MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1275–1293 doi: 10.1111/mec.13570

Genomewide introgressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases

M. Y. OZEROV,*† R. GROSS,‡ M. BRUNEAUX,*[1] J.-P. VÄHÄ,*†[2] O. BURIMSKI,‡ L. PUKK‡ and A. VASEMÄGI*‡

*Department of Biology, University of Turku, 20014 Turku, Finland, †Kevo Subarctic Research Institute, University of Turku, 20014 Turku, Finland, ‡Department of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, 51014 Tartu, Estonia

Abstract

Many salmonid fish populations are threatened by genetic homogenization, primarily due to introgressive hybridization with hatchery-reared conspecifics. By applying genomewide analysis using two molecular marker types (1986 SNPs and 17 microsatellites), we assessed the genetic impacts of inadvertent gene flow via straying from hatchery releases on wild populations of Atlantic salmon in the Gulf of Finland, Baltic Sea, over 16 years (1996–2012). Both microsatellites and SNPs revealed congruent population genetic structuring, indicating that introgression changed the genetic make-up of wild populations by increasing genetic diversity and reducing genetic divergence. However, the degree of genetic introgression varied among studied populations, being higher in the eastern part and lower in the western part of Estonia, which most likely reflects the history of past stocking activities. Using kernel smoothing and permutation testing, we detected considerable heterogeneity in introgression patterns across the genome, with a large number of regions exhibiting nonrandom introgression widely dispersed across the genome. We also observed substantial variation in nonrandom introgression patterns within populations, as the majority of genomic regions showing elevated or reduced introgression were not consistently detected among temporal samples. This suggests that recombination, selection and stochastic processes may contribute to complex nonrandom introgression patterns. Our results suggest that (i) some genomic regions in Atlantic salmon are more vulnerable to introgressive hybridization, while others show greater resistance to unidirectional gene flow; and (ii) the hybridization of previously separated populations leads to complex and dynamic nonrandom introgression patterns that most likely have functional consequences for indigenous populations.

Keywords: Atlantic salmon, genetic homogenization, genomewide, hatchery releases, introgressive hybridization, straying

Received 29 October 2015; revision received 13 January 2016; accepted 26 January 2016

Introduction

An increasing number of species are threatened by the reduction of the spatial component of genetic variability among populations (Olden *et al.* 2004; Vasemägi *et al.* 2005a; Laikre *et al.* 2010). This process, known as genetic homogenization, is often driven by human-mediated translocations and hybridization between

Correspondence: Anti Vasemägi, Fax: +358 2 333 6680; E-mail: antvas@utu.fi

[1]Present address: Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science, University of Jyväskylä, 40014 Jyväskylä, Finland

[2]Present address: The Association for Water and Environment of Western Uusimaa, 08100 Lohja, Finland

native and introduced conspecifics (Rhymer & Simberloff 1996; Allendorf *et al.* 2001; Vasemägi *et al.* 2005a; Edmands 2007). Hybridization between previously isolated taxa may have a negative impact on native populations through the loss of genetic variation, changes in the population's genetic structure and changes in the genetic composition within populations (Utter 2002; Laikre *et al.* 2010). Unidirectional gene flow from one taxon to another, known as introgressive hybridization, can also cause the replacement of alleles conferring local adaptation or cause the break-up of coadapted gene complexes (Edmands 2007; Laikre *et al.* 2010). In addition to direct genetic effects, translocations may cause habitat reduction and fragmentation through the naturalization of introgressed individuals, wasted reproduction through nonintrogressive hybridization and the introduction of novel pathogens and diseases (Utter 2002). However, hybridization has long been recognized as an important driver of evolution (Allendorf *et al.* 2001), as recombinant genotypes might express novel phenotypes important for adaptation to both biotic and abiotic environments (Baack & Rieseberg 2007; Fitzpatrick *et al.* 2010; Hedrick 2013).

During the last century, salmonid species have been a subject of immense anthropogenic pressure that includes pollution, habitat deterioration, commercial and recreational fisheries, aquaculture and rapid contemporary climate change (Susnik *et al.* 2004; Davidson *et al.* 2010; Lamaze *et al.* 2012; Perrier *et al.* 2013). To halt, suspend and reverse population declines, intense artificial breeding programmes have been frequently evoked to enhance the abundance and production of salmonid fish by releasing hatchery-reared conspecifics of native, non-native or mixed origins (Utter 2002; Aprahamian *et al.* 2003; Laikre *et al.* 2010). However, it has been widely recognized that both intentional and unintentional selection during fish rearing can induce fast evolutionary changes (Roberge *et al.* 2008; Tymchuk *et al.* 2009; Sauvage *et al.* 2010) that pose severe threats to indigenous gene pools, such as loss of local adaptation and alteration of fitness-related traits, when these fish are released in the wild (Einum & Fleming 1997; McGinnity *et al.* 2003; Araki *et al.* 2007, 2009; Lamaze *et al.* 2013, 2014). Moreover, hatchery-reared salmonids frequently show higher straying rates compared to their wild conspecifics (Quinn 1993; Jonsson *et al.* 2003; Vasemägi *et al.* 2005a; Jonsson & Jonsson 2006; Hansen *et al.* 2009). Finally, many of the concerns described above also extend to the escape of farmed salmonids, as it has been a growing problem in recent decades (Clifford *et al.* 1998; Utter 2002; Naylor *et al.* 2005; Skaala *et al.* 2006; Bourret *et al.* 2011; Glover *et al.* 2012, 2013; Heino *et al.* 2015). As a result, due to long-term practices of supportive stocking and the propagation of non-native populations, salmonid fishes are among the taxa most genetically affected by introgression (Utter 2002; Lamaze *et al.* 2012).

Because of the potential negative effects of introgressive hybridization, a large number of studies have evaluated both the short- and long-term genetic consequences of stocking in salmonid fishes (Vasemägi *et al.* 2005a; Hansen & Mensberg 2009; Hansen *et al.* 2009, 2010; Lamaze *et al.* 2012; Perrier *et al.* 2013). However, most of these studies relied on a relatively small number of genetic markers (tens of loci, but see Lamaze *et al.* 2012; Jasper *et al.* 2013; Hohenlohe *et al.* 2013), which makes the accurate characterization of introgressive hybridization a challenging task. Thus, the uncertainty of hybrid identification increases when the hybridizing populations are closely related and it is not possible to utilize diagnostic alleles. This is particularly relevant in many salmonid species as genetic differentiation between hatchery-reared salmon stocks and wild populations is typically low (Vähä & Primmer 2006; Hansen & Mensberg 2009). Moreover, introgression patterns may vary across the genome: at some loci, native alleles may be substituted by introgressed alleles easily, but this might be more difficult at other loci, presumably due to the variation of introduced alleles' effect on fitness in different genetic or ecological backgrounds (Baack & Rieseberg 2007). For example, rapid, positive selection-driven introgression may affect relatively few loci, which makes the detection of such genomic changes difficult using low-throughput marker systems (Gompert & Buerkle 2009; Nolte *et al.* 2009; Fitzpatrick *et al.* 2010). Finally, stocked fish often originate from multiple source populations, which complicates the accurate quantification of introgression patterns. As a result, long-term monitoring efforts combined with the use of a large number of polymorphic markers is required to reliably characterize the genome-wide effects of introgressive hybridization and to detect genomic regions that deviate from overall introgression patterns.

As one of the world's most widely recognized and prized salmonid fish species, Atlantic salmon (*Salmo salar* L.) shows a considerable degree of population structuring at varying geographical scales (King *et al.* 2001; Säisä *et al.* 2005; Primmer *et al.* 2006; Vähä *et al.* 2008; Dionne *et al.* 2009; Moore *et al.* 2014). The industrialization of the 1920–1950s led to a rapid decline of Atlantic salmon populations throughout its distribution range, particularly in the Baltic Sea basin. Historically, Atlantic salmon inhabited 84 rivers flowing into the Baltic Sea, but currently, only 10 rivers maintain self-sustaining wild populations in safe numbers (ICES 2014). Earlier studies have demonstrated that Atlantic salmon populations in the Baltic Sea are hierarchically

structured according to the geographical regions showing strong differentiation among the southeastern and northern populations (Koljonen 2001; Nilsson *et al.* 2001; Säisä *et al.* 2005). At smaller geographical scale, significant genetic divergence has been also observed among wild and hatchery populations in the Gulf of Finland (Vasemägi *et al.* 2005b). Currently, self-sustaining native salmon populations exist only in three rivers flowing to the Gulf of Finland, all located in Estonia (rivers Kunda, Keila and Vasalemma), whereas the other rivers are supported by regular releases of hatchery-reared fish (ICES 2014). Stocking programmes directed at re-establishing self-sustainable populations in Estonian rivers that historically supported wild salmon have been implemented since the mid-1990s. The first large-scale releases consisted of hatchery juveniles of captive R. Neva stock origin. Later, from 2000 to 2008, the majority of released fish originated from ascending spawners in R. Narva, and since 2008, the hatchery stock has been based on the R. Kunda population, which has become the main source of released salmon in Northern Estonia (Gross *et al.* 2014; ICES 2014). However, along with attempts to re-establish salmon populations, hatchery releases raise concerns about the status and genetic integrity of the remaining wild populations in the region, given that hatchery fishes can stray in considerable numbers and may potentially reduce the spatial component of genetic variability among populations, thereby causing a loss of intrinsic and extrinsic adaptations (Utter 2002; Vasemägi *et al.* 2005a; Laikre *et al.* 2010). However, despite potential negative impacts, the genetic effects of hatchery stocking programmes on geographically close wild salmon populations in the Baltic Sea have not been analysed at the genomewide level.

To evaluate the genetic impacts of unwanted gene flow via straying from hatchery releases on remaining wild populations of Atlantic salmon in the Gulf of Finland, we carried out a genomewide temporal analysis using two types of molecular marker (SNPs and microsatellites). We evaluated temporal changes over 16 years (*c.* four generations) and quantified the degree of introgression from two hatchery sources. Furthermore, using DNA pooling and allelotyping of a large number of SNP markers, we were able to describe the introgressive hybridization patterns at a genomewide level and identify genomic regions that are susceptible or resistant to introgression.

