SALSEA-MERGE FP7-ENV-2007-1
Grant Agreement No 212529

Work Package 1

Deliverable - D 1.2

Report on new regional markers (D 1.2)

(Month 20)

Table of Contents

I. Microsatellites (Subtask 1.3.2)

Assess existing microsatellite loci for regional differentiation

Overview: *A subset of 400 of the over 1700 microsatellite loci identified in Atlantic to date will be selected based on an assessment of their suitability for population genetics work. Based on screening of the reference collection assembled in Subtask 1.3.1, a suite of 8-12 multiplexable microsatellite loci which give high resolution regional assignment potential will be chosen and optimal conditions for genotyping established.*

Participants: Lead – 1; other 4, 11

Under Sub-task 1.1.1 existing genetic data was critically evaluated and at the Stansted SALSEA Genetics Meeting it was concluded that understanding had advanced over the previous year sufficient to show that the Virginia panel of microsatellite markers would provide a sufficient molecular tool for achieving the basic overriding SALSEA objective. Thus it was concluded that it would be unnecessary to undertake the extensive screening of microsatellites as envisaged in order to identify a sufficient suite to meet basic regional GSI purposes. However, it was recognised that the suite could potentially be refined and made both more efficient and provide higher resolution by either adding additional regionally informative microsatellites, and adding regionally informative mtDNA and SNPs that might be identified under Sub-tasks 1.3.3 and 1.3.4.

Based on the most recent non-SALSEA microsatellite work by consortium partners, reviewed at the Stansted meeting, a smaller group of further microsatellites (N= 49), that had particular promise for improving on the capacities of the Virginia panel, was targeted in further development work. The freed resources were redirected to expand development efforts related to nuclear SNP development under Sub-task 1.3.4 as it was concluded that this would be the most productive way of overall exploring the potential of different markers for increasing the resolution of genetic assignment in the future.

The reference collection assembled for evaluation of additional loci (sub-task 1.3.1) was screened for variation in 64 microsatellite loci. These 64 loci included the common set of 15 microsatellite markers used for genotyping the European baseline, as well as additional neutral microsatellites and EST-microsatellites. In total, 30 neutral microsatellites, 21 EST loci, and 2 MHC-linked loci gave scorable results. Exploratory analyses of genetic differentiation and power for individual assignment of various combinations of loci were conducted, including the standard set of microsatellites. Correct individual assignment to river varied, and was highest when all 53 loci were included in the analysis, resulting in an average of 78% of the individuals assigned to their river of origin (see table below). When using the standard set of microsatellites, average correct assignment was on average 52%. Another analysis was conducted using the 15 loci that WHICHLOCI-analysis indicated

gave greatest power of discrimination and this resulted in on average 61% correct assignment. Assignment to region (country in this context) was much higher; with an average correct assignment of 96% when using all loci, and 73 % and 78% when using the standard set and the Whichloci set respectively. Regional differences were also observed in correct assignment with the various sets of loci tested. While some loci were highly differentiating in some regions and on some spatial scales, others were more informative in other regions and on different spatial scales. Though higher average assignment was achieved for the optimal set compared to the standard set, there were regional differences and none of the sets performed consistently better than the other.

The results from these analyses were also compared in the table below to the results obtained from the SNP analyses on the same samples obtained in sub-task 1.3.4. This shows the percentage correct assignment to river, and to region, for the reference collection of samples using various combinations of loci. This shows the improvement to be variable depending on the river/region, and sometimes negative, but for some cases either markedly better or worse. This suggests that the gains from adding and deleting microsatellites from the marker suite already defined is uncertain and tool development should focus on SNPs where a similar or better geographical resolution and assignment accuracy can be expect but typing efficiencies are likely to be greater. Thus the enhancement of the existing tool should focus on replacing or extending the current tool should focus on SNP markers.

River	**To river**			**To region**		
	All	**Standard set**	**Optimal 15**	**All**	**Standard set**	**Optimal 15**
BlackW	50	25	33	100	59	66
Laxa	100	92	100	100	92	100
MoyTrim	67	33	58	100	83	83
Numedalslågen	70	30	70	90	60	90
Orkla	83	58	50	92	64	50
Shin	92	75	83	92	75	83
Esva	100	67	92	100	67	92
Naatamo/Neiden	75	58	58	100	64	83
Ponoi	50	42	42	100	84	92
Suir	64	55	18	91	82	45

II. MtDNA SNPs (Subtask 1.3.3)

Identify mtDNA SNPs for regional differentiation

Overview: *mtDNA from the D-loop, ND1, Cytb and other gene regions will be sequenced to identify single nucleotide polymorphism (SNPs). Based on screening of the reference collection assembled in Subtask 1.3.1, a suite of SNPs which provide useful regional* assignment capability will be chosen and optimal conditions for typing established.

Participants: Lead – 3; other 8

Regional mtDNA SNP differentiation in European Atlantic salmon (*Salmo salar*): an assessment of potential utility for determination of natal-origin

Eric Verspoor, Sonia Consuegra, Olafur Fridjonsson, Sigridur Hjorleifsdottir, David Knox, Kristinn Olafsson, Scott Tompsett, Vidar Wennevik and Carlos Garciá de Leániz.

E. Verspoor and D. Knox: Freshwater Laboratory, Marine Scotland, Pitlochry, Scotland PH16 5LB, UK. S. Consuegra: Institute of Biological Sciences, University of Wales, Aberystwyth, UK. O Fridjonsson, S. Hjorleifsdottir and K. Olafsson: Matís, Vínlandsleið 12, 113 Reykjavík, Iceland. S. Tompsett and C. Garcia de Leaniz: Department of Pure & Applied Ecology, Swansea University, Swansea SA2 8PP, UK. V. Wennevik: Institute of Marine Research, 5005 Bergen, Norway. E. Verspoor: 12 Dixon Terrace, Pitlochry, Scotland PH16 5QX, UK. Correspondence to E. Verspoor: +44 1796 472945; email: eric.salgen@gmail.com

Abstract

The Atlantic salmon, *Salmo salar*, shows geographically structured differentiation at various classes of molecular genetic variation, among and within river stocks. Nuclear microsatellite locus variation at multiple loci has been exploited as a marker for the continental origin of fish caught at sea in high seas fisheries for over a decade. However, a simpler, more cost-effective, but still accurate, assignment can be obtained using a single microsatellite locus in combination with a mtDNA restriction enzyme detected polymorphisms. Following on from this, a preliminary study was made of the potential for using mtDNA SNP variation to enhance the resolving power and cost-effectiveness of within continent assignment of European salmon based on microsatellites. Variation in 20 mtDNA regions, encompassing ~43% of this genome in 330 individuals from 29 rivers across Europe, was analysed. High levels of inter-individual and inter-river variation were found as well as evidence of regional differentiation paralleling observed microsatellite differentiation. The observations indicate scope for using mtDNA SNPs along with microsatellites for genetically-based assignment of European salmon to region and river of natal origin but further study is needed.

Keywords: mitochondrial DNA, genetic stock identification, marine ecology, 454 sequencing

Introduction

Inherent differences among genetic populations, or phylogeographic groups, can potentially be used as markers or tags in ecological studies, to resolve population structuring and determine the origin of individuals (Swartz *et al.,* 2006; Palsboll *et al.,* 2006). The extent to which this is possible depends on the nature of structuring, including the extent of genetic isolation and evolutionary divergence among populations (Waples and Gaggiotti, 2006). Just as crucially, it depends on identifying DNA loci where differentiation has evolved due genetic drift or selection. In most species the variable loci used as tags represent, at best, an optimized subset of an arbitrary set of available polymorphic loci. Most sets of loci used, given their derivation from arbitrary DNA loci, are unlikely to represent the most divergent loci and best possible set of population markers for resolving population structuring and assignment of natal origin, or be the most cost-effective choice. Yet identifying the best loci would maximize resolving power and assignment success, and finding them poses a significant challenge given the size of most genomes and that different loci may be optimal in different parts of a species range. However, the scope for improving existing sets of marker loci is being facilitated by recent advances in genome sequencing technology, that allow rapid genome scanning for polymorphisms at acceptable levels of cost (Davey *et al.*, 2011).

Most Atlantic salmon (*Salmo salar* L.), a culturally iconic and quintessentially anadromous fish of the Northern Atlantic, spend their early life in rivers, undertake a marine migration, and return to their natal river to spawn and complete their life cycle (Webb *et al*., 2007). Attempts to exploit the potential of molecular markers to ascertain the origin of fish began in the late 1960s, and the ensuing genetic studies have dramatically altered understanding of the structuring of the species into distinct populations and phylogenetic groups. The collective body of work that has emerged makes clear that North American and European stocks represent two essentially isolated phylogenetic groups that, arguably, should be considered distinct subspecies (King *et al.*, 2007), and provides the basis for assigning salmon to their natal continent of natal origin with effectively 100% certainty (Koljonen *et al.*, 2007 and references therein). It also shows clearly further substantive phylogenetic substructuring within these two continental groups as well as phylogenetic and meta-population structuring within rivers (King *et al.*, 2007).

Molecular genetic differentiation among rivers and regions has been exploited for natal assignment of fish on a regional or river specific basis within continental stock groups in a few contexts, (Koljonen *et al*., 2007; Gauthier-Oullet *et al.*, 2009; Griffiths *et al.*, 2010; Sheehan *et al.*, 2010). It is only recently that work has been directed at development of a robust, comprehensive methodology for within continent regional or river-specific assignment. In respect of European salmon, work has been directed at developing a microsatellite based assignment tool (GRAASP) as part of the EU SALSEA-Merge Project (Verspoor *et al*., submitted), aimed at increasing understanding of the marine ecology of this iconic species in the NE Atlantic.

GRAASP as it currently is implemented provides a cost-effective broad-scale assignment of European salmon to broad regions, though in some cases river-specific assignment can be achieved (Gilbey *et al.*, submitted). However, the suite of

microsatellite loci used do not in most cases allow for fine scale regional or river specific assignment. Yet, what work has been carried out (King *et al*., 2007), shows regional differentiation of river stocks at finer scales, even between adjacent rivers, suggesting that accurate river specific assignment may be possible (e.g. Wennevik *et al.*, 2004; Ryynänen *et al.*, 2007; Grandejean *et al.*, 2009; Tonteri *et al*., 2009), if a suitable set of DNA markers can be identified for river stocks and their constituent populations.

Mitochondrial DNA (mtDNA) is an independent, maternally and essentially clonally inherited, haploid component of the salmon's genome. It evolves rapidly due to a high mutation rate and shows higher levels of population differentiation than many nuclear genes due to a lower effective population size (Hansen *et al.*, 2007). Its potential as a population marker was first investigated in respect of continent of origin (Bermingham *et al*., 1991) and a mtDNA restriction fragment length polymorphism (RFLP) were used by Gilbey *et al.* (2005) with a single nuclear microsatellite locus to provide a simple, highly cost-effective marker suite for assigning continent of origin of Atlantic salmon with a projected 99%+ accuracy. Studies of restriction enzyme and sequencing detected polymorphisms shows substantive regional and river-specific differences in variant frequencies (King *et al.*, 2007; Verspoor *et al*., unpublished), suggesting some variation may be suitable for use as intra-continental population markers. However, the full extent of regional and inter-river mtDNA differentiation is unclear as, in most population studies, only a small part of the mtDNA genome (generally <5%) or small part of the species' range has been screened. A complete analysis of the mtDNA genome was carried out by So (2006) but was severely constrained by the number of fish (n=14) and locations (N=9).

Described here is a broad-scale preliminary assessment of the nature and extent of mtDNA single nucleotide polymorphism (SNP) in European Atlantic salmon based on the analysis of ~43% of the mtDNA genome. The aim of the study was to provide an unbiased assessment of mtDNA SNP variation, the extent of population differentiation, and the potential for exploiting this variation as population markers. The study exploits recent advances in enhanced polymorphism screening capacity provided by next generation DNA sequencing methodologies.

Materials and Methods

Samples

The study encompasses the screening of mtDNA variation in 330 individual salmon from 29 rivers across Europe, with numbers analysed ranging from 6 to 12 individuals per river. The rivers selected are broadly geographically representative (Figure 1) and encompass the main phylogeographic regions suggested by allozyme studies (Verspoor *et al.*, 2005). The samples analysed derive from archived ethanol preserved fin tissue collected over the last two decades as part of other studies.

DNA extraction and sequencing

DNA was extracted using commercially available DNA extraction kits (Quiagen). Screening for variation was carried out in a single sequencing run using a novel approach developed by combining the traditional PCR amplification of known gene

regions with 454 Titanium FLX (Roche, 454 Life Sciences) technology (Fridjonsson *et al.*, 2011). The method employs a unique combination of bar-coded primers and a partitioned sequencing plate to associate each sequence read to an individual. The approach allowed sequencing of extensive regions of the mtDNA genome for a large sample group (546 individuals) in a single run, make it both quick and cost-effective. Twenty independent regions of 311 to 384 bp were sequenced for each individual, encompassing a total of 7215 bases (Table 1), ~ 43% of the 16,665 bp Atlantic salmon mtDNA genome (Hurst *et al.*, 1999). The choice of regions was guided by the total mtDNA sequence analysis of 14 salmon from across the species range by So (2006), and focused on regions they showed had the highest levels of polymorphism. Sequence reads were aligned according to the *S. salar* mitochondrial reference sequence (NC_001960.1) and the presence of a SNP was accepted as valid if (i) sequence reads were produced from both DNA strands; (ii) they occurred in a minimum of 90% of replicate sequence reads; and (iii) they occurred in more than one individual. The average number of reads supporting each SNP per individual was 27.3 was with a standard deviation of 11.7 (Fridjonsson *et al,.* 2011).

