The Scottish Government
Riaghaltas na h-Alba

Marine Scotland Science

Marine Scotland Science Report 16/13

An Overview of Population Genetic Structuring of Atlantic Salmon in the Kyle of Sutherland Fisheries Trust

M W Coulson, L M I Webster, I McMyn, A Armstrong,
E Cauwelier, L Stradmeyer, J Gilbey, C Sinclair & E Verspoor

Marine Scotland Science Report 16/13

An Overview of Population Genetic Structuring of Atlantic Salmon in the Kyle of Sutherland Fisheries Trust

M W Coulson, L M I Webster, I McMyn, A Armstrong, E Cauwelier, L Stradmeyer, J Gilbey, C Sinclair & E Verspoor

Prepared as part of the Focusing Atlantic Salmon Management on Populations (FASMOP) Project

Delivered in partnership with the Rivers and Fisheries Trusts of Scotland (RAFTS)

Published by Marine Scotland – Science

Marine Scotland is the directorate of the Scottish Government responsible for the integrated management of Scotland's seas. Marine Scotland Science (formerly Fisheries Research Services) provides expert scientific and technical advice on marine and fisheries issues. *Marine Scotland Science Report* is a series of reports that publish results of research and monitoring carried out by Marine Scotland Science. These reports are not subject to formal external peer-review.

Marine Scotland Science
Freshwater Laboratory
Faskally
Pitlochry
PH16 5LB
Copies of this report are available from the Marine Scotland website at
www.scotland.gov.uk/marinescotland

An Overview of Population Genetic Structuring of Atlantic Salmon in the Kyle of Sutherland Fisheries Trust

M W Coulson[1], L M I Webster, I McMyn[3], A Armstrong[1,2], E Cauwelier[2], L Stradmeyer[2], J Gilbey[2], C Sinclair[1] & E Verspoor[2]

[1]RAFTS, Capital Business Centre, 24 Canning Street, Edinburgh EH3 8EG
[2]Marine Scotland Science, Freshwater Laboratory Faskally, Pitlochry, PH16 5LB
[3]Kyle of Sutherland Fisheries Trust, Dornoch Road, Bonar Bridge, Sutherland, IV2 3EB

Executive Summary

This report describes how genetic information from juvenile Atlantic salmon sampled from 18 sites within the Kyle of Sutherland Fisheries Trust area (Figure 1) has been analysed in order to help inform developing fisheries management activities. The key objective for the Trust was to define the genetic structure of the locations under investigation, in order to determine whether salmon in the area represent distinct breeding populations.

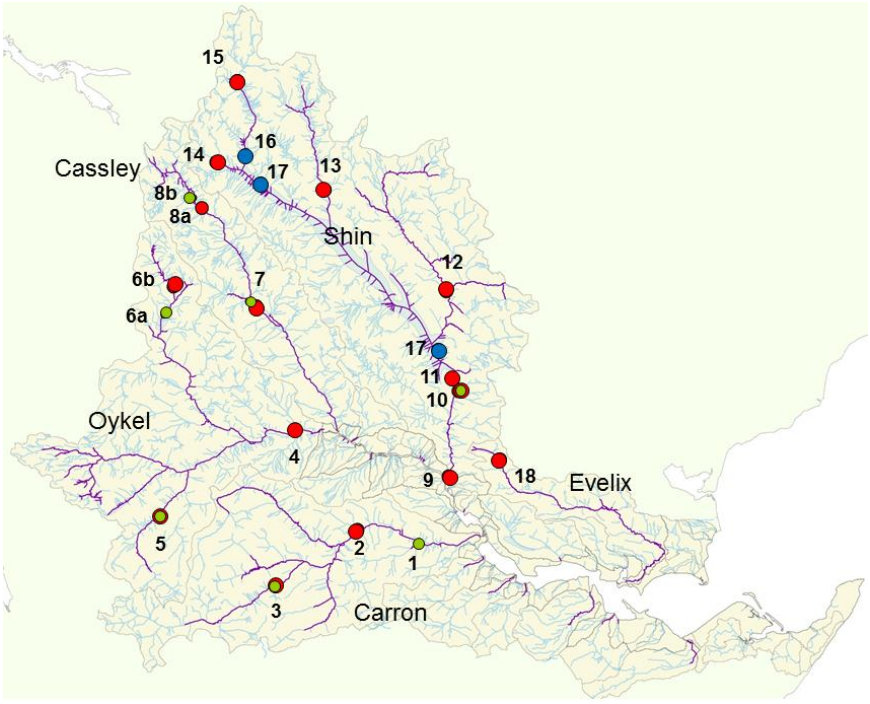

Figure 1 A map of the Kyle of Sutherland Fisheries Trust area, with sample sites that are involved in this report. All sites were screened for microsatellite genetic markers while those in green were additionally screened for SNPs. Sites in blue represent suspected farm escapees collected in the field. Numbers refer to site names in Table 1 (see main text).

Summary of findings

Analysis of the Kyle of Sutherland Fishery Trust samples showed a clear ability for both genetic marker types (microsatellites and SNPs) to resolve significant population structure. Both markers distinguished the Evelix and Shin systems from the Oykel-Cassley-Carron complex. However, SNPs greatly increased levels of genetic differentiation observed and were able to enhance the differentiation of the Oykel, Cassley, and Carron tributaries from one another compared to microsatellites. Within the Oykel-Cassley-Carron complex, the Upper Carron sites was the most differentiated and was also severely impacted by the presence of full-sibling families identified from a small number of breeders.

While the primary focus of the analysis was on the resolution of different stock components, an assessment of the ability to assign to individual sites was carried out, based upon the genetic differences observed. Assignment to site with microsatellites was 45%, reflecting the underlying genetic structure but the genetic differences were not large enough to assign with higher accuracy. However, when assignments were conducted to the three group level comprised of: 1) Shin, 2) Evelix and 3) Oykel-Cassley-Carron, assignment accuracy was on average 89%. With the use of SNPs, assignment to site generally improved but for several sites still remained around 60% or less. However given the ability of SNPs to further distinguish between the Oykel, Cassley and Carron, assignment to these tributaries may be possible with the use of SNPs.

Implications for management

This analysis clearly indicates that both microsatellites and SNPs are able to distinguish between the major systems within the Kyle of Sutherland Fisheries Trust area. However, SNPs offer an improved level of differentiation over microsatellites and further resolve differences that were more limited with the earlier markers (e.g. Oykel-Cassley-Carron complex). Clearly, future genetic work will likely therefore employ current and/or future panels of SNPs. However, microsatellites still have quite high power for particular applications that may be of use for the Trust (e.g. assessing estimating numbers of breeders, pedigree analysis). SNPs offer the ability to generate a clearer picture of population genetic structuring within systems and allow for levels of genetic differences

between sites to be demonstrated where previously they could not have been. Further insights into this structure could be achieved by the addition of new sites for SNP analysis to gain a more robust picture of the different stock components within the Kyle of Sutherland Fisheries Trust area. Additionally, assignment levels to main rivers/tributaries along with a developing Scotland wide river baseline suggests the possibility for accurate genetic assignments back to river of origin for marine caught samples and is likely to have applications in mixed-stock fisheries.

Introduction

Atlantic salmon (*Salmo salar* L.) is one of the world's most widely recognized and prized fish species. However, declines in numbers across much of the species' range have been cause for concern and the focus of intense management schemes and research efforts. Central to these efforts is the recognition of how the species is structured spatially across both broad and fine scales. Given the large native range encompassed by the species and their well-known ability to home to natal rivers and tributaries, it is expected and has been shown that Atlantic salmon demonstrate a considerable degree of population structuring, representing discrete breeding units that are reproductively separated to varying degrees (Webb et al. 2007; King et al. 2007). This reproductive separation underlies the basis for locally adapted traits to establish across a widely variable environment. Indeed, ample evidence exists in favour of local adaptation in many salmonid species (Taylor 1991; Garcia de Leaniz et al. 2007; Fraser et al. 2011).

A principal tool used to resolve population structuring is the use of genetic markers. Such approaches have been used for decades, with most of the contributions towards salmonids occurring within the past 20 years (Verspoor, 2007). Initially, this work focused on range-wide patterns of differentiation (King et al. 2001; Verspoor et al. 2005) or documenting the expected structuring existing between different river systems (O'Reilly et al. 1996; Fontaine et al. 1997; McConnell et al. 1997; Spidle et al. 2003; Castric and Bernatchez 2004; Dillane et al. 2007). More recently, genetic analyses of Atlantic salmon have indicated that rivers may be structured on fine scales, into multiple distinct breeding populations. Such studies have used both neutral genetic markers (Garant et al. 2000; Vaha et al. 2007; Dionne et al. 2008; Dillane et al. 2008), as

well as markers for which there is an underlying basis for natural selection (Landry and Bernatchez 2001).

Ecological studies have shown that different tributaries within a river may exhibit differences in traits, such as run-timing (Stewart et al. 2002, 2006), variation in age at smolting (Englund et al. 1999) and sea-age at maturity (Niemela 2006), among others. When such differences are shown to have a genetic basis (e.g. Stewart et al. 2002, 2006), then salmon breeding in separate locations, for instance, above and below waterfalls or other natural features may often be heritably different in ways that affect their behaviour, survival and reproductive success. This may be true of neighbouring tributary populations and, as such, intermixing of these populations runs a risk of unknown magnitude, and may not be desirable. Indeed, at its worst, mixing in vulnerable populations could have long term negative effects on population viability by reducing survival (McGinnity et al. 2003).

Given the recognition of the 'population' as a focal unit for management, it follows that knowledge of the genetic structuring among sites is required for certain management and conservation schemes. The potential for multiple, distinct populations to become established within a single river means that detailed knowledge needs to be gained regarding the scale at which such population structuring occurs. As a first step in this process, a baseline for systems needs to be constructed and built upon as more information becomes available.

In 2009, a partnership between the Rivers and Fisheries Trusts of Scotland (RAFTS), Marine Scotland Science (MSS), and the participating individual Fisheries Trusts and Boards was established. This project, entitled **F**ocusing **A**tlantic **S**almon **M**anagement **O**n **P**opulations (FASMOP), had, as its central aim, to undertake a program of genetic sampling of Atlantic salmon stocks in river systems across Scotland. The purpose of this sampling scheme was to define the genetic structure among locations, in order to determine whether salmon within and among the various systems in a given area represent distinct breeding populations. This work, alongside the EU SALSEA-MERGE and other MSS projects, is creating a genetic map of salmon populations across Scotland, to help inform management and conservation efforts.

The suite of genetic marker initially used (microsatellites) in the current survey is assumed to be “neutral” (meaning they are not known to be linked to heritable characteristics that may differ among locations, such as run timing, growth rate, etc.). They will, therefore, largely reflect the shared ancestry of salmon among sites rather than make a direct assessment of the heritable trait characteristics that differ among them. Differences at such markers may *imply* that groups are sufficiently distinct for heritable trait differences to have evolved, however, where no difference is observed, we cannot rule out the possibility that these sites differ in heritable traits. Microsatellites have proved useful in resolving population structuring both among and within regions and rivers (King et al. 2001; Fontaine et al. 1997; McConnell et al. 1997; Spidle et al. 2003; Castric and Bernatchez 2004; Dillane et al. 2007; Gilbey et al. in preparation). In certain cases, the level of structuring identified by microsatellites has even allowed for genetic assignment to particular tributaries within a system [see for example the FASMOP reports (www.rafts.org.uk) for the Ness & Beauly Trust (in preparation) as well as the Forth Fisheries Trust (Coulson et al. 2012). Therefore, in some cases, microsatellites are clearly capable of identifying different stock components and will be useful in subsequent applications, in terms of genetic assignment of rod-caught fish. However, this is not universally the case. Particularly within the large, east-coast rivers involved in other FASMOP Scottish surveys, microsatellites showed little to no evidence of genetic structuring (Coulson et al. in preparation). It is unknown, however, if the lack of structure observed with the microsatellites is a true reflection of the genetic population structure in the systems or rather a function of the resolution of the microsatellites markers used. To address this question, a new approach was utilised and is reported here.

