* 2012) and aquaculture species (Tsai *et al.* 2017).

The number of individuals in the reference panel and number of markers in the low-density panel depends on the effective population size of the breeding stock and relationship between reference and test populations. A small effective population size, as present in many aquaculture stocks, will require smaller number of animals in the reference panel and can be imputed with high accuracy with smaller number of SNPs in the low density panel. Moreover, if all the contributing broodstock are genotyped with the high-density panel, the accuracy of imputation in the progeny, genotyped with even smaller SNP panel, could be quite high using a pedigree based imputation approach (Hickey *et al.* 2012). However, accurate imputation requires knowledge about the precise location of SNPs across the genome. For most aquaculture species genetic linkage maps and / or genome assemblies are in the early stages of development (Abdelrahman *et al.* 2017).

NEXT-GENERATION PHENOTYPING

Accurate phenotypes on commercially important traits are critical for any breeding program. This becomes especially challenging in aquaculture where large numbers of animals need to be recorded. Any error in the trait recording will reduce effective estimated heritability and hence realised genetic gain. High-throughput and precise phenotyping strategies are required to supply the large amount of trait data required for commercial scale GS applications. Within this framework, the objective is to increase the accuracy, precision and throughput of phenotypic assessment while reducing costs and minimising labour in an intensive production system. Today, phenotyping is quickly emerging as the major operational bottleneck limiting the power and speed of commercial GS programs (eg., Cobb *et al.* 2013). This problem is compounded in aquaculture where fecundity, progeny numbers from breeding pairs and variable survival rates create circumstances where individual phenotypes and traceability are nearly impossible to obtain without new methodologies. Furthermore, aquaculture does not have the benefit of standardised global phenotyping programs such as in livestock (eg., dairy cattle). Designing effective on-farm phenotyping strategies requires integrated solutions incorporating biologists, computer scientists, statisticians and engineers.

More recently, automation, imaging and software developments have paved the way for many quantitative phenotyping studies. Within these developments, digital imaging has emerged as a cornerstone to capturing quantitative phenotypic information. Visual imaging has already allowed many production traits to be measured efficiently and accurately across different production industries including aquaculture (Cobb *et al.* 2013; Saberioon e*t al.* 2016). For example, fish length has been estimated in Rainbow trout (Miranda and Romero 2017) and fish mass in Jade Perch (Viazzi *et al.* 2015) with very low residual errors using automated computer vision techniques. Furthermore, fish skin colour and pearl quality traits (eg., colour, lustre, completion), which

traditionally are recorded as categorical traits, can now be recorded as highly-reliable continuous quantitative traits based on UV-Vis spectrophotometry (eg., Kustrin and Morton 2015), which ultimately will improve GS predictions. Other emerging aquaculture phenotyping techniques are Near Infra Red (NIR) spectroscopy and Hyperspectral imaging (HIS) which combines spectroscopy with imaging technology. These techniques are able to quantify and evaluate the chemical (eg., fat, protein, moisture) and physical (eg., freshness, texture, colour) attributes of aquatic animals with relatively high accuracies of prediction ($r > 0.8$, see Liu *et al.* 2013; Saberioon *et al.* 2016). All of these machine vision systems (MVS) are able to extract and analyse quantitative information from digital images and have the ability to improve the accuracy of the phenotype by electronically analysing the data at a pixel level across spectral regions not always visible to the human eye.

MVS usually consists of two components, the image acquisition system hardware (ie., UV-Vis, NIR and HIS) and data extraction system software. The latter typically incorporates computer based processing and optimised statistical methods and algorithms specific for the trait of interest, which is often the limiting factor in applying MVS. The development of advanced image analysis software including artificial neural network (ANN) algorithms based on machine learning approaches has been an important step forward in the development of analysis systems for automated MVS phenotyping (eg., Grys *et al.* 2016).

