Supporting Information

Le Luyer et al. 10.1073/pnas.1711229114

Sampling Sites

We collected samples from the Quinsam and Capilano Rivers in British Columbia (Canada), namely (Fig. S1). These systems were chosen because they are well-suited to test specifically for the effect of rearing environment on patterns of methylation, independent of the genetic background between fish born in the wild (thereafter natural origin) vs. those born in hatchery. Indeed, hatchery programs for these rivers were developed as so called "integrated programs," meaning that they are based on local populations and involve spawning in the two environments (hatchery and wild). Thus, hatchery and natural origin fish in each river are not kept separate; hatchery origin fish spawn in both the hatchery and the natural habitat as do natural origin fish. This is made possible by the fact that both hatcheries are run as "swim-in" facilities, meaning that the hatchery water flows through a stream/channel to the river. Following hatching and a rearing period of 2 y, juvenile salmon (thereafter smolts) emigrate and imprint on this hatchery water. While mostly hatchery origin fish tend to "home" back to the hatchery collecting ponds through the channel, nothing prevents gene flow between hatchery and natural origin adults since the latter can also enter the hatchery channel to reproduce. Similarly, nothing prevents hatchery origin salmon to spawn elsewhere in the watershed.

Given this setting, the prediction is that there should be one single panmictic population in each river (that spawns in two different environments), which should result in no overall genetic differentiation between hatchery and wild salmon. Therefore, observed differences in patterns of methylation between wild and hatchery fish would be very unlikely to be associated with genetically distinct populations and should instead reflect the effect of rearing environment. The only chance for differential hatchery/natural selection to occur and to cause changes in allele frequencies within a single generation would be in the eggs/juveniles between adult spawning and emigration of their juvenile (smolt) progeny.

Hatchery Procedures and Sampling

The Salmon Enhancement Program (SEP) hatcheries have standard operating procedures employed across hatcheries, with the primary production strategy (PPS) being used for coho salmon at both Capilano and Quinsam hatcheries. The PPS integrates major improvements with the aim to lessen the genetic and ecological impacts of hatchery releases in the wild, mainly by using local broodstock, maintaining seasonal timing, using an enriched environment, low density, and by promoting volitional release and imprinting (1). Briefly, adult spawning peaks occurred in late October to early November with random single-pair breeding. Pre-eye eggs are kept in troughs with low flow for ∼3 mo and are moved at the eyed stage to heath trays through the alevin stage. The fry are moved to groundwater raceway ponds in April and held until the following late fall, where they are marked and moved to larger gravel substrate ponds through the parr stage until smoltification, which is the physiological change required to allow a salmon to survive in salt water. Coho yearling smolts, as defined 1+ year after hatching, are released over a month. In this study, the progeny of fall-run 2012 Capilano and Quinsam River adult coho salmon were released in each respective river as yearling smolts in 2014. These two river systems were chosen because they are well-suited to test specifically for the effect of rearing environment on patterns of methylation, independent of the genetic background between fish born in the wild and those born in hatchery.

Authorization to collect samples in the study was provided by a scientific license issued under the provisions of the Fisheries Act passed by the Canadian Parliament in 1985 and last amended in 2016. Under the Act, the scientific license was issued by Fisheries and Oceans Canada to allow departmental staff to collect samples in the course of their work. As there is no requirement for an Institutional Animal Care and Use Committee (IACUC) or equivalent under the Act, sampling protocols were neither vetted nor approved by an IACUC.

DNA Extraction and Reduced-Representation Bisulfite Sequencing Library Preparation

