Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Genomics in Selective Breeding of Atlantic Salmon

Thomas Moen[1] and Jørgen Ødegård[1]
[1]AquaGen, Trondheim, Norway

ABSTRACT: Genomics is beginning to have an impact on Atlantic salmon breeding. For the past two decades, a large number of valuable tools have been delivered within the research community, including the recently completed genome sequence. A breakthrough for genomics in Atlantic salmon breeding came with the implementation of MAS for in increased resistance to Infectious Pancreatic Necrosis (IPN). In Norway, MAS for IPN-resistance has contributed strongly to a 75 % decline in the number of IPN-outbreaks since 2009. The putative functional mutations, underlying this QTL, has now been identified, and are being investigated using functional testing. With high-density SNP-chips now being available, breeding companies are also starting to implement genomic selection (GS), shown in simulation studies to have large potential in Atlantic salmon.
Keywords: Atlantic salmon; genomics; marker-assisted selection; genomic selection; infectious pancreatic necrosis

The 'national' Norwegian breeding programme for Atlantic salmon (reviewed in Gjøen and Bentsen (1997)), initiated in the early 1970s by scientist at Norwegian University of Life Sciences (and today run by AquaGen), has been a role model for today's Atlantic salmon breeding programmes across the world. In a typical breeding programme for Atlantic salmon, therefore, a number of siblings groups are produced every generation, using some form of hierarchical design, and kept separate until the fish are large enough to be physically tagged. After tagging, the fish are kept in communal tanks or net pens until they are sexually mature. Selection candidates are recorded for growth rate and other traits that can be measured non-invasively. Slaughter traits (filet colour, fat content etc.) and disease resistance traits (resistance to individual pathogens and parasites) are measured on siblings of the selection candidates. The selection is therefore mainly family-based. Until recently, developments have come particularly in the form of new traits and novel phenotypic measurements, mating designs, innovations related to animal husbandry, and reproductive technologies.

In recent years, genomics has emerged as a potentially major contributor to selective breeding in Atlantic salmon. This prospect has contributed to a build-up of genomic resources for Atlantic salmon seen throughout the past two decades. Microsatellite markers were identified, first from sequencing of repeat-enriched genomic libraries (e.g. McConnell et. al. (1995); Sanchez et. al. (1996); O'Reilly et. al. (1998)), next from mining of Expression Sequence Tag (EST) databases (Ng et. al. (2005); Vasemägi et. al. (2005)) or sequencing of the ends of Bacterial Artificial Chromosomes (BACs) (Danzmann et. al. (2008); Phillips et. al. (2009)). The first large sets of Single Nucleotide Polymorphism (SNP) markers came from alignment of EST sequences (Hayes et. al. (2007); Andreassen et. al. (2010)), made feasible by earlier studies that had delivered large numbers of EST-sequences from different individuals and tissues (Davey et. al. (2001); Martin et. al. (2002); Rise et. al. (2004); Tsoi et. al. (2004); Hagen-Larsen et. al. (2005); Adzhubei et. al. (2007); Koop et. al. (2008); Andreassen et. al. (2009); Leong et al. (2010)). Later, SNPs were also identified using next-generation sequencing of reduced-representational genomic libraries (Lien et. al. (2011)). At the initiation of the Atlantic salmon sequencing project, researchers had access to linkage maps containing hundreds of microsatellite markers and thousands of SNP markers (Danzmann et. al. (2008), Lien et. al. (2011)), integrated with a BAC-based physical map (Ng et. al. (2005); Thorsen et. al. (2005); www.asalbase.org) and with the karyotype (Phillips et. al. (2009)). For gene expression studies, several microarrays had been developed (Rise et. al. (2004); von Schalburg et. al. (2005); Koop et. al. (2008); Taggard et. al. (2008); Krasnov et. al. (2011)).

