ICES Journal of Marine Science Advance Access published March 20, 2012

ICES Journal of Marine Science

ICES International Council for the Exploration of the Sea
CIEM Conseil International pour l'Exploration de la Mer

ICES Journal of Marine Science; doi:10.1093/icesjms/fss029

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

Eric Verspoor[1,2]*, Sonia Consuegra[3], Olafur Fridjonsson[4], Sigridur Hjorleifsdottir[5], David Knox[1], Kristinn Olafsson[4], Scott Tompsett[5], Vidar Wennevik[6], and Carlos Garcia de Leaniz[5]

[1]*Freshwater Laboratory, Marine Scotland, Pitlochry PH16 5LB, Scotland, UK*
[2]*12 Dixon Terrace, Pitlochry PH16 5QX, Scotland, UK*
[3]*Institute of Biological Sciences, University of Wales, Aberystwyth SY23 3DA, UK*
[4]*Matís, Vínlandsleið 12, 113 Reykjavík, Iceland*
[5]*Department of Biosciences, Swansea University, Swansea SA2 8PP, UK*
[6]*Institute of Marine Research, PO Box 1870 Nordnes, 5817 Bergen, Norway*

**Corresponding author: tel: +44 1796 472945; fax: +44 1796 473523; e-mail: eric.salgen@gmail.com.*

Verspoor, E., Consuegra, S., Fridjonsson, O., Hjorleifsdottir, S., Knox, D., Olafsson, K., Tompsett, S., Wennevik, V., and Garciá de Leániz, C. Regional mtDNA SNP differentiation in European Atlantic salmon (*Salmo salar*): an assessment of potential utility for determination of natal origin. – ICES Journal of Marine Science, doi:10.1093/icesjms/fss029.

Received 7 September 2011; accepted 9 January 2012.

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 for more than a decade as a marker for the continental origin of fish caught at sea in distant-water fisheries. However, a simpler, more cost-effective, but still accurate, assignment can be obtained using a single microsatellite locus in combination with a mitochondrial DNA (mtDNA) single-nucleotide polymorphism (SNP) detected by restriction enzyme digestion. 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 as determined using microsatellites. Variation in 20 mtDNA regions, encompassing ~43% of this genome, in 330 salmon 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: Atlantic salmon, genetic stock identification, marine ecology, mitochondrial DNA, 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 (Schwartz *et al.*, 2006; Palsbøll *et al.*, 2007). 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). Equally, it depends on identifying DNA loci at which differentiation has evolved as a result of 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. Because they are derived from arbitrary DNA loci, most of the sets of loci used are unlikely to represent the most divergent loci, to be the best possible set of population markers for resolving population structuring and assignment of natal origin, or to be the most cost-effective choice. Identification of the best loci would maximize the resolving power and assignment success, but finding them poses a significant challenge given the size of most genomes and because different loci may be optimal in different parts of a species' range. However, the scope for improving the existing sets of marker loci is being facilitated by recent advances in genome sequencing technology, which allow rapid genome scanning for polymorphisms at acceptable cost (Davey *et al.*, 2011).

Most Atlantic salmon (*Salmo salar*) are anadromous, spending their early life in rivers, undertaking a marine migration, and returning 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 salmon began in the late 1960s, and the ensuing genetic studies have dramatically improved understanding of the structuring of the species into distinct populations and phylogenetic groups. The collective body of work that has emerged indicates that North American and European salmon 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 origin with effectively 100% certainty (Koljonen *et al.*, 2007, and references therein). It also clearly identifies 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-Ouellet *et al.*, 2009; Griffiths *et al.*, 2010; Sheehan *et al.*, 2010). Recently, efforts have been made to develop a robust, comprehensive methodology for within-continent, regional, or river-specific assignment. For European salmon, a microsatellite-based assignment tool (Genetically based Regional Assignment of Atlantic Salmon Protocol—GRAASP) has been developed through the EU-funded SALSEA-Merge Project with the goal of increasing understanding of the marine ecology of salmon in the Northeast Atlantic (http://www.nasco.int/sas/salseamerge.htm).

GRAASP, as it is currently implemented, provides a cost-effective assignment of European salmon to broad regions using microsatellite markers, although in some cases river-specific assignment can be achieved. The suite of microsatellite loci used does not, usually, allow for fine-scale regional or river-specific

