Identification of a single genomic region associated with seasonal river return timing in adult Scottish Atlantic salmon (Salmo salar L.) identified using a genome-wide association study.

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Fisheries and Aquatic Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>cjfas-2017-0293.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>19-Oct-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Cauwelier, Eef; Marine Scotland Science Freshwater Laboratory Pitlochry Gilbey, John; Marine Scotland Science Freshwater Laboratory Pitlochry Sampayo, James; Marine Scotland Science Freshwater Laboratory Pitlochry Stradmeyer, Lee; Marine Scotland Science Freshwater Laboratory Pitlochry Middlemas, Stuart; Marine Scotland Science Freshwater Laboratory Pitlochry</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue?</td>
<td>N/A</td>
</tr>
<tr>
<td>Keyword:</td>
<td>genome-wide association, life history variation, adaptation, CONSERVATION &lt; General</td>
</tr>
</tbody>
</table>
Identification of a single genomic region associated with seasonal river return timing in adult Scottish Atlantic salmon (*Salmo salar* L.) identified using a genome-wide association study.

Eef Cauwelier*, John Gilbey*, James Sampayo, Lee Stradmeyer, Stuart J. Middlemas

Marine Scotland Science, Freshwater Fisheries Laboratory, Faskally, Pitlochry, PH16 5LB, Scotland, U.K. * Joint first authors

**Keywords:** genome-wide association, life history variation, adaption, conservation

**Author for correspondence:**
Eef Cauwelier, Marine Scotland Science, Freshwater Fisheries Laboratory, Faskally, Pitlochry, PH16 5LB, Scotland, U.K.
Tel: 0044 1224 294445 Fax: 0044 1796 473523
e-mail: Eef.Cauwelier@gov.scot

https://mc06.manuscriptcentral.com/cjfas-pubs
Abstract: Examination of the genetic basis of the timing of the return migration of Atlantic salmon (*Salmo salar* L.) to freshwater from the sea, a trait of economic and conservation interest, was carried out using a genome-wide association study. Genotype data from 52,731 single nucleotide polymorphic (SNP) markers from 73 early and 49 late running two sea-winter salmon from five rivers in eastern Scotland was examined. A single region of the Atlantic salmon chromosome Ssa09 was identified, containing nine SNP markers significantly associated with run timing, a region previously linked to variation in sea age at maturity. Validation of the markers in a group of 233 one and two sea-winter fish, including adults from a novel river, again showed significant associations between the trait and the Ssa09 region, explaining ~24% of the trait variance. The SNP loci identified provide the ability to examine trait variation in populations of Atlantic salmon and so help facilitate conservation management of the differing run timing phenotypes.

Introduction

While Atlantic salmon (*Salmo salar* L.) spawn in late autumn/early winter, their return migration to fresh water from the sea can occur at any time during the year (e.g. Juanes et al. 2004; Quinn et al. 2006; Shearer 1990). Individuals therefore arrive in rivers earlier than is necessary for them to reach their breeding ground. The evolutionary advantages of this strategy are still unclear (Quinn et al. 2016) and, although the mechanisms behind such a variation in return timing are unknown, the consequences for the management and exploitation of stocks are better understood. The diversity in the run timing of Atlantic salmon (defined as the time of return of adults to rivers from the sea) is of great significance to fisheries, as it results in fresh-run fish being spread out through the year, increasing the temporal extent of the catching period for these highly prized fish. In addition to its commercial importance in fisheries, seasonal variation in run timing is also of particular importance for conservation, where phenotypic population diversity is known to have important stabilising effects on stocks, due to a portfolio effect where variability in the overall
aggregated stock complex is significantly reduced compared to the variabilities of the different component stocks (Greene et al. 2009; Hilborn et al. 2003; Schindler et al. 2010). In turn, this has been shown to be important for the longer term sustainability of the larger stock complexes, on which the fisheries ultimately rely (Hilborn et al. 2003).