The RRBS library preparation was adapted from a protocol described elsewhere (2). We used DNeasy blood and tissue kit columns for the extraction of genomic DNA (gDNA) following protocol guidelines (QIAGEN) and Quant-iT PicoGreen dsDNA assays for DNA quantification (Fluoroskan Ascent FL; ThermoFisher Scientific). Genomic DNA (600 ng) was digested using MspI restriction enzymes, cleaned using magnetic beads (volume ratio 1.8×), and rinsed twice with ethanol 80%. End-repair and A-tailing steps were conducted in solution containing Klenow fragments (5 U; New England Bio Labs, Inc.) and dNTPs (0.5 mM dATP, 0.05 m dGTP, 0.05 mM dCTP) with an incubation step (30 °C for 30 min and 37 °C for 20 min). The solution was cleaned with magnetic beads (ratio 1.8×) and rinsed twice with ethanol 80%. NEXTflex illumina DNA barcodes (Bio Scientific) were ligated using overnight incubation at 16 °C in a master mix solution containing the T4 ligase buffer (1×; New England Bio Labs, Inc.) and a T4 ligase (2,000 U). The solution was cleaned with magnetic beads (ratio 1×) and rinsed twice with ethanol 80%. A size-selection step was necessary for capturing specific fragment length (200 to 400 bp), which implied two rinses with magnetic beads (first wash 0.7× and beads were discarded, second wash ratio 0.15×). A bisulfite conversion treatment was conducted following EZ DNA Methylation-Gold Kit protocol recommendations (Zymo Research). Library quality and quantity were verified on a HiSense DNA-Bioanalyzer 2100 chip (Agilent) and with Quant-iT PicoGreen assays (Fluoroskan Ascent FL; Thermo Labsystems), respectively. Libraries were sequenced on a HiSEq. 2000 platform (five individuals by lane) at the McGill University and Génome Québec Innovation Centre (Montréal, QC) using a 100-bp single-end reads approach. In parallel, sex information was inferred by PCR using a method previously described for salmonids (sdY_E2S2 5′-GTGGAGT-ACTGCG-AAGAGGAGGT-3′ and sdY_E2AS4 5′-CTTAAAACCACTCCA-CCCTCCAT-3′ primers) (3). Sex information for each individual is available in Table S4.

Methylation Calling

The RRBS reads were first trimmed using Cutadapt software (4) implemented in Trim_Galore! v0.4.1 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/). After quality filtering and sample quality check, one sample from the Quinsam River hatchery was identified as an outlier in a principal component analysis (PCA) (axes 1 to 4, drives axis 3 alone) and statistically (Grubbs test, G = 3.37, U = 0.7, P value = 0.005179). This sample was characterized by a low coverage (Methylkit *"GetCoverageStats"* function) (5) and a lower total number of unique mapped sequences compared with the other samples, which most likely occurred during library preparation. To avoid the possibility of falsely interpreting existing C-T DNA polymorphism as epigenetic variation, we then masked the draft genome assembly of the coho salmon (GenBank assembly accession no. GCA_002021735.1) for all C-T polymorphism (1,896,050 SNPs markers; maf = 0.05) identified with whole-genome resequencing of 20 salmon from five British Columbia rivers (Fig. 1)

using BEDtools v2.26.0 (6). Both mapping and genotyping were conducted with the BISulfite-seq CUI Toolkit (BISCUIT) suite (7) with default parameters, including an additional prefiltering step for low mapping quality (MAPQ < 10) conducted with SAMtools v1.19 (8).

For methylation calling, trimmed reads were mapped against the masked coho salmon genome using Bismark v0.14.5 aligner (9). Only methylation information for cytosine in a CpG context was extracted using the Bismark "*methyl extraction*" function (9). Only bases with at least 10× coverage over all individuals were conserved for subsequent analysis, and methylation levels on normalized count across individuals were compiled using a tiling window approach (step = 1,000 bp; size = 1,000 bp) with the MethylKit R package (5). Finally, a logistic regression, with the river of origin and sex as covariates, was conducted to identify differentially methylated regions (DMRs) with the "*calculateDiffMeth*" function implemented in the MethylKit R package (5). The DMRs were retained when showing at least 15% of difference between treatment, q-value < 0.001, and when a given 1,000-bp region comprised at least three CpGs.

Population and Rearing Environment Effect on DMR Analysis

We first computed a Euclidian distance matrix on the 131,807 regions and performed a principal coordinates analysis (PCoA) on this Euclidian distance matrix. PCo factors showing a relative eigenvalue higher than 2.75%, which correspond to the meaningful axes based on a broken-stick distribution, were selected as surrogate for multilocus epigenotypes (10). The Euclidean distance and the PCoA were obtained, respectively, with the functions "*daisy*" and "*pcoa*" available in the ape R package (11). To test for the effect of population (river of origin) and rearing environment (NOR or HOR), a distance-based redundancy analysis (db-RDA) was also produced with the retained PCo factors (n = 6) as a response matrix and the variables river of origin, captivity, and sex as the explanatory factors. We first produced a stepwise model selection on variables river of origin, rearing environment (natural or hatchery), and sex using the function "*ordistep*" in the vegan R package (12). Partial db-RDAs were produced to test for the effect of rearing environment or river of origin alone, after controlling for the remaining variables. The effect of a given factor was considered significant when the P value was <0.05.

