Marine Scotland Science Report

The Scottish Government

Marine Scotland Science Report 02/12

AN OVERVIEW OF POPULATION GENETIC STRUCTURING IN THE RIVER DON

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)

M W Coulson, L M I Webster, J Urquhart, A Armstrong,

E Cauwelier, L Stradmeyer, J Gilbey, C Sinclair & E Verspoor

marinescotland science

Marine Scotland Science Report 02/12

An Overview of Population Genetic Structuring in the River Don

M W Coulson, L M I Webster, J Urquhart, 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 Don

M W Coulson[1], L M I Webster[1], J Urquhart[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]The River Don Trust, Old Estate Office, Cluny Castle, Sauchen, Aberdeenshire, AB51 7RT

Executive Summary

This report describes how genetic information from juvenile Atlantic salmon sampled from 9 sites within the River Don (Figure 1) have 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.

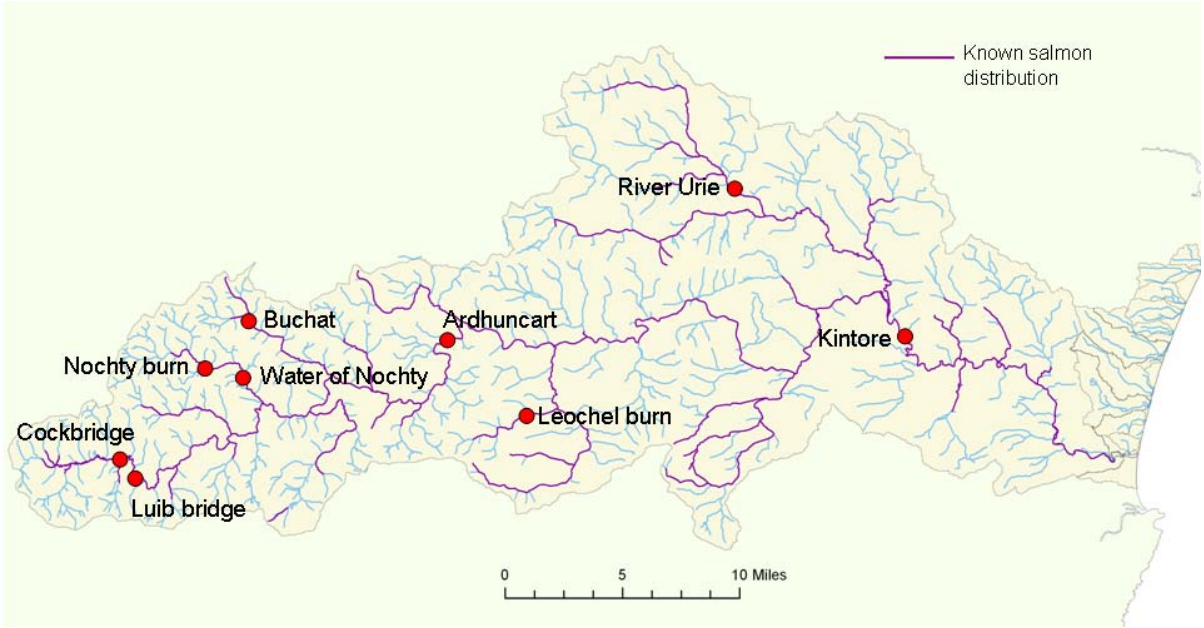

Figure 1. A map of the River Don, with sample sites that are involved in this report indicated in red with associated site names. Both the Cockbridge and Ardhuncart sites were sampled in two years (2008 & 2010).

Summary of findings

The analysis showed that, most sites exhibited weak to no genetic differences from one another with the current set of markers used. Upstream sites generally were the most differentiated, with the exception of the Cockbridge site. The remaining locations showed a mixture of results when compared to each other. For two locations, which were

sampled in different years, there appeared to be stable genetic signatures over time. Overall this suggests a stable, but very weak level of population genetic structuring within the River Don using the current genetic markers.

This weak degree of genetic differentiation observed within the Don is largely reflected by the poor 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 32%, which is higher than one would expect if there was no genetic structure in the data. However, the magnitude of differences observed with the current markers is not large enough to assign fish with higher accuracy

Implications for management

The aim of the current FASMOP project was to identify distinct breeding populations of salmon within the River Don. The results to date suggest that there *may* be distinct breeding populations. However, currently the distinction of these potential populations is limited. While several locations appear to be 'most different' they do not always reflect an obvious geographical pattern.