Analysis

Composite SNP profiles of the individual fish were assembled from sequence data for the 20 amplified fragments. Given the small sample sizes, all individuals were used in the analysis of the distribution of haplotypes among locations, even those with <5% missing sequence data. For these, missing bases were conservatively assumed to be the same as the nearest haplotype in the same or an adjacent population sample. The relatedness of the haplotypes identified assessed based on numbers of pair-wise differences and a minimum evolution (ME) tree constructed for inferring the evolutionary relatedness of haplotypes using Mega4 (Tamura *et al.*, 2007).

A cumulative plot of numbers of haplotypes identified with progressively increasing numbers of amplicons was constructed manually, based on 26 of the 29 populations sampled; the best fit curve fit was determined visually using the SlideWrite Plus (Advanced Graphics Software). Individual plots of haplotype diversity as a function of sample size were generated by the rarefaction function in PAST v2.11 (Hammer *et al*., 2001). The relationship between numbers of populations sampled and numbers of haplotypes observed was generated by manual re-sampling of the populations stratified by the regional groupings as indicated in Figure 1.

Average pair-wise differences within and corrected average pair-wise differences among populations among individuals were calculated and tested for significant differences between samples, and an AMOVA analysis of within and among group variation done. Both tests were carried out using Arlequin v3.5 (Excoffier and Lischer, 2010). Regional groups used in the AMOVA analysis correspond closely with those identified by microsatellite data (Gilbey *et al*., submitted). The groups were 1) Rynda and Teno, 2) Namsen,Eiravassdraget and Bjerkreimselva, 4) Tweed, North Esk, Ugie and Oykel, 5) Laxford, North Uist, Awe and Feochan 6) Stinchar, Eden, Conwy, Blackwater and Taw, with the remaining individual samples treated as distinct groups. A Mantel test of association of genetic and geographic distance calculated using PAST v2.11 (Hammer *et al.*, 2001). For the Mantel test, a geographic distance matrix was generated using the Geographic Distance Matrix Generator (Ersts, 2011) and the pair-wise population genetic

distance matrix generated by Arelquin v3.5. Using the latter matrix, a minimum evolution (ME) clustering tree was generated using MEGA4.

Results

The SNP variation observed within and among the 330 individuals screened defined 139 haplotypes for which DNA sequences are available on GENBANK (Accession numbers xxxxxx – yyyyy; to be submitted on acceptance of paper for publication) Only 7 of the 330 fish were uncertain and could be assigned to either of two closely related haplotypes differing in 1 base pair. Haplotype frequencies observed across samples are set out in Table 2. As summarised in Figure 2, no haplotypes were observed in all samples, only three occurred at ten or more locations, and only 12 were observed in fish from two or more locations; 89 occurred in only one sample.

On the basis of genetic relatedness, the haplotypes clustered into five major groups based on pair-wise differences (Figure 3), with most haplotypes found in one of these, with the other four containing 2-4 types each, of which three clusters are particularly distinctive. The four most common haplotypes, 16, 66, 67 and 96 are found in the largest major cluster. The most distinct grouping is the 136, 137 and 138 cluster, within which haplotypes differ from each other by 1-3 bases. In contrast, they differ from all other haplotypes in all the other clusters by 64-78 base changes, a sequence divergence of 0.89-1.08%. The remaining haplotypes divide into one large and three smaller clusters among which haplotypes differ at 10-20 bases compared to 1-10 bases between haplotypes within these groups. The largest of these three clusters then shows further sub-structuring into three more poorly defined groups and these in turn into smaller groups or more closely related haplotypes with most haplotypes within smaller clusters separated by 1-5 base differences.

The number of haplotypes defined within each amplicon varied from 3 to 11, with a 4-fold variation in the number of haplotypes defined per SNP (Table 1); the number of SNPs per amplicon varied from 4 to 13 which, when corrected for amplicon size, showed a 4-fold variation in SNPs found per base pair sequenced. In some cases, such as one part of the *ND4* gene, only ~1 in 3 of SNPs were associated with a new haplotype, where as in the second part of the *CoxII* gene, the number of haplotypes defined was greater than the number of SNPs, due to the SNPs in this region showing a degree of independent assortment. However, within most regions between 50 and 100% of SNPs defined new haplotypes; across the total sequence analysed ~80% were associated with unique haplotypes.

The number of haplotypes resolved increased progressively with the number of amplicons (Figure 4) across the 20 regions sequenced starting from the D-loop clockwise to the CytoB gene region. The best fit to the cumulative curve is a second order polynomial suggesting that, in general, as the number of amplicons added to the analysis increased, there was a decreasing number of new haplotypes added per base sequenced. However, there was considerable variation in the number of new haplotypes added depending on the amplicon. For example, the addition of amplicons 6 and 9 (Figure4 and Table 1) gave1-2 new haplotypes while including amplicon 10 added approximately 18 new haplotypes.

Stratified sub-sampling of populations shows the numbers of haplotypes to be a direct function of the number of populations screened (Figure 5). The best fit

curve for the observed relationship is also a 2nd order polynomial and suggests numbers detected with each additional population may be decreasing gradually with a possible plateau in haplotype numbers predicted when the numbers reach 50-60. In contrast, the rarefaction curves for haplotype diversity as a function of sample size, with one major exception, show a more or less linear increase in haplotype diversity with increasing size of sample (Figure 6). In the case of the Allier, the curve begins to level out suggesting that the estimate of haplotype diversity from this location is less constrained by sample size than in the case of the other locations.

A high proportion of samples show significant pairwise differences (Table 3). Overall there is no significant association of genetic differentiation with geographic distance among samples (Mantel test R=0.009, p=0.42) and patterns of pairwise differentiation are complex do not appear entirely unlinked to geography. This illustrates that sites which are both geographically distant but proximate in the sampling scheme can be genetically relatively similar (e.g. the Neva and Pechora samples) while those that are geographically close can be relatively highly divergent (e.g. the Hofsa and Olfusa). This apparent randomness is widespread but there is also some evidence of regional patterns of differentiation (e.g. Iceland vs the rest, the close relatedness of the Teno and Rynda, and the close relationship of the Pehcora, Pongoma and Neva). Pairwise differences among geographically close rivers, recognising the somewhat arbitrary nature of the cut-off as to what is included, are graphically summarised in Figure 7 and an overall ME tree based on pairwise differences is shown in Figure 8.

Molecular analysis of variance shows that the frequencies of haplotypes in the samples are highly significantly heterogeneous among the defined groups and approaches significance among samples within groups (Table 4). The Fixation Indices and associated significance are F_{SC} = 0.01292 (*Va*, $p < 10^{-6}$), F_{CT} = 0.16830 (*Vb*, $p < 0.08$), and F_{ST} = 0.17905 (*Vc*, $p < 10^{-6}$), based on 1023 permutations.

Discussion

The assessment of potential for use of mtDNA variation as population markers carried out was made possible by technological advances that allow cost-effective sequencing of a large proportion of the Atlantic salmon mitochondrial genome in a large number of individuals using a novel next generation sequencing protocol (Fridjonsson *et al.*, 2011). Robust assessment of this potential requires screening large numbers of individuals from a representative set of populations across the geographic distribution of the species of interest, for much if not all of the mitochondrial genome. To date, at best, with available technology and the cost of screening has been either possible to characterize 1) large numbers of individuals for a small number of restriction fragment length polymorphisms (RFLPs) or, less commonly, for SNP variation in a small PCR amplified fragment (Verspoor *et al.*, 2006), or 2) small numbers of individuals for large parts of the genome using large numbers of restriction enzymes (REF), or 3) sequence small numbers of salmon for the entire mtDNA (So 200X). However, the potential for using mtDNA variation as a marker in some cases has been demonstrated e.g. continent of origin (Gilbey *et al.*, 2005) and it is known that regional differentiation occurs, both in North America and

Europe (King et al., 2007 and references there in). This suggests there is a potential for its application on smaller regional scales within continents.

The analysis of mtDNA SNP variation in European salmon reported here reinforces this view and significantly advances existing understanding of general levels of diversity and points to a high level of mitochondrial diversity within and among rivers. However, the full extent of regional and inter-river differentiation remains to be elucidated. Given the high levels of diversity and the relatively limited sampling of rivers and of individuals within rivers, the sample numbers and sizes screened are inadequate. They do not provide an accurate and precise account of the number of different haplotypes present or their frequencies, and inter-river differentiation. That said, the results strongly suggest that haplotype distributions and frequencies differ significantly among most river systems and that there is likely to be regional differentiation as well, that can be expected to mirror at least in its broad patterns, that observed at nuclear loci (King et al. 2007).

The fact that there are few haplotypes shared between even geographically adjacent samples adds weight to the view that there is a high level of uniqueness in haplotype frequencies between populations. Three considerations suggest that the numbers of haplotypes identified are likely to be a fraction of the mtDNA variation present in European salmon stocks. Haplotype numbers in almost all populations are a linear function of sample size and do not plateau as sample with increasing numbers of individuals sampled as expected if most haplotypes had been resolved. The same is true of the number of populations sampled. The analysis also suggests that the rate of increase in numbers of haplotypes with increasing numbers of samples is only starting to decline. As such it is expected that the amount of diversity found would be increase substantially by both increasing sample sizes and increasing sample numbers. Finally, only ~43% of the mtDNA was screened and the actual number of haplotypes in the 330 fish examined is undoubtedly higher than this partial analysis of the mtDNA genome shows. Thus a more extensive genomic analysis would be expected to show many of the haplotypes resolved here to represent heterogeneous classes. However, the increase in numbers of haplotypes resolved appears to be starting to decline with increasing numbers of amplicons suggesting that further screening of more mtDNA regions is may not be as useful as extending the number of populations surveyed and the number of individuals screened per population. On the other hand, the data also show that the number of new haplotypes added does vary considerably across the mtDNA molecule.

The analysis suggests that further research will be most productively focused on a more extensive analysis of both populations and individuals within populations. This need not involve the screening of all SNPs as in a number of cases different SNPs are exclusively associated with a single haplotype and only one may be required for its resolution, reducing the number of SNPs to screen without losing information. Extending further work on individuals and populations will provide a baseline to exploit those haplotypes which show regional and river specific variation for assignment purposes.

The findings of the current study are line accord with observations of previous work based on RFLP and sequence analysis of more restricted parts of the Atlantic salmon mtDNA that show that regional differentiation on different spatial scales. Major differences have previously been reported between Baltic and Atlantic

salmon stocks in Europe as well as among regions for restriction enzyme detected SNP variation (Verspoor *et al*., 1999; Nilsson *et al.*, 2000). Regional variation on smaller spatial scales has also been reported within the Baltic (Nilsson et al. 2001) but a detailed analysis of regional RFLP variation among European Atlantic stocks has not been reported. The only report of small scale regional variation is that of Verspoor *et al.* (2002; 2006) whom found that one RFLP identified in the ND1 gene region and resolved by the restriction enzyme *AluI* was only present in populations of salmon in the inner Bay of Fundy. Extending this work, sequence analysis of two 350 base pair regions of this gene in 8xx salmon from YY rivers found evidence that populations of salmon in the Inner and Outer Bay of Fundy as well as along the south and eastern shores of Nova Scotia, showed regional differentiation including some apparently low frequency regionally-specific haplotypes.

Based on the results of the current study, the lack of evidence from existing mtDNA studies for regional structuring probably arises from such work being based on a limited and arbitrary screening of the mtDNA molecule with a few restriction enzymes which resolve widespread polymorphisms, missing most variation and that shows high levels of inter-region or inter-river variation. Given the existence of a high degree of regional variation at nuclear genes (Verspoor *et al.*, 2005; King *et al.*, 2007; Gilbey *et al.*, submitted), it might be expected that the same, or even greater levels of differentiation should be seen in relation to mtDNA given the greater potential for population differentiation inherent to this component of the genome (Hansen *et al.*, 2007). Concordance of small scale regional patterns of differentiation occur in Atlantic salmon stocks in eastern Canada where both classes of variation have been more extensively studied (Verspoor *et al.*, 2002, 2006; O'Reilly, unpublished); the regional differentiation resolved has also more recently been supported by studies of nuclear SNP variation as well (Freamo *et al.*, 2011).

The observations reported here, in so far as they relate to regional and inter-river differentiation, are not inconsistent with significant regional structuring being present in Europe and with the observations of studies to date. However, out with the Baltic, the situation is decidedly inconclusive and the current results change this situation little. The analysis of variation within and among groups, based on the regional groups suggested by the more detailed microsatellite analysis (Gilbey *et al.*, submitted) shows significant differences between these regions in the absence of any general association of genetic and geographic distance. However, the proportion of variation observed within rivers, and the high level of inter-sample variation, preclude the possibility of drawing robust conclusions. Most of the potential regional groups are represented by a single sample, confounding distinguishing between inter-river and inter-regional variation, and many of the differences or not found among samples may be artefacts of sample sizes. The number of rivers screened and the samples sizes used are too small, given the levels of variation observed, to draw specific conclusions and from the current analysis it is only possible to make the general point that the observations are strongly suggest there is substantive regional and inter-river divergence in respect of mtDNA variation.

Despite its limitations and preliminary nature, the current study significantly advances understanding of intra- and inter population mtDNA SNP variation in European Atlantic salmon stocks. It makes more clear the considerable potential for

using mtDNA SNPs to enhance the assignment success and resolution of microsatellite based tools such as the SALSEA-Merge GRAASP (Verspoor *et al*., submitted; Gilbey *et al.*, submitted a, b), alone or in combination with nuclear SNPs (Coughlan *et al*., submitted). Enhancement of the SALSEA-Merge GRAASP, by integrating in the most informative of these two marker types, is likely to become increasingly cost-effective, given on-going advances in the speed and cost of screening SNPs, relative to microsatellite loci. These technological advances will also facilitate the required further exploration of population differentiation to more fully assess the potential offered by SNPs and which SNPs are most useful, as well as the development of the detailed population baseline data for chosen markers required for accurate assignment. However, further work is required to establish the full extent of regional and inter-river mtDNA differentiation in Atlantic salmon stocks and the extent to which could be exploited for assignment of a salmon's natal origin.