Recently, another class of genetic markers, known as Single Nucleotide Polymorphisms (SNPs), have become more widely used and available for a number of species, including Atlantic salmon (e.g. Lien et al. 2011). The use of SNPs, either as an alternative to or in conjunction with microsatellites, has been shown to be capable of enhancing the resolution between different stock components with respect to fisheries management for various species, particularly salmonids (e.g. Smith et al. 2005; Narum et al. 2008, Glover et al. 2010, Beacham et al. 2010, Freamo et al. 2011, Hess et al. 2011, Seeb et al. 2011a,b). Therefore, the aim of the current report was to define population structuring within the Dee using a panel of microsatellite markers and compare,

for a subset of sites, the ability of SNPs to further define or enhance any observed genetic structuring. Observed levels of genetic differentiation were subsequently evaluated for their ability to be applied to fisheries management interests, notably the use of genetic assignment of individuals to population of origin.

Summary of Methods

Juvenile salmon from various locations within the Kyle of Sutherland Fisheries Trust area were sampled for genetic material by the Trust, in order to inform fisheries management, following methods outlined by Verspoor and Laughton (2008). Figure 1 shows the locations of the 18 sites that have been included in the genetic analysis for the Trust. Samples generally consisted of fry and/or parr (n = 31-79, depending on site) and, for each individual, data from 17 genetic markers (microsatellites) were collected. Two samples (sites 16 and 17; Figure 1) consisted of smolts collected by the Trust as they were suspected to be aquaculture escapees. For the Loch Shin smolts (sample 17), scale readings were conducted and generally morphological appearance was assessed by staff at Marine Scotland Science and determined to be consistent with aquaculture origin. A set of scale samples of adults collected from the River Shin (Lairg) in 1987 was additionally included. This sample predates the occurrence of fish farms on Loch Shin and was screened to allow for a temporal comparison of the genetic make-up of Loch Shin fish pre- and post-dating the presence of aquaculture. Finally, from seven of the sites, data from 5568 SNPs were also collected to compare the ability of the different classes of genetic markers to resolve population structure and genetic assignment.

The results from the microsatellite marker SsaF43 allowed us to identify any trout or trout/salmon hybrids that may have been present among samples. These individuals were then removed prior to analysis.

It is possible that samples at a site are more reflective of families rather than populations, given the life-history stage(s) targeted by sampling and the potentially fine-scale geographic coverage (Hansen et al. 1997). This occurrence can alter the genetic signature of the sample and obscure population level differences. Therefore, prior to population level analyses, each site was screened

for the presence of full-siblings, representing family groups and when identified, all but one individual of a full-sib family were removed. Additionally, this analysis can estimate how many breeders contributed to producing each sample, which may include contributions from sexually mature parr. Initial sample sizes, as well as sample sizes after full-siblings were removed, are presented in Table 1.

When samples sites included two life-history stages (i.e. fry and parr), samples taken from a site in different years, or sites sampled in close geographical proximity (< 5km), each of these sub-samples were initially tested for differences using the program CHIFISH (Ryman 2006). Where no significant differences were found, data from these sub-samples were combined; otherwise they were left separate for all further analyses. This resulted in 21 samples for subsequent analyses.

In addition to the microsatellites, for a subset of sites (see Figure 1, Table 1), SNP data were collected. For these individuals, a 5,568 SNP chip assay (Lien et al. 2011) was screened in collaboration with the Centre for Integrative Genomics (CIGENE) in Norway, where the technology was developed. Several methods were used to select a subset of SNPs to carry out further population genetic analysis. These generally fell into two categories: (1) ranking of SNPs based on an overall measure of genetic differentiation (F_{ST}) or (2) the detection of 'outlier' SNPs (i.e. those SNPs that show greater differentiation than would be expected given their levels of variability) using the program ARLEQUIN (Excoffier et al. 2005). Such rankings allowed a sub-set of the 5,568 markers to be identified, which contained those that had the greatest power in resolving population structure among the samples screened. This allowed for a comparison of the two different markers (microsatellites and SNPs) for resolving population genetic structuring within this system.

Data were then analysed (microsatellites and/or SNPs) using standard population genetic methods to evaluate the genetic relationships and groupings among the sample sites, in order to obtain a general overview of population structure and address the objectives of the Trust.

A detailed methods and analysis section can be found in Appendix 1.

Results

Family effects

A total of 1,056 salmon from the Kyle of Sutherland Fisheries Trust area were involved in the genetic analysis. No trout or salmon/trout hybrids were identified at any location. All sites were examined for family effects. The level of family effects differed dramatically among samples, with the largest family group present in the individual samples ranging from 0 to 20 full-siblings and sample sizes subsequently being reduced by 0-70%. The most affected sites were Loch Merkland, both samples from the River Fiag (2009 and 2011), and the Evelix, with the largest full-sib family found within the River Fiag (2009) sample. The remaining sites had moderate to very few full-siblings (Table 1). Family effects were controlled for at each site before all further analyses.

Population structuring – Microsatellite baseline

Several pairs of sites were sampled either in different years and/or in close geographical proximity. These included the two sites in the Upper Oykel (Loch Ailsh 2008 and 2009) the two samples on the Upper Cassley (2008 vs. 2009) and the two samples on the River Fiag in the Shin system (2009 vs. 2011). All three of the comparisons between these sub-samples were significantly different based on the CHIFISH analysis and were therefore kept separate. A fourth comparison, the Shin samples near Lairg (1987 vs. 2008) was not conducted as the interest was to see where the placement of the 1987 sample was in relation to all other sites on the Shin system. This left 21 samples for analysis.

The genetic differences among sites showed a moderate to large level of genetic differentiation; 94% (197 out of 210) of the pairwise comparisons between sites were significantly different (Appendix 2). A visual representation of these relationships among samples can be found in Figure 2, which uses multi-dimensional scaling to represent pairwise estimates of genetic differentiation among sites (Appendix 2). Points that are closer together on the plot have a more similar genetic makeup, while points further apart are more genetically discrete.

The largest differences were seen between the Shin system locations and the rest of the Oykel-Cassley-Carron systems as the Shin samples are all on the

right side of the plot (Figure 2) while the remaining sites are on the left of the plot. The Oykel-Cassley-Carron sites all tend to group relatively close together, suggesting a closer genetic relationship to one another. Among these sites, there does not appear to be an obvious separation based upon the specific tributary (i.e. groupings of Oykel vs. Cassley vs. Carron). However, the upper Cassley (2009) appear to be the most distinct of all Oykel-Cassley-Carron sites screened. This site was also severely affected by the presence of full-siblings and relatively few numbers of breeders. Such family effects likely influence the relative uniqueness of this sample. The last system, the Evelix is quite distinct from all other sites. The two samples of farm escapees are also quite distinct and are located quite far apart from the remaining Shin samples. Among the Shin samples, Loch Merkland is the most distinct, followed by the 2011 sample from the River Fiag. Loch Merkland, in particular, was also heavily influenced by full-siblings with relatively small numbers of breeders. The same was true for both Fiag sites, and more prevalent in the 2009 sample, however this sample was within the central cluster of Shin sites. Finally, the 1987 sample from the River Shin was located within the central Shin cluster of sites.

Table 1

Details of samples used for this analysis, including original sample size, and details of each site following COLONY analysis to re-construct family relationships.

Site	Site ID	Original sample size	Sample size (sibs removed)	Number of breeders contributing to sample	Largest single family	Sample Size (SNPs)	Year sampled
Lower Carron	1	31	30	44	2	32	2009
Middle Carron	2	50	49	66	2	--	2008
Upper Carron	3	50	48	73	2	32	2008
Lower Oykel	4	50	49	70	2	--	2008
Oykel - Corriemulzie	5	50	43	58	5	32	2008
Oykel - inflow Loch Ailsh	6a	32	30	42	2	32	2009
Oykel - above Loch Ailsh	6b	50	43	54	5	--	2008
Cassley – middle mainstem	7	79	78	83	2	30	2008
Cassley – upper mainstem	8a	50	43	58	3	--	2008
Cassley – d/s Fionn Loch Beag	8b	32	16	15	13	32	2009
River Shin – Inveran	9	50	43	50	4	--	2008
River Shin – Lairg	10	50	47	62	2	31	2008
Loch Shin – Lairg	11	94	83	87	3	--	1987
River Tirry	12	50	36	40	5	--	2009
River Fiag	13	44	20	26	20	--	2009
		49	29	30	7	--	2011
Loch Shin – upper end	14	36	36	41	1	--	2009
Loch Merkland	15	50	15	17	17	--	2009
Corriekinloch-Merkland smolts	16	33	32	43	2	--	2011
Loch Shin smolts*	17	76	59	61	4	--	2010
Evelix	18	50	23	32	11	--	2008

*All but three smolts were sampled at the uppermost location #17 (north end of the Loch).

