Investigating genetic change in wild populations: modelling gene flow from farm escapees

F. Besnier*, K. A. Glover, Ø. Skaala

Institute of Marine Research, PO Box 1850, 5817 Nordnes, Bergen, Norway

ABSTRACT: Fish-farm escapees threaten the genetic integrity of wild populations. Because significant genetic differences often occur among aquaculture strains, gene flow from multiple farmed sources may compromise the ability to detect genetic change in wild populations. Here, we investigate this situation by simulating genetic change based upon data from 9 microsatellite loci in 4 wild Atlantic salmon populations receiving variable mixtures of escapees from 5 commercial Norwegian strains. As expected, neutral markers detected genetic change in wild populations when gene flow came from a single and distinct farmed strain. However, the genetic change detected in the wild population was significantly lower when gene flow was simulated from multiple farm strains simultaneously ('concealing effect'). Although the degree of concealing varied among the wild populations, in one of the cases, no significant genetic differentiation was detected when 20% effective migration was simulated from 2 or more farmed strains for up to 10 generations. While individual admixture analysis succeeded in detecting introgression, it was nevertheless strongly underestimated. Where gene flow is expected to originate from multiple farmed sources, analysis of selectively neutral genetic markers is thus likely to underestimate the true level of genetic introgression.

KEY WORDS: Farm escapee · Wildlife forensics · Genetic impact · Microsatellite · Gene flow · Salmon farming

INTRODUCTION

Fish-farm escapees represent a significant threat to the genetic integrity of wild populations where domesticated and wild forms co-exist (http://genimpact. imr.no/). Every year, thousands of individuals of the 3 major farmed species in Norway, Atlantic salmon *Salmo salar*, rainbow trout *Oncorhynchus mykiss*, and Atlantic cod *Gadus morhua*, escape into the wild (Directorate of Fisheries 2009). Specifically for the Atlantic salmon, some of these escapees enter freshwater and have been documented to successfully spawn with wild fish (Saegrov et al. 1997).

Farmed Atlantic salmon have been subject to domestication selection, and differ from wild salmon in allele frequencies for selectively neutral genetic markers (Skaala et al. 2004, 2005, Verspoor et al. 2005, Ferguson et al. 2007), allele frequencies for single-nucleotide polymorphisms potentially under selection (Karlsson et al. 2011), production-related traits including growth (e.g. Glover et al. 2009a), gene expression (Roberge et al. 2006, 2008), behaviour (Einum & Fleming 1997, Flemming & Einum 1997), and physiology (Johnsson et al. 2001, Fleming et al. 2002). In addition, results of field studies conducted in both Irish and Norwegian rivers have demonstrated that offspring of farmed salmon display significantly reduced survival when compared to offspring of native wild salmon (McGinnity et al. 1997, 2003, Fleming et al. 2000). Considering these differences together with the fact that Atlantic salmon displays considerable population genetic struc-

ture (Verspoor et al. 2005), and that populations may show adaptations to their native rivers (Taylor 1991, Garcia de Leaniz et al. 2007), it is generally accepted that farm escapees represent a threat to the fitness of wild Atlantic populations.

One of the first steps in elucidating the potential genetic impact of farm escapees on wild populations is to quantify genetic introgression using molecular genetic markers. Early studies conducted in Ireland using allozyme markers managed to detect genetic changes in wild Atlantic salmon populations (measured as pairwise F_{ST} between wild populations before and after suspected gene flow) in response to fish-farm escapees (Crozier 1993, 2000, Clifford et al. 1998). These studies were conducted in response to a known escapement from a single farm where samples could be taken from the wild population prior to farmed fish spawning. More recent studies also aimed at evaluating the long-term impact of the genetic introgression of farmed fish into wild populations (Koskinen et al. 2002, Hansen et al. 2009, Bourret et al. 2011). In the Finnish lake systems, Koskinen et al. (2002) evaluated the genetic differentiation of grayling *Thymallus thymallus* populations by comparing wild and stocked samples over a period of 15 yr. In North American populations of Atlantic salmon, the temporal change in genetic integrity was evaluated by comparing historical samples from 1980 to contemporary samples from 2005 (Bourret et al. 2011). Even though the above 2 studies (Koskinen et al. 2002, Bourret et al. 2011) differ in their conclusions about the capacity for resilience of the wild population exposed to genetic introgression, they both provided evidence that genetic introgression from farm escapees or from fish stocking did occur in the studied wild populations.

In Norwegian salmon populations, a study of genetic stability over a 20 yr period revealed significant temporal genetic changes in 3 rivers that had received large numbers of farm escapees for many years (Vosso, Opo, Eio) (Skaala et al. 2006). In the same study, no genetic changes were reported in 3 other rivers (Etne, Granvin, Neiden) despite a large number of observed escapes. Over time, it is to be expected that fish-farm escapees will originate from a number of sources displaying different genetic backgrounds. For example, there are often large and statistically highly significant genetic differences observed among groups of salmon reared on commercial farms (e.g. Glover 2008, Glover et al. 2009b, 2010a). This reflects both inter-strain (Skaala et al. 2004, 2005) and intra-strain (Glover et al. 2009b, 2010a, Glover 2010) variation. Furthermore, it is sug-

gested that very distinct genetic diversity among the farm sources might hamper the power to detect the signature of gene flow from farm escapees into the wild populations. This is due to the fact that one is not attempting to quantify the gene flow from a farm strain 'X' to wild population 'Y', but from farmed strains 'X$_1$, X$_2$, X$_3$, X$_4$, X$_5$,…' into wild population 'Y'. However, the possibility of farm escapees originating from multiple sources is not yet explicitly considered in studies of temporal stability of wild populations (Skaala et al. 2006, Bourret et al. 2011).

In the present study, we investigated the potential complexity of quantifying genetic changes in wild populations as a result of gene flow from a variety of farmed strains, as opposed to a single strain. Using data from Skaala et al. (2004), we carried out simulations of gene flow using different contributions of farmed strains and wild populations.

MATERIALS AND METHODS

Data set

The simulations conducted here were based on a data set of 9 microsatellite loci amplified in 832 salmon originating from 5 of the major farmed strains used in Norwegian Atlantic salmon culture (hereinafter referred to as DOM-1.95, DOM-2.96, DOM-5.95, DOM-3.97 and DOM-4.94), and 4 wild Norwegian Atlantic salmon populations sampled between 1995 and 1997 (Namsen, Lone, Vosso and Neiden) (Skaala et al. 2004). The genetic origin of many of these strains has been described previously (Gjedrem et al. 1991, Skaala et al. 2004, Glover et al. 2009a). In short, all of the commercial strains currently reared in Norwegian aquaculture are founded upon wild Atlantic salmon collected from multiple Norwegian rivers or coastal areas in the late 1960s and early 1970s. Individuals from the sampled wild populations were mixed, intercrossed and selectively bred for production-related traits (Gjedrem 2000). Based upon a 4 yr aquaculture generation time, the genetic data for the farmed strains analysed were approximately 6 to 7 generations of selective breeding.

