GENOMIC SIGNATURES OF FINE-SCALE LOCAL ADAPTATION IN ATLANTIC SALMON SUGGEST INVOLVEMENT OF SEXUAL MATURATION, ENERGY HOMEOSTASIS, BEHAVIOUR, AND IMMUNE DEFENCE-RELATED GENES.

Victoria L. Pritchard^{1†}, Hannu Mäkinen^{1,2}, Juha-Pekka Vähä^{3*}, Jaakko Erkinaro⁴, Panu Orell⁴ & Craig R. Primmer^{1,2,5}.

¹ Department of Biology, University of Turku, FI-20014, Finland.

² Department of Biosciences, University of Helsinki, FI-00014, Finland.

³ Kevo Subarctic Research Institute, University of Turku, FI-20014, Finland.

⁴ Natural Resources Institute Finland (Luke), Oulu, FI-90014, Finland.

⁵ Institute of Biotechnology, University of Helsinki, FI-00014, Finland.

† Corresponding author, email: victorialpritchard@gmail.com

*Present address: Association for Water and Environment of Western Uusimaa, POB 51, FI-08101 Lohja, Finland.

Running title: Local Adaptation in Atlantic Salmon

ABSTRACT

Elucidating the genetic basis of adaptation to the local environment can improve our understanding of how the diversity of life has evolved. In this study we used a dense SNP array to identify candidate loci underlying fine-scale local adaptation within a large Atlantic salmon (*Salmo salar*) population. By combining outlier, gene-environment association, and haplotype homozygosity analyses, we identified multiple regions of the genome with strong evidence for diversifying selection. Several of these candidate regions had previously been identified in other studies, demonstrating that the same loci be adaptively important in Atlantic salmon at sub-drainage, regional and continental scales. Notably, we identified signals consistent with local selection around genes associated with variation in sexual maturation, energy homeostasis, behaviour, and immune defence. These included the large-effect age-at-maturity gene *vgl3*, the known obesity gene *mc4r*, and major histocompatibility complex II. Most strikingly, we confirmed a genomic region on Ssa09 that was extremely differentiated among subpopulations, and that is also a candidate for local selection over the global range of Atlantic salmon. This region co-localized with a genomic region strongly associated with spawning ecotype in sockeye salmon (*Oncorhynchus nerka*), with circumstantial evidence that the same gene (*six6*) may be the selective target in both cases. The phenotypic effect of this region in Atlantic salmon remains cryptic, although allelic variation is related to river flow volume and co-varies with timing of the return spawning migration. Our results further inform management of Atlantic salmon and open multiple avenues for future research.

Keywords: Atlantic salmon; sockeye salmon; local adaptation; ecotype.

INTRODUCTION

Understanding how the diversity of life on earth has evolved is one of the central questions in biology. Fundamental to this is elucidating how change at the genomic level underlies phenotypic change as populations adapt to their local environment, and how this change ultimately contributes to evolutionary radiations. Central questions include the relative roles of polygenic variation vs. genes of large effect (Yeaman & Whitlock, 2011), the importance of protein coding vs. gene expression variation (Fraser, 2013), the contribution of chromosome rearrangements (Kirkpatrick & Barton, 2006) and whether parallel, independently evolved adaptations have the same genetic basis across populations (Conte, Arnegard, Peichel, & Schluter, 2012). Resolving such questions contributes to the ultimate aim of linking genotype, phenotype and fitness (Barrett et al. 2011). One limitation until recently has been the lack of dense marker panels and annotated genomes for non-model species, which impedes identification of the loci driving statistical signals of selection (Vilas, Pérez-Figueroa, and Caballero 2012).

Salmonid fishes, which include charr, salmon, trout and whitefish, are splendid models for the study of local adaptation and diversification. Strong philopatry in breeding location and/or separation into different water bodies means that fine-scale genetic differentiation is a characteristic of salmonid populations (Fraser, Weir, Bernatchez, Hansen, & Taylor, 2011). Both allopatric and sympatric populations can express a wide range of phenotypic diversity, including variation in age at maturity, migratory strategy, breeding time and location, and exploitation of different ecological niches (e.g. Bernatchez et al., 2010; Dodson, Aubin-Horth, Thériault, & Páez, 2013; Jonsson & Jonsson, 2001; Quinn, McGinnity, Reed, & Bradford, 2016). Evolutionary diversification may have been facilitated by the whole genome duplication that occurred at the base of the salmonid radiation (Lien et al., 2016). Much of this diversity is heritable (Carlson & Seamons, 2008), and multiple studies have found reduced fitness of salmonids in non-local environments (Fraser et al., 2011; O'Toole et al., 2015). While many fitness-associated traits in salmonids have been well characterized at the quantitative genetic level, most of the underlying loci are yet to be identified (Fraser, Weir, Bernatchez, Hansen, & Taylor, 2011; Garcia de Leaniz et al., 2007).

New genomic tools are enabling us to unpick the molecular genetic basis of local adaptation in salmonids (Elmer, 2016). This can both improve our understanding of evolutionary diversification in general and help guide management of these ecologically and culturally important taxa (e.g. Prince et al., 2017). Recent studies have identified suites of loci potentially involved with adaptation to the natal or migratory environment (e.g. Micheletti, Matala, Matala, & Narum, 2017; Moore et al., 2017), associated with ecotypic differentiation (e.g. Veale & Russello, 2017a), or responding to selection (e.g. Liu et al. 2017). One finding of such studies is that life history diversification, in particular, can be underlain by small genomic changes of large effect. For example, Atlantic salmon (*Salmo salar*) vary in their age at maturity, with some individuals returning to freshwater to spawn after one year feeding at sea ('one-seawinter fish'), and others spending longer in the ocean and returning at a much larger size ('multi-seawinter fish'). Barson et al. (2015) showed that a single locus containing the gene *vgll3* explained nearly 40% of this variation in age at maturity across three evolutionary lineages in Europe (see also Ayllon et al., 2015). This locus appears also to be associated with sea age at maturity in North American Atlantic salmon (Kusche et al., 2017), and is additionally located within a QTL underlying early ('precocious') freshwater maturation (Lepais, Manicki, Glise, Buoro, & Bardonnnet, 2017). The choice between residency and anadromy in many populations of rainbow trout/steelhead (*Oncorhynchus mykiss*) is associated with a genomic region on chromosome Omy05 (Hecht, Campbell, Holecek, & Narum, 2013; Hecht, Thrower, Hale, Miller, & Nichols, 2012; Martínez, Garza, & Pearse, 2011; Nichols, Edo, Wheeler, & Thorgaard, 2008; Pearse, Miller, Abadía-Cardoso, & Garza, 2014). Sockeye salmon (*O. nerka*) exhibit ecotypic variation in spawning site, with both anadromous and resident forms spawning in streams(/rivers) or lake shores(/benthos): different spawning ecotypes can co-occur within the same populations (Larson et al., 2017; Nichols, Kozfkay, & Narum, 2016;

Veale & Russello, 2017b). Veale & Russello (2017a) showed that this ecotypic differentiation is associated with two strongly diverged haplotypes encompassing the *lrr9* gene. This relationship held throughout the species' range: in some populations, it was almost entirely Mendelian, with one homozygote determining shore spawning. Finally, in one case, independently arising mutational variation at the same locus underlies an ecologically important life history trait in two species: variation in the *greb11* gene is associated with premature spawning migration in both steelhead and chinook salmon (*O. tshawytscha*) (Hess, Zendt, Matala, & Narum, 2016; Prince et al., 2017)

One of the world's largest surviving wild Atlantic salmon stocks reproduces in the Teno River (Norwegian: Tana; Sámi: Deatnu, Fig. 1) of subarctic Finland and Norway. The Teno drains a catchment of 16,386km² and includes more than 30 different-sized tributaries accessible to adult salmon: up to 100,000 individuals return to spawn each year. Previous studies have demonstrated significant, temporally stable, population genetic substructure within the Teno stock corresponding to different spawning locations (Vähä, Erkinaro, Falkegård, Orell, & Niemelä, 2017; Vähä, Erkinaro, Niemelä, & Primmer, 2008; global F_{st} from microsatellites ≈ 0.065). This genetic structure is higher than that observed across multiple rivers in other parts of the Atlantic salmon range (Vähä, Erkinaro, Falkegård, Orell, & Niemelä, 2017). Thus, there is strong potential for differential local adaptation within the system. Teno salmon subpopulations vary in the timing of their re-entry to freshwater to spawn ('run timing'; June-August; Vähä et al., 2011), and the proportion of multi-seawinter adults among spawners (Vähä, Erkinaro, Niemelä, & Primmer, 2007). Recently, Aykanat et al. (2015) demonstrated that adult salmon captured in the Teno mainstem belong to two genetically distinct subpopulations that differ in their freshwater and marine growth rate. Otherwise, no detailed studies of phenotypic differentiation among subpopulations have been performed.

In this study, we used a dense marker panel (198,829 SNPs) to examine genomic variation among Teno Atlantic salmon subpopulations, with the aim of identifying regions of the genome that potentially contribute to fine-scale local adaptation. To examine whether these candidate regions might also be under differential selection at broader geographic scales, we compared our results to previous studies of Atlantic salmon throughout their global range. We also examined our candidate regions for any of the large-effect genes implicated in life history diversification in other salmonid species.

MATERIALS AND METHODS

Sample Collection, Preparation and Genotyping

To ensure that samples represented the local breeding subpopulation, we genotyped juvenile salmon collected before their ocean migration (age 0 and 1). Sample collection and DNA extraction is described in Vähä et al. (2017). To minimize sampling of siblings, electrofishing sites were separated by 50-100m and only a single individual of each age was collected at each site. We selected 10 locations in the Teno known to harbour genetically distinct subpopulations (Aykanat et al., 2015; Vähä et al., 2017, 2008) and which represent a range of environmental variation (Vähä et al., 2007): Teno mainstem: Garnjarga (TEMS, n=24), Teno mainstem: Outakoski (OUTA, n=24), Inarijoki (INAR, n=21), lower Iešjohka (IESJ, n=21), Maskejohka (MASK, n=20), Kárásjohka (KARA, n=22), Tsarsjoki/Carsejohka (TSARS, n=21), Utsjoki/Ohcejohka (UTSJ, n=21), Kevojoki/Geavvujohka (KEVO, n=21), Yla Pulmankijarvi/Buolbmátjohka (PULM, n=20) (Fig. 1, Table1). We assessed concentration and degradation of the previously-extracted DNA using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc.) and by agarose gel visualization, and performed re-extractions where necessary. DNA of sufficient quality was standardized to a Nanodrop-estimated concentration of 15ng/ul and sent to the Center for Integrative Genetics (CIGENE), Ås, Norway for genotyping.

Samples were genotyped for 220,000 SNPs using a custom Affymetrix Axiom SNP array on a GeneTitan genotyping platform, as in Barson et al. (2015). SNPs on this 220K array have known locations on the NCBI RefSeq Atlantic salmon genome (Lien et al., 2016, http://www.ncbi.nlm.nih.gov/genome/annotation_euk/Salmo_salar/100/). To ensure correct identification of genotype clusters, we applied the Affymetrix Best Practices Protocol for SNP calling simultaneously to our samples and ≈ 1800 individuals genotyped in previous studies (Barson et al., 2015; Pritchard et al., 2016).

For quality control and analyses we also used two pre-existing 220K SNP datasets: a dataset of known aquaculture escapees ($n=192$, Pritchard et al., 2016), and a dataset of adults collected from the Teno mainstem ($n=463$) and the neighbouring river Borselva ($n=17$) (Barson et al., 2015). The adult Teno mainstem sample included the individuals genotyped with a 7K SNP array by Johnston et al. (2014) and Aykanat et al. (2015).

Genotyping quality control.

Two hundred and two out of 215 samples passed Affymetrix quality thresholds. Subsequent quality control was performed using PLINK v.1.90 (Chang et al., 2015; Purcell et al., 2007). First, we removed 1,112 SNPs not mapped to an assembled *S. salar* chromosome and 35 SNPs with known off-target variants. We then used the dataset of adults collected from the Teno mainstem to identify SNPs that deviated from Hardy-Weinberg equilibrium at $p < 0.001$ in either of the two subpopulations of Aykanat et al. (2015; Pop1: $n=303$; Pop2: $n=121$). We found 2,750 such SNPs; we considered these to have technical genotyping problems and removed them. To identify full sibs within our juvenile samples, we split them by subpopulation, excluded SNPs with $>2\%$ missing data or $MAF < 0.05$ within each subpopulation, performed a linkage disequilibrium (LD) pruning step (PLINK command: `--indep 50 5 1.4`), used the `--genome` function to estimate genome-wide identity-by-descent between each pair of individuals, and graphically explored the results using ggplot2 in R 3.1.2 (R Core Team, 2015; Wickham, 2009). To identify juveniles with possible domestic parentage we added the dataset of aquaculture escapees, performed SNP filtering and LD pruning as described above, applied a two-dimensional multidimensional scaling analysis (MDS) to the genome-wide identity-by-state (IBS) matrix in PLINK, and graphically visualized the results. Following removal of 13 putative sibs and one individual with possible escapee ancestry, we excluded all SNPs with a $MAF < 0.05$ or $>10\%$ missing genotypes. The final dataset included 198,829 SNPs and 188 individuals (TEMS, $n=24$; OUTA, $n=22$; PULM, $n=13$; IESJ $n=21$; INAR $n=20$; KARA $n=19$; TSAR $n=21$; UTSJ $n=21$; KEVO $n=19$; MASK $n=8$). Proportion of missing genotypes per individual ranged from 0.07% to 3.24% (median 0.20%).