Materials and methods

Samples

Atlantic salmon juveniles (0+ and 1+ age) were collected using electrofishing in four rivers flowing into the Gulf of Finland, the Baltic Sea (R. Vasalemma – 2 sites; Keila – 1 site; Kunda – 1 site; and Loobu – 4 sites; Fig. 1, Table 1). The area of sampling sites ranged from 100 to 400 m^2. It should be noted that very small stretches of the studied rivers were accessible for salmon during 1996–2012 (Vasalemma – 4.5 km; Keila – 1.7 km; Kunda – 2.3 km; Loobu – 10.4 km) due to physical obstacles for upstream migration, such as waterfalls or dams. No recent hatchery releases were carried out in the first three populations, while regular releases of hatchery-reared fish to the Loobu River started in 2002. First

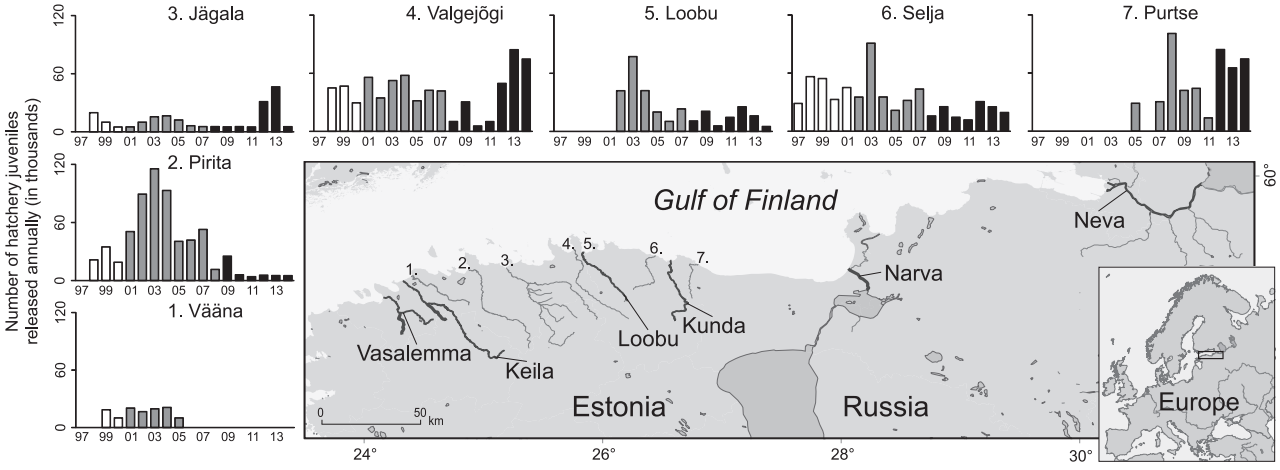

Fig. 1 Map illustrating the studied salmon rivers (bold line) and the rivers in the southern Gulf of Finland, the Baltic Sea, in which regular releases of hatchery-origin Atlantic salmon juveniles were carried out (thin line). Inserted bar plots indicates the number of juveniles released annually (in thousands, all age classes combined). White, grey and black bars indicate the origin of released fish from Neva, Narva and Kunda hatchery stocks, respectively. The years (1997–2014) are indicated below the bars.

Table 1 Sample information and mean genetic diversity indices over 17 microsatellite and 1986 SNP loci

				Microsatellites						SNPs	
Population	Sampling years	Abbr.	Status	*n*	*A*	A_R	H_E	H_O	N_e (95% CI)	*n*	H_E
Vasalemma	1996–99	Vas96-99	Wild	45	7.53	6.98	0.71	0.75	18 (10–35)	45	0.34
	2007–08	Vas07-08	Wild	30	6.88	6.78	0.67	0.69	13 (7–30)	26	0.31***
	2009–10	Vas09-10	Wild	97	9.35	7.78	0.73	0.75	25 (15–43)	48	0.34
	2011–12	Vas11-12	Wild	67	8.59	7.53	0.71	0.71	35 (22–59)	21	0.31***
Keila	1996–97	Kei96-97	Wild	54	7.65	6.95	0.70	0.73	40 (26–64)	39	0.34
	2007–08	Kei07-08	Wild	67	9.12	7.87	0.72	0.75	36 (23–58)	35	0.33*
	2009–10	Kei09-10	Wild	98	10.18	8.30	0.74	0.74	98 (71–133)	48	0.34
	2011–12	Kei11-12	Wild	95	9.82	8.15	0.73	0.73	93 (67–128)	45	0.34
Loobu	1996–99	Loo96-99	Wild	81	4.29	3.92	0.55	0.61	6 (3–21)	45	0.24
	2007–08	Loo07-08	Enhanced	77	9.12	7.65***	0.72**	0.77	32 (20–52)	46	0.32***
	2009–10	Loo09-10	Enhanced	102	9.29	7.83***	0.74**	0.78	25 (15–43)	48	0.34***
	2011–12	Loo11-12	Enhanced	104	9.71	7.83***	0.74***	0.74	59 (41–88)	48	0.35***
Kunda	1996–97	Kun96-97	Wild	71	5.47	4.75	0.59	0.63	22 (13–42)	48	0.28
	2007–08	Kun07-08	Wild	57	8.94	7.69*	0.71*	0.73	56 (38–86)	35	0.34***
	2009–10	Kun09-10	Wild	70	8.71	7.46*	0.71*	0.70	52 (35–81)	48	0.33***
	2011–12	Kun11-12	Wild	120	10.47	8.10**	0.73*	0.73	52 (36–77)	48	0.35***
Narva	1998	Nar98	Hatchery	45	9.00	8.27	0.73	0.76	42 (27–66)	—	—
	2001	Nar01	Hatchery	73	10.24	8.80	0.75	0.75	82 (58–115)	—	—
	2004	Nar04	Hatchery	129	10.24	8.53	0.74	0.72	108 (82–143)	—	—
	2005	Nar05	Hatchery	77	10.12	8.59	0.75	0.75	91 (66—127)	—	—
	2006	Nar06	Hatchery	112	9.88	8.30	0.74	0.75	98 (72–132)	42	0.35
	2007	Nar07	Hatchery	80	9.65	8.33	0.75	0.77	76 (56—109)	—	—
	2008	Nar08	Hatchery	95	9.71	8.22	0.75	0.77	69 (48—98)	—	—
	2009	Nar09	Hatchery	109	10.06	8.33	0.74	0.74	70 (50—99)	—	—
Neva	1997–98	Nev97-98	Hatchery	97	8.76	7.78	0.73	0.73	56 (39–84)	48	0.34

Abbr. – abbreviation; *n* – sample size/pool size; *A* – mean number of alleles; A_R – allelic richness; H_E – expected heterozygosity; H_O – observed heterozygosity; N_e (95% CI) – effective population size and its 95% confidence interval.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ – significance of nonparametric Mann–Whitney *U*-test indicating the change of genetic diversity level in relation to the earliest sample.

samples from each river were collected in 1996–1999 before the start of large-scale releases, while subsequent samples were collected in 2007–2012 (Table 1). Fin clips from sampled individuals were stored in 96% ethanol for further genetic analysis and fish caught from different sites within a river were combined for subsequent analyses. In addition, the samples of adult individuals (adipose fin clip) of two hatchery populations (Neva and Narva) were collected in 1996–2009. We did not include the third hatchery stock of the R. Kunda origin in the analysis, as the large-scale releases of juvenile fish from this broodstock started more recently (Fig. 1). The permits for sample collection were issued by the Ministry of Environment (Estonia).

DNA extraction and microsatellite analysis

Total genomic DNA was extracted according to Laird *et al.* (1991). Altogether 17 microsatellite loci were analysed in two multiplex (9-plex and 8-plex) PCRs (Table S1, Supporting information). Each 10 μL PCR consisted of *c.* 10–20 ng template DNA, 1× Type-it Multiplex PCR Master Mix (Qiagen, Germany) and 50–200 nM of each primer. Forward primers were labelled fluorescently by 6-FAM, VIC, PET or NED. Amplifications were carried out in a Biometra TProfessional Thermal cycler with an initial heat activation at 95 °C for 5 min followed by 28 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C (9-plex) or at 55 °C (8-plex) for 90 s, extension at 72 °C for 30 s and the final extension for 30 min at 60 °C. Multiplex PCR products were electrophoresed on an Applied Biosystems 3500 Genetic Analyzer (Life Technologies, USA) and microsatellite genotypes were scored using GeneMapper v.5 software (Life Technologies, USA). To directly compare the microsatellite and SNP data sets temporal samples were grouped as follows: earlier samples from wild (years 1996–99) before the start of large-scale hatchery releases and more recent samples (years 2007–08, 2009–10 and 2011–12; Table 1).

DNA pool construction

The same individual DNA samples that were used for microsatellite analysis were used to construct DNA pools, grouped as described above, except Narva hatchery stock, from which only Nar06 sample was included for DNA pooling. Given that genetic divergence among Narva hatchery samples of different years based on microsatellite analyses was very low (F_{ST} = 0.000–0.007; Table S2, Supporting information), we used a single temporal sample (2006) from Narva hatchery stock for subsequent DNA pooling and SNP analysis. Equimolar (10 ng/μL) DNA extracts from 21 to 48 individuals (mean = 42) were pooled as described in Ozerov *et al.* (2013a,b) to provide two technical replicates for each year-class sample (Table 1). The pooled DNA samples were analysed in the Center for Integrative Genetics (CIGENE, Norway) using an Illumina infinium assay (Illumina, San Diego, CA, USA) and version 2 of the Atlantic salmon SNP-chip (Lien *et al.* 2011; Bourret *et al.* 2013) carrying probes for 5568 SNP markers. Among our study populations, 3928 of those SNP loci were biallelic and were further analysed in this study. The raw SNP data were analysed using GENOTYPING MODULE v. 1.9.4 (Genome Studio software v. 2011.1; Illumina Inc.).

Estimation of SNP allele frequencies from DNA pools

Allele frequencies for each sample were estimated from DNA pools by comparing pool-specific value of theta with the reference values of theta from individual genotyping of 300 Atlantic salmon specimens, as described in Ozerov *et al.* (2013a,b). Briefly, during Illumina genotyping the raw signal data from two alternate alleles are converted into a theta value which ranges from 0 to 1. In theory, a heterozygous individual would have a theta value of 0.5, whereas theta of 0 or 1 would indicate a homozygote at allele A or B, respectively. However, in reality a SNP's theta for genotype clusters (AA, AB and BB) varies from expected values. Thus, allele frequency in a pooled sample is estimated by comparing the theta value for each SNP to the mean theta values for AA, AB and BB genotypes calculated by genotyping of individual samples using the correction algorithm (method 2 in Janicki & Liu 2009). Stringent quality control filters resulted in selection of 2880 biallelic SNPs showing low error rates (variation of theta among technical replicates ≤0.02) compared to the information content (Ozerov *et al.* 2013b). Population-specific allele frequencies for each SNP were estimated as a mean over two technical replicates. SNPs showing MAF < 0.05 were excluded from the further analysis, thus retaining 1986 SNP loci in the final data set.

