

Original Article

A microsatellite baseline for genetic stock identification of European Atlantic salmon (*Salmo salar* L.)

John Gilbey^{1,*}, Jamie Coughlan², Vidar Wennevik³, Paulo Prodöhl⁴, Jamie R. Stevens⁵, Carlos Garcia de Leaniz⁶, Dennis Ensing⁷, Eef Cauwelier¹, Corrine Cherbonnel⁸, Sofia Consuegra^{6,9}, Mark W. Coulson^{10,11}, Tom F. Cross², Walter Crozier⁷, Eileen Dillane², Jonathan S. Ellis^{5,‡}, Eva García-Vázquez¹², Andrew M. Griffiths⁵, Sigurdur Gudjonsson¹³, Kjetil Hindar¹⁴, Sten Karlsson¹⁴, David Knox¹, Gonzalo Machado-Schiaffino^{12,§}, Dorte Meldrup¹⁵, Einar Eg Nielsen¹⁵, Kristinn Ólafsson¹⁶, Craig R. Primmer^{17,¶}, Sergey Prusov¹⁸, Lee Stradmeyer¹, Juha-Pekka Vähä^{17,**}, Alexey Je. Veselov¹⁹, Lucy M. I. Webster^{10,‡‡}, Philip McGinnity^{2,†}, and Eric Verspoor^{1,11,†}

¹Marine Scotland Science, Freshwater Fisheries Laboratory, Faskally, Pitlochry PH16 5LB, UK

²Aquaculture & Fisheries Development Centre, School of Biological, Earth and Environmental Sciences, University College, Cork, Ireland

³Institute of Marine Research, PO Box 1870 Nordnes, 5817 Bergen, Norway

⁴Institute for Global Food Security, School of Biological Sciences, Queen's University, Belfast BT9 7BL, UK

⁵Department of Biosciences, Geoffrey Pope Building, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

⁶Department of Biosciences, Swansea University, Swansea, UK

⁷Agri-Food and Biosciences Institute Northern Ireland, Fisheries and Aquatic Ecosystems Branch, Newforge Lane, Belfast BT9 5PX, UK

⁸GENINDEXE, 6 rue des Sports, 17000 La Rochelle, France

⁹Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Aberystwyth, UK

¹⁰Rivers and Fisheries Trusts of Scotland (RAFTS), CBC House, 24 Canning Street, Edinburgh EH3 8EG, UK

¹¹Rivers and Lochs Institute, University of the Highlands and Islands, Inverness College, 1 Inverness Campus, Inverness IV2 5NA, UK

¹²Department of Functional Biology, Genetics, Universidad de Oviedo, C/Julian Claveria s/n, 33006 Oviedo, Spain

¹³Marine and Freshwater Research Institute, Skúlagata 4, 101 Reykjavík, Iceland

¹⁴Norwegian Institute for Nature Research (NINA), PO Box 5685 Torgard, 7485 Trondheim, Norway

¹⁵DTU Aqua, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsvøve 39, 8600 Silkeborg, Denmark

¹⁶Matis ohf., Vinlandsleid 12, 113 Reykjavík, Iceland

¹⁷Department of Biology, University of Turku, 20014 Turku, Finland

¹⁸Knipovich Polar Research Institute of Marine Fisheries and Oceanography, 6 Knipovich Street, Murmansk 183763, Russia

¹⁹Institute of Biology, Karelian Research Institute, Pushkinskaya 11, Petrozavodsk 10 185610, Russia

*Corresponding author: tel: +44 1796 472060; fax: +44 1796 473523; e-mail: john.gilbey@gov.scot

†These authors contributed equally to this work.

**Present address: Water and Environment of Western Uusimaa, PO Box 51, 08101 Lohja, Finland.

‡Present address: School of Biological and Marine Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK.

§Present address: Department of Biology, University of Konstanz, 78457 Konstanz, Germany.