Population genomic characteristics

We calculated subpopulation H_e from genome-wide minor allele frequencies returned by PLINK. We also used PLINK to estimate unbiased pairwise F_{st} (Weir & Cockerham, 1984). We explored subpopulation structure among the juveniles by performing an MDS analysis on the LD-pruned final dataset, as described above. We investigated how juvenile samples related to the adult samples from the Teno mainstem by combining the two datasets and performing the same analysis. We assigned adults to the two mainstem subpopulations of Aykanat et al. (2015) using the Structure results from that study (available: <https://doi.org/10.5061/dryad.7t4n0>), with a cut-off of $q=0.9$ for subpopulation membership.

Environmental variation

We investigated four broad-scale environmental characteristics previously shown to influence population structure in the Teno (Vähä, Erkinaro, Niemelä, & Primmer, 2007; Table 1). We estimated waterway distance of the sampling site from the Teno mouth using the package ‘riverdist’ in R (<https://cran.r-project.org/web/packages/riverdist/index.html>). We obtained altitude using Google maps. We estimated

upstream catchment area for each site (a surrogate for flow volume) by summing sub-basin catchment areas obtained from the Finnish Center for Economic Development, Transport and the Environment. We based presence of downstream barriers on data for nearby sites in Vähä, Erkinaro, Niemelä, & Primmer (2007). As these four variables are non-orthologous, we converted them into principle components: standardized values were provided to the function *prcomp* in R with default parameters.

Phenotypic variation

To explore the relationship between these environmental variables and known phenotypic variation within the Teno, we extracted median run time and proportion of multi-seawinter females for each sampled subpopulation from Vähä et al. (2011) (Table 1) and examined their relationship to the principal components using a Kendall rank correlation test in R.

Identification of candidate genome regions responding to local selection

Best practice requires a combination of approaches to identify candidate genome regions responding to local selection (De Villemereuil, Fricot, Bazin, François, & Gaggiotti, 2014). We therefore a) identified markers unusually highly differentiated among subpopulations, based on F_{st} or equivalent statistics; b) examined association of allele frequencies with environmental parameters; and c) examined patterns of haplotype homozygosity within subpopulations, indicative of selective sweeps on particular alleles. We used PLINK or PGD Spider (Lischer & Excoffier, 2012) to convert among input files. We used Beagle 4.1 (Browning & Browning, 2007) with the complete juvenile dataset to impute missing genotypes or infer phasing where necessary. To assess whether several candidate SNPs could be labelling a single selected locus, we identified sets of neighbouring SNPs in high LD ('haploblocks') using *--blocks* in PLINK. Because default parameters in PLINK often returned multiple small haploblocks that appeared to break up a single selective sweep, we used relaxed parameters (*--no-small-max-span --blocks-inform-frac 0.8 --blocks-max-keb 5000 --blocks-strong-lowci 0.55 --blocks-strong-highci 0.85 --blocks-recomb-highci 0.8*). We defined haploblock bounds as the positions halfway between the outermost haploblock SNPs and their closest non-haploblock SNPs. We condensed any abutting haploblocks containing candidate SNPs and within 10kB of each other into a single block.

Outlier approaches: we used three approaches to identify markers with high among-population allelic variation. First, we estimated among-population F_{st} for each marker using OutFank (Whitlock & Lotterhos, 2015). Second, we used Bayescan (Foll & Gaggiotti, 2008), which assumes an island model of migration, with N_e and migration rate allowed to vary among subpopulations; diversifying selection is indicated by positive values of alpha, the locus-specific component of F_{st} . We specified *--prior odds 100*, kept other parameters default, and assessed convergence over three replicate runs. Third, we used BayEnv2 (Coop, Witonsky, Di Rienzo, & Pritchard, 2010; Gunther & Coop, 2013), which accounts for population structure and unequal sampling by estimating a variance-covariance matrix of allele frequencies across populations, and returns the X^TX statistic, which is a measure of the deviation of each locus from the underlying matrix and hence the likelihood that the locus is under diversifying selection. We estimated the population covariance matrix from an LD pruned subset of 33,133 SNPs (PLINK *--indep 50 5 1.4*), using 200,000 iterations. We supplied this matrix in three replicate runs of BayEnv2 for each SNP (100,000 iterations), and took the median of the three X^TX scores.

Association with environmental variables: Two principle components explained 82% of the total environmental variation and were included in the analysis. We used two approaches to identify markers associated with PC1 or PC2. BayEnv2 investigates allele frequency-environment associations accounting for population structure and sampling variance as described above. We used the previous subpopulation variance-covariance matrix, performed 200,000 iterations, and ran the analysis five times. As we observed little concordance between Bayes Factor rankings of SNPs among replicate runs, but much better concordance

between rankings based on Spearman's P , we used the latter as our informative measure. Such Bayes Factor discordance among replicate runs been noted elsewhere, and may not be solved by further increasing run length (Blair, Granka, & Feldman, 2014). We used the median of the absolute values of P over the five runs as our test statistic. Latent factor mixed models (LFMM, Frichot, Schoville, Bouchard, & François, 2013) investigate allele frequency-environment associations using mixed models in which latent variables account for population structure. For the LFMM analyses we imputed missing genotypes, specified 10 latent factors, performed 10,000 iterations with 5,000 burn-in, ran the analysis five times, and took the median z-score. We confirmed that 10 was a suitable number of factors using PCA with a Tracy-Widom test in the R package LEA (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014).

Haplotype homozygosity: We used two approaches to examine subpopulations for elevated haplotype homozygosity, reflecting a selective sweep on an allele within that haplotype. The cross-population extended haplotype homozygosity (EHH) test (XP-EHH, Sabeti et al., 2007) compares EHH at the same site between two populations and thus can be used to identify selective sweeps that have occurred in one population but not in the other. We estimated normalized XP-EHH for each SNP using **selscan** (Szpiech & Hernandez, 2014; *-max-gap* 2Mb, all other parameters default). We selected the Teno mainstem (TEMS) subpopulation as a reference and compared each of the 9 other subpopulations; thus, negative XP-EHH values indicate higher haplotype homozygosity in the TEMS subpopulation and positive values vice-versa. We normalized XP-EHH across all chromosomes within each pairwise comparison. To obtain an overall score for each SNP we took the maximum absolute normalized score over all comparisons.

HapFLK (Fariello, Boitard, Naya, SanCristobal, & Servin, 2013) organizes markers into haplotype clusters using a multipoint linkage disequilibrium model and then measures haplotype frequency differentiation between populations, accounting for population structure using a population tree. For the HapFLK analysis, we included Borselva as the outgroup and inferred the population tree from the same 33,133 SNPs used to estimate the variance-covariance matrix for BayEnv2. We ran HapFLK for each chromosome separately, specifying 10 or 15 haplotype clusters (K) and 10 EM runs (*nfit*). The same outlying regions were identified independent of K , and we retained the HapFLK scores for each SNP from the $K=10$ analysis.

Significance testing and combined evidence: The significance of observed outliers or environmental associations in genome scans is frequently assessed by comparing observed scores to the expected distribution of scores in the absence of selection. However, modelling this neutral distribution requires assumptions about the demographic history of the populations under study that are rarely met. Correspondingly, simulation studies have demonstrated varying levels of Type I and Type II error depending on the true underlying scenario and the sampling scheme (Lotterhos & Whitlock, 2014; Narum & Hess, 2011; De Villemereuil et al., 2014). In a similar vein, the statistical properties of test scores used to investigate haplotype homozygosity are not well characterized (Vatsiou, Bazin, & Gaggiotti, 2015). We did not have a way to a-priori identify a set of selectively neutral SNPs (Lotterhos & Whitlock, 2014) from which we could obtain an expected distribution of scores. Further, neutral or non-locally-adaptive population genetic processes such as allele surfing and purifying selection can generate high test scores, meaning that apparently 'significant' loci still require validation by further studies (Charlesworth, Nordborg, & Charlesworth, 1997; Edmonds, Lillie, & Cavalli-Sforza, 2004). Given these considerations, we selected our loci of interest based on the empirical distribution of test scores. For each test, we ranked SNPs by test score and retained the top-ranked 0.5% (equivalent to empirical $p < 0.005$ considering all 198,829 SNPs). We combined evidence over these nine sets of SNPs to identify 'candidate SNPs' as follows:

- i) SNPs in the top 0.5% set in both the LFMM and BayEnv2 analyses were considered environmentally associated SNPs ('PC1' or 'PC2' SNPs).

- ii) SNPs in the top set in either the LFMM or the BayEnv2 analysis, and also in the top set in at least two of three outlier analyses were also considered environmentally associated SNPs ('PC1' or 'PC2' SNPs).
- iii) SNPs in the top set in at least two of the three outlier analyses, but not in the top set of either environmental analysis, were considered outlier SNPs ('OUT' SNPs).
- iv) SNPs in the top set in both the XP-EHH and HapFLK analysis, but not in other analyses, were considered extended haplotype homozygosity SNPs ('EHH' SNPs).
- v) We identified a subset of haploblocks that were particularly strong candidates to contain loci under diversifying selection by combining results from the outlier/ environmental analyses with results from the XP-EHH and HapFLK analyses. 'Candidate haploblocks' contained at least one PC1, PC2 or OUT SNP and at least one SNP in the top 0.5% set of either the XP-EHH or HapFLK analysis.

Additional analyses: on the basis of our results, we performed an additional, exploratory, LFMM analysis using median run time as the dependent variable, following the method described above.

Annotation of candidate regions

We found genes associated with candidate SNPs and haploblocks, using information from NCBI *Salmo salar* Annotation Release 100 provided in the R package Ssa.RefSeq.db (Grammes, 2016). We excluded non-coding RNAs and pseudogenes from the annotation, and converted gene, SNP and haplotype positions into bed format. To annotate candidate SNPs (PC1, PC2, OUT or EHH), we first used the bedtools function *intersect* to find overlapping genes (Quinlan & Hall, 2010). For SNPs in intergenic regions, we then used the function *closest* to find the nearest downstream gene, assuming that a selected variant could be in a 5' regulatory region. We also identified the intersecting or closest downstream gene for each of the 198,829 SNPs to create the background gene set for a GO term enrichment analysis (see below). For candidate haploblocks, we annotated overlapping and closest downstream genes.

GO term enrichment analysis

We obtained gene ontology (GO) terms for each annotated *S. salar* gene from Ssa.RefSeq.db, and performed GO term enrichment analyses using the R package topGO 2.28.0 (Alexa & Rahnenfuhrer, 2010). We assessed enrichment compared to the background described above using the Fisher exact test with the *neigh01* algorithm. We performed four analyses: i) genes associated with all candidate SNPs; ii) genes associated with PC1 SNPs; iii) genes associated with PC2 SNPs; iv) genes associated with OUT SNPs. Where a gene was associated with both a PC1/PC2 and an OUT SNP, we included it in the PC1 or PC2 group only. We did not apply a multiple-testing correction as the level at which this should be performed for GO term enrichment analyses is unclear (Alexa & Rahnenfuhrer, 2010); instead we examined top-ranking GO terms in each test.

Comparison of outliers to previous results

Several previous studies have used a 7K SNP array to identify regions of the genome putatively under differential selection among *S. salar* populations (Bourret, Kent, et al., 2013; Bourret, Dionne, Kent, Lien, & Bernatchez, 2013; Gutierrez, Yáñez, & Davidson, 2016; Jeffery et al., 2017; Liu et al., 2017; Mäkinen, Vasemägi, McGinnity, Cross, & Primmer, 2015; Moore et al., 2014; Perrier, Bourret, Kent, & Bernatchez, 2013); discriminating populations (Karlsson, Moen, Lien, Glover, & Hindar, 2011) or associated with phenotypic traits (Gutierrez, Yáñez, Fukui, Swift, & Davidson, 2015; Johnston et al., 2014). To compare our results with the location of SNPs of interest identified in these studies, we obtained flanking sequences for the 7K SNPs from NCBI dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/) and aligned them with the *S. salar* genome using bwa-mem with default

parameters (Li, 2013; Li & Durbin, 2009). We only retained unambiguously mapped sequences with $MQ \geq 60$. We considered a SNP to map to one of our candidate regions if it was within the haploblock boundary.