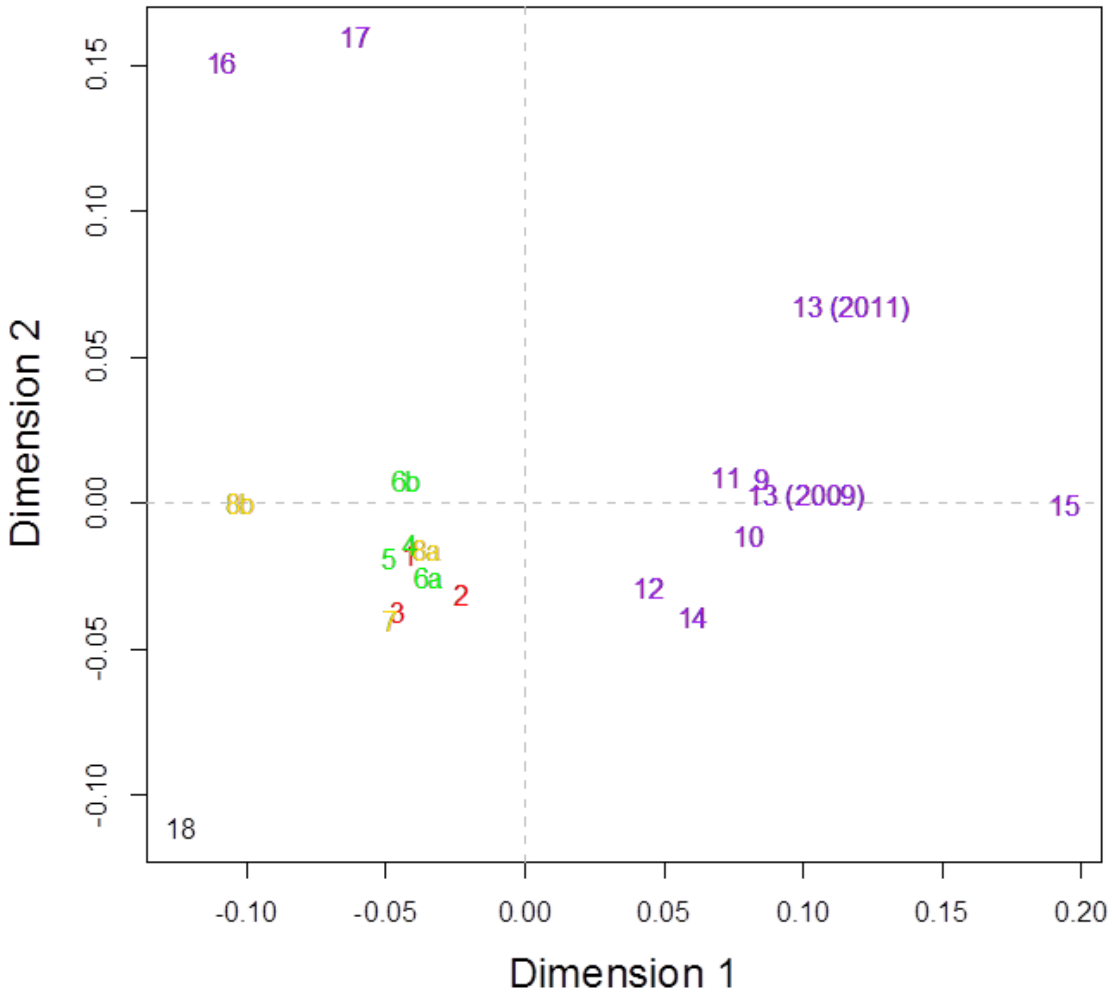

Figure 2 Multidimensional scaling (MDS) plot of genetic relationships among all sites, based on pairwise estimates of genetic differentiation (Jost's *D*; see the appendix for details) for the microsatellite data. Points that are closer together on the plot have a more similar genetic makeup, while points further apart are more genetically discrete. Sample codes are as presented in Table 1 and sites are colour-coded by river/tributary (red=Carron, green=Oykel, gold=Cassley, purple=Shin, black=Evelix). Note: samples 16 and 17 represent the putative aquaculture escape smolts sampled within the Loch Shin system.

Population structuring – SNPs

A total of seven sites included in the full microsatellite baseline were subsequently screened with the 5.5k SNP chip. These sites are indicated in Figure 1 and Table 1. To allow for a direct comparison between marker types, the microsatellite statistics were re-calculated for only the seven sites involved in the SNP analysis. The genetic differences among these seven sites, for the microsatellites, show (as expected) small to moderate levels of differentiation with 90% (19 out of 21) of the pairwise comparisons being significantly different (Table 2). The Upper Cassley and Shin sites were the most different. The two Oykel sites were not significantly different from one another nor were the Lower Carron and Middle Cassley.

For the total panel of SNP markers among the seven sites screened for the Kyle of Sutherland area, 128 'outlier' SNPs were identified and analysis of just these 128 SNPs resulted in 100% (21 out of 21) of the pairwise comparisons among sites being significantly different. The same proportion of comparisons were significantly different using the top 100 F_{ST}-ranked SNPs. Again, the Upper Cassley and Shin sites were the most differentiated. All sites generally showed greater differentiation from one another with SNPs compared to the microsatellites (Figure 3), with an average ~10-fold increase in the levels of genetic differentiation using SNPs (Table 2). A summary of the comparison of the levels of differentiation observed between the microsatellites and SNPs is presented in Table 2. Figure 3A shows the comparison between the microsatellites and the 128 'outlier' SNPs in resolving pairwise genetic differentiation. As can be seen, the Upper Cassley and Shin sites are the most differentiated for both marker types, however the extent of the differentiation is greater for these two sites with the panel of SNPs. Given the greater differentiation of the Upper Cassley and Shin sites to all other sites with the SNPs (Figure 3A), these samples are likely contributing disproportionately to the identification of the top-ranked SNPs, reflected by the fact that the remaining sites (even for SNPs) are clustered close together and do not appear to be grouped geographically. Therefore, the dataset was re-analyzed for 'outlier' SNPs using a method that also identifies markers based upon distinguishing not only specific sites, but additionally broader geographic regions; in this case r river/tributary level. This was made possible by the fact that two sites were screened in each of the Carron, Oykel & Cassley and a single site on the Shin.

This analysis identified 64 ‘outlier’ SNPs (henceforth referred to as ‘group outlier SNPs’) and the pairwise differences are shown in Figure 3B along with the microsatellites for the seven sites. Among these 64 ‘outlier’ SNPs, 41 (64%) were in common with the initial ‘outliers’ detected using only site level. Figure 3B shows the comparison of the ‘group outlier SNPs’ against the microsatellites. Not surprisingly, the SNPs show greater differentiation, but now also appear to reflect a more regional signal. While the Upper Cassley and Shin are still the most differentiated, the two Carron sites (1 and 3) are closest to each other as are the two Oykel sites (5 and 6a). The same is somewhat true for the Cassley, however there is a greater separation between the upper and middle Cassley given the unique nature of the former site.

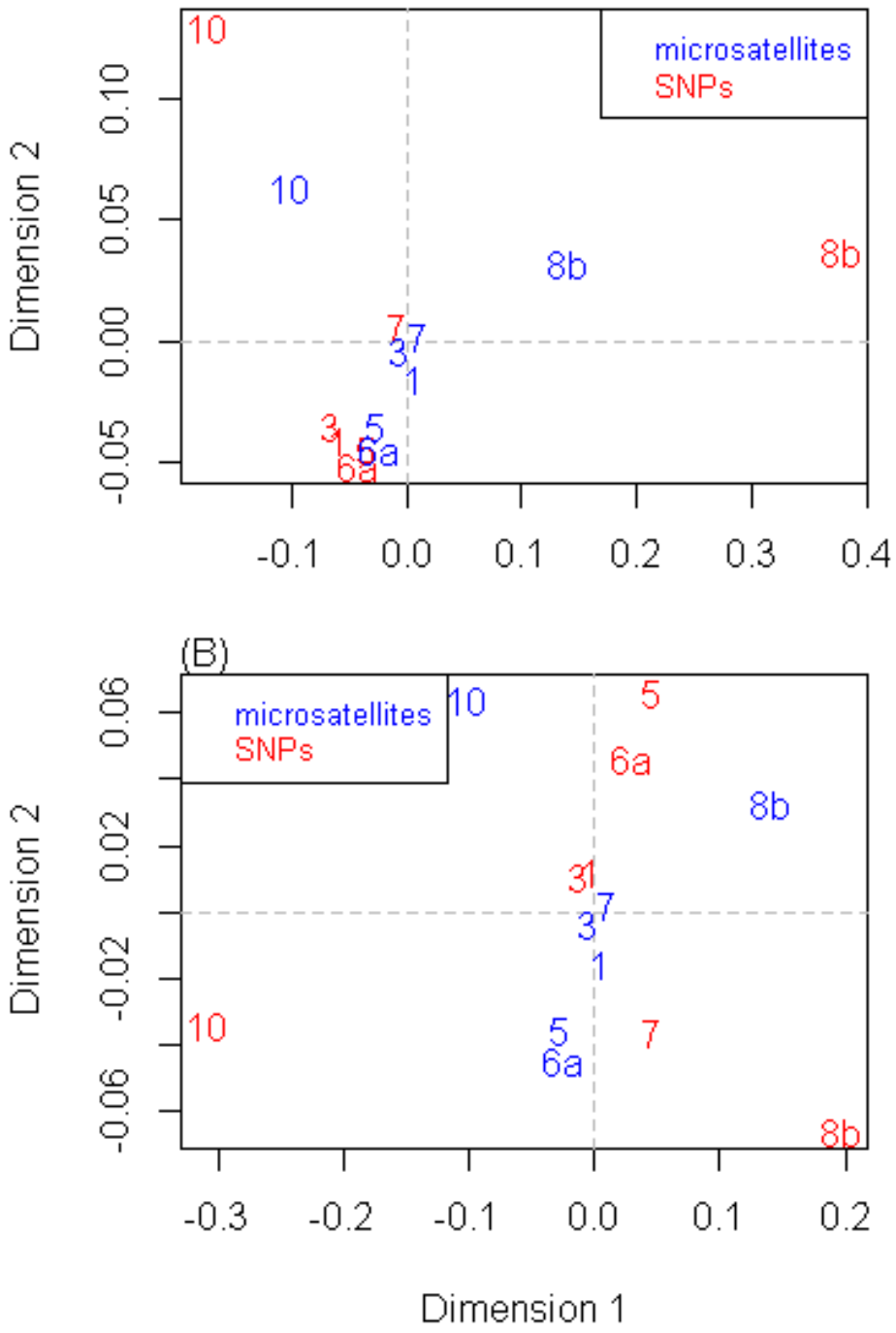

Figure 3 Multidimensional scaling (MDS) plot of genetic relationships among the sites screened for both microsatellites and SNPs based on pairwise estimates of genetic differentiation. Points closer together have a more similar genetic makeup, while points further apart are more genetically different. (A) The set of 128 high-ranked 'outlier' SNPs across all locations compared to the microsatellites. (B) A set of 64 'outlier' SNPs chosen to maximise the differences between the different rivers/tributaries compared to the microsatellites.

Table 2. Comparison of the levels of differentiation for microsatellites versus SNPs

	Microsatellites (F_{ST})	Top 100 F_{ST} ranked SNPs	128 'outlier' SNPs (overall)	64 'outlier' SNPs (among groups)
Overall differentiation	0.017	0.126	0.130	0.111
Range of levels of differentiation	0.005-0.054	0.011-0.354	0.016-0.372	0-0.347
% of significant pairwise comparisons	90%	100%	100%	90%

Genetic assignment of individuals – microsatellites

It may be possible to assign fish back to their population of origin using genetics, once population structure has been identified. While the primary aim of the present report was to identify different stock components within the Kyle of Sutherland Fisheries Trust area, an analysis of this assignment ability was also carried out. The assignment analysis shows how useful this baseline genetic information is in identifying which of the sampled sites a fish of unknown origin is from (Figure 4). Each individual fish is taken in turn and it is assessed from which of the sampling locations that individual is most likely to have originated. It should be noted that for the purposes of genetic assignment discussed below, the two samples of farm escapees were not considered, as they were not consistent with origins from Loch Shin in which they were sampled. However, these fish did assign to their own group of 'farmed' fish with 94-100% accuracy.