Acknowledgements – This work was carried out as part of the NASCO sponsored and EU funded SALSEA-Merge FP7 Project (Contract No.212529) entitled "Advancing understanding of Atlantic salmon at Sea: Merging Genetics and Ecology to resolve Stock-specific Migration and Distribution patterns".

References

Bermingham, E., Forbes, S.H., Friedland, K. and Pla, C. (1991. Discrimination between Atlantic Salmon (Salmo salar) of North American and European Origin using Restriction Analyses of Mitochondrial DNACanadian Journal of Fisheries and Aquatic Sciences, 48: 884-893

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing Nature Reviews Genetics, 12:499-510

Ersts,P.J. 2011. Geographic Distance Matrix Generator(version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Available from http://biodiversityinformatics.amnh.org/open_source/gdmg. Accessed on 2011-8-24.

Excoffier, L. and H.E. L. Lischer (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Freamo, H., O'Reilly, P., Berg, P.R., Lien, S. and Boulding, E.G. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Nature Genetics 11 (Issue Supplement s1): 254–267,

Fridjonsson O, Olafsson K, Tompsett S, Bjornsdottir S, Consuegra S, Knox D, de Leaniz CG, Magnusdottir S, Olafsdottir G, Verspoor E, & Hjorleifsdottir S. (2011). Detection and mapping of mtDNA SNPs in Atlantic salmon using high throughput DNA sequencing. BMC Genomics 12(1):179.

Gauthier-Ouellet, M., Dionne, M., Caron, F., King, T.L. and Bernatchez, L. 2009. Spatio-temporal dynamics of the Atlantic salmon Greenland fishery inferred from mixed-stock analysis. Canadian Journal of Fisheries and Aquatic Sciences. 66: 2040-2051.

Gilbey, J., Knox, D., O'Sullivan, M. & Verspoor, E. (2005). Novel DNA markers for rapid, accurate, & cost-effective discrimination of the continental origin of Atlantic salmon (Salmo salar L.). ICES Journal of Marine Science 62: 1609-1616.

Griffiths, A.M., Machado-Schiaffino, G. Dillane, E. Coughlan, J. Horreo, J. Bowkett, A. Minting, P. Toms, S. Roche, W. Gargan, P. McGinnity, P. Cross, T. Bright, D. Garcia-Vazquez, E. & Stevens, J. 2010. Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in the southern part of the European range. BMC Genetics 11: 31

Grandjean, F., Verne, S., Cherbonnel, C. and Richard, A. 2009. Fine-scale genetic structure of Atlantic salmon (Salmo salar) using microsatellite markers: effects of restocking and natural recolonisation. Freshwater Biology, 54: 417–433,

Hammer, Ø., Harper, D.A.T., and P. D. Ryan, 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1): 9pp.

Hansen, M.M., Villanueva, B., Nielsen, E.E. and Bekkevold, D. 2007. Investigating the Genetics of Populations. Chapter 4, In The Atlantic Salmon: Genetics, Conservation & Management (Verspoor, E., Stradmeyer, L & Nielsen, J.L., eds). Blackwell Publishing, Oxford. pp 86-114.

Hurst, C.D., Bartlett, S.E., Davidson, W.S. and Bruce, I.J, 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239: 237–242.

King, T.L., Verspoor, E., Spidle A. P., Gross, R., Phillips, R. B., Koljonen, M-L., Sanchez, J.A. & Morrison, C.L. (2007). Biodiversity and Population Structure. Chapter 5, In The Atlantic Salmon: Genetics, Conservation & Management (Verspoor, E., Stradmeyer, L & Nielsen, J.L., eds). Blackwell Publishing, Oxford. pp 117-166.

Koljonen, M-L., King, T.L. and Nielsen, E.E. 2007. Genetic identification of Individuals and Populations. Chapter 3, In The Atlantic Salmon: Genetics, Conservation & Management (Verspoor, E., Stradmeyer, L & Nielsen, J.L., eds). Blackwell Publishing, Oxford. pp 270-298.

Palsbøll, P.J., Bérubé, M. and Allendorf, F.W. 2007. Identification of management units using population genetic data. Trends in Ecology and Evolution, 22: 11-6.

Ryynänen, H.J., Tonteri, A., Vasemägi, A. and Primmer, C.R. 2007. A comparison of bi-allelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). Journal of Heredity 98: 692-704 Sheehan et al. 2010

Schwartz, M.K., Luikart, G. and Waples R.S., 2006. Genetic monitoring as a promising tool for conservation and management. Trends in Ecology and Evolution, 22:25‑33.

Tamura,K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. 24(8):1596–1599.

Tonteri, A., Veselov, A.J., Zubchenko, A.V., Lumme, J. and Primmer, C.R. 2009. Microsatellites reveal clear genetic boundaries among Atlantic salmon

(*Salmo salar*) populations from the Barents and White Seas, northwest Russia. Canadian Journal of Fisheries and Aquatic Sciences. 66: 717-735.

Verspoor, E. O'Sullivan, M., Arnold, A.L, Knox, D. and Amiro, P.G. 2002. Restricted matrilineal gene flow & historical population fragmentation in Atlantic salmon (*Salmo salar* L.) within the Bay of Fundy, eastern Canada. Heredity, 89: 465-472.

Verspoor, E., Beardmore, J.A., Consuegra, S., Garcia de Leaniz, C. Hindar, K. Jordan, W. C. Koljonen, M-L. Makhrov, A. A, Paaver, T. Sánchez, 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 (Suppl. A): 3-54.

Verspoor, E. O'Sullivan, M. Arnold, A.M. Knox, D. Curry, A. Lacroix, G. and Amiro, P. 2006. The Nature and Distribution of Genetic Variation at the Mitochondrial Nd1 Gene of the Atlantic Salmon (*Salmo salar* L.) Within and Among Rivers Associated With The Bay Of Fundy and the Southern Uplands of Nova Scotia. FRS Research Services Internal Report No 18/05.

Waples, R.S. and Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology, 15:1419–1439.

Webb, J.H., Verspoor, E., Aubin-Horth, N., Romakkaniemi, A. and Amiro, P. 2007. The Atlantic Salmon, Chapter 2, In *The Atlantic Salmon: Genetics, Conservation & Management* (Verspoor, E., Stradmeyer, L & Nielsen, J.L., eds). Blackwell Publishing, Oxford. pp 17-56.

Wennevik, V., Skaala, Ø,. Titov, S. F., Studyonov, I. and Nævdal, G. 2004. Microsatellite variation in populations of Atlantic salmon from North Europe. Environmental Biology of Fishes, 69: 143-152

Table 1 Amplicons sequenced and levels of polymorphism observed.

Region	Amplicon	Read size	5' base position	Number of SNPs	SNPs per base	Number of Haplotypes	Haplotypes per SNP
DLOOP	1	381	637 to 1059	17	0.044619	9	0.5294
ND1	2	384	3838 to 4260	10	0.026042	8	0.8000
	3	369	4248 to 4654	5	0.013550	5	1.0000
	4	324	4635 to 4998	7	0.021605	6	0.8571
ND2	5	361	5110 to 5510	10	0.027701	7	0.7000
	6	346	5490 to 5879	6	0.017341	3	0.5000
COXI	7	372	6942 to 7351	9	0.024194	5	0.5556
	8	382	7340 to 7762	9	0.023560	8	0.8889
COXII	9	361	8193 to 8594	5	0.013850	4	0.8000
	10	311	8561 to 8907	6	0.019293	8	1.3333
ATP6	11	375	9238 to 9651	11	0.029333	8	0.7273
ND3	12	357	10623 to 11025	9	0.025210	8	0.8889
ND4	13	363	11146 to 11546	8	0.022039	8	1.0000
	14	361	11534 to 11935	11	0.030471	11	1.0000
	15	370	11912 to 12326	13	0.035135	5	0.3846
ND5	16	345	14309 to 14701	7	0.020290	5	0.7143
	17	370	14680 to 15091	10	0.027027	8	0.8000
CYTB	18	366	15376 to 15779	7	0.019126	6	0.8571
	19	352	15765 to 16160	4	0.011364	4	1.0000
	20	365	16133 to 16537	8	0.021918	7	0.8750
Overall		7215	1 to 16,665 bp	172	0.023839	139	0.8081

1 **Table 2** Frequencies of haplotypes of observed in samples.

River	N=	Haplotype: frequency										
Neva	12	**66**: 0.167	**67**: 0.167	**71**: 0.167	**72**: 0.083	**74**: 0.083	**75**: 0.083	**76**: 0.167	**77**: 0.083			
Pechora	12	**66**: 0.167	**67**: 0.167	**78**: 0.250	**79**: 0.333	**80**: 0.083						
Pongoma	12	**40**: 0.083	**43**: 0.333	**44**: 0.083	**66**: 0.500							
Rynda	12	**16**: 0.083	**31**: 0.083	**38**: 0.083	**67**: 0.250	**96**: 0.083	**119**: 0.083	**129**: 0.083	**136**: 0.250			
Teno	12	**16**: 0.083	**39**: 0.083	**47**: 0.083	**67**: 0.167	**96**: 0.167	**98**: 0.083	**105**: 0.083	**113**: 0.083	**137**: 0.083	**138**: 0.083	
Kolmogorov	11	**16**: 0.182	**35**: 0.091	**36**: 0.091	**37**: 0.091	**73**: 0.091	**127**: 0.2727	**128**: 0.091	**130**: 0.091			
Namsen	12	**6**: 0.083	**6/7**: 0.083	**7**: 0.167	**27**: 0.083	**58**: 0.083	**59/60**: 0.083	**90**: 0.083	**96**: 0.083	**118**: 0.083	**123**: 0.083	**124**:
Eiravassdaget	12	**4**: 0.083	**4/5**: 0.083	**5**: 0.083	**16**: 0.167	**33**: 0.083	**59**: 0.083	**59/60**: 0.167	**66**: 0.083	**96**: 0.083	**121**: 0.083	
Bjerkreimselva	12	**31**: 0.083	**32**: 0.083	**33**: 0.083	**48**: 0.083	**49**: 0.333	**81**: 0.083	**98**: 0.083	**108**: 0.167			
Numendalslagen	11	**16**: 0.273	**34**: 0.091	**59**: 0.091	**60**: 0.091	**96**: 0.091	**121**: 0.091	**122**: 0.091	**123/124**:	**125**: 0.091		
Tweed	12	**9**: 0.083	**10**: 0.083	**16**: 0.083	**27**: 0.083	**56**: 0.167	**57**: 0.083	**61**: 0.083	**94**: 0.083	**96**: 0.167	**98**: 0.083	
North Esk	11	**14**: 0.182	**15**: 0.091	**83**: 0.273	**95**: 0.091	**107**: 0.091	**114**: 0.091	**120**: 0.091	**126**: 0.091			
Ugie	10	**2**: 0.100	**26**: 0.100	**86**: 0.100	**88**: 0.100	**91**: 0.100	**104**: 0.100	**116**: 0.100	**117**: 0.200			
Oykel	12	**16**: 0.250	**27**: 0.083	**66**: 0.167	**83**: 0.083	**87**: 0.083	**96**: 0.083	**109**: 0.083	**112**: 0.167			
Laxford	11	**1**: 0.182	**3**: 0.091	**8**: 0.091	**16**: 0.091	**20**: 0.091	**83**: 0.091	**106**: 0.091	**110**: 0.182	**111**: 0.091		
North Uist	12	**12**: 0.083	**16**: 0.333	**24**: 0.167	**25**: 0.083	**96**: 0.250	**102**: 0.083					
Awe	12	**11**: 0.250	**55**: 0.167	**81**: 0.250	**83**: 0.083	**96**: 0.083	**107**: 0.083	**109**: 0.083				
Feochan	12	**1**: 0.250	**13**: 0.083	**23**: 0.083	**56**: 0.083	**67**: 0.083	**96**: 0.083	**101**: 0.083	**102**: 0.083	**103**: 0.167		

River	N=	Haplotype: frequency										
Stinchar	12	**15**: 0.167	**16**: 0.083	**19**: 0.083	**56**: 0.250	**70**: 0.167	**96**: 0.083	**101**: 0.167				
Eden	12	**50**: 0.083	**63**: 0.083	**66**: 0.083	**67**: 0.167	**70**: 0.167	**100**: 0.4167					
Conwy	12	**16**: 0.167	**17**: 0.083	**65**: 0.167	**67**: 0.083	**69**: 0.083	**96**: 0.250	**99**: 0.083	**115**: 0.083			
Blackwater	12	**16**: 0.250	**54**: 0.083	**62**: 0.083	**63**: 0.083	**64**: 0.083	**67**: 0.083	**83**: 0.083	**94**: 0.083	**96**: 0.167		
Taw	6	**16**: 0.167	**20**: 0.167	**70**: 0.167	**82**: 0.167	**97**: 0.167	**98**: 0.167					
Teign	6	**18**: 0.167	**22**: 0.167	**54**: 0.333	**93**: 0.167	**135**: 0.167						
Elorn	12	**16**: 0.083	**51**: 0.083	**52**: 0.083	**53**: 0.083	**54**: 0.333	**67**: 0.083	**89**: 0.083	**96**: 0.083	**100**: 0.083		
Loire-Allier	12	**21**: 0.250	**67**: 0.083	**68**: 0.333	**134**: 0.333							
Ason	12	**16**: 0.583	**23**: 0.083	**92**: 0.083	**95**: 0.083	**96**: 0.167						
Hofsa	12	**41**: 0.083	**46**: 0.083	**74**: 0.083	**84**: 0.083	**85**: 0.083	**98**: 0.083	**131**: 0.167	**132**: 0.083	**133**: 0.250		
Olfusa	12	**15**: 0.083	**16**: 0.083	**28**: 0.083	**29**: 0.083	**30**: 0.083	**42**: 0.083	**45**: 0.083	**46**: 0.4167			