Functional Annotation

For the functional analysis, we first mapped the multitissue reference transcriptome of the coho salmon (13) to the draft genome assembly (GenBank assembly accession no. GCA_002021735.1) using the splicing-tolerant Gmap (release 2016-11-07) aligner (14), and we kept the primary path for subsequent analysis. We used the GenomicRanges R package to find overlap between DMRs and genes (allowing 5 kb up- and downstream of genes) (15). The choice of 5 kb corresponds to the window used in a previous study to locate genes in the vicinity of CpG islands regions (16). Furthermore, detailed information on functional annotation based on gene location information was extracted following methods similar to a previous study on DNA methylation in rainbow trout to enhance comparison (17). The authors defined regions as follows: 3′ UTR < 1.5 kb from transcription start site (TSS); 5′ UTR < 5 kb from TSS, gene body equivalent to the longest ORF identified with the "*LongOrfs*" TransDecoder's function (18). In complement, we used the Gardiner–Garden masked approach described by Bock et al. (19) for CpG islands (CpGi) annotation based on the repeat-masked reference genome, considering CpGs shores (<2 kb up- and downstream of CpGi) and CpG shelves (2 to 4 kb up and downstream of CpGi). For this purpose, the genome was masked for repeats using WindowMasker with default parameters (20). Functional annotations are reported in Table S1 (gene body > UTRs > CpGs > CpGi shores >CpGi shelves). We then used a blastx approach against a Swissprot database (21, 22) to annotate transcript sequences. Only matches with an e-value < 1e−6 were considered. Finally, we used GOAtools v0.6.10 (23) implemented in GO enrichment workflow (https://github.com/enormandeau/go_enrichment) to identify gene ontology and assess gene enrichment. Our background list included genes that could have been localized in the vicinity of methylated regions being considered (at least 3 CpG in a 1,000-bp region and 10× coverage) and represented a total of 20,817 genes. Only GO terms with P value < 0.05 and including at least three genes were considered (Dataset S1).

Genotyping for Genetic Data

Variation at the genome level was also quantified between HOR and NOR fish in each river to test the prediction that coho salmon in each river should comprise a single panmictic population, which should translate in the overall absence of genetic differences between HOR and NOR populations within a river, as well as similar measures of genetic diversity. However, significant population structure and associated genetic differences are expected between rivers. For population genomics analysis, only biallelic markers with minimum and maximum depth of coverage between 5× and 100×, minor allele frequency (maf) > 0.05, minimum quality of 5, maximum missing of 20%, and in Hardy–Weinberg equilibrium (P value > 0.05) were conserved. Markers with statistical linkage disequilibrium (LD) above R^2 0.8 were also orphaned (one SNP dropped) (24). All filtering was conducted with VCFtools v0.1.14 software (25). From the initial 12,375,758 SNPs, only 15,044 were retained for subsequent population genomics analysis after applying these filtering criteria.

Genomic Differentiation Between Hatchery and Wild Origin Fish from Each River

Similarly to DMR analysis, we first computed a Euclidian distance matrix using the 15,044 filtered SNPs that were used to perform a principal coordinates analysis (PCoA). This procedure allowed us to avoid any effect of imputing data on the genotype database, even with our observed low global missing genotypes (2.09%) (10). Because no axis could be selected according to the broken-stick distribution, we selected all axes explaining at least 2.75% of the variation (10 axes explaining 33.9% of the variance), as previously performed with epigenetic markers (10). The Euclidean distance and the PCoA were obtained, respectively, with the functions "*daisy*" and "*pcoa*" available in the ape R package (11). We first produced a stepwise model selection on variables river of origin, rearing environment (natural or hatchery), and sex using the function "*ordistep*" in the vegan R package (12). A distance-based redundancy analysis (db-RDA) was then produced with the retained PCo factors (n = 10) as the response matrix and the variables population, rearing environment, and sex as the explanatory matrix. A db-RDA is a variant of canonical correlation analysis (CCA), which gives less weight to rare alleles that can be caused by genotypic error (26). Partial db-RDAs were produced to test for the effect of sex or river of origin alone, after controlling for the other variables. The effect of a given factor was considered significant when the P value was < 0.05. Pairwise genetic differentiation (Fst) was quantified with the multilocus AMOVAs function implemented in GENODIVE v2.0b27 software with 999 iterations (27, 28) (Table S2). Individual coefficients of inbreeding (Gis) and observed and expected heterozygosity within samples were estimated using GENODIVE v2.0b27 (27) (Table S3).

