Managing Interactions Aquaculture Project 2011/12

Report on Genetic Tool Development for Distinguishing Farmed *vs*. Wild Fish in Scotland

Mark Coulson

February 2013

RAFTS
Rivers And Fisheries Trusts Of Scotland

Contents Page

Executive Summary

The purpose of this report is to present the results of the development and application of genetic tools to support the identification of wild and aquaculture origin fish from west coast catchments in Scotland.

Four main objectives have been outlined and will be discussed in turn. These are as follows:

1. Development of a cost effective and Scotland specific tool to allow wild and aquaculture strains to be identified from tissue sample analysis. The results here demonstrate the potential to identify aquaculture strains of Norwegian origin from Scottish fish, however, as a reduced set of markers was utilized, there were mixed results for distinguishing between wild Norwegian and farmed Norwegian origin. Furthermore, the current maker set is unable to identify Scottish aquaculture fish of Scottish origin from wild Scottish fish.
2. Develop an annual sampling network across the west coast that allows these catchments to be sampled systematically to assess the extent or presence of genetic materials of aquaculture origin. Each of the participating fisheries trusts collected samples from sites throughout their region for screening with the current SNP marker set.
3. Support fishery trusts in the gathering of samples from an agreed network across west coast catchments.
4. Provide resources for samples gathered to be analysed, and results reported for prospective application in policy and management practice. Sites collected by the trusts were screened an analysed to assess whether there is evidence of Norwegian origin among Scottish samples. Results to date demonstrate varying levels of the presence of Norwegian genetic signatures from most of the sites screened and confirmed several cases of putative direct escapees sampled in the wild.

The genetic markers developed in Norway, and implemented here, provide strong resolution between Scottish fish and Norwegian farmed strains. This offers a powerful tool for the identification of fish as either a Scottish fish or Norwegian origin fish. The identification of a single sample of a Scottish aquaculture strain (Loch Duart Ltd) was not possible at present with the current marker set, and was indistinguishable from wild Scottish samples. Furthermore, these markers offer the possibility of being used in identifying samples where Norwegian origin genetic material is still present after initial introgression in the past. To this end, further sampling of both wild and a robust farmed strain baseline in Scotland, is needed. Additionally, the particular panel of genetic markers can be revised as more baseline data is accumulated. For instance, additional markers may be of use in separating Scottish aquaculture strains from wild Scottish fish.

1. Project Background

In 2011, the Rivers and Fisheries Trusts of Scotland (RAFTS) and its member fishery trusts and partner district salmon fishery boards on the west coast of Scotland began a programme of work funded by the Scottish Government associated with the interactions between aquaculture and wild fish populations. The Managing Interactions Aquaculture Project is designed to support the better coordination and management of wild fisheries and stocks with the aquaculture industry. Underpinning this programme of work were the wild fish priorities of protecting sensitive and high value fresh water sites, improving practice and management at existing aquaculture sites and finally informing decisions on the location and biomass production at aquaculture sites both current and proposed. To achieve these strategic objectives three projects were identified as key priorities and work streams within the overall Project.

These were:

- Strategic programme of post smolt sweep netting and analysis;
- Programme of genetic sampling and analysis; and

- Locational guidance and zones of sensitivity analysis.

The three Managing Interaction Aquaculture projects are overseen by a Steering Group, chaired by RAFTS, which includes representatives from a range of west coast fishery trusts and boards, Marine Scotland Science and Marine Scotland Policy.

The participating fishery trusts and boards are:

- Argyll Fisheries Trust
- Argyll District Salmon Fishery Board
- Lochaber Fisheries Trust
- Wester Ross Fisheries Trust
- Wester Ross District Salmon Fishery Board
- Skye Fisheries Trust
- Skye District Salmon Fisheries Board
- West Sutherland Fisheries Trust
- Outer Hebrides Fisheries Trust
- Western Isles Salmon Fisheries Board

This paper will discuss further the programme of genetic sampling and analysis which was organised to develop and explore the applicability of genetic tools in the aspect of distinguishing farmed *vs*. wild fish in Scotland and assess their utility for identifying individuals of mixed farm and wild ancestry. Further details on the other two Managing Interactions projects are available on the RAFTS website (www.rafts.org.uk) and are reported separately.

2. Introduction

The Atlantic salmon genome consists of approximately 6 billion DNA base pairs, which is about 2x the size of the human genome (Moran et al. 2007). Differences that occur among these base pairs allow for the identification of

different individuals as well as populations or ‘stocks’. Indeed such differences have been used in the identification of genetically differentiated stocks in different regions and rivers (Garant et al. 2000; King et al. 2001; Landry & Bernatchez 2001; Verspoor et al. 2005; Dillane et al. 2007; Vaha et al. 2007; Dionne et al. 2008), as well as the reconstruction of parent-offspring relationships in supportive breeding programmes (Villanueva et al. 2002; Herbinger et al. 2006; Horreo et al. 2012). These differences evolve either as random processes among groups of individuals, which, are to a greater or lesser degree, reproductively isolated, or as a result of direct selective processes acting upon characteristics that differentially affect individual survival and reproduction.

One area where such differences are apparent is in the distinction between wild fish and fish from domesticated, aquaculture strains (Youngson et al. 1991; Skaala et al. 2004, 2005; Karlsson et al. 2011; Vasemagi et al. 2012). Selective processes are involved in the domestication process that may differ from those in the wild (either intentionally or unintentionally). Furthermore, genetic drift (random differences) occurs not only as a result of domestication, but also within different cohorts or even family groups of individual aquaculture strains. Such domestication effects can lead to differences in the type and frequency of genetic variants within the aquaculture strains and as such potentially allow them to be genetically differentiated from their wild originator stocks (Skaala et al. 2004; Glover et al. 2010; Vasemagi et al. 2012). These differences have been shown to potentially occur within a very small number of generations of domestication (indeed has been seen within a single generation, Christie et al 2012). Such differences have been used to both identify farm of origin of aquaculture escapes (Glover et al. 2008, 2009) as well as assess the potential and degree of interbreeding between aquaculture escapes and wild fish (Clifford et al. 1998a,b; Bourret et al. 2011; Besnier et al. 2011; Glover et al. 2012).

Single nucleotide polymorphisms (SNPs) are a class of genetic markers that differ by a single base change at a given location in the genome. Recently, a set of 60 SNPs has been identified that distinguish between Norwegian wild fish versus Norwegian farmed strains (Karlsson et al. 2011), with high accuracy. These SNPs differ in the frequencies of the genetic variants rather than being diagnostic for 'farmed' or 'wild' origins. Therefore across all 60 SNPs, a probability is associated with any individual as coming from either of these sources depending on the variants it possesses and the frequency of these in the different potential originator strains. Karlsson et al. (2011) demonstrated, with individuals of known source, a high accuracy of these markers to correctly identify individuals of either purebred Norwegian wild or purebred Norwegian farm origin.