Microsatellite data analysis

The basic descriptive statistics for each microsatellite locus and sample (allelic richness, observed and expected heterozygosity) were obtained using FSTAT 2.93 (Goudet 1995). The same software was used to test for deviation from the Hardy–Weinberg equilibrium and to calculate Weir & Cockerham (1984) within-population inbreeding coefficients (f) and between-population fixation indices (F_{ST}). The Brookfield 1 estimator (Brookfield 1996) implemented in MICRO-CHECKER 2.2.3 (Van Oosterhout *et al.* 2004) was used to assess the potential presence of genotyping errors due to scoring of stutter peaks, large allele dropouts and null alleles. Effective population size (N_e) for each sample was estimated using the sibship assignment approach implemented in COLONY 2.0.5.0 (Jones & Wang 2010). The method estimates demographic parameters from the multilocus genotypes of a sample allowing to calculate the probabilities that a pair of offspring taken at random from the population are half- and full-sibs. These sibship probabilities were subsequently used to estimate N_e (equation 10 in Wang 2009). The COLONY runs were performed with the following options: female and male polygamy, random mating, full-likelihood method and medium-length run.

Population relationships and within-population diversity

To infer the genetic relationships among populations, pairwise D_A (Nei *et al.* 1983) distances and mean pairwise F_{ST} (Reynolds *et al.* 1983) were calculated from allele frequency estimates derived from allelotyping of 1986 SNPs with the POWERMARKER v3.25 software package (Liu & Muse 2005). The same approach was applied for the 17 microsatellite markers. The D_A genetic distances (Nei *et al.* 1983) were used to construct neighbour-joining trees based on 1986 SNPs and 17 microsatellites with 1000 bootstrap replicates. The consensus dendrogram was constructed using the program SPLITSTREE4 v4.11.3 (Huson & Bryant 2006).

Admixture analysis and detection of hybrid categories using microsatellite data

Individual admixture proportions (Q) were estimated using STRUCTURE 2.3.4 (Pritchard *et al.* 2000). Each pairwise analysis included fish sampled in a given river and one of the hatchery samples (Nev97-98 or Nar98-09). An admixture model with correlated allele frequencies was applied for each STRUCTURE run with 250 000 MCMC repeats preceded by a burn-in of 100 000 steps. The number of clusters was assumed to range from

$K = 1$ to $K = 6$. The ad hoc statistics ΔK (Evanno *et al.* 2005) calculated in STRUCTURE HARVESTER 0.6.94 (Earl & vonHoldt 2012) was used to estimate the most probable K for each pairwise comparison. The most probable number of clusters was $K = 2$ for all pairwise comparisons (Table S3, Supporting information). The membership of each individual in each of the two clusters (wild or hatchery) was averaged across five replicate runs and the average admixture proportions for each wild sample were estimated.

To estimate the probability of each individual belonging to a certain hybrid group, the method of Anderson & Thompson (2002) implemented in NEWHYBRIDS 1.1 was applied. In a Bayesian framework, the model infers each individual's genotype frequency, and a Markov chain Monte Carlo (MCMC) simulation approach is then used to obtain estimates from the posterior distribution reflecting the level of confidence of an individual belonging to one of six categories: pure (wild or hatchery), F1, F2 and the backcrosses (F1 × wild or F1 × hatchery). The samples of hatchery stocks (Nev97-98 or Nar98-09) and earlier samples from wild populations before the start of large-scale releases were used as reference samples to assign more recent samples (2007–2012) to hybrid categories. The highest posterior probability value was used as a threshold for assigning an individual to that specific category. Jeffreys-like priors were applied for both allele frequency and mixture proportions. All results are based on 100 000 MCMC repeats following a burn-in period of 50 000.

To further evaluate the power of 17 microsatellite loci to classify individuals to different hybrid categories, we carried out simulations as in Kovach *et al.* (2015). In brief, one hundred multilocus genotypes of each pure (wild × wild; hatchery × hatchery), F1 (wild × hatchery), F2 (F1 × F1) and backcross (F1 × wild; F1 × hatchery) category were generated using the software HYBRIDLAB 1.0 (Nielsen *et al.* 2006). Simulated genotypes were analysed in NEWHYBRIDS as described above using hatchery and the earlier wild samples as references. Individuals were considered accurately classified when the highest posterior probability was associated with the correct hybrid category.

Quantification of genetic introgression

Genetic introgression was quantified using two approaches. First, microsatellite-based introgression was estimated using STRUCTURE 2.3.4 as the proportion of the hatchery genome (I) in the recent samples by comparing the admixture proportions to the earliest wild reference sample. Given that wild reference and hatchery samples may have an average membership in the wild cluster (Q) less than one and above zero, respectively, the following expression was applied [modified from Karlsson *et al.* (2014): $I = 1 - ((Q_w - Q_h) \times (Q_n - Q_h)^{-1})$], where Q_w, Q_h and Q_n are the average membership in the wild cluster of the wild sample, the hatchery stock and the native population sample collected before large-scale releases, respectively. I estimates less than zero (three of 24 cases) were subsequently adjusted to zero. Second, introgression was quantified from allele frequency shifts for each SNP marker based on the expression: $\hat{S} = (p_w - p_n) \times (p_h - p_n)^{-1}$, where p_w, p_h and p_n are the allele frequencies of the wild population, the hatchery stock and the native population sample collected before large-scale releases, respectively (Taggart & Ferguson 1986; Almodovar *et al.* 2001). To ensure that the locus-specific and genomewide introgression estimates were as accurate as possible, for subsequent analysis, we used only a subset of SNPs (upper 30th percentile) that showed the highest genetic differentiation (F_{ST}) between hatchery stock and the earliest wild samples. $\hat{S}$ estimates >1 or < 0 were adjusted to one and zero, respectively, before calculating median $\hat{S}$ across all studied loci and kernel smoothing.

Detection of genomic regions showing elevated or reduced introgression

To identify genomic regions, rather than single markers, with elevated and reduced levels of introgression compared to the random distribution in the genome, we applied a kernel smoothing and permutation test approach similar to Hohenlohe *et al.* (2010) and Bruneaux *et al.* (2013). Kernel smoothing allows to reduce the noise of individual introgression estimates ($\hat{S}$) and permutation testing enables identification of regions where smoothed introgression estimates differ significantly from expectation under a random distribution. The applied approach takes the differences in marker density into account and increases the statistical power for detecting regions when multiple adjacent markers show consistently high or low introgression values. Smoothed introgression profiles were calculated within each chromosome using a nonparametric regression with an adaptive local plug-in bandwidth (LOKERN R package). P-values indicating the departure from the random distribution of $\hat{S}$ estimates along the chromosome were calculated by permuting single-locus $\hat{S}$ estimates across all chromosomes (10 000 permuted replicates) and comparing the original kernel regression profile to the distribution of profiles obtained from the permuted data sets (Flori *et al.* 2009; Hohenlohe *et al.* 2010). Genomic regions with P-values ≤0.025 were considered to significantly deviate from the random introgression pattern. The false discovery rate was estimated as the corresponding Q-value using the QVALUE package

ver. 1.0 in R (R Core Team 2015) with a conservative bootstrap method (Storey *et al.* 2004). Kernel smoothing was performed using R ver. 3.1.2 and the LOKERN package ver. 1.1-5 (Herrmann 2013) and permutation tests were implemented with custom-made scripts. To ensure a reliable run of the kernel smoothing algorithm, data were filtered prior to the analysis to exclude chromosomes with a low number of SNPs (≤5).

Results

Genetic diversity at microsatellite loci

In total, 234 alleles were observed across 17 microsatellite loci, ranging from two (Ssa14) to 27 (SSsp2201) alleles per locus (mean = 14; median = 15). The mean level of genetic diversity over 17 microsatellite loci varied considerably between populations and temporal replicates; it was the lowest in Loo96-97 (H_O = 0.61, A_R = 3.92) and the highest in Kun09-10 (H_O = 0.78, A_R = 7.83) and Nar01 (H_O = 0.75, A_R = 8.88; Table 1).

Deviations from the Hardy–Weinberg equilibrium ($P < 0.05$) were detected in 51 of 423 locus–sample combinations. After correcting for multiple significance tests, 19 of the combinations remained significant at the 5% level (Benjamini & Yekutieli (2001), $\alpha = 0.00755$; Table S1, Supporting information). All of them involved negative f values, indicating heterozygote excess. The potential presence of low-frequency null alleles (0.04–0.10) was suggested at nine locus–sample combinations out of 425 tests (Table S4, Supporting information). However, as none of the loci showed consistent deviations across multiple populations, we retained all 17 microsatellite loci for further analysis.

The earliest wild samples contained small number (13 of 168) of unique microsatellite alleles at relatively low frequencies (average frequency: 1.95%) that were not observed in hatchery stocks. On the other hand, much larger number of hatchery-specific alleles (60 of 215) were observed in two reared stocks that were missing in earliest wild samples (average frequency: 1.67%) reflecting the higher variability of hatchery stocks (Table S5, Supporting information).

Comparison of genetic variation revealed by microsatellite and SNP loci

As expected, genetic diversity estimates within populations, measured as the mean expected heterozygosity (H_E), were significantly correlated between the two marker types (Spearman's $r = 0.81$, $P < 0.0001$); they were higher for microsatellite loci (H_E = 0.55–0.74) and lower for SNPs (H_E = 0.24–0.35; Mann–Whitney U-test, $P < 0.001$; Table 1). Pairwise genetic differentiation estimates (F_{ST}) over 17 microsatellites varied from zero (Nar01 vs. Nar05) to 0.243 (Loo96-99 vs. Nev97-98), whereas mean pairwise F_{ST} values over 1986 SNPs ranged from 0.021 (Kun07-08 vs. Kun11-12) to 0.267 (Loo96-99 vs. Nev97-98; Tables S2 and S6, Supporting information). Similar to genetic diversity, genetic differentiation estimates among populations were highly correlated between two marker types ($r_{xy} = 0.95$, $P < 0.01$). Thus, both individual genotyping of microsatellite markers and allelotyping of SNPs from DNA pools revealed similar population genetic structuring, as illustrated by neighbour-joining trees (Fig. 2).

Temporal changes of genetic variation and divergence

Both marker types showed a significant temporal increase in H_E in the Kunda and Loobu populations (Mann–Whitney U-test $P < 0.001$–0.045; Table 1). By contrast, only minor temporal changes of genetic diversity were observed in Keila or Vasalemma (Table 1). Both marker types indicated a sharp decrease in genetic divergence (F_{ST}) between hatchery stocks and the Loobu and Kunda populations, while for two other populations (Vasalemma and Keila), the genetic divergence did not drastically change in time (Fig. 3). Overall genetic divergence among four wild populations declined nearly twofold in recent years compared to the earliest samples before the start of large-scale releases (Fig. 4). Evaluation of the temporal changes within wild populations showed a tendency of increased effective population size estimates (N_e) in the recent samples based on microsatellite data (Table 1). N_e estimates of hatchery samples (median = 76.0) were higher than those of wild populations (median = 35.5; Mann–Whitney U-test, $P < 0.01$).