To further explore one strong candidate region for local selection, we aligned two *O. nerka* RAD-tag sequences to the annotated *S. salar* and/or *O. mykiss* genomes using bwa-mem (Omyk_1.0 available: https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1). RAD-tag 68810/57884 was associated with spawning ecotype by Veale & Russello (2017b) and Nichols, Kozfkay, & Narum (2016), and maps to the *hrr9* gene (Veale & Russello 2017a). RAD-tag 24343/64477/41305 was associated with spawning ecotype by Veale & Russello (2017b), Nichols, Kozfkay, & Narum (2016), and Larson et al. (2017), but not previously annotated. Finally, we used bwa-mem to align the zebrafish variant of the *six6* retinal enhancer from Conte et al. (2010) to the *S. salar* genome.

RESULTS

Environmental variation and population structure

PC1 explained 58% of the variation observed within the environmental data and was related to the correlated variables distance (loading: -0.58), elevation (-0.63) and accessibility (-0.47); thus we considered it to broadly represent adult migration distance. PC2 explained a further 24% of variation and was strongly related to upstream catchment area (loading: 0.96); thus we considered it to represent river flow volume (Table S1). Position of sampling sites along the two environmental PCs is shown in Fig. S1. Median runtime was positively related to PC2 (Fig. S2).

Pairwise F_{st} between samples is shown in Table S2; global F_{st} was 0.067, similar to previous microsatellite-based analyses (Vähä et al. 2017). We observed slightly lower genome-wide heterozygosity in the tributary subpopulations (H_e , KEVO: 0.338; MASK: 0.349; PULM: 0.320; TSAR: 0.294; UTSJ: 0.343) than in the mainstem and headwater subpopulations (IESJ: 0.370; INAR: 0.362; KARA: 0.363; OUTA: 0.372; TEMS: 0.373). Correspondingly, MDS visualization of population structure revealed differentiation at two hierarchical levels. Analysis of the entire dataset showed the tributary subpopulations KEVO, TSAR, UTSJ and PULM to be strongly differentiated from one another and the mainstem, headwater and MASK subpopulations (Fig. 2). Excluding these tributary subpopulations and repeating the analysis also confirmed genomic divergence among the remaining six subpopulations (Fig. 2), although at least two individuals from OUTA were possible migrants from other sites (Fig. 2).

Comparing Teno mainstem and headwater juveniles to adults caught in the mainstem revealed, clearly, that the two genetically differentiated subpopulations described in Aykanat et al. (2015) derived from different spawning locations (Fig. S3). ‘Subpopulation 1’ overlapped with juveniles caught in the Gamjarga area of the Teno mainstem (TEMS), while ‘Subpopulation 2’ overlapped juveniles caught 150km upstream in Inarijoki (INAR). Adults not assigned to either subpopulation cluster overlapped juveniles collected from other Teno locations, including the OUTA site midway between INAR and TEMS. Thus, the subpopulation structuring observed in the Teno mainstem by Aykanat et al. (2015) was generated by the sampling of migratory adults originating from two subpopulations with geographically distinct spawning and/or juvenile rearing sites.

Regions of the genome potentially responding to local selection

Distributions of test scores for the top-ranked 10,000 SNPs, and the top 0.5% that were retained, are shown in Fig. S4. By combining results from the outlier and environmental association tests, we identified 1,409 SNPs that, under our criteria, were candidates to be linked to variants under local selection (0.71% of total SNPs examined; 676 ‘OUT’ SNPs, 227 ‘PC1’ SNPs, 460 ‘PC2’ SNPs, Table S3). These ‘candidate SNPs’ occurred

within 654 different haploblocks, and were associated with 522 overlapping and 258 downstream genes (Table S4; 1.9% of all genes annotated on the *S. salar* chromosomes; OUT: 258 overlapping/ 139 downstream; PC1: 93 overlapping/ 41 downstream; PC2: 168 overlapping/ 80 downstream). An additional 46 SNPs were supported as candidates by both XP-EHH and HapFLK results but no other tests ('HH' SNPs), however these were associated with only one additional gene (Tables S3 & S4). Thirty nine haploblocks (5%) contained SNPs on the 7K array that were identified as candidates for divergent selection or associated with traits of interest in previous studies (Table S3).)

Thirty-seven haploblocks had evidence for local selection both from outlier and/or environmental association analyses and from haplotype frequency/homozygosity analyses (Table 2; Fig. S5; 25 supported by HapFLK, 8 supported by XPEHH; 4 supported by both). These 'candidate haploblocks' were distributed over 17 of the 29 *S. salar* chromosomes. Three of the candidate haploblocks on Ssa09 (Fig. S5.5), five on Ssa12 (Fig. S5.8), two on Ssa17 (Fig. S5.15) and two on Ssa18 (Fig. S5.16) were located closely together on the chromosome and supported by a HapFLK signal that could be driven by a single selected locus; thus we considered there to be 28 independent candidate regions. Haploblock size ranged from 1.8Kb to 2.4Mb, reflecting substantial variation in local recombination rates around the candidate loci.

Six of these candidate haploblocks included loci of interest from previous studies (Table 2). A single haploblock with particularly strong evidence for divergent selection was located on Ssa09. *O. nerka* ecotype-associated RAD-tag 68810 mapped unambiguously to the *S. salar lrr9* gene within this haploblock. Ecotype-associated RAD-tag 24343 did not align with the *S. salar* genome, but mapped unambiguously on the *O. mykiss* genome $\approx$ 36kb from 68810 between the genes *lrr9* and *dhfr7*. The zebrafish retinal *six6* enhancer mapped unambiguously 5' of the *six6* gene within the haploblock.

GO enrichment analysis

Extended GO term enrichment results are provided in Table S5. We restrict our discussion to GO enrichments significant at $p < 0.001$ (Table 3). We emphasize that GO term enrichment results, in general, can be skewed by the presence of a small number of well-studied genes in the dataset, and that gene functions in non-model organisms such as salmonids are not well characterized. Nevertheless, several general trends emerge from the results. Corresponding to expectations that adaptive divergence is likely to be mediated by changes in gene expression our total set of candidate loci is enriched for transcription factors (GO:0006367; GO:0035326). We find an over-representation of various neuronal components (including the those of the neuronal cell body, growth cone, and dendritic shaft), and a corresponding over-representation of processes involved in nervous system development and function, particularly in respect to the sensory system. We also find over-representation of cell surface/plasma member proteins, and corresponding enrichment of GO terms related to cell-cell adhesion. Finally, we observe a few highly enriched GO terms related to muscle and skeletal differentiation.

DISCUSSION

In this study, we used a 'bottom up' approach – examining patterns of variation across the genome - to identify loci that may underlie fine-scale local adaptation of Atlantic salmon in a large river system. By combining results from outlier and environmental association analyses, we identified 1,409 candidate SNPs, annotated with 780 protein coding loci. While the relatively relaxed empirical p threshold ($p < 0.005$ based on the entire distribution) means that these candidate genes may include numerous false positives, this threshold also increases our power to detect genes underlying polygenic adaptive traits (De Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014). By adding evidence from patterns of haplotype homozygosity, we found 28 regions of the genome that are particularly likely to contain loci under differential selection (Table 2, Fig. S5). Six of these regions have been documented as potential selective targets among other Atlantic salmon populations and/or contain candidate

genes known to underlie ecologically relevant phenotypic variation (Table 2, Fig. S5). This observation both increases our confidence that the other regions identified using the same criteria also harbour genes under divergent selection, and also suggests that certain loci are under repeated selection among Atlantic salmon at local, regional and continental scales. Of particular note, we found a single genomic region that was highly differentiated among Teno River subpopulations, a candidate selective target throughout the range of Atlantic salmon, and which co-localized with an ecologically important haplotype within a different salmonid genus.

A large-effect gene underlying age-at-maturity appears under differential selection among populations: A candidate haplotype for diversifying selection on Ssa25 contains the known large-effect locus underlying variation in age at maturity in Atlantic salmon (Ayllon et al., 2015; Barson et al., 2015; candidate gene *vgl3*) (Table 2, Fig. S5.26). Our results lend further support to the evidence in Barson et al. (2015) that variants at this locus may be differentially selected among rivers. This genomic region is strongly associated with PC2 (approximating flow volume), an association that suggests different trade-offs between size at maturity and reproductive success in different size rivers. At the simplest level, mechanical constraints could limit the access of larger, later-maturing fish to smaller tributaries, while only larger females may be able to successfully construct redds in higher-flow locations with coarser substrate (Kondolf & Wolman, 1993).

A strong signature of divergent selection among Atlantic salmon populations co-localizes with a locus associated with spawning site selection in Sockeye salmon: We observe an extremely strong signal of diversifying selection approximately 2.5Mb along Atlantic salmon chromosome Ssa09 (Fig. 3, Table 2, Fig. S5.5). SNPs in this region are among the most extreme outliers in all three outlier analyses, are robustly associated with PC2 (approximating flow volume) in both environmental association analyses, and are clearly indicated as co-locating with a selective sweep by XP-EHH and HapFLK results. The selective signal centres on a cluster of closely linked SNPs spanning ≈ 7 kb between the genes *protein phosphatase 1a* (*ppm1a*) and *SIX homeobox 6* (*six6*). The TEMS sample, from the large Teno mainstem, is almost fixed for one allele across these SNPs, and XP-EHH results indicate a selective sweep on this allele. Conversely, the samples from the smaller tributaries KEVO, TSAR and PULM are almost fixed for the alternate allele (Table S3). The strong relationship of allele frequencies at this location with PC2 infers that this outlying locus is also strongly associated with subpopulation run time, a relationship that we confirmed by performing a supplementary LFMM analysis with median run time as the dependent variable (Table S6).

Differentiation of this Ssa09 region among Atlantic salmon populations was previously shown by Barson et al. (2015), where the candidate SNPs were F_{st} outliers and strongly associated with variation in age-at-maturity in genome-wide association studies before correction for population stratification. This pattern was observed both in the Teno and throughout Norway. Barson et al. (2015) also found a small but significant effect of genotype at this locus on length of returning adults. Recently, Cauwelier, Gilbey, Sampayo, Stradmeyer, & Middlemas (2017) found that the locus was associated with intra-population variation in run timing of Atlantic salmon in Scotland. Further, Bourret, Dionne, et al. (2013) and Moore et al. (2014), using a 7K SNP chip, found evidence that this region of Ssa09 was under diversifying selection among Atlantic salmon populations in North America. This suggests that the same variant may be involved in adaptive divergence throughout the range of the species. Intriguingly, this strongly outlying region is close to the locus associated with spawning site selection in *O. nerka* (Veale & Russello, 2017a). The relevant RAD-tags map within our candidate haplotype but $\approx 75 - 100$ Kb away from our strongest candidate markers (Fig. 3). Given the much lower density of the Veale & Russello (2017b), Nichols, Kozfkay, & Narum (2016), and Larson et al. (2017) RAD-tags compared to our SNPs, the same locus may be involved in local adaptation in both cases. Supporting this, SNPs on the 7K array found to be associated with age-at-maturity in Teno salmon by Johnston et al. (2014) map ≈ 700 Kb distant (Table S3) but are clearly labelling our candidate locus. Further, Veale & Russello (2017a), sequencing through *lm9*, found that haplotype divergence increased towards our candidate region.