The prevalence of Norwegian farmed strains in the Scottish aquaculture industry allows for the development and application of these markers in the Scottish context. Given the ability of these markers to distinguish Norwegian farmed strains from wild Norwegian fish, such markers might confidently be expected to distinguish more readily between Norwegian farmed and wild Scottish fish, as greater genetic differentiation occurs between these regions (Gilbey et al. in preparation) and so wild Scottish fish might be expected to be more differentiated from the Norwegian farmed strains than wild Norwegian fish.

The purpose of this study was two-fold. Firstly, to confirm as expected, that the set of farm-wild markers developed by Karlsson et al. (2011) would allow for differentiation between Norwegian strains of farmed fish versus wild Scottish fish (see Objective 1 below). The second aim was to screen a number of fish from the west coast of Scotland to distinguish between wild and farmed fish and assess the potential of further distinguishing fish with mixed ancestry (i.e. introgression) (Objective 4 below).

3. Summary of Methods

DNA was extracted and quantified for all samples prior to SNP processing. Samples that met quality and quantity controls were subsequently sent to CIGENE (Norway) where they were assayed for either a V2 Illumina panel of 5,500 SNPs (see Objective 1) or the set of 60 farm-wild SNPs previously identified (Karlsson et al. 2011) (see Objective 4).

The raw data was subsequently returned to RAFTS staff for analysis and interpretation. SNPs common to both the V2 and farm-wild panels were extracted from the database and included in the analysis for all individuals. As further quality control, individuals that failed at more than 10% of the SNPs were excluded for analysis. Analysis consisted of two parts: 1) a preliminary exploration of the separation of the 3 defined groups (Norwegian wild, Norwegian farmed & Scottish) and 2) an individual level analysis to explore the potential of identifying fish of mixed ancestry (i.e. hybrids). Initially, a discriminant analysis of principal components (DAPC) was conducted on the raw genotype frequencies from the three baselines using the program adegenet (Jombart 2008). Groups to which individuals were assigned based on DAPC were compared to their known original baseline group.

Subsequently, individual-level analysis was conducted using the program STRUCTURE (Pritchard et al. 2000). This program uses the raw genotype data initially without considering to which group an individual belongs. The analysis determines the number of distinct clusters or groups of individuals in the dataset by evaluating the fit of the data to a particular model. To do this, the program evaluates a range of possible groups from a single group to some upper defined limit of the number of groups, set by the user (in this case 10 groups). By comparing the fit of each scenario (1 to 10 groups), the ‘most-likely’ number of groups is returned. Additionally, for each fish, in each of these scenarios, a group membership coefficient (between 0 and 1) is calculated for it belonging to each of

the groups. Values close to 1 indicate 'pure' individuals from a particular group whereas intermediate values suggest an intermediate genetic make-up for that individual (i.e. a hybrid between groups) and allows for an estimate as to whether individuals are purely from a single group or represent individuals with mixed ancestry. As the approach is a probabilistic one, ten replicate runs of the STRUCTURE analysis were conducted and results represent the consensus among these runs. This analysis was first run on the baseline data (Norwegian farmed, Norwegian wild, and Scottish wild) to establish the group identifications. Subsequently, the analysis was re-run with the west coast samples included as 'test' samples against the baseline data identified from the previous run. The first run, on baseline samples only, consisted of two main groups: (1) Norwegian (farm + wild) and (2) Scottish. Looking at the result for three putative groups showed two possible solutions among the 10 runs. In 60% of the runs the wild Norwegian samples grouped with the Norwegian farmed strains, while the other 40% of the runs, they formed a separate, third group.

Several key points need to be made here. Firstly, the inconsistency of this particular analysis to separate the wild Norwegian vs. farmed Norwegian is likely impacted by the fact that the current analysis uses only ~60% (35 of 60) of the SNPs identified by Karlsson et al. (2011), and therefore would be expected to suffer from reduced power. Secondly, both the Scottish baseline and Norwegian farm baseline contain 100s-1000s of individuals whereas the Norwegian wild samples total 75 individuals. Such skewed differences in sample sizes have been shown to affect the clustering ability of this program (Kalinowski 2011). Therefore it would be desirable to obtain a larger Norwegian wild baseline. Thirdly, these markers have already been shown to distinguish between Norwegian wild and Norwegian farm fish (Karlsson et al. 2011) when the full set of 60 SNPs is available along with representative baselines for each. Finally, as the main focus is on the detection of introgression between Norwegian aquaculture strains and Scottish fish, these two baselines are of most interest. For these reasons, the Norwegian wild baseline was omitted from the subsequent introgression analysis.

The membership coefficients for the Norwegian farm samples and known Scottish wild samples were used to apply cut-off values for distinguishing between ‘pure’ and ‘hybrid’ individuals (see result below). Given that the dataset does not contain individuals that are known to be hybrids, this scenario was simulated using the Norwegian farmed and Scottish baseline data. The program Hybridlab (Nielsen et al. 2006) was used to simulate first-generation hybrid individuals by using the estimated genetic variant frequencies in the baseline populations. To this end, nine crosses were simulated, each cross using a different Norwegian farm strain and a different Scottish east-coast wild sample. These crosses employed a different farm strain to encompass the variability observed among the strains. The Scottish samples for the crosses were selected at random. These individuals were then included along with the west coast test samples in the STRUCTURE analysis of introgression. Given the outcome mentioned above relating to the separation of Norwegian farmed vs. Norwegian wild baselines, the interpretation of the results is discussed in terms of ‘Scottish’ vs. ‘Norwegian’ origin genetic signatures rather than Norwegian ‘farm’ vs. ‘wild’, at present.

4. Objectives and Results

4.1. Development of a cost effective and Scotland specific tool to allow wild and aquaculture strains to be identified from tissue sample analysis.

To date, a set of 60 SNP markers has been developed by Karlsson et al. (2011) in Norway [Norwegian Institute for Nature Research (NINA) & Centre for Integrative Genomics (CIGENE)] to distinguish between wild and farmed Norwegian salmon. The aim of the current project was to build upon and develop these markers in a Scottish context. This involved the determination whether the markers would be largely applicable in Scotland followed by subsequent sample screening to distinguish between fish of wild and farmed origin. Given the extent

of Norwegian strains of salmon used by the Scottish aquaculture industry, the first step was to verify, as expected, that such markers distinguish Scottish versus Norwegian fish.

