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

Victoria L. Pritchard[1†], Hannu Mäkinen[1,2], Juha-Pekka Vähä[3*], Jaakko Erkinaro[4], Panu Orell[4] & 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 Selection 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 potentially 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 could 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, and immune defence. These

included the large-effect age-at-maturity gene *vgll3*, 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 haplotype 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 upstream catchment area 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 selection; microgeographic 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). 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, & 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 helping 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 directional selection (e.g. Liu et al. 2017). One finding of such studies is that life history diversification, in particular, can be underlain by single genomic regions 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, & Bardonnet, 2017). The choice between residency and anadromy in many populations of rainbow trout/steelhead (*Oncorhynchus mykiss*) is associated with a large non-recombining 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 *lrrc9* 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 *greb1l* 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 16,386km^2 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 selection and 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 and search for regions of the genome which exhibit signatures of differential selection and therefore potentially underlie fine-scale local adaptation. We compared our results to previous studies of Atlantic salmon throughout their global range. We also examined our candidate regions for genes strongly implicated in life history diversification or local adaptation in other salmonid species. The aims of this study were first, to assess whether differential local selection was acting among Teno River subpopulations; second, to examine whether candidate regions were also locally selected at broader geographic scales; and third, to identify focal genomic regions that could be the focus of future work validating their potential role in local adaptation.

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 (Finland and Norway) that are 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 (TSAR, 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. 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 minor allele frequency (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. We removed of 13 putative sibs and one individual with possible escapee ancestry. Finally, we excluded SNPs with a MAF < 0.05 or >10% missing genotypes, leaving a final juvenile dataset that comprised 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 expected heterozygosity (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 and phenotypic variation

We investigated three 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 (freshwater migration distance) 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 tested for correlations among these three environmental variables using a Kendall rank correlation test in R.

To explore the relationship between these environmental variables and known phenotypic variation among the subpopulations, we extracted values for the median run time of one-seawinter fish and the proportion of females that are multi-seawinter fish (rounded to the nearest 5%) from Vähä et al. (2011) (Table 1). We applied Kendall rank correlation tests as above.

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, Frichot, 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-kb 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 OutFlank (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 *–pr-odds* 100, kept other parameters default, and took the median value of alpha 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: We used two approaches to identify markers associated with standard-normalized environmental variables. 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 (indicated by extreme negative or positive scores). We used selscan (Szpiech & Hernandez, 2014; --*max-gap* 2Mb, all other parameters default) to estimate XP-EHH for each SNP for each of the 45 pairwise population comparisons. We standard-normalized XP-EHH across all chromosomes within each comparison (selscan function --norm), and took the maximimum absolute normalized score over all comparisons as our test statistic for each SNP.

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 clusters (*K*) and 10 EM runs (*nfit*). We confirmed that 10 was a sufficient number of haplotype clusters by performing a cross validation procedure using fastPhase 1.4 with an R wrapper script (Scheet & Stephens 2006; Khvorykh 2017), and applying 10, 15 or 20 clusters.

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 any candidate loci identified in genome scans 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 SNPs). We then 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.

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.

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.

iv) SNPs in the top set in both the XP-EHH and HapFLK analysis, but not in other analyses, were termed '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 environmentally associated or outlier SNP, and at least one SNP in the top 0.5% set of either the XP-EHH or HapFLK analysis.

For reference, results from significance tests packaged with Bayescan, OutFLANK, LFMM and HapFLK are provided in archived data files and summarized for candidate SNPs in Tables S2 and S3.

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 candididate 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 haploblock positions into bed format. To annotate candidate SNPs, we first used the BEDtools 2.26.0 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. Where a single gene was labelled with both environmentally associated SNPs and outlier and/or 'EHH' SNPs, we considered that gene to be environmentally associated.

297 *Comparison of outliers to previous results*

298 Several previous studies have used a 7K SNP array to identify regions of the genome putatively under
299 differential selection among *S. salar* populations (Bourret, Kent, et al., 2013; Bourret, Dionne, Kent,
300 Lien, & Bernatchez, 2013; Gutierrez, Yáñez, & Davidson, 2016; Jeffery et al., 2017; Liu et al., 2017;
301 Mäkinen, Vasemägi, McGinnity, Cross, & Primmer, 2015; Moore et al., 2014; Perrier, Bourret, Kent,
302 & Bernatchez, 2013); discriminating populations (Karlsson, Moen, Lien, Glover, & Hindar, 2011) or
303 associated with phenotypic traits (Gutierrez, Yáñez, Fukui, Swift, & Davidson, 2015; Johnston et al.,
304 2014). To compare our results with the location of SNPs of interest identified in these studies, we
305 obtained flanking sequences for the 7K SNPs from NCBI dbSNP
306 (www.ncbi.nlm.nih.gov/projects/SNP/) and aligned them with the *S. salar* genome using bwa-mem
307 with default parameters (Li, 2013; Li & Durbin, 2009). We only retained unambiguously mapped
308 sequences with MQ ≥ 50. We considered a 7K SNP to map to one of our candidate regions if it was
309 within the haploblock boundary.

310 To further explore one strong candidate region for local selection, we aligned *O. nerka* RAD-tag
311 sequences from the study of Veale & Russello (2017b)to the annotated *S. salar* and *O. mykiss*
312 genomes (Omyk_1.0 available: https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1). We
313 again used bwa-mem with default parameters, and retained unambiguously mapped sequences with
314 MQ ≥ 25.

315 **RESULTS**

316 *Environmental variation, phenotypic variation, and population structure*

317 Freshwater migration distance was strongly correlated with elevation (τ = 0.78, p = 0.0009), but
318 neither were correlated with upstream catchment area (τ = 0.20, p= 0.4843; τ = -0.02, p=1.000; Fig.
319 S1). We therefore retained 'Distance' and 'Catchment' as the environmental parameters in our
320 analysis. Median run time not correlated with migration distance (τ = 0.09, p = 0.7165; Fig S1), and
321 was positively but not significantly correlated with upstream catchment area (τ = 0.46, p = 0.0694; Fig

S1). Proportion of multi-seawinter females was not correlated with either environmental variable (Fig S1).