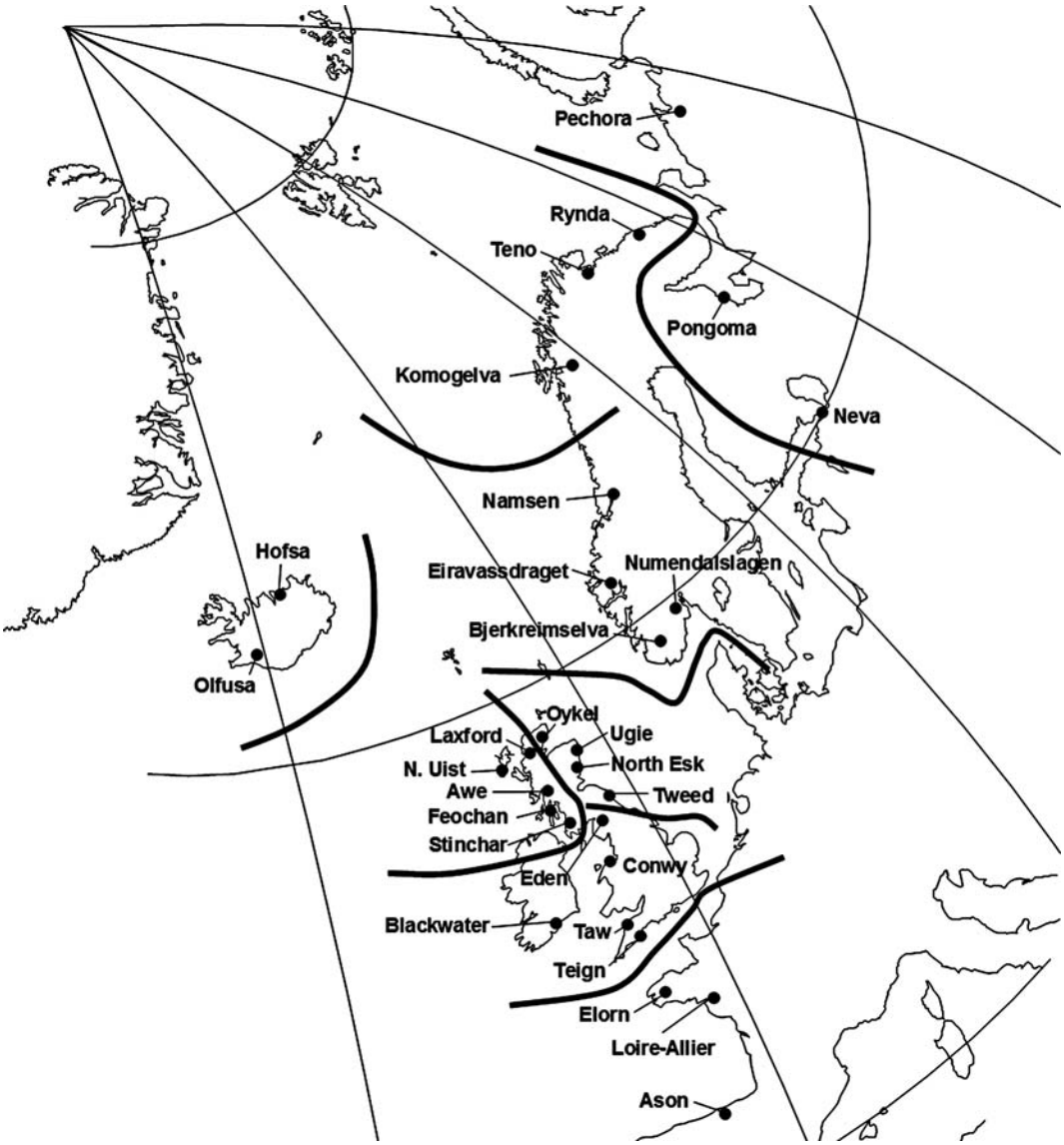

Figure 1. Map of the locations of rivers from which samples were analysed; heavy lines delineate regional groupings of samples used for stratified resampling (see text).

assignment. However, the work that has been carried out does show regional differentiation of river stocks at finer scales (King *et al.*, 2007), even between adjacent rivers, suggesting that accurate river-specific assignment may be possible (e.g. Wennevik *et al.*, 2004; Ryynänen *et al.*, 2007; Grandjean *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 as a consequence of a high mutation rate and shows higher levels of population differentiation than many nuclear genes owing 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 an mtDNA restriction fragment length polymorphism (RFLP) was 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 accuracy of >99%. Studies of restriction enzyme- and sequencing-detected polymorphisms show substantive regional and river-specific differences in variant frequencies (Verspoor *et al.*, 2002, b; King *et al.*, 2007), suggesting that 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 because, in most population studies, only a small part of the mtDNA genome (generally <5%) or a 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 ($n = 9$) screened.

Here, we describe a broad-scale preliminary assessment of the nature and extent of mtDNA single-nucleotide polymorphism (SNP) in European Atlantic salmon. 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.

Material and methods

Samples

The study involved the screening of mtDNA variation in 330 individual salmon from 29 rivers across Europe, with numbers analysed ranging from 6 to 12 fish per river. The rivers selected are broadly geographically representative of (Figure 1) and encompass the main phylogeographic regions suggested by allozyme studies (Verspoor *et al.*, 2005a). The samples analysed were derived from archived fin tissue that had been collected during studies over the past two decades and preserved in ethanol.

DNA extraction and sequencing

DNA was extracted using commercially available DNA extraction kits (Qiagen). Screening for variation was carried out in a single sequencing run using a novel approach developed by combining the traditional polymerase chain reaction (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 with an individual. The approach allowed sequencing of extensive regions of the mtDNA genome for a large sample group (546 individuals) in a single run, making it both quick and cost-effective. In all, 20 independent regions of 311–384 bp were sequenced for each individual, encompassing a total of 7215 bp (Table S1), ~43% of the 16 665 bp in the 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 shown to have the highest levels of polymorphism. Sequence reads were aligned according to the *S. salar* mitochondrial reference sequence (NC_001960.1), and the presence of an 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 were in more than one individual. The average number of reads supporting each SNP per individual was 27.3, with an s.d. of 11.7 (Fridjonsson *et al.*, 2011).