Given the on-going development of SNP markers as part of FASMOP and various Marine Scotland Science (MSS) internal projects, a baseline of Scottish samples had been screened for a larger panel of 5,500 SNPs on a V2 Illumina Array at CIGENE. This larger panel includes a subset of the 60 farm-wild SNPs identified by Karlsson et al. (2011). However, a noticeable gap in the geographical coverage of Scotland for these markers was identified along the west coast. To this end, the first phase of this project involved the processing of 2-3 sites from each of 4 west coast Scottish rivers (Snizort, Carnoch, Moidart, Ghriomarstaidh and the Gruinard) for this SNP panel. These sites are identified in Figure 1 along with the wider geographical coverage mentioned above. This allowed for a more comprehensive baseline of the variability and applicability of these markers for distinguishing among Scottish versus Norwegian fish.

Data from the sites illustrated in Figure 1 were combined with data from three Norwegian rivers (Gaula, Laerdalselva, & Numedalslagen), which were screened with the 5,500 SNP panel. These sites are used to represent a wild Norwegian baseline. Additionally, the genetic profiles from 756 individuals representing 12 samples of Norwegian farmed fish [Aqua Gen, SalmoBreed and Marine Harvest (Mowi strain)] were provided by Sten Karlsson (NINA, Trondheim, Norway) for comparison against the Scottish baseline. Two samples from Scottish fish farms were also included, and thought to be predominantly of Norwegian origin (referred to as Scottish ‘Norwegian’ farm in Figure 2). A total of 49 of the 60 farm-wild SNPs identified by Karlsson et al. (2011) were in common across these samples and so it should be noted that the power of the analysis is likely reduced due to it being performed using 18% less markers than the full set.

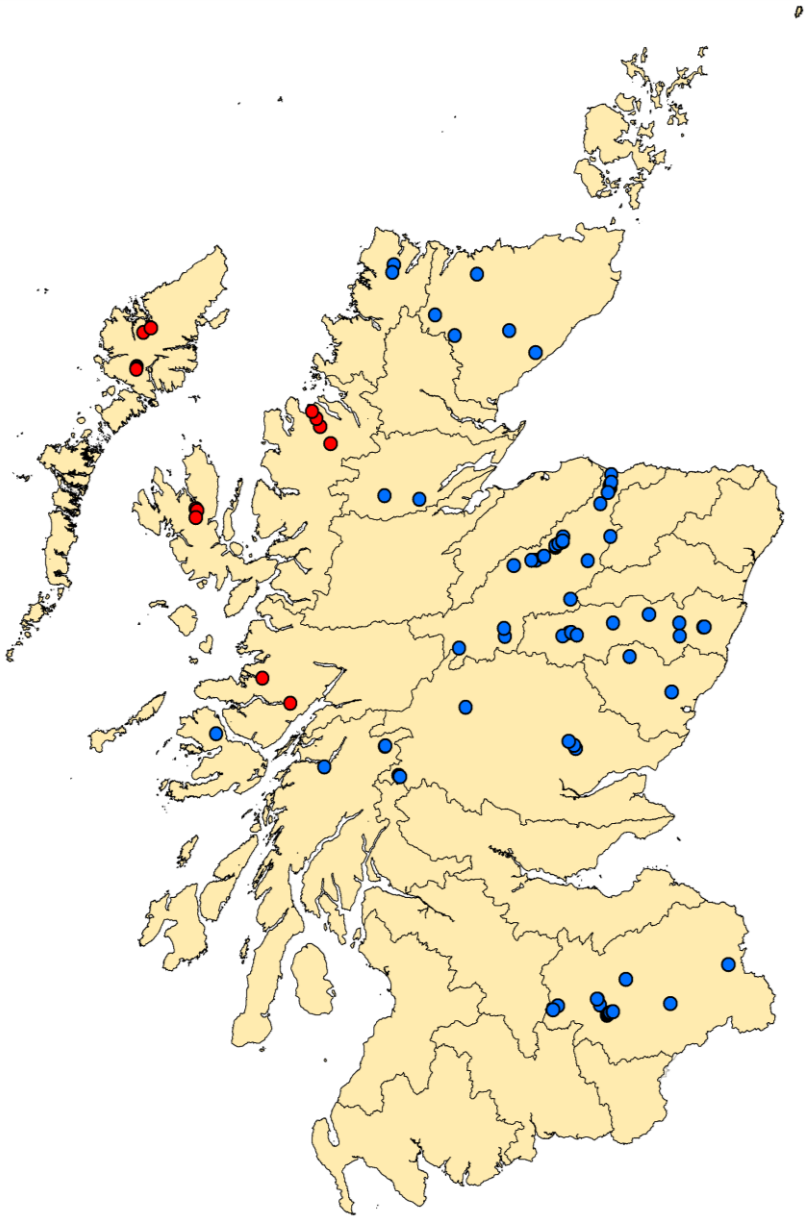

Figure 1. Location of sites screened for the V2 Illumina SNP panel (5,500 genetic markers). Sites in blue represent those screened as part of FASMOP and/or internal MSS projects. Sites in red are those screened as part of the Managing Interactions project to extend the geographical coverage of this panel.

A principal component analysis was conducted on the samples and results are displayed in Figure 2. Each point represents an individual and points are colour-coded by group. The three baseline groups can clearly be seen to be separated: Norwegian farmed strains (including the two samples taken from Scottish fish farms), the three Norwegian rivers representing wild Norwegian fish, and the Scottish samples (which include the Loch Duart Scottish aquaculture strain). Figure 3 shows the correct assignment of individuals from the different baselines to their correct cluster. As can be seen most individuals are correctly assigned with very few cases of mis-assignments. These results confirm expectations that the markers are able to distinguish Scottish fish from Norwegian fish of either farmed or wild ancestry. However further analysis (see Methods and Discussion sections for further explanation), was unable to consistently distinguish between Norwegian farmed and Norwegian wild origin. Furthermore, at present it is not possible to distinguish between wild Scottish fish and a Scottish aquaculture strain. However, as the Loch Duart samples were screened at this larger panel of SNPs, future work could determine if other SNPs may resolve Scottish wild vs. Scottish farmed samples. However, this would best be accomplished using more than one sample representing farmed strains of Scottish origin.