Local Selection on Loci of Major Effect and Polygenic Selection

To detect outlier loci between sexes (Fig. S1) and test for possible selective effect within a single generation between HOR and NOR fish within each river (Fig. S2), we first conducted a standard genome scan approach. We ran Bayescan v1.2 (29) on the 15,044 markers with 5,000 iterations and a burn-in length of 100,000. A permissive value of prior-odds of 10 was chosen to ensure detecting

all putative outliers. No outlier loci associated with differences between sexes were detected; thus, it is unlikely that spurious genetic differentiation could have been created from sex differentiation in subsequent population genomics analysis (30). We also tested for polygenic selection using a multilocus analysis while accounting for population structure (rivers). Thus, we first quantified the admixture coefficient for each sample with ADMIXTURE v1.23 software using k = 2 and 20,000 iterations (31, 32). To test for polygenic selection, we used a random forest framework, implemented in the randomForest R package (33), that is well-suited for a large number of genetic markers and a low number of individuals (34–37). Given the small number of missing data (2.09%), we used the "*na.roughfix*" function implemented in the randomForest R package to impute the missing data (33). To account for population structure, we used a general linear model to fit the genotype against the admixture proportion. Residuals of the models were used for the random forest analysis. A series of three random forest analyses were run with 1,000,000 trees for each run. We used the "permuted importance" statistic automatically calculated by the random forest algorithm as an indicator of the significance of the marker (in interaction with others) in discriminating HOR and NOR fish. At the end of the first two runs, the markers having an importance of <0 were removed as they are not useful in discriminating hatchery from wild fish (38). We chose the total number of random forest analyses so the final out-of-bag (OOB) error would be stable and close to 5% (38). Finally, as there is no consensus method to set the important markers from the nonimportant ones, we fixed the threshold according to the upper end elbow of the importance distribution (markers with importance of >25) (35, 38).

We used permutations (n = 1,000) to assess whether a signal of polygenic selection similar to the one that was detected (*Results*) could be obtained by chance (e.g., due to genetic drift or sampling error). Within each river, the factor hatchery or wild was randomly permutated using the "schuf" function in Linux (https://github.com/jleluyer/permutation2randomforest). As conducted above, we assessed the coefficient of membership to each population using ADMIXTURE v1.23 software (k = 2) (32). The coefficients of admixture were used to remove the river effect in the random forest analysis (1,000,000 trees, three successive runs with removal of markers with importance of <0 between each run) (33). To evaluate the probability of identifying markers under selection, we compiled the final out-of-bag (OOB) error statistics for each run of simulation and compared it with the final OOB statistics in our empirical dataset using permutations tests (Fig. S3).

1. Flagg TA, Nash CE, eds (1999) *A Conceptual Framework for Conservation Hatchery Strategies for Pacific Salmonids* (US Department of Commerce, Washington, DC), NOAA Technical Memorandum NMFS-NWFSC-38.
2. Gu H, et al. (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* 6:468–481.
3. Yano A, et al. (2013) The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol Appl* 6: 486–496.
4. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.
5. Akalin A, et al. (2012) methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13:R87.
6. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
7. Zhou W (2016) biscuit-0.1.3. Zenodo. Available at doi.org/10.5281/zenodo.48262. Accessed October 3, 2016.
8. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
9. Krueger F, Andrews SR (2011) Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27:1571–1572.
10. Legendre P, Legendre L (2012) *Numerical Ecology* (Elsevier, Amsterdam), 3rd Ed, Vol 24.
11. Paradis E, Claude J, Strimmer K (2004) Ape: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
12. Oksanen J, et al. (2012) Vegan: Community Ecology Package. R package. Version 2.1-16/r2187.
13. Kim J-H, Leong JS, Koop BF, Devlin RH (2016) Multi-tissue transcriptome profiles for coho salmon (*Oncorhynchus kisutch*), a species undergoing rediploidization following whole-genome duplication. *Mar Genomics* 25:33–37.
14. Wu TD, Watanabe CK (2005) GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875.
15. Lawrence M, et al. (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol* 9:e1003118.
16. Guo JU, et al. (2011) Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci* 14:1345–1351.
17. Baerwald MR, et al. (2016) Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. *Mol Ecol* 25:1785–1800.
18. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512.
19. Bock C, Walter J, Paulsen M, Lengauer T (2007) CpG island mapping by epigenome prediction. *PLoS Comput Biol* 3:e110.
20. Morgulis A, Gertz EM, Schäffer AA, Agarwala R (2006) WindowMasker: Window-based masker for sequenced genomes. *Bioinformatics* 22:134–141.
21. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
22. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28:45–48.
23. Klopfenstein DV, et al. (2016) tanghaibao/goatools: GOATOOLS v0.6.10. Zenodo. Available at doi.org/10.5281/zenodo.159493. Accessed December 5, 2016.
24. Larson WA, et al. (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol Appl* 7:355–369.
25. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
26. Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271–280.
27. Meirmans PG, Van Tienderen PH (2004) Genotype and genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* 4:792–794.
28. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
29. Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180:977–993.
30. Benestan L, et al. (June 28, 2017) Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Mol Ecol*, 10.1111/mec.14217.
31. Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246.
32. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.
33. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2: 18–22.
34. Holliday JA, Wang T, Aitken S (2012) Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. *G3 (Bethesda)* 2: 1085–1093.
35. Laporte M, et al. (2016) RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Mol Ecol* 25:219–237.
36. Pavey SA, et al. (2015) RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Curr Biol* 25:1666–1671.
37. Brieuc MSO, Ono K, Drinan DP, Naish KA (2015) Integration of random forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Mol Ecol* 24: 2729–2746.
38. Goldstein BA, Polley EC, Briggs FBS (2011) Random forests for genetic association studies. *Stat Appl Genet Mol Biol* 10:32.
39. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.