Table 3 Corrected average pairwise differences in base composition among haplotypes within (diagonal) and among populations (below diagonal), and the significance of differences among populations (above diagnonal) – ns: not significant, #: significant, **: significant after Bonferroni correction for multiple tests.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Neva	1.9	**	#	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	#	**	**	**	**	**	**	**	**	**	**
Pongoma	1.0	3.0	**	**	**	**	#	**	**	**	**	**	**	#	**	**	**	**	**	**	**	**	**	#	**	**	**	**	**
Pechora	0.3	1.1	1.4	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**
Rynda	4.0	3.9	4.5	31.0	ns	**	#	#	ns	ns	ns	ns	ns	ns	ns	**	ns	ns	#	ns	ns	ns	ns	ns	#	#	#	**	#
Teno	2.2	1.9	2.5	-1.3	23.2	**	#	#	#	#	ns	ns	ns	ns	ns	**	#	ns	#	ns	ns	ns	#	ns	**	#	#	**	#
Komogorov	4.0	3.5	4.6	4.2	3.5	9.6	**	#	**	**	#	**	**	**	#	#	#	**	**	**	**	#	**	#	**	**	#	**	**
Namsen	1.7	1.4	2.3	3.0	1.3	2.4	5.5	#	ns	ns	#	ns	**	ns	ns	#	ns	ns	**	ns	ns	ns	ns	ns	#	#	#	**	**
Eirva	1.5	1.0	2.0	3.2	1.6	2.4	-0.1	5.1	**	#	**	#	**	#	ns	ns	**	ns	**	**	#	ns	#	#	**	**	ns	**	**
Bjerkriemselvaa	3.0	1.8	3.5	3.8	1.9	2.7	0.7	0.7	4.8	ns	ns	ns	#	ns	ns	ns	ns	ns	#	ns	ns	ns	ns	ns	#	**	ns	**	#
Numendalslagen	1.2	1.2	1.8	3.1	1.6	2.4	-0.1	-0.1	1.0	5.4	#	ns	**	ns	ns	#	#	ns	#	ns	ns	ns	ns	ns	#	#	#	**	**
Tweed	1.4	1.5	1.9	2.8	1.3	1.4	0.3	0.4	1.1	0.3	8.2	ns	ns	ns	ns	**	ns	ns	#	ns	ns	ns	ns	ns	#	#	#	**	#
NorthEsk	1.3	1.2	1.9	2.9	1.1	2.3	0.0	0.1	0.7	0.1	0.0	4.9	#	ns	ns	ns	ns	ns	#	ns	ns	ns	ns	ns	ns	#	ns	**	**
Ugie	2.3	2.1	2.9	3.0	1.3	3.1	0.6	1.0	1.2	1.0	0.5	0.6	6.3	ns	ns	#	#	ns	**	ns	#	ns	#	#	**	#	**	**	**
Oykel	1.2	0.8	1.6	2.9	0.9	2.3	0.1	0.2	0.5	0.2	0.1	-0.1	0.4	4.7	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	#	#	ns	**	#
Laxford	2.5	1.6	3.0	3.0	1.2	2.4	0.1	0.4	0.5	0.5	0.4	0.2	0.4	0.1	5.1	ns	#	ns	#	ns	ns	ns	ns	ns	**	#	ns	**	**
NorthUist	2.7	1.6	3.3	3.4	1.5	2.4	0.3	0.4	0.3	0.6	0.7	0.3	0.8	0.2	0.1	3.2	#	ns	**	ns	#	ns	ns	#	**	**	ns	**	**
Orchy	1.3	1.5	1.6	3.1	1.5	2.6	0.3	0.4	1.2	0.2	0.0	0.0	0.7	0.1	0.6	0.7	5.4	ns	ns	ns	ns	ns	ns	ns	#	#	#	**	**
Feochan	1.9	1.4	2.4	3.0	1.0	2.4	0.0	0.3	0.5	0.3	0.3	-0.1	0.3	-0.1	-0.1	0.1	0.3	4.3	ns	ns	ns	ns	ns	ns	**	#	ns	**	**
Stinchar	1.3	1.4	1.9	3.0	1.3	2.5	0.1	0.2	0.8	0.1	0.1	-0.2	0.7	0.0	0.3	0.4	0.1	0.0	4.2	ns	ns	ns	ns	#	#	**	**	**	**
Eden	1.0	1.5	1.4	3.2	1.2	3.2	0.7	0.9	1.6	0.6	0.5	0.4	0.8	0.3	1.0	1.1	0.4	0.5	0.4	3.5	ns	ns	ns	ns	**	#	ns	**	**
Conwy	1.3	1.3	1.8	2.9	1.0	2.5	0.1	0.3	0.8	0.2	0.2	0.0	0.4	-0.1	0.2	0.4	0.2	-0.1	0.0	0.1	4.4	ns	ns	ns	ns	#	**	**	**
Blackwater	0.9	1.1	1.4	2.9	1.2	2.3	0.1	0.1	1.0	0.0	-0.1	-0.2	0.7	0.0	0.4	0.6	-0.1	0.1	0.0	0.2	-0.1	4.4	ns	ns	ns	ns	ns	**	#

Taw	1.1	0.9	1.5	2.5	0.7	2.2	-0.1	0.0	0.5	0.0	-0.1	-0.2	0.2	-0.3	-0.2	0.0	0.0	-0.3	-0.2	0.0	-0.3	-0.2	4.7	ns	**	#	ns	**	**
Teign	1.6	1.5	2.1	2.9	1.5	2.4	0.2	0.2	1.0	0.2	0.2	-0.1	0.9	0.3	0.6	0.7	0.1	0.4	0.2	0.8	0.3	-0.2	0.1	10.1	ns	ns	#	**	**
Elorn	1.5	2.2	2.0	4.0	2.3	3.6	1.0	1.0	2.3	0.8	0.8	0.4	1.8	1.0	1.6	1.8	0.6	1.2	0.8	1.0	1.0	0.2	0.8	-0.2	4.3	**	**	**	**
Allier	2.5	2.6	3.0	4.0	2.8	3.5	1.8	1.7	2.6	1.6	1.6	1.6	2.1	1.7	2.2	2.2	1.7	1.8	1.6	2.0	1.7	1.5	1.5	0.3	2.4	10.1	**	**	**
Ason	2.6	1.5	3.2	3.4	1.5	2.4	0.2	0.4	0.2	0.6	0.6	0.3	0.8	0.1	0.1	-0.1	0.7	0.1	0.4	1.1	0.3	0.5	0.0	0.6	1.8	2.1	2.5	**	**
Olfusa	4.0	2.2	4.5	4.5	2.9	3.3	1.5	1.4	1.1	1.8	2.1	1.7	2.6	1.5	1.4	1.0	2.3	1.6	1.8	3.1	2.0	2.0	1.6	1.9	3.4	3.4	0.9	2.1	**
Hofsa	2.2	2.4	2.7	3.2	2.0	3.9	1.6	1.6	2.4	1.5	1.2	1.2	2.1	1.3	1.9	2.1	1.4	1.7	1.3	1.9	1.6	1.2	1.3	1.5	1.9	3.0	2.0	2.9	7.5

9

Table 4 Results of AMOVA analysis for within group and among group variation of haplotype frequencies; groups are defined in text.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	15	256.682	0.67135 (Va)	16.68
Within groups	13	48.901	0.04287 (Vb)	1.07
Within populations	301	985.680	3.27469 (Vc)	82.09
Total	329	1291.264	3.98891	

Figure 1 Map of the locations of rivers from which samples were analysed; heavy lines delineate regional groupings of samples used for stratified resampling – see text.

Figure 2 The number of locations with a haplotype plotted against the total number observed for that haplotype, for all 139 haplotype detected in samples; numbers indicate number of haplotypes with a given value.

Figure 3 Minimum evolution (ME) tree of the relatedness of the haplotypes based on number of pair-wise differences; the most common haplotypes are highlighted.

Figure 4 Cumulative number of haplotypes defined with the sequential addition of amplicons clockwise from D-loop to CytoB gene; based on data for 26 of 29 locations; the best fit curve shown is a second order polynomial.

Figure 5 Relationship between number of populations and number of haplotypes, based on a geographically structured re-sampling of the 29 populations; the best fit curve shown is a second order polynomial.

Figure 6 Rarefaction curves for individual samples showing the relationship between samples size and haplotype diversity. Curves shown are mean and standard deviation

Figure 7 Relative degree of similarity between geographically neighbouring samples based on mean pairwise differences between haplotypes in samples. Dotted lines show separation of samples into regional groups based on microsatellite data set (Gilbey et al. submitted).

Figure 8 Minimum evolution (ME) tree of the relatedness of the populations based on number of pair-wise differences.

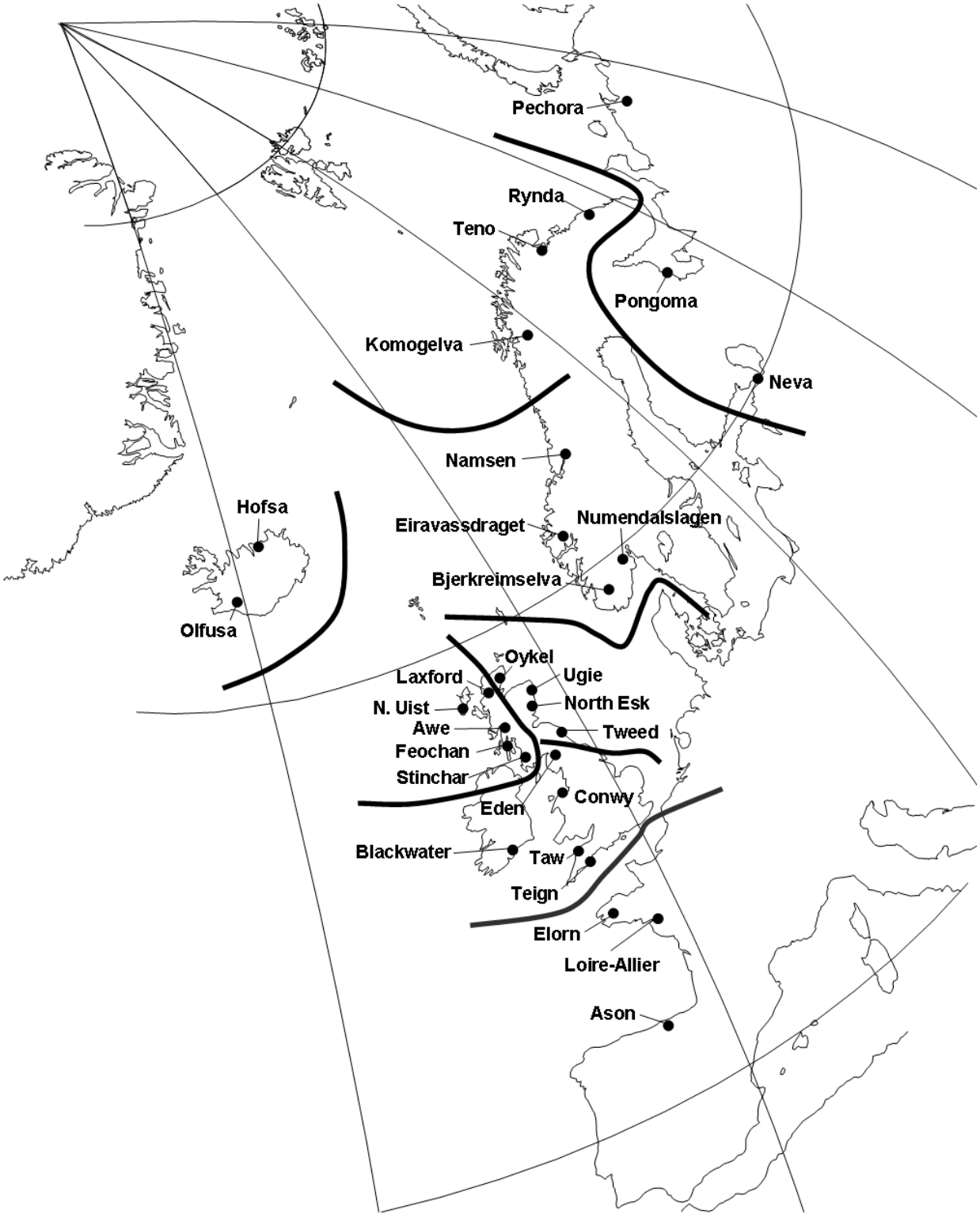

Pechora
Rynda
Teno
Pongoma
Komogelva
Neva
Namsen
Hofsa
Eiravassdraget
Numendalslagen
Bjerkreimselva
Olfusa
Oykel
Laxford
Ugie
N. Uist
North Esk
Awe
Tweed
Feochan
Stinchar
Conwy
Eden
Blackwater
Taw
Teign
Elorn
Loire-Allier
Ason