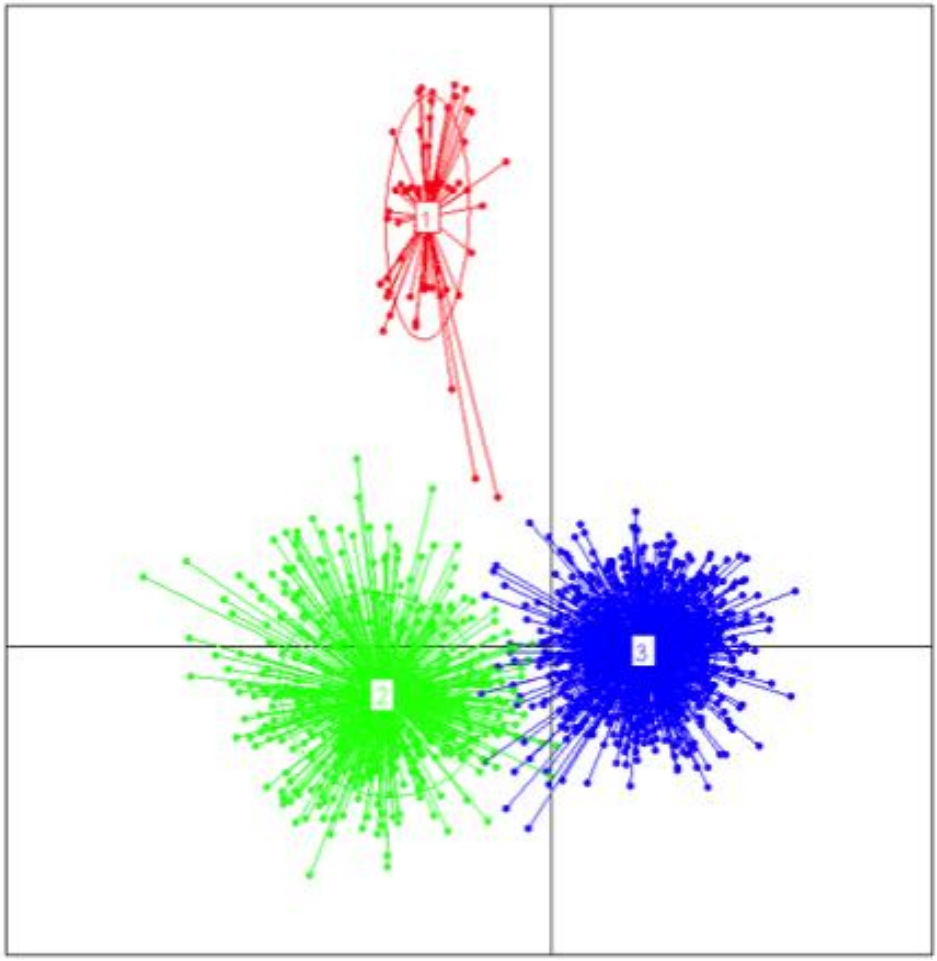

Figure 2. Principal component analysis based on 49 farm-wild SNPs from Karlsson et al. (2011). Group 1 (red) = Norwegian wild, Group 2 (Green) = Norwegian farmed and Group 3 (blue) = Scottish. Points represent individual fish. Note: the Scottish samples (blue) also include the Scottish aquaculture strain supplied by Loch Duart Ltd.

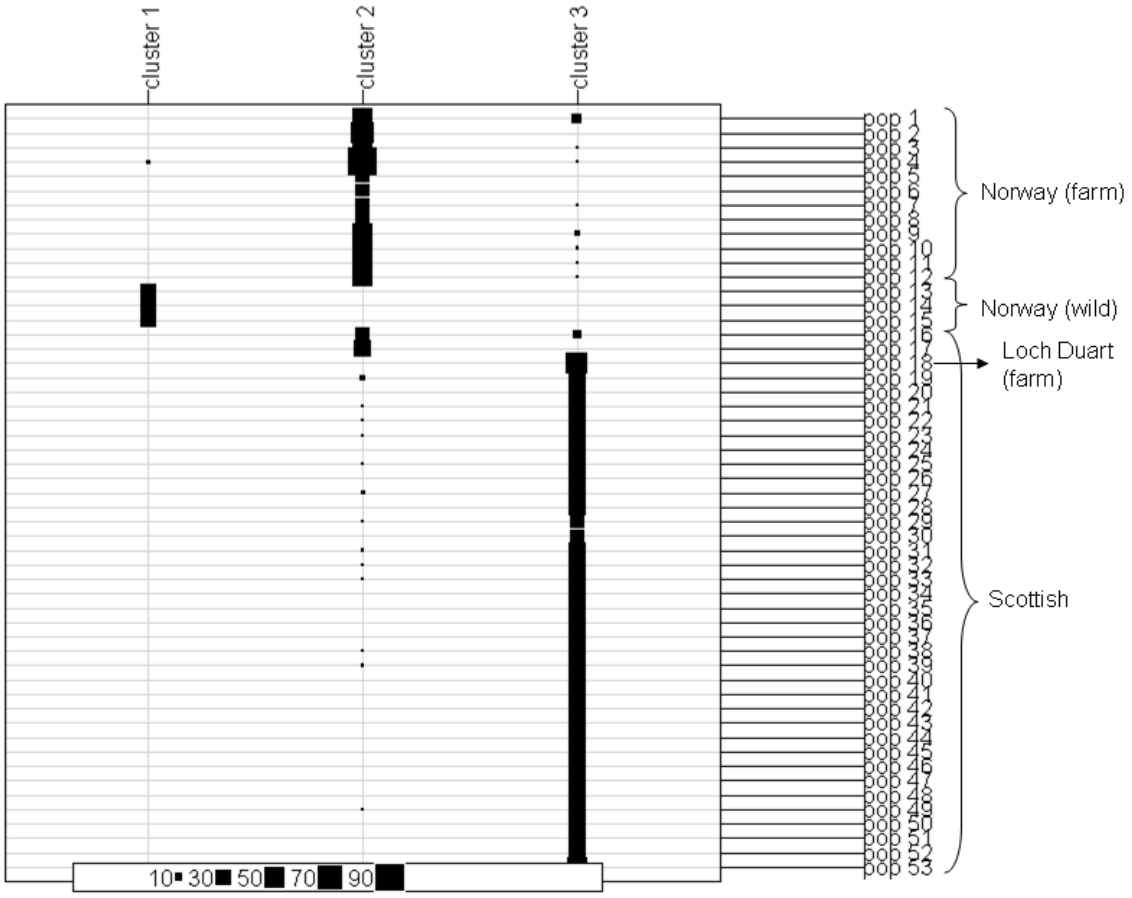

Figure 3. Correct assignment back to individual clusters. Cluster 1 = Wild Norwegian, Cluster 2 = Norwegian Farm, Cluster 3 = Scottish. The size of the boxes represents the number of fish.

4.2. Develop an annual sampling network across the west coast that allows these catchments to be sampled systematically to assess the extent or presence of genetic materials of aquaculture origin.

Each of the participating fisheries trusts was allocated a total of 100 samples for subsequent screening with the developed 60 farm-wild SNPs. The locations of these samples are shown in Figure 4. Additionally a number of locations have been screened with these SNPs as part of the **F**ocusing **A**tlantic **S**almon **M**anagement **O**n **P**opulations (FASMOP) project, thereby extending the geographical coverage.

The locations surveyed aimed to initiate a robust, pan-west coast sampling network, with trusts focusing on areas of particular concern or interest. In addition to samples collected in areas near fish farming operations, sites located further away were also targeted to represent a wide range of the genetic diversity present. The choice of sites was agreed upon by the Trusts and RAFTS staff and in some cases, prior results (e.g. samples of known direct aquaculture escapes). A summary of all samples screened (including those processed as part of FASMOP) is presented in Table 1.