Pairwise F_{st} between samples is shown in Table S1; 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. S2). 'Subpopulation 1' overlapped with juveniles caught in the Garnjarga 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. S3. By combining results from the outlier and environmental association tests, we identified 1,534 SNPs that, under our criteria, were candidates to be linked to variants under local

selection (0.77% of total SNPs examined; 604 outlier SNPs, 437 SNPs associated with distance, 394 SNPs associated with catchment area, 99 'extended haplotype homozygosity' SNPs, Table S2). These 'candidate SNPs' occurred within 675 different haploblocks, and were associated with 531 overlapping and 286 downstream genes (Table S3; 1.9% of all genes annotated on the *S. salar* chromosomes; Outlier: 231 overlapping/135 downstream; Distance: 163 overlapping/90 downstream; Catchment: 129 overlapping/58 downstream; 'EHH': 9 overlapping/2 downstream) (Tables S2 & S3). Thirty one haploblocks (4.6%) contained SNPs on the 7K array that were identified as candidates for divergent selection or associated with traits of interest in previous Atlantic salmon studies (Table S2)

Forty-three haploblocks had evidence for local selection both from outlier and/or environmental association analyses and from haplotype frequency/homozygosity analyses (Table 2; Fig. S4; 26 supported by HapFLK, 11 supported by XP-EHH; 6 supported by both). These 'candidate haploblocks' were distributed over 18 of the 29 *S. salar* chromosomes. Two of the candidate haploblocks on Ssa04 (Fig S4.2), three (Fig S4.5) and three (Fig S4.6)on Ssa09, six on Ssa12 (Fig. S4.10_11), and two on Ssa27 (Fig. S4.31) were located closely together on the chromosome and supported by a HapFLK or XP-EHH signal that could be driven by a single selected locus; thus we considered there to be 32 independent candidate regions. Haploblock size ranged from 1.8kb to 2.4Mb.

Eight of these candidate regions included loci of interest from previous salmonid studies (Table 2). A single haploblock with particularly strong evidence for divergent selection was located on Ssa09, and contained the *lrrc9* locus shown by Veale & Russello (2017a) to be associated with spawning ecotype in *O. nerka*. Ecotype associated RAD-tag 66810 of Veale & Russello (2017b, corresponding to RAD tag 57884 of Nichols, Kozfkay, & Narum 2016) mapped to *lrrc9* within this haploblock. RAD-tag 24343, also found to be associated with spawning ecotype by Veale & Russello (2017b), Nichols, Kozfkay, & Narum (2016, RAD-tag 64477), and Larson et al. (2017, RAD-tag 41305) did not align with the *S. salar* genome, but mapped unambiguously on the *O. mykiss* genome ≈ 36kb from 68810 between the genes *lrr9* and *dhrs7*. No other (ecotype associated or non-associated) RAD-tag from

Veale & Russello (2017b) mapped within this haploblock or adjacent candidate haploblocks on Ssa09.

DISCUSSION

In this study, we used a 'bottom up' approach – examining patterns of variation across the genome - to identify loci that may be involved in fine-scale local adaptation of Atlantic salmon in a large river system. By combining results from outlier and environmental association analyses, we identified 1,480 candidate SNPs, annotated with 807 protein coding loci. While the relatively relaxed empirical p threshold ($p<0.005$ based on the observed distribution of test statistics) 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 32 regions of the genome that are particularly likely to contain loci under differential selection (Table 2, Fig. S4). Eight 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. S4). 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 haploblock 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 genes *vgll3 and akap11*) (Table 2, Fig. S4.26). 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 upstream catchment area, 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. S4.5). SNPs in this region are among the most extreme outliers in all three outlier analyses, are robustly associated with upstream catchment area 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 ≈7kb 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 allelic variant across these SNPs, and XP-EHH results indicate a positive selective sweep on this allele. Conversely, the samples from the smaller tributaries KEVO, TSAR and PULM are almost fixed for the alternate allele (Table S2). The strong relationship of allele frequencies at this location with catchment area infers that this outlying locus is also 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 S4).

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 could 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 haploblock but ≈ 75 – 100kb 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 causal locus may be implicated 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 ≈700kb distant (Table S3) but are clearly labelling our candidate locus. Further, Veale & Russello (2017a), sequencing through *lrrc9*, found that haplotype divergence increased towards our candidate region.

One gene flanking this region, *ppm1a*, is a broad-specificity enzyme whose potential selective importance in 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 an orthologue 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). 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. Many putative *six6* enhancer elements are located within our candidate region in other species. Several have directly been shown to regulate *six6*

expression in different tissues and at different developmental stages, while others remain poorly characterized (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 reducing gene flow and altering the selective landscape for the Ssa25 locus among populations – remains unknown Given the observation that the same genomic region underlies spawning site differentiation in *O. nerka*, we hypothesize that *six6* could mediate spawning site selection 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 annotated to candidate SNPs associated with upstream catchment area: *neuronal growth regulator 1* (*negr1*, Ssa10), linked to both obesity and sexual maturation in several taxa (Lee et al., 2012); *trmt11* (Ssa05); *ptprf* (Ssa10); *ntrk2* (Ssa10); and *h6st1* (Ssa29) (Table S3). 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 upstream catchment area (and run timing) from the combined LFMM and BayEnv results is located ≈ 19Mb along Ssa11 (Table S2, Table S3, Table S4). While this region is not supported as a 'candidate haploblock' given our current thresholds, it contains SNPs within the top 2% 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

480 locus is zinc finger homeobox 3 (*zfhx3*), a transcription factor expressed in the suprachiasmatic
481 nucleus with a role in mammalian circadian rhythms, including sleep (Balzani et al., 2016). This
482 might therefore be considered a candidate locus for co-selection with *six6*.

483 A candidate haploblock at on Ssa13, associated with freshwater migration distance, overlaps *neuronal*
484 *PAS domain protein 2* (*npas2*) (Table 2, Fig. S4.10). *Npas2* is a paralogue of *clock*, which is well
485 known for regulating circadian rhythms and reproductive cycles in diverse taxa. In mammals, the two
486 genes have similar roles in the suprachiasmatic nucleus, and *npas2* can compensate if *clock* is silenced
487 (DeBruyne, Weaver, & Reppert, 2007). *Clock* polymorphisms are associated with the timing of
488 reproduction in rainbow trout and chinook salmon, and show a latitudinal cline consistent with local
489 selection in the latter (Leder, Danzmann, & Ferguson, 2006; O'Malley, Camara, & Banks, 2007;
490 O'Malley & Banks, 2008). Few salmonid studies have examined *npas2*, but O'Malley, Jacobson,
491 Kurth, Dill, & Banks (2013) found that *npas2*-linked variation discriminated O. *tshawytscha*
492 populations with different migratory timing. In our analysis, however, this candidate haploblock is not
493 associated with run timing.