The 9 microsatellite markers upon which the farmed and wild data set were established were shown to fit Hardy-Weinberg proportions (Skaala et al. 2004), and a new analysis of the data using BayeScan software (Foll & Gaggiotti 2008) indicated that 8 out of the 9 loci were selectively neutral, whereas locus SSaF43 (Sanchez et al. 1996) was possibly under mild positive selection (Log(Bayes Factor) = 1.83).

Simulations

To investigate the effect of migration from multiple as opposed to single farmed strains into a given wild population, we implemented a simulation routine in R (R Development Core Team 2010). The approach chosen for this simulation was to assume discrete and non-overlapping generations with an effective population size of 200 individuals for the wild population. A given proportion of farmed migrants were then introduced into the initial wild population, and the next generation was simulated by assuming random mating, equal sex proportion and equal mating probability for wild and farmed conspecifics. The proportion of farmed migrants introduced into the wild population represents the effective migration rate, or gene flow, i.e. animals that physically migrate and are able to breed with the wild population.

In turn, each wild population received the same proportion of farmed migrants per generation: for a population size of 200, an effective migration rate of e.g. 10% would bring 20 farm animals every generation into the wild population. The migration-mating routine was simulated for 10 consecutive generations.

Simulations were repeated for each wild population receiving 0, 5, 10, 15 and 20% farmed migrants (for graphical reasons, data only depict 0, 10 and 20% migration in the associated illustrations). Together with the proportion of migrants, we also changed the composition of the migrant population from 1 single farmed strain to a mixture of the 5 farmed strains in equal proportions. To ease the comparison among scenarios, this was computed by introducing the genetically most similar farmed strain first, and adding the next most similar farmed strains subsequently until genetic introgression was simulated for equal proportions of the 5 strains (i.e. 10% gene flow from all 5 strains would be represented by 4 salmon from each of the 5 strains). Under the hypothesis that the genetic change occurring in a wild population is directly linked to the genetic distance between the wild and immigrant farm populations, we created a set of scenarios where the expected genetic change in a given wild population receiving escapees from i different farms is always smaller than the genetic change expected in the same wild population receiving migrants from j different farms, where $i < j$.

Genetic differentiation among populations

Despite the recent development of statistical and computational approaches to evaluate the genetic composition of a population (Pritchard et al. 2000) or the migration rate between populations (Dyer & Nason 2004, Wilson & Rannala 2003), the measure of population differentiation by pairwise F_{ST} is still commonly used alone (Skaala et al. 2006), or along with other approaches (Bourret et al. 2011), to evaluate the impact of gene flow from farm or hatchery fish into wild populations. In the present study, the genetic distance between historical (i.e. any given wild population prior to gene flow) and contemporary (i.e. the same wild population following various gene flow simulations) populations was estimated using pairwise F_{ST} (Weir & Cockerham 1984). After each step of the migration-mating routine of our simulations, F_{ST} (P$_0$, P$_g$) was measured as the genetic distance between the initial wild population at Generation 0 (P0) and the wild population at Generation g (P$_g$) as illustrated (Fig. 1). When the migration rate is set at zero, the measured F_{ST} (P$_0$, P$_g$) represents the genetic differentiation that can be expected in the specific wild populations due to genetic drift. One hundred replicates of each simulation scenario were performed to obtain the average and standard deviation for F_{ST} values.

Multivariate analysis

The use of a single summary statistic such as F_{ST} (Weir & Cockerham 1984) or Nei's D (Nei 1972, 1978) is not necessarily sufficient to adequately capture inter-population relationships (Dyer & Nason 2004).
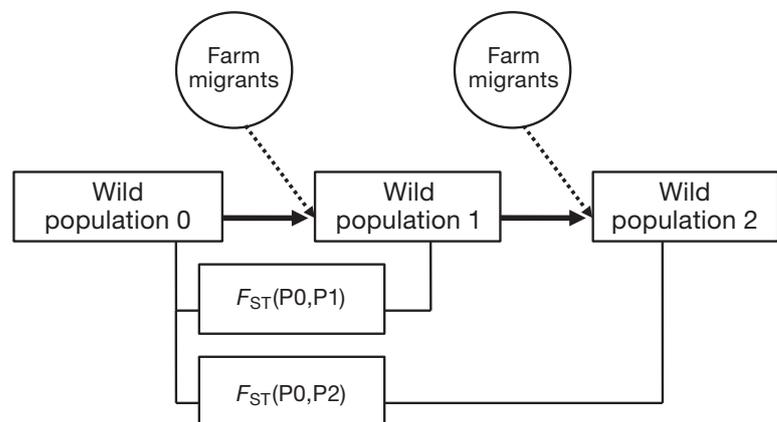

Fig. 1. Workflow of the simulation process that iterates migration and random mating, and calculation of F_{ST} between the initial population and each new generation (g)

As the interrelationship interrelation between populations is measured using their genotypes at a number n of loci with m independent combination of alleles, each population can be represented in a multidimensional space consisting of m orthogonal axes. To summarise this data in a comprehensible format, and better understand the structure of the set of populations studied here, we performed a multivariate analysis using the R package 'adegenet' (Jombart 2008).

Assignment methods

In addition to genetic statistics based on average group values, i.e. F_{ST}, we also explored the possibility of detecting introgression by methods of individual assignment, following the approach of Pritchard et al. (2000) implemented in the software STRUCTURE. For each wild population, 2 scenarios were compared: first, genetic introgression originating from 1 single farm strain, and second, genetic introgression originating from all 5 farm strains in equal proportions, both with 20% gene flow per generation. The data analysed in each scenario consisted of 8 populations: the 5 farm strains and 1 wild population at 3 different stages (historical, post-5 and post-10 generations of introgression, respectively). For each scenario, the number K of groups included in the assignment model was tested for $K = 1$ to $K = 8$, without using *a priori* population information. All simulations were conducted with a 100 000 burn-in, 1 000 000 iterations and correlated allele frequencies. After estimating the value of K that provided the best fit for the present data, 50 replicates of the model assuming $K = 6$ were computed to test the power of detecting gene flow by use of an admixture model. The average assignment of the animals into each group was then compared for the wild population before and after migration.

The same simulated data were also analysed using the method of Wilson & Rannala (2003) implemented in the software BayesAss (http://rannala.org) to estimate the migration rate into the wild population.

RESULTS

Genetic differentiation among populations

When simulating gene flow from a single farmed strain at a time, taking the genetically most similar farmed strain, DOM-2.96 (Fig. 2A), and the genetically most distant farmed strain, DOM-1.95 (Fig. 2B),

the genetic change as measured by F_{ST} was readily detected in the Vosso population for both simulations. Furthermore, the rate of F_{ST} change was, as expected, correlated with the level of gene flow. Without migration, the wild population did not change significantly after 10 generations of random mating. For both DOM-2.96 and DOM-1.95, an effective gene flow of 20% per generation for 10 generations was enough for the wild population to be almost as distant from the initial population as from the farm strain (i.e. there is almost total fixation of allele frequencies with the farmed strain from which the simulated migrants are taken).