Admixture analyses and classification of hybrids

Both STRUCTURE and NEWHYBRIDS produced comparable results, indicating the temporal increase in unidirectional gene flow from hatcheries to wild populations (Fig. 5, Fig. S1, Supporting information). As expected, R. Loobu was the most affected by introgression, evidenced by the high frequency of admixed individuals observed in the recent samples. The second most affected population was Kunda (Fig. S1, Supporting information). While STRUCTURE indicated the occurrence of introgression, NEWHYBRIDS provided further insight into the hybridization process by allowing further separation of admixed individuals to specific hybrid categories. Whereas pure hatchery individuals of Neva origin were not observed in the recent wild juvenile samples, the genetic impact of Narva hatchery stock to studied populations was much more visible (Fig. 5). As

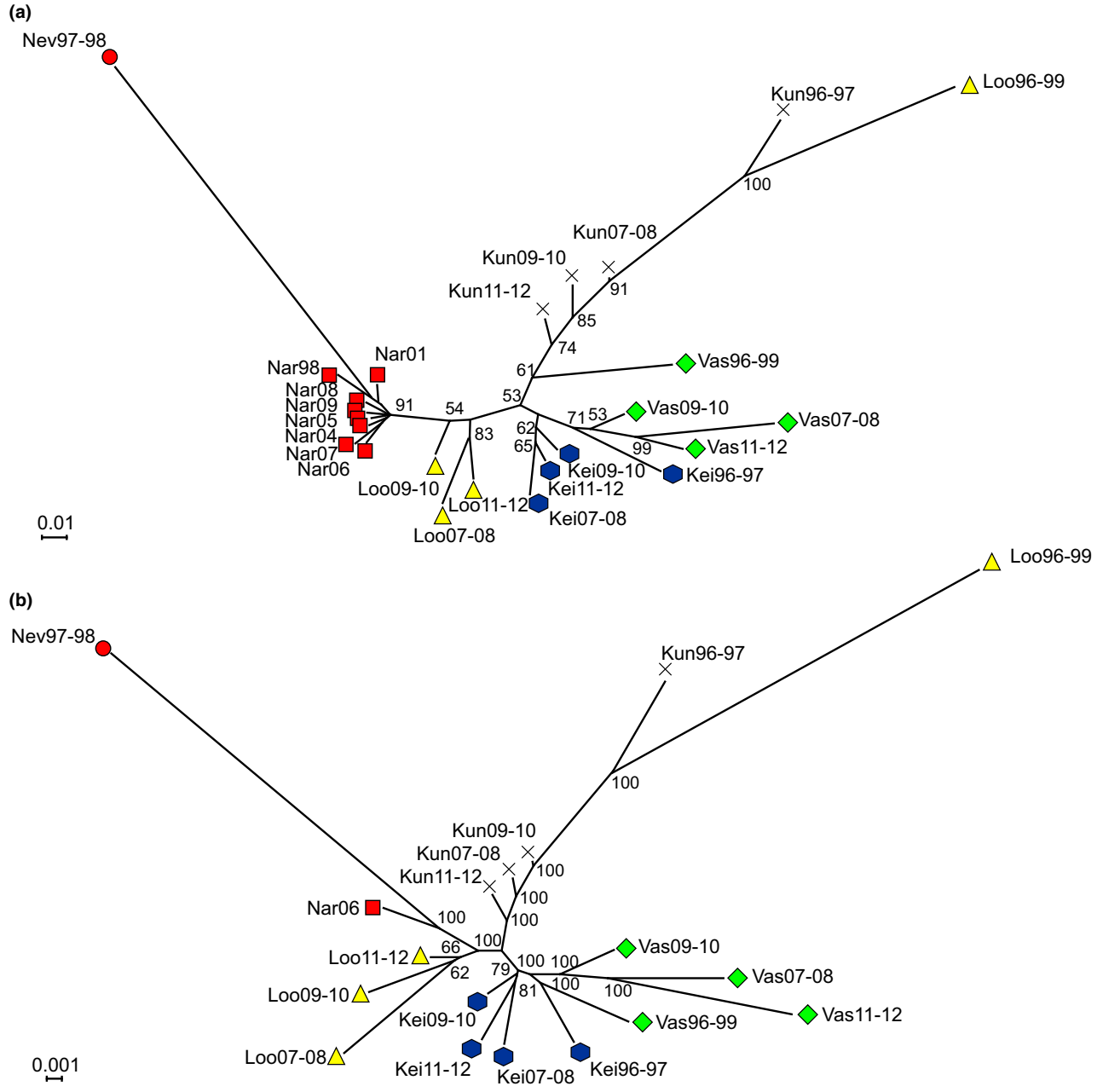

Fig. 2 Neighbour-joining consensus dendrograms representing relationships among Estonian Atlantic salmon populations based on Nei's D_A genetic distances over (a) 17 microsatellites or (b) 1986 SNPs. The number on the nodes indicates the bootstrap values (percentage) obtained after 1000 replicates.

expected, the most drastic changes were observed in the R. Loobu, where the native population has been largely replaced with hatchery fish of Narva origin. Interestingly, a large proportion of juveniles collected during 2007–2008 in the R. Loobu represented F1 hybrids, while the increase in backcrosses (F1 × Narva) was evident in later samples (2009–2012; Fig. 5b). In Kunda, a large proportion of juveniles in recent samples corresponded to putatively pure Narva fish and backcrosses (F1 × wild). Similarly, putatively pure hatchery juveniles were also present in Keila and Vasalemma. In contrast to Vasalemma and Kunda, the recent samples of Keila also included F1 hybrids (Fig. 5b).

Simulations indicated that the accuracy of hybrid classification depended on genetic divergence (F_{ST}) between reference samples (Fig. S2; Table S7,

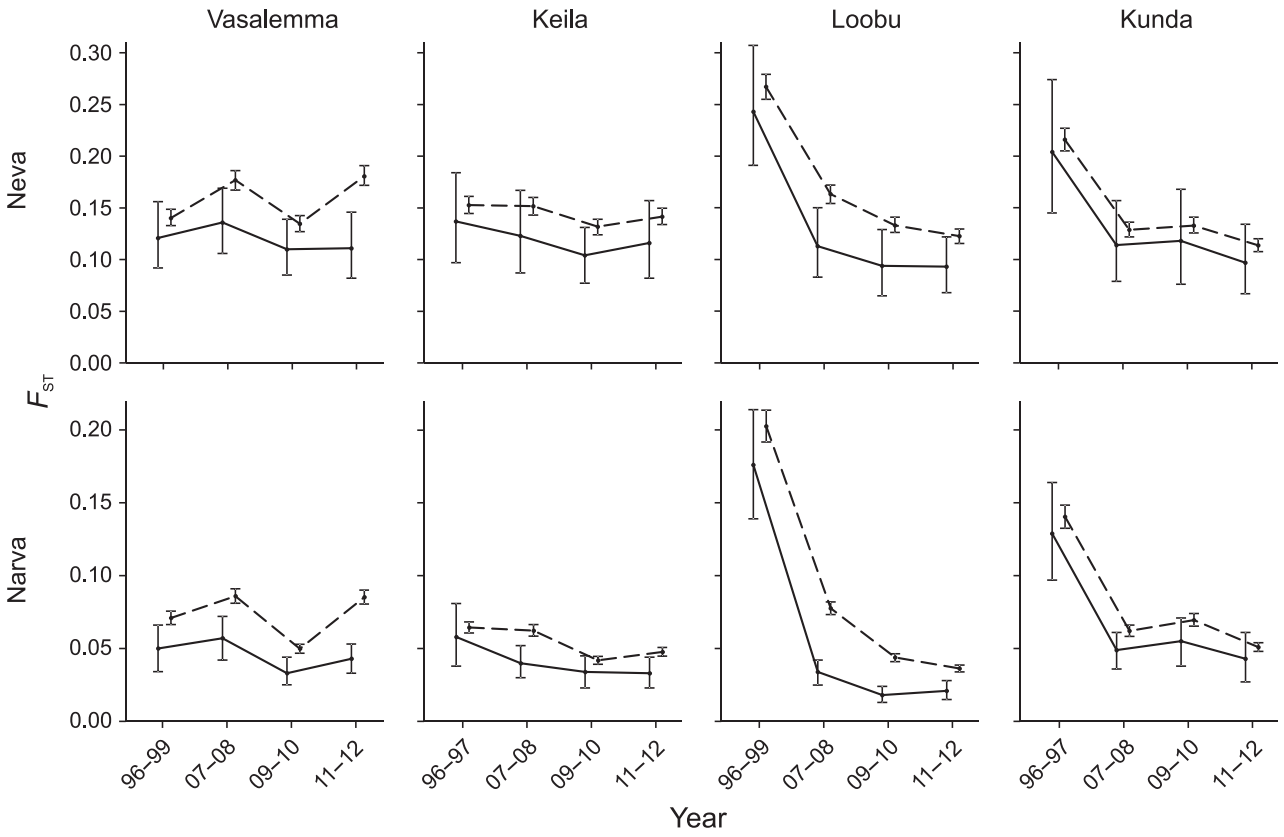

Fig. 3 Temporal changes of genetic divergence (estimated as F_{ST}) between hatchery and wild populations. Solid and dashed lines correspond to F_{ST} estimates based on 17 microsatellites and 1986 SNPs, respectively.

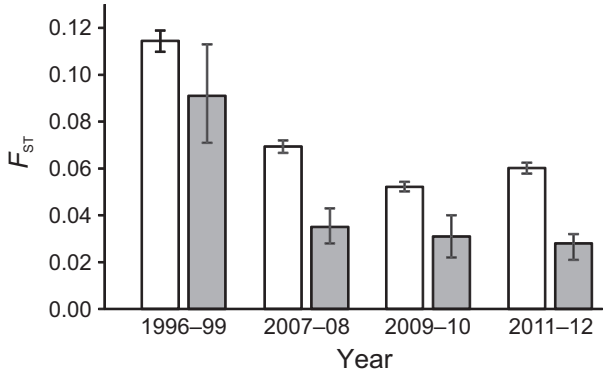

Fig. 4 Temporal change in overall genetic divergence (estimated as F_{ST}) among four wild populations. F_{ST} estimates based on SNP and microsatellite loci are presented as white and grey rectangles, respectively. Error bars indicate 95% confidence interval estimates.

Supporting information). While the majority of pure wild and hatchery individuals (97–100%) were correctly classified irrespective of genetic divergence between two reference populations, the classification accuracy of other hybrid categories was dependent on genetic divergence. For example, a large proportion of F1 hybrids was correctly assigned when $F_{ST} > 0.05$ (78–97%), whereas the assignment accuracy decreased to 52% at $F_{ST} < 0.05$. Similarly, the proportion of correctly classified F2 hybrids decreased when the genetic divergence between two reference populations decreased. The majority of misclassified backcross individuals were assigned to pure reference populations.

Quantification of genetic introgression

Two different genetic introgression estimates showed similar temporal patterns within populations, although SNPs generally resulted in higher introgression estimates than microsatellites (Table 2). However, our SNP-based introgression estimates ($\hat{S}$) were highly correlated with microsatellite-based estimates, I (introgression from Neva: Pearson's $r = 0.93$, $P < 0.0001$; introgression from Narva: Pearson's $r = 0.90$, $P < 0.0001$). Consistent with the results of STRUCTURE and NEWHYBRIDS, introgression estimates from Neva were lower compared to Narva hatchery stock. As expected, the highest introgression levels were detected in the R. Loobu, where regular releases of hatchery-reared fish began in 2002.