494 *Evidence for local selection on genes mediating energy homeostasis:* We found a large (2.3Mbp)
495 candidate haploblock on Ssa19 that was associated with freshwater migration distance (Table 2, Fig.
496 S5.20). This haploblock overlaps a region of the genome responding to domestication selection in
497 North America *S. salar* (Liu et al., 2017; Mäkinen et al., 2015), and also contains a SNP
498 differentiating wild and domestic Norwegian *S. salar* (Karlsson et al., 2011). A homologous region
499 was recently found to be associated with migratory difficulty (a function of distance and altitude gain)
500 in Arctic char (Moore et al., 2017). Following Liu et al. (2017), we hypothesize that the target of
501 selection is a regulatory element for downstream *melanocortin receptor 4* (*mc4r*). This gene, well
502 known for its association with human obesity, is a controller of energy homeostasis and somatic
503 growth in fish and other vertebrates via its influence on food intake and energy expenditure (Krashes,
504 Lowell, & Garfield, 2016; Metz, Peters, & Flik, 2006). Given its function, it is expected to be
505 differentially selected in the wild vs. domestic environment, and by varying energy requirements

during the upstream spawning migration (when Atlantic salmon do not feed). Several other obesity-associated genes are candidate selective targets, including *negr1*; four copies of *neurexin 3* (*nrxn3*, (Heard-Costa et al., 2009)Heard-Costa *et al.* 2009) on a large candidate haploblock on Ssa09 (Table 2, Fig. S4); *lingo2* (Ssa05 and Ssa09); and *arid5b* (Ssa01) (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 *vgll3* and *six6* to influence variation in life-history dynamics among Atlantic salmon populations.

A signature of directional selection on MhcII: A clear signal consistent with directional selection occurs around the single Atlantic salmon copy of the classical major histocompatibility complex (MHC) II (*dab/daa*, Ssa12, Table 2, Fig. S5.; Gómez, Conejeros, Marshall, & Consuegra, 2010). Allelic variation within this haploblock correlates with freshwater migration distance. *MhcII* initiates the adaptive immune response to pathogens by binding foreign peptides and presenting them to T cells (Piertney & Oliver, 2006). Elevated *Mhc* diversity confers broader pathogen resistance: thus, over evolutionary time, *Mhc* alleles are maintained by balancing selection (Piertney & Oliver, 2006). However, pathogen pressure within a population can generate directional selection 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). The signal of differential selection around *MhcII* supports a model of different pathogen pressures among our sampled subpopulations, despite their co-occurrence in the same drainage basin. Similar evidence for localized directional selection on *MhcII* variants has been observed in sockeye salmon at both broad and fine spatial scales (Gomez-Uchida et al., 2011; Larson, Seeb, Dann, Schindler, & Seeb, 2014; McClelland et al., 2013).

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 may be common among Atlantic salmon populations, even those in geographic proximity. This should be taken into account in stocking and restoration programs.

Summary and prospects: Here, we have identified numerous regions of the genome which exhibit signatures of differential selection among Teno River Atlantic salmon subpopulations, suggesting that these regions harbour genes involved in local adaptation at a microgeographic scale. Our results open up multiple avenues for future research, with the ultimate aim of validating candidate genes as adaptively important by linking genotype, phenotype and fitness (Barrett & Hoekstra, 2011). Initially, this could include work to further characterize these regions at the molecular genetic level, including replicate studies on different populations, identification of putative binding sites for regulatory molecules, examination of how variants influence gene transcription, and generation of comparative sequence data for other taxa. Importantly, it should also include and field and common-garden studies to identify the phenotypic effects of candidate regions -in particular those associated with the candidate genes *six6* and *mc4r* – within Atlantic salmon.

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 and Heli Junes for laboratory assistance, and several field workers at the Teno Research Station (LUKE) who helped in sampling the juvenile salmon. Computing resources were provided by *CSC – IT Center for Science Ltd* (Finland). Comments from three anonymous reviewers improved the manuscript.

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

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 *vgll3* 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 Zfhx3-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

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

Castillo, J. J., Hazlett, Z. S., Orlando, R. A., & Garver, W. S. (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

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*

606 *Fisheries and Aquatic Sciences*, cjfas-2017-0293. http://doi.org/10.1139/cjfas-2017-0293

607 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-
608 generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7.
609 http://doi.org/10.1186/s13742-015-0047-8

610 Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced
611 polymorphism and background selection on equilibrium patterns of genetic diversity in
612 subdivided populations. *Genetical Research*, *70*(2), S0016672397002954.
613 http://doi.org/10.1017/S0016672397002954

614 Clark, D. D., Gorman, M. R., Hatori, M., Meadows, J. D., Panda, S., & Mellon, P. L. (2013). Aberrant
615 development of the suprachiasmatic nucleus and circadian rhythms in mice lacking the
616 homeodomain protein *Six6*. *Journal of Biological Rhythms*, *28*(1), 15–25.
617 http://doi.org/10.1177/0748730412468084

618 Claussnitzer, M., Dankel, S. N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., … Kellis, M.
619 (2015). *FTO* obesity variant circuitry and adipocyte browning in humans. *New England Journal*
620 *of Medicine*, *373*(10), 895–907. http://doi.org/10.1056/NEJMoa1502214

621 Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic
622 parallelism and convergence in natural populations. *Proceedings of the Royal Society B:*
623 *Biological Sciences*, *279*(1749), 5039–5047. http://doi.org/10.1098/rspb.2012.2146

624 Conte, I., Marco-Ferreres, R., Beccari, L., Cisneros, E., Ruiz, J. M., Tabanera, N., & Bovolenta, P.
625 (2010). Proper differentiation of photoreceptors and amacrine cells depends on a regulatory loop
626 between NeuroD and *Six6*. *Development (Cambridge, England)*, *137*(14), 2307–17.
627 http://doi.org/10.1242/dev.045294

628 Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to
629 identify loci underlying local adaptation. *Genetics*, *185*(4).
630 http://doi.org/10.1534/genetics.110.114819

631 De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan
632 methods against more complex models: When and how much should we trust them? *Molecular*
633 *Ecology*, *23*(8), 2006–2019. http://doi.org/10.1111/mec.12705

634 DeBruyne, J. P., Weaver, D. R., & Reppert, S. M. (2007). CLOCK and NPAS2 have overlapping
635 roles in the suprachiasmatic circadian clock. *Nature Neuroscience*, *10*(5), 543–545.
636 http://doi.org/10.1038/nn1884

637 Dionne, M., Miller, K. M., Dodson, J. J., & Bernatchez, L. (2009). MHC standing genetic variation
638 and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal*
639 *Society B: Biological Sciences*, *364*(1523), 1555–1565. http://doi.org/10.1098/rstb.2009.0011

640 Dodson, J. J., Aubin-Horth, N., Thériault, V., & Páez, D. J. (2013). The evolutionary ecology of
641 alternative migratory tactics in salmonid fishes. *Biological Reviews*, *88*(3), 602–625.
642 http://doi.org/10.1111/brv.12019

643 Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of
644 an expanding population. *Proceedings of the National Academy of Sciences*, *101*(4), 975–979.
645 http://doi.org/10.1073/pnas.0308064100