Moving from the simulations of gene flow from single farmed strains into the Vosso population (Fig. 2A,B) to the more complicated situation where gene flow was simulated from multiple farmed strains simultaneously (Fig. 2C–F), we observed that, despite the fact that each additional strain included in the migrating pool was more distinct from Vosso than the previous one, the resultant genetic change did not increase past the level caused by migration from the genetically most similar strain, DOM-2.96, alone (Fig. 2A,F). Furthermore, the rate of change (i.e. number of generations needed to attain a given F_{ST} value for any given level of migration) was slower for the situation where gene flow was simulated from the combination of either 2, 3, 4 or all 5 farmed strains as opposed to the single most similar strain. This indicates a concealing effect, where genetic change detected after genetic introgression from a heterogeneous pool of migrants appears smaller than the genetic changes detected after genetic introgression from the same number of migrants originating from 1 single farm. Nevertheless, in all the combinations of farmed strains tested, F_{ST} ranging between 0.067 ± 0.006 and 0.016 ± 0.003 was observed in the Vosso population following 5 generations of 20% gene flow (Fig. 2, Table 1). These values differ significantly (p < 0.01) from the null model where genetic difference is only caused by genetic drift. This means that in the River Vosso, genetic introgression of farm escapees into the wild population would always be detected with neutral markers; however, the level of gene flow would likely be underestimated.

The extent of the concealing effect observed when adding gene flow from multiple farmed strains as opposed to single strains is population dependent (Table 1). In 3 wild populations (Neiden, Lone and Vosso), the concealing effect exists: the scenario where DOM-2.96 is the only source of migrants leads to a genetic change equivalent to that in the scenario where all 5 farm strains release migrants in equal
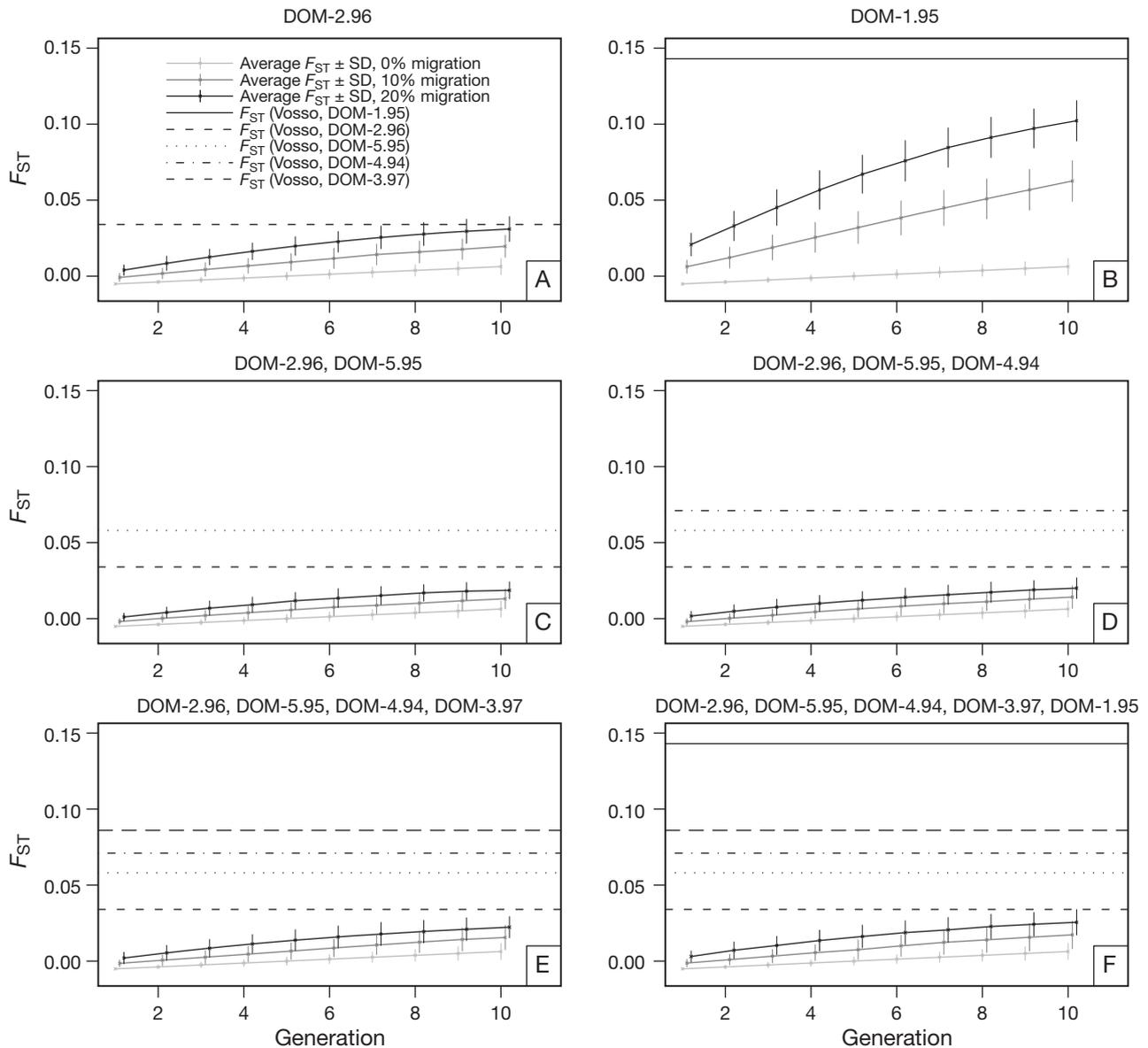
Fig. 2. *Salmo salar*. Genetic differentiation (F_{ST}) of the Vosso population over 10 generations of migration from (A) the genetically closest farm population, (B) the genetically farthest farm population and (C–F) several sources simultaneously, starting from (C) the 2 closest populations, (D) the 3 closest, (E) the 4 closest and (F) all 5 farm populations

proportions. In these situations, genetic change in the wild populations could be detected with neutral markers after 5 to 10 generations of introgression and 10 to 20% gene flow (Table 1). In the case of the Namsen population, however, the concealing effect was considerably stronger (Fig. 3). Here, the genetic change induced by migrants coming from all 5 farm strains was significantly smaller than that induced by DOM-2.96 alone (i.e. the genetically most similar farmed strain). As a consequence, in any scenario where the River Namsen received migrants from >1 farmed strain simultaneously, no genetic change

could be detected with neutral markers, even in the most extreme case of our simulations (10 generations and 20% gene flow).