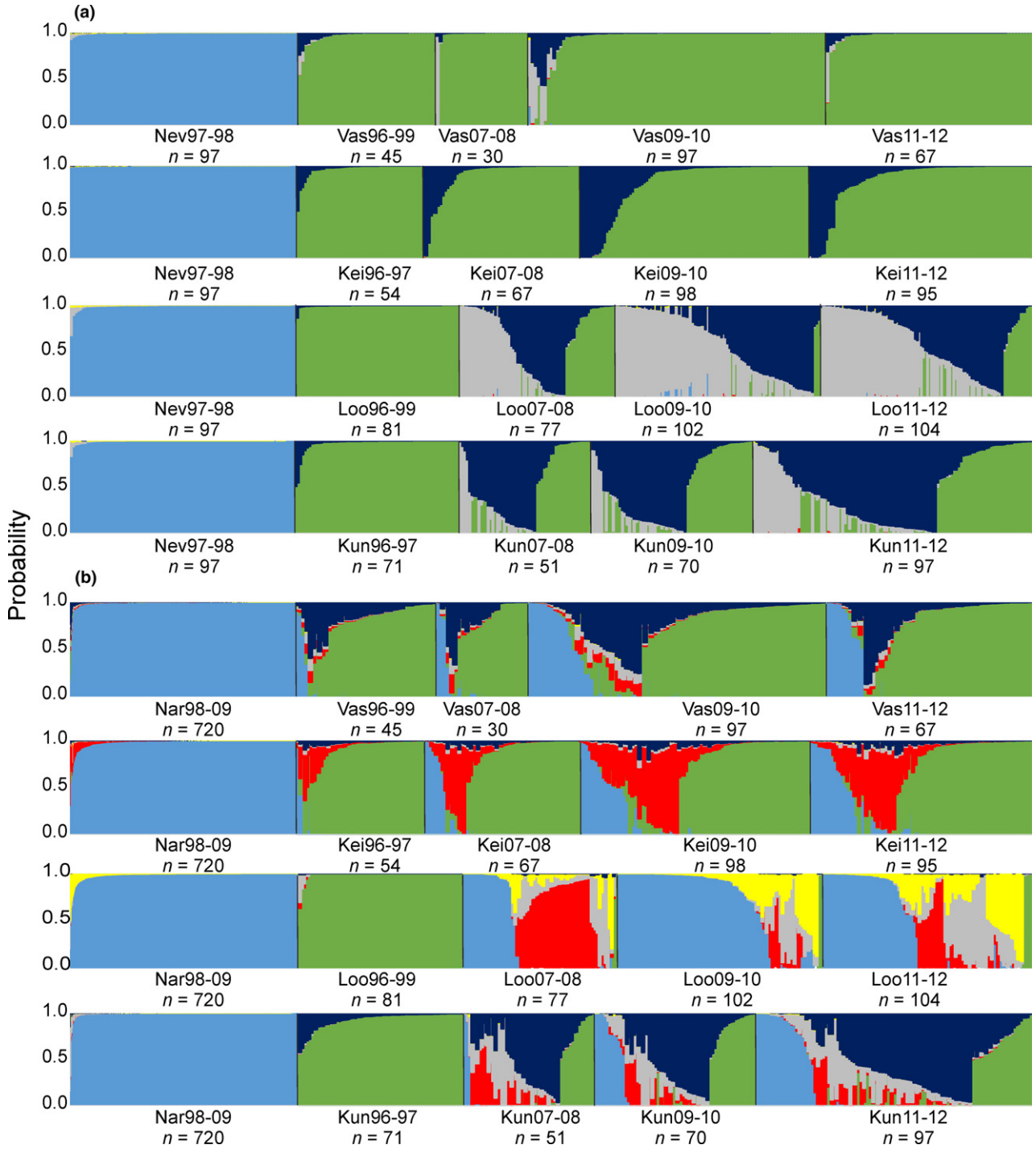

Fig. 5 Probability of an individual to be a member of six categories for each of the four wild populations affected by introgression from (a) Neva and (b) Narva hatchery stocks based on 17 microsatellite loci. Each individual is presented as a vertical bar. The probability of each individual belonging to pure hatchery (blue), pure wild (green), F1 (red), F2 (grey), and backcross (F1 × hatchery, yellow; F1 × wild, dark blue) hybrid categories. Populations and temporal replicates are separated by vertical solid lines.

Similarly, relatively high levels of introgression from Narva hatchery to Keila and Kunda populations were detected (Kunda: I = 0.218–0.266, $\hat{S}$ = 0.433–0.504; Keila: I = 0.097–0.146, $\hat{S}$ = 0.309–0.458), indicating that hatchery releases to geographically adjacent rivers resulted in inadvertent gene flow to native salmon populations in the Gulf of Finland. In contrast, estimated introgression levels to the R. Vasalemma were much lower, and 95%

confidence intervals for microsatellite-derived introgression estimates (I) always included zero (Table 2).

Heterogeneous distribution of introgression along the chromosomes

Permutation analysis indicated that single-locus introgression estimates ($\hat{S}$) were not randomly distributed along the chromosomes (Fig. 6). The genomic regions showing elevated or reduced introgression (kernel permutation test, $P < 0.025$) represented only a small proportion (0.04–4.50%; Table S8, Supporting information) of the whole genome (in relation to the Atlantic salmon female map length), and the size of most of the regions deviating from random introgression pattern was rather small (< 5–10 cM; Fig. S3, Supporting information). Across all populations, the size of the regions showing elevated introgression was significantly higher for Narva than for Neva hatchery stock (Wilcoxon test $P < 0.01$; Table S8, Supporting information). Similarly, across all populations, the proportion of the genome that showed elevated introgression was larger compared to reduced introgression from Narva hatchery (Wilcoxon test $P < 0.05$).

The comparison of temporal introgression dynamics within populations indicated that only a small number of regions that deviated from random introgression patterns remained stable over time (e.g. elevated introgression: Nar-Kun on Chr. 25; reduced introgression: Nar-Kun on Chr. 6), whereas in several cases, nonrandom introgression patterns were observed only for two temporal samples of three (Fig. 6; Fig. S4; Table S9, Supporting information). Similarly, the majority of genomic regions deviating from random introgression showed little overlap among populations (Fig. 6; Fig. S4; Table S9, Supporting information). However,

Table 2 Mean individual admixture (Q; Pritchard *et al*. 2000) and introgression estimates based on microsatellites (I; Karlsson *et al*. 2014) and SNPs ($\hat{S}$; Taggart & Ferguson 1986; Almodovar *et al*. 2001)

From:	To:	Year	Microsatellites Q (95% CI)	I (95% CI)	SNPs $\hat{S}$ (95% CI)
Neva	Vasalemma	1996–99	0.031 (0.015–0.053)		
		2007–08	0.017 (0.004–0.040)	0.000 (0.000–0.019)	0.054 (0.024–0.091)
		2009–10	0.040 (0.021–0.066)	0.009 (0.000–0.042)	0.156 (0.133–0.181)
		2011–12	0.020 (0.009–0.036)	0.000 (0.000–0.015)	0.096 (0.048–0.140)
	Keila	1996–97	0.009 (0.005–0.013)		
		2007–08	0.021 (0.011–0.034)	0.012 (0.001–0.027)	0.106 (0.082–0.133)
		2009–10	0.032 (0.021–0.045)	0.023 (0.012–0.038)	0.171 (0.142–0.185)
		2011–12	0.022 (0.013–0.034)	0.001 (0.004–0.027)	0.165 (0.141–0.190)
	Loobu	1996–99	0.022 (0.019–0.026)		
		2007–08	0.442 (0.370–0.513)	0.444 (0.373–0.511)	0.336 (0.312–0.375)
		2009–10	0.686 (0.645–0.724)	0.703 (0.661–0.746)	0.498 (0.458–0.536)
		2011–12	0.585 (0.540–0.628)	0.596 (0.550–0.641)	0.475 (0.442–0.505)
	Kunda	1996–97	0.017 (0.014–0.021)		
		2007–08	0.158 (0.119–0.200)	0.148 (0.107–0.191)	0.290 (0.277–0.315)
		2009–10	0.154 (0.119–0.191)	0.144 (0.106–0.183)	0.273 (0.253–0.290)
		2011–12	0.200 (0.168–0.234)	0.192 (0.158–0.230)	0.348 (0.328–0.367)
Narva	Vasalemma	1996–99	0.135 (0.066–0.210)		
		2007–08	0.120 (0.046–0.204)	0.000 (0.000–0.113)	0.342 (0.289–0.411)
		2009–10	0.176 (0.122–0.234)	0.050 (0.000–0.149)	0.481 (0.438–0.522)
		2011–12	0.147 (0.084–0.216)	0.015 (0.000–0.118)	0.449 (0.390–0.547)
	Keila	1996–97	0.081 (0.042–0.126)		
		2007–08	0.165 (0.107–0.235)	0.097 (0.011–0.180)	0.309 (0.263–0.354)
		2009–10	0.199 (0.148–0.251)	0.136 (0.066–0.206)	0.411 (0.378–0.446)
		2011–12	0.207 (0.155–0.255)	0.146 (0.070–0.221)	0.458 (0.415–0.509)
	Loobu	1996–99	0.015 (0.008–0.025)		
		2007–08	0.606 (0.542–0.671)	0.616 (0.552–0.680)	0.567 (0.533–0.601)
		2009–10	0.838 (0.797–0.873)	0.858 (0.817–0.896)	0.781 (0.753–0.810)
		2011–12	0.718 (0.675–0.766)	0.733 (0.686–0.787)	0.751 (0.735–0.773)
	Kunda	1996–97	0.008 (0.007–0.008)		
		2007–08	0.219 (0.155–0.287)	0.218 (0.153–0.294)	0.433 (0.410–0.455)
		2009–10	0.254 (0.179–0.333)	0.254 (0.180–0.332)	0.406 (0.387–0.431)
		2011–12	0.265 (0.208–0.321)	0.266 (0.204–0.329)	0.504 (0.477–0.526)

