The Scottish Government
Riaghaltas na h-Alba

Marine Scotland Science

Marine Scotland Science Report 04/13

An Overview of Population Genetic Structuring in the River Tay

M W Coulson, D Summers, A Armstrong, E Cauwelier,
L Stradmeyer, J Gilbey, C Sinclair & E Verspoor

Marine Scotland Science Report 04/13

An Overview of Population Genetic Structuring in the River Tay

M W Coulson, D Summers, 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 in the River Tay

M W Coulson[1], D Summers[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]Tay DSFB, Site 6, Cromwell Park, Almondbank, Perth, Perthshire PH1 3LW

Executive Summary

This report describes how genetic information from juvenile Atlantic salmon sampled from 12 sites within the River Tay (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.

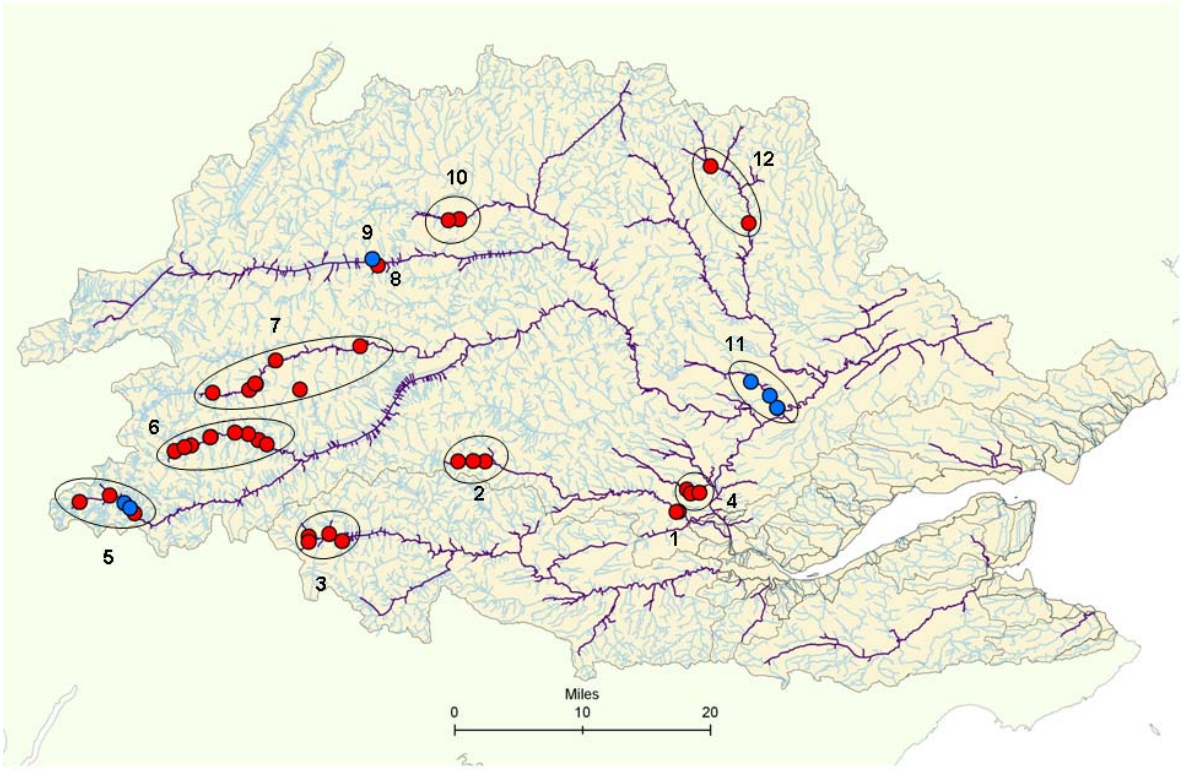


Figure 1. A map of the River Tay, with sample sites that are involved in this report. All sites were screened for microsatellite genetic markers while those in blue were additionally screened for SNPs. Numbers refer to site names in Table 1 (see main text). Ellipses represent samples grouped for analysis.

Summary of findings

The analysis showed that most sites exhibited moderate genetic differences from one another with the markers used, indicating genetic structuring among these sites. The most differentiated sites were the Inner Haddon Burn (8), Tummel (9), Crom Allt-Cononish-Fillan (5) and Kendrum-Ogle Burn (3), which all showed reduced genetic diversity at each of these locations compared to other sites throughout the system.

This degree of genetic differentiation observed within the Tay is largely reflected by the ability to predict where a sample is from, using only genetic information (genetic assignment); where genetic signatures are strongly related to location, individuals are more likely to be assigned to the location from which they were originally sampled. The average value of correct assignment to site was only 49%, which is higher than one would expect if there was no genetic structure in the data. However, assignment to several of the most differentiated sites approached 70%.

For three of the sites (blue dots on Figure 1), a new set of markers, single nucleotide polymorphisms (SNPs) was utilised to compare the levels of differentiation with the current suite of microsatellite markers. SNPs greatly increased the level of genetic differentiation among sites, by up to 7-fold, despite both marker types being able to resolve these locations as genetically differentiated. Addition of further sites within the Tay, screened with SNPs, would allow for further resolution regarding the different stock components within the system, and allow for a more comprehensive appraisal regarding the ability of SNPs to be used in genetic assignment.