To investigate the influence of multiple farmed-strain gene flow on the estimate of genetic distance among the 4 wild populations, we compared the pairwise genetic distance between them prior to and post-gene flow. Simulations were conducted for 100 replicates, where each wild population received migrants from the 5 farm strains in equal proportions, during 10 generations, and with 20% gene flow. The results of these simulations (Table 2) indicate that

Table 1. *Salmo salar*. Genetic differentiation between the 4 historical (pre-gene flow) and simulated (post-gene flow) wild populations (Namsen, Neiden, Lone, Vosso) after introgression by farm migrants (DOM-2.96 alone; DOM-1.95 alone; or all 5 farm strains: DOM-2.96, DOM-1.95, DOM-5.95, DOM-3.97 and DOM-4.94) for 5 or 10 generations of either 10% or 20% gene flow. The p-values (in parentheses) are for the test of whether genetic distance induced by introgression is significantly higher than when only due to drift. Average of the 5 pairwise F_{ST} between the wild populations prior to gene flow and each of the 5 farm strains

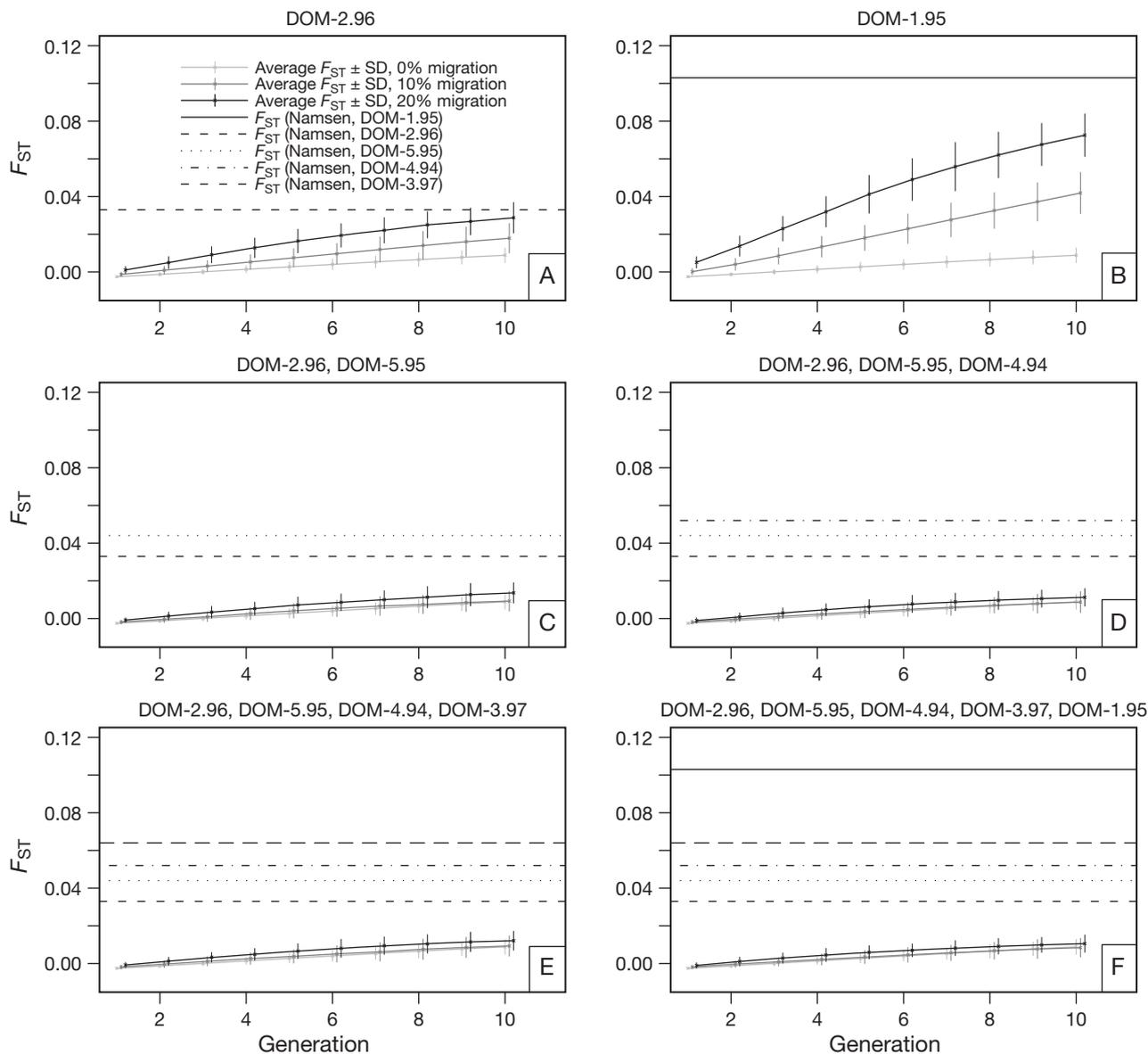| Population	% Gene flow	F_{ST} change after 5 generations			F_{ST} change after 10 generations			Average pairwise F_{ST}
		DOM-2.96	DOM-1.95	5 farms	DOM-2.96	DOM-1.95	5 farms	
Namsen	10	0.0075 (0.023)	0.0180 ($<10^{-6}$)	0.0034 (0.371)	0.0178 (0.012)	0.0418 ($<10^{-6}$)	0.0083 (0.573)	0.0592
Namsen	20	0.0164 ($<10^{-6}$)	0.0412 ($<10^{-6}$)	0.0059 (0.051)	0.0287 ($<10^{-6}$)	0.0725 ($<10^{-6}$)	0.0105 (0.296)	
Neiden	10	0.0149 ($<10^{-6}$)	0.0258 ($<10^{-6}$)	0.0113 ($<10^{-3}$)	0.0338 ($<10^{-6}$)	0.0560 ($<10^{-6}$)	0.0248 ($<10^{-3}$)	0.0963
Neiden	20	0.0318 ($<10^{-6}$)	0.0561 ($<10^{-6}$)	0.0242 ($<10^{-6}$)	0.0528 ($<10^{-6}$)	0.0931 ($<10^{-6}$)	0.0404 ($<10^{-6}$)	
Lone	10	0.0215 ($<10^{-6}$)	0.0339 ($<10^{-6}$)	0.0131 ($<10^{-6}$)	0.0396 ($<10^{-6}$)	0.0626 ($<10^{-6}$)	0.0259 ($<10^{-6}$)	0.0942
Lone	20	0.0421 ($<10^{-6}$)	0.0691 ($<10^{-6}$)	0.0262 ($<10^{-6}$)	0.0602 ($<10^{-6}$)	0.1009 ($<10^{-6}$)	0.0373 ($<10^{-6}$)	
Vosso	10	0.0092 ($<10^{-3}$)	0.0319 ($<10^{-6}$)	0.0074 (0.005)	0.0196 (0.002)	0.0625 ($<10^{-6}$)	0.0173 (0.014)	0.0783
Vosso	20	0.0198 ($<10^{-6}$)	0.0670 ($<10^{-6}$)	0.0162 ($<10^{-6}$)	0.0309 ($<10^{-6}$)	0.1022 ($<10^{-6}$)	0.0255 ($<10^{-6}$)	

Fig. 3. *Salmo salar.* Genetic differentiation (F_{ST}) of the Namsen population over 10 generations of migration. Panels (A) to (F) as in Fig. 2 but for the Namsen population

Table 2. *Salmo salar*. Genetic differentiation (pairwise F_{ST}) between historical (prior to gene flow) wild populations (below diagonal) and simulated wild populations after 10 generations and 20 % gene flow (above diagonal). p-values are given in parentheses

	Namsen	Neiden	Lone	Vosso
Namsen	0	0.007 (0.0080)	0.006 (0.0260)	0.006 (0.0046)
Neiden	0.038 (0.0008)	0	0.007 (0.0080)	0.008 (0.0110)
Lone	0.026 (0.0178)	0.066 (0.0010)	0	0.007 (0.0130)
Vosso	0.014 (0.0856)	0.050 (0.0084)	0.038 (0.0384)	0

introgression of farmed salmon, even originating from multiple farmed strains, considerably reduced the pairwise genetic distance between all wild populations. For example, the distance between the Namsen and Lone historical populations was $F_{ST} = 0.026$,

whereas the average distance between them after 10 generations of introgression was only $F_{ST} = 0.006$.