Among the Kyle of Sutherland locations, assignment of fish back to their specific site of collection was, on average, correct 45% of the time (Figure 4A). This average is significantly greater than would be expected if assignments were purely random (15 sites, random = ~7%), and reflects the degree of population genetic structure underlying the data. However, given the closer genetic

relationships among some locations (i.e. the Oykel-Cassley-Carron sites and most sites within Loch Shin), it is not surprising that assignment success is not higher. A second level of assignment was conducted whereby fish were assigned to a reporting group, rather than a site, as larger scale differences may be more pronounced. The reporting groups used were the following: 1) Carron, 2) Oykel, 3) Cassley, 4) Shin and 5) Evelix, thereby corresponding to river/tributaries. Average assignment to these five reporting groups was 69% (Figure 4B). Assignment was highest to the Shin (89.4%), followed closely by the Evelix (81.0%) Assignment was much lower to the Cassley (67.8%), It may be possible to even improve this accuracy by implementing a cut-off rule for the probability that an individual gets assigned or by allowing individuals to not be assigned to any of the sites in the baseline (i.e. came from an unsampled population). For the Ness & Beauly Fisheries Trust sites, a cut-off does marginally improve assignments. For example, if we assign only fish that have a minimum of 80% assignment probability, overall correct assignment, to reporting group level, increases assignment success by 1.2% (Enrick) to 9.8% (Moriston). However, applying such a cut-off may come at a potential cost, as not all fish in the baseline will be assigned. For the sites examined here, an 80% cut-off still resulted in 90% of fish being assigned.Oykel (58.8%) and Carron (45.9%). Given the relatively high assignment success for the Shin and Evelix, coupled with the lower population structure observed among the Oykel-Cassley-Carron sites, a final assignment analysis was conducted on three major groupings (Oykel-Cassley-Carron, Shin, Evelix; Figure 4C). Average assignment to the three main systems was 89% and was now highest to the Oykel-Cassley-Carron (96.1%).

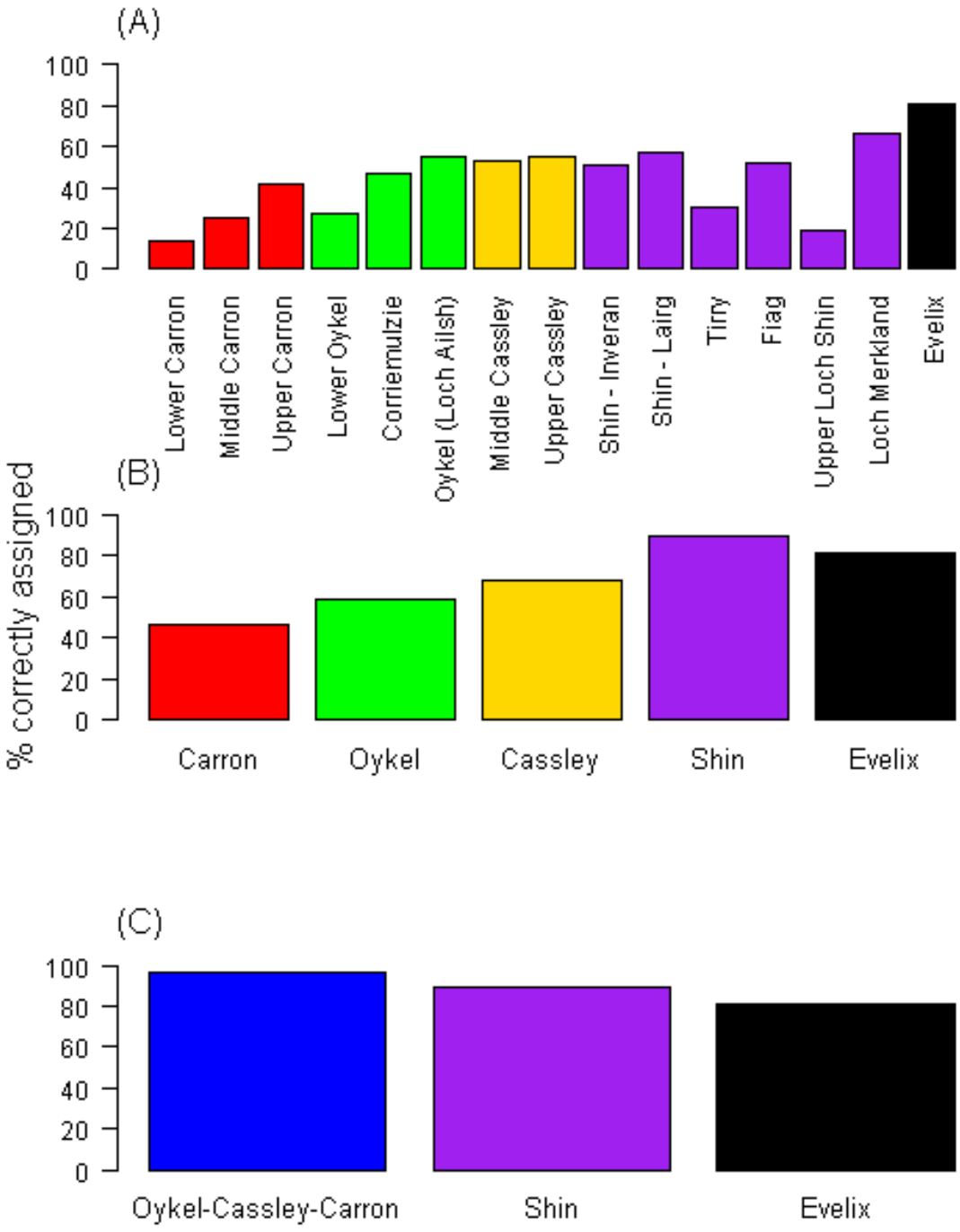

Figure 4 Percentage of fish sampled from each site that correctly assign back to that site, based on a suite of 17 microsatellite markers.

Genetic assignment - SNPs

Unlike with the microsatellites, where a limited number (17) of genetic markers had previously been identified for widespread salmon genotyping, SNPs provide a much larger panel (~5,500) from which to choose a subset of markers for a specific purpose (in this case, defining stock structure within the Kyle of Sutherland Fisheries Trust area). In situations with large numbers of markers available, the use of the same samples to both pick the subset that best distinguishes between sites, as well as to evaluate genetic assignment to those sites can lead to over-estimates of assignment accuracy (Anderson 2010; Waples 2010). To address this bias, each site was split in half, with one half used as a 'training set' and the other half as a 'holdout' set. The training set was used to identify the subset of markers that best distinguished among the sites; the holdout set was then used to assess the accuracy of the assignment. While this approach reduced the bias of assignment accuracy, it also halved both the number of samples used to pick the most informative markers (i.e. the 'training' set), as well as the number of samples from a given site to be assigned (i.e. the 'holdout' set (on average now 15 as opposed to the original 30).

To asses the degree to which the SNPs could be used to assign fish to site, a two-step approach was taken. The first was to assess the impact of increasing numbers of SNP on assignment success, followed by looking at site-specific assignment with the set of SNPs resulting in the highest levels of correct assignment. Six sets of top-ranked SNPs, based upon the eight sites included here, were assessed (50, 100, 200, 300, 500 and 1000 SNPs). It is expected that increasing numbers of markers should contribute toward increased levels of assignment. As can be seen in Figure 5A, levels of assignment increase from 50 SNPs to about 300 SNPs, with minor variation beyond this point. A cut-off of 80% (meaning that only fish that assigned to a site with an assignment score of 80% or greater were considered 'assigned') marginally improved assignments for the l300 SNPs or more but noticeably improved for the lowest number of SNPs (Figure 5A). Average overall assignment for these seven sites was ~80% with a cut-off. Subsequently, the 300 SNP set was chosen to evaluate the site-specific levels of assignment compared with the microsatellites (Figure 5B). Assignment to site was 60% or more, with the lowest assignment to the Corriemulzie followed by the lower Cassley. The highest assignment was to the Shin (100%) followed by the Upper Oykel (91%). The Upper Cassley assigned with 76%, which may be

surprisingly low given the degree of differentiation observed with both marker types to this site. This observation could however be the result of a reduced sample size due to significant family effects at this site, resulting in a poorer baseline for assignment purposes.

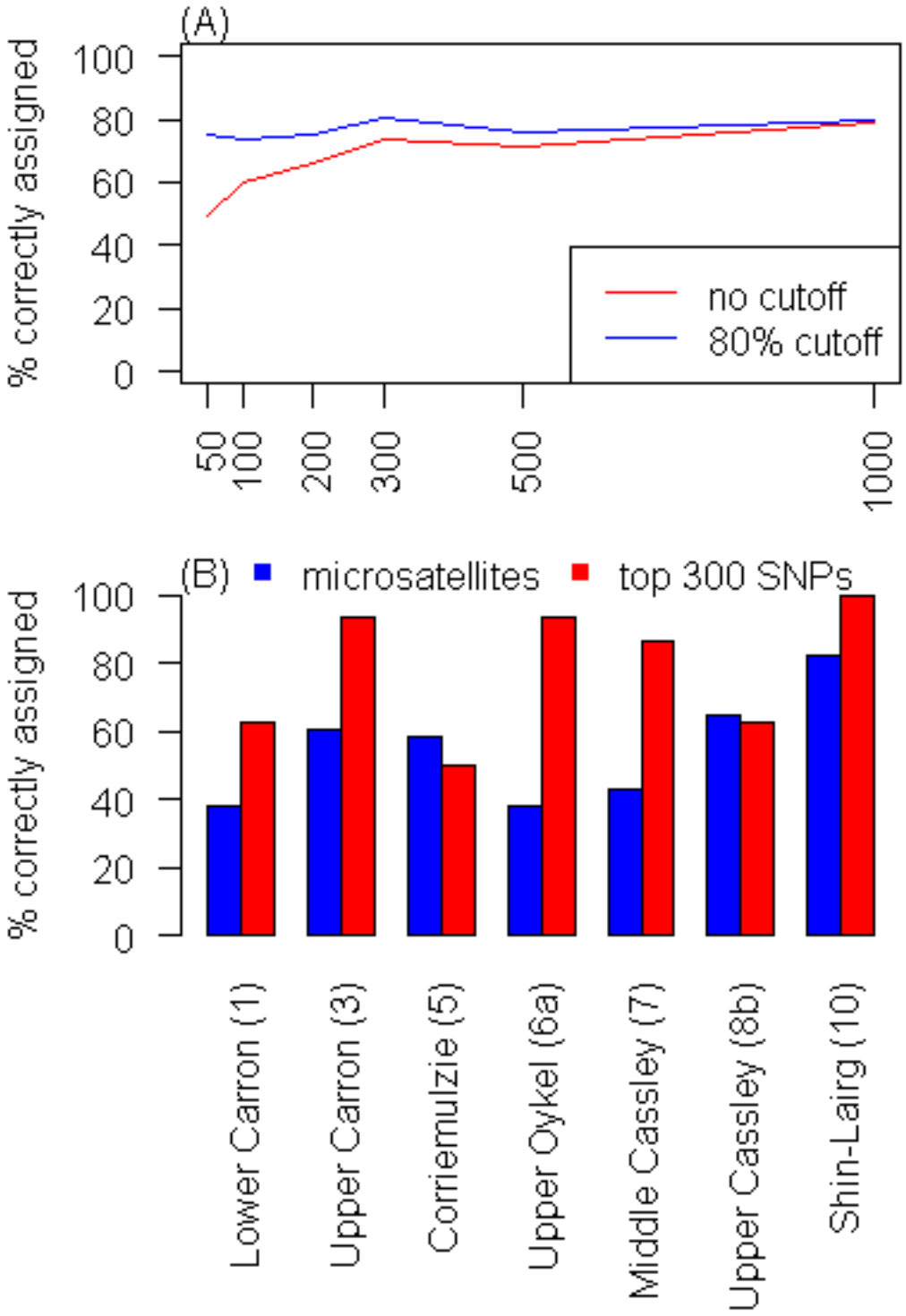

Figure 5 (A) Average overall assignment to the seven sites for varying number of top-ranked SNPs using both a no-cutoff and an 80% cutoff for assignment accuracy. (B) Correct assignment levels to individual sites for both the top 300 ranked SNPs and the reduced microsatellite-screened baseline.