Atlantic salmon have a strong tendency to home to natal rivers to reproduce, often targeting natal breeding streams (Quinn et al. 1999; Youngson et al. 1994). As a consequence, genetic differences between these reproductively isolated populations evolve as a result of limited gene flow (Dionne et al. 2008; Elo 1993; Youngson et al. 2003), bottlenecks, genetic drift and natural selection; the latter leading to genetic adaptation of populations to local environmental conditions (Garcia de Leaniz et al. 2007). In Atlantic salmon, heritability of seasonal return timing has been shown using translocation experiments involving different populations (Hansen and Jonsson 1991; Stewart et al. 2002). However, direct estimates of the levels of heritability are not available for the species, but estimates from both pink (Oncorhynchus gorbuscha) and chinook (O. tshawytscha) salmon suggest it is likely to be relatively high (Quinn et al. 2000; Smoker et al. 1998).

Across the various salmonids, the genetic basis of a number of traits, including migratory run timing has been examined using both candidate loci and genome-wide association studies (GWAS) (Bush and Moore 2012; McCarthy et al. 2008); the latter made possible through the advances in next generation sequencing technologies, which has resulted in the inexpensive production of large quantities of sequence data (Metzker 2010) and the associated identification of large numbers of novel genetic markers including single nucleotide polymorphisms (SNPs). Candidate loci investigations into the genomic basis of migration timing in salmonids, based on the circadian Clock genes, have identified associations between genetic variation at these loci and run-timing in Chinook (O’Malley et al. 2007; O’Malley et al. 2013), chum (O. keta) Atlantic (O’Malley et al. 2014) and possibly pink salmon (O’Malley et al. 2010). However, the difficulty of disentangling other factors,
such as competing selection and epistatic or pleiotropic interactions suggest that caution should be
taken when predicting the function of these loci across species (O’Malley et al. 2010).

GWAS studies focussing on migratory-related traits in salmonids have, as in other ontogenetic
life history traits (Ayllon et al. 2015; Gutierrez et al. 2015; Johnston et al. 2014; Kusche et al. 2017),
reported variation in the genomic architectures underlying them. A complex polygenic basis for the
propensity to migrate in rainbow/steelhead trout (O. mykiss) (Hecht et al. 2013; Hecht et al. 2012)
has been identified, with several loci of small effect distributed throughout the genome and some
evidence of master control regions on two chromosomes (Hecht et al. 2012). In contrast to this
polygenic genomic architecture, both Hess et al. (2016) and Prince et al. (2017) reported a number
of SNPs associated with seasonal return timing in steelhead trout and Chinook salmon, all of which
mapped to a small single chromosome region that aligned to Atlantic salmon chromosome 3 and
overlapped the GREB1L gene, an oestrogen target gene. Considering the complexity of the seasonal
migratory trait, and the power of the analysis undertaken, the strength of the associations observed
were surprising (Prince et al. 2017).

A further factor adding complexity to the determination of the genomic architecture of
various life history traits is the sometimes inconsistent patterns of association between marker and
trait across different populations (Hess et al. 2016, Prince et al. 2017, Nichols et al. 2016, Showalter
2017). Both Hess et al. (2016) and Prince et al. (2017) reported a consistent pattern of control of the
GREB1L gene influencing seasonal return timing across some populations, but Showalter (2017)
found population-specific genetic effects in this marker-trait interaction. Similar population-specific
inconsistencies have been observed with differing migratory strategies in Sockeye salmon (O. nerka)
(Nichols et al. 2016).

Genomic control of complex traits is thus seen to vary greatly across traits, species and
populations. However, understanding such controls and identifying markers linked to traits of
interest is important when managing structured populations and trying to identify conservation
units within rivers, which may be defined by their heritable trait differences (Allendorf et al. 2010;
In this study, we used genotype data from a 220,000 SNP array (Barson et al. 2015) to perform a genome-wide association study (Lewis 2002; McCarthy et al. 2008) to identify SNPs associated with seasonal return timing in two-sea-winter Atlantic salmon. Chromosome maps were used to identify genomic regions and screen for candidate loci associated with the trait. Significantly associated markers were validated in a new set of adults of both one- and two-sea-winter and including individuals from a novel river to confirm the observed genetic architecture across different life-history types and novel populations.

Materials and Methods

Study sites and sample information

The Atlantic salmon analysed in the present study comprised of adult fish from seven rivers located along the east coast of Scotland (Fig. 1). Fin clips, muscle tissue or scales were taken from returning adults either rod-caught within rivers or from within-river and/or coastal netting stations (Table 1, with full details of fish screened contained in Supplementary data S1). The seasonal return timing phenotype of each fish was based on date of capture. For fish caught within rivers, external features characteristic of freshly run fish, such as a silver colour and presence of sea lice (Youngson and Hay 1996), were identified. The sea age of each fish was defined either directly from scale readings or, where such readings were not available, using the predictive algorithms based on fish length and/or weight and date of capture, as defined by Bacon et al. (2010).