Multivariate analysis

Fig. 4 represents all 9 populations in the 6 main axes of the principal component analysis (PCA), which summarises 85 % of the total variation. The first 2 principal components (Fig. 4A) revealed that 3 of the wild populations (Namsen, Lone and Vosso) gathered in the same cluster, whereas the farmed strains were more clearly differentiated and positioned around the first cluster. Only the wild population Neiden was posi-
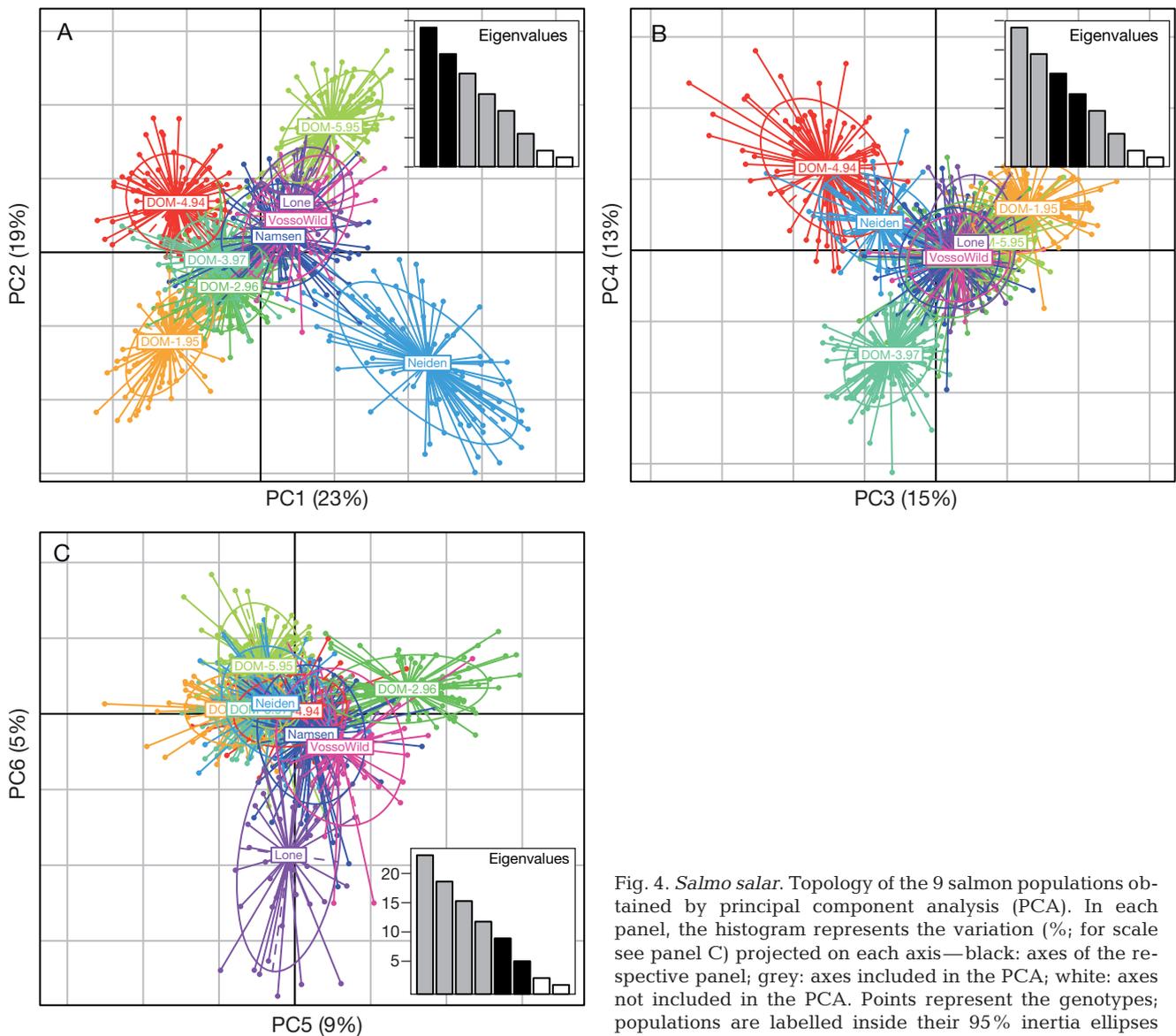

Fig. 4. *Salmo salar*. Topology of the 9 salmon populations obtained by principal component analysis (PCA). In each panel, the histogram represents the variation (%; for scale see panel C) projected on each axis — black: axes of the respective panel; grey: axes included in the PCA; white: axes not included in the PCA. Points represent the genotypes; populations are labelled inside their 95 % inertia ellipses

tioned further from the cluster and the farm strains, both on the 1st and 2nd axis. The 3rd and 4th principal components (Fig. 4B) presented a similar picture to that presented in Fig. 4A, whereas the 6th principal component (Fig. 4C) was mostly discriminating the wild population Lone from the other populations. The PCA supports earlier observations that the genetic distance is larger among farm strains than wild populations (Skaala et al. 2004), but importantly, there is no 'common direction' to the genetic differentiation between wild and farmed samples analysed for the neutral markers here. Indeed, the wild and farm populations do not form 2 distinct groups; instead, the farm strains form a cloud around the wild populations that are grouped together.

Assignment methods

Individual assignment tests were performed on a set of populations consisting of 5 farm strains and 1 wild population after 0, 5 and 10 generations of migration from all 5 farm strains. The results indicated that the most likely number of populations was

$K = 6$, where each farm strain was represented by a different population, and where the wild population appeared as a mixture of genetic backgrounds from the different farm strains and a 6th population (Fig. 5). This result corroborates the PCA where all farm strains are clearly differentiated from each other and from the wild populations, whereas wild populations sit on an intermediate position between the different farm strains.

The migration of animals from all 5 farm strains into Vosso (Fig. 5A) and Namsen (Fig. 5B) led to a progressive replacement of the main local group (in red in Fig. 5A and in purple in Fig. 5B) by a blend of genetic backgrounds of all farm and wild origins. When migration originated from 1 single farm, as illustrated in Fig. 5C, where only animals from DOM-2.96 migrated into Namsen, the local population (in pink in Fig. 5C) is progressively replaced by a homogeneous population similar to DOM-2.96.