A final assignment was done using the 64 ‘group outlier SNPs’ compared to the microsatellites for assigning to river/tributary level (Figure 6). As can be seen these 64 SNPs improve assignment to river/tributary compared to the microsatellites with three of the four systems assigning with nearly 80% or more. Assignment was lowest to the Oykel with higher assignment to both the Carron and Cassley and the Shin with 100% assignment.

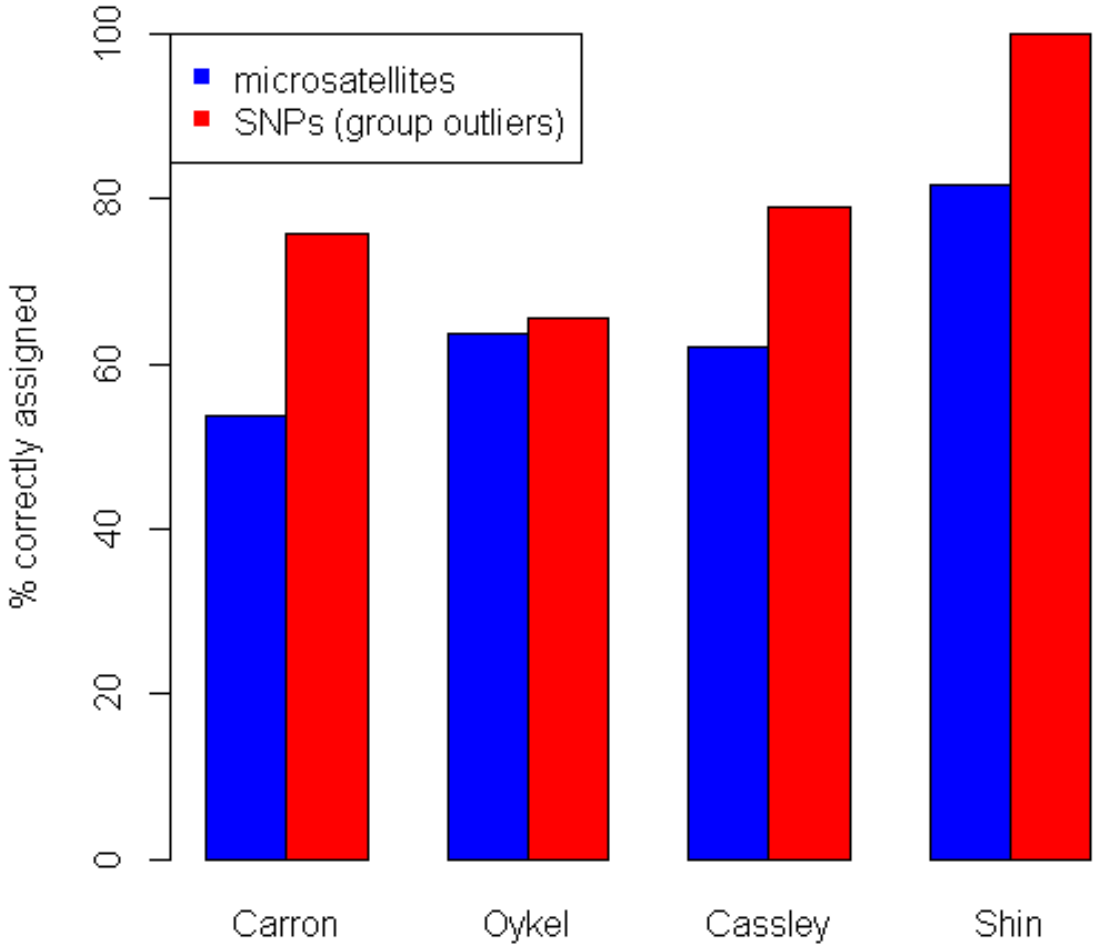

Figure 6 Tributary level assignment comparing the reduced microsatellite screened baseline against the 64 ‘outlier’ SNPs identified maximising for among reporting group (i.e. tributary) differences.

River-level assignment (SNPs)

The SNP data presented here are part of a larger SNP dataset encompassing an increasingly comprehensive number of rivers around Scotland. This is allowing for the ongoing identification of a set of SNPs that may be useful in assigning fish back to their river of origin. To date, this work involves 12 rivers, which are mostly on the east-coast, shown in Figure 7A. A preliminary analysis (Gilbey & Coulson, 2013) demonstrates an ability to assign fish back to river of origin with ~80% accuracy, on average, using a set of 200-300 SNPs (Figure 7B). However, there is considerable variation in the figure among rivers. For the Oykel-Cassley-Carron-Shin rivers combined, at present, this figure is ~90%, which is above the average river-level assignment. It should be noted that particular effort is being focused on expanding the geographical coverage of this baseline and this figure is therefore likely to change.

In order to test whether the set of SNPs identified from the 12 sample rivers could be applied more widely, the accuracy of applying this set to samples from the South Esk, a river not in the original sample, was assessed. The low levels of correct assignments (a maximum of 33% of South Esk samples were assigned back to the South Esk) suggested that, for this process be most effective, rivers of interest should be in the original baseline used to choose the subset of SNPs. However, the SNPs clearly demonstrate a significant improvement to river-level assignments than is possible with previously used markers (e.g. microsatellites). This level of analysis was made possible, in large part, by the individual within-river SNP screening. These results have applications for assigning fish caught at sea to river level and suggest that with further development they could produce a genetic tool for application in mixed-stock analysis.

While river-level assignment was not the focus of the Kyle of Sutherland Fisheries Trust SNP work necessarily, it does demonstrate the types of applications and analyses that may be achieved by combining individual river system analyses into a wider Scottish context. While some additional screening of samples from the Kyle of Sutherland rivers/tributaries may help to continue to improve assignments (e.g. for the Oykel), more importantly, at present, is the addition of more rivers to cover the baseline of genetic variation throughout Scotland. The expansion of this Scottish baseline is currently underway.

(A) (B)

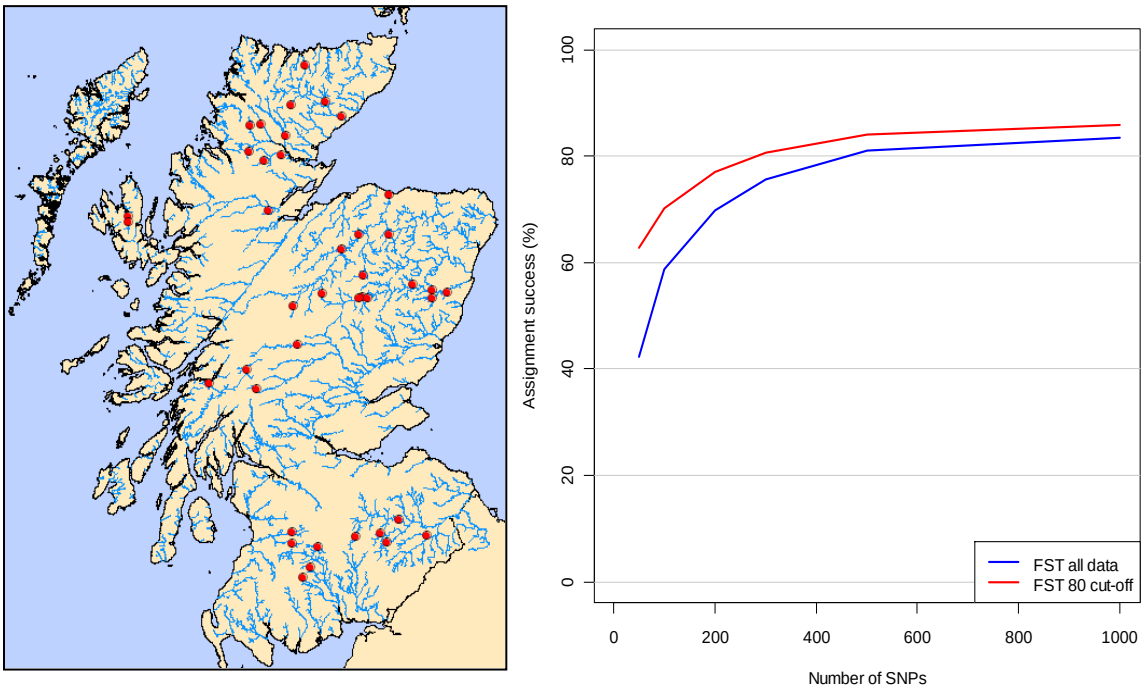

Figure 7 (A) Sites screened for the 5,500 SNP chip to date. (B) Assignment accuracy to river of origin for a preliminary baseline of Scottish rivers with increasing number of SNPs (Figures courtesy of Gilbey & Coulson, 2013).

Discussion

Population structure

The aim of the FASMOP project for the Trust was to identify distinct breeding populations of salmon within the Kyle of Sutherland Fisheries Trust area. The results to date suggest that there are distinct breeding populations. Such a conclusion was strengthened with the use of SNPs, as for the Oykel-Cassley-Carron systems in particular, the microsatellites revealed weaker levels of differences among these sites.

The microsatellites demonstrated the ability to distinguish two samples of farm escapees as distinct from all other sites screened within the Kyle of Sutherland Fisheries Trust. This latter point demonstrate the ability of genetic to identify farm escapees of foreign origin in cases where field identification may be difficult or need corroboration.

Both marker types (microsatellites and SNPs) showed a clear ability resolve significant population structure among the Kyle of Sutherland Fisheries Trust samples. Both markers distinguished the Evelix and Shin systems from the Oykel-Cassley-Carron complex. However, SNPs greatly increased levels of genetic differentiation observed and were able to enhance the differentiation of the Oykel, Cassley, and Carron tributaries from one another compared to microsatellites. The use of a set of 'group outlier SNPs' appeared to perform better in overall distinction of the seven sites screened than simply ranking the top performing SNPs at site level. This is likely due to the fact that the more unique nature of both the Shin and Upper Cassley site were biasing the top ranking SNPs. In other words most of the top ranking SNPs were likely distinguishing these two sites at the expense of poorer resolution among the remaining five sites. Therefore the 'group outlier SNPs' allows for each site to be treated separately but also uses some information about the larger scale groups (i.e. tributaries or rivers) to which these sites belong. The analysis then aims to find the set of markers maximising these 'group' differences. Such an improvement was observed on the MDS plots whereby sites within the same tributary generally clustered together with the 'group outlier SNPs', which was not the case for the top ranking SNPs.

Not all sites were screened for SNPs, and we are therefore limited to the microsatellite data for inferences about their population structure. While a lower degree of differentiation (as seen among several of these sites) may be the result of moderate levels of exchange of spawning adults among sites, caution should still be used in making such an interpretation. This lack of differences may be due, at least in part, to the current set of genetic markers. For instance, adaptive differences may be present (e.g. for run timing behaviour), which the current panel of SNPs may not include. Therefore, other markers associated with adaptive traits, may help to further address the degree to which these locations represent distinct breeding populations and may affect levels of assignment. Overall, this type of pattern supports the idea of a meta-population structuring rather than distinct phylogenetic lineages. This implies that spatially separated populations are connected by different degrees of interactions or exchange of individuals over time rather than being composed of multiple, distinct evolutionary lineages that have persisted for extended periods of time.