Samples were divided into two groups: 122 two sea-winter fish (2SW) from five rivers were used for the GWAS, whilst 233 2SW fish and 212 one sea-winter fish (1SW) from four rivers were used in the validation process (Table 1). In order to maximise the power of detection of markers linked to genes influencing run timing, GWAS fish were selected from the extremes of the phenotypic distribution available (Lander and Botstein 1989; Li et al. 2011). This resulted in fish
classified and grouped as either early (caught in March and April) or late running (caught between July and September) (Table 1).

**DNA extraction and SNP array genotyping**

DNA of fish used in the GWAS was purified and quantified as described in Gilbey et al. (2016), whilst DNA from samples used in the validation process were extracted using a Chelex extraction protocol (Walsh et al. 1991). GWAS fish were screened at the Centre for Integrative Genetics (CIGENE, Norway), using a custom 220,000 single nucleotide polymorphism (SNP) Affymetrix Axiom array according to the manufacturer’s instructions. These SNPs were a subset of those on the 930,000 SNP XHD Ssal array (Barson et al. 2015).

Initial quality control (QC) of the SNPs and samples was carried out at CIGENE. Only truly polymorphic SNPs were retained for the analysis. Multiple steps were taken to assess sample quality. An initial dish QC (DQC) analysis examined signal strength and colour resolution of ~7,000 monomorphic SNPs and samples with a DQC < 0.82 were removed. A second step examined call rates of all samples for ~20,000 SNPs, and only samples with a call rate >0.97 were included in the GWAS. It was found that, following this initial high threshold sample QC, all but one of the old scale samples from the rivers Tay and Spey had failed at the DQC stage. As such, a second QC process was carried out by lowering the thresholds and examining samples with a DQC > 0.5 and call rate > 0.85 and assigning genotypes based on the cluster patterns observed in the samples included after the initial high threshold QC. From that, a multidimensional scaling analysis was carried out to investigate the correspondence of the data with juveniles from the same rivers that had been screened and genotyped for the same Affymetrix Axiom array (unpublished data). A Fisher’s exact test of allele counts for each SNP was carried out and those SNPs showing significant deviation from the expected distribution, based on the juvenile data, were excluded. Finally, SNPs with a minor allele frequency < 0.05 were removed (Houston et al. 2014).

**Genome-wide association study (GWAS)**
GWAS was performed using the R-package (R Core Team 2015) GAPIT (Lipka et al. 2012), whereby a linear mixed effects model (Zhang et al. 2010) was used to investigate the association between each SNP locus and return timing, with return timing treated as a continuous variable based on Julian return day. River of origin and year of capture were included as fixed effects in the initial model, and the significance of the inclusion of principle component (PC) scores to account for underlying population genetic structure determined (Lipka et al. 2012). Relatedness between individuals was accounted for by the inclusion of a kinship matrix (VanRaden 2008) calculated within GAPIT and included as a random effect. Model selection for fixed effects was undertaken using Bayesian information criterion (BIC) based model selection within GAPIT. The significance of the final model fit to the return timing data was assessed using the results of likelihood ratio tests controlled for multiple tests using a false discovery rates (FDR) correction (Benjamini and Hochberg 1995). Differentiation of run timing phenotypes using significantly associated loci was examined using principal component analysis (PCA).

**Genome mapping**

The mapping positions of all SNP loci used in the analysis remain unpublished. However, following the GWAS, mapping information was obtained (from correspondence with CIGENE and from Johnston et al. 2014) for the top ranked 496 SNPs, together with a further 500 SNPs that were available from a previous analysis (full details in Supplementary data S2). Associations of these 996 SNPs with map positions and return timing were visualised using Manhattan plots created using the R package qqman (Turner 2014).

**Alignment and candidate genes**

Genomic regions containing SNPs that showed significant associations with run timing were examined for candidate genes using alignments against the latest assembly of the Atlantic salmon genome sequencing project (Davidson et al. 2010), available at www.salmobase.org (Samy et al. 2014).
The flanking sequence of each SNP was compared to the genome using BLAST searches, as well as regions containing multiple significant SNPs.