Fig 1

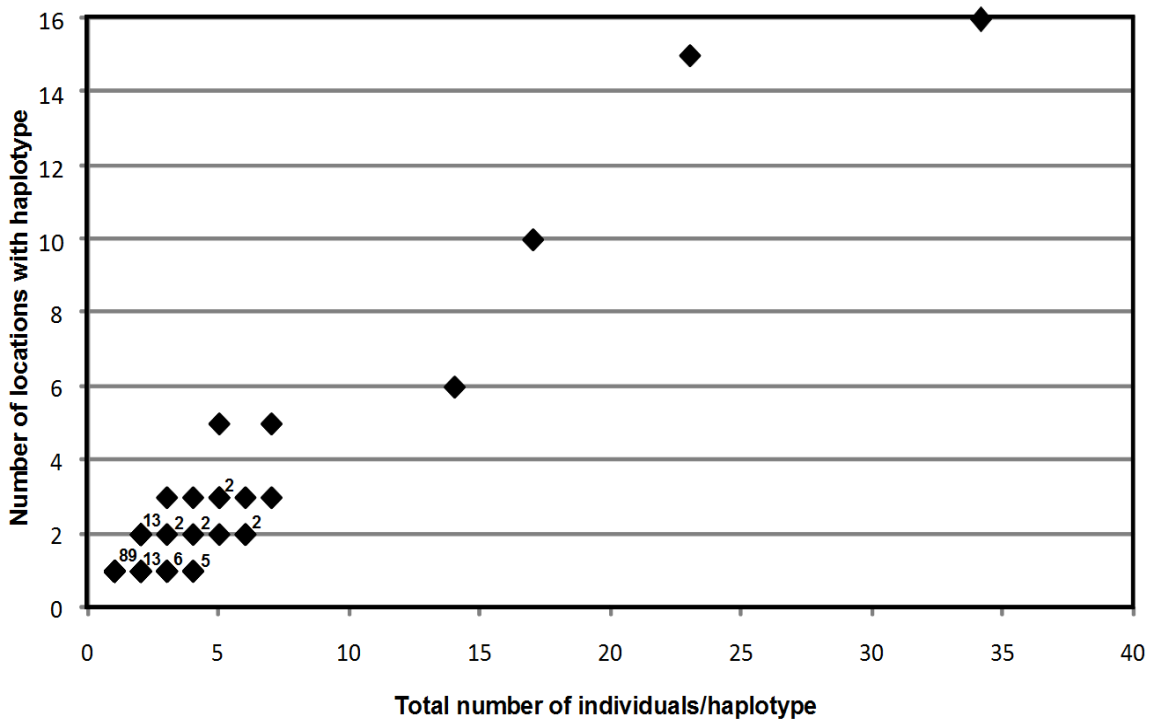

Number of locations with haplotype
Total number of individuals/haplotype
0
2
4
6
8
10
12
14
16
0
5
10
15
20
25
30
35
40
89
13
13
2
2
2
6
5
2

Fig 2

Fig 3

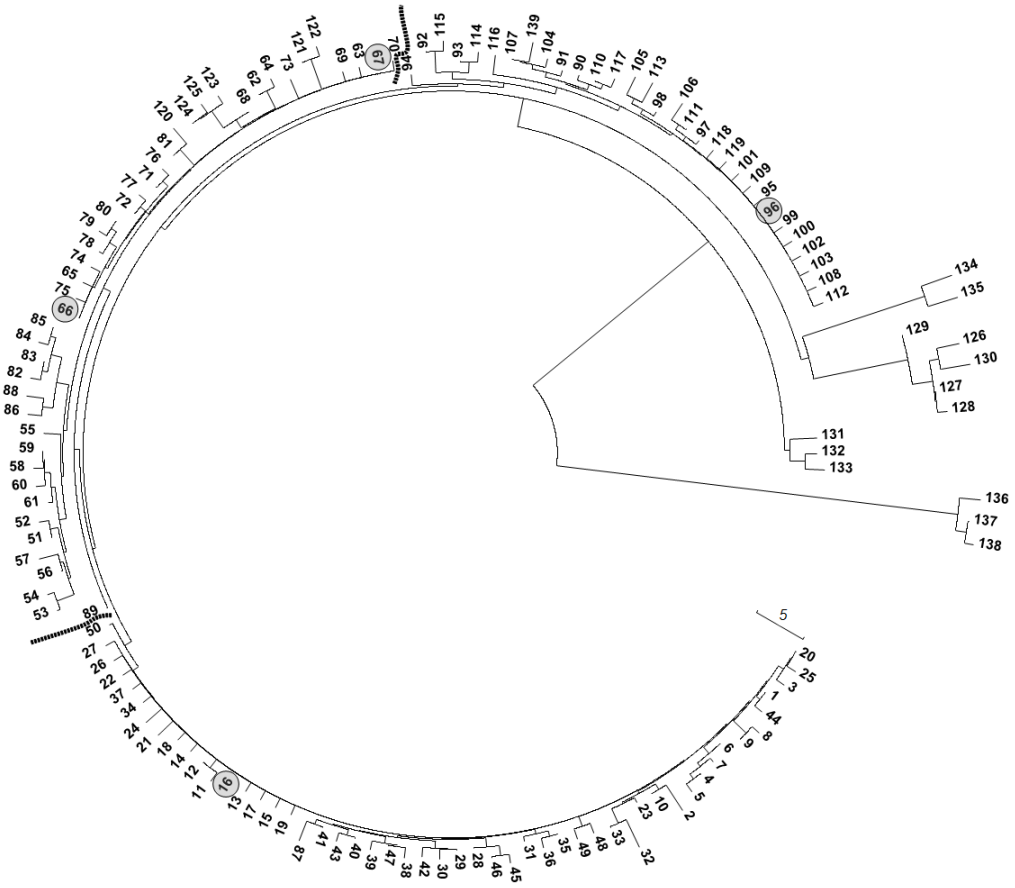


Fig 4

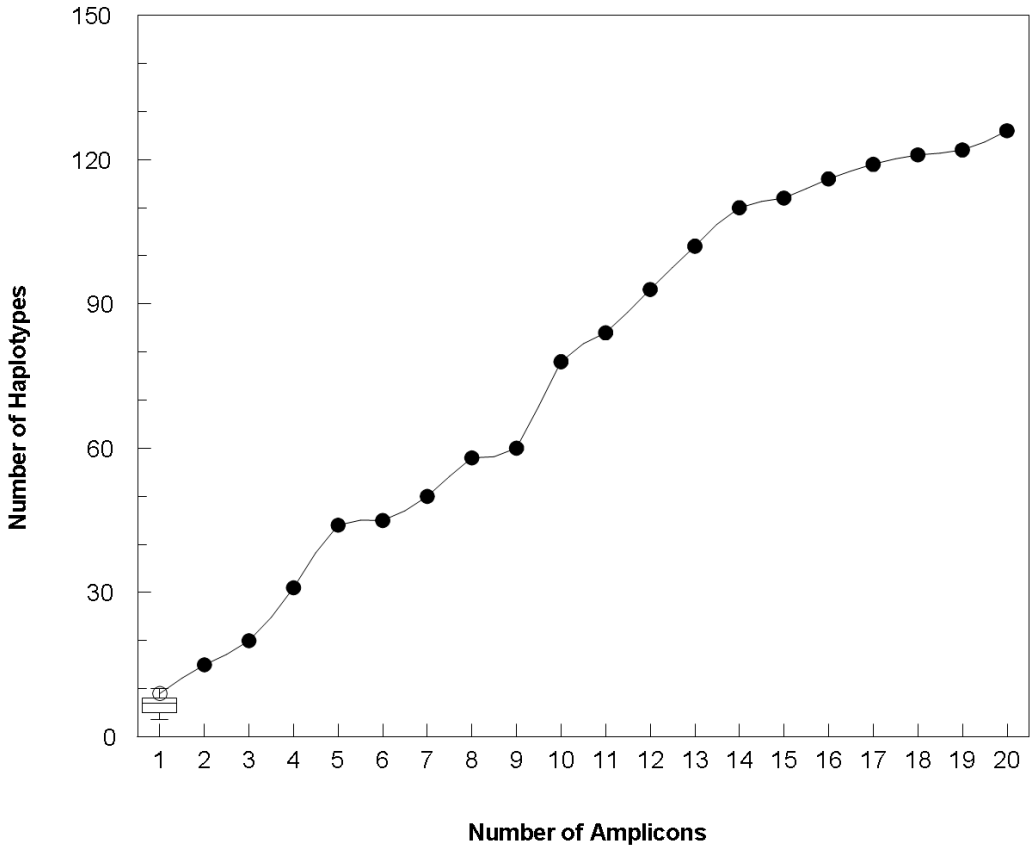

150
120
90
60
30
0
Number of Haplotypes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of Amplicons

Fig
5

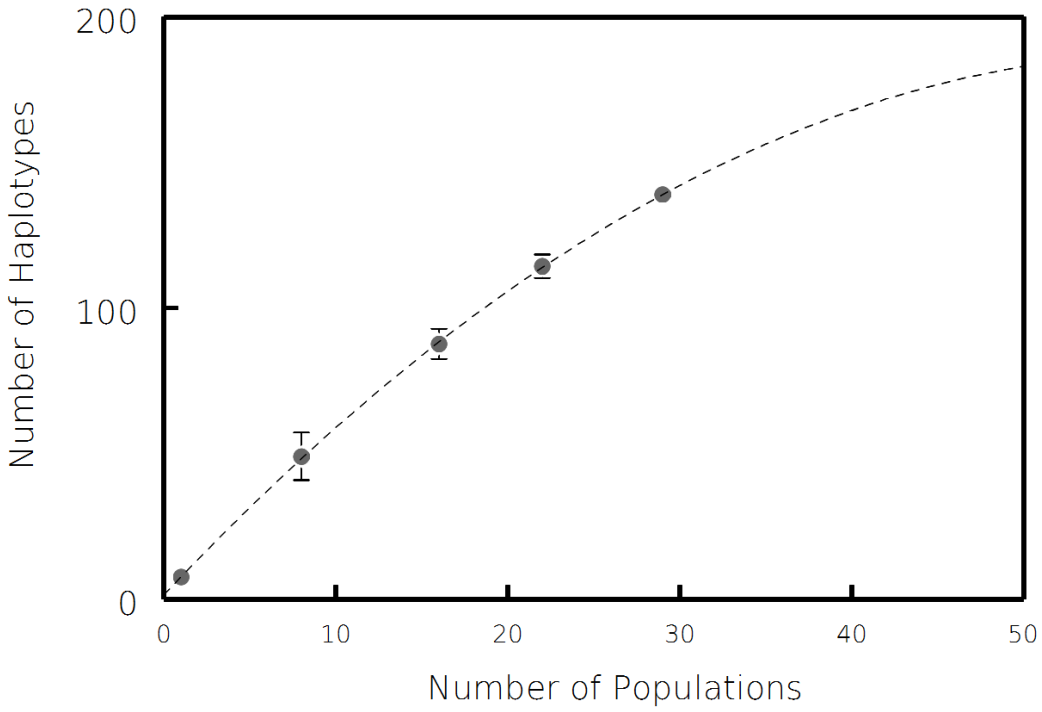

200
100
0
Number of Haplotypes
0
10
20
30
40
50
Number of Populations

Fig 6

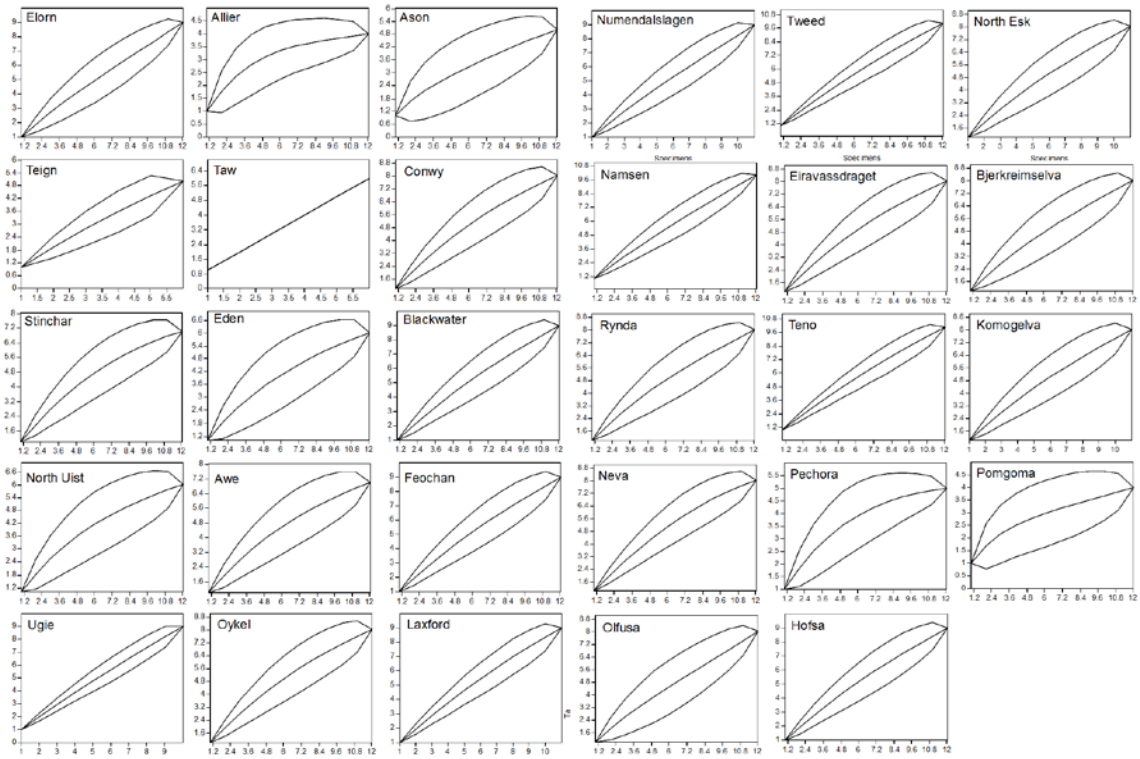

Elorn
Allier
Ason
Numendalslagen
Tweed
North Esk
Teign
Taw
Conwy
Namsen
Eiravassdraget
Bjerkreimselva
Stinchar
Eden
Blackwater
Rynda
Teno
Komogelva
North Uist
Awe
Feochan
Neva
Pechora
Pomgoma
Ugie
Oykel
Laxford
Olfusa
Hofsa