Table 1. Amplicons sequenced and the 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	637–1 059	17	0.044619	9	0.5294
ND1	2	384	3 838–4 260	10	0.026042	8	0.8000
	3	369	4 248–4 654	5	0.013550	5	1.0000
	4	324	4 635–4 998	7	0.021605	6	0.8571
ND2	5	361	5 110–5 510	10	0.027701	7	0.7000
	6	346	5 490–5 879	6	0.017341	3	0.5000
COXI	7	372	6 942–7 351	9	0.024194	5	0.5556
	8	382	7 340–7 762	9	0.023560	8	0.8889
COXII	9	361	8 193–8 594	5	0.013850	4	0.8000
	10	311	8 561–8 907	6	0.019293	8	1.3333
ATP6	11	375	9 238–9 651	11	0.029333	8	0.7273
ND3	12	357	10 623–11 025	9	0.025210	8	0.8889
ND4	13	363	11 146–11 546	8	0.022039	8	1.0000
	14	361	11 534–11 935	11	0.030471	11	1.0000
	15	370	11 912–12 326	13	0.035135	5	0.3846
ND5	16	345	14 309–14 701	7	0.020290	5	0.7143
	17	370	14 680–15 091	10	0.027027	8	0.8000
CYTB	18	366	15 376–15 779	7	0.019126	6	0.8571
	19	352	15 765–16 160	4	0.011364	4	1.0000
	20	365	16 133–16 537	8	0.021918	7	0.8750
Overall		7 215	1–16 665	172	0.023839	139	0.8081

Table 2. Frequencies of haplotypes observed in samples; haplotype numbers are the same as in Figure 3.

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: 0.083
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: 0.091	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		
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			

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, including those with <5% missing sequence data. For those fish, 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 was assessed based on numbers of pairwise 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 the number of haplotypes identified with progressively increasing numbers of amplicons was constructed manually, based on 26 of the 29 populations sampled; the best-fit curve was determined visually using SlideWrite Plus software (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 the numbers of populations sampled and the numbers of haplotypes observed was generated by manual resampling of the populations stratified by the regional groupings as indicated in Figure 1.

Average pairwise differences among individuals within, and corrected average pairwise differences among, populations were calculated and tested for significant differences between samples. Additionally, an AMOVA of within- and among-group variation was conducted. Both tests were carried out using Arlequin v3.5 (Excoffier and Lischer, 2010). The regional groups used in the AMOVA correspond closely with those identified by microsatellite data (J. Gilbey, pers. comm.) and fit within those observed using allozymes. These were: (i) Rynda and Teno; (ii) Namsen, Eiravassdraget, and Bjerkreimselva; (iii) Tweed, North Esk, Ugie, and Oykel; (iv) Laxford, North Uist, Awe, and Feochan; (v) Stinchar, Eden, Conwy, Blackwater, and Taw; the remaining individual samples were each treated as a distinct group. A Mantel test of association of genetic and geographic distance was 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 pairwise population genetic distance matrix generated by Arlequin v3.5. Using the latter matrix, an ME clustering tree was generated using MEGA4.

Results

The SNP variation observed within and among the 330 individuals screened defined 139 haplotypes for which SNPs are defined in NCBI and haplotype composition in Supplementary Table S1. There was uncertainty in the assignments of just 7 of the 330 fish to haplotypes based on missing information. Those fish could have been assigned to either of two closely related haplotypes differing in 1 bp. The haplotype frequencies observed across samples are shown in Table 2. As summarized in Figure 2, no haplotypes were observed in all samples, only three occurred at ten or more locations, only 12 were observed in fish from two or more locations, and 89 were in just one sample.

Based on genetic relatedness, the haplotypes clustered into five major groups based on pairwise differences (Figure 3), with most haplotypes found in one of these groups and with the other four groups containing 2–4 haplotypes 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 this group, 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 with 1–10 bases between haplotypes within these groups. The largest of these three clusters shows further substructuring into three more poorly defined groups, and these in turn are divided into smaller groups of more closely related haplotypes, with most haplotypes within smaller clusters separated by 1–5 base differences.

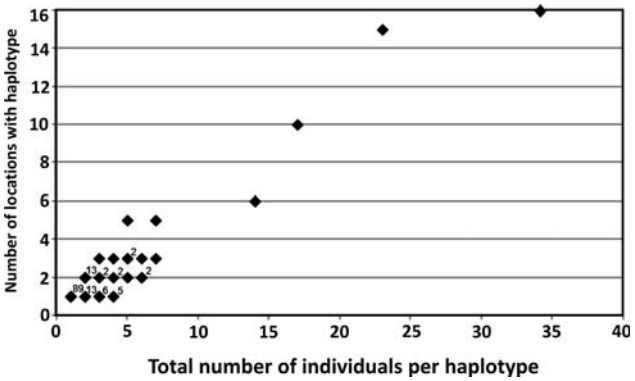

Figure 2. The number of locations with a haplotype plotted against the total number observed for that haplotype, for all 139 haplotypes detected in samples; numbers indicate the number of haplotypes with a given value.

The number of haplotypes defined within each amplicon varied from 3 to 11, with a fourfold variation in the number of haplotypes defined per SNP (Table 1); the number of SNPs per amplicon varied from 4 to 13 that, when corrected for amplicon size, showed a fourfold 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, whereas in the second part of the *CoxII* gene, the number of haplotypes defined was greater than the number of SNPs, owing 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 and adding regions 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 (Figure 4, Table 1) gave 1–2 new haplotypes, whereas including amplicon 10 added ~18 new haplotypes.