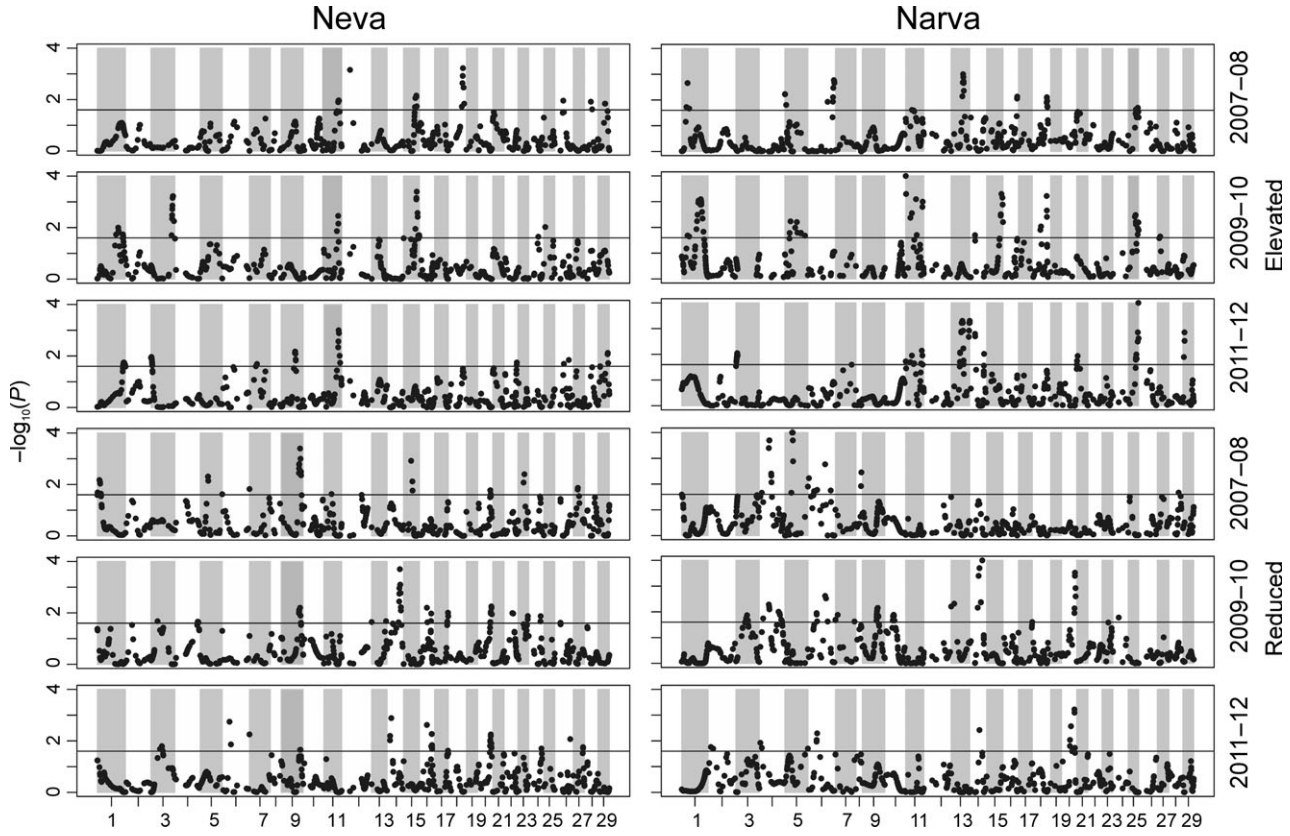

Fig. 6 Genomic regions showing elevated and reduced introgression from two hatchery stocks (Neva and Narva) to the wild Kunda population. *P*-values for each SNP ($-\log_{10}$-transformed) were plotted against their respective positions on the autosomal chromosomes.

there were few observations of elevated or reduced introgression within the same genomic region among three or four populations. For example, a large genomic region showing elevated introgression from Neva was observed in at least one temporal sample of each studied population on Chr. 15 (52.1–65.8 cM). Similarly, a region of elevated introgression from the Narva hatchery was observed on Chr. 4 in three populations (51.4–58.3 cM: Vas, Kei and Loo), whereas regions of reduced introgression were consistently observed on Chr. 8 (18.8–20.6 cM: Vas, Kei and Kun) and Chr. 20 (54.7–55.1 cM: Kei, Loo and Kun; Fig. S4; Table S9, Supporting information).

Discussion

Since the initiation of the Atlantic salmon restoration programme in Estonia nearly two decades ago, regular spawning of salmon currently occurs in a number of rivers (HELCOM 2011). This, together with increasing densities of wild-born salmon parr (ICES 2014), indicates that hatchery releases can be an effective way to re-establish salmon populations when the environmental conditions have become suitable for Atlantic salmon. However, increased straying from hatchery releases may also threaten the genetic integrity of the remaining indigenous salmon populations in the region. Our study provides genetic evidence of significant introgression from hatchery stocks into wild populations using both microsatellite and SNP markers. In addition, we detected nonrandom patterns of introgression at the genomewide level, indicating that some genomic regions may be more vulnerable to introgressive hybridization, while others show more resistance to unidirectional gene flow.

Temporal changes of genetic variation and structure

Both microsatellite and SNP markers revealed congruent temporal patterns, suggesting that introgression has changed the genetic make-up of wild populations by reducing genetic divergence and increasing genetic diversity. The increase in genetic diversity as a result of unidirectional gene flow from hatchery releases contradicts the typical perception that farmed strains often possess reduced genetic diversity relative to wild popu-

lations (Ryman & Laikre 1991; Blanchet *et al.* 2008; Hutchings & Fraser 2008; Araki & Schmid 2010); thus, the introgression of reared fish is expected to reduce the genetic variability of wild populations (Eldridge *et al.* 2009; Bourret *et al.* 2011; Jasper *et al.* 2013). However, it is not always the case, as hatchery fish may carry unique alleles not observed in wild populations (Verspoor 1998); thus, hatchery releases may also increase the genetic variability of wild populations, particularly if the source populations for hatchery stocks are larger than the wild populations. All wild salmon populations in the Gulf of Finland are characterized by small population sizes (ICES 2014) because of the very limited area available for reproduction (0.017–0.104 km^2) as a result of man-made (Vasalemma, Loobu and Kunda) and natural (Keila) obstacles for upstream salmon migration (HELCOM 2011). In contrast, reproduction of both the Neva and Narva hatchery stocks is based on hundreds of spawners. Nevertheless, the earliest samples from the rivers Keila and Vasalemma showed relatively high levels of genetic variation comparable to that of hatchery samples, whereas genetic diversity of the earlier samples from the rivers Loobu and Kunda was nearly two times lower than that of hatchery samples. Thus, despite the increase in allelic richness in all four wild populations, significant temporal increases in genetic diversity were detected only in Loobu and Kunda. It is therefore likely that sharp increases in genetic diversity in R. Loobu and Kunda can be explained by relatively low variability of the earliest samples combined with the strong impact of hatchery releases. In contrast, the more subtle impact of hatchery releases via straying to wild Vasalemma and Keila populations is most likely related to the smaller number of fish released to adjacent rivers. For example, R. Vääna, which is closest to the rivers Vasalemma and Keila, was stocked over a relatively short period of time (1999–2005), and the number of annually released juveniles was relatively low (10 000–20 000 individuals). However, direct stocking to R. Loobu and to neighbouring rivers of R. Kunda began earlier (1996–1997), and the number of annually released juveniles has been much higher (Fig. 1). Similar patterns were observed by Marie *et al.* (2010) and Lamaze *et al.* (2012), showing that the increase in within-population genetic diversity in wild brook charr (*Salvelinus fontinalis*) was related to stocking efforts in the lakes. Additionally, Perrier *et al.* (2013) found an increase in within-population genetic diversity in wild Atlantic salmon in France due to long-term stocking activities in the region.

In addition to temporal shifts in diversity, the impact of hatchery releases on wild populations was also evident from the decrease in genetic divergence between wild populations and hatchery stocks. As expected, the sharpest temporal decline in genetic divergence was observed for the Loobu population, where the native stock has been practically replaced with hatchery fish of Narva origin. However, the decrease in genetic divergence was also observed between wild Kunda, Keila and hatchery samples, while the genetic divergence between hatchery and wild Vasalemma samples did not decline over time. The convergence of allele frequencies of wild populations towards hatchery samples also resulted in the temporal reduction of overall genetic divergence among four wild populations. Although it was high among the earlier samples (F_{ST} = 0.091 and F_{ST} = 0.114 for microsatellites and SNPs, respectively), genetic divergence decreased considerably among recent samples (F_{ST} = 0.028 for microsatellites and F_{ST} = 0.060 for SNPs). Our results therefore demonstrate that straying and subsequent gene flow from hatchery releases have reduced the genetic divergence among remaining wild Atlantic salmon populations in the Gulf of Finland. Similar alterations of genetic structure due to unidirectional gene flow from reared conspecifics as a result of straying (Vasemägi *et al.* 2005a; Jasper *et al.* 2013), direct stocking (Hansen 2002; Eldridge *et al.* 2009; Lamaze *et al.* 2012; Marie *et al.* 2010; Perrier *et al.* 2013) or farm escapes (Skaala *et al.* 2006; Bourret *et al.* 2011; Glover *et al.* 2012) have also been observed in earlier studies on Atlantic salmon and other salmonids.

Bayesian analyses further demonstrated that a substantial degree of hybridization between hatchery stocks and wild populations has occurred over time. As expected, the highest levels of admixture were observed in the recent Loobu samples, where the majority of individuals currently comprise different types of hybrids as well as pure hatchery fish. Similarly, recent samples from the R. Kunda consisted of a significant proportion of different hybrids and even juveniles of putatively pure hatchery origin. The populations of the R. Vasalemma and Keila were affected by introgression to a lesser degree, as the majority of recent samples were classified as pure wild individuals. However, it should be noted that the power to classify individuals into correct hybrid categories varies depending on the F_{ST} between populations. For example, Vähä & Primmer (2006) showed that F1 hybrids can be correctly detected when the level of divergence between samples is moderate to high (F_{ST} = 0.12–0.21) and at least 12–24 microsatellite loci are used (~76 and 153 independent alleles, respectively). However, the separation of backcrosses from pure individuals generally requires even larger numbers of loci or, alternatively, the presence of high genetic divergence between samples (Vähä & Primmer 2006). Our data set consisted of 17 microsatellite loci (217 independent alleles) and the accuracy of the assignment of pure hatchery or wild individuals to

the correct hybrid category was high in all hatchery–wild combinations (97–100%). The identification of F1 hybrids was also relatively accurate (>78%) in most of the hatchery–wild comparisons, except for Nar-Vas, due to the low genetic divergence between samples (F_{ST} = 0.047). As expected, the highest level of misclassifications was observed for backcrosses, particularly for F1 × Narva hybrids. Thus, the proportion of pure hatchery fish in the recent wild samples may be overestimated, and the occurrence of putatively pure hatchery juveniles in wild rivers should be interpreted with caution.