One gene flanking this region, *ppm1a*, is a broad-specificity enzyme whose potential role in mediating local adaptation is unclear. In contrast *six6* is an evolutionarily conserved transcriptional co-regulator with well characterized roles in development of the eye and establishment of the pituitary-hypothalamic axis in multiple vertebrates (Jean, Bernier, & Gruss, 1999; Seo, Drivenes, Ellingsen, & Fjose, 1998; Toy, Yang, Leppert, & Sundin, 1998). It is a paralogue of the invertebrate gene *optix*, which underlies locally adaptive variation in wing pigmentation across multiple butterfly genera (Zhang, Mazo-Vargas, & Reed, 2017). In the developing vertebrate eye, *six6* interacts with other genes to mediate both early initiation of the eye field and later establishment of the mature retina (Conte et al., 2010; Seo et al., 1998). It continues to be expressed in the fully developed eye (Aijaz et al., 2005), and is associated with thickness of the retinal nerve fibre layer and primary open-angle glaucoma in humans (Ulmer-Carnes et al., 2014). Through its role in hypothalamic development, it is required for proper development of the suprachiasmatic nucleus, which is the central regulator of circadian timing in mammals (Clark et al., 2013), and may have a similar role in fish (Watanabe et al., 2012). *Six6* is also an important regulator of fertility in mammals of both sexes, via its effect on gonadotropin-releasing hormone production by the hypothalamus (Larder, Clark, Miller, & Mellon, 2011). Correspondingly, like *vgll3*, it is implicated in human pubertal timing (Hou et al., 2017). Barson et al. (2015) did not find any nonsynonymous mutations in the *six6* coding region, and results indicate that the selective target could be a regulatory element for this gene. Studies in various species have identified multiple 5' evolutionarily conserved enhancer elements which regulate *six6* expression in different tissues and at different development stages (Conte et al., 2010; Ledford et al., 2017; Lee et al., 2012). *Six6* regulation also involves at least one 3' enhancer (Conte et al., 2010; Ledford et al., 2017), and may also be mediated by local chromatin structure (Gómez-Marín et al., 2015). We found that one conserved 5' enhancer, responsible for *six6* expression in the maturing medaka retina (Conte et al., 2010) maps between two candidate SNPs (ss1868043109 and ss1868793364). Barson et al. (2015) showed that a second conserved enhancer, necessary for *six6* expression in the mouse forebrain (Lee et al., 2012) maps close to candidate SNP ss1868315338. Additional putative enhancers, with poorly characterized functions, are also located within our candidate region in other species (Conte et al., 2010; Ledford et al., 2017; Lee et al., 2012).

How this region interacts with the Ssa25 (*vgll3*) locus to influence age-at-maturity – whether through a direct functional relationship, and/or indirectly by altering the selective landscape for the Ssa25 locus among populations – remains unknown. Given the implication that the same genomic region may be involved in spawning site differentiation in *O. nerka*, we hypothesize that *six6* could mediate local adaptation/spawning site differentiation in both species by modulating aspects of the sensory and/or reproductive system, including reproductive timing. Notably, both Atlantic salmon in larger rivers and sockeye salmon on lake shores tend to be spawning at increased depths (with consequent altered light regimens) and on coarser substrates (Frazer & Russello, 2013; Louhi, Mäki-Petäys, & Erkinaro, 2008). Further, *O. nerka* spawning ecotypes also differ in their reproductive timing (Frazer & Russello, 2013). As with its invertebrate paralogue (Zhang et al., 2017), *six6* may have pleiotropic effects on multiple aspects of the phenotype by acting as ‘master switch’ across different regulatory networks during development.

In addition to *vgll3* and *six6*, several other genes implicated in human pubertal timing (Hou et al., 2017) are associated with PC2 SNPs. These include *neuronal growth regulator 1* (*negr1*, Ssa10), linked to both obesity and sexual maturation in several taxa (Lee et al., 2012); *trmt11* (Ssa05); *ptprf* (Ssa10); *ntkr2* (Ssa10); *lgr4* (Ssa11); *ptprd* (Ssa18); *kcnk9* (Ssa27); and *b6st1* (Ssa29) (Table S4). These genes merit further investigation as potential interactors involved in the relationship between *six6*, *vgll3* and life-history variation among different Atlantic salmon populations.

Possible selection on genes involved in circadian timing: After the Ssa09 locus, the genomic region most clearly associated with PC2 (and run timing) in the LFMM analysis is located $\approx$ 19Mb along Ssa11 (Table S3, Table S4, Table S6).

While this region is not supported as a ‘candidate haploblock’ given our current thresholds, it contains SNPs within the top 1% of XP-EHH scores. This region is also highly differentiated between northern and southern populations of *S. salar* in Norway (Kjærner-Semb et al., 2016). One of two candidate genes at this locus is zinc finger homeobox 3 (*zfx3*), a transcription factor expressed in the suprachiasmatic nucleus with a role in mammalian circadian rhythms, including sleep (Balzani et al., 2016). This must therefore be considered a candidate locus for co-selection with *six6*.

A candidate haploblock at on Ssa13 overlaps *neuronal PAS domain protein 2* (*npas2*) (Table 2, Fig. S5.10). *Npas2* is a paralogue of *clock*, which is well known for regulating circadian rhythms and reproductive cycles in diverse taxa. In mammals, the two genes have similar roles in the suprachiasmatic nucleus, and *npas2* can compensate if *clock* is silenced (DeBruyne, Weaver, & Reppert, 2007). *Clock* polymorphisms are associated with the timing of reproduction in rainbow trout and chinook salmon, and show a latitudinal cline consistent with local selection in the latter (Leder, Danzmann, & Ferguson, 2006; O’Malley, Camara, & Banks, 2007; O’Malley & Banks, 2008). Few salmonid studies have examined *npas2*, but O’Malley, Jacobson, Kurth, Dill, & Banks (2013) found that *npas2*-linked variation discriminated *O. tshawytscha* populations with different migratory timing. In our analysis, however, this candidate haploblock is not associated with PC2, meaning that there is no clear link with run time.

Evidence for local selection on genes mediating energy homeostasis: The extremely large candidate haploblock on Ssa19 (Table 2, Fig. S5.20) overlaps a region of the genome responding to domestication selection in North America *S. salar* (Liu et al., 2017; Mäkinen et al., 2015). This region also contains a SNP differentiating wild and domestic Norwegian *S. salar* (Karlsson et al., 2011). Following Liu et al. (2017), we hypothesize that the target of selection is a regulatory element for the downstream *melanocortin receptor 4* (*mc4r*) gene. This gene is a controller of energy homeostasis and somatic growth in fish and other vertebrates via its influence on food intake and energy expenditure (Krashes, Lowell, & Garfield, 2016; Metz, Peters, & Flik, 2006); it is a well-characterized human obesity gene and expected to be differentially selected in the wild vs. domestic environment. Several other obesity-associated genes are candidate selective targets: four copies of *neurexin 3* (*nr3n3*, Heard-Costa et al. 2009) on a large candidate haploblock on Ssa09 (Table 2, Fig. S5.6); *negr1* (discussed above); *lingo2* (Ssa05 and Ssa09); *arid5b* (Ssa01); and *kedt15* (Ssa11) (Claussnitzer et al., 2015; Castillo, Hazlett & Orlando, 2017; Table S4). The observation of these genes as selective candidates conforms to a model in which different Teno Atlantic salmon sub-populations are adapted to different energy-balance optima. Further, as threshold levels of fat reserves at specific times of the year are thought influence sexual maturity in salmon (Thorpe, Mangel, Metcalfe, & Huntingford, 1998), such obesity-associated genes may further interact with *vgl3* and *six6* to influence variation in life-history dynamics among Atlantic salmon populations.

Signatures of directional selection on MHCII and other genes involved in pathogen response: A clear signal consistent with directional selection occurs around the *major histocompatibility complex II* (*mhcII*) gene on Ssa12 (Table 2, Fig. S5.8). Allelic variation within this haploblock correlates with PC1 (approximating distance from the ocean). Salmonid genomes contain only a single classical *mhcII* locus (Gómez, Conejeros, Marshall, & Consuegra, 2010), which mediates an immune response to bacteria and parasites (Piertney & Oliver, 2006). Elevated *mhc* diversity confers broader pathogen resistance: thus, over evolutionary timescales, *mhc* alleles are maintained by balancing selection (Piertney & Oliver, 2006). However, pathogen pressure within a population can generate directional selection on particular *mhc* alleles over shorter time scales. In salmonids, specific *mhcII* alleles have been associated with resistance to the bacterial diseases piscirickettsiosis (Gómez, Conejeros, Consuegra, & Marshall, 2011), furunculosis (Kjøglum, Larsen, Bakke, & Grimholt, 2008; Langefors, Lohm, Grahn, Andersen, & von Schantz, 2001) and infectious salmon anaemia (Kjøglum, Larsen, Bakke, & Grimholt, 2006), and the parasite *Myxobolus cerebralis* (Dionne, Miller, Dodson, & Bernatchez, 2009). Similar evidence for localized directional selection on *mhcII* variants has been observed in sockeye salmon at both broad and fine spatial scales (Larson, Seeb, Dann, Schindler, & Seeb, 2014; McClelland et al., 2013).

Signatures of diversifying selection around other genes involved in pathogen response supports this model of different pathogen pressures among our sampled subpopulations, despite their co-occurrence in the same drainage basin. A strong XP-EHH signal suggesting a recent selective sweep, in particular in the Yla Pulmankijarvi (PULM) subpopulation, occurs on Ssa04 around a copy of *cox sackie virus and adenovirus receptor* (*cxadr*) (Table 2, Fig. S5.2). Although this gene has multiple functions, this may indicate virus-mediated selection. Other immune associated genes located in candidate haploblocks include two important mediators of the inflammatory response: *beta 1 integrin* (*itgb1*), on Ssa14 and *interleukin 1 beta* (*il1b*) on Ssa24 (Table 2).

Other candidate selective targets: The candidate haploblock on Ssa17, associated with PC2, is immediately upstream of *aryl hydrocarbon receptor 2 gamma* (*ahr2g*) (Table 2, Fig. S5), and we observed candidate SNPs upstream of two of the three other *ahr2* genes annotated on the *S. salar* genome (*ahr2a* and *ahr2d*, Table S4). Although genetic variation in the aryl hydrocarbon receptor pathway underlies adaptation to anthropogenic pollution (Reid et al., 2016; Wirgin et al., 2011), this is unlikely to be a selective force in the Teno. Regulatory variation in the Ahr2 pathway is associated with ecotypic diversity in craniofacial morphology in Arctic charr (*Salvelinus alpinus*, Ahi et al., 2014).

The small candidate haploblock on Ssa05 overlaps one of three conserved copies of *5-hydroxytryptamine receptor 4* (*htr4*) annotated on the Atlantic salmon genome (Table 3, Fig. S5). The neurotransmitter 5-HT (serotonin) is an important modulator of vertebrate behavioural state; in Atlantic salmon, for example, it mediate the expression of different behavioural syndromes between early and later-emerging fry, likely an adaptive response to varying levels of competition (Thornqvist, Hoglund, & Winberg, 2015). The class 4 receptors of 5-HT are not well characterized in fish, however in mammals they are expressed in the nervous and gastrointestinal systems, and have specific roles in memory formation and feeding behaviour (Bockaert, Claeyens, Compan, & Dumuis, 2008).

Examination of GO terms implied an enrichment of nervous system genes overlapping or downstream of candidate SNPs (Table 3). Further, a number of candidate haploblocks and SNPs are annotated with genes whose adaptive function is not clear but which code proteins important for neuronal function. These include *amphiphysin* (*amph*), the only gene annotated on the candidate haploblock near the start of Ssa14 (Table 2; Fig. S5.11), and two copies of *neurophilin 1a* (*nrf1a*) associated with Ssa14 and Ssa19 haploblocks (Table 2; Figs. S5.12 & S5.19). We tentatively hypothesize that aspects of the nervous system, perhaps underlying sensory systems involved in natal homing, may be under differential selection among different Atlantic salmon populations. This suggests several interesting avenues for future research.

Management Implications: Our results strongly indicate that Teno River Atlantic salmon subpopulations identified on the basis of microsatellite variation are also differentiated at the functional genetic level, and so are unlikely to be ecologically interchangeable. This underscores the recommendations of Vähä, Erkinaro, Niemelä, & Primmer (2007) and Vähä, Erkinaro, Falkegård, Orell, & Niemelä (2017) that the stock be managed at the subpopulation level. On a broader scale, our results support the hypothesis that local adaptation is common among Atlantic salmon populations, even those in geographic proximity. This should be taken into account, for example, in stocking and restoration programs.

ACKNOWLEDGEMENTS

This work was supported by Academy of Finland Grants 284941 and 314254 to CRP and a grant from the Finnish Cultural Foundation to VLP. We thank Sigbjørn Lien, Matthew Kent, and Silje Karoliussen (CIGENE) for SNP Array genotyping, Kristiina Haapanen for laboratory assistance, and Matthieu Bruneaux and Ksenia Zueva for help with GO analyses. Computing resources were provided by CSC – IT Center for Science Ltd (Finland).