To better quantify the rate of gene flow captured by the admixture model in those 3 scenarios, we compared, over 50 replicates, the average individual assignment of the animals into each group for the wild population before and after migration (Table 3).
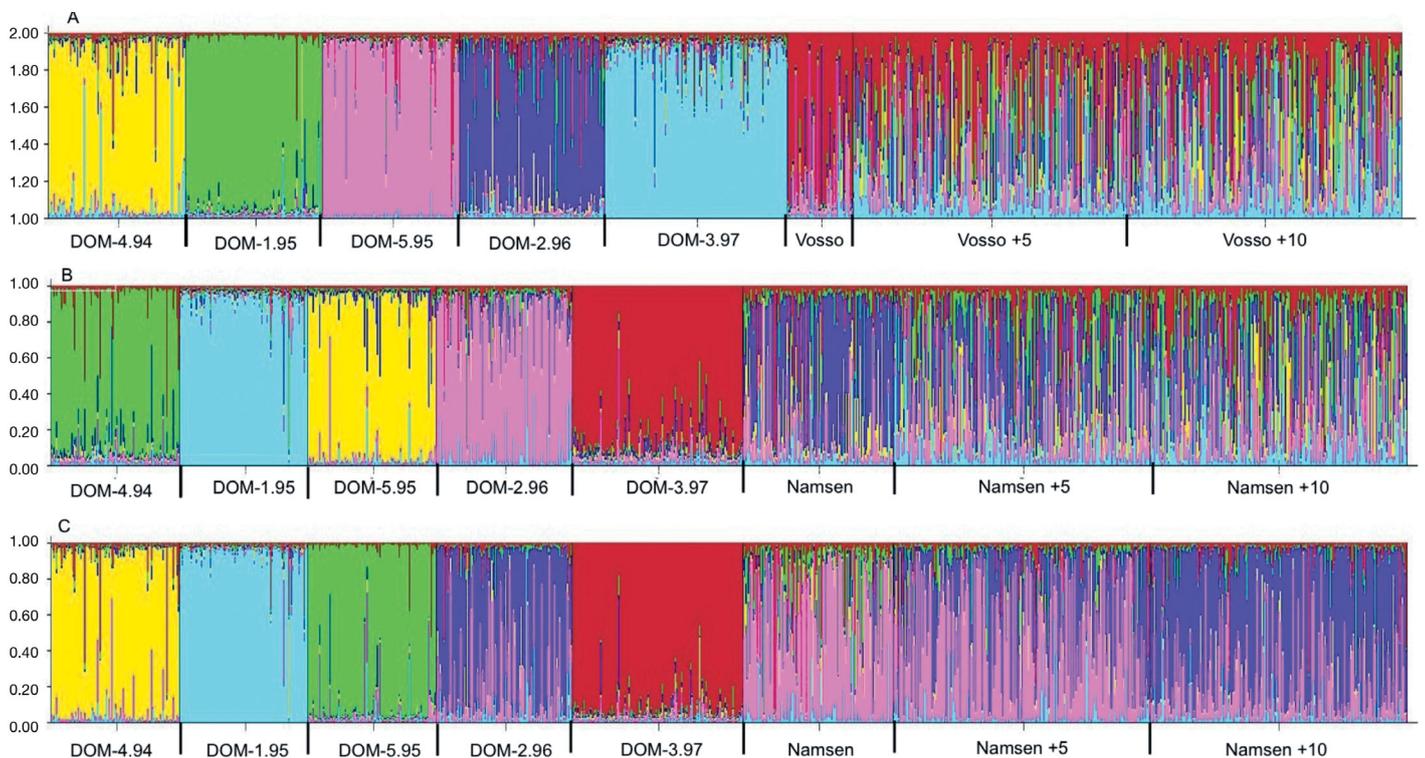

Fig. 5. *Salmo salar*. Bar plots of individual assignment of 8 groups consisting of 5 farm strains (DOM-4.94 … DOM-3.97) and 1 wild population (Vosso or Namsen, respectively) after 0, 5 and 10 generations of migration from farms into 6 population groups (named 1 to 6; see Table 3) determined using the software STRUCTURE. The assignment probability is given on the *y*-axis. The most likely number of populations is 6, where each farm strain is represented by a different population, and the wild population is a mixture of the different farm strains and a 6th population. (A) Vosso and (B) Namsen population after introgression from all 5 farm strains in equal proportions and (C) Namsen population after introgression from DOM-2.96

Table 3. *Salmo salar*. Average individual assignment of the populations into 6 groups determined using the software STRUCTURE. Numbers are proportions of the total for any given column; those in **bold** represent the strongest assignment. 'Historical' indicates prior to any gene flow, and +5 and +10 indicate after 5 and 10 generations of 20% gene flow, respectively

Population group	Farm strains					Gene flow from all 5 farms to Vosso			Gene flow from all 5 farms to Namsen			Gene flow from DOM-2.96 to Namsen	
	DOM-4.94	DOM-1.95	DOM-5.95	DOM-2.96	DOM-3.97	Historical	+5	+10	Historical	+5	+10	+5	+10
1	**0.83**	0.02	0.01	0.04	0.02	0.03	0.12	0.13	0.07	0.13	0.16	0.04	0.03
2	0.02	**0.90**	0.02	0.05	0.02	0.02	0.08	0.10	0.05	0.10	0.11	0.05	0.04
3	0.03	0.01	**0.85**	0.02	0.02	0.14	0.13	0.14	0.13	0.12	0.14	0.05	0.04
4	0.02	0.02	0.02	**0.75**	0.03	0.11	0.16	0.16	0.11	0.16	0.17	0.37	**0.65**
5	0.05	0.02	0.03	0.03	**0.85**	0.03	0.14	0.17	0.10	0.13	0.16	0.05	0.05
6	0.04	0.02	0.06	0.09	0.05	**0.63**	**0.33**	**0.27**	**0.52**	**0.33**	**0.25**	**0.43**	0.19

For each farm strain, 75 to 90% of the animals were associated with 1 corresponding population (Groups 1 to 5 in Table 3). For the historical wild populations, however, the average association of the animals into the majority group (Group 6) is only 63% for Vosso and 52% for Namsen.

When animals originating from DOM-2.96 migrated into Namsen, the average assignment of the Namsen population to Group 6 decreased from 52 to 19% after 10 generations. The main group assigned to the Namsen population was no longer Group 6, but Group 4, with 65% individual assignment (Table 3). In contrast, when animals originating from all 5 farm strains migrated into 1 wild population, Namsen and Vosso remained mainly assigned to Group 6. The assignment of Namsen to Group 6 decreased from 52 to 25%, and the average assignment of Vosso animals into Group 6 decreased from 63 to 27%.

Additionally, the analysis of our simulated data using the method of Wilson & Rannala (2003) implemented in the software BayesAss provided an estimated migration rate that was 30% larger when migrants originated from 1 single farm than when they originated from 5 different farms in equal proportions. Besides the estimation of migration rate, results from BayesAss were similar to those from STRUCTURE.

DISCUSSION

The simulations conducted here demonstrate that the rate and magnitude of genetic change detected by the analysis of neutral markers is subject to a concealing effect when wild populations are exposed to migrants originating from multiple as opposed to single farmed strains. In all 4 wild populations, the genetic difference observed between the historical and the contemporary population was always smaller when the population received migrants from multiple sources than when it received migrants from a single farmed strain. Despite admixture providing greater power to detect genetic introgression, the concealing effect was also observed when using the software STRUCTURE (Pritchard et al. 2000) as well as BayesAss (Wilson & Rannala 2003). The majority of individuals from the Namsen population were indeed assigned to the same group before and after migration when migrants originated from multiple sources, whereas Namsen's individual assignment clearly changed from Group 6 to Group 4 (Table 3) when migration was simulated from 1 unique farm strain. It is therefore concluded that the analysis of neutral genetic makers is likely to lead to an underestimation of the level of gene flow when migrants originate from several distinct populations. On the other hand, the information provided by the loci under selection is expected to provide better power to discriminate farm and wild conspecifics (Bourret et al. 2011).