Fig 7

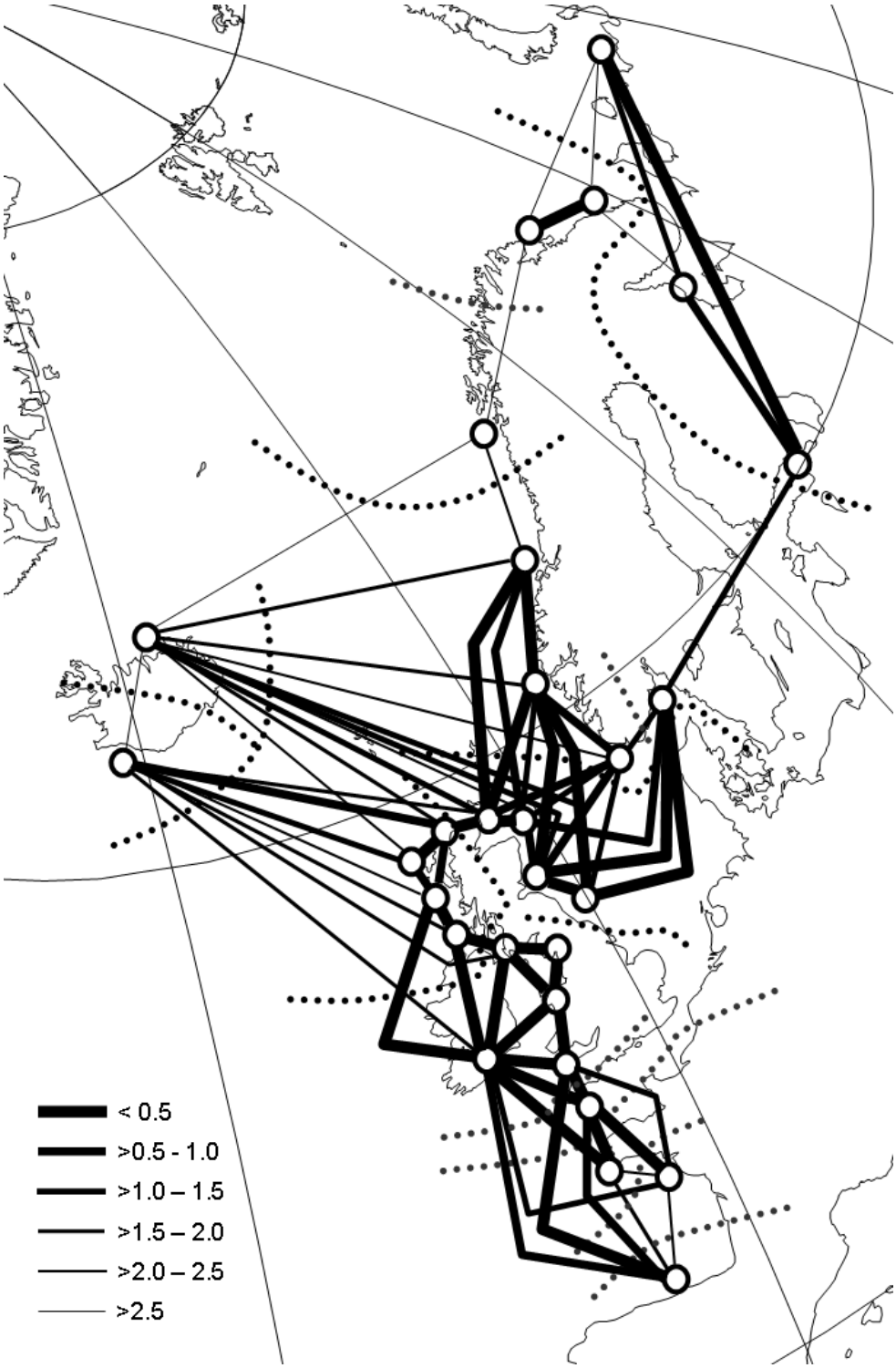
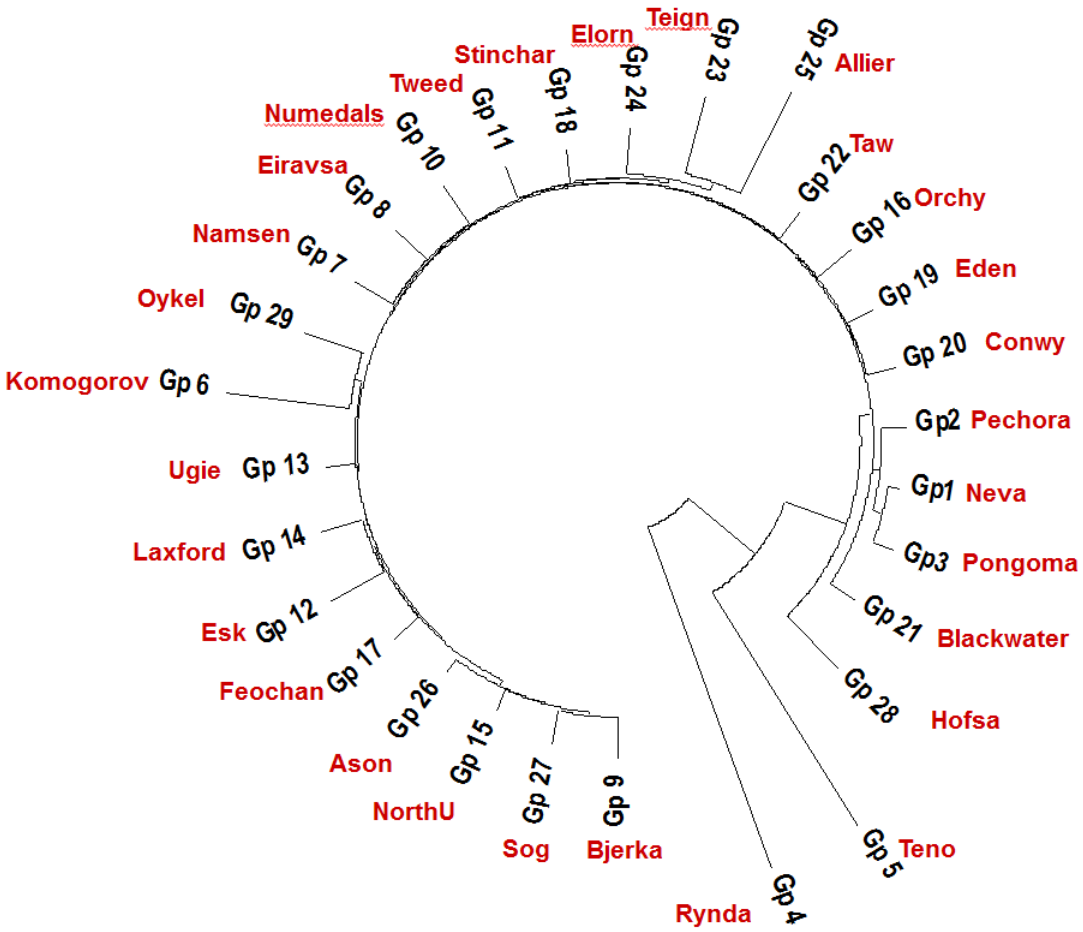


Fig 8


Teign
Elorn
Stinchar
Tweed
Numedals
Eiravsa
Namsen
Oykel
Komogorov
Ugie
Laxford
Esk
Feochan
Ason
NorthU
Sog
Bjerka
Rynda
Teno
Hofsa
Blackwater
Pongoma
Neva
Pechora
Conwy
Eden
Orchy
Taw
Allier
Gp 23
Gp 25
Gp 24
Gp 18
Gp 11
Gp 10
Gp 8
Gp 7
Gp 29
Gp 6
Gp 13
Gp 14
Gp 12
Gp 17
Gp 26
Gp 15
Gp 27
Gp 9
Gp 4
Gp 5
Gp 28
Gp 21
Gp3
Gp1
Gp2
Gp 20
Gp 19
Gp 16
Gp 22

III. Nuclear DNA SNPs (Subtask 1.3.4)

Identify and develop nDNA SNPS

***Overview*:** *Known nDNA sequences identified in electronic databases such as GENBANK and cGRASP will be reviewed and the most promising selected for screening for SNPs. Based on screening of the reference collection assembled in Subtask 1.3.1, a suite of 100 SNPs which provide useful regional assignment capability will be chosen and optimal conditions for typing established.*

Participants: Lead – 6; other 2,7

1. Identification and evaluation of an arbitrary selection of existing nuclear SNPs

Jamie Coughlan1, J-P Vähä2, Paul R. Berg3, Paulo A Prodöhl4, John Gilbey5, Jens Carlsson1, Phil McGinnity1, Dennis Ensing6, Sigbjørn Lien7, Craig Primmer2, Eric Verspoor5, Vidar Wennevik8, and Tom Cross1

1 School of Biological, Earth and Environmental Sciences, University College Cork, IRELAND 2 Department of Biology, University of Turku, FINLAND 3 CEES, University of Oslo, NORWAY 4 School of Biological Sciences, Queen's University Belfast, NORTHERN IRELAND 5 Marine Scotland - Science, Freshwater Laboratory Pitlochry, SCOTLAND 6 Fisheries & Aquatic Ecosystems Branch, Agri-Food & Biosciences Institute, NORTHERN IRELAND 7 CIGENE, Norwegian University of Life Sciences, NORWAY 8 Institute of Marine Research, Bergen, NORWAY

The identification of nuclear single nucleotide polymorphism (nSNPs) loci useful for assignment was undertaken in collaboration with CIGENE (Norway). A panel of 388 EST-DNA derived SNP loci were optimised for screening (using multiplexes) on the Sequenom platform. Samples from 84 rivers/locations across the species range were screened (Figure 1 and Table 1) for an average of 5.5 individuals (range 1-26) from each river. Also, included were nine individuals sampled from an Irish fish farm (strain of Norwegian ancestry) and four known salmon/trout hybrids (based on microsatellite DNA profiles). This was a specially designed reference collection of samples rather than those assembled in Subtask 1.3.3.

Data quality was extensively tested and resulted in the loss of approximately 5% of samples (due to poor DNA concentration/quality) which left 477 samples for further analysis. Examination of individual SNP loci revealed that 305/388 loci worked consistently and showed polymorphisms among the total data set. The remaining loci appeared to be monomorphic in all samples screened or failed to resolve genotypes in at least 90% of the samples (perhaps due to technical difficulties associated with multiplexing) and these were excluded from further analysis.

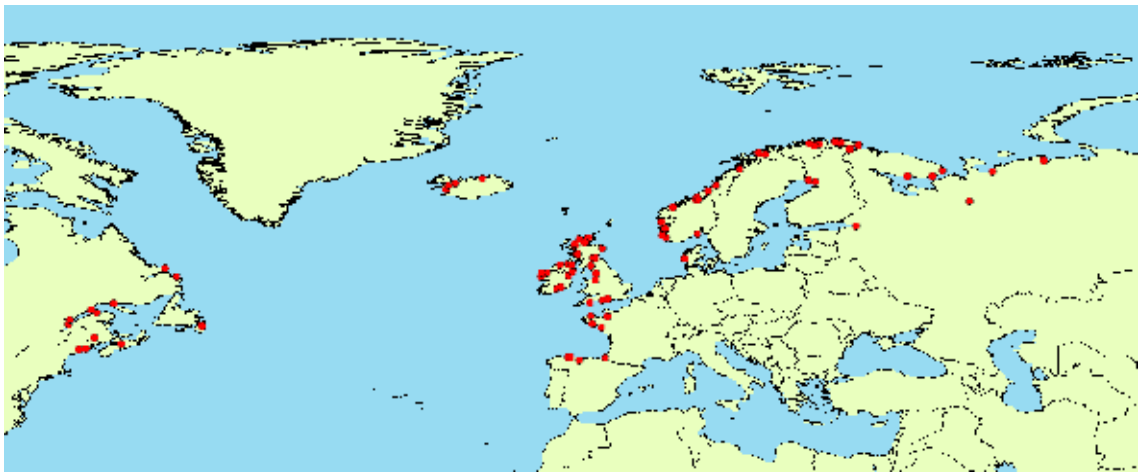


Figure 1 Approximate locations of origin of Atlantic salmon samples used in study

Bayesian and non-parametric clustering techniques were used to identify genetic structures and regionality among the samples. The STRUCTURE and BAPS software packages revealed five and eleven major genetic clusters, respectively. STRUCTURE analysis identified genetic clusters that corresponded to geographically defined regions of 1) North American, 2) Icelandic, 3) Baltic/Russian/northern Norwegian, 4) western Norwegian/Swedish and 5) Danish/British Isles/French/Spanish groups (Figure 2). BAPS defined similar genetic/regional clusters which were partitioned samples into 8 eight major groups (1 -mainland North America, 2 - Iceland, 3 -Baltic, 4 - Kola Peninsula, 5 - North Norway & Russia, 6 - West Norway & Sweden, 7 - extended British Isles (Denmark/Britain/Ireland/Northern France) and 8 - Southern France & Spain) and three minor genetic groups (two of which were river-specific (Sandhill in Newfoundland and Pechora in Russia) and one of which was composed of hybrid samples) (see Figure 3).

Non-parametric clustering methods revealed similar genetic clusters as the Bayesian approaches although these were less well defined and therefore excluded. For both STRUCTURE and BAPS analysis, the identified broad genetic clusters/regions are in agreement with previous findings using microsatellites and mtDNA and other genetic markers. Because of the higher number of clusters detected using the BAPS software, these regions were used for further testing of variability and usefulness for of SNPs assignment.