Implications for management

The aim of the current analysis was to identify distinct breeding populations of salmon within the Tay. The results to date suggest that there are distinct breeding populations within the system. These results suggest caution should be taken regarding stocking, as underlying heritable, adaptive differences may be present among locations. The moderate to strong genetic structuring observed is likely influenced by the large catchment size, the dendritic nature of the system and the presence of multiple lochs, among other factors. Further geographical coverage of the system, along with the use of SNPs, will offer the ability to generate a clearer picture of population genetic structuring within the Tay and allow for an assessment into the ability to use these genetic differences in further management applications (e.g. genetic assignment). Furthermore, the SNP data presented here has also contributed to a preliminary analysis on a Scotland-wide scale, demonstrating the possibility for accurate genetic assignment back to river of origin. While it is not yet possible to produce a cost effective genetic method of identifying the origins of salmon caught in coastal waters, with further development, the

between river differences in SNPs have the potential to have applications in mixed-stock fisheries and assignment of marine-caught fish.

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 Tay using a panel of microsatellite markers and compare, in a limited extent, 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 River Tay 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 12 sites that have been included in the genetic analysis for the Trust. Samples generally consisted of fry and/or parr (n= 24-80, depending on site) and, for each individual, data from 17 genetic markers (microsatellites) were collected. From three of these sites, data from 5568 SNPs were collected for the same individuals as screened for microsatellites.

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.

As many individual sites contained small numbers of fish (n = 1 – 20), sites within a given area were pooled for analysis and this pooling scheme is represented in Figure 1. With such small sample sizes, testing for genetic differences of allele frequencies prior to pooling can be unreliable (Ryman 2006). Additionally, pooling of small samples (<25) has been shown to produce more informative analyses of genetic structuring (Gomez-Uchida & Banks, 2005). This resulted in 12 samples for subsequent analyses.

In addition to the microsatellites, for a subset of sites (see 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 594 juvenile salmon from the River Tay were involved in the genetic analysis. There was a single salmon/trout hybrid identified at one location. All sites were examined for family effects, with relatively few samples being removed due to full-sibling relationships (Table 1). The level of family effects differed between samples, with the largest family group present in the individual samples ranging from 1 to 3 full-siblings

and sample sizes subsequently being reduced by 0-16%. Family effects were controlled for at each site before all further analyses.

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 analyzed (sibs removed)†	Number of breeders contributing to sample	Largest single family	Sample Size (SNPs)	Year sampled
Lower Almond	1	30	29	51	2	--	2009
Upper Almond	2	50	44	55	3	--	2009
Kendrum[1]-Ogle	3	24	20	25	2	--	2009
Schochie Burn	4	24	24	40	1	--	2009
Crom Allt, Cononish & Fillan	5	75	48 (*71*)	86	2	32	2007/ 2009
River Lochay	6	79	56 (*75*)	83	3	--	2007/ 2009
River Lyon	7	80	65 (*77*)	88	2	--	2007/ 2009
Inner Haddon Burn	8	50	49	59	2	--	2006
Tummel	9	55	54	59	2	21	2006
Errochty Water	10	50	48	68	2	--	2009
Lunan Water	11	47	44	67	2	27	2009
Shee	12	30	30	50	1	--	2007

1. One sample from this location was a salmon/trout hybrid

†**Note**: Sample size analysed reflects the final sample size included in the genetic analysis, while sample sizes in parentheses indicate the sample size after removing full-sibs but before removing samples as part of quality controlling (samples typed for fewer than 14 of the 17 microsatellites were not included in the analysis).

Population structuring – Microsatellite baseline

The genetic differences among sites showed a modest degree of genetic differentiation, and 95% (63 out of 66) of the pairwise comparisons between sites were significantly different (Appendix 2). The three comparisons that were not different from one another were: (1) the Inner Haddon Burn vs. Tummel, (2) the Schochie site vs. the Upper Almond and (3) the Schochie site vs. the lower Almond. A visual representation of these relationships among locations 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 to the Inner Haddon Burn (8), Tummel (9), Crom Allt-Cononish-Fillan (5) and

Kendrum-Ogle Burn (3) sites and, to a lesser extent, the Lunan (11). The two Almond sites (1 & 2) and the Schochie Burn (4) were closest to each other, as were the Lochy (6), Lyon (7), Errochty (10) and Shee sites (12) (Figure 2). However, all of these sites were still significantly different from one another.