Stratified subsampling of populations shows the number 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 second-order polynomial and suggests that the numbers detected with each additional population may be decreasing gradually, with a possible plateau in haplotype numbers

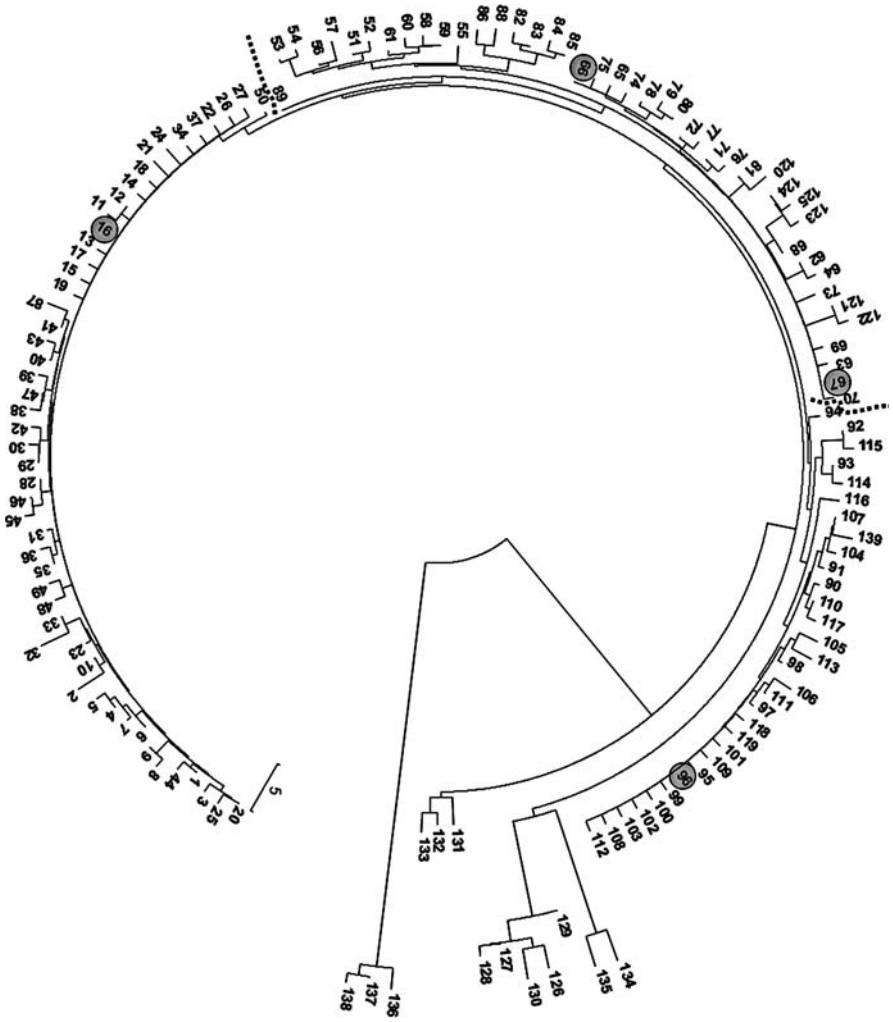

Figure 3. ME tree of the relatedness of the haplotypes based on the number of pairwise differences; the four most common haplotypes (Table 2) are highlighted.

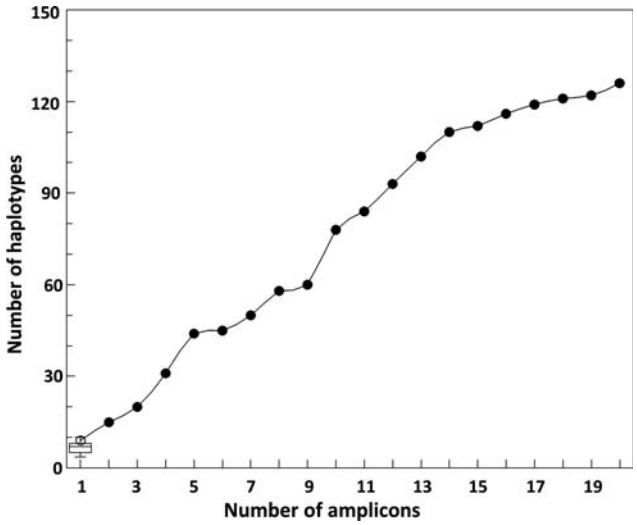

Figure 4. Cumulative number of haplotypes defined with the sequential addition of amplicons clockwise from D-loop to CytoB gene, based on data for 26 of 29 locations; the best-fit curve shown is a second-order polynomial.

predicted when the numbers reach 50–60. In contrast, with one major exception, the rarefaction curves for haplotype diversity as a function of sample size show a more or less linear increase in haplotype diversity with increasing size of sample (Figure 6). For the River Allier, the curve begins to level out, suggesting that the estimate of haplotype diversity from that location is less constrained by sample size than is the case for the other locations.

A large 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 and do not appear entirely unlinked to geography. This arises because sites which are both geographically distant but proximate in the sampling scheme can be genetically relatively similar (e.g. the Neva and Pechora samples), whereas 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. all other regions, the close relatedness of the Teno and Rynda, and the close relationship of the Pechora, Pongoma, and Neva). Pairwise differences among geographically close rivers, recognizing the somewhat arbitrary

nature of the cutoff as to what is included, are graphically summarized in Figure 7, and an overall ME tree based on pairwise differences is shown in Figure 8.