REFERENCES

- Ahi, E., Kapralova, K., Pálsson, A., Maier, V., Gudbrandsson, J., Snorrason, S. S., ... Franzdóttir, S. (2014). Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr. *EvoDevo*, 5(1), 40. <http://doi.org/10.1186/2041-9139-5-40>
- Aijaz, S., Allen, J., Tregidgo, R., van Heyningen, V., Hanson, I., & Clark, B. J. (2005). Expression analysis of SIX3 and SIX6 in human tissues reveals differences in expression and a novel correlation between the expression of SIX3 and the genes encoding isocitrate dehydrogenase and cadherin 18. *Genomics*, 86(1), 86–99. <http://doi.org/10.1016/j.ygeno.2005.03.002>
- Alexa, A., & Rahnenfuhrer, J. (2010). topGO: Enrichment analysis for Gene Ontology. R package version 2.18.0. Available:
- Aykanat, T., Johnston, S. E., Orell, P., Niemelä, E., Erkinaro, J., & Primmer, C. R. (2015). Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population. *Molecular Ecology*, 24(20), 5158–5174. <http://doi.org/10.1111/mec.13383>
- Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M. F., Dahle, G., ... Wargelius, A. (2015). The *vgl3* locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. *PLoS Genetics*, 11(11), e1005628. <http://doi.org/10.1371/journal.pgen.1005628>
- Balzani, E., Lassi, G., Maggi, S., Sethi, S., Parsons, M. J., Simon, M., ... Tucci, V. (2016). The Zfh3-mediated axis regulates sleep and interval timing in mice. *Cell Reports*, 16(3), 615–621. <http://doi.org/10.1016/j.celrep.2016.06.017>
- Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, 12(11), 767–780. <https://doi.org/10.1038/nrg3015>
- Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., ... Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, 528(7582), 405–408. <http://doi.org/10.1038/nature16062>
- Bernatchez, L., Renaut, S., Whiteley, A. R., Derome, N., Jeukens, J., Landry, L., ... St-Cyr, J. (2010). On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1547), 1783–800. <http://doi.org/10.1098/rstb.2009.0274>
- Blair, L. M., Granka, J. M., & Feldman, M. W. (2014). On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, 8(1), 1. <http://doi.org/10.1186/1479-7364-8-1>
- Bockaert, J., Claeysen, S., Compan, V., & Dumuis, A. (2008). 5-HT₄ receptors: History, molecular pharmacology and brain functions. *Neuropharmacology*, 55(6), 922–931. <http://doi.org/10.1016/j.neuropharm.2008.05.013>
- Bourret, V., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation. *Evolution*, 67(12), 3469–3487. <http://doi.org/10.1111/evo.12139>
- Bourret, V., Kent, M. P., Primmer, C. R., Vasemägi, A., Karlsson, S., Hindar, K., ... Lien, S. (2013). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 22(3), 532–51. <http://doi.org/10.1111/mec.12003>
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, 81(5), 1084–1097. <http://doi.org/10.1086/521987>

- Carlson, S. M., & Seamons, T. R. (2008). SYNTHESIS: A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. *Evolutionary Applications*, 1(2), 222–238. <http://doi.org/10.1111/j.1752-4571.2008.00025.x>
- Cauwelier, E., Gilbey, J., Sampayo, J., Stradmeyer, L., & Middlemas, S. J. (2017). 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. *Canadian Journal of Fisheries and Aquatic Sciences*, cjfas-2017-0293. <http://doi.org/10.1139/cjfas-2017-0293>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. <http://doi.org/10.1186/s13742-015-0047-8>
- Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, 70(2), S0016672397002954. <http://doi.org/10.1017/S0016672397002954>
- Clark, D. D., Gorman, M. R., Hatori, M., Meadows, J. D., Panda, S., & Mellon, P. L. (2013). Aberrant development of the suprachiasmatic nucleus and circadian rhythms in mice lacking the homeodomain protein *Six6*. *Journal of Biological Rhythms*, 28(1), 15–25. <http://doi.org/10.1177/0748730412468084>
- Claussnitzer, M., Dankel, S. N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., ... Kellis, M. (2015). *FTO* obesity variant circuitry and adipocyte browning in humans. *New England Journal of Medicine*, 373(10), 895–907. <http://doi.org/10.1056/NEJMoa1502214>
- Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), 5039–5047. <http://doi.org/10.1098/rspb.2012.2146>
- Conte, I., Marco-Ferreres, R., Beccari, L., Cisneros, E., Ruiz, J. M., Tabanera, N., & Bovolenta, P. (2010). Proper differentiation of photoreceptors and amacrine cells depends on a regulatory loop between *NeuroD* and *Six6*. *Development (Cambridge, England)*, 137(14), 2307–17. <http://doi.org/10.1242/dev.045294>
- Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 185(4). <http://doi.org/10.1534/genetics.110.114819>
- De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, 23(8), 2006–2019. <http://doi.org/10.1111/mec.12705>
- DeBruyne, J. P., Weaver, D. R., & Reppert, S. M. (2007). *CLOCK* and *NPAS2* have overlapping roles in the suprachiasmatic circadian clock. *Nature Neuroscience*, 10(5), 543–545. <http://doi.org/10.1038/nn1884>
- Dionne, M., Miller, K. M., Dodson, J. J., & Bernatchez, L. (2009). MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1523), 1555–1565. <http://doi.org/10.1098/rstb.2009.0011>
- Dodson, J. J., Aubin-Horth, N., Thériault, V., & Páez, D. J. (2013). The evolutionary ecology of alternative migratory tactics in salmonid fishes. *Biological Reviews*, 88(3), 602–625. <http://doi.org/10.1111/brv.12019>
- Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences*, 101(4), 975–979. <http://doi.org/10.1073/pnas.0308064100>

- Elmer, K. R. (2016). Genomic tools for new insights to variation, adaptation, and evolution in the salmonid fishes: a perspective for charr. *Hydrobiologia* 783, 191. <http://doi.org/10.1007/s10750-015-2614-5>
- Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3). <http://doi.org/10.1534/genetics.112.147231>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977–93. <http://doi.org/10.1534/genetics.108.092221>
- Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. (2011). Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, 106(3), 404–420. <http://doi.org/10.1038/hdy.2010.167>
- Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research* 23, 1089–1096
- Frazer, K. K., & Russello, M. A. (2013). Lack of parallel genetic patterns underlying the repeated ecological divergence of beach and stream-spawning kokanee salmon. *Journal of Evolutionary Biology*, 26(12), 2606–2621. <http://doi.org/10.1111/jeb.12250>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983. <http://doi.org/10.1534/genetics.113.160572>
- Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30(7), 1687–1699. <http://doi.org/10.1093/molbev/mst063>
- Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., ... Quinn, T. P. (2007). A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews of the Cambridge Philosophical Society*, 82(2), 173–211. <http://doi.org/10.1111/j.1469-185X.2006.00004.x>
- Gómez, D., Conejeros, P., Consuegra, S., & Marshall, S. H. (2011). MHC mediated resistance to *Piscirickettsia* salmonis in salmonids farmed in Chile. *Aquaculture*, 318(1–2), 15–19. <http://doi.org/10.1016/j.aquaculture.2011.04.023>
- Gómez, D., Conejeros, P., Marshall, S. H., & Consuegra, S. (2010). MHC evolution in three salmonid species: a comparison between class II alpha and beta genes. *Immunogenetics*, 62(8), 531–542. <http://doi.org/10.1007/s00251-010-0456-x>
- Gómez-Marín, C., Tena, J. J., Acemel, R. D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., ... Gómez-Skarmeta, J. L. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24), 7542–7. <http://doi.org/10.1073/pnas.1505463112>
- Grammes, F. (2016). Ssa.RefSeq.db: Annotation package for the *Salmo salar* RefSeq (ICSASG_v2). R package version 1.2. Available: github.com/FabianGrammes/Ssa.RefSeq.db.
- Gunther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1), 205–220. <http://doi.org/10.1534/genetics.113.152462>
- Gutierrez, A. P., Yáñez, J. M., & Davidson, W. S. (2016). Evidence of recent signatures of selection during domestication in an Atlantic salmon population. *Marine Genomics*, 26, 41–50.

- <http://doi.org/10.1016/j.margen.2015.12.007>
- Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., & Davidson, W. S. (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One*, 10(3), e0119730. <http://doi.org/10.1371/journal.pone.0119730>
- Hecht, B. C., Campbell, N. R., Holecek, D. E., & Narum, S. R. (2013). Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, 22(11), 3061–3076. <http://doi.org/10.1111/mec.12082>
- Hecht, B. C., Thrower, F. P., Hale, M. C., Miller, M. R., & Nichols, K. M. (2012). Genetic architecture of migration-related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*. *G3: Genes | Genomes | Genetics*, 2(9), 1113–1127. <http://doi.org/10.1534/g3.112.003137>
- Hess, J. E., Zendt, J. S., Matala, A. R., & Narum, S. R. (2016). Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society of London B: Biological Sciences*, 283(1830). <http://doi.org/10.1098/rspb.2015.3064>
- Hou, H., Uusküla-Reimand, L., Makarem, M., Corre, C., Saleh, S., Metcalf, A., ... Wilson, M. D. (2017). Gene expression profiling of puberty-associated genes reveals abundant tissue and sex-specific changes across postnatal development. *Human Molecular Genetics*, 26(18), 3585–3599. <http://doi.org/10.1093/hmg/ddx246>
- Jean, D., Bernier, G., & Gruss, P. (1999). *Six6* (Optx2) is a novel murine Six3-related homeobox gene that demarcates the presumptive pituitary/hypothalamic axis and the ventral optic stalk. *Mechanisms of Development*, 84(1–2), 31–40. [http://doi.org/10.1016/S0925-4773\(99\)00068-4](http://doi.org/10.1016/S0925-4773(99)00068-4)
- Jeffery, N. W., Stanley, R. R. E., Wringe, B. F., Guijarro-Sabaniel, J., Bourret, V., Bernatchez, L., ... Bradbury, I. R. (2017). Range-wide parallel climate-associated genomic clines in Atlantic salmon. *Royal Society Open Science*, 4(11), 171394. <http://doi.org/10.1098/rsos.171394>
- Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., ... Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 23(14), 3452–68. <http://doi.org/10.1111/mec.12832>
- Jonsson, B., & Jonsson, N. (2001). Polymorphism and speciation in Arctic charr. *Journal of Fish Biology*, 58(3), 605–638. <http://doi.org/10.1111/j.1095-8649.2001.tb00518.x>
- Joseph J. Castillo, Zachary S. Hazlett, Robert A. Orlando, W. S. G. (2017). A global evolutionary and metabolic analysis of human obesity gene risk variants. *Gene*, 627, 412–419. <http://doi.org/10.1016/j.GENE.2017.07.002>
- Karlsson, S., Moen, T., Lien, S., Glover, K. A., & Hindar, K. (2011). Genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular Ecology Resources*, 11(SUPPL. 1), 247–53. <http://doi.org/10.1111/j.1755-0998.2010.02959.x>
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–34. <http://doi.org/10.1534/genetics.105.047985>
- Kjærner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemelä, E., ... Edvardsen, R. B. (2016). Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. *BMC Genomics*, 17(1), 610. <http://doi.org/10.1186/s12864-016-2867-z>
- Kjøglum, S., Larsen, S., Bakke, H. G., & Grimholt, U. (2006). How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*).

- Fish & Shellfish Immunology*, 21(4), 431–441. <http://doi.org/10.1016/j.fsi.2006.02.001>
- Kjøglum, S., Larsen, S., Bakke, H. G., & Grimholt, U. (2008). The effect of specific MHC class I and class II combinations on resistance to furunculosis in Atlantic salmon (*Salmo salar*). *Scandinavian Journal of Immunology*, 67(2), 160–168. <http://doi.org/10.1111/j.1365-3083.2007.02052.x>
- Kondolf, G. M., & Wolman, M. G. (1993). The sizes of salmonid spawning gravels. *Water Resources Research*, 29(7), 2275–2285. <http://doi.org/10.1029/93WR00402>
- Krashes, M. J., Lowell, B. B., & Garfield, A. S. (2016). Melanocortin-4 receptor-regulated energy homeostasis. *Nature Neuroscience*, 19(2), 206–219. <http://doi.org/10.1038/nn.4202>
- Kusche, H., Côté, G., Hernandez, C., Normandeau, E., Boivin-Delisle, D., & Bernatchez, L. (2017). Characterization of natural variation in North American Atlantic salmon populations (Salmonidae: *Salmo salar*) at a locus with a major effect on sea age. *Ecology and Evolution*, 7(15), 5797–5807. <http://doi.org/10.1002/ece3.3132>
- Langefors, a, Lohm, J., Grahn, M., Andersen, O., & von Schantz, T. (2001). Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proceedings. Biological Sciences / The Royal Society*, 268(1466), 479–485. <http://doi.org/10.1098/rspb.2000.1378>
- Larder, R., Clark, D. D., Miller, N. L. G., & Mellon, P. L. (2011). Hypothalamic dysregulation and infertility in mice lacking the homeodomain protein *Six6*. *The Journal of Neuroscience*: The Official Journal of the Society for Neuroscience, 31(2), 426–38. <http://doi.org/10.1523/JNEUROSCI.1688-10.2011>
- Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., & Seeb, L. W. (2017). Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular Ecology*, 26(2), 554–570. <http://doi.org/10.1111/mec.13933>
- Larson, W. A., Seeb, J. E., Dann, T. H., Schindler, D. E., & Seeb, L. W. (2014). Signals of heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye salmon. *Molecular Ecology*, 23(22), 5448–5461. <http://doi.org/10.1111/mec.12949>
- Leder, E. H., Danzmann, R. G., & Ferguson, M. M. (2006). The candidate gene, clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity*, 97(1), 74–80. <http://doi.org/10.1093/jhered/esj004>
- Ledford, K. L., Martinez-De Luna, R. I., Theisen, M. A., Rawlins, K. D., Viczian, A. S., & Zuber, M. E. (2017). Distinct cis-acting regions control *six6* expression during eye field and optic cup stages of eye formation. *Developmental Biology*, 426(2), 418–428. <http://doi.org/10.1016/j.ydbio.2017.04.003>
- Lee, A. W. S., Hengstler, H., Schwald, K., Berriel-Diaz, M., Loreth, D., Kirsch, M., ... Schäfer, M. K. E. (2012). Functional inactivation of the genome-wide association study obesity gene Neuronal Growth Regulator 1 in mice causes a body mass phenotype. *PLoS ONE*, 7(7), e41537. <http://doi.org/10.1371/journal.pone.0041537>
- Lee, B., Rizzoti, K., Kwon, D. S., Kim, S. Y., Oh, S., Epstein, D. J., ... Jeong, Y. (2012). Direct transcriptional regulation of *Six6* is controlled by SoxB1 binding to a remote forebrain enhancer. *Developmental Biology*, 366(2), 393–403. <http://doi.org/10.1016/j.ydbio.2012.04.023>
- Lepais, O., Manicki, A., Glise, S., Buoro, M., & Bardonnnet, A. (2017). Genetic architecture of threshold reaction norms for male alternative reproductive tactics in Atlantic salmon (*Salmo salar* L.). *Scientific Reports*, 7, 43552. <http://doi.org/10.1038/srep43552>

- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Retrieved from <http://arxiv.org/abs/1303.3997>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–60. <http://doi.org/10.1093/bioinformatics/btp324>
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., ... Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602), 200–205. <http://doi.org/10.1038/nature17164>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299. <http://doi.org/10.1093/bioinformatics/btr642>
- Liu, L., Ang, K. P., Elliott, J. A. K., Kent, M. P., Lien, S., MacDonald, D., & Boulding, E. G. (2017). A genome scan for selection signatures comparing farmed Atlantic salmon with two wild populations: Testing colocalization among outlier markers, candidate genes, and quantitative trait loci for production traits. *Evolutionary Applications*, 10(3), 276–296. <http://doi.org/10.1111/eva.12450>
- Lotterhos, K. E. K., & Whitlock, M. C. M. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23(9), 2178–92. <http://doi.org/10.1111/mec.12725>
- Louhi, P., Mäki-Petäys, A., & Erkinaro, J. (2008). Spawning habitat of Atlantic salmon and brown trout: General criteria and intragravel factors. *River Research and Applications*, 24(3), 330–339. <http://doi.org/10.1002/rra.1072>
- Mäkinen, H., Vasemägi, A., McGinnity, P., Cross, T. F., & Primmer, C. R. (2015). Population genomic analyses of early-phase Atlantic Salmon (*Salmo salar*) domestication/captive breeding. *Evolutionary Applications*, 8(1), 93–107. <http://doi.org/10.1111/eva.12230>
- Martínez, A., Garza, J. C., & Pearse, D. E. (2011). A microsatellite genome screen identifies chromosomal regions under differential selection in steelhead and rainbow trout. *Transactions of the American Fisheries Society*, 140(3), 829–842. <http://doi.org/10.1080/00028487.2011.588094>
- McClelland, E. K., Ming, T. J., Tabata, A., Kaukinen, K. H., Beacham, T. D., Withler, R. E., & Miller, K. M. (2013). Patterns of selection and allele diversity of class I and class II major histocompatibility loci across the species range of sockeye salmon (*Oncorhynchus nerka*). *Molecular Ecology*, 22(18), 4783–4800. <http://doi.org/10.1111/mec.12424>
- Metz, J. R., Peters, J. J. M., & Flik, G. (2006). Molecular biology and physiology of the melanocortin system in fish: a review. *General and Comparative Endocrinology*, 148(2), 150–162. <http://doi.org/10.1016/j.ygcen.2006.03.001>
- Micheletti, S. J., Matala, A. R., Matala, A. P., & Narum, S. R. (2017). Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Molecular Ecology*. <http://doi.org/10.1111/mec.14407>
- Moore, J.-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., ... Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology*, 23(23), 5680–5697. <http://doi.org/10.1111/mec.12972>
- Moore, J.-S., Harris, L. N., Le Luyer, J., Sutherland, B. J. G., Rougemont, Q., Tallman, R. F., ... Bernatchez, L.

- (2017). Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic char. *Molecular Ecology*.
<http://doi.org/10.1111/mec.14393>
- Narum, S. R., & Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11 Suppl 1, 184–94. <http://doi.org/10.1111/j.1755-0998.2011.02987.x>
- Nichols, K. M., Edo, A. F., Wheeler, P. A., & Thorgaard, G. H. (2008). The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. *Genetics*, 179(3). <http://doi.org/10.1534/genetics.107.084251>
- Nichols, K. M., Kozfkay, C. C., & Narum, S. R. (2016). Genomic signatures among *Oncorhynchus nerka* ecotypes to inform conservation and management of endangered Sockeye Salmon. *Evolutionary Applications*, 9(10), 1285–1300. <http://doi.org/10.1111/eva.12412>
- O'Malley, K. G., & Banks, M. A. (2008). A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) Clock gene: evidence for selection on PolyQ length variants. *Proceedings of the Royal Society B: Biological Sciences*, 275(1653), 2813–2821. <http://doi.org/10.1098/rspb.2008.0524>
- O'Malley, K. G., Camara, M. D., & Banks, M. A. (2007). Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, 16(23), 4930–4941. <http://doi.org/10.1111/j.1365-294X.2007.03565.x>
- O'Malley, K. G., Jacobson, D. P., Kurth, R., Dill, A. J., & Banks, M. A. (2013). Adaptive genetic markers discriminate migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*) amid continued gene flow. *Evolutionary Applications*, 6(8), 1184–1194. <http://doi.org/10.1111/eva.12095>
- O'Toole, C. L., Reed, T. E., Bailie, D., Bradley, C., Cotter, D., Coughlan, J., ... McGinnity, P. (2015). The signature of fine scale local adaptation in Atlantic salmon revealed from common garden experiments in nature. *Evolutionary Applications*, 8(9), 881–900. <http://doi.org/10.1111/eva.12299>
- Pearse, D. E., Miller, M. R., Abadía-Cardoso, A., & Garza, J. C. (2014). Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/ rainbow trout. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1783).
<http://doi.org/10.1098/rspb.2014.0012>
- Perrier, C., Bourret, V., Kent, M. P., & Bernatchez, L. (2013). Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Molecular Ecology*, 22(22), 5577–5593. <http://doi.org/10.1111/mec.12500>
- Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96(1), 7–21. <http://doi.org/10.1038/sj.hdy.6800724>
- Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., ... Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, 3(8), e1603198. <http://doi.org/10.1126/sciadv.1603198>
- Pritchard, V. L., Erkinaro, J., Kent, M. P., Niemelä, E., Orell, P., Lien, S., & Primmer, C. R. (2016). Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. *Evolutionary Applications*, 9(8), 1017–1031. <http://doi.org/10.1111/eva.12407>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–75. <http://doi.org/10.1086/519795>

- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <http://doi.org/10.1093/bioinformatics/btq033>
- Quinn, T. P., McGinnity, P., Reed, T. E., & Bradford, M. (2016). The paradox of “premature migration” by adult anadromous salmonid fishes: patterns and hypotheses. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(7), 1015–1030. <http://doi.org/10.1139/cjfas-2015-0345>
- R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna Austria ISBN, 1(10), 1–2896.
- Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., ... Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, 354(6317), 1305–1308. <http://doi.org/10.1126/science.aah4993>
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. <http://doi.org/10.1038/nature06250>
- Seo, H. C., Drivenes, O., Ellingsen, S., & Fjose, A. (1998). Expression of two zebrafish homologues of the murine Six3 gene demarcates the initial eye primordia. *Mechanisms of Development*, 73(1), 45–57. [http://doi.org/10.1016/S0925-4773\(98\)00028-8](http://doi.org/10.1016/S0925-4773(98)00028-8)
- Szpiech, Z. A., & Hernandez, R. D. (2014). Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, 31(10), 2824–2827. <http://doi.org/10.1093/molbev/msu211>
- Thornqvist, P.-O., Hoglund, E., & Winberg, S. (2015). Natural selection constrains personality and brain gene expression differences in Atlantic salmon (*Salmo salar*). *Journal of Experimental Biology*, 218(7), 1077–1083. <http://doi.org/10.1242/jeb.114314>
- Thorpe, J. E., Mangel, M., Metcalfe, N. B., & Huntingford, F. A. (1998). Modelling the proximate basis of salmonid life-history variation, with application to Atlantic salmon, *Salmo salar* L. *Evolutionary Ecology*, 12(5), 581–599. <http://doi.org/10.1023/A:1022351814644>
- Toy, J., Yang, J. M., Leppert, G. S., & Sundin, O. H. (1998). The optx2 homeobox gene is expressed in early precursors of the eye and activates retina-specific genes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), 10643–8. <http://doi.org/10.1073/pnas.95.18.10643>
- Ulmer Cames, M., Liu, Y. P., Allingham, R. R., Whigham, B. T., Havens, S., Garrett, M. E., ... Hauser, M. A. (2014). Discovery and Functional Annotation of SIX6 Variants in Primary Open-Angle Glaucoma. *PLoS Genetics*, 10(5), e1004372. <http://doi.org/10.1371/journal.pgen.1004372>
- Vähä, J.-P. P., Erkinaro, J., Niemelä, E., & Primmer, C. R. (2007). Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, 16(13), 2638–2654. <http://doi.org/10.1111/j.1365-294X.2007.03329.x>
- Vähä, J.-P. P., Erkinaro, J., Niemelä, E., Primmer, C. R., Saloniemi, I., Johansen, M., ... Brors, S. (2011). Temporally stable population-specific differences in run timing of one-sea-winter Atlantic salmon returning to a large river system. *Evolutionary Applications*, 4(1), 39–53. <http://doi.org/10.1111/j.1752-4571.2010.00131.x>
- Vähä, J.-P., Erkinaro, J., Falkegård, M., Orell, P., & Niemelä, E. (2017). Genetic stock identification of Atlantic salmon and its evaluation in a large population complex. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(3), 327–338. <http://doi.org/10.1139/cjfas-2015-0606>

- Vähä, J.-P., Erkinaro, J., Niemelä, E., & Primmer, C. R. (2008). Temporally stable genetic structure and low migration in an Atlantic salmon population complex: implications for conservation and management. *Evolutionary Applications*, 1(1), 137–154. <http://doi.org/10.1111/j.1752-4571.2007.00007.x>
- Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2015). Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular Ecology*, 25(1), 89–103. <http://doi.org/10.1111/mec.13360>
- Veale, A. J., & Russello, M. A. (2017a). An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Scientific Reports*, 7(1), 1747. <http://doi.org/10.1038/s41598-017-01890-2>
- Veale, A. J., & Russello, M. A. (2017b). Genomic changes associated with reproductive and migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). *Genome Biology and Evolution*. <http://doi.org/10.1093/gbe/evx215>
- Vilas, A., A. Pérez-Figueroa, and A. Caballero. 2012. A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology* 25 (7): 1364–1376. doi:10.1111/j.1420-9101.2012.02526.x.
- Watanabe, N., Itoh, K., Mogi, M., Fujinami, Y., Shimizu, D., Hashimoto, H., ... Suzuki, T. (2012). Circadian pacemaker in the suprachiasmatic nuclei of teleost fish revealed by rhythmic period2 expression. *General and Comparative Endocrinology*, 178(2), 400–407. <http://doi.org/10.1016/j.ygcen.2012.06.012>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358. <http://doi.org/10.2307/2408641>
- Whitlock, M., & Lotterhos, K. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST} . *The American Naturalist*, 186 (Suppl 1), S24–36. <http://doi.org/10.1086/682949>
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.
- Wirgin, I., Roy, N. K., Loftus, M., Chambers, R. C., Franks, D. G., & Hahn, M. E. (2011). Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. *Science (New York, N.Y.)*, 331(6022), 1322–5. <http://doi.org/10.1126/science.1197296>
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65(7), 1897–1911. <http://doi.org/10.1111/j.1558-5646.2011.01269.x>
- Zhang, L., Mazo-Vargas, A., & Reed, R. D. (2017). Single master regulatory gene coordinates the evolution and development of butterfly color and indescence. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), 10707–10712. <http://doi.org/10.1073/pnas.1709058114>

AUTHOR CONTRIBUTIONS

CRP, HM & VLP designed the study. JE, J-PV & PO collected tissue samples. HM & J-PV performed laboratory work. VLP analysed the data and wrote the manuscript, with input from CRP, HM, JE & PO. All authors approved the manuscript.