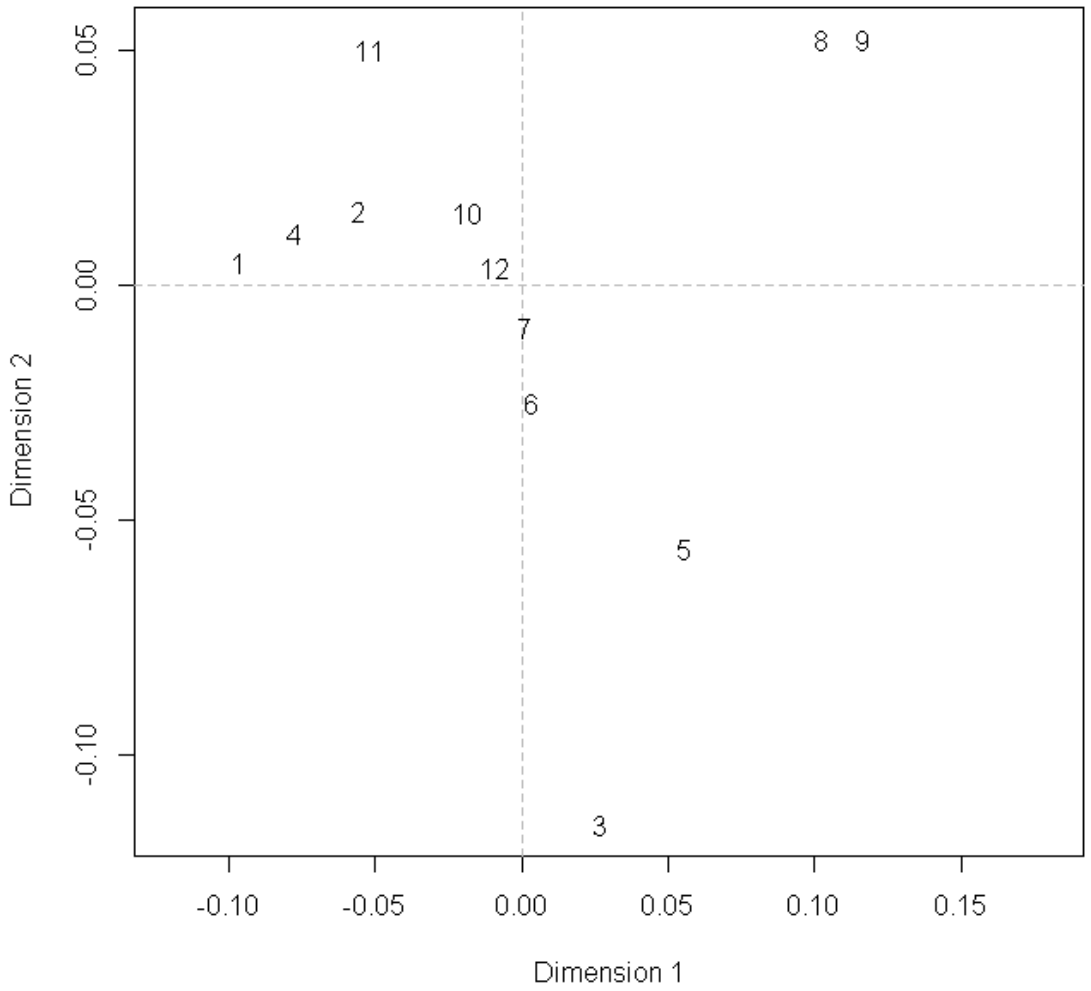
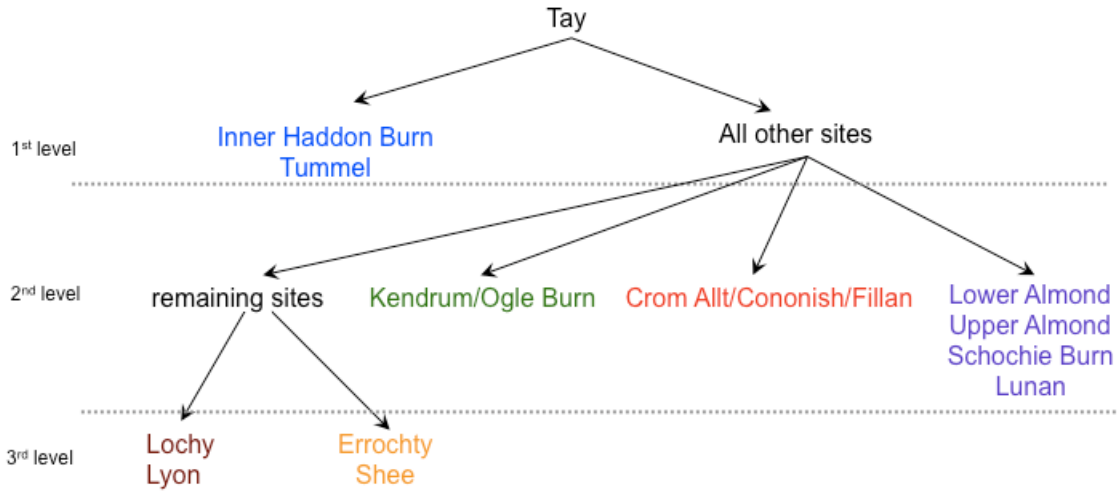


Figure 2. Multidimensional scaling (MDS) plot of genetic relationships among all sites, based on pairwise estimates of genetic differentiation (Jost's *D*; see the appendix for details) for the microsatellite data. Points that are closer together on the plot have a more similar genetic makeup, while points further apart are more genetically discrete. Sample codes are as presented in Table 1.