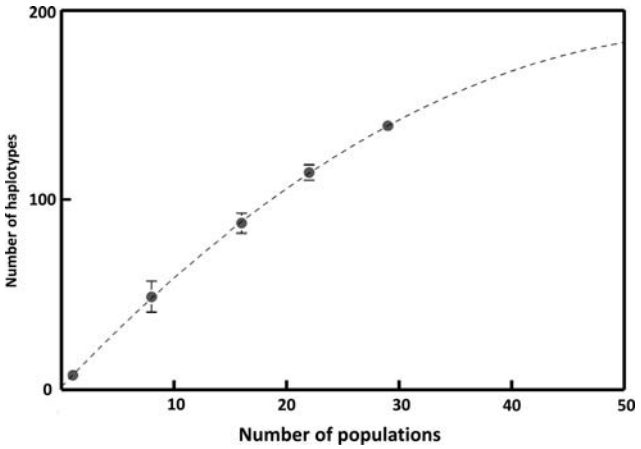

Figure 5. Relationship between the number of populations and the number of haplotypes, based on a geographically structured resampling of the 29 populations; the best-fit curve shown is a second-order polynomial.

Molecular analysis of variance shows that the frequencies of haplotypes in the samples are highly significantly heterogeneous among the defined groups, and approach 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 the potential for using mtDNA variation as population markers 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. Previously, given available technology and the cost of screening, studies have only been able to examine: (i) large numbers of individuals for a small number of RFLPs (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001) or, less commonly, for SNP variation in a small PCR-amplified fragment (Verspoor *et al.*, 2005b);

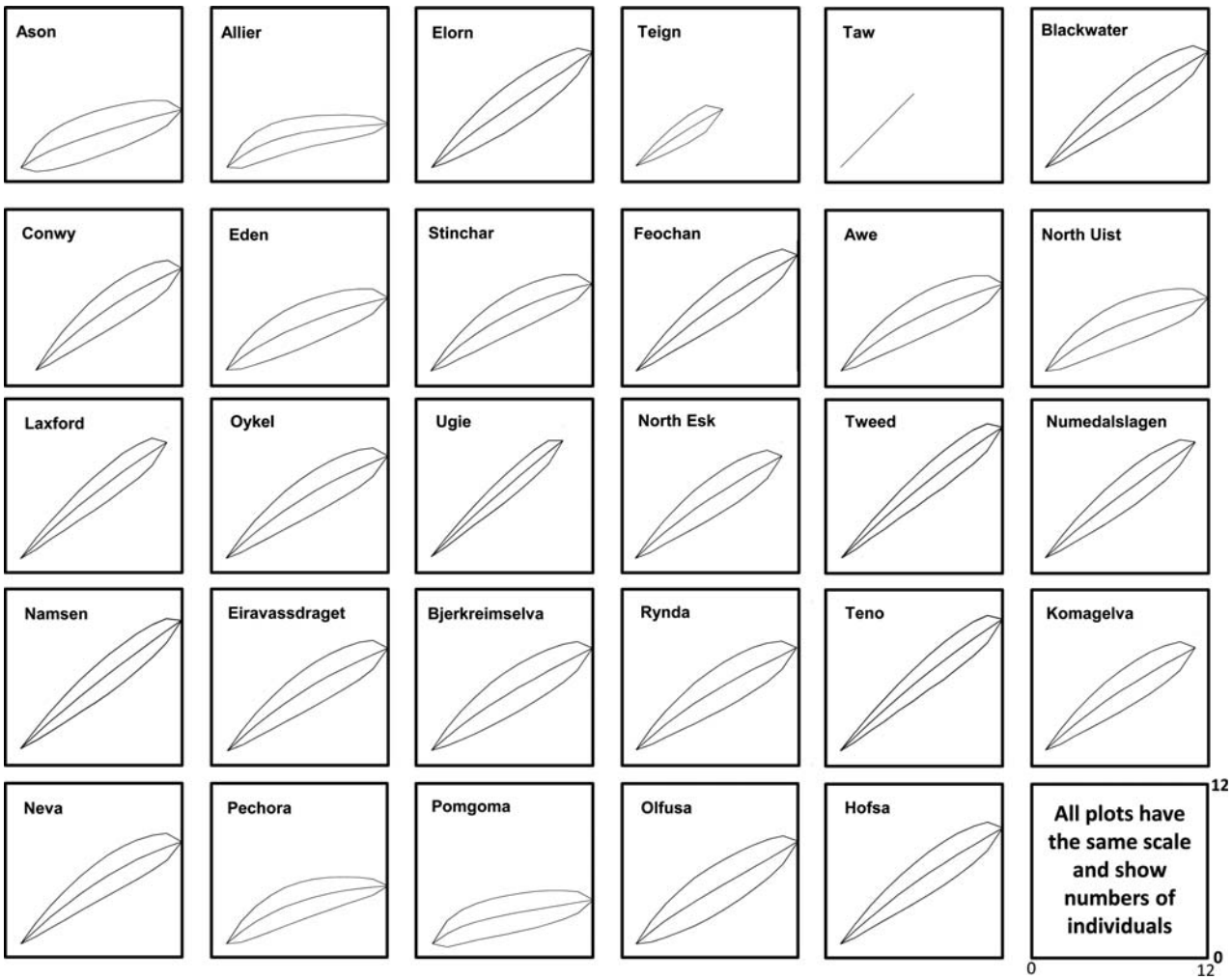

Figure 6. Rarefaction curves for individual samples showing the relationship between sample size and haplotype diversity. Curves shown are mean and s.d.

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 diagonal).