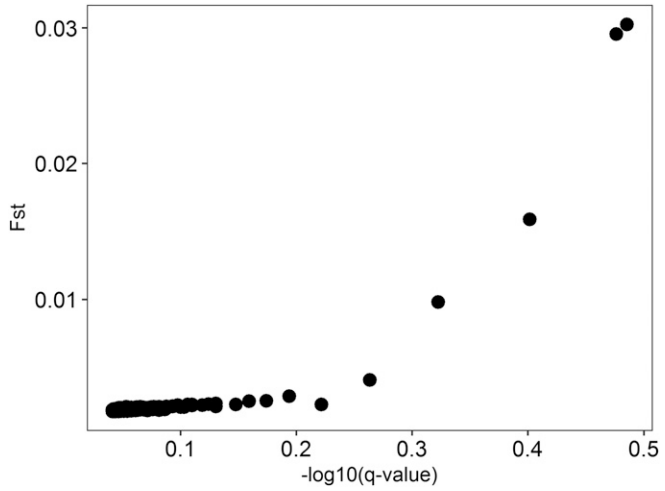

Fig. S1. BayeScan 2.0 plot of the 15,044 biallelic markers for the comparison between sexes. Fst are plotted against the −log10 q-value. The prior odd was fixed at 10. Outlier was considered significant when −log10(q-value) > 1.3 (q-value < 0.05).

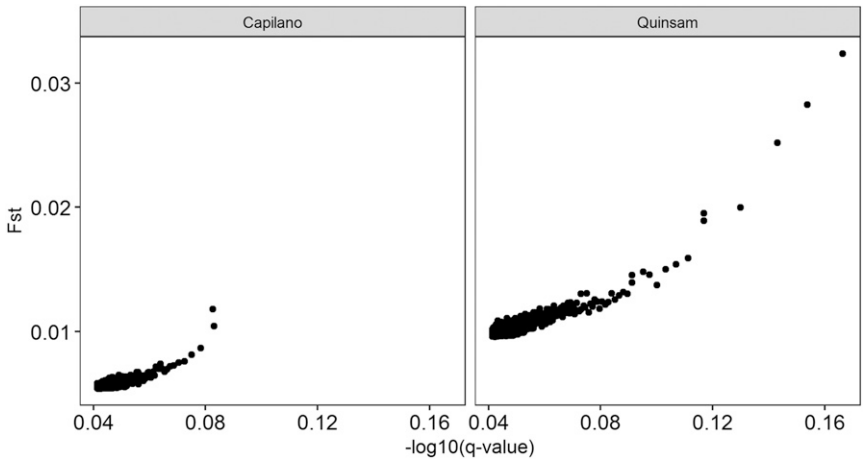

Fig. S2. BayeScan 2.0 plot of 15,044 biallelic markers for the comparison between hatchery and natural origin fish for each river. Fst are plotted against the −log10 q-value. The prior odd was fixed at 10. Outliers were considered significant when −log10 (q-value) > 1.3 (q-value < 0.05), but no marker passed this threshold.