**SNP validation**

In order to validate the SNPs identified as being significantly associated with return timing, a new set of fish was sampled and screened (Table 1). Primers for the SNPs (detailed in Supplementary data S2) were designed by Fluidigm (Fluidigm, San Francisco, CA, USA) and optimised for running on the Fluidigm EP1 system, with samples processed according to the manufacturer’s protocols. In order to standardise between the different genotyping platforms (Affymetrix Axiom array and Fluidigm EP1 System) a sample of 178 individuals was screened on both platforms. Any SNP with either a low correspondence between the two platforms (< 98%) or failing to amplify was removed from further analysis.

The validation procedure was based on adults (n = 425) caught in the rivers North Esk, Tweed and South Esk (a river not involved in the GWAS). The fish consisted of 213 two sea winter (2SW) salmon and, although not used in the GWAS, 212 one sea winter (1SW) fish (Table 1), which had been captured at or soon after returning to the rivers. Together with these fish, where timing of river entry was known reasonably well, a second set of twenty fish caught in the autumn in an upland tributary (Girnock) of the river Dee was also examined. River entry dates for these fish were unknown, but the tributary is known to produce a high proportion of early running fish (Buck and Hay 1984) (Table 1).

The fish were screened on the Fluidigm EP1 platform for the SNPs identified as being significantly associated with return timing. Individual assignment of fish to return timing group was then carried out in ONCOR (Kalinowski et al. 2007), with probabilities estimated using the method of Rannala and Mountain (1997). Fish from the GWAS were used as the reference baseline groups, split into a binary Early/Late return phenotype determined by their date of river return (early returning Feb/Mar and late returning Aug/Sept/Oct, see Table 1). The probability associated with each
individual being an early running fish was then analysed in R in relation to day of capture (Julian day), sea age, river of origin and their interactions, based on a generalised linear model assuming binomial errors. Model selection was undertaken using a backwards stepwise approach using likelihood ratio tests to assess the significance of changes to the model.

Results

Quality Control

All samples from the rivers North Esk, Oykel and Tweed passed the stages of the higher threshold level QC, whilst the samples from the rivers Spey and Tay passed the QC stages of the lower threshold levels; so all samples were available for analysis. Once the Spey and Tay data had been examined, 56,048 of the 202,706 markers were available for GWAS. Out of those, 3,318 SNPs had a MAF < 0.05 and were also removed, resulting in a final set of 52,731 SNPs.

Genome-wide SNP association

Model selection resulted in all fixed effect covariates being removed, resulting in a final model consisting of just the relationship matrix and locus effect. Nine loci were found to be significantly associated with return timing using a FDR of 0.05 (Table 2). Full GWAS results for all loci are detailed in Supplementary data S2. Although all fixed terms were removed from the final model, comparative results for models including either the PC, covariates or both were calculated, with the same nine loci being identified in all models (see QQ plots in Supplementary data S3). The differentiation of return phenotypes observed using the nine significant loci was visualised using principle component analysis (Fig. 2) with the first PC explaining 53% of the total variance for the loci significantly associated with run timing.

Genome mapping, alignment and candidate genes

Of the nine significant loci, all but one were located on chromosome Ssa09 (Table 2, Fig. 3). These eight were found to be within a 9109 kb region, within which two smaller regions of 11 kb and 945 kb contained two and six SNPs, respectively (Fig. 3A and 3C). BLAST searches of the salmon
genome encompassing these two regions revealed a single gene associated with smaller of the two regions and 24 associated with larger region (full details in Supplementary data S4). The putative functions of the loci identified include roles in regulating cell stress response pathways induced by environmental stress, regulating cellular and systemic homeostatic responses to hypoxia, metabolism of xenobiotics and/or other compounds, centrosome stabilisation, eye development, voltage-gated potassium channel function, and activation of cyclin-associated kinases.

The chromosome position of the remaining locus, the top ranked AX-87609372, was reported by CIGENE as unknown, but a BLAST search resulted in a top match to chromosome Ssa15. There were, however, a number of other matches with regions on different chromosomes (including chromosome Ssa09), suggesting a possible degree of genomic duplication and meaning definitive positioning is problematic (full details of all matches in Supplementary data S5). The map position of AX-87609372 was thus retained as unknown.