646 Elmer, K. R. (2016). Genomic tools for new insights to variation, adaptation, and evolution in the
647 salmonid fishes: a perspective for charr. *Hydrobiologia 783*, 191. http://doi.org/10.1007/s10750-
648 015-2614-5

649 Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of
650 selection through haplotype differentiation among hierarchically structured populations.
651 *Genetics*, *193*(3). http://doi.org/10.1534/genetics.112.147231

652 Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both
653 dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977–93.
654 http://doi.org/10.1534/genetics.108.092221

655 Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. (2011). Extent and scale of
656 local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, *106*(3), 404–420.
657 http://doi.org/10.1038/hdy.2010.167

658 Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research 23*, 1089–
659 1096

660 Frazer, K. K., & Russello, M. A. (2013). Lack of parallel genetic patterns underlying the repeated
661 ecological divergence of beach and stream-spawning kokanee salmon. *Journal of Evolutionary*
662 *Biology*, *26*(12), 2606–2621. http://doi.org/10.1111/jeb.12250

663 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). fast and efficient
664 estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.
665 http://doi.org/10.1534/genetics.113.160572

666 Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between
667 loci and environmental gradients using latent factor mixed models. *Molecular Biology and*
668 *Evolution*, *30*(7), 1687–1699. http://doi.org/10.1093/molbev/mst063

669 Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., …
670 Quinn, T. P. (2007). A critical review of adaptive genetic variation in Atlantic salmon:
671 implications for conservation. *Biological Reviews of the Cambridge Philosophical Society*,
672 *82*(2), 173–211. http://doi.org/10.1111/j.1469-185X.2006.00004.x

673 Gómez, D., Conejeros, P., Consuegra, S., & Marshall, S. H. (2011). MHC mediated resistance to
674 Piscirickettsia salmonis in salmonids farmed in Chile. *Aquaculture*, *318*(1–2), 15–19.
675 http://doi.org/10.1016/j.aquaculture.2011.04.023

676 Gómez, D., Conejeros, P., Marshall, S. H., & Consuegra, S. (2010). MHC evolution in three salmonid
677 species: a comparison between class II alpha and beta genes. *Immunogenetics*, *62*(8), 531–542.
678 http://doi.org/10.1007/s00251-010-0456-x

679 Gomez-Uchida, D., Seeb, J. E., Smith, M. J., Habicht, C., Quinn, T. P., & Seeb, L. W. (2011). Single
680 nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among
681 Alaskan sockeye salmon (Oncorhynchus nerka) populations. *BMC Evolutionary Biology*, *11*(1),
682 48. http://doi.org/10.1186/1471-2148-11-48

683 Grammes, F. (2016). Ssa.RefSeq.db: Annotation package for the *Salmo salar* RefSeq (ICSASG_v2).
684 R package version 1.2. Available: github.com/FabianGrammes/Ssa.RefSeq.db.

685 Gunther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies.
686 *Genetics*, *195*(1), 205–220. http://doi.org/10.1534/genetics.113.152462

687 Gutierrez, A. P., Yáñez, J. M., & Davidson, W. S. (2016). Evidence of recent signatures of selection
688 during domestication in an Atlantic salmon population. *Marine Genomics*, *26*, 41–50.
689 http://doi.org/10.1016/j.margen.2015.12.007

690 Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., & Davidson, W. S. (2015). Genome-wide
691 association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon
692 (*Salmo salar*). *PloS One*, *10*(3), e0119730. http://doi.org/10.1371/journal.pone.0119730

693 Heard-Costa, N. L., Zillikens, M. C., Monda, K. L., Johansson, Å., Harris, T. B., Fu, M., … North, K.
694 E. (2009). NRXN3 is a novel locus for waist circumference: A genome-wide association study
695 from the CHARGE consortium. PLoS Genetics, 5(6), e1000539.
696 http://doi.org/10.1371/journal.pgen.1000539

697 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

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

Karlsson, S., Moen, T., Lien, S., Glover, K. A., & Hindar, K. (2011). Generic 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

Khvorykh, G. (2017) Inzilico/kselection v1.0. Available: github.com/inzilico/kselection. doi:10.5281/zenodo.1173276

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), 431–41. 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

744 Krashes, M. J., Lowell, B. B., & Garfield, A. S. (2016). Melanocortin-4 receptor–regulated energy
745 homeostasis. *Nature Neuroscience*, *19*(2), 206–219. http://doi.org/10.1038/nn.4202

746 Kusche, H., Côté, G., Hernandez, C., Normandeau, E., Boivin-Delisle, D., & Bernatchez, L. (2017).
747 Characterization of natural variation in North American Atlantic salmon populations
748 (Salmonidae: *Salmo salar*) at a locus with a major effect on sea age. *Ecology and Evolution*,
749 *7*(15), 5797–5807. http://doi.org/10.1002/ece3.3132

750 Langefors, a, Lohm, J., Grahn, M., Andersen, O., & von Schantz, T. (2001). Association between
751 major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in
752 Atlantic salmon. *Proceedings. Biological Sciences / The Royal Society*, *268*(1466), 479–485.
753 http://doi.org/10.1098/rspb.2000.1378

754 Larder, R., Clark, D. D., Miller, N. L. G., & Mellon, P. L. (2011). Hypothalamic dysregulation and
755 infertility in mice lacking the homeodomain protein *Six6*. *The Journal of Neuroscience : The*
756 *Official Journal of the Society for Neuroscience*, *31*(2), 426–38.
757 http://doi.org/10.1523/JNEUROSCI.1688-10.2011

758 Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., & Seeb, L. W. (2017).
759 Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular*
760 *Ecology*, *26*(2), 554–570. http://doi.org/10.1111/mec.13933

761 Larson, W. A., Seeb, J. E., Dann, T. H., Schindler, D. E., & Seeb, L. W. (2014). Signals of
762 heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye
763 salmon. *Molecular Ecology*, *23*(22), 5448–5461. http://doi.org/10.1111/mec.12949

764 Leder, E. H., Danzmann, R. G., & Ferguson, M. M. (2006). The candidate gene, clock, localizes to a
765 strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity*, *97*(1),
766 74–80. http://doi.org/10.1093/jhered/esj004

767 Ledford, K. L., Martinez-De Luna, R. I., Theisen, M. A., Rawlins, K. D., Viczian, A. S., & Zuber, M.
768 E. (2017). Distinct cis-acting regions control *six6* expression during eye field and optic cup
769 stages of eye formation. *Developmental Biology*, *426*(2), 418–428.
770 http://doi.org/10.1016/j.ydbio.2017.04.003