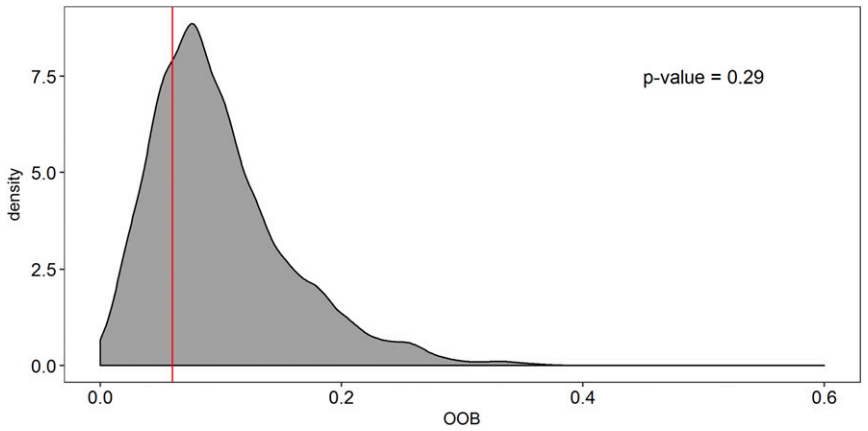

Fig. S3. Distribution of the final out-of-bag (OOB) errors from random forest analysis. The final OOB were calculated after the three runs of random forest with 1,000,000 trees for each of the 1,000 permutations. The red line indicates the final OOB value for the empirical dataset (OOB = 0.06) after three successive runs of random forest with 1,000,000 trees for each run.

Table S1. Differentially methylated regions (DMRs) and their association with Uniprot entries between hatchery and wild smolt coho salmon

Symbol	Uniprot ID	Transcript ID	Gene name	Chr./Scaff.	Met. diff.	Localization
5HT2C	Q5IS66	GDQG01000256.1	5-Hydroxytryptamine receptor 2C	Scaffold04777	23.3	Shelves
ANK1	Q02357	GDQG01029546.1	Ankyrin-1	Okis08	21.1	Gene body
AT2A2	Q03669	GDQG01041157.1	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	Okis23	15.7	Gene body
BCR	P11274	GDQG01010189.1	Breakpoint cluster region protein	Okis23	15.7	3′ UTR
BCR	Q6PAJ1	GDQG01010373.1	Breakpoint cluster region protein	Okis23	15.7	Gene body
BEGIN	Q9BUH8	GDQG01032671.1	Brain-enriched guanylate kinase-associated protein	Okis21	17.3	3′ UTR
BTAF1	O14981	GDQG01038303.1	TATA-binding protein-associated factor 172	Okis11	15.6	3′ UTR (1.5 to 5 kb)
BTAF1	O14981	GDQG01038304.1	TATA-binding protein-associated factor 172	Okis11	15.6	Gene body
CHKA	P35790	GDQG01024040.1	Choline kinase alpha	Okis04	15.6	Gene body
CSK21	Q60737	GDQG01021514.1	Casein kinase II subunit alpha	Okis17	19.6	5′ UTR
DDX53	Q86TM3	GDQG01005021.1	Probable ATP-dependent RNA helicase DDX53	Okis04	16.1	5′ UTR
DJC17	Q91WT4	GDQG01018295.1	DnaJ homolog subfamily C member 17	Okis14	17.3	Gene body
DUS12	Q9UNI6	GDQG01021214.1	Dual specificity protein phosphatase 12	Okis28	−18.9	3′ UTR (1.5 to 5 kb)
F172A	Q7T297	GDQG01019163.1	Protein FAM172A	Okis08	20.3	Gene body
HXB3A	O42368	GDQG01036924.1	Homeobox protein Hox-B3a	Okis10	15.2	Gene body
HYAL2	Q12891	GDQG01026988.1	Hyaluronidase-2	Okis05	−17.6	3′ UTR
HYAL2	Q12891	GDQG01026990.1	Hyaluronidase-2	Okis05	−17.6	Gene body
HYAL2	Q12891	GDQG01026993.1	Hyaluronidase-2	Okis05	−17.6	3′ UTR
HYAL2	Q12891	GDQG01027002.1	Hyaluronidase-2	Okis05	−17.6	Gene body
KCC1A	Q63450	GDQG01026516.1	Calcium/calmodulin-dependent protein kinase type 1	Okis17	17.9	3′ UTR (1.5 to 5 kb)
KCC2B	P28652	GDQG01028157.1	Calcium/calmodulin-dependent protein kinase type II subunit beta	Okis29	18.9	Gene body
LRC47	Q505F5	GDQG01001482.1	Leucine-rich repeat-containing protein 47	Okis17	15.6	3′ UTR (1.5 to 5 kb)
OARD1	Q9Y530	GDQG01002423.1	O-acetyl-ADP ribose deacetylase 1	Okis17	−17.1	Gene body
P73	Q9XSK8	GDQG01009063.1	Tumor protein p73	Okis17	17.5	Gene body
PCDH8	O95206	GDQG01028291.1	Protocadherin-8	Okis26	−15.9	3′ UTR (1.5 to 5 kb)
PHB2	Q5XIH7	GDQG01016094.1	Prohibitin-2	Okis30	15.5	3′ UTR (1.5 to 5 kb)
PKHA1	Q9HB21	GDQG01013238.1	Pleckstrin homology domain-containing family A member 1	Okis11	27.7	5′ UTR
SAM12	Q0VE29	GDQG01021692.1	Sterile alpha motif domain-containing protein 12	Okis17	15.1	Gene body
SGK2	Q9HBY8	GDQG01021555.1	Serine/threonine-protein kinase Sgk2	Okis17	18.8	Gene body
SRSF9	Q5PPI1	GDQG01030164.1	Serine/arginine-rich splicing factor 9	Okis23	19.2	5′ UTR
SRSF9	Q5PPI1	GDQG01030165.1	Serine/arginine-rich splicing factor 9	Okis23	19.2	5′ UTR
STX16	Q8BVI5	GDQG01021767.1	Syntaxin-16	Okis01	15.3	3′ UTR
TMC5	Q6UXY8	GDQG01032265.1	Transmembrane channel-like protein 5	Scaffold04350	27.4	Shores
TSH2	Q9NRE2	GDQG01039096.1	Teashirt homolog 2	Okis03	22.2	3′ UTR (1.5 to 5 kb)
UBE2K	P61087	GDQG01019710.1	Ubiquitin-conjugating enzyme E2 K	Okis19	15.3	5′ UTR
Unknown	Unknown	GDQG01000757.1	Unknown	Okis07	15.6	Gene body
Unknown	Unknown	GDQG01002050.1	Unknown	Scaffold07390	15.2	3′ UTR (1.5 to 5 kb)
Unknown	Unknown	GDQG01003870.1	Unknown	Okis21	17.3	Shelves
Unknown	Unknown	GDQG01005352.1	Unknown	Okis19	16.5	3′ UTR
Unknown	Unknown	GDQG01007903.1	Unknown	Okis04	16.1	3′ UTR (1.5 to 5 kb)
Unknown	Unknown	GDQG01008658.1	Unknown	Scaffold04821	18.0	3′ UTR
Unknown	Unknown	GDQG01009276.1	Unknown	Okis13	15.6	5′ UTR
Unknown	Unknown	GDQG01009277.1	Unknown	Okis13	15.6	5′ UTR
Unknown	Unknown	GDQG01020612.1	Unknown	Scaffold03114	15.0	Gene body
Unknown	Unknown	GDQG01023154.1	Unknown	Okis04	16.1	3′ UTR (1.5 to 5 kb)
Unknown	Unknown	GDQG01023155.1	Unknown	Okis04	16.1	3′ UTR (1.5 to 5 kb)
Unknown	Unknown	GDQG01023157.1	Unknown	Okis04	16.1	3′ UTR (1.5 to 5 kb)
Unknown	Unknown	GDQG01025110.1	Unknown	Okis17	19.6	5′ UTR
Unknown	Unknown	GDQG01026416.1	Unknown	Scaffold04821	18.0	3′ UTR
Unknown	Unknown	GDQG01027613.1	Unknown	Scaffold00446	17.9	3′ UTR (1.5 to 5 kb)
Unknown	Unknown	GDQG01033116.1	Unknown	Scaffold04821	18.0	5′ UTR
Unknown	Unknown	GDQG01042549.1	Unknown	Scaffold02804	19.2	Gene body