Validation

The nine significant SNPs from the GWAS all successfully transferred to the Fluidigm platform and so were all used in the validation analysis. For the validation fish, the probability of each individual belonging to the early running phenotype was found to be significantly related to Julian day ($\chi^2 = 67.88, df = 1, p = 2.2 \times 10^{-16}$) and sea age ($\chi^2 = 24.89, df = 1, p = 6.07 \times 10^{-7}$) with the significant differences also found between fish from the North and South Esk compared to those from the Tweed ($\chi^2 = 7.66, df = 2, p = 0.0056$). There were no significant interactions between these three variables (all $p > 0.1$). The probability decreased through the year, and, for a given day, was higher for 1SW fish compared to 2SW, although there was still considerable individual variation around the fitted relationship, as reflected by Nagelkerke’s $R^2$ value of 0.245 (Figure 4). All but one of the fish from the Girnock tributary of the river Dee assigned as early running, with a mean (± SD) assignment probability score of 0.95 ± 0.09.
Discussion

A GWAS approach identified nine SNPs associated with return timing in 2SW Atlantic salmon. Of these, eight were found to be in a 9109 kb region of chromosome Ssa09, within which two smaller regions of 11 kb and 945 kb contained two and six SNPs, respectively. Successful validation of the SNP markers was achieved in a novel group of fish including individuals from a river not used in the GWAS, together with 1SW fish. Approximately 24% of phenotypic variation in return timing was controlled by a single locus or a small group of closely linked loci located on chromosome Ssa09, and this genomic control appeared to be conserved across rivers and life-history phenotypes.

The region of chromosome Ssa09 associated with run timing identified here has also been found to be involved in age at maturity in Norwegian/Finnish Atlantic salmon (Barson et al. 2015). The region was significantly associated in a GWAS model that did not take underlying population structure into account. However, when this structure was included, significance was lost, suggesting potential between-population variation in a correlated trait. Barson et al. (2015) suggest that this trait may be size rather than age at maturity but here we found it to be associated with seasonal return timing, which is itself related to size (Bacon et al. 2009). Disentangling the various related migratory traits in detail would be very interesting and worthy of further investigation.

The identification of a small region of a single chromosome having a significant role in the genetic control of a complex ontogenetic trait has previously been reported in other salmonids. (Ayllon et al. 2015; Barson et al. 2015) both found that between 33-40% of the phenotypic variation in age of maturity of Atlantic salmon was explained by a single region of chromosome Ssa25, which contained four SNPs (including two missense mutations) covering 2386 bp and was linked to the *vgll3* gene (an adiposity regulator). Similarly, a small region of a single chromosome has been identified which controls return timing in natural populations of steelhead trout (*O. mykiss*) exhibiting distinct summer and winter run-timing life histories (Hess et al. 2016; Prince et al. 2017; Showalter 2017). A 46 kb chromosome region was identified containing three SNP loci that explained...
up to 46% of the trait variation. Alignment of the region to the Atlantic salmon genome showed localisation to a section of the salmon chromosome Ssa03 which overlapped with the *GREB1-like* gene (Hess et al. 2016).

Migration is a common but complex life history trait with an incredibly large number of different strategies, from the longest 8,500 km humpback whale (*Megaptera novaeangliae*) breeding migration to the resource migration of the slime mold (*Dictyostelium discoideum*) (Bowlin et al. 2010). Within salmonids, the various migration components are complex, with differing strategies based on size (Hutchings and Jones 1998), growth rates (Okland et al. 1993), age (Johnston et al. 2014), sex (Barson et al. 2015), season (Waples et al. 2004), duration (Gurney et al. 2012) and even whether to migrate at all (Hecht et al. 2012). It might be expected, therefore, that the genomic control of the various aspects of the migratory habit might themselves be complex. The identification of a single locus explaining a significant proportion of the total trait variation (Prince et al. 2017, this study) is therefore quite surprising, especially given that migration is the result of complex interactions between intrinsic (genetics, physiology, and behavior) and extrinsic factors (weather, habitat conditions, food availability, predation, topography) (Åkesson and Hedenström 2007). Fully understanding such findings would require examination of the selective and evolutionary mechanisms involved in the development of the genomic control of seasonal returning timing. Indeed, increasing our understanding of the genetic control of migration has been identified as one of the main challenges in enhancing our knowledge of migration across species (Bowlin et al. 2010) and salmonids could be ideal model species due to the increasing availability of genetic resources and the large variation in their migratory life history characteristics.