¶Present address: Department of Biosciences and Biotechnology Institute, University of Helsinki, 00014 Helsinki, Finland.

**Present address: Science and Advice for Scottish Agriculture, Roddinglaw Road, Edinburgh EH12 9FJ, UK.

We dedicate this paper to the memory of our co-author Kristinn Ólafsson who died 22 March 2017, aged 39. Kristinn's contribution to the paper was seminal and crucial, both in leading the technical development of the screening protocols for the SALSEA panel of microsatellites, and in the generation of the Icelandic component of the baseline. He is sorely missed by us, his colleagues, and friends, and his passing is a loss to the European fisheries science community. Though only in the early stages of his career, his contribution to European fisheries science was already highly significant, and well beyond just the SALSEA-MERGE project.

Gilbey, J., Coughlan, J., Wennevik, V., Prodöhl, P., Stevens, J. R., Garcia de Leaniz, C., Ensing, D., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coulson, M. W., Cross, T. F., Crozier, W., Dillane, E., Ellis, J. S., García-Vázquez, E., Griffiths, A. M., Gudjonsson, S., Hindar, K., Karlsson, S., Knox, D., Machado-Schiaffino, G., Meldrup, D., Eg Nielsen, E., Ólafsson, K., Primmer, C. R., Prusov, S., Stradmeyer, L., Vähä, J.-P., Veselov, A. J., Webster, L. M. I., McGinnity, P., Verspoor, E. 2017. A microsatellite baseline for genetic stock identification of European Atlantic salmon (*Salmo salar* L.). – ICES Journal of Marine Science, 75: 662–674.

Received 6 April 2017; revised 17 August 2017; accepted 18 August 2017; advance access publication 23 September 2017.

Atlantic salmon (*Salmo salar* L.) populations from different river origins mix in the North Atlantic during the marine life stage. To facilitate marine stock identification, we developed a genetic baseline covering the European component of the species' range excluding the Baltic Sea, from the Russian River Megra in the north-east, the Icelandic Ellidaar in the west, and the Spanish Ulla in the south, spanning 3737 km North to South and 2717 km East to West. The baseline encompasses data for 14 microsatellites for 26 822 individual fish from 13 countries, 282 rivers, and 467 sampling sites. A hierarchy of regional genetic assignment units was defined using a combination of distance-based and Bayesian clustering. At the top level, three assignment units were identified comprising northern, southern, and Icelandic regions. A second assignment level was also defined, comprising eighteen and twenty-nine regional units for accurate individual assignment and mixed stock estimates respectively. The baseline provides the most comprehensive geographical coverage for an Atlantic salmon genetic data-set, and a unique resource for the conservation and management of the species in Europe. It is freely available to researchers to facilitate identification of the natal origin of European salmon.

Keywords: Atlantic salmon, genetic stock identification, individual assignment, marine ecology, microsatellites.

Introduction

Homing of Atlantic salmon to natal rivers, in combination with factors such as founder effects, isolation, selection, and genetic drift, and broad scale phylogeographic processes, has resulted significant population structuring at a hierarchy of levels from intra-river to inter-continental (King *et al.*, 2001) and locally adapted populations (Garcia de Leaniz *et al.*, 2007) including variations in marine migratory patterns amongst populations from different parts of the species range (Webb *et al.*, 2007). However, the full extent of differences in migratory patterns amongst populations and how this may be changing in response to shifting environmental conditions remains to be resolved (Jonsson *et al.*, 2016).

Advancing understanding of population and stock-specific migration, distribution and feeding patterns, and their implications for marine mortality rates, and the impact of climate change, is hampered by a lack of information relating to the marine-phase of the lifecycle (Crozier *et al.*, 2004). This situation makes it difficult to appropriately target actions to mitigate anthropogenic influences on different stock components, e.g. the impacts of mixed-stock fisheries and bycatches. Thus a tool that allows the accurate identification of genetically distinct populations and regional entities (MacKenzie *et al.*, 2011) and discrimination of the stock origins of fish in mixed feeding aggregations or during migratory phases would be invaluable in species' and North Atlantic marine ecosystem management.