In 2010, the sequencing of the Atlantic salmon genome project was initiated, in a collaborative project funded by Canadian, Chilean, and Norwegian funding agencies as well as by four private aquaculture companies (Davidson et. al. (2010)). In 2014, the project will release a high-quality genome reference of the Atlantic salmon, the end product of a project that turned out to be very challenging due to the partly duplicated and repeat-rich nature of the Atlantic salmon genome. The availability of a genome reference, and of next-generation sequencing technologies, have facilitated identification of truly large sets of DNA markers. Thus, a SNP-chip containing more than 657k polymorphic SNPs was recently produced as a collaboration between AquaGen, CIGENE, and Affymetrix. This chip was produced on the basis of whole-genome Illumina-sequencing of 28 normal AquaGen-salmon and 3 double haploids, and there are reasons to believe that the chip incorporates a significant fraction of the SNPs segregating in the AquaGen population. A SNP-chip, containing 132k polymorphic SNPs was developed by Roslin Institute, Landcatch, and Affymetrix, based on the sequencing of restriction-site associated DNA (Houston et. al. (2014)), and a linkage map has been made using this SNP-chip (Gonen et. al. (2014)). With these developments, Atlantic salmon researchers have access to resources that are comparable to those available for major livestock species.

Initially, genomics was seen as contributor to breeding in aquaculture mainly through the identification of Quantitative Trait Loci (QTL) and the implementation of such QTL in Marker-Assisted Selection (MAS). Ge-

nome scan have been conducted, searching for QTL for many different traits, including growth rate (Reid et. al. (2005); Boulding et. al. (2008); Houston et. al. (2009); Baranski et. al. (2010); Gutierrez et. al. (2012)), fillet colour (Houston et. al. (2009); Baranski et. al. (2010)), fillet texture and fat content (Sodeland et. al. (2013)), age at sexual maturation (Gutierrez et. al. (2014)), resistance to infectious salmon anaemia (Moen et. al. (2004); Moen et. al. (2007)), and infectious pancreatic necrosis (IPN) (Houston et. al. (2008); Moen et. al. (2009)). Most of these studies have been performed on Atlantic salmon derived from European aquaculture populations, while some have been performed on crosses between populations. Some of the studies have utilised a SNP-chip with approximately 6k working assays (Lien et. al. (2011)), others have been using microsatellites. Some studies are based on linkage analysis (interval mapping), others are based on association-testing of large numbers of individual markers (genome-wide association studies).

The identification of a major QTL for resistance to IPN (Houston et. al. (2008); Moen et. al. (2009) turned out to be a breakthrough for the application of genomics in aquaculture breeding. Two research groups independently discovered a major QTL for IPN-resistance on Atlantic salmon chromosome 26, employing 10 large full-sib groups coming from a Scottish and a Norwegian breeding population, respectively. The QTL turned out to be responsible for most or all of the genetic variation in IPN resistance both at the fry and the post-smolt life stages in Atlantic salmon (Moen et. al. (2009)). A haplotype of microsatellites located within the QTL region was found to be highly predictive of the individual animals' genotypes at the QTL, and thus strongly associated with the trait at the population level (Moen et. al. (2009)). Since breeding companies were directly involved in the identification of the QTL, the road to commercialization of the results was short. AquaGen started employing a haplotype-based test for identifying animals that were homozygous for the high-resistance QTL allele, marketing eggs coming from these animals as 'IPN-QTL eggs'. In this way, the genetic improvement could be delivered to the industry without delay. MAS for IPN resistance was also implemented in the breeding nucleus and in selection of parents for multiplier lines (the latter already in 2007). Interest in AquaGen's IPN-resistant salmon eggs selected using MAS was substantial already from the first season (2009), and today, most of the company's customers choose eggs from MAS-selected parents. A monitoring project conducted by AquaGen has shown that their MAS-selected salmon are highly resistant to IPN, both in controlled challenge experiments and in the industry. Other breeding companies have chosen similar or other strategies for implementing the IPN-QTL in their breeding programmes. The number of IPN outbreaks on Norwegian Atlantic salmon has currently dropped by 75 % relative to the period prior to the implementation of MAS for IPN resistance (Figure 1), attributable, according to Norwegian fish health authorities, in a large part to the effectiveness of MAS for IPN-resistance (Norwegian Veterinary Institute (2012); Norwegian Veterinary Institute (2013)).