The percentage of trait variation explained in the validation fish is relatively high (24.5%), although apparently lower than some observations of large gene effects in other studies (e.g. Ayllon et al. 2015; Barson et al. 2015). This value, however, was calculated in a novel group of fish not included in the GWAS and, as such, not likely to suffer from inflated levels due to possible ascertainment bias, which may otherwise be seen. Further, such a level still represents a relatively
high effect size for what might be considered a complex trait, with the potential for mutations at many different genes being able to influence the phenotype. The levels of variation explained indicate that return timing is substantially controlled either by a single gene or complex of closely linked genes and differentiating between these two possibilities would require further fine-scale mapping.

The mechanisms controlling the remaining ~75% of variation shown in run timing are unknown. However, considering the complexity of the trait, run timing is likely to be influenced by many loci of varying effects, as has been observed in other migratory related salmonid gene mapping studies (e.g. Hecht et al. 2013). Furthermore, many heritable traits are also, to a greater or lesser extent, the product of both genetic and environmental effects (Garcia de Leaniz et al. 2007), with environmental influences shown to influence run timing in translocation experiments (Hansen and Jonsson 1991; Stewart et al. 2002). Again, investigating the full extent of the genomic architecture and genotype-by-environment interactions would require further analysis.

The various functions of the genes located in the region of chromosome Ssa09 associated with run timing included regulatory roles in stress and hypoxia responses and various metabolic and developmental functions. The position is different to that of the Clock genes (Paibomesai et al. 2010) and GREB1L (Prince et al. 2017), which have previously been linked to migratory associated traits. The region was, however, the same as that identified in Barson et al. (2015), linked to age at maturity and which included a transcription factor of the hypothalamus–pituitary–gonadal axis, SIX6, associated with size and age at maturity in humans (Perry et al. 2014) and a conserved non-coding element that aligns to a candidate distal forebrain enhancer of SIX6 (Lee et al. 2012). The actual genes involved in seasonal return timing were not identified here and further fine scale mapping would help in identifying those involved.

The SNPs identified as being associated with return timing may be linked either with the trait itself or could be the result of different selection regimes experienced by early and late returning fish at some period/s in their marine phase. Differences in potential selective agents have been seen in
fish differing in sea ages and were related to influences such as diet, migration routes, parasites and predation (Johnston et al. 2014). Fitness differences influenced by such agents may result in a change in allele frequencies within the returning individuals which was not present before migration. It is difficult to definitively distinguish between the two hypotheses, but the fact that, during the validation process, the SNP panel was found to be effective in distinguishing between the run timing groups in both 1SW fish and in adults caught in novel rivers suggest that true associations of marker/traits has been identified. Notwithstanding these observations, further analysis to definitively distinguish between the two hypotheses is required, as selective pressures will vary between and within rivers, and 1SW and 2SW fish may migrate to different areas of the marine environment (Menzies and Shearer 1957; Thorstad et al. 2011).

The initial genotyping phase using the Affymetrix Axiom array resulted in a ~ 75% loss of available SNPs. This was solely due to poor DNA quality obtained from the scale samples used to represent fish from the rivers Tay and Spey and so no bias might be expected in the remaining SNPs. The samples represented ~38% of the total number of fish available for the GWAS and increased the geographic coverage from three to five rivers. It was therefore decided to maximise power using the ~50K remaining SNPs across multiple rivers rather than using all SNPs for a reduced number of sites and fish. The remaining SNPs provided sufficient power to identify a small chromosomal region with a significant effect on seasonal return timing and the available mapping information allowed this to be positioned on chromosome Ssa09. It should be noted, however, that the power of any GWAS analysis is limited by the interaction between numbers of individuals, numbers of markers and strength of the loci effects influencing the trait of interest. Larger numbers of individuals and/or markers may therefore identify other regions of smaller effect not observed here.