Annotation was based on a blastx approach against the Uniprot–Swissprot database (e-value < 10−6). Only significant DMRs were included [methylation difference (Met. diff.) between hatchery and wild (Meth. diff.) >15%; q-value < 0.01]. Positive values are associated with hypermethylation relative to natural origin salmon. Transcript IDs correspond to the multitissue reference transcriptome for the coho salmon (13). Each region represents a 1,000-bp portion of one of the 30 chromosomes (Chr.) (Okis) or additional scaffolds (Scaff.) from the draft coho salmon genome assembly (GenBank assembly accession no. GCA_002021735.1)

Table S2. Fst values for all pairwise comparisons

Pairwise comparisons	Fst	*P* value
Comparisons between rearing environments		
Qui-Hat/Qui-Wld	0.005	0.087
Cap-Wld/Cap-Hat	0.002	0.301
Comparisons between samples from different rivers		
Qui-Hat/Cap-Hat	0.041	0.001
Qui-Hat/Cap-Wld	0.035	0.001
Qui-Wld/Cap-Hat	0.041	0.001
Qui-Wld/Cap-Wld	0.037	0.001

Cap-Hat, Capilano-hatchery; Cap-Wld, Capilano-wild; Qui-Hat, Quinsam-hatchery; Qui-Wld, Quinsam-wild. Fst was calculated with 999 iterations using Weir and Cockerham's (39) parameter estimation (related to *No Evidence for Genome-Wide Genetic Differentiation Between HOR and NOR Salmon*).