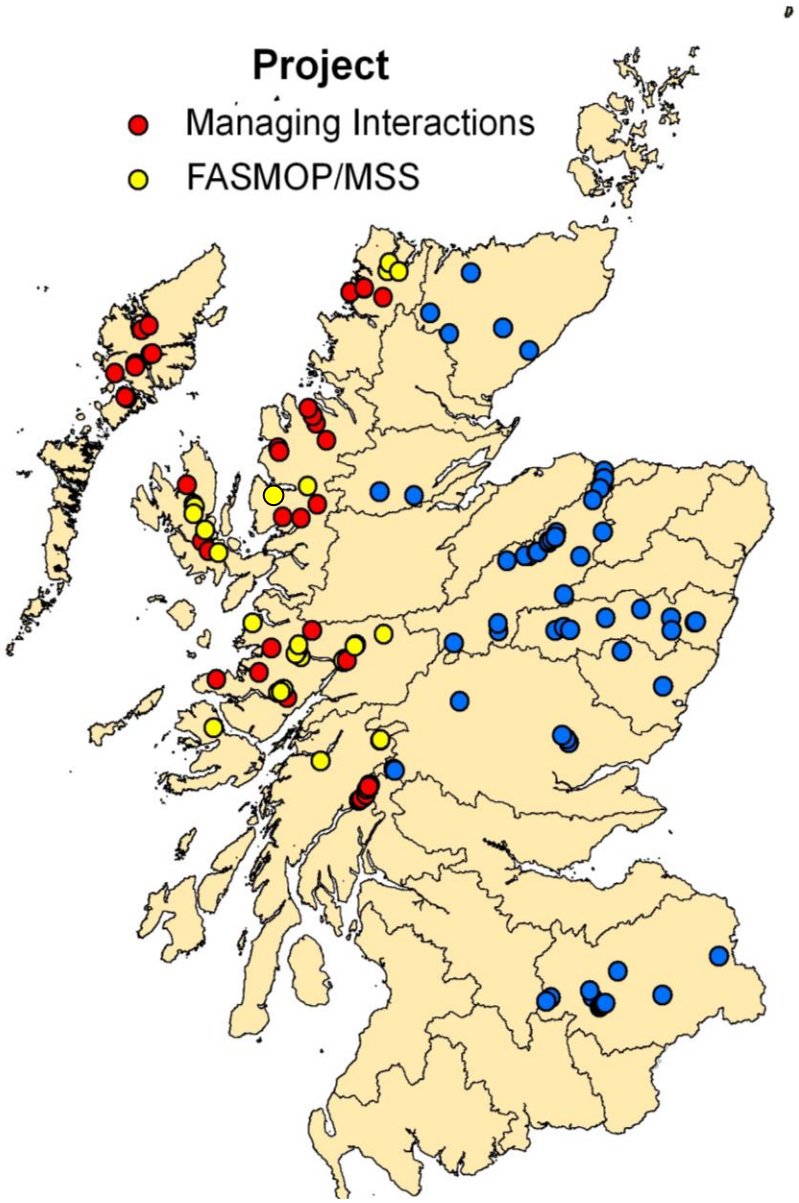

Figure 4. Map of locations sampled screened for farm vs. wild SNPs. West coast samples are coded by the project with which they were screened. Locations in blue represent east coast sampling sites used as a 'wild Scottish' baseline (see text for details).

Table 1. West coast of Scotland sampling locations analysed for farm-wild SNPs.

Trust	River	Site	Year of Collection	Number of Samples
Argyll	Aros River	Loch Frisa	2010	32
	River Awe	Lower Awe	2010	32
		Upper Orchy	2010	32
	River Fyne	various locations	2011	100
Lochaber	Achateny Water	Achateny Burn	2006	15
	River Ailort	mainstem	2011	30
	Carnoch River	various locations	2010	32
	River Lochy	Lower Lochy	2008	29
		Lundy	2011	31
		Pean	2011	24
		Roy	2008	43
		Upper Lochy	2005	34
		Loch Lochy (farm escapes)	2010	35
	River Moidart	mainstem	2006	32
	River Morar	Loch an Nostarie	2005	21
	River Shiel	River Callop	2010	30
		River Finnan	2008	12
			2010	22
		River Shlatach	2005	52
			2008	24
			2010	8
	Strontian River	various locations	2008	23
			2010	8
Outer Hebrides	Kintaravay	mainstem	2011	33
	Ghriomarstaidh	Langadale River	2009	22
		Langavat – Grimersta	2005	21
		Langavat – March Burn	2005	21
	Laxadale	mainstem	2011	34
	Loch Leosaid	River Leosaid	2011	33

Trust	River	Site	Year of Collection	Number of Samples
Skye	River Drynoch	mainstem	2011	33
	River Hinnisdal	mainstem	2011	31
	River Sligachan	mainstem	2010	19
			2011	35
	River Snizort	Lower Snizort	2010	40
		Upper Snizort	2010	32
	Varagill River	mid river	2010	20
Wester Ross	Balgy River (MSS Sheildag)	smolt trap	2006	59
			2007	20
	River Carron	River Lair	2011	33
	Loch Carron	Tullich burn (escapes?)	2011	7
	Gruinard River	Lower river	2008	21
		Mid river	2005	20
		Upper river	2007	23
	River Kerry	Mid river	2011	33
	River Kishorn	Lower river	2011	27
	River Torridon	Mainstem	2007	45
West Sutherland	River Dionard	Mainstem	2006	19
		Rhigolter Burn	2006	19
	River Laxford	Allt Horn	2011	32
		Bad na Baighe	2011	34
	Allt a Mhuilinn	Bhadaidh Daraich	2010	34
	River Polla	Allt Coire an Uinnseinn	2008	32

5. Support fishery trusts in the gathering of samples from an agreed network across west coast catchments.

The above sampling network (Table 1) was supported by a £2,000 payment to each of the participating trusts and was a combination of newly acquired samples on the part of the trust and/or existing samples being stored at

the Marine Scotland Freshwater Laboratory, on behalf of the trusts. All trusts have invoiced and been paid for their sampling contributions.

6. Provision of resources for samples gathered to be analysed, and results reported for prospective application in policy and management practice.