	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	n.s.	*	#	#	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	#	#	#	*	#
Teno	2.2	1.9	2.5	−1.3	23.2	*	#	#	#	#	n.s.	n.s.	n.s.	n.s.	n.s.	*	#	n.s.	#	n.s.	n.s.	n.s.	#	n.s.	*	#	#	*	#
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	#	n.s.	n.s.	#	n.s.	*	n.s.	n.s.	#	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	#	#	#	*	*
Eirva	1.5	1.0	2.0	3.2	1.6	2.4	−0.1	5.1	*	#	*	#	*	#	n.s.	n.s.	*	n.s.	*	*	#	n.s.	#	#	*	*	n.s.	*	*
Bjerkriemselvaa	3.0	1.8	3.5	3.8	1.9	2.7	0.7	0.7	4.8	n.s.	n.s.	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	#	*	n.s.	*	#
Numendalslagen	1.2	1.2	1.8	3.1	1.6	2.4	−0.1	−0.1	1.0	5.4	#	n.s.	*	n.s.	n.s.	#	#	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	#	#	#	*	*
Tweed	1.4	1.5	1.9	2.8	1.3	1.4	0.3	0.4	1.1	0.3	8.2	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	#	#	#	*	#
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	#	n.s.	n.s.	n.s.	n.s.	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	#	n.s.	*	*
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	n.s.	n.s.	#	#	n.s.	*	n.s.	#	n.s.	#	#	*	#	*	*	*
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	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	#	#	n.s.	*	#
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	n.s.	#	n.s.	#	n.s.	n.s.	n.s.	n.s.	n.s.	*	#	n.s.	*	*
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	#	n.s.	*	n.s.	#	n.s.	n.s.	#	*	*	n.s.	*	*
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	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	#	#	#	*	*
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	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	#	n.s.	*	*
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	n.s.	n.s.	n.s.	n.s.	#	#	*	*	*	*
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	n.s.	n.s.	n.s.	n.s.	*	#	n.s.	*	*
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	n.s.	n.s.	n.s.	n.s.	#	*	*	*
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	n.s.	n.s.	n.s.	n.s.	n.s.	*	#
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	n.s.	*	#	n.s.	*	*
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	n.s.	n.s.	#	*	*
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

n.s., not significant.
#Significant.
*Significant after Bonferroni correction for multiple tests, for $p < 0.05$.

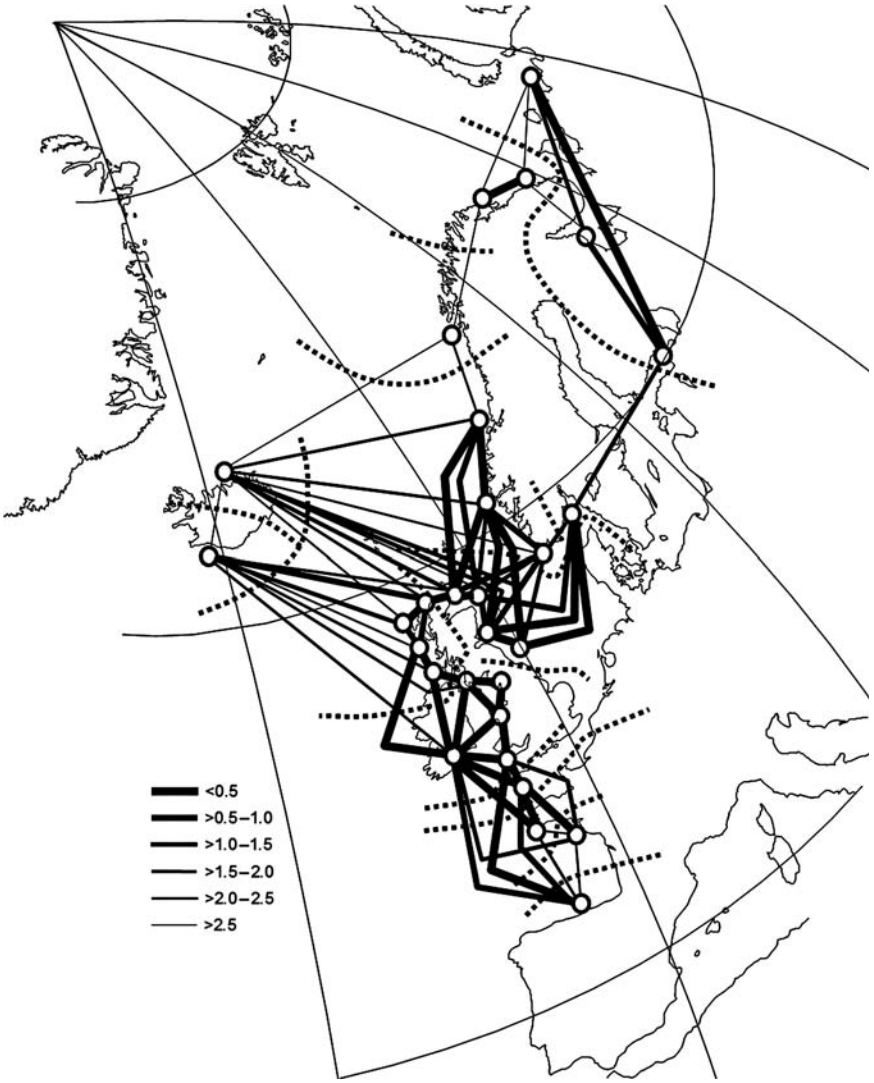

Figure 7. Relative degree of similarity between geographically neighbouring samples based on mean pairwise base differences between haplotypes in samples. Dotted lines show the separation of the samples into regional groups based on the microsatellite dataset.

(ii) a small number of individuals for large parts of the genome using large numbers of restriction enzymes (Bermingham *et al.*, 1991); or (iii) small numbers of salmon for the entire mtDNA genome (So, 2006). However, the potential for using mtDNA variation as a marker has been demonstrated in some cases, e.g. continent of origin (Gilbey *et al.*, 2005), and regional differentiation is known, both in North America and Europe (King *et al.*, 2007, and references therein). This suggests that there is potential for its application on smaller regional scales within continents.