Two different estimates of genetic introgression provided additional insights into introgressive hybridization patterns. Except for five comparisons involving Loobu samples, introgression estimate $\hat{S}$ (Taggart & Ferguson 1986; Almodovar *et al.* 2001) based on SNPs resulted in higher values compared to the alternative estimate I (Karlsson *et al.* 2014) based on admixture modelling and multilocus microsatellite genotype data. There are several potentially nonexclusive explanations for this. First, it is possible that temporal allele frequency shifts measured at SNP loci are influenced by genetic drift and the small sizes of wild populations and may contribute to the introgression estimate ($\hat{S}$), as shifts in allele frequency towards hatchery samples increase $\hat{S}$. Similarly, both technical and sampling variation may influence allele frequency estimates from DNA pools and therefore inflate the introgression estimate ($\hat{S}$). However, we believe that the stringent quality control filters applied to the SNP data minimize these technical effects. Moreover, using a much larger number of SNPs (~700 most informative SNPs for each pairwise comparison) in contrast to microsatellites (17 loci) is expected to reflect the average effect of introgression over a much larger genomic area and may therefore be closer to the true genomewide estimate.

Genomewide patterns of genetic introgression

We found considerable heterogeneity in introgression signatures across the genome both among and within wild populations, including tens of loci showing elevated or reduced introgression. Regions exhibiting nonrandom introgression were widely dispersed across the genome, rather than colocalized in a few discrete genomic regions. A similar heterogeneity of genomewide introgression patterns has been recently observed in other species, for example between manakin birds (*Manacus candei* and *M. vitellinus*; Parchman *et al.* 2013) or different mice subspecies (*Mus musculus musculus* and *M. m. domesticus*; Staubach *et al.* 2012; *M. spretus* and *M. m. domesticus*; Liu *et al.* 2015). Such a genomewide distribution of introgression patterns is expected if the fitness is shaped by many widely dispersed alleles across the genome rather than a few loci with major effects (Parchman *et al.* 2013). In addition, we observed considerable variation in nonrandom introgression patterns within populations, as genomic regions showing elevated or reduced introgression were not consistently detected among all temporal samples. This suggests that recombination, selection and stochastic processes (e.g. a small number of breeders) may contribute to complex nonrandom introgression patterns (Fitzpatrick 2013). Therefore, for a deeper understanding of the genomic consequences of introgression, it is important to evaluate the temporal dynamics of the hybridization process of two previously separated gene pools. Alternatively, the temporal aspect of the introgression process may be captured along the geographical cline (Barton & Hewitt 1985), while further insights into hybrid fitness and gene exchange may be gained by genomic cline analysis that enables one to disentangle the effects of selection and drift (Gompert & Buerkle 2009; Fitzpatrick 2013). However, the analysis of genomic clines was not possible in this study, as our SNP data set consisted of allele frequency information obtained from pooled DNA rather than individual genotype data. Nevertheless, the combination of the kernel smoothing approach with permutation testing allowed us to shift the focus of the analysis from single markers to a more genomewide context.

Despite the fact that most genomic regions showing nonrandom introgression patterns were scattered throughout the genome, few regions showing elevated or reduced introgression were more regularly detected across several populations and temporal samples. For example, the genomic region on Chr. 16 showed elevated introgression from the Neva hatchery stock in all wild populations. Interestingly, this genomic region harbours a QTL that controls the resistance to infectious salmon anaemia (ISA) in Atlantic salmon (Moen *et al.* 2007). This region also contains several genes that are associated with immune response to viral infections, including the HIV-EP2/MBP-2 gene, which was recently identified as a strong candidate gene involved in the resistance to the ISA virus in Atlantic salmon (Li *et al.* 2011). We also identified one particular SNP on Chr. 25 (GCR_cBin47052_Ctg1_234) that showed signs of elevated introgression from the Narva hatchery in two wild populations (Keila and Kunda). Intriguingly, this marker located in proximity (1.62 Mb) to the vestigial-like family member three gene that was recently shown to control a large proportion of the variation (~39%) at the age of maturity of Atlantic salmon in a sex-specific manner (Barson

et al. 2015). Therefore, it is possible that the elevated introgression of the hatchery genome into wild populations on Chr. 25 is related to fitness effects associated with the timing of maturation. Finally, regions showing elevated or reduced introgression from the Narva hatchery that were consistently observed on chromosomes 4, 8 and 20 have been previously shown to harbour the QTL related to growth and flesh colour in Atlantic salmon (Baranski *et al.* 2010; Gutierrez *et al.* 2012). At the same time, growth-related QTL are known to be widely distributed across the genome, and given the large genomic regions associated with most of the identified QTL, it is not clear whether the regions exhibiting nonrandom introgression patterns are in fact functionally related to differences in growth.

Taken together, this study described the dynamics and patterns of introgressive hybridization of indigenous Atlantic salmon populations in the Gulf of Finland as a result of inadvertent gene flow from hatchery releases over 16 years. We showed that the impact of hatchery releases varied among studied populations, most likely because of the differences in stocking effort and the geographical location of releases. We detected nonrandom patterns of introgression at the genome-wide level, indicating that some genomic regions are more vulnerable to introgressive hybridization while others show more resistance to unidirectional gene flow. Our results suggest that hybridization of previously separated populations leads to complex and dynamic nonrandom introgression patterns across the genome that most likely have functional consequences for indigenous populations.

Acknowledgements

We thank Ene Saadre, Kunnar Klaas and others from the Põlula fish rearing department (State Forest Management Centre), and Martin Kesler, late Mart Kangur, Mari Liis Viilmann, Marje Aid, Tiit Paaver and many volunteers for assistance with field and hatchery sampling; and Matthew P. Kent for running Illumina infinium assays. The research was funded by the Estonian Ministry of Education and Research (institutional research funding project IUT8-2 led by RG), the Academy of Finland (Grant No. 266321 to AV), Estonian Science Foundation (Grant No. 8215 to AV) and the Estonian Environmental Investment Centre.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Almodovar A, Suarez J, Nicola GG, Nuevo M (2001) Genetic introgression between wild and stocked brown trout in the Douro River basin, Spain. *Journal of Fish Biology*, **59**, 68–74.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Aprahamian MW, Smith KM, McGinnity P, McKelvey S, Taylor J (2003) Restocking of salmonids – opportunities and limitations. *Fisheries Research*, **62**, 211–227.

Araki H, Schmid C (2010) Is hatchery stocking a help or harm? Evidence, limitations and future directions in ecological and genetic surveys. *Aquaculture*, **308**, S2–S11.

Araki H, Cooper B, Blouin MS (2007) Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. *Science*, **318**, 100–103.

Araki H, Cooper B, Blouin MS (2009) Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. *Biology Letters*, **5**, 621–624.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, **17**, 513–518.

Baranski M, Moen T, Vage DI (2010) Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, **42**, 17.

Barson N, Aykanat T, Hindar K *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in Atlantic salmon. *Nature*, **528**, 405–408.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

Blanchet S, Paez DJ, Bernatchez L, Dodson JJ (2008) An integrated comparison of captive-bred and wild Atlantic salmon (*Salmo salar*): implications for supportive breeding programs. *Biological Conservation*, **141**, 1989–1999.

Bourret V, O'Reilly PT, Carr JW, Berg PR, Bernatchez L (2011) Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity*, **106**, 500–510.

Bourret V, Kent MP, Primmer CR *et al.* (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**, 532–551.

Brookfield JF (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Bruneaux M, Johnston SE, Herczeg G *et al.* (2013) Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Molecular Ecology*, **22**, 565–582.

Clifford SL, McGinnity P, Ferguson A (1998) Genetic changes in an Atlantic salmon population resulting from escaped juvenile farm salmon. *Journal of Fish Biology*, **52**, 118–127.

Davidson WS, Koop BF, Jones SJ *et al.* (2010) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology*, **11**, 403.

Dionne M, Caron F, Dodson JJ, Bernatchez L (2009) Comparative survey of within-river genetic structure in Atlantic salmon; relevance for management and conservation. *Conservation Genetics*, **10**, 869–879.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and imple-

menting the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Edmands S (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, **16**, 463–475.

Einum S, Fleming IA (1997) Genetic divergence and interactions in the wild among native, farmed and hybrid Atlantic salmon. *Journal of Fish Biology*, **50**, 634–651.

Eldridge WH, Myers JM, Naish KA (2009) Long-term changes in the fine-scale population structure of coho salmon populations (*Oncorhynchus kisutch*) subject to extensive supportive breeding. *Heredity*, **103**, 299–309.

Evanno G, Regnaut S, Goudet G (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Fitzpatrick BM, Johnson JR, Kump DK *et al.* (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences United States of America*, **107**, 3606–3610.

Flori L, Fritz S, Jaffrezic F *et al.* (2009) The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE*, **4**, e6595.

Glover KA, Quintela M, Wennevik V *et al.* (2012) Three decades of farmed escapees in the wild: a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway. *PLoS ONE*, **7**, e43129.

Glover KA, Pertoldi C, Besnier F *et al.* (2013) Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC Genetics*, **14**, 74.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Gross R, Paaver T, Aid M, Burimski O, Pukk L (2014) Kalade taastootmise alased uuringud. Keskkonnaministeeriumi töövõtulepingu 4-1.1/13/238, 62 p. EMÜ Veterinaarmeditsiini ja Loomakasvatuse Instituut, Tartu, Estonia (in Estonian).

Gutierrez AP, Lubieniecki KP, Davidson EA *et al.* (2012) Genetic mapping of quantitative trait loci (QTL) for bodyweight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array. *Aquaculture*, **358**, 61–70.

Hansen MM (2002) Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology*, **11**, 1003–1015.

Hansen MM, Mensberg KL (2009) Admixture analysis of stocked brown trout populations using mapped microsatellite DNA markers: indigenous trout persist in introgressed populations. *Biology Letters*, **5**, 656–659.

Hansen MM, Fraser DJ, Meier K, Mensberg KL (2009) Sixty years of anthropogenic pressure: a spatio-temporal genetic analysis of brown trout populations subject to stocking and population declines. *Molecular Ecology*, **18**, 2549–2562.

Hansen MM, Meier K, Mensberg KL (2010) Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Molecular Ecology*, **19**, 1787–1800.

Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.

Heino M, Svasand T, Wennevik V, Glover KA (2015) Genetic introgression of farmed salmon in native populations: quantifying the relative influence of population size and frequency of escapees. *Aquaculture Environment Interactions*, **6**, 185–190.

HELCOM (2011) Salmon and Sea Trout Populations and Rivers in the Baltic Sea – HELCOM assessment of salmon (*Salmo salar*) and sea trout (*Salmo trutta*) populations and habitats in rivers flowing to the Baltic Sea. 24 pp.

Herrmann E (2013) lokern: Kernel Regression Smoothing with Local or Global Plug-in Bandwidth.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hohenlohe PA, Day MD, Amish SJ *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002–3013.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Hutchings JA, Fraser DJ (2008) The nature of fisheries- and farming-induced evolution. *Molecular Ecology*, **17**, 294–313.

ICES (2014) Report of the Baltic Salmon and Trout Assessment Working Group (WGBAST), 26 March–2 April 2014, Aarhus, Denmark. ICES CM 2014/ACOM:08, 347 pp.