All sites in Figure 4 plus the Norwegian farm baseline were analysed for a final, reduced set of 35 farm-wild SNPs in common. Given the 40% reduction of the number of SNPs (35 out of 60) utilized compared to the full, potential set, it should be anticipated that a consequence of this will be reduced power of the analysis. A principal component analysis of all sites is shown in Figure 5 and represented as sample means. As before, there is strong separation between the Norwegian wild, Norwegian farm and Scottish wild baselines. The sites in blue represent east coast Scottish samples, while those in red represent west coast Scottish samples. Three west coast samples fall well within the Norwegian farm group: Shlatach (2008), Loch Carron (Tullich Burn) and Loch Lochy. Two of these sites (Loch Lochy and Tullich Burn) were known or suspected to be direct escapes and previous work (FASMOP) suggested there may be a farm effect among the temporal replicates from the Shlatach (2005 vs. 2008).

Although a number of farm companies were approached only one provided samples to be used in the project. The farm sample provided by Loch Duart Ltd represents a Scottish aquaculture strain. This sample grouped most closely with the Scottish wild samples and could not be distinguished from them based on group or individual-level analyses. However it may be possible to separate local strains in the event the full set of 60 SNPs (Karlsson et al. 2011) is available across all baselines. Even if this is not the case, given the Loch Duart sample was screened for the full V2 5,500 SNP chip, it remains to be determined if other SNPs (apart from those used by Karlsson et al. 2011) may be useful in distinguishing local strains from wild Scottish fish. These efforts would be greatly

improved if other local strains could be acquired and incorporated into the analysis.

A number of other west coast Scottish samples (labelled in Figure 5) appear to fall between the large Scottish group and the Norwegian farmed strains. This would suggest that these fish represent admixed individuals, a mixture of pure farmed and pure wild individuals, or some combination of both. To look at this issue further, the results of the individual-level analysis are presented in Table 2. Given the purpose of the individual level analysis is to determine if fish captured within the west coast aquaculture zone are Scottish, Norwegian or potentially admixed, only the east coast Scottish samples were used as a 'wild' Scottish baseline. The reason for this is that if there has been a long history of introgression from aquaculture into west coast fish, then they may not actually represent truly wild Scottish fish. Therefore, the east coast, which is further removed from aquaculture was kept as the wild baseline and all west coast samples were treated as test cases.

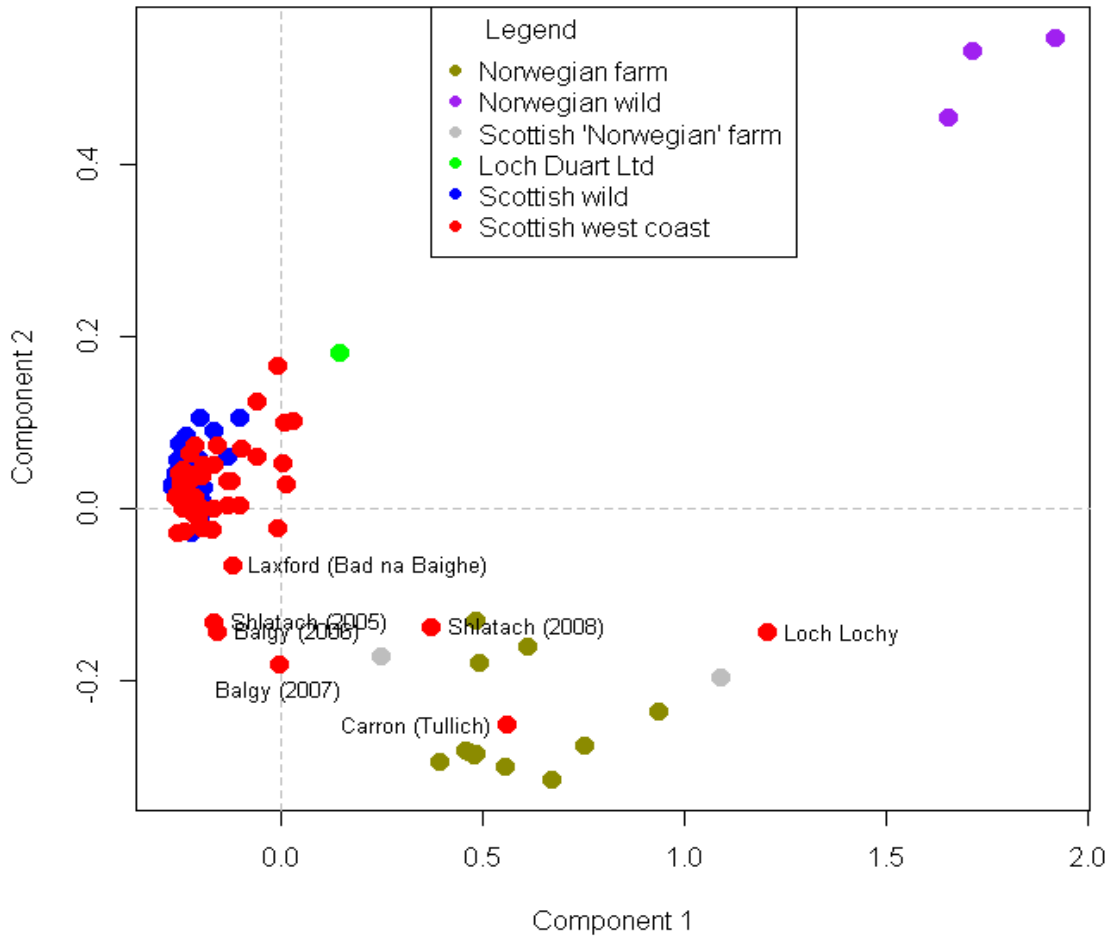

Figure 5. Principal component analysis based on 35 farm-wild SNPs including all west coast samples screened (Managing Interactions & FASMOP).

For the blind, individual-level analysis on the two known baselines (Norwegian farmed & Scottish), STRUCTURE identified these two groups as expected. The average membership coefficient of Norwegian farmed fish to this group was 0.95 ± 0.09 and for Scottish fish was 0.97 ± 0.08. Therefore, given the errors around these averages, a cut-off of 0.85 was chosen. This means that a fish with a membership coefficient of 0.85 or greater to one of the two groups would be classified as 'pure' while a value of <0.85 would infer mixed ancestry (i.e. hybrid). This approach has been used extensively in cases of hybridization between source populations or species (e.g. Sanz et al. 2009; Schwartz & Beheregaray 2008, Taylor et al. 2008; Consuegra et al. 2011). Applying such a cut-off on the Norwegian farm and Scottish baselines resulted in 70 out of 1596

(4.4%) of fish being mis-classified and therefore represents an estimate of the error rate. This error rate applies to the baseline samples (Norwegian farmed and Scottish east coast) and therefore may not represent the true error rate on the west coast, which will be unknown as these are the test samples. However the 4.4% error rate is an average of the error rates for Norwegian farmed fish (6.7%) and Scottish wild fish (3.3%). As the average error rate is higher than that for Scottish fish, this estimate of error should be considered conservative, resulting in added caution being applied to the analysis in identifying a pure Scottish fish as having Norwegian ancestry.