When there is clear evidence of distinct breeding populations, then a continued caution with respect to sourcing brood stock would be desirable in respect of stocking programmes. Even though some of these significant differences are weaker, the same caution should be exercised. As mentioned above, locations may still differ with respect to heritable adaptive traits and until such issues can be addressed, locally sourced brood stock should reduce the risk of disrupting any local adaptations that lead to increased survival.

Genetic assignment

The power to assign fish of unknown origin to their location of origin with high accuracy is possible where candidate locations show strong genetic differentiation. Such an approach is useful for discriminating the composition of mixed-stock fisheries, for instance, in assigning rod-caught adults to their particular stock component. It may also be possible to use genetic assignments to determine whether salmon returning to a river at different time points are destined for different parts of the catchment, if there is well defined structuring between these components and with genetic markers that may be associated with that particular trait. Genetic assignment allows one to calculate an assignment score for each individual fish as having originated from each of the sampled locations. Then the location with the highest assignment score is taken as the baseline site from where that individual is most likely to have originated. This is done for each individual and Figures 4, 5B, and 6 show the proportion of individuals from a given site/group that was assigned back to that site/group, based on their genetic profile using either microsatellites or SNPs. If each location exhibits large differences from everywhere else, one would expect the accuracy to be high. While SNPs improved the average level of assignment compared to microsatellites (average of 80% with a cut-off for SNPs vs. 55% for microsatellites), not all sites assigned equally well as several sites had assignment accuracy of 60% or less (Figure 5B). However, while a fish may not assign to its particular site, it may still assign to other sites within the same tributary, reflecting larger-scale genetic differences. Indeed, even with the microsatellites, assignment was very accurate in distinguishing Shin and Evelix fish from one another and from the combined Oykel-Cassley-Carron baseline. The assignment results with SNPs demonstrated increased accuracy of assignment to most sites, but additionally suggests a greater ability to further assign on a river/tributary level. As observed with the microsatellites, SNPs were

highly accurate in assigning to the Shin and Evelix. However, SNPs improved even further on the microsatellites by suggesting an ability to assign as well among the Oykel, Cassley and Carron.

As many assignments will try to assign individuals to sites represented in the baseline, if the 'true' site has not been sampled, fish from these missing sites will be forced to be incorrectly assigned. At present, these assignments represent our best estimates, since all fish assigned were known to originate from sites in the baseline. For the most accurate and complete picture of assignments within a system, detailed knowledge of *all* breeding populations is required. It should be noted, however, that at certain geographical scales or for certain systems, assignment may not be possible with high accuracy, regardless of the samples and markers employed. If there is exchange of even modest amounts of spawning individuals over time between sites, then the genetic make-up of these sites will look relatively similar and prevent assignment to defined groups with high accuracy.

Implications for Management / Future Work

This analysis clearly indicates that both microsatellites and SNPs are capable of identifying discrete breeding units within the Kyle of Sutherland Fisheries Trust area. However, the use of SNPs allows for both a greater resolution of different breeding populations as well as increased accuracy of assignment to river/tributary of origin. Clearly, future genetic work along either of these lines should employ current and/or future panels of SNPs. However, microsatellites still have quite high power for particular applications that have may continue to be of use for the Trust (e.g. estimating numbers of breeders, pedigree analysis). SNPs offer the ability to generate a clearer picture of population genetic structuring within systems and allow for levels of genetic differences between sites to be demonstrated where previously they could not have been. This could be developed by addition of new sites for SNP analysis to gain a more robust picture of the different stock components within and among rivers within the Kyle of Sutherland Fisheries Trust area. Preliminary analysis demonstrates the possibility for accurate river-level assignments the Scottish baseline used continues to be expanded. Ultimately, the identification of genetic markers underlying particular traits of interest (e.g. sea age and run-timing) would be a

valuable tool. At present, this would involve some development work to identify candidate genetic markers before routine screening could be applied.

A number of factors may underlie population genetic structuring. At least two more, not addressed here, are the potential contribution or impact made by the stocking practices of fishery managers and that of aquaculture. The stocking of Atlantic salmon has been a common management intervention practice by fishery managers for many years within and between river systems. Such practices could influence why certain locations appear genetically distinct if the fish stocked to these locations were derived from brood fish or sourced directly from locations genetically different to the surrounding stock components. In recent and current times such hatchery and stocking activities have largely used brood fish sourced from the river where fish are to be introduced or from close by catchments. In the past there is evidence that introductions from far flung catchments or using fish donated by fish farms and of Norwegian origin may have been used; such past practice is more likely to have introduced genetically different strains to the wild populations. Alternatively, if stocking was widespread in an area, this could result in a more similar genetic make-up among stock components than would otherwise be the case.

However, in order to address the degree, if any, to which stocking plays a role in genetic structuring, detailed knowledge of the stocking history and records in the relevant catchments are essential; such information was not available to the current study. Furthermore, including genetic samples from possible donor sources, such as the strains of fish farmed in Scotland and which were on occasion donated to fishery managers for stocking activities by aquaculture companies or from catchments known to have provided previous broodstock, ova or juvenile fish in the past, would provide an insight into whether those donors had made a lasting impact on the local stock. Additionally, the availability of historical samples that pre-date the stocking history of an area would be of particular value in addressing the impact as it would offer a comparison of the genetic make-up pre- and post-stocking levels of differentiation.

Secondly, selective processes are involved in the domestication process that may differ from those in the wild (either intentionally or unintentionally). Furthermore, genetic drift (random differences) occurs not only as a result of domestication, but also within different cohorts or even family groups of individual

aquaculture strains. Such domestication effects can lead to differences in the type and frequency of genetic variants within the aquaculture strains and as such potentially allow them to be genetically differentiated from their wild originator stocks (Skaala et al. 2004; Glover et al. 2010b; Vasemagi et al. 2012). These differences have been shown to potentially occur within a very small number of generations of domestication (indeed has been seen within a single generation, Christie et al 2012). Two samples of aquaculture escapees were screened in the current study and demonstrate that for highly divergent genetic strains (e.g. Norwegian), direct escapees are clearly genetically identifiable from native stocks. However, in many instances such identification is possible in the field or with scale reading. However, of interest is still to what degree, if any, do farm escapees interbreed with native stocks. Such differences have been used elsewhere to both identify farm of origin of aquaculture escapes (Glover et al. 2008, 2009) as well as assess the potential and degree of interbreeding between aquaculture escapes and wild fish (Clifford et al. 1998a,b; Bourret et al. 2011; Besnier et al. 2011; Glover et al. 2012).

Summary

With the use of SNPs, clearer signals of genetic differentiation have begun to emerge for particular areas (i.e. Oykel-Cassley-Carron) than was possible with the earlier microsatellite markers. These results suggest that there are distinct breeding populations both between and within systems in the Kyle of Sutherland Fisheries Trust area. However, the degree of these differences (even with the SNPs) is not sufficient to allow for high levels of assignment to all sites. Clearly, there are exceptions (evidenced by the Water of Dye) suggesting that assignment may work well for the *most* differentiated sites within the system. Screening of further sites within the Kyle of Sutherland area with SNPs may help to further clarify the extent of within-river genetic stock components and the relationships among them, as well as increase assignment accuracy to individual tributaries. Combined with preliminary work aimed at resolving accurate assignments to river of origin, these results will likely have applications in assigning marine-caught fish back to natal river as well as in mixed-stock fisheries. As newer, and more targeted (e.g. adaptive markers) genetic tools become available, they will offer further insights into salmon population structure and, in turn, assist the efficient management and conservation of this valuable resource.

References

Anderson, E.C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701-710.

Beacham, T.D., McIntosh, B. and Wallace, C. 2010. A comparison of stock and individual identification for sockeye salmon (*Oncorhynchus nerka*) in British Columbia provided by microsatellites and single nucleotide polymorphisms. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 1274-1290.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme F. 2004. GENETIX 4.05, logiciel sous Windows TM pour la genetique des populations. Laboratoire Genome, Populations, Interactions, CNRS UMR 5171, Universite de Montpellier II, Montpellier.

Castric, V. and Bernatchez, L. 2004. Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology*, **13**, 1299-1312.

Coulson, M.W., Steuart-Fothringham, P., Girvan, J., Armstrong, A., Cawelier, E., Stradmeyer, L., Gilbey, J., Sinclair, C. and Verspoor, E. 2012. An overview of Atlantic salmon population genetic structuring in the Forth Fisheries Trust. Marine Scotland Science Report 21/12.

Crawford, N.G. 2010. SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556-557.

Dillane, E., McGinnity, P., Coughlan, J.P., Galvin, P.T., Wilkins, N.P. and Cross, T.F. 2007. Spatial and temporal patterns in microsatellite DNA variation of wild Atlantic salmon, *Salmo salar*, in Irish Rivers. *Fisheries Management and Ecology*, **14**, 209-219.

Dillane, E., McGinnity, P., Coughlan, J.P., Cross, M.C., De Eyto, E., Kenchington, E., Prodohl, P. and Cross, T.F. 2008. Demographics and landscape features determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. *Molecular Ecology*, **17**, 4786-4800.

Dionne, M., Caron, F., Dodson, J.J. and Bernatchez, L. 2008. Comparative survey of within-river genetic structure in Atlantic salmon; relevance for management and conservation. *Conservation Genetics*, **10**, 869-879.

Englund, V., Niemela, E., Lansman, M. and Heino, M. 1999. Variations in Atlantic salmon, *Salmo salar* L., smolt age in tributaries of the River Teno, Finland. *Fisheries Management and Ecology*, **6**, 83-86.

Excoffier, L., Laval, G. and Schneider, S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47-50.

Fontaine, P.-M., Dodson, J.J., Bernatchez, L. and Slettan A. 1997. A genetic test of metapopulation structure in Atlantic salmon (*Salmo salar*) using microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 2434-2442.

Fraser, D.J., Weir, L.K., Bernatchez, L., Hansen, M.M. and Taylor, E.B. 2011. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**, 404-420.

Freamo, H., O'Reilly, P., Berg, P.R., Lien, S. and Boulding, E.G. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources*, **11**, 254-267.

Garant, D., Dodson, J.J. and Benatchez, L. 2000. Ecological determinants and temporal stability of the within-river population structure in Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **9**, 615-628.

Garcia de Leaniz, C., Fleming, I.A., Einum, S., Verspoor, E., Consuegra, S., Jordan, W.C., Aubin-Horth, N., Lajus, D.L., Villanueva, B., Ferguson, A., Youngson, A.F. and Quinn, T.P. 2007. Local adaptation. In: *Verspoor, E., Stradmeyer, L. and Nielsen, J.L. eds. The Atlantic Salmon: Genetics, Conservation and Management*. Oxford, Blackwell Publishing, 195-235.

Gilbey, J., and Coulson, M.W. 2013. Determination of the utility of using SNP markers for assigning Atlantic salmon to their Scottish river of origin. Marine Scotland Science Report. Vol xx No xx.

Glover, K.A., Hansen, M.M., Lien, S., Als, T.D., Hoyheim, B. and Skaala, O. (2010) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics*, **11**, 2.

Hansen, M.M., Nielsen, E.E. and Mensberg, K.-L.D. 1997. The problem of sampling families rather than populations: relatedness among individuals in samples of juvenile brown trout *Salmo trutta* L. *Molecular Ecology*, **6**, 469-474.