DNA profiling methods for identifying the region or river/tributary of origin of salmonids have advanced over recent decades and are widely applied to Pacific salmon (*Oncorhynchus* spp.) stock management (e.g. Shaklee *et al.*, 1999; Beacham *et al.*, 2004, 2006; Shedd *et al.*, 2016). Their application to Atlantic salmon stock management has provided valuable insights into stock mixing at several spatial scales, including intercontinental (e.g. North American and European stocks in the West Greenland fishery: Gauthier-Ouellet *et al.*, 2009), regional (e.g. stock composition in Canadian gill-net fisheries: Bradbury *et al.*, 2016), and river level (e.g. population structuring in the River Teno/Tana: Vähä *et al.*, 2016). However, overall, its use has been more limited due to the lack of useful genetic baselines for many parts of the species range.

Genetic baselines are available for the western side of the Atlantic (e.g. Sheehan *et al.*, 2010; Bradbury *et al.*, 2015), including a recently developed fine scale range-wide North American microsatellite baseline (Bradbury *et al.*, 2016), that facilitate within-region identification of fish originating from Western Atlantic populations at high geographic resolution. In contrast, only partial baselines have been developed for the eastern side of the Atlantic (e.g. Griffiths *et al.*, 2010; Verspoor *et al.*, 2012; Ensing *et al.*, 2013; Gilbey *et al.*, 2016a; Vähä *et al.*, 2016) and no high-resolution baseline exists for the species' non-Baltic, eastern Atlantic range. Such a baseline would allow a DNA-based approach to the genetic stock identification (GSI) of marine samples from the Eastern Atlantic and, in conjunction with ecological studies, would help to provide a more detailed understanding of variations in the North Atlantic migration and distribution patterns of different European Atlantic salmon stocks. Such insight could improve understanding of the factors conditioning marine mortality, and facilitate the implementation of more effective management programmes (Crozier *et al.*, 2004).

GSI has been carried out using various genetic markers, with early work successfully using allozymes (Koljonen and McKinnell, 1996) and mitochondrial DNA (Moriya *et al.*, 2007) for salmonid species in some contexts, including for Atlantic salmon. However, higher levels of resolution and more widespread application has been subsequently achieved using microsatellite loci and they became the genetic marker most widely used in studies of Atlantic salmon stock differentiation. Even though, more recently, attention has turned to Single Nucleotide Polymorphisms (SNPs), the existing large body of microsatellite data available remains a unique and powerful resource that can be exploited for GSI in Atlantic salmon. However, it also has limitations (reviewed in Moran *et al.*, 2006) related to laboratories using different sets of markers, variations in allele-calling with different size markers or allele-size bins, different screening platforms, differences in chemistry, differences in the fluorophore markers across loci and whether the forward or reverse primer is labelled as well as differences in primer sizes. All of these can result in inconsistent allele-size designations across data sets generated by different laboratories. Nevertheless, evidence from large-scale standardization projects for salmonid species such as

Oncorhynchus mykiss (Stephenson *et al.*, 2009) and *Oncorhynchus tshawytscha* (Seeb *et al.*, 2007), as well as Atlantic salmon (e.g. Ellis *et al.*, 2011), indicate these issues can be addressed and comprehensive, large scale integrated genetic baselines constructed (Moran *et al.*, 2006).

