Marine Genomics 26 (2016) 41–50

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Evidence of recent signatures of selection during domestication in an Atlantic salmon population

A.P. Gutierrez [a,b], J.M. Yáñez [c], W.S. Davidson [a,*]

[a] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
[b] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK
[c] Faculty of Veterinary and Animal Sciences, University of Chile, Av. Santa Rosa, 11735, La Pintana, Santiago, Chile

ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form 25 November 2015
Accepted 16 December 2015
Available online 23 December 2015

Keywords:
Atlantic salmon
SNP array
Domestication
Signatures of selection
Grilsing

ABSTRACT

Selective breeding practices in Atlantic salmon aquaculture have been carried out intensively since the 1970s. Along with the phenotypic improvement of fish, we expect to observe genomic regions showing evidence of selection for traits related to growth and age at sexual maturation, as well as traits involved in the domestication process. This is mainly linked to the increase in the frequency of favourable alleles at loci that affect the traits of interest in the breeding population. In this study we searched for signatures of selection in the Cermaq Atlantic salmon broodstock, a Mowi strain, which was derived from wild Norwegian populations, and is now farmed in British Columbia, Canada. A 6.5K SNP array was used to genotype 202 fish from the Cermaq population, and the genotypes were compared with four wild populations from Norway. We used three methods based on F_{ST} values to detect signatures of selection. Forty four markers showing divergence in allele frequency were identified as outliers by the three detection methods, suggesting the presence of signatures of selection in the Cermaq population relative to their wild counterparts. Markers identified as outliers are associated with molecular functions that could be related to the selection for economically important traits (e.g., growth) as well as the domestication process (e.g., response to pathogens and environmental stressors). Of particular interest were three outlier markers that had been previously associated with grilsing (i.e., early sexual maturation) an undesirable trait, which has been heavily selected against in Atlantic salmon aquaculture. This study provides clear evidence of the presence of signatures of selection and domestication in a farmed Atlantic salmon population.

1. Introduction

Atlantic salmon genetic improvement has been practiced since the early 1970s when the first breeding trials began in Norway (Gjedrem, 2012). Since then, more than a dozen breeding programs have been established (Rohe et al., 2009), and traits such as growth rate, age at sexual maturity, pathogen resistance, flesh colour and fat content have been included in the breeding goal. The selection responses in salmonids have been higher than in other animal species, due to higher genetic variability and greater fertility, which allow the application of higher selection intensity. For instance, in Norwegian domesticated strains, the response to selection for growth related-traits was found to be greater than 10% in the first few generations (Gjedrem and Baranski, 2010). After more than a half a century of Atlantic salmon aquaculture and even considering the short period of time during which they have experienced intense artificial selection (approximately 12 generations of captive breeding for one Norwegian strain), it can be assumed that the allele frequency of selected loci and the levels of genetic variation have been affected within populations as a consequence of the domestication process. This includes the adaptation to new farming environments and intensive selection for economically important production traits.

Genetic improvement in livestock is mainly driven by increasing the frequency of favourable alleles at loci that affect the traits of interest in populations (Bijma, 2012). The magnitude of these increases is mainly determined by allele substitution effects and allele frequencies at these loci, along with the intensity and accuracy of artificial selection (Falconer and Mackay, 1996). Accordingly, if a population is heavily selected for a particular trait, then there is a higher chance of some alleles reaching fixation. At the same time, aquaculture practices may inadvertently decrease the genetic variation present in farmed stocks. Unless accurate pedigree records are maintained, there is a probability of selecting related individuals as parents, thereby increasing inbreeding.

A series of events may occur in genomic regions of populations affected by intensive selection. For instance a "hard sweep" is the process that occurs when the genetic variation in selected regions is disrupted leading to an association between an adjacent locus and the selected site. On the other hand, there is another scenario known as a "soft sweep", in which more than one positive allele can be present within

* Corresponding author.
E-mail address: wdavidso@sfu.ca (W.S. Davidson).

http://dx.doi.org/10.1016/j.margen.2015.12.007

the selected locus and a drastic reduction in genetic variation in the genomic region does not occur. Soft sweeps involve variants at a selected locus, and therefore they may produce many different haplotypes at the closely linked sites. Additionally, when adaptation occurs by selection of polygenic traits, it generally induces an increase in the allelic frequency of several loci, which have a favourable effect on the trait. These alleles do not necessarily reach fixation and the haplotype pattern corresponds to several partial selection signatures or multiple "partial sweeps" (Pritchard et al., 2010). Signatures of selection are genomic regions which contain DNA sequences affecting genetic variation of characters that have undergone natural or artificial selection (Qanbari et al., 2012). Such signatures may be identified using genomic information and different analytical approaches (Lopez et al., 2015). This is possible because selection can affect the DNA sequence at a particular region. This is known as the "hitchhiking effect" (Smith and Haigh, 1974). This leaves a "footprint" or "signature" around the selected gene variant, which yields a specific and detectable genomic sequence pattern (Pennings and Hermisson, 2006).

Previous studies have shown evidence of directional selection in genomic regions of farmed Atlantic salmon when compared to their wild counterparts (Vasemägi et al., 2012) or when comparisons are made among pairs of farmed populations (Martinez et al., 2013), even when a limited number of markers is used. Using a SNP chip, Karlsson et al. (2011) were able to identify a panel of 60 markers capable of differentiating farmed Atlantic salmon from wild populations in Norway. However, they did not provide evidence of significant genetic differentiation between them. Based on these observations it has been suggested that genetic differences on a genomic scale, between Norwegian farmed and wild populations are the result of small allele frequency changes at a large number of loci, rather than large allele changes at few loci.