There are two possible reasons for the observed low levels of genetic structuring seen:

- There is reproductive mixing of individuals between the different parts of the system. This could include possible stocking events in the past.
- The microsatellites in the study do not give the resolution required to adequately describe population structuring within the river.

The current genetic markers show overall weak genetic differentiation. However, this observation cannot be used to rule out the possibility of locally adapted traits being present within the system. This may be further clarified with the development and application of newer, more targeted, genetic markers. To determine if it is possible to improve assignments and gain better distinction for potential breeding populations, larger sample sizes and/or newer genetic markers will be required and possibly a more complete baseline of potential populations sampled.

Introduction

Atlantic salmon (*Salmo salar* L.) are 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, it is expected that Atlantic salmon will 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 (e.g. King et al. 2001; Verspoor et al. 2005) or documenting the expected structuring existing between different river systems (e.g. 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). However, not as much is known about structuring within river systems and few examples exist for Scottish rivers (but see for example, Verspoor et al., 1991; Jordan et al., 2005).

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, since intermixing of these populations runs a risk of unknown magnitude, 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). Recent 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).

The suite of genetic markers used in the current survey are 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.

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 populations occur. 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. It set out to undertake a Scotland-wide survey of genetic structuring within all Scotland’s major salmon-producing rivers. 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.

Summary of Methods

Juvenile salmon from various locations within the River Don 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 9 sites that have been included in the genetic analysis for the Trust. Two of these sites (Cockbridge & Ardhuncart) were each sampled in two different years, thereby giving 11 samples across the 9 sites. Samples generally consisted of fry and mostly parr (n= 11-52, depending on site) and for each individual, data from 17 genetic markers (microsatellites) were collected. The results from the microsatellite marker SsaF43 allowed us to identify any

trout or trout/salmon hybrids that may be present among samples. These individuals were then removed prior to analysis.

It is possible that samples 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 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 precocious parr. Initial sample sizes as well as sample sizes after full-siblings were removed are presented in Table 1.

When sample sites included fish collected in different years (Cockbridge & Ardhuncart), each of these sub-samples was 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. Among the remaining samples multiple age classes were present at several sites, however the numbers of each age class after family effects had been removed were not large enough within a site to warrant testing for differences among age classes. As such, all age classes within a site were grouped for further analysis. This resulted in 9 samples for subsequent analyses.

Data were then analysed 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

Broadly speaking, most sites exhibited weak or no significant genetic differences from one another, indicating low levels of genetic structuring among these sites with the current set of markers. The interpretation of the pattern and degree of differences in terms of the relationships among populations, combined with the known history and geographical proximity of sites can be useful to inform fisheries management decisions. Here we discuss the results of the FASMOP project summarizing the main genetic findings in terms of population genetic structuring within the River Don.

Family effects

A total of 440 juvenile salmon from the River Don were involved in the genetic analysis. All sites were examined for family effects with a modest number of samples being removed from several sites due to full-sibling relationships (Table 1). The level of family effects differed between samples with the largest family present in the individual samples ranging from 1 to 10 full-siblings and sample sizes subsequently being reduced by 0-55%. Family effects were controlled for at each site before all further analyses. There were 12 trout identified at a single site (Cockbridge 2008) and a single salmon/trout hybrid sample identified at a two location(s) (Kintore & Cockbridge (2008)).

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 analysed (sibs removed)	Number of breeders contributing to sample	Largest single family	Year sampled
Kintore[1]	KIN	49	40	56	3	2009
Urie	URI	50	47	61	3	2009
Leochel	LEO	48	42	56	4	2009
Ardhuncart†	ARD	49	49	65	1	2008
Ardhuncart†	ARD	43	34	52	7	2010
Buchat	BUC	52	31	41	8	2009
Nochty burn	NOB	31	23	33	4	2010
Water of Nochty	NOW	47	21	23	10	2009
Cockbridge[2]‡	COC	36	31	46	3	2008
Cockbridge‡	COC	11	11	45*	1	2010
Luib bridge	LUB	24	22	45*	2	2010

1. One sample from this site was identified as a trout/salmon hybrid.
2. One sample from this site was identified as a trout/salmon hybrid and 12 were identified as trout.

† These two time points were combined as there was no significant difference between them.

‡ These two time points were combined as there was no significant difference between them.