The Kendrum-Ogle, Crom Allt-Cononish-Fillan, Inner Haddon Burn and Tummel sites all showed a lower degree of genetic diversity at each location compared to the remaining sites. On average, these four locations had two fewer genetic variants (termed allelic richness) at individual markers, resulting in up to a 6% decrease in genetic diversity compared to the remaining sites.

A clustering analysis that explores possible groupings of individuals other than the defined sampling sites was also carried out. This analysis was done in a hierarchical fashion, as larger genetic differences among groups may obscure weaker differences among groups. For each level, the analysis aims to determine, from a given number of samples, the most likely number of groups and the membership of each individual into those groups. Each of the groupings identified by the first round are then separately analysed in a second round and this process is repeated until further identification of groupings is not possible. At the broadest level, the analysis revealed two most likely groups: (1) Inner Haddon Burn and the Tummel sites and (2) all remaining sites. These remaining sites were then analysed in a second round and four groups were determined to be most likely: (1) a group comprising the Lower & Upper Almond, Schochie Burn and the Lunan, (2) the Crom Allt/Cononish/Fillan sites, (3) Kendrum/Ogle Burn sites and (4) all remaining sites. Among this latter group, a third round of analysis revealed a final two groups: (1) a group comprising the Lochay and Lyon and (2) a group consisting of the Errochty Water and Shee sites. Therefore, overall the clustering analysis identified six genetic groupings. This does not mean, however, that there are not significant genetic differences within each of these groups, but rather, using this clustering approach, these smaller differences are more difficult to tease apart and the distinction for further splitting is less obvious. A schematic of the levels of analysis at which each group was resolved is shown in Figure 3A with the geographical representation of these groups depicted in Figure 3B.

(A)



(B)

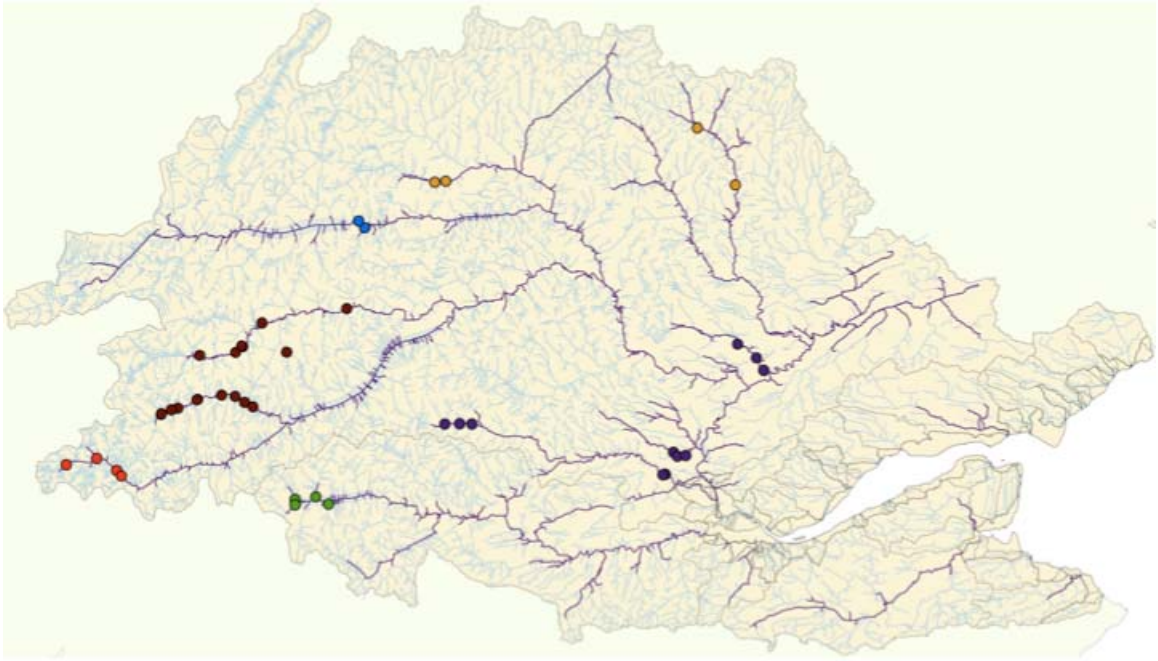


Figure 3. (A) Schematic summarizing the order in which the different groups were resolved according to the hierarchical clustering analysis. (B) Geographic representation of the genetic relationships among sites, following the cluster analysis. Locations with the same colour are more similar to one another and belong in the same cluster.

Population structuring – SNPs

Three 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, along with their respective sample sizes. To allow for a direct comparison between marker types, the microsatellite statistics were re-calculated for only the three sites involved in the SNP analysis. The genetic differences among these three sites, for the microsatellites, show (as expected) a moderate magnitude of genetic differentiation (Table 2), and all three of the pairwise comparisons are significantly different.