The broadly used F_{ST} and individual-assignment approaches can only provide a proxy for the true migration rate, whereas specific assignment methods that were developed to answer the problem of identifying the composition of mixture stocks, e.g. in fishery management (Wilson & Rannala 2003), can provide a direct estimate of migration rate. The aim of this approach is to estimate the stock composition of a fishery, in order to take advantage of abundant populations without imposing excessive pressure on less abundant ones (Manel et al. 2005). Such methods are, however, not currently used in studies for conservation of wild populations exposed to farm migrants. Moreover, the possibility of farm escapees originating from multiple sources is not yet explicitly considered in studies of temporal stability of wild populations (Skaala et al. 2006, Bourret et al. 2011). In our simulated data, such an approach also

produced an underestimated migration rate when migrants originated from multiple farms.

Skaala et al. (2006) investigated temporal genetic stability in Atlantic salmon populations affected to varying degrees by farm escapees. By analysing 8 microsatellite markers in historic (prior to major expansion of the aquaculture industry; i.e. from 1971 to 1995) and contemporary samples, these authors were able to detect significant temporal genetic changes in 3 of the rivers studied (Vosso, Opo, Eio). All of these rivers had reported large numbers of farm escapees on the spawning grounds, and the authors concluded that the observed changes most likely represented genetic introgression. For the River Vosso, Skaala et al. (2006) reported an F_{ST} of 0.025 over the approximately 20 yr period. This is slightly higher than the level of genetic change observed in the River Vosso in the present study (F_{ST} of 0.016) when simulating 20% effective gene flow from multiple farmed strains over 5 generations (equals approximately 20 yr). Therefore, the simulations reported in the present study support the conclusions by Skaala et al. (2006) that the River Vosso has experienced considerable gene flow from farmed salmon. In contrast to the 3 rivers that did experience genetic change (Vosso, Opo, Eio), Skaala et al. (2006) did not observe genetic change in 3 other rivers (Etne, Granvin, Namsen), all of which had reported significant numbers of escapees over several years (up to 79% farm escapees for the River Etne in 1 yr). The same authors concluded that this might reflect a lack of genetic introgression in these populations. While this remains a possibility, due to the fact that farmed escaped salmon display significantly reduced spawning success compared to wild conspecifics (Fleming et al. 1996, 2000), the concealing effect observed in the present study's simulations suggests that an alternative explanation might be that the markers used by Skaala et al. (2006) did not accurately detect the gene flow that might have occurred. Directly supporting this idea is the fact that the River Namsen was included in both studies, showing no genetic change either in the empirical data set reported by Skaala et al. (2006) or in the present study's simulations when gene flow was simulated from multiple farmed strains (even after 10 generations and 20% gene flow). While it is important to bear in mind that our simulations are intended to compare the qualitative genetic change of populations in different scenarios of genetic introgression rather than to predict a true level of gene flow, the results of the present study clearly overlap with the observed F_{ST} change reported by Skaala et al. (2006).

The use of non-overlapping generations made the simulation procedure simpler despite the life-history trait of salmon which, with 1 breeding period per year and an average of approximately 4 yr as age of sexual maturity, shows clear overlapping generations. The dynamic of the process of genetic change described in the present study may thus be biased. However, our aim was not to provide an exact estimate of parameters, such as time and amount of gene flow, but rather to explore the precision in currently employed methods for detecting gene flow and give a qualitative description of the phenomenon when a wild population receives escapees from several different farm sources as opposed to a single strain/ farm. Our simulations therefore represent a tool to explore and compare different scenarios of migration rate and composition of migrant populations. The effective population size (Ne) for the simulations in the present study was fixed at Ne = 200. This is within the range of values estimated from other Atlantic salmon populations (Borrell et al. 2008, Dillane et al. 2008, Palstra et al. 2009). The effective population size varies among rivers, and likewise, the spawning success of farm migrants is expected to vary between rivers. For example, it is likely that the spawning success of farm escapees is higher in rivers with a low density of wild spawners as opposed to a river with a high density of farm escapees due to competition (Fleming et al. 1996, 2000). In our simulations, the variability of spawning success of the farm migrants was not taken into account; only the effective migration was considered. Nevertheless, it is possible that 10% effective migration (i.e. gene flow) can be caused by 10% of farm escapees on the spawning ground of a given river where conditions favour spawning success of farm escapees; however, it is also possible that far higher percentages of farm escapees on the spawning grounds of another river, where conditions do not favour spawning success of farmed fish, may be required to result in 10% effective migration (gene flow).

All strains reared in Norwegian fish farms originate from Norwegian rivers (Gjedrem et al. 1991, Skaala et al. 2004, Glover et al. 2009a). Although the source wild populations were intercrossed and selected for production-related traits (Gjedrem 2000), they still display overlapping allele frequencies with wild populations as opposed to fixed allele differences (Skaala et al. 2004). Thus, it is likely that the magnitude of the concealing effect reported in the present study is magnified by the particular situation of salmon aquaculture in Norway. Nevertheless, the concealing effect described here is likely to also

occur in other studies, where one wild population receives migrants from multiple distinct sources. Studies of farmed Atlantic cod (Glover et al. 2010b) and farmed rainbow trout (Glover 2008) have revealed large genetic differences among groups of fish reared in different farms. This is similar to the situation observed for Atlantic salmon, where inter-cage variation reflects both inter- and intra-strain genetic differentiation (Glover 2008, 2010, Glover et al. 2009b, 2010a). While rainbow trout is not native to Europe, thus circumventing any direct farmed–wild genetic interactions, Atlantic cod is reared in cages that are often located in coastal regions where wild cod exist. Atlantic cod are notoriously effective at escaping from net pens (Moe et al. 2007), and consequently represent a threat to the genetic integrity of the wild populations. The simulations conducted here, and importantly, the observed concealing effect on detection of genetic introgression in wild populations, are thus likely to occur for other aquaculture species such as Atlantic cod where genetic differences exist among the farmed strains. Indeed, a recent study of farmed escaped cod in the wild illustrated some of the complexities of detecting genetic changes in wild populations (Glover et al. 2011).

The results of the present study indicate that accurate quantification of genetic introgression of farmed salmon in wild populations is likely to be strongly underestimated when based on neutral markers. This highlights the need for the identification of genetic markers that are diagnostic on the farmed–wild interface irrespective of the population of origin in order to provide a tool capable of effectively quantifying the true level of genetic introgression in wild populations. For Atlantic salmon, which has been under domestication since breeding programmes were started in the early 1970s, a set of single-nucleotide polymorphisms has recently been identified to discriminate this species along the wild and farmed interface (Karlsson et al. 2011). However, for other aquaculture species where the level of domestication is not so advanced (e.g. Atlantic cod), the identification of domestication-linked markers will be more challenging.