* These two sites were run for COLONY together due to small sample sizes (for Cockbridge) and their close proximity, having been collected in the same year.

Population structuring

Among sites there is a limited range of genetic differentiation, with 47% (17 out of 36) of the pairwise comparisons being significantly different (Appendix 2). Among these comparisons, the Luib bridge and Water of Nochty sites were the only sites that differed from all others (Appendix 2). Most sites show a close genetic relationship. 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 which are closer together on the plot have a more similar genetic makeup while points further apart are more genetically discrete.

The largest differences are seen to the Water of Nochty and Luib bridge sites (as mentioned above) as well as the Nochty Burn, Kintore and Buchat sites. All of these are plotted apart from the central cluster of the remaining sites (Figure 2) but are not always significant different from all other sites (Appendix 2). The central cluster of sites are generally not significantly different from one another, however they are also located throughout the system, rather than being more geographically close to one another. For instance this central group includes both the River Urie (a tributary off the lower mainstem) as well as Cockbridge (uppermost site).

Two sites were sampled in different years; Cockbridge & Ardhuncart were each sampled in 2008 and 2010. The CHIFISH analysis showed no significant differences between the two sample years at either of these sites. Therefore samples from different years within each of these sites were combined for all other analyses. This suggests relatively stable temporal genetic signatures at least for these two locations for which this could be tested.

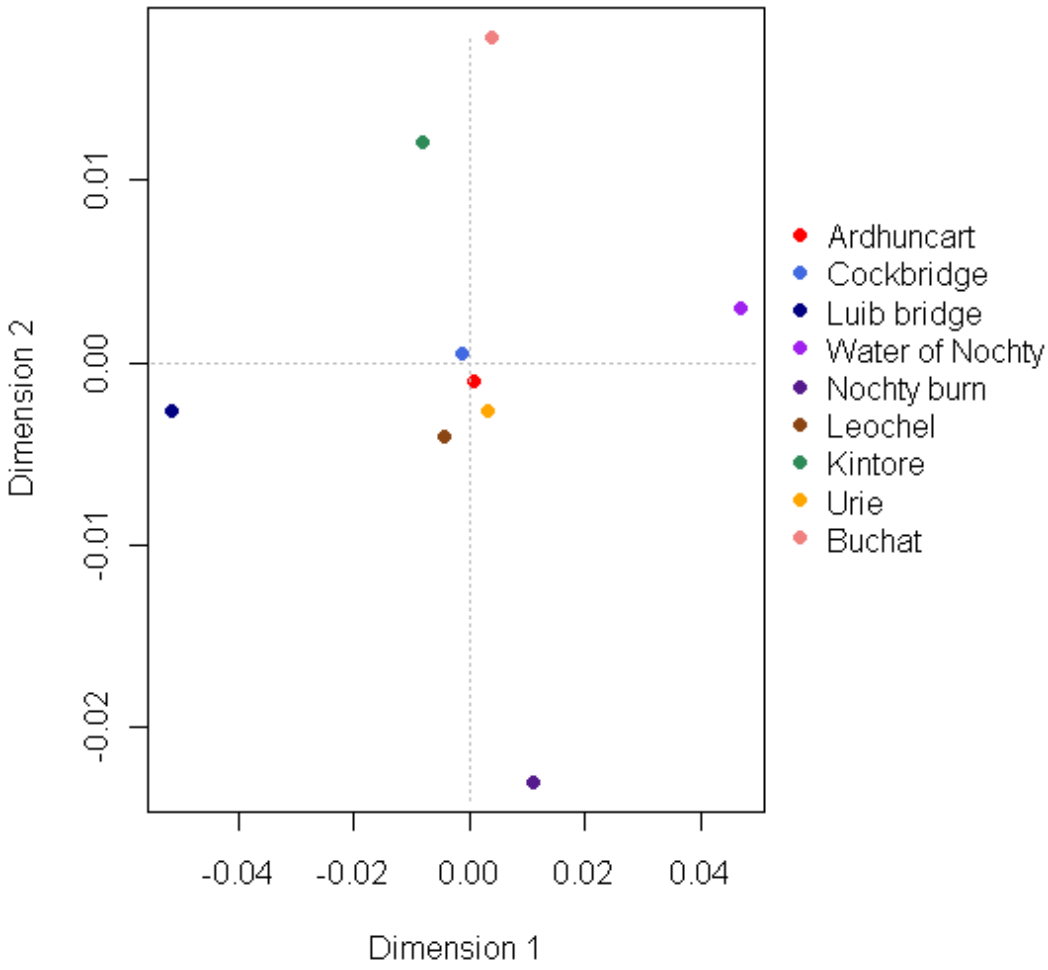

Figure 2

Multi-dimensional scaling (MDS) plot of genetic relationships among all sites based on pairwise estimates of genetic differentiation (Jost's D; see the appendix for details). Points which are closer together on the plot have a more similar genetic makeup while points further apart are more genetically discrete.