Described here is a trans-European GSI baseline for Atlantic salmon (excluding Baltic salmon stocks which do not migrate to the North Atlantic) constructed by linking existing national and international microsatellite screening programmes. Baltic salmon populations are excluded from the baseline, as they do not migrate outside the Baltic Sea (Karlsson and Karlstrom, 1994; Torniainen *et al.*, 2013). Data were integrated for a common set of 14 microsatellite loci for a geographically representative set of rivers spanning the species' Eastern Atlantic European range from the Russian River Megra in the north-east (66.151N 41.484W), to the Icelandic Ellidaar in the west (64.117N 21.833E), and the Spanish Ulla in the south (42.639N 8.761E). Baseline samples encompassed rivers responsible for about $\approx 85\%$ of wild-salmon production in the study region (based on rod-catch data derived from numerous sources). Existing and new data supplied by partners in a multi-laboratory trans-European consortium were calibrated (Ellis *et al.*, 2011), subjected to stringent quality control and integrated to produce the new baseline. A hierarchical assignment unit approach was used and the baseline resolved into genetically distinctive regional assignment units. Assignment power and accuracy to these units were assessed, using both simulations and test samples, the latter constructed by removing fish from the dataset, to establish the utility of the baseline for regional assignment of marine-phase European origin salmon in the North Atlantic.

Material and methods

Baseline samples

Samples were collected from 32 888 Atlantic salmon from 551 sites representing 325 rivers in 13 countries across Europe [Denmark, England, Finland (two rivers with outlets in Norway), France, Iceland, Ireland, Northern Ireland, Norway, Russia, Scotland, Spain, Sweden, and Wales; Figure 1, Table 1, Supplementary Data S1 and S2], including the Baltic River Torne to act as a genetic out-group. Sampled sites spanned the species' entire eastern Atlantic range and spanned 3737 km from North to South and 2717 km from East to West.

Samples were collected from 1994 to 2010, with the majority collected in 2008–2009. Mainly juvenile fish were sampled, mostly parr and fry, but in some cases tissues from smolts or mature salmon returning to fresh water to spawn were sampled. Numbers sampled at a site ranged from 11 to 300 with a mean of 58, and rivers were characterized by 1–12 sites, depending largely on river size, with a mean number of sample sites per river of 1.7. Full details of sites are given in the Supplementary Data S1 and S2.

Genotyping

Microsatellite data were obtained from DNA extracted from tissue samples (typically fin clips or scales) screened by a consortium of 11 laboratories located across Europe (Table 1) for 14 of the 15 loci identified by a consortium of researchers and described by Olafsson *et al.* (2010). *SsaD486* (King *et al.*, 2005) was excluded from the analysis due to its lack of variation over much of the European range. The panel of 14 loci used here were *SsaF43* (Sanchez *et al.*, 1996), *Ssa14*, *Ssa289* (McConnell *et al.*, 1995), *Ssa171*, *Ssa197*, *Ssa202* (O'Reilly *et al.*, 1996), *SSsp1605*,

SSsp2201, *SSsp2210*, *SSsp2216*, *SSspG7* (Paterson *et al.*, 2004), *SsaD144*, *SsaD157* (King *et al.*, 2005), and *SSsp3016* (unpublished, GenBank number AY37820).

PCR conditions, thermocyclers and multiplexes varied across laboratories, as did genotyping platforms, size standards and other chemistry employed. Genotyping details and standardization of genotype assignments amongst laboratories appear in Ellis *et al.* (2011). In summary, two 96-well “control plates” were prepared (Matis, Iceland) containing template DNA extracted from samples representing the widest coverage of the range of *Salmo salar* as was practicable and which covered sites from both the Eastern and Western Atlantic. These were subsampled and typed by each laboratory. Genotypes were submitted by each member of the consortium to a single depository (Exeter University) where conversion algorithms and standardized nomenclature were applied. For each locus, lists of allele counts and sizes for each laboratory were aligned and cross-referenced for the sample genotypes in the control plates. Standard allele scores were designated for each locus and size differences between allele lists from each laboratory were determined, which allowed laboratory specific standardization rules to be defined. It should be noted that using this approach not every possible allele was screened, but the approach did allow the individual microsatellite bin ladders to be defined at each location. It cannot be ruled out therefore that rare alleles or alleles affected by regional indels may have been missed using such an approach, although the coherence of the reference baseline produced (see below) suggests this is unlikely to have been a major influencing factor.