For the total panel of SNP markers among the three sites screened for the Tay, 46 'outlier' SNPs were identified and analysis of just these 46 SNPs resulted in all three pairwise comparisons among sites being significantly different. Even though these sites were relatively well differentiated with the microsatellites, the 'outlier' SNPs showed ~7-fold increase in the overall level of differentiation, as measured by F_{ST} (Table 2). So, while both marker types resolved significant structuring, the SNPs demonstrated a strengthening in the signal separating these sites (Figure 4). Furthermore, these 46 'outlier' SNPs showed a slightly higher level of differentiation than the top 100 ranked SNPs. Clearly, all three sites were well differentiated from one another, however due to the limited sampling involved in the SNP analysis, further exploration of the data was limited. It would be of interest to include further sites, screened for SNPs, within the Tay to assess the broader levels of structuring within the system. The three sites chosen are situated far apart throughout the system and two of them are above lochs. As such, it might be expected to show significant degrees of structuring. It would be of interest to perform the same comparison at some of the sites where the microsatellites showed lower differentiation to determine if the SNPs could enhance detection of population structure at such sites.

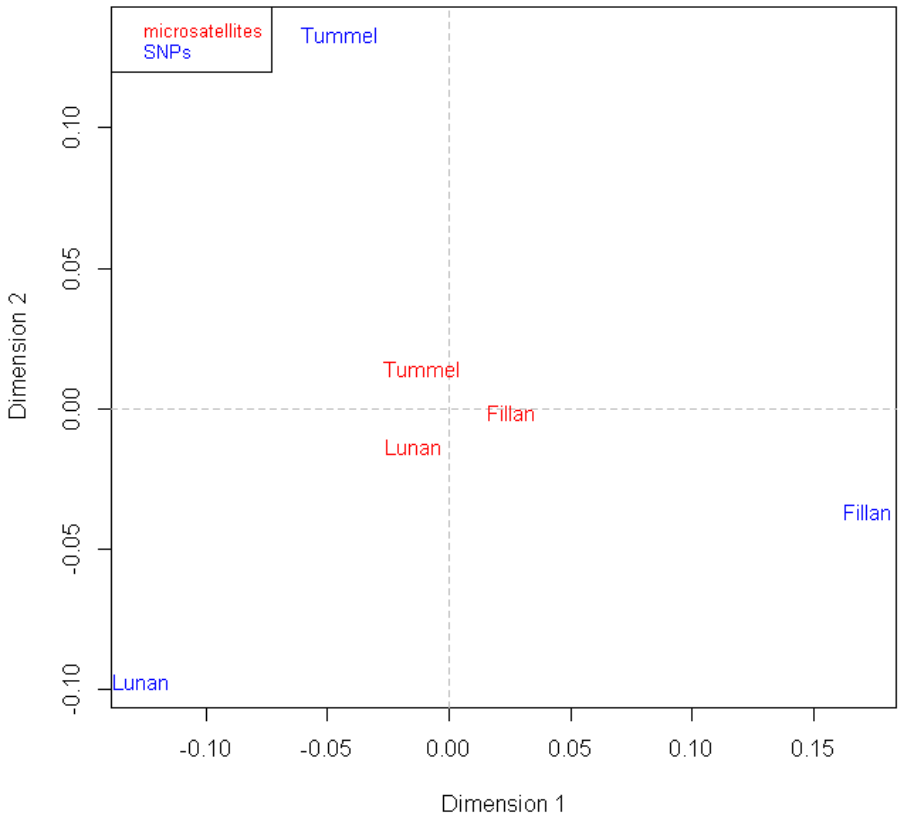


Figure 4. Multidimensional scaling (MDS) plot of genetic relationships among the sites screened for both 17 microsatellites and a set of 46 high-ranked 'outlier' SNPs (based on pairwise F_{ST}). Points closer together have a more similar genetic makeup, while points further apart are more genetically different.

Table 2. Comparison of the levels of differentiation for microsatellites versus SNPs for sites in Figure 3.

	Microsatellites (3 sites) (F_{ST})	Top 100 F_{ST} ranked SNPs (all)	46 'outlier' SNPs (all)
Overall differentiation	0.036	0.199	0.278
Range of levels of differentiation	0.028-0.042	0.159-0.234	0.245-0.305
% of significant pairwise comparisons	100%	100%	100%

Genetic assignment – 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 River Tay, 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 5). 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. Given the moderate level of differentiation defined with the microsatellites, assignment of fish back to their specific site of collection was, on average, correct 49% of the time. This is higher than would be expected if assignments were purely random (12 sites, random = ~8%). Assignment success was highest for Kendrum-Ogle, Crom Allt-Cononish-Fillan, Lyon & Errochty, which was between 60-75%. Despite the Inner Haddon Burn and Tummel sites being most distinct, correct assignment to each of these sites was ~50%. However, looking at the mis-assigned fish revealed that more than 45% of the Inner Haddon Burn fish sampled were being assigned to the Tummel and vice versa. Therefore, combining these two locations (as there was no observable genetic differences between them and they are geographically close) into a single group (Haddon-Tummel) resulted in 94% of fish from these two sites being assigned correctly (data not shown).