DATA ACCESSIBILITY

SNP genotypes, raw analysis results, code and other relevant files will be deposited in the Dryad Digital Repository.

Figure 1: Location of juvenile sampling sites

Table 1: Sample site environmental and phenotypic variables. Median run time is estimated from 1SW adults only (Vähä et al., 2011). Percentage of multi-seawinter females is rounded to the nearest 5%.

SITE	Latitude	Longitude	Catch- ment (km ²)	Distance (km)	Elevation (m)	Access- ibility	% MSW females	Median run time (day)
IESJ	69° 24' 23.49"	24° 40' 50.93"	1900	259	230	4	95	185
INAR	69° 8' 13.67"	25° 44' 33.39"	1257	245	159	4	45	185
KARA	69° 10' 28.08"	24° 41' 39.30"	1235	285	244	4	90	184
KEVO	69° 42' 11.52"	26° 57' 27.49"	494	127	94	5	10	178
MASK	70° 12' 43.50"	27° 56' 36.71"	463	41	56	2	60	187.5
OUTA	69° 38' 0.45"	25° 57' 41.63"	8387	175	115	4	45	185
PULM	69° 53' 36.56"	28° 1' 29.77"	598	76	22	2	20	172
TEMS	69° 55' 54.79"	27° 8' 58.64"	11035	93	60	3	75	190
TSAR	69° 46' 58.03"	26° 54' 25.39"	236	125	170	6	20	173
UTSJ	69° 39' 51.21"	27° 4' 16.28"	441	131	107	4	85	184.5

Figure 2: Two-dimensional MDS plot based on genome-wide IBS. (A) All juvenile samples; (B) all samples except KEVO, PULM, TSAR and UTSJ.

Table 2: Candidate haploblocks supported by both outlier or environmental association analyses and HapFLK or XP-EHH analyses. ‘Type’ – haploblock associated with PC1, PC2, or outlying (OUT), based on haploblock consensus SNP signal; ‘Annotated Genes’: gene symbols based on NCBI annotated gene products. Overlapping and selected downstream (ds) genes are shown. Genes in bold are those that are closest to the most likely selective targets (if any) based on visual observation of SNP score distribution (Fig. S5).

Chrom	Type		Start	End	Annotated Genes	Notes
ssa03	OUT	HapFLK	54440792	54662849	<i>mpst</i> (2), <i>mgat3</i> (2), <i>tex33</i> , <i>mid51</i> , <i>ube2i</i> , <i>c1qtnf6</i> , <i>il2b</i> , uncharacterized (1)	-
ssa04	OUT	HapFLK, XP-EHH	49548061	49670244	<i>btg3</i> , <i>cxadr</i>	Outlier among N. American <i>S. salar</i> populations (Moore <i>et al.</i> 2014)
ssa05	OUT	HapFLK	39939921	39945242	<i>htr4</i>	-
ssa09	PC2	HapFLK, XP-EHH	24690259	24953243	<i>ppm1a</i> , <i>six6</i> , <i>dhrs7</i> , <i>rtn1</i> , <i>lrrc9</i> , <i>pcnx4</i> , <i>pgbd4</i> , <i>six1</i>	Associated with age-at-maturity and run timing in <i>S. salar</i> (Barson <i>et al.</i> 2015; Cauwelier <i>et al.</i> 2017) Outlier among N. American <i>S. salar</i> populations (Bourret, Dionne, <i>et al.</i> , 2013; Moore <i>et al.</i> , 2014) Associated with spawning location in <i>O. nerka</i> (Veale & Russello, 2017a, 2017b)
ssa09	OUT	HapFLK	24998704	25337622	<i>mnat1</i> , <i>slc38a6</i> , <i>prkch</i> , <i>hif1a</i> , <i>sacs</i> , <i>znf239</i> , <i>wdr89</i> , <i>ppp2r5e</i>	-
ssa09	OUT	HapFLK	25427680	25766954	<i>kcnh5</i> , <i>gphb5</i> , <i>jag2</i>	-
ssa09	PC1	HapFLK	26371376	26399217	<i>plcb1</i> , <i>tmx1</i> (ds), <i>cdca4</i> (ds)	-
ssa09	OUT	HapFLK	42158973	43699365	<i>nrxn3</i> (4), <i>slirp</i> , <i>alkbh1</i> , <i>sptlc2</i> , <i>ism2</i> , <i>xlcgf42.1</i> , <i>fam186a</i> , <i>znf239</i> , <i>znf585a</i> ,	-

					<i>emilin1, snap25</i> , uncharacterized (1)	
ssa10	OUT	XP-EHH	18878599	19137683	<i>midn, dos, fstl3, rnf126, bsg, atp5d, stk11, hcn2</i>	-
ssa10	OUT	HapFLK	19419205	19484226	<i>spin1, qil1, hsd11b1l, polrmt</i> (ds), <i>rgma</i> (ds)	-
ssa12	PC2	XP-EHH	60592476	60953722	<i>klf15, nucks1, slc45a3, elk4, mfsd4a, cdk18, klhdc8a, slc41a1</i>	-
ssa12	PC1	HapFLK	60985380	61042362	<i>slc41a1, etnk1, sox13</i>	-
ssa12	PC1	HapFLK	61059163	61122037	<i>sme, sox13, cdc42</i> , uncharacterized (1)	-
ssa12	PC1	HapFLK, XP-EHH	61343872	61703592	<i>mhcII-dab, mdfi, tmem183a, tfef, rab40c</i>	Associated with age-at-maturity in <i>S. salar</i> (Johnston et al., 2014)
ssa12	OUT	HapFLK	61735215	61917647	<i>perk8, foxp4</i>	-
ssa12	PC1	HapFLK	61945904	62288997	<i>mrps16, tas1r3, errfi1, srp, pax7</i>	-
ssa13	OUT	XP-EHH	9604538	9606700	<i>trpc4ap</i>	-
ssa13	PC1	HapFLK	82012742	82114498	<i>npas2, aifm1, nono, vma21</i> , uncharacterized (1)	-
ssa14	OUT	XP-EHH	3730963	3779453	<i>amph</i>	-
ssa14	PC2	XP-EHH	18263799	18453977	<i>itgb1, nrp1a, epc1</i>	-
ssa16	PC2	HapFLK	69506105	69700244	<i>mid1, arhgap6, msl3</i> (ds)	-
ssa17	PC2	XP-EHH	6734823	6862460	<i>ahr2g</i> (ds), <i>s100b, cntnap5</i> , uncharacterized (1)	Outlier between Baltic anadromous and landlocked <i>S. salar</i> (Bourret, Kent, et al., 2013)
ssa17	PC2	HapFLK	18258926	18332523	<i>gpr180, ddx3y, tgds, slc5a3, ptgfrn</i>	-
ssa17	OUT	HapFLK	18450541	18743279	<i>kpna4, cdadc1, cab39, fafx, unc93, tubgcp3, ddx3y, atp11a</i>	-

ssa18	PC1	HapFLK	7977374	8044809	<i>hs3st1</i> , <i>blnk</i> (ds)	-
ssa18	PC2	HapFLK	26409568	26428879	<i>slc22a7</i> , <i>zcchc24</i> (ds), <i>nphp1</i> (ds)	-
ssa18	OUT	HapFLK	45989781	46042995	<i>taf6l</i> , <i>mta2</i> , <i>ecsit</i> (ds), <i>ehbp1</i> , <i>kcnk1</i>	-
ssa18	PC2	HapFLK	46146694	46150451	<i>men1</i> , <i>map4k2</i> (ds), <i>rtn2</i> (ds)	-
ssa19	OUT	HapFLK	17946362	18012584	<i>nrp1a</i> (ds), <i>crem</i> (ds)	
ssa19	OUT	HapFLK	21194663	23553773	<i>cdh12</i> , <i>ttyh2</i> , <i>btbd3</i> , <i>f13a</i> , <i>cdh10</i> , <i>cdh6</i> , <i>mc4r</i> (ds), uncharacterized (1)	Outlier between wild and domesticated <i>S. salar</i> , N. America and Norway (Karlsson et al., 2011; Liu et al., 2017; Mäkinen et al., 2015).
ssa20	OUT	HapFLK	43187801	43605543	<i>gdpd1</i> , <i>lipt2</i> , <i>abcg1</i> , <i>spsc2</i> , <i>rnf169</i> , <i>tmem120b</i> , <i>por</i> , <i>fus</i> , <i>mmp28</i>	-
ssa20	PC2	XP-EHH	65649774	66094034	<i>lrx2</i> , <i>mrps23</i> , <i>msi2</i> , <i>nlk</i>	-
ssa23	OUT	XP-EHH	44140557	44159611	<i>prrtgb</i> , <i>myo5c</i> (ds), <i>tcf12</i> (ds)	-
ssa24	OUT	HapFLK	1264159	1272401	<i>cnm4</i> , <i>tjp2</i> (ds), <i>il1b</i> (ds)	-
ssa25	PC2	HapFLK, XP-EHH	24013779	24066599	<i>lztr</i> (ds), <i>pofut2</i> (ds), <i>myo1b</i>	-
ssa25	PC2	HapFLK	28639461	28832410	<i>vgll3</i> , <i>akap11</i> , <i>tnfsf11</i> , <i>epsti1</i> , <i>dnajc15</i>	Strongly associated with age-at-maturity (Ayllon et al., 2015; Barson et al., 2015)
ssa27	OUT	HapFLK	9990038	9991883	<i>znf384</i> , <i>kif2</i> (ds)	-

Table 3. GO Enrichment Analysis. Only terms enriched at $p < 0.001$ are shown. Where GO terms are enriched under these criteria in more than one analysis, statistics show results from the largest set of examined genes, and additional analyses are shown in brackets. ALL: All candidate genes; PC1: candidate genes associated with PC1 SNPs; PC2: candidate genes associated with PC2 SNPs; OUT: candidate genes associated with OUT SNPs; BP: Biological Process; CC: Cellular Component; MF: Molecular Function.

Test	GO category	GO Term	GO Term Description	Annotated	Hits	Expected	P
ALL (OUT, PC2)	BP	GO:0097374	sensory neuron axon guidance	19	6	0.45	3.70E-06
ALL (PC1)	BP	GO:0007156	homophilic cell adhesion via plasma membrane cell adhesion molecules	231	17	5.5	4.30E-05
ALL (OUT)	BP	GO:0030947	regulation of vascular endothelial growth factor receptor signaling pathway	158	11	3.76	7.70E-05
PC2	BP	GO:0072193	ureter smooth muscle cell differentiation	32	4	0.24	1.00E-04
ALL	BP	GO:0048842	positive regulation of axon extension involved in axon guidance	33	6	0.79	1.10E-04
ALL (OUT)	BP	GO:1902336	positive regulation of retinal ganglion cell axon guidance	12	4	0.29	1.40E-04
ALL (PC2)	BP	GO:0007185	transmembrane receptor protein tyrosine phosphatase signaling pathway	35	6	0.83	1.60E-04
ALL (OUT)	BP	GO:0034109	homotypic cell-cell adhesion	1599	43	38.06	2.20E-04
ALL	BP	GO:0051149	positive regulation of muscle cell differentiation	398	19	9.47	2.50E-04
ALL	BP	GO:0001922	B-1 B cell homeostasis	25	5	0.59	2.70E-04
ALL (OUT)	BP	GO:2000599	negative regulation of cyclin catabolic process	25	5	0.59	2.70E-04
ALL (PC2)	BP	GO:0035872	nucleotide-binding domain, LRR containing receptor signaling pathway	134	8	3.19	3.40E-04
ALL	BP	GO:0006367	transcription initiation from RNA polymerase	634	32	15.09	3.50E-04
ALL (OUT)	BP	GO:1902946	protein localization to early endosome	25	4	0.59	4.40E-04
ALL (OUT)	BP	GO:1902378	VEGF-activated neuropilin signaling pathway involved in axon guidance	7	3	0.17	4.40E-04
ALL (OUT)	BP	GO:0021649	vestibulocochlear nerve structural organization	7	3	0.17	4.40E-04
PC1	BP	GO:0016199	axon midline choice point recognition	84	4	0.36	4.60E-04
PC2	BP	GO:1902443	negative regulation of ripoptosome assembly involved in necroptotic process	5	2	0.04	5.70E-04
PC2	BP	GO:0090557	establishment of endothelial intestinal barrier	51	4	0.39	6.20E-04
ALL (OUT)	BP	GO:0033689	negative regulation of osteoblast proliferation	45	6	1.07	6.60E-04
ALL (OUT)	BP	GO:0071673	positive regulation of smooth muscle cell chemotaxis	8	3	0.19	6.90E-04
ALL (PC2)	BP	GO:2000374	regulation of oxygen metabolic process	8	3	0.19	6.90E-04
ALL (OUT)	CC	GO:0009986	cell surface	2162	86	51.81	1.20E-07

ALL (OUT)	CC	GO:0030426	growth cone	725	34	17.37	6.80E-05
ALL (PC2)	CC	GO:0043025	neuronal cell body	2293	82	54.95	2.20E-04
PC2	CC	GO:0043198	dendritic shaft	298	9	2.25	4.70E-04
OUT	CC	GO:0002116	semaphorin receptor complex	31	4	0.37	5.00E-04
PC2	CC	GO:0032279	asymmetric synapse	91	5	0.69	6.50E-04
ALL	MF	GO:0030275	LRR domain binding	82	7	1.98	2.10E-04
PC1	MF	GO:0045543	gibberellin 2-beta-dioxygenase activity	36	3	0.16	5.20E-04
PC2	MF	GO:0043273	CTPase activity	5	2	0.04	5.90E-04
ALL	MF	GO:0042802	identical protein binding	3919	118	94.86	6.20E-04
OUT	MF	GO:0017154	semaphorin receptor activity	35	4	0.42	8.00E-04
PC2	MF	GO:0035326	enhancer binding	475	7	3.7	8.40E-04
PC2	MF	GO:0035373	chondroitin sulfate proteoglycan binding	25	3	0.19	9.40E-04

Figure 3: Signatures of local selection on the *Ssa09* region containing the *Six6* gene. Dashed and dotted lines indicate empirical $p < 0.005$ and $p < 0.001$ respectively. The grey rectangle shows the boundaries of the associated haplotype block. Black vertical lines indicate the mapping position *O. nerka* ecotype-associated RAD tags 68810 (left) and 24343 (right) (Veale & Russello, 2017b).