On the basis of the standardization rules, all data generated for baseline sites were converted to the standard size ranges and stored in a single bespoke database for further analysis (see Ellis *et al.*, 2011 for full details). Sib-ship analysis amongst individuals in each sample was investigated using the pedigree-likelihood approach implemented within the program COLONY (Jones and Wang, 2010) and used to exclude all but one fish from each full-sib family in each sample prior to inclusion in the database. Fish with <10 loci genotyped were removed from further analysis due to concerns with DNA and genotype quality. Sites with more than half of the loci out of Hardy–Weinberg equilibrium (examined in GENEPOP 4.2.2; Rousset, 2008; potentially not representative of a single population), those that had <70% of fish scored at all loci (potentially poor quality DNA and genotypes), and those consisting of <30 individuals after quality control checks listed above (potential failure to provide accurate estimates of allele frequencies), were also removed. We estimated descriptive statistics with GenALEx 6 (Peakall and Smouse, 2006).

Assignment units

Assignment units were defined in an iterative way similar to that employed by Gilbey *et al.* (2016a). Units were first defined by a combination of distance-based and Bayesian clustering. Individual assignment accuracies using these units were then examined and units where accuracies did not meet a pre-defined threshold were combined with units that saw reciprocal misassignments, until all units had accuracies at or above the threshold level.

The distance-based approach was based on a neighbour-joining tree (Saitou and Nei, 1987) constructed using Nei's genetic distance D_A (Nei *et al.*, 1983) calculated in POPTREE2 (Takezaki *et al.*, 2010) and visualized in MEGA7 (Kumar *et al.*, 2016). The clustering approach was carried out in STRUCTURE