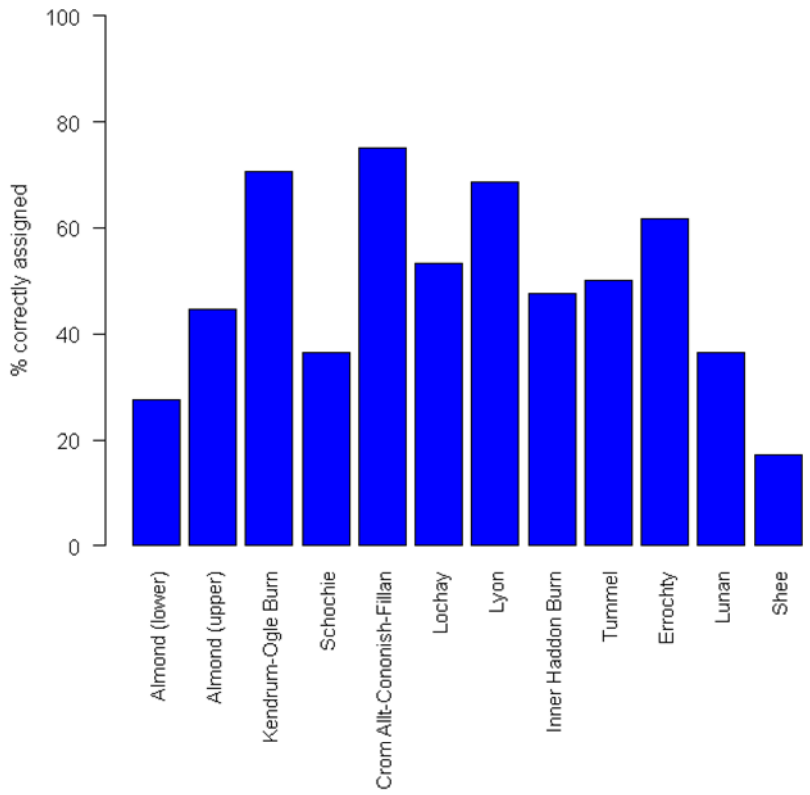


Figure 5. Percentage of fish sampled from each site that correctly assign back to that site, based on a suite of 17 microsatellite markers.

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 Tay). 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. This problem is also most pronounced in cases with few locations sampled and few individuals per location (Anderson 2010; Waples 2010), as was the case for the present analysis. 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; now 10 to 16 as opposed to the original 21 to 32).

Despite only having three sites screened for SNPs, an assignment exercise was nevertheless conducted. Six sets of top-ranked SNPs, based upon the three 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 there is more genetic information available to utilise. All sets of SNPs had 100% assignment accuracy to site for the holdout samples, with the exception of the top 50 ranked SNPs, which had 98.8% correct assignment. However, microsatellite assignment success considering only these three sites was also high, at 92.3%.

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 6A. 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 6B). However, there is considerable variation in the figure among rivers. For the Tay, at present, this figure is ~90%. It should be noted that particular effort is being focused on separating the big east coast rivers, and, as more rivers/sites are included in the baseline, this figure is 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 Tay SNP work necessarily, it does demonstrate the types of applications and analyses that can be achieved by combining individual river system analyses into a wider Scottish context. Given the limited sampling of sites screened for SNPs in the Tay to date, the addition of further sites within the system will help to extend the knowledge of the genetic diversity within the system and aid in assignment both within and between rivers.

(A) (B)