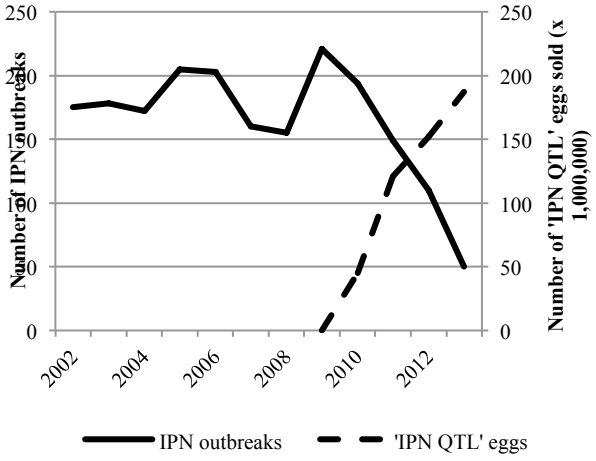

Figure 1: Number of IPN-outbreaks in Norwegian Atlantic salmon farms versus number of "IPN-QTL" eggs sold by AquaGen (total number of eggs across egg providers is ~350 million eggs per year).

A collaborative project, involving AquaGen, CIGENE, Nofima (Ås, Norway), and Simon Fraser University (Burnaby, Canada) has been searching for the causative mutation(s) underlying the QTL for IPN resistance. This project has been making use of many of the genome resources mentioned above. Based on the BAC-based physical map, a minimum tiling path of BAC clones, covering the QTL region, was made. These clones were sequenced using Illumina technology, and a reference sequence of the QTL region was constructed. Simultaneously, animals deduced to be homozygous for the high-resistance or the low-resistance allele, respectively, were whole-genome sequenced using Illumina technology, and reads were aligned to the reference sequence of the QTL region in order to identify DNA variations displaying strong contrast between the QTL genotype groups. The publicly available collections of EST- and cDNA sequences were used in order to annotate the QTL region and assign putative functions to the identified DNA variations. After the commencement of the Project to Sequence to Atlantic Salmon Genome, reads and assemblies from this project were made use of in order to improve the reference sequence of the QTL region and to extend the search to a larger genomic region. In 2013, a breakthrough came when an amino-acid-shifting mutation, displaying a strong contrast between the QTL genotype groups, was discovered in a gene located within the QTL region. A little later, utilising the 930k SNP-chip, a second amino-acid-shifting mutation was discovered within the same gene was discovered, explaining the segregation patterns of the QTL not explained by the first-identified mutation. Functional studies have strongly indicated that the gene harbouring these two mutations is functionally connected with resistance to IPN. Simultaneously, microarray- and RNA-seq studies have revealed genes that are functionally connected with the gene harbouring the causative mutations.

The QTL for IPN-resistance has had an impact on the Atlantic salmon selective breeding sector in several ways. The success of the 'IPN-QTL' eggs has contributed to the development of further 'specialised products', in

most cases based on specific QTL. Different egg producers now offer eggs selected (by MAS) for high performance on IPN, pancreas disease (PD), fillet colour, and salmon lice. Coming along with this focus on QTL are a number of interesting ongoing research projects, and an increasing number of QTL that have been detected/validated in multiple experiments with high experimental power. The success of the 'IPN-QTL' eggs has led to an increased awareness and increased expectations of what can be achieved using genetics. The QTL for IPN-resistance is in many ways ideal for MAS, explaining a large fraction of the phenotypic variance, still segregating with an intermediate allele frequency in many populations (presumably because it was not under natural selection prior to salmon aquaculture), the high-resistance allele(s) being partly dominant over the low-resistance allele, and the trait having high economic value. Other, more complex traits will obviously be more challenging to deal with.