Janicki P, Liu J (2009) Accuracy of allele frequency estimates in pool DNA analyzed by high-density Illumina Human 610-Quad microarray. *The Internet Journal of Genomics and Proteomics*, **5**, 6857.

Jasper JR, Habicht C, Moffitt S *et al.* (2013) Source-sink estimates of genetic introgression show influence of hatchery strays on wild chum salmon populations in Prince William Sound, Alaska. *PLoS ONE*, **8**, e81916.

Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551–555.

Jonsson B, Jonsson N (2006) Cultured Atlantic salmon in nature: a review of their ecology and interaction with wild fish. *ICES Journal of Marine Science*, **63**, 1162–1181.

Jonsson B, Jonsson N, Hansen LP (2003) Atlantic salmon straying from the River Imsa. *Journal of Fish Biology*, **62**, 641–657.

Karlsson S, Diserud OH, Moen T, Hindar K (2014) A standardized method for quantifying unidirectional genetic introgression. *Ecology and Evolution*, **4**, 3256–3263.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, **10**, 807–821.

Koljonen M-L (2001) Conservation goals and fisheries management units for Atlantic salmon in the Baltic Sea area. *Journal of Fish Biology*, **59**(Suppl. A), 269–288.

Kovach AI, Walsh J, Ramsdell J, Kelley Thomas W (2015) Development of diagnostic microsatellite markers from whole-genome sequences of Ammodramus sparrows for

assessing admixture in a hybrid zone. *Ecology and Evolution*, **5**, 2267–2283.

Laikre L, Schwartz MK, Waples RS, Ryman N, Ge MWG (2010) Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. *Trends in Ecology and Evolution*, **25**, 520–529.

Laird PW, Zijderveld A, Linders K *et al.* (1991) Simplified mammalian DNA isolation procedure. *Nucleic Acids Research*, **19**, 4293.

Lamaze FC, Sauvage C, Marie A, Garant D, Bernatchez L (2012) Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, **21**, 2877–2895.

Lamaze FC, Garant D, Bernatchez L (2013) Stocking impacts the expression of candidate genes and physiological condition in introgressed brook charr (*Salvelinus fontinalis*) populations. *Evolutionary applications*, **6**, 393–407.

Lamaze FC, Pavey SA, Normandeau E *et al.* (2014) Neutral and selective processes shape MHC gene diversity and expression in stocked brook charr populations (*Salvelinus fontinalis*). *Molecular ecology*, **23**, 1730–1748.

Li J, Boroevich KA, Koop BF, Davidson WS (2011) Comparative genomics identifies candidate genes for infectious salmon anemia (ISA) resistance in Atlantic salmon (*Salmo salar*). *Marine Biotechnology (NY)*, **13**, 232–241.

Lien S, Gidskehaug L, Moen T *et al.* (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, **12**, 615.

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**, 2128–2129.

Liu KJ, Steinberg E, Yozzo A *et al.* (2015) Interspecific introgressive origin of genomic diversity in the house mouse. *Proceedings of the National Academy of Sciences USA*, **112**, 196–201.

Marie AD, Bernatchez L, Garant D (2010) Loss of genetic integrity correlates with stocking intensity in brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, **19**, 2025–2037.

McGinnity P, Prodohl P, Ferguson A *et al.* (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 2443–2450.

Moen T, Sonesson AK, Hayes B *et al.* (2007) Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics*, **8**, 53.

Moore JS, Bourret V, Dionne M *et al.* (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**, 5680–5697.

Naylor R, Hindar K, Fleming IA *et al.* (2005) Fugitive salmon: assessing the risks of escaped fish from net-pen aquaculture. *BioScience*, **55**, 427–437.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, **6**, 971–973.

Nilsson J, Gross R, Dove O *et al.* (2001) Matrilinear phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area. *Molecular Ecology*, **10**, 89–102.

Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.

Olden JD, Leroy Poff N, Douglas MR, Douglas ME, Fausch KD (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology and Evolution*, **19**, 18–24.

Ozerov M, Vasemagi A, Wennevik V *et al.* (2013a) Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS ONE*, **8**, e82434.

Ozerov M, Vasemagi A, Wennevik V *et al.* (2013b) Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon (*Salmo salar* L.). *BMC Genomics*, **14**, 12.

Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Perrier C, Guyomard R, Bagliniere JL, Nikolic N, Evanno G (2013) Changes in the genetic structure of Atlantic salmon populations over four decades reveal substantial impacts of stocking and potential resiliency. *Ecology and Evolution*, **3**, 2334–2349.

Primmer CR, Veselov AJ, Zubchenko A *et al.* (2006) Isolation by distance within a river system: genetic population structuring of Atlantic salmon, Salmo salar, in tributaries of the Varzuga River in northwest Russia. *Molecular Ecology*, **15**, 653–666.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quinn TP (1993) A review of homing and straying of wild and hatchery-produced salmon. *Fisheries Research*, **18**, 29–44.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the Coancestry coefficient: basic for a short-term genetic distance. *Genetics*, **105**, 767–779.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Roberge C, Normandeau E, Einum S, Guderley H, Bernatchez L (2008) Genetic consequences of interbreeding between farmed and wild Atlantic salmon: insights from the transcriptome. *Molecular Ecology*, **17**, 314–324.

Ryman N, Laikre L (1991) Effects of supportive breeding on the genetically effective population size. *Conservation Biology*, **5**, 325–329.

Säisä M, Koljonen M-L, Gross R *et al.* (2005) Population genetic structure and postglacial colonization of Atlantic salmon (*Salmo salar*) in the Baltic Sea area based on microsatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 1887–1904.

Sauvage C, Derome N, Normandeau E *et al.* (2010) Fast transcriptional responses to domestication in the brook charr *Salvelinus fontinalis*. *Genetics*, **185**, 105–112.

Skaala O, Wennevik V, Glover K (2006) Evidence of temporal genetic change in wild Atlantic salmon, *Salmo salar* L., populations affected by farm escapees. *ICES Journal of Marine Science*, **63**, 1224–1233.

Staubach F, Lorenc A, Messer PW *et al.* (2012) Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genetics*, **8**, e1002891.

Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **66**, 187–205.

Susnik S, Berrebi P, Dovc P, Hansen MM, Snoj A (2004) Genetic introgression between wild and stocked salmonids and the prospects for using molecular markers in population rehabilitation: the case of the Adriatic grayling (*Thymallus thymallus* L. 1785). *Heredity (Edinb)*, **93**, 273–282.

Taggart JB, Ferguson A (1986) Electrophoretic evaluation of a supplemental stocking programmes for brown trout (*Salmo trutta* L.). *Aquaculture and Fisheries Management*, **17**, 155–162.

Tymchuk W, Sakhrani D, Devlin R (2009) Domestication causes large-scale effects on gene expression in rainbow trout: analysis of muscle, liver and brain transcriptomes. *General and Comparative Endocrinology*, **164**, 175–183.

Utter FM (2002) Kissing cousins: Genetic interactions between wild and cultured salmon. In: *Ghost Runs: The Future of Wild Salmon on the North and Central Coasts of British Columbia*(eds Harvey B, MacDuffee M), pp. 119–135. Rainforest Conservation Society, Victoria, British Columbia.

Vähä J-P, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Vähä J-P, Erkinaro J, Niemelä E, Primmer CR (2008) Temporally stable genetic structure and low migration in an Atlantic salmon population complex: implications for conservation and management. *Evolutionary Applications*, **1**, 137–154.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vasemägi A, Gross R, Paaver T, Koljonen ML, Nilsson J (2005a) Extensive immigration from compensatory hatchery releases into wild Atlantic salmon population in the Baltic sea: spatio-temporal analysis over 18 years. *Heredity*, **95**, 76–83.

Vasemägi A, Gross R, Paaver T *et al.* (2005b) Analysis of gene associated tandem repeat markers in Atlantic salmon (*Salmo salar* L.) populations: implications for restoration and conservation in the Baltic Sea. *Conservation Genetics*, **6**, 385–397.

Verspoor E (1998) Genetic impacts on wild Atlantic salmon (*Salmo salar* L.) stocks from escaped farm conspecifics: An assessment of risk. In: *Canadian Stock Assessment Secretariat Research Document 98/156* 20 p., Canadian Stock Assessment Secretariat, Halifax, NS.

Wang J (2009) A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular Ecology*, **18**, 2148–2164.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution*, **38**, 1358–1370.

M.Y.O., A.V. and R.G. conceived and designed the study. M.B. wrote the R script for the detection of nonrandom introgression patterns. M.Y.O. and A.V. analysed the data. L.P., O.B. and J.-P.V. prepared DNA pools. O.B., R.G. and L.P. were involved in sample collection and microsatellite analysis. M.Y.O. and A.V. wrote the manuscript with contributions from all co-authors.

Data accessibility

Individual genotypes (17 microsatellite loci), SNP allele frequencies of pooled DNA samples and custom-made script to perform kernel smoothing and permutation tests are available under Dryad accession: doi:10.5061/dryad.p00gd.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Individual admixture proportions and 90% confidence intervals based on 17 microsatellite loci for each of the four wild populations introgressed by (a) Neva and (b) Narva hatchery stocks. The *y*-axis depicts the proportion of the genome from hatchery stock.

Fig. S2 The effect of the level of genetic divergence on the accuracy of identification of hybrid categories: pure (green squares), F1 (red triangles), F2 (grey diamonds), F1 × hatchery (orange circles) and F1 × wild (blue crosses) backcross. The trends for each hybrid category are presented as logarithmic trend lines of the same colours.

Fig. S3 Estimated size of genomic regions (in cM) showing (a) elevated and (b) reduced introgression in four studied salmon populations.

Fig. S4 Genomic regions showing elevated and reduced introgression from two hatchery stocks (Neva and Narva) to wild Vasalemma, Keila and Loobu populations.

Table S1 Microsatellite diversity indices of the studied Atlantic salmon populations and the results of the Hardy–Weinberg equilibrium test.

Table S2 Pairwise genetic distances (F_{ST}) among salmon samples based on 17 microsatellite loci.

Table S3 Structure statistic output and the ad hoc statistic of Evanno *et al.* (2005) to determine the most probable *K* for all comparisons.

Table S4 Estimation of null allele frequencies (Brookfield 1) across studied microsatellite loci and samples.

Table S5 Allele frequencies (%) for all samples by microsatellite locus.

Table S6 Pairwise genetic distances (F_{ST}) among salmon samples based on 1986 SNP loci.

Table S7 Power assessment of 17 microsatellite markers for correctly assigning simulated salmon individuals to pure (Neva or Narva hatchery and wild Vasalemma, Keila, Loobu or Kunda), F1 hybrid, F2 hybrid, and backcrossed (F1 × hatchery and F1 × wild) categories based on NEWHYBRIDS.

Table S8 Genomewide genetic divergence (F_{ST}) and introgression between hatchery and wild populations estimated from SNP data.

Table S9 The level of genetic introgression ($\hat{S}$) and its respective *P*- and *Q*-values for each SNP locus.