Hess, J.E., Matala, A.P. and Narum, S.R. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, **11**, 137-149.

Jones, O.R. and Wang, J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551-555.

Jost, L. 2008. GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015-4026.

Kalinowski, S.T. 2005. HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187-189.

Kalinowski, S.T., Manlove, K.R. and Taper, M.L. 2007. ONCOR a computer program for genetic stock identification. Montana State University, Bozeman. Available from www.montana.edu/kalinowski/Software/ONCOR.htm

King, T.L., Eackles, M.S. and Letcher, B.H. 2005. Microsatellite DNA markers for the study of Atlantic salmon (Salmo salar) kinship, population structure, and mixed-fishery analyses. *Molecular Ecology Notes*, **5**, 130-132.

King, T.L., Kalinowski, S.T., Schill, W.B., Spidle, A.P. and Lubinski, B.A. 2001. Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, **10**, 807-821.

King, T.L., Verspoor, E., Spidle, A.P., Gross, R., Phillips, R.B., Koljonen, M-L., Sanchez, J.A. and Morrison, C.L. 2007. Biodiversity and population structure. In: *Verspoor, E., Stradmeyer, L. and Nielsen, J.L. eds. The Atlantic Salmon: Genetics, Conservation and Management*. Oxford, Blackwell Publishing, 117-166.

Landry, C. and Bernatchez, L. 2001. Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **10**, 2525-2539.

Lien, S., Gidskehaug, L., Moen, T., Hayes, B.J., Berg, P.R., Davidson, W.S., Omholt, S.W. and Kent, M.P. 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.

McConnell, S.K., O'Reilly, P., Hamilton, L., Wright, J.M. and Bentzen, P. 1995. Polymorphic microsatellite loci from Atlantic salmon (Salmo salar): genetic differentiation of North American and European populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 1863-1872.

McConnell, S.K.J., Ruzzante, D.E., O'Reilly, P.T., Hamilton, L. and Wright, J.M. 1997. Microsatellite loci reveal highly significant genetic differentiation among Atlantic salmon (*Salmo salar* L.) stocks from the east coast of Canada. *Molecular Ecology*, **6**, 1075-1089.

McGinnity, P., Prodohl, P., Ferguson, A., Hynes, R., Maoileidigh, N.O., Baker, N., Cotter, D., O'Hea, B., Cooke, D., Rogan, G., Taggart, J. and Cross, T.F. 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 of London Series B – Biological Sciences*, **270**, 2443-2450.

Narum, S.R. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783-787.

Narum, S.R., Banks, M., Beacham, T.D., Bellinger, M.R., Campbell, M.R., Dekoning, J., Elz, A., Guthrie, C.M., Kozfkay, C., Miller, K.M., Moran, P., Phillips, R., Seeb, L.W., Smith, C.T., Warheit, K., Young, S.F. and Garza, J.C. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, **17**, 3464-3477.

Niemela, E., Erkinaro, J., Julkunen, M., Hassinen, E., Lansman, M. and Brors, S. 2006. Temporal variation in abundance, return rate and life histories of previously spawned Atlantic salmon in a large subarctic river. *Journal of Fish Biology*, **68**, 1222-1240.

O'Reilly, P.T., Hamilton, L.C., McConnell, S.K. and Wright, J.M. 1996. Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 2292-2298.

Paterson, S., Piertney, S.B., Knox, D., Gilbey, J. and Verspoor, E. 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes*, **4**, 160-162.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L. and Estoup A. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536-539.

Rannala, B. and Mountain, J. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9197-9201.

Rousset, F. 2008. Genepop '007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

Ryman, N. 2005. CHIFISH: a computer program testing for genetic heterogeneity at multiple loci using chi-square and Fisher's exact test. *Molecular Ecology Notes*, **6**, 285-287.

Sánchez, J.A., Clabby, C., Ramos, D., Blanco, G., Flavin, F., Vázquez, E. and Powell, R. 1996. Protein and microsatellite single locus variability in Salmo salar L. (Atlantic salmon). *Heredity*, **77**, 423-432.

Seeb, L.W., Templin, W.D., Sato, S., Abe, S., Warheit, K., Park, J.Y. and Seeb, J.E. 2011a. Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Molecular Ecology Resources*, **11**, 195-217.

Seeb, L.W., Seeb, J.E., Habicht, C., Farley, E.V. and Utter, F.M. 2011b. Single-nucleotide polymorphic genotypes reveal patterns of early juvenile migration of sockeye salmon in the eastern Bering Sea. *Transactions of the American Fisheries Society*, **140**, 734-748.

Smith, C.T., Templin, W.D., Seeb, J.E. and Seeb, L.W. 2005. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries. *North American Journal of Fisheries Management*, **25**, 944-953.

Spidle, A.P., Kalinowski, S.T., Lubinski, B.A., Perkins, D.L., Beland, K.F., Kocik, J.F. and King, T.L. 2003. Population structure of Atlantic salmon in Maine with reference to populations from Atlantic Canada. *Transactions of the American Fisheries Society*, **132**, 196-209.

Stewart, D.C., Smith, G.W. and Youngson, A.F. 2002. Tributary-specific variation in timing of return of adult Atlantic salmon (*Salmo salar*) to fresh water has a genetic component. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 276-281.

Stewart, D.C., Middlemas, S.J. and Youngson, A.F. 2006. Population structuring in Atlantic salmon (*Salmo salar*): evidence of genetic influence on the timing of smolt migration in sub-catchment stocks. *Ecology of Freshwater Fish*, **15**, 552-558.

Taylor, E.B. 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, **98**, 185-207.

Vaha, J.-P., Erkinaro, J., Niemela, E. and Primmer, C.R. 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, **16**, 2638-2654.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P. and Shipley, P. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Verspoor, E. 2007. Introduction. In: *Verspoor, E., Stradmeyer, L. and Nielsen, J.L. eds. The Atlantic Salmon: Genetics, Conservation and Management*. Oxford, Blackwell Publishing, 1-13.

Verspoor, E., Beardmore, J.A., Consuegra, S., Garcia de Leaniz, C., Hindar, K., Jordan, W.C., Koljonen, M.-L., Mahkrov, A.A., Paaver, T., Sanchez, J.A., Skaala, O., Titov, S. and Cross, T.F. 2005. Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *Journal of Fish Biology*, **67**, 3-54.

Verspoor, E. and Laughton, R. 2008. Sampling river stocks of Atlantic salmon for genetic analysis to understand structuring into breeding populations. FRS Internal Report 2008.

Wang, J. and Santure, A.W. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*, **181**, 1579-1594.

Waples, R.S. 2010. High-grading bias: subtle problems with assessing power of selected subsets of loci for population assignment. *Molecular Ecology*, **19**, 2599-2601.

Webb, J., Verspoor, E., Aubin-Horth, N., Romakkaniemi, A. and Amiro, P. 2007. The Atlantic salmon. In: *Verspoor, E., Stradmeyer, L. and Nielsen, J.L. eds. The Atlantic Salmon: Genetics, Conservation and Management*. Oxford, Blackwell Publishing, 17-56.

Weir, B.S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Appendix 1

Laboratory Procedures

Microsatellites

DNA was extracted from individual fin clips using a standard proteinase K digestion (Fisher Scientific UK). The crude DNA extract was diluted (1 in 10) in 1xTE (Tris-EDTA) buffer for all further work. Seventeen microsatellite markers that have previously been developed for Atlantic salmon were amplified from each DNA extract by polymerase chain reaction (PCR) using fluorescently-labelled primers. The microsatellite markers used were: SP2201, Sp2210, SPG7, SP1605, SP1608, SP2216, SP3016 (Paterson et al., 2004), SsaD144, SsaD157, SsaD48, SsaD71 (King et al., 2005), Ssa14, Ssa289 (McConnell et al., 1995), Ssa202, Ssa171, Ssa197 (O'Reilly et al., 1996), SsaF43 (Sánchez et al., 1996). These 17 markers were amplified in three multiplex reactions according to the mixtures in Table 1 of this appendix. PCR reactions were conducted using the Type-it Microsatellite PCR kit (Qiagen). Cycle conditions were as follows : an initial denaturation at 95^{o}C for 5 min followed by 32 cycles of 94^{o}C for 30 s, annealing at either 58^{o}C (mix A and C) or 55^{o}C (mix B) for 90 s and extension at 72^{o}C for 60 s. After cycling, a final extension was completed at 60^{o}C for 30 min.

PCR products were run on a MegaBACE capillary sequencer (Amersham Biosciences) and compared against a size standard of Et Rox 550 (GE Healthcare) run along with each sample. Fragment sizes were scored with Fragment Profiler version 1.2 software (GE Healthcare). For data quality control, all results were independently checked by two people and in addition one in five results were scored “double-blind” and the results compared to calculate error rates.

SNPs

DNA was extracted using a QIAGEN DNeasy extraction kit, following the manufacturer’s recommendations. Extracted DNA was quantified on a Nanodrop DNA quantification system to ensure samples met the minimum required DNA concentration of 50 ng/ul. Samples meeting this requirement were subsequently sent to the Centre for Integrative Genomics (CIGENE) in Norway for SNP

screening on a V2 Illumina chip. Data calls and quality control were carried out by CIGENE staff and the raw data and quality control results were returned to FASMOP staff for population genetic analyses.

Data Analysis

Microsatellites

The results from the microsatellite marker SsaF43 allowed us to identify any trout/salmon hybrids that may be present among samples, and also any mis-identified trout. The genetic information from these individuals was then removed from further analysis.

In order to remove bias in the data due to over-representation of family groups, an analysis of family relationships was performed using the software COLONY (Wang & Santure, 2009, Jones & Wang 2010) to identify full-sibling individuals. Furthermore, this analysis allowed for a prediction as to the number of breeders that contributed to each sample. For each location sampled, all but one member of a full-sibling group were removed from analysis.

Where there was more than one site sampled within a 5-km distance, two life-history stages (i.e. fry and parr) sampled at the same site and/or a site was sampled in different years, the data were initially tested for differences using the program CHIFISH (Ryman 2006). Where no significant differences were found, data from these sites or time points were combined; otherwise they were left separate for all further analyses.

The program MICROCHECKER (Van Oosterhout et al. 2004) was used to screen for genotyping errors and non-amplifying variants (null alleles) in the raw data. In addition markers were checked for conformity to linkage equilibrium (probability test) and Hardy-Weinberg equilibrium (exact test), as implemented by GENEPOP version 4 (Rousset, 2008). In all cases, inference of significance was corrected for multiple-testing using the false discovery rate (FDR) method (Narum, 2006). Allelic richness is an estimate of the number of genetic variants found in a sample after controlling for sample size. This was calculated using the program HP-Rare (Kalinowski, 2005), and allows an assessment of differences in genetic diversity among samples standardized to a common sample size (N = 32).

The genetic structure between groups was examined using two measures of genetic differentiation - pairwise F_{ST} (calculated as θ; Weir & Cockerham 1984) calculated in the program GENETIX (Belkhir et al. 2004) and assessed for significance with permutation tests using 500 randomizations. The second measure of differentiation, pairwise Jost's D (Jost, 2008), was calculated with the program SMOGD (Crawford 2010). A pairwise matrix of both distance measures is presented in Table 2 of this appendix. A multi-dimensional scaling (MDS) plot was drawn to illustrate the relationships among sites using the Jost's D measure of differentiation.