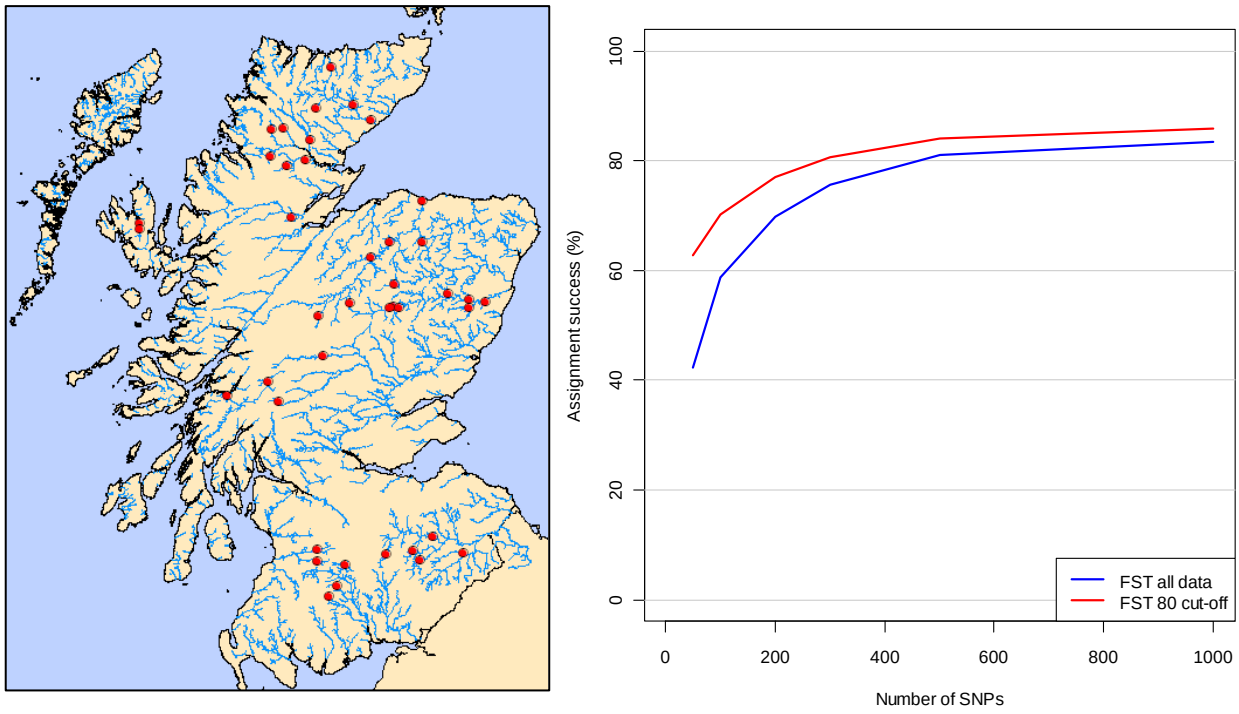


Figure 6. (A) Sites screened for the 5,500 SNP chip to date. (B) Assignment accuracy to river of origin for a preliminary baseline of Scottish rivers with increasing number of SNPs (Figures courtesy of Gilbey & Coulson, 2013).

Discussion

Population structure

The aim of the FASMOP project for the Trust was to identify distinct breeding populations of salmon within the River Tay. The results to date suggest that there are distinct breeding populations. Significant differences among locations were observed with both microsatellites and SNPs, however, as the latter was carried out on only three locations, discussion of the patterns of structuring with SNPs is limited. What is clear from the use of SNPs is a much higher degree of structuring as compared to microsatellites, suggesting that SNPs will be more useful in a catchment-wide survey of genetic variation. This may be most important for sites where the microsatellites revealed lower levels of structure than the present three included here, as even in the full microsatellite analysis, these three sites were among the most differentiated.

One might expect sites that are more isolated in the system, potentially with a smaller number of breeders, to be more likely to be distinct from other, more accessible, locations further downstream. The Tummel, Inner Haddon Burn, Kendrum-Ogle, and

Crom Allt-Cononish-Fillan sites are all located quite far up in the catchment and are all above lochs, factors implicated in promoting and/or maintaining genetic structuring in other systems (see FASMOP reports; www.rafts.org.uk). These sites also showed a reduced level of genetic diversity compared to all other sites, which may be a function of their more remote location or smaller local population sizes. However, none of the current sites showed a significant difference in terms of family effects. An exception to the above pattern was the Lochay site, which is also above a large loch and quite far upstream but did not show reduced genetic diversity. Most remaining sites were significantly different from one another and reflected moderate to strong levels of genetic differentiation throughout the system. The only non-significant differences were those in which sites were located in close proximity (Inner Haddon Burn vs. Tummel, and the Schochie site vs. both the Almond sites), possibly reflecting a greater exchange of individuals among those locations.

Several points need to be considered when interpreting the current pattern of genetic structuring within the system. Several sites were sampled over multiple years or samples consisted of multiple age classes. In these cases, most samples were of limited size (often < 10 individuals) and were therefore combined for analysis. Furthermore, as seen in Figure 1, some sites were composed of many small samples taken over a much larger area. Again, the limited sample sizes required that they be combined prior to analysis. This sampling scheme could therefore consist of some temporal and/or smaller-scale variation that could not be robustly taken into account prior to analysis. However, the assumption that has been made is that these differences will be minor compared to the larger geographical scale upon which sites were considered here. A more local sampling scheme with a sampling design to allow for assessment of the temporal component would help to clarify these considerations.

The degree of differentiation observed within the Tay may indicate that populations may differ in heritable traits and be locally adapted. For example, it has been documented that heritable differences for run-timing exist among different tributaries within the Tay (Stewart et al., 2002). Whether any of the current genetic markers (in particular, the SNPs) reflect such underlying adaptive traits is uncertain at present. Discovery of markers associated with particular adaptive traits may help to further address the degree to which locations represent distinct breeding populations and may offer a very powerful management tool. Overall, the pattern resolved with the microsatellites 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, such as in the present study, then a continued caution with respect to sourcing brood stock would be desirable in respect of stocking programmes. As mentioned above, locations may 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.

Finally, a number of factors underlie population genetic structuring. At least one of these, not addressed here, is the potential impact played by stocking practices. Stocking in many areas has been common for Atlantic salmon both within and between systems. Such practices may influence why certain locations appear very distinct if they were sourced from a different location compared to the surrounding stock components. 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 are essential. Furthermore, including genetic samples from possible donor sources 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.

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 example 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 Figure 5 shows the proportion of individuals from a given site that was assigned back to that site, based on their genetic profile, using microsatellites. If each location exhibits large differences from everywhere else, one would expect the accuracy of assigning individuals to the location from which they were sampled to be high. While the average assignment success with microsatellites was 49%, several sites did indeed show a higher success rate (60-70%). Given that the assignment success is better than

random expectations, this supports the conclusion that there is genetic differentiation among locations, indicative of separate breeding populations. However, depending upon the degree of the genetic differences for a particular site, assignment success can be quite variable.