The development of technologies for typing dense marker genotypes provides the opportunity to simultaneously analyse thousands of SNPs and more precisely identify regions of the genome that show evidence of selection. These tools have already been used for the analysis of many livestock species such as cattle, sheep, pig and chicken (Ai et al., 2013; Kijas et al., 2012; Qanbari et al., 2012, 2014; Ramey et al., 2013), by performing genome-wide scans comprising from hundreds of thousands to millions of SNPs. The identification of selection signatures may help to unravel the genetic factors and mechanisms involved in important biological traits, because these regions might have adaptive and functional relevance underlying their selection (Nielsen et al., 2007). Atlantic salmon provide an excellent model for studying the effects of early selection and the domestication process, as some of their populations have been domesticated very recently in their evolutionary history and there is the availability of both domesticated and wild populations from the same lineage. This study was designed to detect signatures of domestication and early selection processes in Atlantic salmon using an Illumina iSelect SNP-array (Kent et al., 2009) in order to get a better understanding of Atlantic salmon adaptation to a farming environment and high selection pressures for production traits.

2. Materials and methods

2.1. Farmed fish samples and wild fish genotype data

The Atlantic salmon used in these analyses were part of a commercial broodstock program initiated by Cermaq Canada in 1995 and based on the Mowi strain of Atlantic salmon (imported to British Columbia, Canada in the mid 1980s). The Mowi strain was established in the late 1960s with a major contribution from the River Bolstad, in the Vosso watercourse, and the River Aaroy with additional contributions from wild salmon captured in the sea near Osterfjord and Sotra in western Norway (Glover et al., 2009; Verspoor et al., 2007). Considering the time since the Mowi strain was established, we estimate that it has gone through approximately 12 generations of selection, seven of these in Norway and five in Canada. The Cermaq population has been selectively bred for faster growth and reduced early sexual maturation. We chose 202 parents from the 2005 broodstock year. DNA from all 202 fish was obtained from fin-clips using the DNeasy Blood & Tissue Kit (Qiagen) and then sent for genotyping. SNP genotyping was carried out at CIGENE, Norwegian University of Life Sciences, Ås using an Atlantic salmon 6.5K Illumina iSelect SNP-array (Kent et al., 2009), described by Lien et al. (2011). The level of relatedness of the 202 fish was estimated from a genomic kinship matrix based on identity-by-state (IBS) values obtained using the *GenABEL* package implemented in R. The mean value was −0.0025 with a median of −0.0069, indicating low levels of relatedness.

To perform comparative analyses we used the genotype data from four Norwegian wild Atlantic salmon populations available at the Dryad Digital Repository (http://datadryad.org), obtained and made publicly available by Bourret et al. (2013). The wild populations were identified according to the region of origin (Fig. 1), as follows: Tana (TAN) with 29 individuals (1), Gaula (GAU) with 43 individuals (2), Laerdaselva (LAR) with 25 individual (3) and Numedalslagen (NUM) with 43 individuals (4), giving a total of 140 wild samples, which together comprised the WILD dataset. More details regarding these wild populations can be found in Bourret et al. (2013). These populations were chosen to provide a good representation of the wild populations in Norway (based on their geographical provenance) for comparisons with the Mowi strain, which can be considered as initially being a composite strain established with contributions from different Norwegian rivers.

For comparative purposes, the data-sets from both, the WILD set and the Cermaq populations were filtered to keep only the shared markers. These markers were used in the subsequent analyses, after filtering for a call rate threshold of 95%.

2.2. Basic population genetic statistics and structure

Deviations from Hardy–Weinberg equilibrium (HWE) were tested with the exact test (Guo and Thompson, 1992), as implemented in GENEPOP 3.4 (Rousset, 2008) and Arlequin 3.5 (Excoffier and Lischer, 2010). Genetic differentiation between populations was measured with pairwise F_{ST} estimates (Weir and Cockerham, 1984), using Arlequin 3.5 (Excoffier and Lischer, 2010). Genetic distances between populations were estimated based on Nei (1972), implemented in the R package *adegenet* (Jombart, 2008). Inbreeding coefficients (F_{IS}) were calculated using the R package *Demerelate* (Kraemer and Gerlach, 2013).

Population structure was inferred from the SNP markers using a hierarchical Bayesian modelling construct in the program *STRUCTURE* (Pritchard et al., 2000; Falush et al., 2007), using a burn-in of 100,000 iterations and running 100,000 iterations ($K = 2$). Then we conducted an individual-based principal component analysis (PCA) implemented in the R package *adegenet* (Jombart, 2008). We used the function *find.clusters* to estimate the optimal number of groups with the Bayesian Information Criterion (BIC) method, and we used the function *a.score* to determine the optimal number of discriminant functions to retain (Jombart et al., 2010).

2.3. Detection of F_{ST} outlier loci

Three different tests for the detection of loci subject to directional selection during domestication were used. These tests are based on different assumptions, but rely on the rationale that directional selection increases genetic differentiation between populations and reduces variation at linked loci, providing additional support for the identification of outlier loci. For the first stage of the analyses, the data from the four wild Atlantic salmon populations were grouped into one set of 140 samples (WILD) that was compared against the dataset from the 202 farmed samples.