Rotation:				
	PC1	PC2	PC3	PC4
Catchment	0.191	0.957	0.199	0.091
Distance	-0.582	0.266	-0.427	-0.639
Elevation	-0.632	0.095	-0.192	0.744
Accessibility	-0.474	-0.068	0.861	-0.172

Importance:				
	PC1	PC2	PC3	PC4
Standard_deviation	1.522	0.986	0.799	0.274
Proportion_variance	0.579	0.243	0.160	0.019
Cumulative_prop	0.579	0.822	0.981	1.000

Table S2: Pairwise unbiased F_{st} among samples, based on all SNPs.

	IESJ	INARI	KARA	KEVO	MASK	OUTA	PULM	TEMS	TSAR	UTSJ
IESJ	0.000	0.024	0.016	0.068	0.035	0.013	0.084	0.017	0.129	0.062
INARI		0.000	0.028	0.075	0.046	0.016	0.095	0.024	0.136	0.071
KARA			0.000	0.075	0.044	0.018	0.092	0.023	0.136	0.069
KEVO				0.000	0.083	0.064	0.128	0.065	0.059	0.084
MASK					0.000	0.033	0.104	0.033	0.148	0.080
OUTA						0.000	0.084	0.008	0.124	0.059
PULM							0.000	0.084	0.188	0.124
TEMS								0.000	0.125	0.057
TSAR									0.000	0.140
UTSJ										0.000

Table S3 (separate .csv file): Candidate SNP annotation, overlapping or closest downstream genes. **SNP:** Candidate SNP ss#; **Chr:** Chromosome (ssa); **Position:** Position on chromosome (bp); **EmpP:** Empirical P-value based on BayesCan; **EmpI:** Empirical I-value based on BayesCan. Empirical P-value based on BayesCan; Empirical I-value based on BayesCan.

OutFLANK: ; **XtX:** BayEnv2 XtX score; **BayenvPC1:** BayEnv2 |p| - association with PC1; **BayenvPC2:** BayEnv2 |p| - association with PC2; **LFMMPC1:** LFMM z-score - association with PC1; **LFMMPC2:** LFMM z-score - association with PC2; **HapFLK:** HapFLK score; **XPEHH:** Maximum |XP_EHH| over nine inter-population comparisons; **Type:** SNP classification, OUT, PC1, PC2, EHH; **XPEHH/HapFLK:** SNP is in the top 0.05% for either XP-EHH or HapFLK (1), or both (2); **IESJ_maf;** **INAR_maf;** **KARA_maf;** **KEVO_maf;** **MASK_maf;** **OUTA_maf;** **PULM_maf;** **TEMS_maf;** **TSAR_maf;** **UTSJ_maf:** Minor allele frequency in each sample (minor allele determined over all samples). **HB_start;** **HB_end:** Haploblock boundaries (bp); **7K_Outlier:** name of 7K SNP within haploblock; **7K_Study:** study in which 7K SNP was identified, (1) Bourret et al. 2013a; (2) Bourret et al. 2013b; (3) Moore et al. 2014; (4) Jeffery et al. 2017; (5) Mäkinen et al. 2015; (6) Liu et al. 2017; (7) Karlsson et al. 2011; (8) Gutierrez et al. 2016; (9) Johnston et al. 2014; (10) Perrier et al. 2013.

Table S4 (separate .csv file): Candidate SNP annotation, overlapping or closest downstream genes. **SNP:** Candidate SNP ss#; **Chr:** Chromosome (ssa); **Position:** Position on chromosome (bp); **Type:** Candidate SNP classification (OUT, PC1, PC2, EHH); **HapFLK/XPEHH:** Number of HapFLK/XP-EHH tests in which SNP is in the top 0.5%; **Hit:** Overlapping or closest downstream gene; **Strand:** gene strand; **Distance:** gene distance from SNP; **Product:** gene product inferred by NCBI; **Gene_symbol:** gene symbol; **SNP_location:** SNP location in respect to gene; **Gene_Location:** gene location in respect to SNP; **Alt_Hit:** Closest gene, if not downstream; **Alt_Strand:** gene strand; **Alt_Distance:** gene distance from SNP; **Alt_Product:** gene product inferred by NCBI; **Alt_Symbol:** gene symbol; **Alt_Location:** gene location in respect to SNP; **Noncoding_hit:** closest or overlapping non-coding locus, if closer than coding loci; **NC_strand:** locus strand; **NC_distance:** locus distance from SNP; **NC_type:** locus type; **NC_Location:** locus location in respect to SNP.

Table S5 (separate .csv file): Extended GO term analyses. All GO terms enriched at $p > 0.01$ are reported. **Test:** whether result was for genes annotated on all SNPs, PC1 candidate SNPs, PC2 candidate SNPs, or OUT candidate SNPs; **Aspect:** BP, Biological Process; CC, Cellular Component; MF, Molecular Function; **Annotated:** number of background genes annotated with the GO term; **Significant:** number of candidate genes annotated with the GO term; **Expected:** expected number of candidate genes with the GO term; **classicFisher:** uncorrected significance of GO term enrichment.

Table S6. Top 50 SNPs associated with population stratification in the population as IEMM analysis. SNPs are ordered by genomic position. **SNP**: dbSNP ID; **Chr**: chromosome; **Position**: Position (bp); **|z-score|**: Absolute z-score, median over five runs; **p**: Probability adjusted for λ ; **Hit**: overlapping or closest downstream gene; **Strand**: gene strand; **Distance**: distance of gene from SNP; **Symbol**: gene symbol.

SNP	Chr	Position	z-score	p	Hit	Strand	Distance	Symbol
ss1868188845	6	52437286	5.956	2.96E-05	mei4	+	25239	<i>mei4</i>
ss1867931876	7	39743181	6.457	5.96E-06	atp5c1	+	0	<i>atp5c1</i>
ss1868320062	9	24836460	7.001	9.13E-07	LOC106610975	+	0	<i>ppm1a</i>
ss1868103516	9	24836551	5.942	3.09E-05	LOC106610975	+	0	<i>ppm1a</i>
ss1868573231	9	24837700	7.342	2.62E-07	LOC106610975	+	0	<i>ppm1a</i>
ss1868320064	9	24837864	7.177	4.82E-07	dhrs7	-	-17949	<i>dhrs7</i>
ss1868546336	9	24841906	7.210	4.28E-07	dhrs7	-	-21991	<i>dhrs7</i>
ss1868076674	9	24841963	7.215	4.21E-07	dhrs7	-	-22048	<i>dhrs7</i>
ss1867970200	9	24842377	6.962	1.05E-06	dhrs7	-	-22462	<i>dhrs7</i>
ss1868310195	9	24844380	7.213	4.22E-07	dhrs7	-	-24465	<i>dhrs7</i>
ss1868360711	9	24847995	7.038	8.00E-07	dhrs7	-	-28080	<i>dhrs7</i>
ss1868315338	9	24870077	7.124	5.86E-07	six6	+	32701	<i>six6</i>
ss1868378525	9	24874391	7.108	6.20E-07	six6	+	28387	<i>six6</i>
ss1868643202	9	24875802	6.859	1.51E-06	six6	+	26976	<i>six6</i>
ss1868098595	9	24876478	7.074	7.01E-07	six6	+	26300	<i>six6</i>
ss1867998455	9	24878676	7.143	5.47E-07	six6	+	24102	<i>six6</i>
ss1868467979	9	24878713	7.199	4.45E-07	six6	+	24065	<i>six6</i>
ss1868806132	9	24879959	7.138	5.57E-07	six6	+	22819	<i>six6</i>
ss1868266633	9	24880995	7.190	4.60E-07	six6	+	21783	<i>six6</i>
ss1868806133	9	24881067	7.221	4.10E-07	six6	+	21711	<i>six6</i>
ss1868187813	9	24885705	7.098	6.44E-07	six6	+	17073	<i>six6</i>
ss1868336574	9	24886574	7.152	5.30E-07	six6	+	16204	<i>six6</i>
ss1867967563	9	24899756	6.986	9.61E-07	six6	+	3022	<i>six6</i>
ss1868793364	9	24902474	7.132	5.69E-07	six6	+	304	<i>six6</i>
ss1867989992	9	24902641	6.206	1.35E-05	six6	+	137	<i>six6</i>
ss1868459579	9	24903234	6.135	1.69E-05	six6	+	0	<i>six6</i>
ss1868459580	9	24903650	7.047	7.73E-07	six6	+	0	<i>six6</i>
ss1868259754	9	24904601	6.002	2.57E-05	six6	+	0	<i>six6</i>
ss1868107586	9	24905454	6.184	1.45E-05	six6	+	0	<i>six6</i>
ss1868323907	9	24905643	7.190	4.60E-07	mnat1	+	59647	<i>mnat1</i>
ss1868793365	9	24906492	6.124	1.75E-05	mnat1	+	58798	<i>mnat1</i>
ss1868459581	9	24906993	6.233	1.24E-05	mnat1	+	58297	<i>mnat1</i>
ss1868301468	9	24944806	6.304	9.83E-06	mnat1	+	20484	<i>mnat1</i>
ss1868172009	9	25482113	6.017	2.45E-05	LOC106610918	+	0	<i>kcnf5</i>
ss1868722897	9	1.02E+08	6.538	4.53E-06	LOC106612713	-	-13641	<i>hhipl1</i>
ss1868856494	10	35755853	6.010	2.50E-05	ptprf	+	0	<i>ptprf</i>
ss1868007037	11	19205606	6.539	4.52E-06	LOC106562084	-	0	<i>numa1</i>
ss1868133927	11	19205828	6.466	5.78E-06	LOC106562084	-	0	<i>numa1</i>
ss1868819149	11	19210153	6.208	1.34E-05	LOC106562084	-	0	<i>numa1</i>
ss1868668419	11	19248995	5.973	2.80E-05	LOC106562084	-	0	<i>numa1</i>
ss1868137711	11	19250875	6.080	2.01E-05	LOC106562084	-	0	<i>numa1</i>
ss1868162370	11	19272988	6.183	1.45E-05	LOC106562084	-	-6694	<i>numa1</i>
ss1868698436	11	19299154	5.949	3.02E-05	zfhx3	-	0	<i>zfhx3</i>
ss1868809730	15	49289033	6.180	1.47E-05	LOC106571749	+	0	<i>klhl29</i>
ss1868287346	15	49324191	6.109	1.84E-05	LOC106571749	+	0	<i>klhl29</i>
ss1868029361	15	49326015	6.131	1.71E-05	LOC106571749	+	0	<i>klhl29</i>
ss1868295098	15	49333951	6.103	1.87E-05	LOC106571749	+	0	<i>klhl29</i>
ss1867970967	15	49338778	6.090	1.94E-05	LOC106571749	+	0	<i>klhl29</i>
ss1868334451	15	49359773	6.273	1.09E-05	LOC106571749	+	0	<i>klhl29</i>
ss1868801922	29	23144883	5.970	2.84E-05	LOC106590351	+	43840	<i>sf3a2</i>