Clustering of individuals based on genetic relatedness was also carried out as well as that for samples. This analysis aims to determine the most likely number of genetic groupings of individuals, allowing for groupings other than the pre-defined sampling sites. Subsequently the analysis then calculates the membership for each individual among each of the defined groups. For the River Don samples, this analysis determined the most likely number of groups to be one.

Genetic assignment of individuals

The assignment analysis shows how useful this baseline genetic information is to identify which of the sampled sites a fish of unknown origin is from (Figure 3). Each individual fish is taken in turn and it is assessed from which of the sampling locations provided in the baseline, that individual is most likely to have originated. Assignment of fish back to their specific site of collection was, on average, correct 32 % of the time. This average is greater than would be expected if assignments were purely random (9 sites, random = ~11%).

It may be possible to improve 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). However, for the Don, a cut-off does not appear to dramatically improve assignments. For example, if we assign only fish that have a minimum of 70% assignment probability, overall correct assignment is still only ~36%. Individual-site assignment increases at most ~14%, however this is only for a single location (Kintore) with the remaining locations increasing by 1-7%. Applying such a cut-off comes at a potential coast as not all fish in the baseline will be assigned. However the above example for the Don (70% cutoff) still resulted in over 77% of fish being assigned. This suggests that many fish are being assigned with high probability to sites other than those from which they were sampled.