Table S3. Measures of genetic diversity by sampling locations for 15,044 filtered SNPs

Sampling location	Ho	SD	He	SD	Gis	SD
Qui-wild	0.237	0.002	0.247	0.001	0.040	0.003
Qui-hatchery	0.241	0.002	0.249	0.001	0.031	0.003
Cap-wild	0.242	0.002	0.249	0.001	0.029	0.003
Cap-hatchery	0.327	0.002	0.248	0.001	0.045	0.003

SDs were obtained by jackknifing over loci. Gis, inbreeding coefficient; He, expected frequency of heterozygotes within samples; Ho, observed frequency of heterozygotes within samples (related to *No Evidence for Genome-Wide Genetic Differentiation Between HOR and NOR Salmon*).

Table S4. Table of individual information and bisulfite conversion results

Sample	Uniquely mapped sequence	Methylated cytosines in CpG context	Unmethylated cytosines in CpG context	Sex	River	Captivity
c1794	18,745,108	44,822,828	14,163,057	Male	Capilano	Hatchery
c1795	22,245,802	54,695,255	18,880,282	Female	Capilano	Hatchery
c1803	15,075,530	39,914,114	12,822,778	Female	Capilano	Hatchery
c1804	17,668,794	44,981,567	15,211,068	Male	Capilano	Hatchery
c1810	20,533,309	53,009,956	20,616,271	Male	Capilano	Hatchery
c1813	18,332,412	45,177,880	14,758,058	Male	Capilano	Hatchery
c1817	19,635,361	49,485,493	19,539,160	Male	Capilano	Hatchery
c1820	17,161,389	44,125,902	16,360,997	Male	Capilano	Hatchery
c1821	19,136,087	47,345,316	16,200,291	Female	Capilano	Hatchery
c1822	17,258,064	45,173,398	16,193,714	Male	Capilano	Hatchery
c4231	14,590,959	35,206,270	17,901,987	Female	Quinsam	Hatchery
c4232	16,599,541	42,783,196	15,366,059	Female	Quinsam	Hatchery
c4233	14,542,311	28,905,502	23,029,703	Female	Quinsam	Hatchery
c4234	17,493,722	46,814,905	15,350,392	Male	Quinsam	Hatchery
c4237	15,252,011	39,965,992	14,837,504	Female	Quinsam	Hatchery
c4242	20,069,147	51,358,688	18,692,043	Male	Quinsam	Hatchery
c4244	17,050,661	45,269,417	14,391,076	Male	Quinsam	Hatchery
c4245	15,763,567	41,480,785	13,747,503	Male	Quinsam	Hatchery
c4247	17,153,814	45,542,096	14,823,065	Male	Quinsam	Hatchery
c4251	15,095,394	38,971,558	15,978,062	Female	Quinsam	Wild
c4252	15,898,404	42,566,157	14,410,810	Female	Quinsam	Wild
c4254	14,509,279	39,839,585	12,777,768	Female	Quinsam	Wild
c4255	15,721,649	41,974,619	13,868,434	Male	Quinsam	Wild
c4257	20,199,795	50,191,624	18,629,320	Male	Quinsam	Wild
c4258	15,868,459	41,957,069	14,278,354	Female	Quinsam	Wild
c4260	15,750,731	41,372,870	14,961,624	Male	Quinsam	Wild
c4262	14,742,275	37,499,561	13,756,859	Male	Quinsam	Wild
c4263	16,789,054	35,586,662	25,947,096	Male	Quinsam	Wild
c4268	16,344,630	40,671,905	19,507,202	Male	Quinsam	Wild
cap06	15,702,751	40,656,552	14,070,745	Male	Capilano	Wild
cap10	16,173,511	38,457,794	20,107,091	Female	Capilano	Wild
cap11	18,215,774	46,861,209	16,394,616	Female	Capilano	Wild
cap13	15,398,650	36,558,507	12,114,735	Female	Capilano	Wild
cap17	18,409,033	45,626,631	15,919,577	Male	Capilano	Wild
cap19	18,756,014	45,577,492	21,374,067	Female	Capilano	Wild
cap21	16,506,006	40,453,837	17,214,330	Female	Capilano	Wild
cap22	16,536,356	40,163,921	17,949,988	Female	Capilano	Wild
cap23	19,567,836	50,367,430	17,878,701	Male	Capilano	Wild
cap26	18,265,660	40,606,297	26,403,499	Male	Capilano	Wild

Dataset S1. Gene ontology enrichment analysis

Dataset S1