The analysis of mtDNA SNP variation in European salmon presented here reinforces this view, significantly advances existing understanding of general levels of diversity, and suggests a high level of mitochondrial diversity within and among rivers. However, the full extent of regional and inter-river differentiation remains to be elucidated. Because of the high levels of diversity and the relatively limited sampling of rivers and of individuals within rivers, the number of samples and the sample sizes screened are inadequate. They do not provide an accurate and precise account either of the number of different haplotypes present or their frequencies, or of inter-river differentiation. That said, the results strongly suggest that haplotype distributions and frequencies differ significantly among most river systems and that there is also likely to be regional differentiation 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 supports the view that there is a high level of uniqueness in haplotype frequencies between populations. Three considerations suggest that the number of haplotypes identified is likely to be a fraction of the mtDNA variation present in European salmon stocks. First, haplotype numbers in almost all populations are a linear function of sample size and do not plateau with increasing numbers of individuals sampled as would be expected if most haplotypes had been resolved. The same is true of the number of populations sampled. Second, the analysis suggests that the rate of increase in the number of haplotypes with increasing numbers of samples is only starting to decline. As such, it is expected that the amount of diversity found would increase substantially by both increasing sample sizes and increasing sample numbers. Third, 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. Hence, a more extensive genomic analysis would be expected to show many of the haplotypes resolved here to represent heterogeneous classes. However, the increase in the number of haplotypes resolved appears to be starting to decline with increasing numbers of amplicons, suggesting that further screening of more mtDNA regions may not be as useful as extending the number of populations surveyed and the number of individuals screened per population. Conversely, the data also show that the number of new haplotypes added does vary considerably across the mtDNA molecule.

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 because often different SNPs are exclusively associated with a single haplotype and only one may be required for its resolution, reducing the number of SNPs that need to be screened 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.

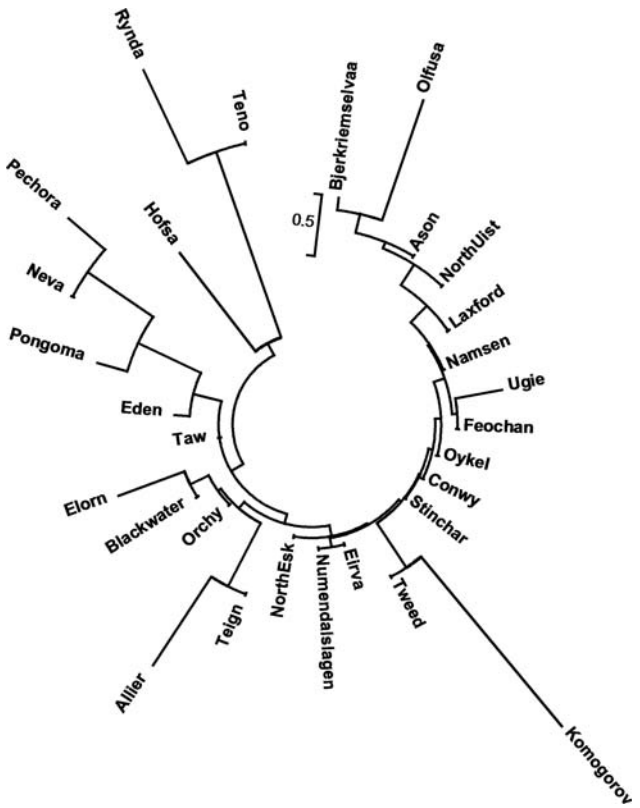

Figure 8. ME tree of the relatedness of the population samples based on the number of pairwise base differences between individuals in samples.

The findings of the current study accord with observations of previous work based on RFLP and sequence analysis of more restricted parts of the Atlantic salmon mtDNA genome that reveal 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.*, 2001). 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 salmon stocks has not been conducted. The only report of small-scale regional variation is that of Verspoor *et al.* (2002), who found that one RFLP identified in the ND1 gene region and resolved by the restriction enzyme *Alu*I was only present in populations of salmon in the inner Bay of Fundy. This work was extended by sequence analysis of two 350 bp regions of this gene in 743 salmon from 26 rivers (Verspoor *et al.*, 2005b). That study found regional differentiation encompassing low frequency regionally restricted haplotypes among salmon populations in the inner and outer Bay of Fundy, and along the south and eastern shores of Nova Scotia.

Based on the results of the current study, the lack of evidence for regional structuring seen in existing mtDNA studies most probably arises because it is based on a limited and arbitrary screening of the mtDNA molecule, with a few restriction enzymes that resolve widespread polymorphisms. This misses most variation that is local in its occurrence and that shows high levels of inter-region or inter-river differentiation. Given the existence of a high degree of regional variation at nuclear genes (Verspoor *et al.*, 2005a; King *et al.*, 2007), 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 owing to its lower effective population size and higher mutation rate (Hansen *et al.*, 2007). There is concordance of small-scale regional patterns of differentiation in Atlantic salmon stocks in eastern Canada, where both classes of variation have been more extensively studied (O'Reilly *et al.*, 1996; Verspoor *et al.*, 2002, 2005b), and it has recently been supported by studies of nuclear SNP variation (Freamo *et al.*, 2011).