Table 1

List of microsatellites used in the genetic survey with primer sequences, multiplex mixture, final primer concentration in the PCR and the reference reporting the microsatellite locus.

Microsatellite marker	Sequence forward primers 5'-3'	Sequence reverse primers 5'-3'	Multiplex mixture	Final primer concentration (µM)	Reference
Sp2201	TTTAGATGGTGGGATACTGGGAGGC	CGGGAGCCCCATAACCCTACTAATAAC	A	0.02	Paterson et al., 2004
Sp2210	AAGTATTCATGCACACACATTCACTGC	CAAGACCCTTTTTCCAATGGGATTC	A	0.02	Paterson et al., 2004
SPG7	CTTGGTCCCGTTCTTACGACAACC	TGCACGCTGCTTGGTCCTTG	A	0.02	Paterson et al., 2004
Ssa 202	CTTGGAATATCTAGAATATGGC	TTCATGTGTTAATGTTGCGTG	A	0.02	O'Reilly et al., 1996
SsaD144	TTGTGAAGGGGCTGACTAAC	TCAATTGTTGGGTGCACATAG	A	0.03	King et al., 2005
SsaD157	ATCGAAATGGAACTTTTGAATG	GCTTAGGGCTGAGAGAGGATTAC	A	0.03	King et al., 2005
Sp1605	CGCAATGGAAGTCAGTGGACTGG	CTGATTTAGCTTTTTAGTGCCCAATGC	B	0.015	Paterson et al., 2004
Sp1608	AGCACACTCATCATCTTACCTAGAG	ATGGACAGAAAGATAATGAGGG	B	0.015	Paterson et al., 2004
Sp2216	GGCCCAGACAGATAAACAAACACGC	GCCAACAGCAGCATCTACACCCAG	B	0.015	Paterson et al., 2004
Ssa171	TTATTATCCAAAGGGGTCAAAA	GAGGTCGCTGGGGTTTACTAT	B	0.015	O'Reilly et al., 1996
Ssa14	CCTTTTGACAGATTTAGGATTTC	CAAACCAAACATACCTAAAGCC	B	0.02	McConnell et al., 1995
Ssa289	GTTTCTTTACAAATAGACAGACT	TCATACAGTCACTATCATC	B	0.02	McConnell et al., 1995
Sp3016	GACAGGGCTAAGTCAGGTCA	GATTCTTATATACTCTTATCCCCAT	C	0.02	Paterson et al., 2004
Ssa197	GGGTTGAGTAGGGAGGCTTG	TGGCAGGGATTTGACATAAC	C	0.02	O'Reilly et al., 1996
SsaF43	AGCGGCATAACGTGCTGTGT	GAGTCACTCAAAGTGAGGCC	C	0.02	Sánchez et al., 1996
SsaD48	GAGCCTGTTCAGAGAAATGAG	CAGAGGTGTTGAGTCAGAGAAG	C	0.03	King et al., 2005
SsaD71	AACGTGAAACATAAATCGATGG	TTAAGAATGGGTTGCCTATGAG	C	0.03	King et al., 2005

The utility of the data to assign fish of unknown origin to sample site was examined by running individual assignment tests using the program ONCOR (Kalinowski et al. 2007). Assignments were conducted using the method of Rannala & Mountain (1997). Location of assignment was taken as the site with the highest probability. Only individuals with a complete multi-locus genotype (i.e. all 17 microsatellites) were chosen for assignment as comparing criterion values for individuals with differing number of markers typed is difficult (Piry et al 2004). Caution should be used when interpreting these results as the locations used may not represent the full spread of genetic diversity or populations present within the catchment, as well as differences in sample size, may affect the results.

SNPs

Prior to analysis, several quality control measures were applied to the dataset. Firstly, SNPs with cluster patterns that could not be reliably scored were removed. Secondly, individuals with successful call rates at less than 98% of the SNPs screened, were also removed from the analysis. Finally, SNPs that were monomorphic (i.e. invariant across the sites being considered) or had a minimum allele frequency of less than 5% overall, were removed. These measures left a dataset containing 221 individuals for 3802 variable SNPs.

A test was used to identify so-called ‘outlier’ SNPs, which may be indicative of markers under the influence of selection. This followed the method of Beaumont & Nichols (1996), as implemented in ARLEQUIN v3.5 (Excoffier *et al.* 2005). This method detects loci with significantly high or significantly low F_{ST} values, controlled for locus-specific heterozygosity. Markers were classified as outliers if they fell within the top 1% of the distribution.

In order to find a subset of SNPs useful for further analyses, loci were ranked according to a measure of genetic differentiation (F_{ST}) as this was found to provide the highest levels of accuracy of assignment to river (Gilbey *et al*. in preparation). When selecting panels of markers for assignment, Anderson (2010) demonstrated an upward bias in assignment success when the same samples are used for both locus selection and evaluation of assignment. Therefore we adopted the ‘training-holdout-leave-one-out’ (THL) procedure described in Anderson (2010). Each site was first divided in half and one half used as a

‘training’ set to pick the loci. Loci were ranked (based upon F_{ST}) from highest to lowest and the top 50, 100, 200, 300, 500, & 1000 SNPs were selected for further investigation from the training set. Both the training and holdout sets were used as a reference baseline for assignments, but only the holdout set was used to evaluate assignment accuracy (see Anderson 2010 for more details). Pairwise F_{ST} and assignments were conducted on the SNP datasets as described above for the microsatellites.

Appendix 2

Pairwise estimates of genetic differentiation among microsatellite baseline groups as defined in Table 1 (main text). Jost's D above diagonal, F_{ST} below diagonal. Significant pairwise F_{ST} values are indicated in italics and shaded in gray.

	1	2	3	4	5	6a	6b	7	8a	8b	9	10	11	12	13 (2009	13 (2011	14	15	16	17	18
1	--	0.002	0.026	0.004	0.029	0.029	0.052	0.019	0.045	0.120	0.114	0.102	0.097	0.109	0.129	0.173	0.083	0.247	0.172	0.184	0.143
2	0.001	--	0.018	0.002	0.021	0.019	0.046	0.014	0.039	0.134	0.130	0.085	0.100	0.098	0.121	0.171	0.093	0.224	0.216	0.186	0.165
3	*0.007*	*0.006*	--	0.033	0.041	0.037	0.024	0.031	0.044	0.124	0.153	0.135	0.131	0.135	0.162	0.182	0.132	0.300	0.237	0.235	0.237
4	0.003	0.001	*0.007*	--	0.010	0.004	0.021	0.002	0.037	0.108	0.120	0.105	0.127	0.079	0.133	0.160	0.086	0.259	0.188	0.179	0.172
5	*0.007*	0.004	*0.009*	*0.005*	--	0.010	0.032	0.020	0.045	0.155	0.138	0.126	0.132	0.107	0.160	0.174	0.119	0.275	0.211	0.173	0.176
6a	*0.009*	*0.008*	*0.009*	0.004	*0.006*	--	0.023	0.028	0.045	0.154	0.123	0.095	0.113	0.082	0.153	0.169	0.094	0.256	0.212	0.181	0.171
6b	*0.010*	*0.008*	*0.007*	*0.005*	*0.008*	*0.006*	--	0.038	0.037	0.149	0.144	0.115	0.128	0.126	0.177	0.191	0.110	0.287	0.217	0.162	0.257
7	*0.005*	0.003	*0.008*	0.001	*0.006*	*0.007*	*0.007*	--	0.019	0.086	0.154	0.123	0.133	0.085	0.143	0.178	0.107	0.255	0.217	0.195	0.147
8a	*0.011*	*0.008*	*0.009*	*0.006*	*0.010*	*0.010*	*0.008*	*0.005*	--	0.103	0.156	0.125	0.121	0.122	0.146	0.164	0.101	0.266	0.226	0.190	0.223
8b	*0.027*	*0.029*	*0.033*	*0.025*	*0.033*	*0.031*	*0.032*	*0.022*	*0.024*	--	0.244	0.230	0.210	0.195	0.234	0.278	0.213	0.311	0.221	0.209	0.257
9	*0.027*	*0.022*	*0.031*	*0.025*	*0.026*	*0.029*	*0.032*	*0.026*	*0.030*	*0.058*	--	0.015	0.021	0.039	0.038	0.080	0.009	0.149	0.243	0.213	0.259
10	*0.026*	*0.017*	*0.026*	*0.020*	*0.023*	*0.023*	*0.025*	*0.022*	*0.024*	*0.056*	*0.004*	--	0.000	0.012	0.010	0.045	0.000	0.137	0.259	0.211	0.245
11	*0.025*	*0.020*	*0.025*	*0.022*	*0.025*	*0.024*	*0.025*	*0.025*	*0.024*	*0.054*	*0.008*	0.000	--	0.018	0.036	0.073	0.007	0.164	0.256	0.183	0.262
12	*0.023*	*0.019*	*0.026*	*0.016*	*0.023*	*0.020*	*0.025*	*0.019*	*0.023*	*0.050*	*0.012*	0.005	*0.009*	--	0.034	0.092	0.017	0.181	0.249	0.200	0.179
13 (2009	*0.033*	*0.024*	*0.034*	*0.027*	*0.032*	*0.033*	*0.037*	*0.030*	*0.032*	*0.066*	*0.011*	0.006	*0.011*	*0.011*	--	0.033	0.014	0.162	0.266	0.236	0.291
13 (2011	*0.039*	*0.032*	*0.040*	*0.034*	*0.037*	*0.041*	*0.042*	*0.036*	*0.035*	*0.067*	*0.023*	*0.014*	*0.016*	*0.023*	*0.012*	--	0.063	0.185	0.260	0.188	0.341
14	*0.022*	*0.018*	*0.026*	*0.015*	*0.021*	*0.022*	*0.023*	*0.019*	*0.023*	*0.052*	*0.009*	0.002	0.004	0.005	*0.009*	*0.017*	--	0.156	0.247	0.237	0.192
15	*0.065*	*0.056*	*0.071*	*0.064*	*0.065*	*0.067*	*0.074*	*0.061*	*0.068*	*0.099*	*0.042*	*0.040*	*0.046*	*0.051*	*0.051*	*0.062*	*0.051*	--	0.348	0.317	0.364
16	*0.037*	*0.039*	*0.047*	*0.032*	*0.039*	*0.042*	*0.040*	*0.037*	*0.045*	*0.059*	*0.054*	*0.058*	*0.057*	*0.052*	*0.064*	*0.063*	*0.048*	*0.106*	--	0.060	0.264
17	*0.038*	*0.036*	*0.044*	*0.031*	*0.037*	*0.034*	*0.033*	*0.035*	*0.038*	*0.053*	*0.047*	*0.047*	*0.044*	*0.040*	*0.056*	*0.050*	*0.045*	*0.093*	*0.017*	--	0.286
18	*0.030*	*0.032*	*0.043*	*0.027*	*0.034*	*0.034*	*0.041*	*0.028*	*0.040*	*0.055*	*0.054*	*0.052*	*0.054*	*0.044*	*0.063*	*0.065*	*0.041*	*0.097*	*0.050*	*0.054*	--

Marine Scotland Science
Freshwater Laboratory
Faskally
Pitlochry
PH16 5LB

Copies of this report are available from the Marine Scotland website at
www.scotland.gov.uk/marinescotland