771 Lee, A. W. S., Hengstler, H., Schwald, K., Berriel-Diaz, M., Loreth, D., Kirsch, M., … Schäfer, M. K.
772 E. (2012). Functional inactivation of the genome-wide association study obesity gene Neuronal
773 Growth Regulator 1 in mice causes a body mass phenotype. *PLoS ONE*, *7*(7), e41537.
774 http://doi.org/10.1371/journal.pone.0041537

775 Lee, B., Rizzoti, K., Kwon, D. S., Kim, S. Y., Oh, S., Epstein, D. J., … Jeong, Y. (2012). Direct
776 transcriptional regulation of *Six6* is controlled by SoxB1 binding to a remote forebrain enhancer.
777 *Developmental Biology*, *366*(2), 393–403. http://doi.org/10.1016/j.ydbio.2012.04.023

778 Lepais, O., Manicki, A., Glise, S., Buoro, M., & Bardonnet, A. (2017). Genetic architecture of
779 threshold reaction norms for male alternative reproductive tactics in Atlantic salmon (*Salmo*
780 *salar* L.). *Scientific Reports*, *7*, 43552. http://doi.org/10.1038/srep43552

781 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
782 Retrieved from http://arxiv.org/abs/1303.3997

783 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
784 *Bioinformatics (Oxford, England)*, *25*(14), 1754–60.
785 http://doi.org/10.1093/bioinformatics/btp324

786 Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., … Davidson, W. S. (2016).
787 The Atlantic salmon genome provides insights into rediploidization. *Nature*, *533*(7602), 200–
788 205. http://doi.org/10.1038/nature17164

789 Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for
790 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

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

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

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*

838 *Royal Society B: Biological Sciences*, *275*(1653), 2813–2821.
839 http://doi.org/10.1098/rspb.2008.0524

840 O'Malley, K. G., Camara, M. D., & Banks, M. A. (2007). Candidate loci reveal genetic differentiation
841 between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*).
842 *Molecular Ecology*, *16*(23), 4930–4941. http://doi.org/10.1111/j.1365-294X.2007.03565.x

843 O'Malley, K. G., Jacobson, D. P., Kurth, R., Dill, A. J., & Banks, M. A. (2013). Adaptive genetic
844 markers discriminate migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*) amid
845 continued gene flow. *Evolutionary Applications*, *6*(8), 1184–1194.
846 http://doi.org/10.1111/eva.12095

847 O'Toole, C. L., Reed, T. E., Bailie, D., Bradley, C., Cotter, D., Coughlan, J., … McGinnity, P. (2015).
848 The signature of fine scale local adaptation in Atlantic salmon revealed from common garden
849 experiments in nature. *Evolutionary Applications*, *8*(9), 881–900.
850 http://doi.org/10.1111/eva.12299

851 Pearse, D. E., Miller, M. R., Abadía-Cardoso, A., & Garza, J. C. (2014). Rapid parallel evolution of
852 standing variation in a single, complex, genomic region is associated with life history in
853 steelhead/rainbow trout. *Proceedings of the Royal Society of London B: Biological Sciences*,
854 *281*(1783). http://doi.org/10.1098/rspb.2014.0012

855 Perrier, C., Bourret, V., Kent, M. P., & Bernatchez, L. (2013). Parallel and nonparallel genome-wide
856 divergence among replicate population pairs of freshwater and anadromous Atlantic salmon.
857 *Molecular Ecology*, *22*(22), 5577–5593. http://doi.org/10.1111/mec.12500

858 Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility
859 complex. *Heredity*, *96*(1), 7–21. http://doi.org/10.1038/sj.hdy.6800724

860 Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., … Miller,
861 M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the
862 utility of genomics for informing conservation. *Science Advances*, *3*(8), e1603198.
863 http://doi.org/10.1126/sciadv.1603198

864 Pritchard, V. L., Erkinaro, J., Kent, M. P., Niemelä, E., Orell, P., Lien, S., & Primmer, C. R. (2016).
865 Single nucleotide polymorphisms to discriminate different classes of hybrid between wild
866 Atlantic salmon and aquaculture escapees. *Evolutionary Applications*, *9*(8), 1017–1031.
867 http://doi.org/10.1111/eva.12407

868 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C.
869 (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses.
870 *American Journal of Human Genetics*, *81*(3), 559–75. http://doi.org/10.1086/519795

871 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic
872 features. *Bioinformatics*, *26*(6), 841–842. http://doi.org/10.1093/bioinformatics/btq033

873 Quinn, T. P., McGinnity, P., Reed, T. E., & Bradford, M. (2016). The paradox of "premature
874 migration" by adult anadromous salmonid fishes: patterns and hypotheses. *Canadian Journal of*
875 *Fisheries and Aquatic Sciences*, *73*(7), 1015–1030. http://doi.org/10.1139/cjfas-2015-0345

876 R Core Team (2015). R: A Language and Environment for Statistical Computing. *R Foundation for*
877 *Statistical Computing Vienna Austria ISBN*, *1*(10), 1–2896.

878 Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Stewart, J. (2007).
879 Genome-wide detection and characterization of positive selection in human populations. *Nature*,
880 *449*(7164), 913–918. http://doi.org/10.1038/nature06250

881 Scheet, P, &Stephens, M (2006). A fast and flexible statistical model for large-scale population
882 genotype data: applications to inferring missing genotypes and haplotypic phase. *American*
883 *Journal of Human Genetics*, *78*(4):629-644. http://doi.org/10.1086/502802

884 Seo, H. C., Drivenes, Ø., Ellingsen, S., & Fjose, A. (1998). Expression of two zebrafish homologues
885 of the murine Six3 gene demarcates the initial eye primordia. *Mechanisms of Development*,
886 *73*(1), 45–57. http://doi.org/10.1016/S0925-4773(98)00028-8

887 Szpiech, Z. A., & Hernandez, R. D. (2014). Selscan: An efficient multithreaded program to perform
888 EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827.
889 http://doi.org/10.1093/molbev/msu211

890 Thornqvist, P.-O., Hoglund, E., & Winberg, S. (2015). Natural selection constrains personality and
891 brain gene expression differences in Atlantic salmon (*Salmo salar*). *Journal of Experimental*
892 *Biology*, *218*(7), 1077–1083. http://doi.org/10.1242/jeb.114314

893 Thorpe, J. E., Mangel, M., Metcalfe, N. B., & Huntingford, F. A. (1998). Modelling the proximate
894 basis of salmonid life-history variation, with application to Atlantic salmon, *Salmo salar* L.
895 *Evolutionary Ecology*, *12*(5), 581–599. http://doi.org/10.1023/A:1022351814644