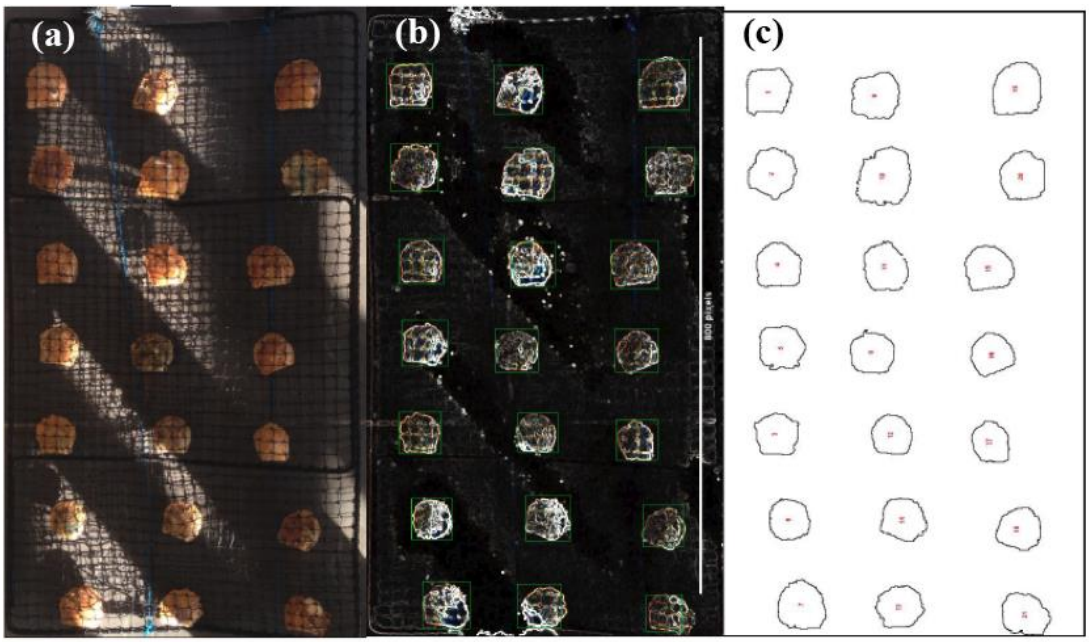

Figure 2. (a) Oyster net image depicts one of the most difficult tested situations. (b) Oysters and net have low contrast from the background and lighting is variable. (c) Sliding windows CNNs correctly identified and measured oysters with >93% accuracy.

Within our own research programs (ie., for marine shrimp and pearl oyster), machine learning algorithms have allowed precise inexpensive phenotyping across diverse production traits. For example, MVS systems have been used for pearl oyster growth data as well as pearl quality traits (eg., colour, size, lustre, completion). Although still in development, sliding window algorithms and Convolutional Neural Network (CNN) with rule-association based clustering yielded high accuracy

(exceeding 93%) in Object Character Recognition (OCR) for the oysters in nets within the full spectrum of commercial situations (Figure 2; P. Toole unpub. data). By definition, CNN learning algorithms get more precise when presented with more data. This supervised learning approach has been undertaken with developing methodologies on how to automate the entry of commercial data into a noSQL or graph-based database.

IMPLEMENTING GENOMIC SELECTION ON FARM – LESSONS LEARNT

Greatest immediate value from genomic selection is realised where genomic breeding values can be targeted against traits that drive economic returns to commercial farmers. Typically such traits are based on yields of harvested product. Although this sounds straightforward enough, practical limitations become immediately apparent in situations where traits under commercial grow-out conditions vary substantially from performance recording environments in often pathogen-free central nucleus breeding facilities (as used in specific pathogen free shrimp breeding programs for instance). For most aquaculture systems the Genotype by Environment (GxE) interactions are largely unknown and limit the value of GS training data if the genetic correlation between the central nucleus breeding values and on-farm breeding values is significantly less than unity (ie., < 0.6). Fortunately, genomic selection platforms allow for field data to be linked to nucleus broodstock through DNA derived genomic relationships and on-farm phenotyping. Secondly, genomic selection programs become increasingly more complex when harvest yields are determined by diverse genomes, as is the case of pearl oyster, with a host recipient seeded with the saibo of a donor. The need to have accurate breeding values for both host and donor oyster may eventually result in the need of separate breeding lines for both. Unknown interactions between host and donor further complicate the application of genomic selection if such epistatic effects are significantly greater than zero. In the case of pearl oyster the multi-factorial nature of pearl value adds to the complexity of setting up multi-trait genomic selection. Thirdly, and potentially of greatest commercial appeal for genomic selection is to build disease resistance into the genetic improvement program as has been highlighted above. Most central nucleus breeding programs are pathogen free and breeding decisions are based on family sib-selection, but commercial grow out environments are under constant disease challenge. It is unlikely that simply screening commercial stocks will yield data of sufficient quality to obtain genomic breeding values for disease resistance, since most disease field challenges are uncontrolled, and often resistance to multiple pathogens is of interest. One potential solution is to expose large mixed-family progeny cohorts to standardised disease challenge and ascertain survival statistics from pooled genotype data pre- and post-challenge. Finally, it is almost certain that for most genomic selection programs, there will be a need for ongoing phenotyping to update the training sets, and cross validate data collected under diverse commercial environments and to monitor unfavourable genetic correlated responses.