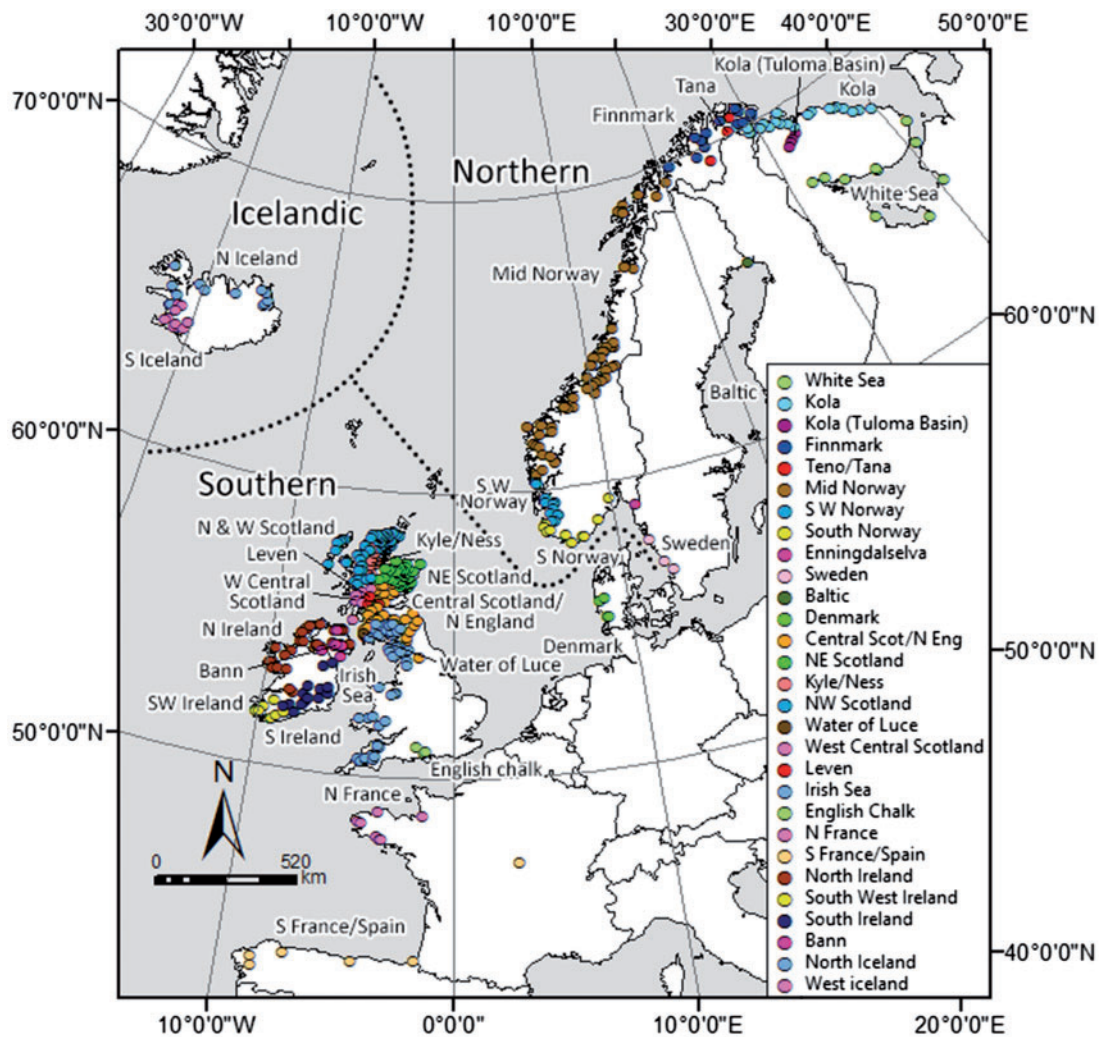


Figure 1. Map of sampling region. Points represent sample sites and/or river mouths. Full site information is contained in [Supplementary Data S1](#) and an expanded map with all rivers identified is in [Supplementary Data S2](#). Regions noted are all those referred to in the text. The Level 1 assignment units (see text) are delineated by the dashed line and the initial Level 2 units by coloured points.

(Pritchard *et al.*, 2000), using a burn-in of 100 000 and a run phase of 300 000 iterations during each application. Three replicates for each cluster number (K) were run with values of K from 1 to 10. $K = 10$ emerged as an upper limit after monitoring of the results of the runs while they were underway. In each case stable estimates of true K at the level under analysis had been identified by this point (see results). Prior site information was incorporated into the analysis using the LOCPRIOR option. The smallest K capturing the major structure in the dataset was defined by the ΔK method of Evanno *et al.* (2005), which was calculated using STRUCTURE HARVESTER (Earl and vonHoldt, 2012). Replicate membership coefficients were combined with CLUMPP (Jakobsson and Rosenberg, 2007) using the Full Search method.

The Bayesian clustering was carried out using a hierarchical approach, starting with the full dataset. Evanno *et al.* (2005) showed that STRUCTURE tends to capture the major structure in a reference dataset but that more fine scale structure may become evident if a hierarchical analysis is performed. In the current analysis, at each hierarchical level a STRUCTURE analysis was performed and the minimum best K identified. The data

were then split up into the cluster units and further STRUCTURE analysis performed on each one independently. This was repeated at each hierarchical split until either single-river structuring was observed or geographical coherence of the clusters was lost.

Once both the distance-based and clustering analysis had been performed, the degree to which the assignment units identified by each technique corresponded was examined. Where the same units were identified these were incorporated into the initial assignment unit panel. Where the two approaches had identified different units the smallest unit from either approach was incorporated into the initial assignment unit panel, for example in a situation where one technique had identified a single unit and another had identified sub-units the sub-units were added to the initial assignment panel. In this way, the smallest units identified by one or both technique were incorporated into the initial assignment unit test panel.