896 Toy, J., Yang, J. M., Leppert, G. S., & Sundin, O. H. (1998). The optx2 homeobox gene is expressed
897 in early precursors of the eye and activates retina-specific genes. *Proceedings of the National*
898 *Academy of Sciences of the United States of America*, *95*(18), 10643–8.
899 http://doi.org/10.1073/pnas.95.18.10643

900 Watanabe, N., Itoh, K., Mogi, M., Fujinami, Y., Shimizu, D., Hashimoto, H., … Suzuki, T. (2012).
901 Circadian pacemaker in the suprachiasmatic nuclei of teleost fish revealed by rhythmic period2
902 expression. *General and Comparative Endocrinology*, *178*(2), 400–407.
903 http://doi.org/10.1016/j.ygcen.2012.06.012

904 Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2015). Detection of selective sweeps in structured
905 populations: a comparison of recent methods. *Molecular Ecology*, *25*(1), 89–103.
906 http://doi.org/10.1111/mec.13360

907 Veale, A. J., & Russello, M. A. (2017a). An ancient selective sweep linked to reproductive life history
908 evolution in sockeye salmon. *Scientific Reports*, *7*(1), 1747. http://doi.org/10.1038/s41598-017-
909 01890-2

910 Veale, A. J., & Russello, M. A. (2017b). Genomic changes associated with reproductive and
911 migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). *Genome Biology and Evolution*.
912 http://doi.org/10.1093/gbe/evx215

913 Vilas, A., Pérez-Figueroa, A., & Caballero, A. (2012). A simulation study on the performance of
914 differentiation-based methods to detect selected loci using linked neutral markers. *Journal of*
915 *Evolutionary Biology*, *25*(7), 1364–1376. http://doi.org/10.1111/j.1420-9101.2012.02526.x

916 Vähä, J.-P. P., Erkinaro, J., Niemelä, E., & Primmer, C. R. (2007). Life-history and habitat features
917 influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, *16*(13),
918 2638–2654. http://doi.org/10.1111/j.1365-294X.2007.03329.x

919 Vähä, J.-P. P., Erkinaro, J., Niemelä, E., Primmer, C. R., Saloniemi, I., Johansen, M., … Brørs, S.
920 (2011). Temporally stable population-specific differences in run timing of one-sea-winter
921 Atlantic salmon returning to a large river system. *Evolutionary Applications*, *4*(1), 39–53.
922 http://doi.org/10.1111/j.1752-4571.2010.00131.x

923 Vähä, J.-P., Erkinaro, J., Falkegård, M., Orell, P., & Niemelä, E. (2017). Genetic stock identification
924 of Atlantic salmon and its evaluation in a large population complex. *Canadian Journal of*
925 *Fisheries and Aquatic Sciences*, *74*(3), 327–338. http://doi.org/10.1139/cjfas-2015-0606

926 Vähä, J.-P., Erkinaro, J., Niemelä, E., & Primmer, C. R. (2008). Temporally stable genetic structure
927 and low migration in an Atlantic salmon population complex: implications for conservation and
928 management. *Evolutionary Applications*, *1*(1), 137–154. http://doi.org/10.1111/j.1752-
929 4571.2007.00007.x

930 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population

931 structure. *Evolution*, *38*(6), 1358. http://doi.org/10.2307/2408641

932 Whitlock, M., & Lotterhos, K. (2015). Reliable detection of loci responsible for local adaptation:
933 Inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist, 186*
934 *(Suppl 1),* S24-36. http://doi.org/10.1086/682949

935 Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.

936 Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-
937 selection balance. *Evolution*, *65*(7), 1897–1911. http://doi.org/10.1111/j.1558-
938 5646.2011.01269.x

939 Zhang, L., Mazo-Vargas, A., & Reed, R. D. (2017). Single master regulatory gene coordinates the
940 evolution and development of butterfly color and iridescence. *Proceedings of the National*
941 *Academy of Sciences of the United*
942

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.

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. See Table 1 for details.

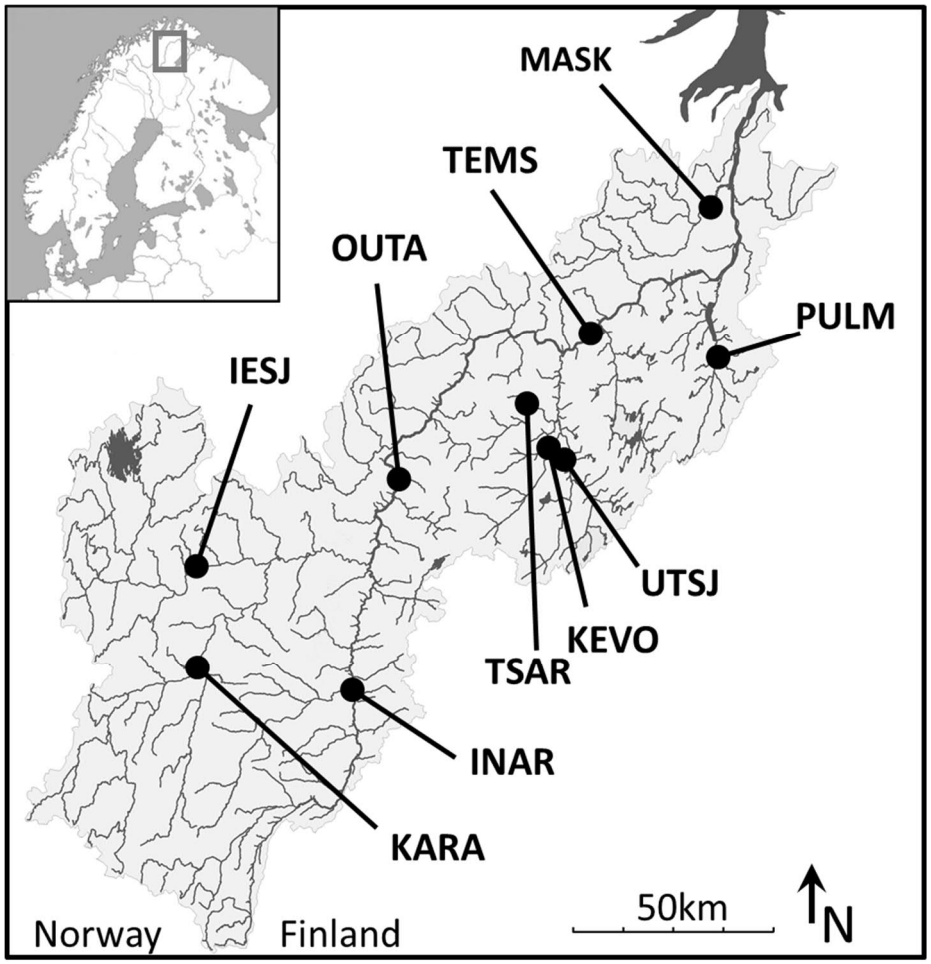

Table 1: Sample site environmental and phenotypic variables. MSW: multi-seawinter.