Identification of SNP loci associated with run timing and validation of these SNPs in fish of different sea ages and from both original and new rivers has important implications for conservation and fisheries management (Hess et al. 2016). In order to maintain the wide diversity observed in salmon fisheries, it is important to maintain the diversity of the populations that underpin them (e.g.
Hilborn et al. 2003; Potter et al. 2003; Schindler et al. 2010). However, achieving this management aim is only possible if the biological structure of the resource is known. The multiple run timing groups present in some rivers produce an elongated fishing season and are known to show different trends in abundance, and may therefore have different conservation needs (Anon 2016; Quinn et al. 2016; Youngson et al. 2003). Tracking studies have been used to determine where salmon returning at different times of the year spawn (Laughton and Smith 1992; Webb and Campell 2000) but such information is spatially limited. It may be more practical to obtain greater spatial coverage for such information by using the genetic markers discovered here to characterise salmon populations and producing run timing maps both within and between rivers. At present salmon are often managed at the river level, due to the difficulties in defining and accounting for within-river structuring (Potter et al. 2003). The ability to characterise areas of the catchment associated with differences in run timing characteristics will allow sub-catchment management of specific run timing stocks.

Previous translocation experiments have shown a heritable component to run timing of adult salmon in Scotland (Stewart et al. 2002) and here we identified a small panel of markers linked to a single region of the genome which significantly influences the trait. The ability to identify and quantify the different run timing components with this panel allows a prerequisite to maintaining the diversity of run timing phenotypes seen in rivers, which is important for both economic and ecological reasons. In addition, the results enhance our understanding of the genomic architecture of important life history traits in salmonids and so, again, help in our understanding of the complex life history variations this species exhibits.

Acknowledgements

We are grateful to S. Karoliussen and M. Kent from the Centre for Integrative Genetics (Cigene) for screening the samples and providing early access to the chromosome position. The manuscript was considerably improved by the helpful comments of anonymous reviewers who we would like to thank. We would also like to express our thanks to Marine Scotland Science staff at Montrose and
Pitlochry together with personnel from the Kyle of Sutherland Fisheries Trust and the Tweed Foundation for their help in sample collection, and to J. Henry and D. Stewart and staff at the Tweed Foundation for reading the scales.

References


Laughton, R., and Smith, G.W. 1992. The relationship between date of river entry and the estimated spawning position of adult Atlantic salmon (Salmo salar) in two major Scottish east coast rivers. In


https://mc06.manuscriptcentral.com/cjfas-pubs


Shearer, W.M. 1990. The Atlantic salmon (Salmo salar L.) of the North Esk with particular reference to the relationship between river and sea age and time of return to home waters. Fish Res 10: 93-123.

Showalter, B. 2017. Measuring associations between polymorphisms in the GREB1-L gene and the development of different migratory phenotypes in Oncorhynchus mykiss, Department of Biology, Texas Christian University.


Table 1. Details of samples used in the genome wide association study (GWAS) and validation process. All fish used in the GWAS were two sea winter (2SW). The validation analysis was carried out on using one sea winter (1SW) and 2SW fish. Month refers to month of capture and source of both DNA (Tissue) and fish are described.

<table>
<thead>
<tr>
<th>River</th>
<th>Month</th>
<th>Total</th>
<th>Tissue</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>North Esk</td>
<td>16</td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Oykel</td>
<td>8</td>
<td>8</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Spey</td>
<td>15</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Tay</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tweed</td>
<td>5</td>
<td>12</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>All Total</td>
<td>35</td>
<td>38</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Validation**

<table>
<thead>
<tr>
<th>1SW</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>North Esk</td>
<td>2</td>
<td>23</td>
<td>35</td>
<td>30</td>
<td>43</td>
<td>133</td>
<td>muscle</td>
<td>Net caught from river estuary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Esk</td>
<td>36</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td>46</td>
<td>fin clip</td>
<td>Net caught from estuarine or in-river nets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tweed</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>2</td>
<td></td>
<td>33</td>
<td>fin clip</td>
<td>Net caught from estuarine or in-river nets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1SW Total</td>
<td>2</td>
<td>23</td>
<td>41</td>
<td>30</td>
<td>89</td>
<td>25</td>
<td>2</td>
<td></td>
<td>212</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2SW</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>North Esk</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>15</td>
<td>20</td>
<td>88</td>
<td>muscle</td>
<td>Net caught from river estuary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Esk</td>
<td>1</td>
<td>14</td>
<td>32</td>
<td>15</td>
<td>1</td>
<td>15</td>
<td>9</td>
<td>fin clip</td>
<td>Net caught from estuarine or in-river nets</td>
<td></td>
</tr>
<tr>
<td>Tweed</td>
<td>17</td>
<td>1</td>
<td>8</td>
<td>11</td>
<td>1</td>
<td>38</td>
<td>fin clip</td>
<td>Net caught from estuarine or in-river nets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dee</td>
<td>11</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td>20</td>
<td>fin clip</td>
<td>Trap caught in Girnock Burn(^1) in upper river</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2SW Total</td>
<td>1</td>
<td>14</td>
<td>49</td>
<td>33</td>
<td>35</td>
<td>17</td>
<td>43</td>
<td>31</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>All Total</td>
<td>1</td>
<td>14</td>
<td>51</td>
<td>56</td>
<td>76</td>
<td>47</td>
<td>132</td>
<td>56</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