The observations reported here, insofar 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, outside the Baltic, the situation is decidedly inconclusive, and the current findings do

Table 4. Results of AMOVA for within-group and among-group variation in haplotype frequencies, with the groups being (i) Rynda and Teno, (ii) Namsen, Eiravassdraget, and Bjerkreimselva, (iii) Tweed, North Esk, Ugie, and Oykel, (iv) Laxford, North Uist, Awe, and Feochan, (v) Stinchar, Eden, Conwy, Blackwater, and Taw, and the remaining individual samples being each treated as a distinct group.

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	1 291.264	3.98891	

not significantly change this situation. The analysis of variation within and among groups, based on the regional groups suggested by the more detailed microsatellite analysis (J. Gilbey, pers. comm.), 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 any distinguishing between inter-river and inter-regional variation, and many of the differences or lack of differences found among samples may be artefacts of sample size. The number of rivers screened and the samples sizes used in the current analysis are too small, given the levels of variation observed, to draw specific conclusions, and it is only possible to make the general point that the observations strongly suggest that there is substantive regional and inter-river divergence in respect of mtDNA variation.

Despite its limitations and preliminary nature, this study significantly advances understanding of intra- and interpopulation mtDNA SNP variation in European Atlantic salmon stocks. It clarifies the considerable potential for using mtDNA SNPs to enhance the assignment success and resolution of microsatellite-based tools such as the SALSEA-Merge GRAASP, alone or in combination with nuclear SNPs. 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 ongoing advances in the speed and cost of screening SNPs, relative to microsatellite loci. These technological advances will also facilitate the further exploration of population differentiation required to assess more fully the potential offered by SNPs and to identify the SNPs that 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 this could be exploited for assignment of natal origin.

Supplementary material

Supplementary material is available at the *ICESJMS* online version of the paper. This contains the composite SNP profile for each of the 139 haplotypes observed and provides the NCBI RefSNP(rs) for each SNP.

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 DNA. Canadian 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. http://biodiversityinformatics.amnh.org/open_source/gdmg (last accessed 24 August 2011).

Excoffier, L., and Lischer, H. E. L. 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(Suppl. S1): 254–267.

Fridjonsson, O., Olafsson, K., Tompsett, S., Bjornsdottir, S., Consuegra, S., Knox, D., de Leániz, C. G., *et al.* 2011. Detection and mapping of mtDNA SNPs in Atlantic salmon using high throughput DNA sequencing. BMC Genomics, 12: 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., and Verspoor, E. 2005. Novel DNA markers for rapid, accurate, and cost-effective discrimination of the continental origin of Atlantic salmon (*Salmo salar* L.). ICES Journal of Marine Science, 62: 1609–1616.

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.

Griffiths, A. M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J., Bowkett, A., Minting, P., *et al.* 2010. Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in the southern part of the European range. BMC Genetics, 11: 31.

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

Hansen, M. M., Villanueva, B., Nielsen, E. E., and Bekkevold, D. 2007. Investigating the genetics of populations. *In* The Atlantic Salmon: Genetics, Conservation and Management, pp. 86–114. Ed. by E. Verspoor, L. Stradmeyer, and J. L. Nielsen. Blackwell Publishing, Oxford. 520 pp.

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., *et al.* 2007. Biodiversity and population structure. *In* The Atlantic Salmon: Genetics, Conservation and Management, pp. 117–166. Ed. by E. Verspoor, L. Stradmeyer, and J. L. Nielsen. Blackwell Publishing, Oxford. 520 pp.

Koljonen, M-L., King, T. L., and Nielsen, E. E. 2007. Genetic identification of individuals and populations. *In* The Atlantic Salmon: Genetics, Conservation and Management, pp. 270–298. Ed. by E. Verspoor, L. Stradmeyer, and J. L. Nielsen. Blackwell Publishing, Oxford. 520 pp.

Nilsson, J., Gross, R., Asplund, T., Dove, O., Jansson, H., Kellonieni, J., Kohlmann, K., *et al.* 2001. Matrilineal phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and post-glacial colonization of the Baltic Sea area. Molecular Ecology, 10: 89–102.

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

Palsbøll, P. J., Bérubé, M., and Allendorf, F. W. 2007. Identification of management units using population genetic data. Trends in Ecology and Evolution, 22: 11–16.

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.

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.

Sheehan, T. F., Legault, C. M., King, T. L., and Spidle, A. P. 2010. Probabilistic-based genetic assignment model: assignments to subcontinent of origin of the West Greenland Atlantic salmon harvest. ICES Journal of Marine Science, 67: 537–550.

So, M. S. Y. 2006. Evolution of mitochondrial DNA in the genus *Salmo*. MSc thesis, Simon Fraser University. 93 pp.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 24: 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., Beardmore, J. A., Consuegra, S., de Leániz, C. G., Hindar, K., Jordan, W. C., Koljonen, M-L., *et al.* 2005a. Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. Journal of Fish Biology, 67: 3–54.

Verspoor, E., McCarthy, E., Knox, D., Bourke, E., and Cross, T. F. 1999. The phylogeography of European Atlantic salmon (*Salmo salar* L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. Biological Journal of the Linnaean Society, 68: 129–146.

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

Verspoor, E., O'Sullivan, M., Arnold, A. M., Knox, D., Curry, A., Lacroix, G., and Amiro, P. 2005b. 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, 18/05. 17 pp.

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. *In* The Atlantic Salmon: Genetics, Conservation and Management, pp. 17–56. Ed. by E. Verspoor, L. Stradmeyer, and J. L. Nielsen. Blackwell Publishing, Oxford. 520 pp.

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.