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. 'Region#' indexes the 32 genomic inferred to be under diversifying selection, which may include >1 haploblock. 'Haplo' indicates whether the selective signal was supported by XP-EHH, HapFLK, or both. 'Type' indicates whether a haploblock is an outlier or environmentally associated, based on the consensus SNP signal; 'Annotated Genes': NCBI annotated gene products.

Region#	Chr	Start	End	Haplo	Type	Annotated genes	Notes
1	ssa03	54440792	54662849	HapFLK	Outlier	3-mercaptopyruvate sulfurtransferase-like (mpst, 2 copies); beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase-like (mgat3, 2copies); C1q and tumor necrosis factor related protein 6 (c1qtnf6); interleukin 2 receptor, beta (il2rb); mitochondrial dynamics protein mid51; SUMO-conjugating enzyme UBC9-like (ube2i); testis expressed 33 (tex33); LOC106600886.	
2	ssa04	49356149	49385031	XPEHH	Catchment	uromodulin-like 1 (umod); zinc finger and BTB domain-containing protein 21-like (zbtb2); autophagy-related protein 101-like (atg101).	
2	ssa04	49548061	49670244	HapFLK, XPEHH	Outlier	protein btg3-like; coxsackievirus and adenovirus receptor homolog (cxadr)	Outlier among N. American S. salar populations[1]
3	ssa05	20377175	20401922	XPEHH	Catchment	ecto-NOX disulfide-thiol exchanger 2-like (enox2)	
4	ssa05	39939921	39945242	HapFLK, XPEHH	Outlier	5-hydroxytryptamine receptor 4-like (htr4)	5-hydroxytryptamine mediates expression of behavioural syndromes in S. salar fry[2]
5	ssa09	24690259	24966282	HapFLK, XPEHH	Catchment	reticulon 1(rtn1); leucine-rich repeat-containing protein 9-like (lrrc9); pecanex-like 4 (pcnx4); dehydrogenase/reductase (SDR family) member 7 (dhrs7); protein phosphatase 1A-like (ppm1a); SIX homeobox 6 (six6); piggyBac transposable element-	Associated with age-at-maturity and run timing in S. salar [3,4] Outlier among N. American S. salar populations[1,5]

						derived protein 4-like (pgbd4); SIX homeobox 1 (six1); SIX homeobox 4 (six4)	Associated with spawning location in O. nerka[6]
5	ssa09	24998704	25337622	HapFLK, XPEHH	Catchment	MNAT CDK-activating kinase assembly factor 1 (mnat1); solute carrier family 38, member 6 (slc38a6); protein kinase C eta type (prkch); hypoxia-inducible factor 1-alpha-like (hif1a); sacsin-like (sacs); zinc finger protein 239-like (znf239); WD repeat domain 89 (wdr89); Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon (ppp2r5e)	
5	ssa09	25427680	25766954	HapFLK, XPEHH	Outlier	potassium voltage-gated channel subfamily H member 5-like (kcnh5); glycoprotein hormone beta 5 (gphb5); protein jagged-2-like (jag2).	
6	ssa09	26292242	26359364	HapFLK	Distance	protein lbh-like; neuroblast differentiation-associated protein AHNAK-like (ahnak)	
6	ssa09	26371376	26399217	HapFLK	Distance	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1-like (plcb1); thioredoxin-related transmembrane protein 1-like (tmx1)	
6	ssa09	26461807	26480728	HapFLK	Distance	suppressor of Ty 3 homolog (supt3h); runt-related transcription factor 2 (runx2); chloride intracellular channel 5 (clic5)	
7	ssa09	42158973	43699365	HapFLK	Distance	neurexin-3-like (nrxn3, 4 copies); LOC106611372; SRA stem-loop-interacting RNA-binding protein, mitochondrial-like (slirp); alkB homolog 1 (alkbh1), histone H2A dioxygenase; serine palmitoyltransferase 2-like (sptlc2, 2 copies); isthmin-2-like (ism2); gastrula zinc finger protein xlcgf42.1-like; protein fam186a-like; zinc finger protein 239-like (znf239); zinc finger protein 585A-	

						like (znf585a); emilin-1-like; synaptosomal-associated protein 25-B-like (snap25b).	
8	ssa10	14054529	14098744	XPEHH	Catchment	growth arrest and DNA damage-inducible protein gadd45b-like; survival motor neuron protein-like (smn); microtubule-associated protein 1B-like (map1b).	
9	ssa10	19419205	19484226	HapFLK	Outlier	spindlin-Z-like (spin1); protein qil1-like; hydroxysteroid 11-beta-dehydrogenase 1-like protein (hsd11b1l).	
10	ssa12	60592476	60953722	XPEHH	Catchment	nuclear ubiquitous casein and cyclin-dependent kinase substrate 1-like (nucks1); solute carrier family 45 member 3-like (slc45a3); Krueppel-like factor 15 (klf15); ETS domain-containing protein Elk-4-like (elk4); major facilitator superfamily domain-containing protein 4-A-like (mfsd4a); cyclin-dependent kinase 18-like (cdk18); kelch domain containing 8A (klhdc8a); solute carrier family 41 member 1-like (slc41a1).	
11	ssa12	60985380	61042362	HapFLK	Distance	ethanolamine kinase 1-like (etnk1); transcription factor sox-13-like.	
11	ssa12	61059163	61122037	HapFLK	Distance	small nuclear ribonucleoprotein E (sme); cell division control protein 42 homolog (cdc42); LOC10656561	
11	ssa12	61136096	61165293	HapFLK	Distance	LOC106565687	
11	ssa12	61343872	61703592	HapFLK	Distance	ras-related protein rab-40c-like; myoD family inhibitor-like (mdfi); transmembrane protein 183A-like (tmem183a); transcription factor EB-like (tfeb); MHC class II alpha chain (daa) & beta chain (dab).	Associated with age-at-maturity in S. salar [7]