Perhaps one of the greatest advantages offered by application of genomic selection over conventional breeding programs, is that large-scale multi-family data can be resolved retrospectively through genomic relationships. This has two immediate and highly significant advantages. Firstly, the predicted genetic response and realised inbreeding are far superior over the management of multiple single-family lines. Simple simulation shows that a cohort of 100 families in a single line outperforms the average of 100 single-family lines and creates long-term sustainable value for the industry (Khatkar *et al.* 2017b, these proceeding). Secondly, the enormous costs in establishing and maintaining single-family mating, spawning and rearing facilities are not required under a genomic selection program using a large scale multi-family breeding program. In many cases the commercial infrastructure for propagation is sufficient, and the cost saving outweighs the cost incurred for genotyping.

In our experience, the transition from existing/traditional selection programs into a genomic selection program is challenging since most mating and infrastructure designs in central nucleus

breeding facilities do not capture the advantages offered by genomic selection programs. In the case where simple mass produced commercial stocks are produced, or where no genetic improvement programs are in place, imposing a genomic selection program is potentially straightforward. The main requirement is that the species is domesticated, since lifecycles need to be closed for ongoing selection and capture of genetic gain. Where source broodstock has been harvested from wild stock, the base generation needs to be adequately represented in the foundation stocks, and inclusion of "new" ongoing sampling of wild stocks limited. Once an adequate training data set against commercially well-defined breeding objectives has been completed, a robust test-set and validation phase is required to determine the accuracy of the genomic predictions. For easy to measure traits of moderate to high heritability, this is relatively easy to achieve; however, for most, if not all diseases, and complex multi-factorial traits, the development of adequate training data sets will remain a logistical challenge. Of practical concern is also how best to use available information. For most applications, genotyping potential candidates under selection remains a significant cost. The use of multi-stage selection, based on simple phenotypic selection as a primary selection, followed by genomic sampling (DNA sampling genotyping and tracking tagged individuals) and selection is likely the most cost-effective application of this technology (Khatkar *et al*. 2017b these proceeding). Other applications of genomic selection include the genomic management to minimize inbreeding by candidate selection and mate allocation to maximize genomic diversity. Genomic selection also offers an additional commercial benefit, to pre-screen females and males in the current generation for production of commercial animals, given that relatively few females are needed to generate the many millions of larvae for commercial production. The exact benefits of GS breeding programs will be dependent on the species and nature of the aquaculture enterprise.

REFERENCES

Abdelrahman H., ElHady M., Alcivar-Warren A., Allen S., et al. (2017). *BMC Genomics*, **18**:191, doi:10.1186/s12864-017-3557-1.

Cobb J.N., Declerck G., Greenberg A., Clark R., McCouch S. (2013). *Theoretical Applied Genetics*, **126**:867–887.

Daetwyler H.D., Villanueva B., Woolliams J.A. (2007). *Journal of Animal Breeding Genetics*, **124**:369-376.

Dou J., Li X., Fu Q., Jiao W., Li Y., et al. (2016). *Scientific Reports*, **6***:*19244. doi:10.1038/srep19244.

Eggen A. (2012). Animal Frontiers 2:10–15.

Forni, S., Aguilar I., Misztal I. (2011). *Genet Sel Evol*, **43**:1, doi:10.1186/1297-9686-43-1.

Gjedrem, T. (2005) In: 'Selection and Breeding Programs in Aquaculture', editor Gjedrem T., Springer, Berlin, Heidelberg.