The west coast test samples were compared to the two reference baselines mentioned above and the same cut-off value (0.85) was applied. Therefore any fish with a 0.85 or greater membership value was classified as 'pure' to that particular group (either Norwegian or Scottish) and individuals with a value < 0.85 as hybrids. However, given the error rate mentioned above (4.4%) it stands to reason that some individuals identified on the west coast as admixed will also be false positives. Therefore this error rate was used to estimate per site, how many 'hybrid' individuals would be expected, based on the error rates, and this was compared to the number observed and assessed for statistical significance by Chi-square tests (i.e. were there significantly more admixed fish identified than could be explained by the known accuracy of the analysis). Results on a site-by-site basis are tabulated below. Significant values indicate that there is a greater signal of admixture than would be expected given the calculated error rate from known baseline samples.

Table 2. Sample sites with the number of individuals classified as Scottish, farmed ('Norwegian') or admixed. For admixed individuals, values in parentheses indicate if the level of admixture is significantly higher than expected by the estimated error rate.

Trust	River	Site	Scottish	Norwegian	admixed
Argyll	Aros River	Loch Frisa (farm escapes)	0	32	0 (N/A)
	River Awe	Lower Awe	28	0	4 (Y)
		Upper Orchy	22	0	10 (Y)
	River Fyne	various locations	63	0	37 (Y)
Lochaber	Achateny Water	Achateny Burn	12	0	3 (Y)
	River Ailort	mainstem	11	0	19 (Y)
	Carnoch River	various locations	25	0	7 (Y)
	River Lochy	Lower Lochy	23	0	6 (Y)
		Lundy	26	0	3 (N)
		Pean	13	0	7 (Y)
		Roy	34	0	9 (Y)
		Upper Lochy	30	0	3 (N)
		Loch Lochy (farm escapes)	0	34	0 (N/A)
	River Moidart	mainstem	29	0	3 (N)
	River Morar	Loch an Nostarie	18	0	1 (Y)
	River Shiel	River Callop	21	0	9 (Y)
		River Finnan 2008	8	0	3 (Y)
		River Finnan 2010	17	0	4 (Y)
		River Shlatach 2005	21	6	20 (Y)
		River Shlatach 2008	0	5	19 (Y)
		River Shlatach 2010	6	1	1 (N)
	Strontian River	2008	23	0	0 (N)
		2010	4	0	2 (Y)
Outer Hebrides	Kintaravay	mainstem	28	0	5 (Y)
	Ghriomarstaidh	Langadale River	14	0	8 (Y)

		Langavat – Grimersta	21	0	0 (N)
		Langavat – March Burn	18	0	3 (Y)
	Laxadale	mainstem	30	0	4 (Y)
	Loch Leosaid	River Leosaid	30	0	3 (N)
Skye	River Drynoch	mainstem	26	0	7 (Y)
	River Hinnisdal	mainstem	23	0	8 (Y)
	River Sligachan	2010	13	0	5 (Y)
		2011	29	0	6 (Y)
	River Snizort	Lower Snizort	27	0	12 (Y)
		Upper Snizort	29	0	3 (N)
	Varagill River	mid river	16	0	4 (Y)
Wester Ross	Balgy River	2006	15	3	39 (Y)
		2007	8	7	6 (Y)
	River Carron	River Lair	21	0	11 (Y)
	Loch Carron	Tullich burn	0	6	1 (N/A)
	Gruinard River	Lower river	19	0	2 (N)
		Mid river	17	0	3 (Y)
		Upper river	19	0	2 (N)
	River Kerry	Mid river	24	0	9 (Y)
	River Kishorn	Lower river	19	0	8 (Y)
	River Torridon	Mainstem	31	0	11 (Y)
West Sutherland	River Dionard	Mainstem	15	0	4 (Y)
		Rhigolter Burn	17	0	2 (N)
	River Laxford	Allt Horn	27	0	5 (Y)
		Bad na Baighe	16	0	17 (Y)
	Allt a Mhuilinn	Bhadaidh Daraich	34	0	0 (N)
	River Polla	Allt Coire an Uinnseinn	19	3	13 (Y)

Note: due to quality control, numbers of samples per site here may be lower than in Table 1.

As can be seen from the table above most sites had a signature of hybridization that was significantly higher than expected by chance. Across all sites, 369 out of 1472 (25.1%) individuals were identified as hybrids, which is significantly higher than that seen for the east coast 'wild' baseline. Furthermore, the three cases of putative escapees that were sampled (Loch Frisa, Loch Lochy & Tullich Burn), all but one individual were identified as pure Norwegian fish. Otherwise, very few pure Norwegian fish were identified. The 2006 samples from the Balgy showed most individuals (39 out of 57) to be admixed. However, for some of these fish, they were confirmed in the field as being farmed fish (e.g. presence of injection marks). However, it is known that some wild fish used in the hatchery broodstock were subsequently determined to be of farm origin and therefore it is possible that these escapees could have a mixed ancestry. A similar genetic signature has been found for these samples using microsatellites (Cauwelier et al., in prep, Marine Scotland Science).

Finally, for the 270 simulated F1 hybrids, 228 (84.4%) were correctly identified as such. The remaining individuals (42) were equally distributed between being classified as pure 'Scottish' or 'Norwegian'.

7. Discussion and recommendations

These results show the ability of this panel of farm-wild SNPs to reliably distinguish between direct escapes of Norwegian origin from wild Scottish fish. However, the grouping of the Loch Duart farm strain with the Scottish wild samples based on the individual-level assignment clearly demonstrates that these markers may not work for all domesticated strains if such strains have been derived from native Scottish fish. Given the current panel was developed specifically to distinguish Norwegian farm strains from Norwegian wild fish, the make-up of the panel may need to be revised if local sources of aquaculture strains become available for SNP screening. Given that most farmed fish in Scotland are of Norwegian origin, however, this allows for this tool to be widely, but not universally, applied in Scotland, at present.

The results of the individual-level analysis demonstrate several points. Scottish fish were largely classified as wild and the three examples of known direct escapes (Loch Frisa, Loch Lochy & Tullich Burn) were each confirmed by the STRUCTURE results, with one individual exception. This demonstrates that pure Norwegian farmed escapees can clearly be distinguished from individuals of mixed ancestry. Most fish from west coast sites were identified as Scottish, with varying numbers classified as ‘hybrids’. The levels of hybridization observed on the west coast were significantly greater than the estimated error rate using known baselines from the east coast of Scotland. Only a few sites had hybridization levels that were not significantly different from the error rate, indicating no detectable level of hybridization at those sites. The 85% correct classification of simulated F1 hybrids between Norwegian farm strains and Scottish fish suggests a relatively robust success rate in identifying ‘true’ hybrid individuals and lends further support for the ability of these markers to distinguish intermediate versus pure genotypes.