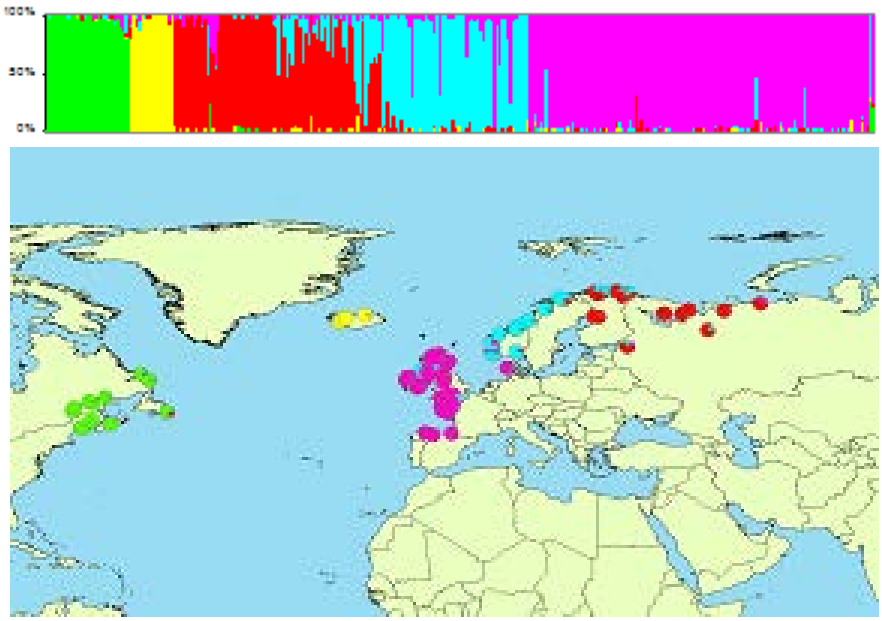


Figure 2 Estimated genetic structure (revealed by STRUCTURE) where each individual is partitioned into five clusters. The map shows the geographical distribution of individual partitioning (averaged for each river).

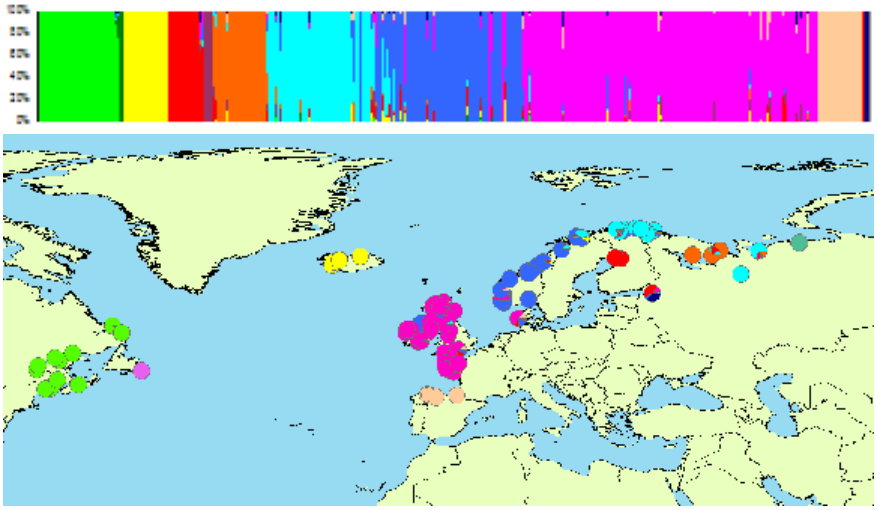


Figure 3 Estimated genetic structure (revealed by BAPS) where each individual is partitioned into one of eleven clusters. The map shows the geographical distribution of individual partitioning (averaged for each river).

The geographical distribution of BAPS identified clusters agrees with previous findings using other genetic marker types although it fails to detect some fine-scale structure as

revealed by microsatellites (Subtask 1.3.2) which may be the result of the small number of samples/rivers included here. However, the geographical location of the minor, river-specific clusters Sandhill (Newfoundland) and Pechora (furthest east in salmon distribution), suggest that additional structure could be identified if other rivers and individuals were analysed.

The geographical boundaries of the identified genetic clusters appear to be very robust with most genetically overlapping rivers/samples being typically (although not exclusively) located in geographically adjacent areas (see Figure 3, in particular between the Russian/northern Norway and western Norway/Sweden clusters). It is also noteworthy that Irish farmed salmon appear to cluster best with the western Norway/Sweden region, which is likely to be the result of their Norwegian ancestry. This has important implications for the assignment of farmed escapes (or the progeny of these) to region of natal origin

Variability in terms of potential utility to identify fine-scale population structure in each of the eight major genetic clusters across this panel of 305 loci was also assessed. Three measures of variability were used; number of monomorphic loci, number of loci where the minor allele frequency (MAF) was less than 0.05 and the number of loci where heterozygosity was less than 10%. There were dramatic differences between genetic clusters using all these measures which were lowest in the West Norway & Sweden cluster in all cases (see figure 4 where variability measures are expressed as proportions of the total 305 loci used). Variability was lowest in the North American group and also appeared to be related to geographic distance from western Norway (which was the location of samples used to discover these SNPs.

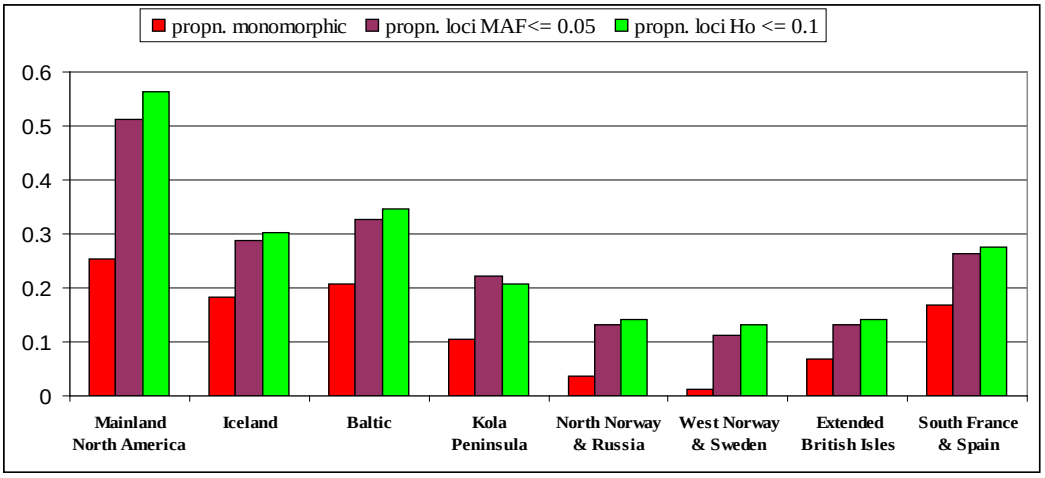


Figure 4 SNP locus variability as revealed by proportion of loci monomorphic, minor allele frequency (MAF)<0.05 and observed heterozygosity <0.01 in each of the eight major genetic clusters

Table 1 Sample details of individuals used in SNP analysis

Country	River	n	Country	River	n
Canada	Malbaie	5	Norway	Etneelva	4
Canada	Michael	3	Norway	Figgjo	3
Canada	Sandhill	1	Norway	Komag	4
Canada	Seal Cove	3	Norway	Lakselv	4
Canada	Ste Anne	5	Norway	Langfjordelva	3
Canada	Stewiacke	1	Norway	Laukhalle	4
Canada	St Jean	5	Norway	Loneelva	4
Canada	St John	7	Norway	Naatamo	8
Canada	Ste Marguerite	5	Norway	Neiden	4
Canada	Trinite	5	Norway	Numedalslagen	10
Denmark	Skjern	1	Norway	Orkla	12
England	Dart	3	Norway	Reppasfjord	4
England	Esk	1	Norway	Saltdalselva	5
England	Frome	4	Norway	Skauga	4
England	Itchen	1	Norway	Stordalselva	5
England	Lune	1	Norway	Tana	26
Finland (Baltic)	Simojoki	5	Russia	Pecha	2
France	Allier	5	Russia	Pechora	5
France	Leguer	1	Russia	Ponoi	14
France	Nivelle	6	Russia	Pulonga	2
France	Scourff	3	Russia	Varzuga	14
France	See	5	Russia	Vigda	5
France	Selune	2	Russia (Baltic)	Neva	3
Iceland	Langa	4	Scotland	Almond	5
Iceland	Laxa i Aldal	5	Scotland	Awe	4
Iceland	Laxa I Dolum	11	Scotland	Coulin	5
Iceland	Nupsa	5	Scotland	Don	4
Ireland	Blackwater	15	Scotland	Ewe	1
Ireland	Boyne	5	Scotland	Halladale	8
Ireland	Burrishoole	5	Scotland	Laxford	10
Ireland	Dawros	5	Scotland	Nith	4
Ireland	Fanad (farmed)	9	Scotland	Orchy	6
Ireland	Moy	11	Scotland	Shin	11
Ireland	Owenmore	7	Spain	Esva	12
Ireland	Suir	10	Spain	Narcea	5
Northern Ireland	Bush	7	Spain	Sella	3
Northern Ireland	Glendun	6	Sweden	Altran	5
Northern Ireland	Shimna	7	Sweden	Ura	4
Norway	Åelva	4	Sweden (Baltic)	Tornijoki	13
Norway	Bjerkreimselva	2	USA	Narragus	4
Norway	Bogna	4	USA	Penobscot	5
Norway	Borselv	4	Wales	Dee	7
Norway	Målselv	4	Multiple	Hybrids	4

The utility of these loci for assignment to region was assessed using GENECLASS software using “leave-one-out” and uniquely “leave-one-river-out” approaches. Individual “Leave-

one-out" assignment among the eight major clusters indicates 97.2% assignment correct to cluster of origin (mis-assignment was typically but not exclusively to rivers on the geographical fringes of the clusters). "Leave-one-river-out" analysis reveals 95.0% correct assignment to region of origin (again most mis-assignment was in overlap or contact areas of cluster distribution). Assignment scores were very high for these samples in both cases (average 99.5). Overall, these genetic clusters are robust for regional assignment.

To assess power of individual loci and identify the most informative SNPs for assignment, loci were ranked according region-specific F_{ST} (e.g. the loci which demonstrated the highest value between the focal region and to the remaining regions pooled) using a hierarchical structure based on the most divergent clusters.

The number of loci required to discriminate at least 95% of samples on a hierarchical basis is shown in Table 2. For example, a single SNP can distinguish North American and Icelandic salmon from the remainder of the baseline in 98.8 and 97.3% of cases, respectively. However, 10-30 loci may be required to discriminate between regions in the lower levels of the hierarchical structure at 95% correct assignment. Also included in Table 2 is total assignment success using the complete set of 305 loci.

Additionally, a number of SNPs were found to be informative at multiple levels of the hierarchical structure. Exploiting the potential for synergistic effects of these loci, region-specific assignment success was estimated (see Figure 5).

Total assignment success clearly improves as more SNPs markers are included. However, the gain in assignment success tends to level-out as increasingly larger numbers of markers are included (e.g. approximately 90% of all samples can be correctly assigned using 51 SNP loci compared to 97.2% with all 305 loci). However, more than 80% of the loci contribute, cumulatively, less than 7.2 per cent-units to total success. Assignment success for different numbers of loci also varies greatly among regions with the most genetically distinct regions requiring the lowest number of loci for maximum correct assignment (e.g. assignment highest among North America, Iceland, Baltic and South France/Spain but lower amongst Kola Peninsula, Russia/North Norway, West Norway and the Extended British Isles for panels of 7-51 loci). However, as overall self-assignment for each region was maximal at 97% (excluding Mainland North America and Iceland that showed 100% assignment) and this was achieved using 305 loci, indicating that the entire locus panel should be used for assigning marine samples using this baseline data.

The SNP analysis of relatively few individuals from across the species range reveals substantial population structure and genetic clusters similar to those previously identified using other genetic markers such as mtDNA and microsatellites. Likely due to smaller sample sizes, this study has failed to resolve some of the fine-scale structure indicated by

Table 2 Number of SNP loci and correct assignment success in terms of discriminating major genetic regions hierarchically

# SNP loci	1	5	10	15	20	25	30	305
North America	98.8							100
Iceland	97.3							100
Baltic	94.4	93.9	97.5					99.2
Kola Peninsula	88.6	88.6	91.2	96.2				99.7
South France/Spain	91.8	92.7	96.5					100
Russia/North Norway	81.1	89.6	92.1	89.1	94.0	94.6	96.2	96.8
West Norway *v* British Isles	76.4	89.4	92.1	94.1	96.1			97.6

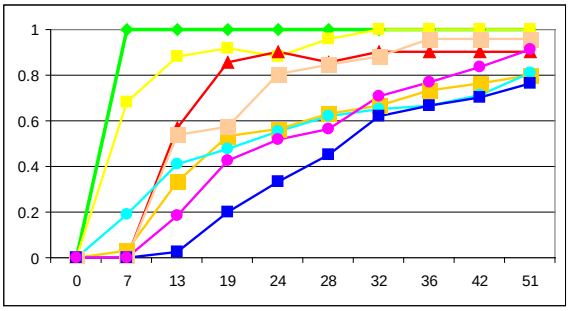


Figure 5 Graph of proportional assignment success (vertical) for each region using specifically selected panels of SNP loci in numbers from one to 60 (horizontal) ●Mainland North America, ● Iceland, ● Baltic, ● Kola Peninsula, ● Russia/North Norway, ● West Norway/Sweden, ● Extended British Isles, ● South France/Spain.

other marker types screened on larger numbers of individuals. These SNP loci show substantial ascertainment bias which is apparent in terms of different levels of genetic variability within clusters. SNP loci seem to be of great utility and reliability (in terms of correctness) for assigning samples to broad geographical region of origin. The number of SNPs required for correct assignment to region of origin can vary from just a few to tens of loci although for this baseline, all available loci were used.

2. Identification and evaluation of nuclear SNP potentially affected by direction selection

Paulo A Prodöhl[1] & Jamie Coughlan[2]

1. School of Biological Sciences, Queen's University Belfast, NORTHERN IRELAND
2. School of Biological, Earth and Environmental Sciences, University College Cork, IRELAND

SNPs affected by directional selection often display high levels of differentiation among populations in comparison to their neutral counterparts and, thus, can be particularly valuable for genetic stock identification. Two parallel approaches were used to identify SNPs that might potentially be influenced by natural selection.