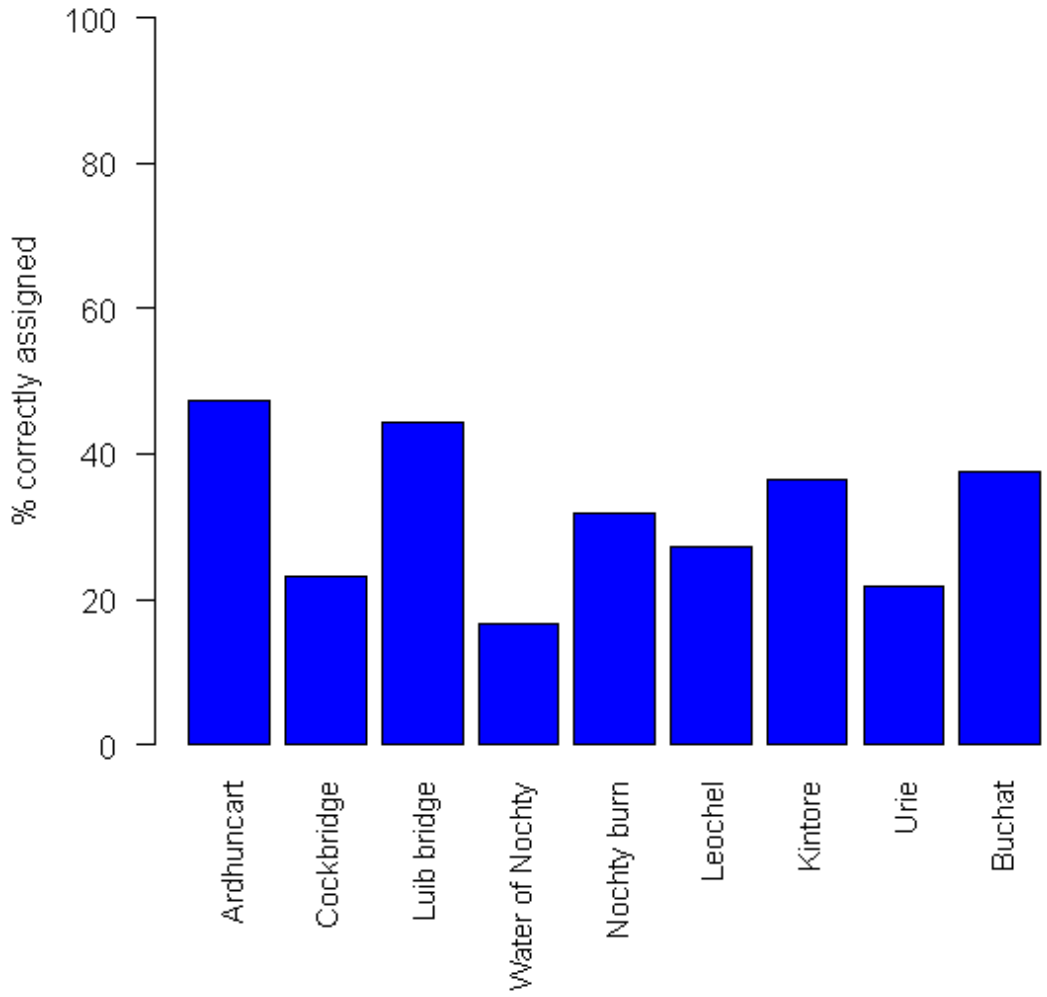

Figure 3
Percentage of fish sampled from each site that correctly assign back to that site.

Discussion

Population structure

The aim of the FASMOP project for the Trust was to identify distinct breeding populations of salmon within the River Don. The results to date suggest that there *may be* distinct breeding populations within the Don. However, using our current set of genetic markers the distinction of these potential breeding populations is not easily defined. The magnitude of differences observed among sites is low, and accounts for the low accuracy of assignment of individuals to the location from which they were originally sampled.

The sites that separated out most on the MDS plot (Luib bridge, Nochty Burn, Water of Nochty, Kintore and Buchat) are mostly upstream sites, with the exception of Kintore which is on the mainstem and the furthest downstream of all sample sites. However, the most upstream site (Cockbridge) is located in the central cluster of sites, similar to several sites further downstream (Ardhuncart, Leochel burn & River Urie). Therefore, while there appears to be a tendency of the uppermost sites to be the most genetically differentiated, there are exceptions to this pattern.

Several of the upstream sites were most affected by the presence of full-sibs (Table 1). The Nochty burn, water of Nochty and Buchat sites had their sample sizes reduced by ~26-55% after full-sibs were removed. Part of the observed differentiation of these sites may be influenced by the presence of closely related individuals. Despite having removed all but one individual of a full-sib family, there could still remain other family-level relationships which are less certain in this analysis and more difficult to account for, such as half-siblings, cousins, etc. Given these three samples had the fewest numbers of breeders, the genetic signature at these sites may not be best represented in the presence of large family effects. On the other hand, small numbers of breeders (and associated family effects) may be typical of these sites and therefore one would expect the genetic signature at these sites to be more highly differentiated. The Nochty burn, for instance, is a spate burn with a lot of wash out (J. Urquhart, personal communication), which may affect the number of adults breeding there. It would be of interest to assess the temporal stability of the genetic signatures at these locations.

While the clustering program showed the most likely number of groups to be one, this does not necessarily mean that there are not significant genetic differences. Using this approach the weak differences observed are more difficult to tease apart and the distinction for splitting individuals into more than one group is less obvious. While a lack of differentiation 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 may be due, at least in part, to the current set of “neutral” genetic markers, which may not resolve weak but still biologically significant reproductive isolation where differentiation may only be dramatic in respect of genetic variation associated with adaptive differentiation. For instance, adaptive differences may be present (e.g. for run timing behaviour) which our neutral genetic markers could not detect. Therefore, more and/or other types of markers, some of which may be associated with adaptive traits, may help to further address the degree to which these locations represent distinct breeding populations.

Two locations (Cockbridge & Ardhuncart) allowed for an assessment of the temporal stability of the genetic signature at these sites. Both sites were sampled in two different years and in both cases there were no differences in genetic signature between years at

a given site. At least for these two sites, there appears to be relative temporal stability suggesting that there is a stronger signal of differentiation among sites (albeit still quite weak) compared to a given site over time. Whether this pattern holds throughout the system (or over longer periods of time) needs to be determined. However, this type of pattern supports the idea of weak meta-population structuring within the system, whereby spatially separated populations are connected by different degrees of interactions or exchange of individuals over time.