Of course, the ability to classify individuals as either ‘wild’, ‘farmed’ or ‘admixed’ will be affected by the extent to which each of the pure baselines are represented. A wide and robust baseline has been, and continues to be, developed for Scottish ‘wild’ fish, however such a baseline is still extremely sparse for aquaculture strains being used in Scotland. Indeed, if such a baseline can be improved, this would significantly aid in strengthening the robustness of this type of analysis. It has been clearly shown that the lack of a comprehensive baseline (in this case of both wild and farmed fish) can have a significant and negative impact on the ability of the analysis to be able to examine the questions of interest (Karlsson et al. 2011).

Additionally the current analysis uses only 35 of the 60 SNPs. Given that the data was a combination of several SNP chips (V2 Illumina vs. farm-wild panel) not all 60 SNPs were in common across all individuals. As future samples would be processed at the full 60 SNPs, this would allow for farmed and wild baselines as well as new test samples to be screened for almost twice the number of SNPs, which will likely improve resolution and increase the accuracy around these assignments even further. This reduction in SNP number will also have affected the inconsistency of the clustering analysis in separating the Norwegian farm strains from Norwegian wild fish. Indeed, it has been shown the full SNP panel does separate these two groups (Karlsson et al. 2011). Therefore, future work utilizing the full (and possibly increased) SNP panel will allow for the identification of Norwegian-Scottish hybrid fish to be traced to either wild Norwegian or farmed Norwegian sources. As mentioned earlier, as other non-Norwegian sources of aquaculture strains become available for analysis, the specific set of SNPs used is likely to be refined. This would be of particular relevance to strains such as the Loch Duart samples.

Future work should therefore focus on at least three main recommendations:

- Widen the Scottish baseline samples to be screened. This would obviously include greater sampling along the west coast of Scotland in areas of interest and of primary concern. Additionally, areas removed from aquaculture (e.g. east coast sites) should be screened for the full 60 SNP panel.
- Increased coverage of the various farmed strains utilized by the Scottish aquaculture industry. This is also particularly relevant for local strains. This coverage would ideally be repeated at regular time intervals and would aim to capture the diversity of the different strains being utilized. Furthermore, this coverage would allow for the specific make-up of the SNPs involved to be revisited to improve resolution in the Scottish context.
- Characterization of local (i.e. Scottish) aquaculture strains. Even if the full set of 60 SNPs (Karlsson et al. 2011) does not resolve these strains from wild Scottish fish, there is still scope for choosing further SNPs that may work for a Scottish comparison. To date, only the Loch Duart strain has been sampled for genetics and this was screened at the full V2 5,500 SNP chip. This sample, and any others that may become available, could be analysed to add further SNPs to the full panel of Karlsson et al. (2011).

8. References

Besnier F, Glover KA, Skaala O (2011) Investigating genetic change in wild populations: modelling gene flow from farm escapees. *Aquaculture Environment Interactions*, **2**, 75-86.

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

Christie MR, Marine ML, French RA, Blouin MS (2012) Genetic adaptation to captivity can occur in a single generation. *Proceedings of the National Academy of Sciences*, 109, 238-242.

Clifford SL, McGinnity P, Ferguson A (1998a) Genetic changes in Atlantic salmon (*Salmo salar*) populations of northwest Irish rivers resulting from escapes of adult farm salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 358-363.

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

Consuegra S, Phillips N, Gajardo G, Garcia de Leaniz C (2011) Winning the invasion roulette: escapes from fish farms increase admixture and facilitate establishment of non-native rainbow trout. *Evolutionary Applications*, 4, 660-671.

Dillane E, McGinnity P, Coughlan JP, Galvin PT, Wilkins NP, Cross TF (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.

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

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

Glover KA, Skilbrei OT, Skaala O (2008) Genetic assignment identifies farm of origin for Atlantic salmon *Salmo salar* escapees in a Norwegian fjord. *ICES Journal of Marine Science*, **65**, 921-930.

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

Glover KA, Skaala O, Sovik AGE, Helle TA (2010) Genetic differentiation among Atlantic salmon reared in sea-cages reveals a non-random distribution of genetic material from a breeding programme to commercial production. *Aquaculture Research*, **2010**, 1-9.

Glover KA, Quintela M, Wennevik V, Besnier F, Sorvik AGE, Skaala O (2012) Three decades of farmed escapees in the wild: a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway. *PLoS One*, **7**, e43129.

Herbinger CM, O'Reilly PT, Verspoor E (2006) Unraveling first-generation pedigrees in wild endangered salmon populations using molecular genetic markers.

Horreo JL, de la Hoz J, Gonzalez Pola I, Machado-Schiaffino G, Garcia-Vazquez E (2012) Ecological and economic costs of supportive breeding: Atlantic salmon (*Salmo salar*) as a case study. *Aquaculture*, **356/357**, 1-6.

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

Karlsson S, Moen T, Lien S, Glover K, Hindar K (2011) Generic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular Ecology Resources*, 11, 247-253.

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

Landry C, 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.

Moran P, Verspoor E, Davidson WS (2007) The Atlantic salmon genome. In: *Verspoor E, Stradmeyer L and Nielsen JL (eds). The Atlantic Salmon: Genetics, Conservation and Management*. Oxford, Blackwell Publishing, 57-85.

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

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

Sanz N, Araguas RM, Fernandez R, Vera M, Garcia-Marin J-L (2009) Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. *Conservation Genetics*, 10, 225-236.

Schwartz TS, Beheregaray LB (2008) Using genotype simulations and Bayesian analyses to identify individuals of hybrid origin in Australian bass: lessons for fisheries management. *Journal of Fish Biology*, 72, 435-450.

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

Skaala O, Taggart JB, Gunnes K (2005) Genetic differences between five major domesticated strains of Atlantic salmon and wild salmon. *Journal of Fish Biology*, **67**, 118-128.

Taylor EB, Lowery E, Lilliestrale A, Elz A, Quinn TP (2008) Genetic analysis of sympatric char populations in western Alaska: Arctic char (*Salvelinus alpinus*) and Dolly Varden (*Salvelinus malma*) are not two sides of the same coin. *Journal of Evolutionary Biology*, 21, 1609-1625.

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

Vasemagi A, Nilsson J, McGinnity P, Cross T, O'Reilly P, Glebe B, Berg PR, Primmer CR (2012) Screen for footprints of selection during domestication/captive breeding of Atlantic salmon. *Comparative and Functional Genomics*, **2012**, Art. ID 628204.

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

Villanueva B, Verspoor E, Visscher PM (2002) Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. *Animal Genetics*, **33**, 33-41.

Youngson AF, Martin SAM, Jordan WC, Verspoor E (1991) Genetic protein variation in Atlantic salmon in Scotland: comparison of wild and farmed fish. *Aquaculture*, **98**, 231-242.