11	ssa12	61735215	61917647	HapFLK	Distance	proline-rich receptor-like protein kinase perk8; forkhead box protein P4-like (foxp4)	
11	ssa12	61945904	62288997	HapFLK	Distance	taste receptor type 1 member 3-like (tas1r3); ERBB receptor feedback inhibitor 1-like (errfi1); serine-rich adhesin for platelets-like (srap); 28S ribosomal protein S16, mitochondrial-like (mrps16); paired box protein 7 (pax7).	
12	ssa13	9604538	9606700	XPEHH	Outlier	short transient receptor potential channel 4-associated protein-like (trpc4ap)	
13	ssa13	82012742	82114498	HapFLK, XPEHH	Distance	neuronal PAS domain-containing protein 2-like (npas2); vacuolar ATPase assembly integral membrane protein vma21-like; LOC106567768; non-POU domain-containing octamer-binding protein-like (nono); apoptosis-inducing factor, mitochondrion-associated (aifm1); glycine receptor subunit alpha-4-like (glra4, downstream)	Npas2-linked variation discriminates O. tshawytscha populations with different migratory timing[8]
14	ssa14	3730963	3779453	XPEHH	Outlier	amphiphysin (amph)	
15	ssa16	69280327	69501948	HapFLK	Distance	male-specific lethal 3 homolog (msl3, downstream); SLIT and NTRK-like protein 6 (slitrk6, downstream)	
16	ssa17	6846387	6862460	XPEHH	Catchment	aryl hydrocarbon receptor 2 gamma (ahr2g, downstream)	Outlier between Baltic anadromous and landlocked S. salar[9] Ahr2 regulatory variation is associated with craniofacial differentiation between S. alpinus ecotypes[10]

17	ssa17	18258926	18332523	HapFLK	Catchment	ATP-dependent RNA helicase ddx3y-like; TDP-glucose 4,6-dehydratase (tgds); integral membrane protein gpr180-like; sodium/myo-inositol cotransporter-like (slc5a3); prostaglandin F2 receptor negative regulator-like (ptgfrn).	
18	ssa17	41531495	41549890	HapFLK	Distance	centrosomal protein of 83 kDa-like (cep83); host cell factor 2-like (hcfc2); ETS domain-containing protein elk3-like.	
19	ssa18	7977374	8045085	HapFLK	Outlier	heparan sulfate glucosamine 3-O-sulfotransferase 1-like (hs3st1)	
20	ssa18	26414764	26428879	HapFLK	Catchment	solute carrier family 22 member 7-like (slc22a7)	
21	ssa18	45989781	46042995	HapFLK	Outlier	EH domain-binding protein 1 (ehbp1); potassium channel subfamily K member 1-like (kcnk1); TAF6-like RNA polymerase II, p300/CBP-associated factor-associated factor, 65kDa (taf6l); metastasis-associated protein mta2-like.	
22	ssa19	17946362	18012584	HapFLK	Outlier	neuropilin-1a (nrp1a, downstream)	
23	ssa19	21194663	23553773	HapFLK	Distance	cadherin-12-like (cdh12); protein tweety homolog 2-like (ttyh2); BTB/POZ domain-containing protein 3-like (btbd3); coagulation factor XIII A chain-like (f13a); LOC106578580; cadherin-10-like(cdh10); cadherin-6-like (cdh6); melanocortin receptor 4-like (mc4r, downstream)	Outlier between wild and domesticated S. salar, N. America and Norway[11,12,13] Associated with migration difficulty in S. alpinus[14]
24	ssa20	43187801	43605543	HapFLK	Outlier	ATP-binding cassette sub-family G member 1-like (abcg1); signal peptidase complex subunit 2 homolog (spcs2); E3 ubiquitin-protein ligase rnf169-like; putative lipoyltransferase 2, mitochondrial (lipt2); glycerophosphodiester	

						phosphodiesterase domain containing 1 (gdpd1); transmembrane protein 120B-like (tmem120b); NADPH--cytochrome P450 reductase-like (por); RNA-binding protein fus-like; matrix metalloproteinase-28-like (mmp28); cullin-3 (cul3); dedicator of cytokinesis protein 10-like (dock10)	
25	ssa20	65599645	65628606	XPEHH	Catchment	RNA-binding protein Musashi homolog 2-like (msi2)	
26	ssa23	44140557	44159611	XPEHH	Outlier	protogenin B-like (prtgb).	
27	ssa24	1264159	1272401	HapFLK	Outlier	metal transporter cnnm4-like; interleukin-1 beta (il1b, downstream)	Cnnm4 is a strong candidate gene for local selection in O.mykiss[15]
28	ssa25	24013779	24066599	HapFLK	Catchment	GDP-fucose protein O-fucosyltransferase 2-like (pofut2, downstream)	
29	ssa25	28639461	28832410	HapFLK	Catchment	transcription cofactor vestigial-like protein 3 (vgll3); A kinase (PRKA) anchor protein 11 (akap11); tumor necrosis factor (ligand) superfamily, member 11 (tnfsf11); epithelial stromal interaction 1 (epsti1); DnaJ (Hsp40) homolog, subfamily C, member 15 (dnajc15).	Strongly associated with age-at-maturity in S. salar [3,16]
30	ssa25	35970829	36174533	XPEHH	Distance	amyloid beta A4 protein-like (app); GA-binding protein alpha chain-like (gabpa); splicing factor U2AF 35 kDa subunit-like (u2af1); cystathionine beta-synthase-like (cbs); ATP synthase subunit a-like (atp5a1); serine/threonine-protein kinase sik2-like.	
31	ssa27	9946286	9966206	HapFLK	Distance	LOC106588410; tubulin beta chain (tub2).	
31	ssa27	9990038	9991883	HapFLK	Catchment	zinc finger protein 384-like (znf384).	

32	ssa28	6256290	6284228	XPEHH	Catchment	myoferlin (myof); centrosomal protein 55 (cep55); retinol binding protein 4 (rbp4); cone cGMP-specific 3',5'-cyclic phosphodiesterase subunit alpha-like (pde6c, downstream)	Outlier between wild and domesticated Norwegian S. salar[17]

1. Moore et al., 2014; 2. Thornqvist, Hoglund, & Winberg, 2015; 3. Barson et al. 2015; 4. Cauwelier, Gilbey, Sampayo, Stradmeyer, & Middlemas, 2017; 5. Bourret, Dionne, Kent, Lien, & Bernatchez, 2013; 6. Veale & Russello 2017a; 7. Johnston et al. 2014; 8. O'Malley, Jacobson, Kurth, Dill, & Banks, 2013; 9. Bourret, Kent, et al., 2013; 10. Ahi et al., 2014; 11. Karlsson, Moen, Lien, Glover, & Hindar, 2011; 12. Mäkinen, Vasemägi, McGinnity, Cross, & Primmer, 2015; 13. Liu et al., 2017; 14. Micheletti, Matala, Matala, & Narum, 2017; 15. Moore et al., 2017; 16. Ayllon et al., 2015; 17. Gutierrez, Yáñez, & Davidson, 2016.

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 (where empirical p = SNP rank / total number of tests). The grey rectangle shows the boundaries of the associated haploblock. Black vertical lines indicate the mapping position *O. nerka* ecotype-associated RAD tags 68810 (left) and 24343 (right) (Veale & Russello, 2017b).