\(^1\) For description see Buck and Hay (1984).
Table 2. SNP loci showing significant association with run timing using GAPIT modelling of Julian return day.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>P-value</th>
<th>Sample number</th>
<th>R-square of Model without SNP</th>
<th>R-square of Model with SNP</th>
<th>FDR adjusted P-value</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>AX-87609372</td>
<td>-</td>
<td>-</td>
<td>7.07E-09</td>
<td>122</td>
<td>0.19583</td>
<td>0.453992</td>
<td>0.000373</td>
<td>1</td>
</tr>
<tr>
<td>AX-87209310</td>
<td>9</td>
<td>25749686</td>
<td>1.65E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.401358</td>
<td>0.004066</td>
<td>2</td>
</tr>
<tr>
<td>AX-87233345</td>
<td>9</td>
<td>25214157</td>
<td>2.31E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.395913</td>
<td>0.004066</td>
<td>3</td>
</tr>
<tr>
<td>AX-87714789</td>
<td>9</td>
<td>25739448</td>
<td>3.73E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.388255</td>
<td>0.004285</td>
<td>4</td>
</tr>
<tr>
<td>AX-87323325</td>
<td>9</td>
<td>16641095</td>
<td>4.8E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.384218</td>
<td>0.004285</td>
<td>5</td>
</tr>
<tr>
<td>AX-87134825</td>
<td>9</td>
<td>25213625</td>
<td>4.89E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.383909</td>
<td>0.004285</td>
<td>6</td>
</tr>
<tr>
<td>AX-87832982</td>
<td>9</td>
<td>24803954</td>
<td>5.69E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.38152</td>
<td>0.004285</td>
<td>7</td>
</tr>
<tr>
<td>AX-87027895</td>
<td>9</td>
<td>16651966</td>
<td>8.66E-07</td>
<td>122</td>
<td>0.19583</td>
<td>0.374879</td>
<td>0.005708</td>
<td>8</td>
</tr>
<tr>
<td>AX-87748269</td>
<td>9</td>
<td>25213391</td>
<td>1.58E-06</td>
<td>122</td>
<td>0.19583</td>
<td>0.365437</td>
<td>0.009277</td>
<td>9</td>
</tr>
</tbody>
</table>
Fig. 1. Map of Scotland with major salmon rivers denoted and locations of rivers from which samples were used. Points represent river mouths. © Crown copyright and database right 2017. All rights reserved. Ordnance Survey Licence number 100024655. Scottish Salmon Rivers © Marine Scotland 2017.
Fig. 2. Plots of the first two principle components calculated using SNPs found to be significantly associated with run timing in the GWAS. Black points represent early returning (Feb/Mar) and grey late returning (Aug/Sept/Oct) fish.
Fig. 3. Combining significance levels of SNPs for return timing GWAS with available map positions (see text). A) Manhattan plot of $-\log_{10}(p$-value) for Atlantic salmon genome position of the SNP markers. B) QQ plot showing expected versus observed $-\log_{10}(p$-values). C) Manhattan plot of $-\log_{10}(p$-value) focusing on SNPs positioned on chromosome Ssa09. Horizontal lines in the Manhattan plots represent the 0.05 significance threshold after FDR $p$-value correction.
Fig. 4. Relationship between the probability of assigning as early running fish as defined by individual ONCOR assignments and day of the year for both one sea-winter and multi sea-winter validation fish using the nine SNPs found to be associated with return timing for A: North and South Esk fish and B: Tweed fish. Points represent mean probability of assigning as early running fish at a particular 10 day Julian day bin and the lines the fitted relationships.