AquaGen and other salmon breeding companies are currently implementing genomic selection (GS), which is a methodology well adapted to utilization of genomic information in highly polygenic traits (Meuwissen et. al. (2001)). Rather than aiming for specific QTL or associated markers with high effect, GS uses dense genome-wide markers jointly for calculation of genomic estimated breeding values (GEBV). One of the main advantages of this methodology is that it provides a method for calculation of individual breeding values, even for fish without own phenotype, based on marker-phenotype associations among phenotyped and genotyped fish (training data). For aquaculture species, GS methodology has its highest potential for traits that are typically measured in sibs of the selection candidates (Ødegård et. al. (2009)). In a traditional family-based breeding program, this implies that selection can only be performed across families, but will be random within full-sib families, as the non-phenotyped selection candidates are evaluated based on their sibs. Using genomic selection, one can perform combined across and within-family selection, giving both faster genetic gain and reduced rates of inbreeding (as superior individuals can be selected even from less favourable families). The improved reliability of GS relative to classical pedigree-based selection (given the same data structure), can be explained by its ability to utilize three major sources of information from the genomic data (Habier et. al. (2013)): 1) Population-wide linkage disequilibrium (LD), 2) Co-segregation of loci, and 3) Additive-genetic relationships. LD is the statistical dependency between alleles at different loci in the base generation (i.e. the generation with unknown parents), while co-segregation is here defined as the deviation from independent segregation of alleles as a result of linkage (i.e., deviations between pedigree-based and linkage-analysis-based relationships across the genome), while the additive-genetic relationships are the classical pedigree-based relationships, which are implicitly included in dense marker data although pedigree information is not necessarily used directly. In contrast, classical pedigree-based selection can only utilize the last source of information (additive-genetic relationships).

Aquaculture populations, such as Atlantic salmon, are typically characterized by enormous male and female fecundity, resulting in large full-sibs families. Furthermore, partly factorial mating designs are often used, so that selection candidates may have large number of full-sibs as well as both maternal and paternal half-sibs. Using classical relationships, full-and half-sibs are assumed to be related by factors of ½ and ¼, respectively, while the genomic realized relationships (due to co-segregation) varies around these values, and will be reflected by the actual inheritance of dense marker genotypes. Furthermore, presumably "unrelated" animals may share smaller fractions of DNA, which will be captured by dense markers. Hence, GS is likely to have a substantial potential for selective breeding of Atlantic salmon. So far, documented genetic benefits from GS in aquaculture is largely absent for real populations, but there are a number of simulations studies indicating substantial increase in genetic gain expcially in sib-evaluated traits (e.g. Nielsen et. al. (2009); Ødegård et. al. (2009); Ødegård and Meuwissen (2014).

Literature Cited

Adzhubei, A. A., Vlasova, A. V., Hagen-Larsen, H. et. al. (2007). BMC Genomics, 8:209

Andreassen, R., Lunner, S., and Høyheim B. (2009). BMC Genomics, 10:502.

Andreassen, R., Lunner, S., and Høyheim, B. (2010). BMC Genomics, 11:706

Baranski, M., Moen, T., and Våge, D. I. (2010). Genet. Sel. Evol., 42:17.

Boulding, E. G., Culling, M., Glebe, B. et. al. (2008). Heredity, 101:381-391.

Danzmann, R. G., Davidson, E. A., Ferguson, M. M. et. al. (2008). BMC Genomics, 9:557.

Davidson, W.S, Koop, B. F., Jones, S. J. M. et. al. (2001). Genome Biology, 11: 403.

Davey, G. C., Caplice, N. C., Martin, S. A. et. al. (2001). Gene, 263:121-30.

Gonen, S., Lowe, N. R., Cézard, T. et. al. (2014). BMC Genomics, 15: 166.

Gjøen, H. M., and Bentsen, H. B. (1997). ICES J. Mar. Sci., 54:1009-1014.

Gutierrez, A. P., Lubieniecki, K. P., Davidson, E. A. et. al. (2012). Aquaculture, 358-359: 61-70.

Gutierrez, A. P., Lubieniecki, K. P., Fukui, S. et. al. (2014). Marine Biotech., 16:103-110.