While SNPs have been shown to improve the level of assignment, compared to microsatellites, in other catchments (see, for example, the Spey FASMOP report; www.rafts.org.uk), this does not always translate to levels that may be of value for particular management applications. Given only three locations were screened for SNPs in the Tay, an assessment of the relative performance of the marker types is difficult, at present. The 100% assignment success with most panels of SNPs evaluated may seem promising, however for the three sites screened (Crom Allt-Cononish-Fillan, Tummel and Lunan), even with microsatellites, the assignment was 92%. If however, one compares the assignment success for these sites with the full microsatellite baseline, encompassing 12 sites, the assignment success is now 75%, 50% and 36%, respectively. As the assignment analysis determines from which of the sites in the baseline an individual is most likely to originate, then, with only three sites, one would expect assignment success to be relatively high to begin with, as there are fewer sites available for individuals to be mis-assigned. As this analysis 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. Therefore, a more accurate appraisal of assignment success with SNPs will be obtained with the addition of further sites throughout the catchment.

Implications for Management / Future Work

The aim of the current analysis was to identify distinct breeding populations of salmon within the Tay. The results show that there are distinct breeding populations within the system. These findings suggest caution should be taken regarding stocking, as underlying heritable, adaptive differences may be present among locations. The moderate to strong genetic structuring observed is likely influenced by the large catchment size, the dendritic nature of the system and the presence of multiple lochs, among other factors. Further geographical coverage of the system, along with increasing

use of SNPs, will offer the ability to generate a clearer picture of population genetic structuring within the Tay and allow for an assessment into the ability to use these genetic differences in further management applications (e.g. genetic assignment). Furthermore, the SNP data presented here has also contributed to a preliminary analysis on a Scotland-wide scale, demonstrating the possibility for accurate genetic assignment back to river of origin and will likely have applications in mixed-stock fisheries and assignment of marine-caught fish. Finally, 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.

Conclusion

This analysis demonstrated moderate to strong population structuring within the Tay using microsatellites. However, with the use of SNPs, stronger signals of genetic differentiation have begun to emerge. These results suggest that there are distinct breeding populations within the Tay. However, the degree of these differences is variable in their ability to allow for levels of assignment to sites. Screening of further sites within the Tay for SNPs (and possibly microsatellites) will help to clarify the extent of within-river genetic stock components and the relationships among these. Furthermore, preliminary work is beginning to resolve accurate assignments to river of origin that 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.

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

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.

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.

Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1332.

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.

Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

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 (ILLUMINA INFO). 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 = 36).

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.

Micro-satellite 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 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.

Clustering of individuals among potential distinct groups was undertaken with STRUCTURE 2.3.3, using the admixture model with correlated alleles (Pritchard et al., 2000). Briefly, this method assumes the number of distinct groups (*K*) in turn to be from 1 to some defined upper limit (i.e. the number of sites sampled). The analysis then determines which *K* is most consistent with the observed data,and assigns each individual to one of the defined groups. Furthermore, prior information on sampling sites was used to initiate the analysis with the LOCPRIOR option available in STRUCTURE 2.3.3 (Hubisz et al. 2009). A burn-in phase of 150,000 iterations was followed by a run phase of 250,000, using a minimum of five independent runs for each number of groups (*K*) being tested. Both the log-likelihood probabilities and the delta *K* method (Evanno et al., 2005) were examined to find the most likely *K*.

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 80 individuals (out of 96 screened for the Spey) for 3832 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	6	7	8	9	10	11	12
1	-	0.019	0.185	0.001	0.184	0.061	0.056	0.200	0.222	0.055	0.016	0.042
2	*0.008*	-	0.178	0.022	0.156	0.073	0.072	0.172	0.176	0.093	0.042	0.070
3	*0.031*	*0.032*	-	0.159	0.180	0.123	0.128	0.191	0.194	0.149	0.201	0.149
4	0.001	0.005	*0.029*	-	0.201	0.077	0.071	0.183	0.194	0.052	0.032	0.050
5	*0.035*	*0.033*	*0.047*	*0.040*	-	0.047	0.080	0.177	0.160	0.149	0.167	0.093
6	*0.012*	*0.015*	*0.029*	*0.015*	*0.012*	-	0.021	0.118	0.136	0.057	0.073	0.027
7	*0.013*	*0.012*	*0.028*	*0.012*	*0.020*	*0.007*	-	0.108	0.130	0.056	0.031	0.030
8	*0.035*	*0.035*	*0.041*	*0.037*	*0.038*	*0.025*	*0.020*	-	0.000	0.131	0.167	0.130
9	*0.037*	*0.035*	*0.042*	*0.039*	*0.037*	*0.026*	*0.023*	0.001	-	0.139	0.157	0.125
10	*0.012*	*0.018*	*0.030*	*0.013*	*0.031*	*0.011*	*0.013*	*0.027*	*0.028*	-	0.032	0.032
11	*0.006*	*0.008*	*0.035*	*0.007*	*0.032*	*0.013*	*0.009*	*0.030*	*0.029*	*0.011*	-	0.037
12	*0.011*	*0.014*	*0.032*	*0.013*	*0.023*	*0.008*	*0.006*	*0.027*	*0.025*	*0.010*	*0.009*	-

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