LITERATURE CITED

Borrell YJ, Bernardo D, Blanco G, Vazquez E, Sanchez JA (2008) Spatial and temporal variation of genetic diversity and estimation of effective population sizes in Atlantic salmon (*Salmo salar* L.) populations from Asturias (Northern Spain) using microsatellites. Conserv Genet 9:807−819

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

Clifford SL, McGinnity P, Ferguson A (1998) Genetic changes in an Atlantic salmon population resulting from escaped juvenile farm salmon. J Fish Biol 52:118−127

Crozier WW (1993) Evidence of genetic interaction between escaped farmed salmon and wild Atlantic salmon (*Salmo salar* L.) in a northern Irish river. Aquaculture 113:19−29

Crozier WW (2000) Escaped farmed salmon, *Salmo salar* L., in the Glenarm River, Northern Ireland: genetic status of the wild population 7 years on. Fish Manag Ecol 7: 437−446

Dillane E, McGinnity P, Coughlan JP, Cross MC and others (2008) Demographics and landscape features determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. Mol Ecol 17:4786−4800

Directorate of Fisheries (2009) Key figures from aquaculture industry. Directorate of Fisheries, Bergen

Dyer RJ, Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. Mol Ecol 13:1713−1727

Einum S, Fleming IA (1997) Genetic divergence and interactions in the wild among native, farmed and hybrid Atlantic salmon. J Fish Biol 50:634−651

Ferguson A, Fleming IA, Hindar K, Skaala Ø, McGinnity P, Cross T, Prodöhl P (2007) Farm escapes. In: Verspoor E, Nielsen J, Stradmeyer L (eds) The genetics of Atlantic salmon: implications for conservation. Blackwell, Oxford, p 363

Fleming IA, Einum S (1997) Experimental tests of genetic divergence of farmed from wild Atlantic salmon due to domestication. ICES J Mar Sci 54:1051−1063

Fleming IA, Jonsson B, Gross MR, Lamberg A (1996) An experimental study of the reproductive behaviour and success of farmed and wild Atlantic salmon (*Salmo salar*). J Appl Ecol 33:893−905

Fleming IA, Hindar K, Mjølnerød IB, Jonsson B, Balstad T, Lamberg A (2000) Lifetime success and interactions of farm salmon invading a native population. Proc R Soc B 267:1517−1523

Fleming IA, Agustsson T, Finstad B, Johnsson IJ, Bjornsson BT (2002) Effects of domestication on growth physiology and endocrinology of Atlantic salmon (*Salmo salar*). Can J Fish Aquat Sci 59:1323−1330

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

Garcia de Leaniz C, Fleming IA, Einum S, Verspoor E and others (2007) A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. Biol Rev Camb Philos Soc 82:173−211

Gjedrem T (2000) Genetic improvement of cold-water fish species. Aquacult Res 31:25−33

Gjedrem T, Gjoen HM, Gjerde B (1991) Genetic origin of Norwegian farmed Atlantic salmon. Aquaculture 98: 41−50

Glover KA (2008) Genetic characterisation of farmed rainbow trout in Norway: intra- and inter-strain variation

reveals potential for identification of escapees. BMC Genet 9:87 doi:10.1186/1471-2156-9-87

Glover KA (2010) Forensic identification of farmed escapees: a review of the Norwegian experience. Aquacult Environ Interact 1:1–10

Glover KA, Ottera H, Olsen RE, Slinde E, Taranger GL, Skaala Ø (2009a) A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. Aquaculture 286:203–210

Glover KA, Hansen MM, Skaala Ø (2009b) Identifying the source of farmed escaped Atlantic salmon (*Salmo salar*): Bayesian clustering analysis increases accuracy of assignment. Aquaculture 290:37–46

Glover KA, Hansen MM, Lien S, Als TD, Høyheim B, Skaala Ø (2010a) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genet 11:2 doi:10.1186/1471-2156-11-2

Glover KA, Dahle G, Westgaard JI, Johansen T, Knutsen H, Jørstad KE (2010b) Genetic diversity within and among strains of farmed Atlantic cod (*Gadus morhua*): a proof-of-concept study for the identification of escapees. Anim Genet 41:515–522

Glover KA, Dahle G, Jørstad E (2011) Genetic identification of farmed and wild Atlantic cod, *Gadus morhua*, in costal Norway. ICES J Mar Sci 68:901–910

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

Johnsson JI, Hojesjo J, Fleming IA (2001) Behavioural and heart rate responses to predation risk in wild and domesticated Atlantic salmon. Can J Fish Aquat Sci 58:788–794

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011) Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol Ecol Resour 11(Suppl 1):247–253

Koskinen MT, Sundell P, Piironen J, Primmer CR (2002) Genetic assessment of spatiotemporal evolutionary relationships and stocking effects in grayling (*Thymallus thymallus*, Salmonidae). Ecol Lett 5:193–205

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. Trends Ecol Evol 20:136–142

McGinnity P, Stone C, Taggart JB, Cooke D and others (1997) Genetic impact of escaped farmed Atlantic salmon (*Salmo salar* L.) on native populations: use of DNA profiling to assess freshwater performance of wild, farmed, and hybrid progeny in a natural river environment. ICES J Mar Sci 54:998–1008

McGinnity P, Prodohl P, Ferguson K, Hynes R and others (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. Proc R Soc Lond B 270:2443–2450

Moe H, Dempster T, Sunde LM, Winther U, Fredheim A (2007) Technological solutions and operational measures to prevent escapes of Atlantic cod (*Gadus morhua*) from sea cages. Aquacult Res 38:91–99

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Palstra FP, O'Connell MF, Ruzzante DE (2009) Age structure, changing demography and effective population size in Atlantic salmon (*Salmo salar*). Genetics 182:1233–1249

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

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.R-project.org

Roberge C, Einum S, Guderley H, Bernatchez L (2006) Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. Mol Ecol 15:9–20

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. Mol Ecol 17:314–324

Saegrov H, Hindar K, Kalas S, Lura H (1997) Escaped farmed Atlantic salmon replace the original salmon stock in the River Vosso, western Norway. ICES J Mar Sci 54:1166–1172

Sanchez JA, Clabby C, Ramos D, Blanco G, Flavin F, Vázques E, Powell R (1996) Protein and microsatellite single locus variability in *Salmo salar* L (Atlantic salmon). Heredity 77:423–432

Skaala Ø, Hoyheim B, Glover KA, Dahle G (2004) Microsatellite analysis in domesticated and wild Atlantic salmon (*Salmo salar* L.): allelic diversity and identification of individuals. Aquaculture 240:131–143

Skaala Ø, Taggart JB, Gunnes K (2005) Genetic differences between five major domesticated strains of Atlantic salmon and wild salmon. J Fish Biol 67(Suppl A):118–128

Skaala Ø, Wennevik V, Glover KA (2006) Evidence of temporal genetic change in wild atlantic salmon, *Salmo salar* L., populations affected by farm escapees. ICES J Mar Sci 63:1224–1233

Taylor EB (1991) Behavioural interaction and habitat use in juvenile chinook, *Oncorhynchus tshawytscha*, and coho, *Oncorhyncus kisutch*, salmon. Anim Behav 42:729–744

Verspoor E, Beardmore JA, Consuegra S, Garcia de Leaniz C and others (2005) Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. J Fish Biol 67(Suppl A):3–54

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

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191