Habier, D., Fernando, R. L, and Garrick, D. J. (2013). Genetics, 194:597-607.

Hagen-Larsen, H., Lærdahl, J. K., Panitz, F. et. al. (2005). BMC Genomics, 6:171.

Hayes, B., Lærdahl, J. K., Lien, S. et. al. (2007). Aquaculture, 265:82-90.

Houston, R. D., Haley, C. S., Hamilton, A. et. al. (2008). Genetics, 178:1109-1115.

Houston, R. D, Bishop, S. C., Hamilton, A. et al. (2009). Anim. Genet., 40:753-755.

Houston R. D., Taggart, J. B., Cézard, T. et. al. (2014). BMC Genomics, 15:90.

Koop, B. F., von Schalburg, K. R., Leong, J. et. al. (2008). BMC Genomics, 9:545.

Krasnov, A., Timmerhaus, G., Afanasyev, S. et. al. (2011). Comp. Biochem. Physiol. Part D, 6:31-38.

Leong, J. S., Jantzen, S. G., Schalburg, K. R. (2010). BMC Genomics, 11:279.

Lien, S., Gidskehaug, L., Moen, T. et. al. (2011). BMC Genomics, 12:615.
Martin, S. A., Caplice, N. C., Davey, G. C. et. al. (2002). Biochem. Biophys. Res. Comm., 293:578-585.
Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Genetics, 157:1819-1829.
McConnell, S. K., O'Reilley, P., and Hamilton, L. (1995). Can. J. Fish. Aquat. Sci., 52:1863-1872.
Moen, T., Fjalestad, K. T., Munck, H. et. al. (2004). Genetics, 167:851-858.
Moen, T., Sonesson, A. K., Hayes, B. et. al. (2007). BMC Genetics, 8:53.
Moen, T., Baranski, M., Sonesson, A.K. et. al. (2009). BMC Genomics, 10:368.
Nielsen, H. M., Sonesson, A. K., Yazdi, H. et. al. (2009). Aquaculture, 289: 259-264.
Norwegian Veterinary Institute (editor Johansen, R.) (2012). Fish Health Report 2012.
Norwegian Veterinary Institute (editor Hjeltnes, B.) (2013). Fiskerihelserapporten 2013 (Fish Health Report 2013, in Norwegian).
Ng, S. H. S., Artieri, C. G., Bosdet, I. E. et. al. (2005). Genomics, 86:396-404.
O'Reilly, P. T. O., Herbinger, C., and Wright, J. M. (1998). Anim. Genet., 29:363–370.
Phillips, R. B., Keatley, K. A., Morasch, M. R. et.al. (2009). BMC genetics, 10:46.
Reid, D. P., Szanto, A., Glebe, B. et. al. (2005). Heredity, 94:166-172.
Rise, M. L., von Schalburg, K. R., Brown, G. D. et. al. (2004). Genome Res., 14:478–490.
Sanchez, J. A., Clabby, C., Ramos, D. et. al. (1996). Heredity, 77:423-432.
Sodeland, M., Gaarder, M., Moen, T. et. al. (2013). Aquaculture, 408-409:169-174.
Taggart, J. B., Bron, J. E., Martin, S. A. et. al. (2008). J. Fish Biol., 72:2071-2094.
Thorsen, J., Zhu, B., Frengen, E. et. al. (2005). BMC Genomics, 6:50.
Tsoi, S. C. M, Ewart K. V., Penny, S. et. al. (2004). Marine Biotech., 6:199-214.
Vasemägi, A., Nilsson, J., and Primmer, C. R. (2005). Mol. Ecol. Notes, 5:282–288.
von Schalburg, K. R., Rise, M. L., Cooper, G. A. et. al. (2005). BMC Genomics, 6:126.
Ødegård, J., M. H. Yazdi, A. K. Sonesson et. al. (2009). Genetics, 181:737-745.
Ødegård, J., and Meuwissen, T. H. E. (2014). Genet. Sel. Evol., 46:3.