Grys B.T., Lo D.S., Sahin N., Kraus O.Z., Morris Q., Boone C., Andrews B. (2016). *Journal of Cell Biology*, doi:10.1083/jcb.201610026.

Hayes B.J., Pryce J., Chamberlain A.J., Bowman P.J., Goddard M.E. (2010). *PLoS Genet* **6**(9): e1001139.

Hickey, J.M., Kinghorn, B.P., Tier, B., van der Werf, J.H.J., Cleveland, M.A. (2012*). Genet Sel Evol*, **44**:9, doi:10.1186/1297-9686-44-9.

Jones D.B., Toole P.B., Khatkar M.S., Raadsma H.W., Jerry D.R., Zenger K.R. (2017). Proceedings of the AAABG 22nd Conference 2017, 2nd-5th July 2017, Townsville, QLD, Australia.

Khatkar, M. S., G. Moser, B. J. Hayes and H. W. Raadsma (2012). *BMC Genomics* **13**: 538, doi:10.1186/1471-2164-13-538.

Khatkar M.S., Zenger K.R., Jones D.B., van der Steen H.A.M., Jerry D.R., Raadsma H.W. (2017a). Proceedings of the AAABG 22nd Conference 2017, 2nd-5th July 2017, Townsville, QLD, Australia.

Khatkar M.S., Coman,G.J., Thomson, P.C., Raadsma H.W. (2017b). Proceedings of the AAABG 22nd Conference 2017, 2nd-5th July 2017, Townsville, QLD, Australia.

Kustrin S.A., Morton D.W. (2015) *Modern Chemistry Applications*, **3**:152, doi:10.4172/2329-6798.1000152.

Lal M.M., Southgate P.C., Jerry D.R., Zenger K.R. (2016). *Marine Genomics* **25**:57–68.

Lind, C. E., Kilian A., Benzie J.A.H. (2017). *Animal Genetics* **48**:362-364.

Liu D., Zeng X.A., Sun D.W. (2013). *Applied Spectroscopy Reviews*, **48**:609-628.

Ødegård, J., Moen T., Santi N., Korsvoll S.A., Kjøglum S., Meuwissen T.H.E. (2014) *Frontiers in Genetic*s, **5**: 402

Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001). *Genetics*, **157**:1819-1829.

Miranda J.M. and Romero M. (2017). *Aquaculture Engineering*, **76**:41-49.

Nielsen H.M., Sonesson A.K., Yazdi H., Meuwissen T.H.E. (2009). *Aquaculture*, **289**:259-264.

Palaiokostas C., Ferraresso S., Franch R., Houston R.D., Bargelloni L. (2016) *G3 (Bethesda)* **6**: 3693–3700.

Robledo D., Palaiokostas C., Bargelloni L., Martínez P., Houston R. D. (2017)**.** *Reviews in Aquacultur*e. **0**: 1-13. doi:10.1111/raq.12193.

Sonesson A.K., Meuwissen T.H.E. (2009). *Genet Sel Evol* **41**:37, doi:10.1186/1297-9686-41-37

Tsai H.Y., Hamilton A., Tinch A.E., Guy D.R., Bron J.E., Taggart J.B. et al. (2016). *Genet Sel Evol* **48**:47, doi:10.1186/s12711-016-0226-9.

Tsai, H.Y., Matika O., Edwards S. M., Antolin-Sanchez R., Hamilton A., Guy D. R., Tinch A. E. K. Gharbi, M. J. Stear, J. B. Taggart, J. E. Bron, J. M. Hickey and R. D. Houston (2017). *G3 (Bethesda)* **7**(4): 1377-1383.

Vallejo R.L., Leeds T.D., Gao G., Parsons J.E., Martin K.E., Evenhuis J.P., Fragomeni B.O., Wiens G.D., Palti Y. (2017). *Genet Sel Evol*. **49**:17, doi:10.1186/s12711-017-0293-6.

Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A. and Urban, J. (2016). *Reviews in Aquaculture*, doi:10.1111/raq.12143.

Viazzi S., Van Hoestenberghe S., Goddeeris B.M., Berckmans D. (2015). *Aquacultural Engineering*, **64***:* 42–48.

Yu Y., Zhang X., Yuan J., Li F., Chen X., et.al. (2015). *Scientific Reports*, **5**:15612, doi:10.1038/srep15612.