The first approach involved a literature review to identify promising nuclear coding genes for SNP development. A number of candidate genes (e.g. IDH, MEP, transferrin, and 'executioner' caspases) were initially selected. Among these, the NADP dependent malic enzyme (MEP) was chosen for further developmental work as previous protein electrophoresis studies have clearly demonstrated the potential of this gene coding locus for discriminating Atlantic salmon populations and/or major regional groups (e.g. Verspoor 1994, 1997).

The methodological approach employed for identifying and characterising the MEP polymorphism in Atlantic salmon focused on detecting the causative sequence variation underlying the known allozyme polymorphism. It involves: 1) identifying sequences from the gene of interest; 2) designing PCR primers to assay the whole gene; and 3) comparing sequence data from fish of known MEP genotype to identify the mutation. A number of Atlantic salmon EST sequences/contigs were identified in various genomic databases with annotations suggesting relatedness to potential NADP- dependent ME loci. One ASGI Tentative Consensus (TC68582) was selected for further investigation, and a number PCR primers sets designed to span various regions of the tentatively identified ME open reading frame. cDNA template was prepared from RNA extracted from both muscle & liver tissue from a salmon parr (of unknown ME genotype) for testing of the primer pairs. Since there is no information on the number or position of intron/exon boundaries, and no idea of intron sizes, use of genomic DNA was not appropriate.

Two of the three primer sets produced amplicons (see Figure 6). In contrast to the 75 bp PCR primer set, which yielded a single fragment of expected size, the 833 bp primer set gave two distinct bands per tissue (more prominent from muscle cDNA) – though neither band was of the expected size. Among the possible explanations for this discrepancy is the occurrence of different splice variants, the presence of products from more than one locus or may reflect erroneous sequence alignment in the original ASGI TC. Both bands were isolated and sequenced. BLASTN & BLASTX analyses confirmed

that they were both NADP-dependent malic enzyme related sequences. Subsequent sequencing of these regions using fish on known MEP-1 genotype failed to resolve that target polymorphism.

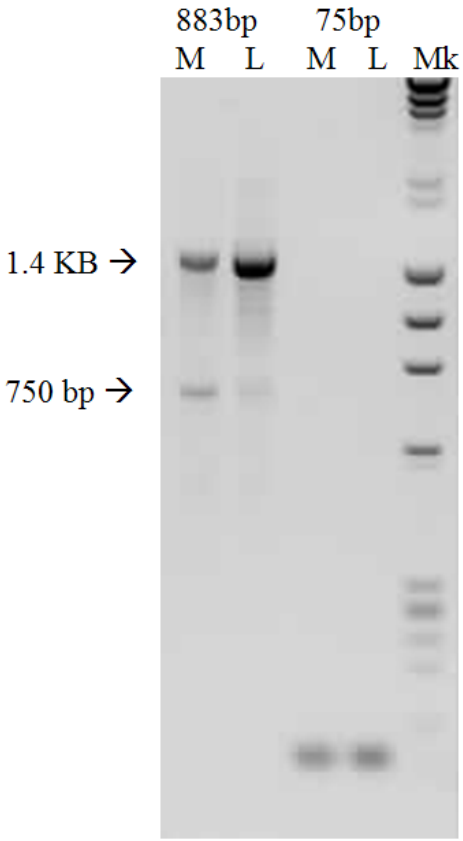


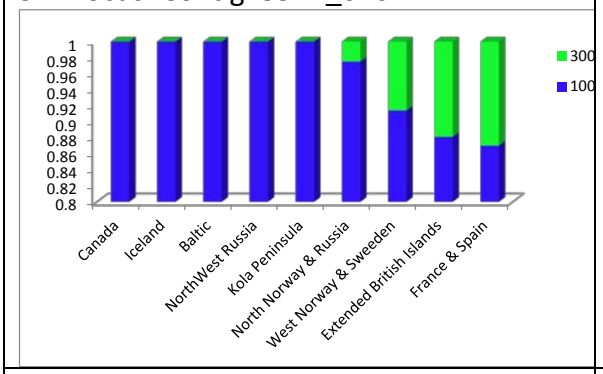
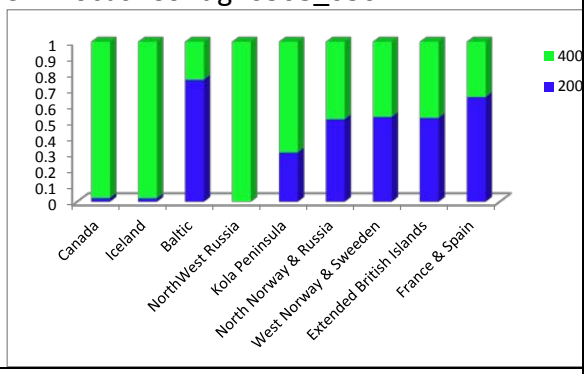
Figure 6 Agarose gel of amplified DNA: to left – M = muscle; L = liver. (Mk is lambda *Hind*III + phiX174 *Hae*III size marker).

The parallel approach to identify SNPs potentially influenced by natural selection proved more successful. The data set consisting of samples from 84 rivers/locations across the species range described earlier (Figure 1 and Table 1) and reliably genotyped for the panel consisting of 306 EST-DNA derived SNP loci provided a unique opportunity to identify SNPs affected by selection. To this end, the F_{ST} outlier approach (Beaumont and Nichols 1996), implemented in the program LOSITAN, was used. Since selection is a population dependent process (i.e. directional changes in the genetic makeup of populations in response to environmental changes), analyses were carried out based upon population sample pairwise comparisons. The following groups, defined in previous SNP STRUCTURE/BAPS analyses, were used for pairwise comparisons: 1)

Canada; 2) Iceland; 3) Baltic; 4) North West Russia; 5) Kola Peninsula; 6) North Norway & Russia; 7) West Norway & Sweden; 8) Extended British Islands and 9) France & Spain).

In total, 36 pairwise comparisons were carried out (i.e. independent runs of LOSITAN). For each run, all outlier loci were recorded into an Excel database. To identify EST-DNA derived SNP loci under selection, the outlier loci were summed over the loci for each population sample pairwise comparison. A particular EST-DNA derived SNP locus was considered to be under selective constrains when it was found to be an outlier in over 25% of the independent population pair-wise comparisons. While this is an 'ad-hoc' approach, it does allow for some measure of confidence (i.e. identification of same outlier SNP locus over multiple 'more or less' independent tests provide confidence in results).

Out of the 306 SNPs loci, 88 SNPs (28.7%) were found to be potentially under the influence of directional selection. Of these, 41 were found to be outliers in pairwise comparisons involving both European and North American samples (i.e. within and between groups), 32 were found to be outliers in pairwise comparisons involving European samples only, and the remaining 15 were found to be outliers in comparisons involving samples between North America and Europe (i.e. between groups only). Examples of allelic frequency distributions among samples for these loci are displayed in Figure 7 All SNP loci identified as outliers are reported in Table 1. Interestingly, in many cases, SNP loci potentially affected by selection displayed an obvious "gradient profile" often related to geographical origin of samples.

<table>
<tr><td>SNP locus: Contig15917_670</td><td>SNP locus: Contig16308_0302</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>SNP locus: BASS129-B7-H01_304</td><td>SNP locus: Contig15447_0553</td></tr>
</table>

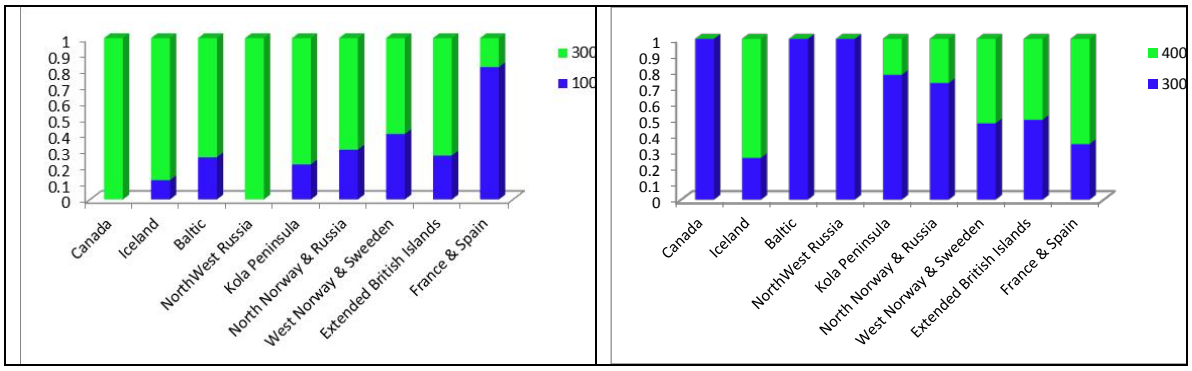


Figure 7 Allelic frequency distribution for 4 of the 90 identified SNP loci potentially under the influence of selection.

Table 1. 88 identified outlier SNP loci potentially under the influence of selection. Empty grey empty cells indicate instances of no evidence for selection for particular group comparison.

Canada v.s. Europe	Europe
	Contig16240_0204
Contig16475_1011	Contig16475_1011
Contig15447_0553	Contig15447_0553
Contig17112_0405	Contig17112_0405
	Contig16654_489
Contig14291_364	
	Contig14058_0333
	Contig14342_489
	BASS114-B7-B09_399
	Contig12386_223
	BASS127-B7-A09_539
	Contig16361_472
Contig17081_268	Contig17081_268
Contig17611_0091	Contig17611_0091
Contig14333_465	
BASS10-B7-F03_268	BASS10-B7-F03_268
	BASS10-B7-F03_425
	Contig16731_596
Contig16780_0490	Contig16780_0490
Contig14161_559	Contig14161_559
Contig15535_270	Contig15535_270
Contig16677_0620	
Contig15097_0126	Contig15097_0126
Contig15674_404	Contig15674_404
BASS119-B7-A09_482	BASS119-B7-A09_482
	Contig17291_925
Contig14157_248	
Contig14157_291	Contig14157_291
	Contig16109_0535
	Contig16109_0600
Contig13513_0493	Contig13513_0493
BASS17-B7-C04_472	BASS17-B7-C04_472
Contig14800_360	Contig14800_360
Contig13981_262	Contig13981_262
	Contig16053_552

	Contig15230_258
BASS129-B7-H01_304	BASS129-B7-H01_304
Contig13800_0155	Contig13800_0155
Contig15119_356	Contig15119_356
BASS138-B7-B10_225	BASS138-B7-B10_225
BASS138-B7-B10_200	BASS138-B7-B10_200
Contig16466_1044	Contig16466_1044
BASS111-B7-D03_407	BASS111-B7-D03_407
Contig13218_0324	Contig13218_0324
BASS120-B7-F09_587	BASS120-B7-F09_587
Contig16938_271	
Contig16938_888	
Contig14634_0088	Contig14634_0088
	BASS141-B7-B10_480
BASS113-B6A-F03_685	BASS113-B6A-F03_685
	Contig17429_1139
Contig16973_559	
Contig16967_0522	Contig16967_0522
Contig16855_383	Contig16855_383
	Contig15118_153
Contig12810_0361	Contig12810_0361
Canada v.s. Europe	Europe
Contig17071_586	Contig17071_586
Contig15486_0436	
Contig16034_0989	Contig16034_0989
Contig16686_0431	Contig16686_0431
Contig15393_948	Contig15393_948
Contig16308_0302	Contig16308_0302
	Contig13579_599
	BASS117-B7-E02_480
	Contig16260_0757
	Contig15690_536
BASS139-B7-E01_105	BASS139-B7-E01_105
	Contig14035_0356
Contig16763_445	
	Contig16532_248
Contig16378_0529	
	Contig13615_543
Contig16221_0769	Contig16221_0769
Contig15360_138	Contig15360_138
	Contig15360_434
	Contig15482_813
BASS112-B7-H03_670	
BASS126-B7-D12_1074	
	Contig15918_614
Contig15977_0274	Contig15977_0274
Contig17164_89	
BASS111-B7-D09_174	BASS111-B7-D09_174
	BASS115-B6A-H02_496
BASS121-B7-G07_759	
	BASS132-B7-C12_203
	Contig14835_0237
	Contig16055_561
Contig16405_0153	

To verify the usefulness of the panel comprised of 88 nuclear SNPs potentially under the influence of directional selection to identify genetic structures and regional groupings among the samples, the Bayesian approach implement within the programme STRUCTURE was used as for the whole SNP panel described above. The genetic clusters/regions identified with this subpanel of marker (Figure 8) are virtually identical to those identified with the full maker panel comprising 306 SNPs for both STRUCTURE and BAPS analyses (Figures 2 & 3). It is clear that these markers under directional selection should provide more cost effective, useful and precise tool for individual assignment.

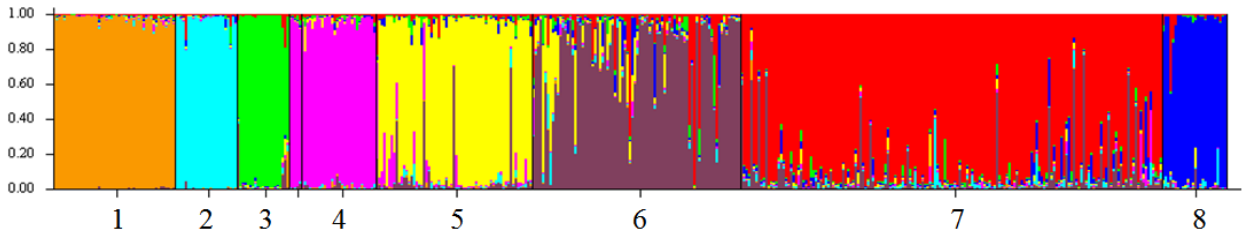


Figure 8 Estimated genetic structure (revealed by STRUCTURE) where each individual is partitioned into eight clusters as follows: 1 - North America, 2 - Iceland, 3 -Baltic, 4 - Kola Peninsula, 5 - North Norway & Russia, 6 - West Norway & Sweden, 7 - extended British Isles (Denmark/Britain/Ireland) and 8 - Southern France & Spain.