If 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. However, even with weak to little observed differentiation, the same caution should be exercised. As mentioned above, a lack of genetic differences with a given set of markers does not necessarily imply a single breeding population. Locations may still differ with respect to adaptive traits and until such issues can be addressed, then 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. This could be applied, for instance, in assigning rod caught adults to their particular stock component to determine whether salmon returning to a river at different time points are destined for different parts of the catchment. Of course, this relies upon well defined structuring between these components and perhaps even genetic markers which may be associated with that particular trait. Genetic assignment allows one to calculate the probability that a given fish originated from a particular location. Then the location with the highest probability is taken as the site from where that individual originated. This is done for each individual and Figure 3 shows the proportion of individuals from a given site, which was assigned back to that site based on their genetic profile. 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 (e.g. 90-100%). The average value of correct assignment to site is 32% (Figure 3), which is somewhat higher than one would expect if there was no genetic structure in the data. This supports the conclusion that there may be genetic differentiation among some locations, indicative of separate breeding populations, but the data do not at present have the power to assign fish of unknown origin (e.g. rod caught adults) to their location with high accuracy.

In order to improve assignments and gain better distinction for potential breeding populations, larger sample sizes and/or additional genetic markers may be required as well as a more complete baseline of potential populations sampled. The assignment of individuals in the analysis was only to sites represented in the baseline. If the 'true' site has not been sampled, fish from these missing sites will be forced to be assigned incorrectly to some site that is in the baseline. For the most accurate and complete picture of assignments within a system, detailed knowledge of *all* breeding groups defined by the set of markers used is required, so these would need to be resolved first. 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, and there is only weak population structuring, then the genetic make-up of these sites will be relatively similar and prevent assignment to defined groups with high accuracy.

Future work

There is some suggestion of genetic structuring within the Don but the evidence from the current marker set is weak and prevents robust conclusions. Such an outcome is not unique to the Don and has been observed in several other systems throughout Scotland. However, before it can be concluded that there are little to no genetic differences within these systems, a much more detailed survey is warranted. Currently, the development and application of a different class of genetic marker (**S**ingle **N**ucleotide **P**olymorphisms, or SNPs) is underway to address the resolution of population structuring in more detail and provide a more robust assessment. This approach offers at least two distinct advantages over the current suite of markers in that (1) the number of markers screened for SNPs is much larger than that for microsatellites (100s - 1000s vs. 10s, respectively) and (2) that while microsatellites are selectively "neutral", SNP markers should be associated with both "neutral" as well as actual traits, the latter of which some may be adaptive. The use of SNPs, either as an alternative to or in conjunction with microsatellites, has been shown to be promising for resolving different stock components with respect to fisheries management for various salmonid species (e.g. Narum et al. 2008, Glover et al. 2010, Beacham et al. 2010). Trying to target the underlying genetic differences that are associated with known biological (e.g. run-timing) or habitat (e.g. pH, elevation) differences will help to shed light on different stock components. For instance, finding a genetic marker associated with run-timing would allow for direct application toward the identification of spring vs. late-running stock components. This would allow for a more diagnostic application rather than using a set of random, 'neutral' genetic markers.

A number of factors may 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, is 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 levels of differentiation..

Summary

This analysis found overall weak levels of population structuring within the River Don and suggests that there may be distinct breeding populations. The degree of these differences, as revealed by the genetic markers so far employed, is not sufficient to allow for robust application to management at present. For several of the sites where there was temporal sampling, the analysis suggests that there is stability of the genetic make-up at these sites as they were less differentiated across years than they were differentiated from other sites. Clearly more work is needed to clarify the extent of genetic structuring within Don. This work should focus on the use of newer genetic tools and take a more targeted approach, to make the greatest contribution to our overall understanding of the underlying salmon population structure and to the efficient management and conservation of this valuable resource.

References

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.

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.

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.

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.

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.

Jordan, W.C., Cross, T.F., Crozier, W.W., Ferguson, A., Galvin, P., Hurrell, R.H., McGinnity, P., Martin, S.A.M., Moffett, I.J.J., Price, D.J., Youngson, A.F. and Verspoor, E. 2005. Allozyme variation in Atlantic salmon from the British Isles: associations with geography and the environment. *Journal of Fish Biology*, **67**, 146-168.

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.

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.

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., Fraser, N.H.C. and Youngson, A.F. 1991. Protein polymorphism in Atlantic salmon within a Scottish river: evidence for selection and estimates of gene flow between tributaries. *Aquaculture*, **98**, 217-230.

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.

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

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.

Data Analysis

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.

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

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.

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.

Appendix 2

Pairwise estimates of genetic differentiation among 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.

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