Once the initial assignment unit panel had been identified, individual assignment accuracy was calculated for each of these units (see below). If the assessed accuracy to a unit was at or above 80% the unit was retained in the panel. If accuracy was below this level the unit was

Table 1. Sample baseline coverage pre- and post-genotype quality control (see text for details).

Country	Pre-QC			Post-QC		
	Rivers	Sites	Fish	Rivers	Sites	Fish
Denmark ^a	3	6	253	2	4	189
England ^{b,c}	24	38	1 652	23	35	1 498
Finland ^d	2	5	395	2	5	393
France ^{b,c,e,f}	13	16	759	9	9	450
Iceland ^g	17	25	2 352	16	22	1 986
Ireland ^h	29	45	2 345	29	40	2 053
Northern Ireland ⁱ	9	20	1 469	7	18	1 302
Norway ^{d,j,k}	90	109	7 749	81	99	7 008
Russia ^{d,j,l}	33	36	2 506	30	33	2 350
Scotland ^c	87	230	11 625	69	185	8 884
Spain ^f	7	7	342	4	4	190
Sweden ^{a,d}	4	4	180	4	4	172
Wales ^b	7	10	375	6	9	347
Total	325	551	32 002	282	467	26 822

Institutions contributing data: ^aDanish Institute for Fisheries Research, Denmark; ^bUniversity of Exeter, England; ^cMarine Scotland Science, Scotland; ^dUniversity of Turku, Finland; ^eGeneindex, France; ^fUniversity of Oviedo, Spain; ^gMarine and Freshwater Research Institute, Iceland; ^hUniversity College Cork, Ireland; ⁱQueen's University Belfast & Agri-Food and Biosciences Institute Northern Ireland, Northern Ireland; ^jInstitute of Marine Research, Norway; ^kNorwegian Institute for Nature Research, Norway; ^lKnipovich Research Institute of Marine Fisheries & Oceanography, Russia.

combined with other units to which reciprocal misassignments were occurring. Accuracies were tested again and the process repeated until all units in the panel had individual assignment accuracies at or above the 80% level. Nei's genetic distance D_A (Nei *et al.*, 1983) was again calculated for all pairwise final assignment combinations using the POPULATIONS 1.2.3 software package (Langella, 1999).

Assignment analysis

Individual assignment

Individual assignment accuracy was calculated using maximum likelihood-based mixture analyses carried out using ONCOR (Kalinowski *et al.*, 2007) with mixture proportions estimated using the EM algorithm and genotype probabilities calculated by the method of Rannala and Mountain (1997). In order to estimate unbiased assignment accuracies using fish not represented in the baseline, assignment tests were based on fish randomly removed from the reference baseline and combined into a mixture file. A randomly selected 10% of fish were removed from each of the three top level assignment units identified (see results) resulting in a total of 2682 fish in the mixture file. For each fish the most likely assignment unit of origin and associated assignment probability was calculated. Fish with assignment probabilities below 0.8 were classified as unassigned and excluded from the analysis. Accuracy to the assignment units was then calculated with the remaining fish. Using such a cut-off meant that fish whose origin was difficult to determine (low probability) were removed from the analysis and so potential accuracy could be increased (Bekkevold *et al.*, 2015; Gilbey *et al.*, 2016a). However, the application of cut-off scores also increased the proportion of unassigned fish (Gilbey *et al.*, 2016a) and can thus influence apparent stock proportions if calculated from the individual assignments. As such, this should not be performed for this purpose and so, in order to estimate accurate stock proportions a Mixed Stock Analysis (MSA) approach was utilized (see below).

100% simulations

Simulated fishery mixtures were analysed in ONCOR and comprised sets of 100% simulated samples of fish from each assignment unit. Genotypic frequencies for each locus in each unit were re-sampled following Anderson *et al.* (2008). The 100% simulations were based on 1000 simulations of 200 fish per hierarchical assignment unit and the same simulated reference sample sizes as in the actual dataset.

Mixed Stock Analysis

Mixed stock proportions were calculated for each assignment unit. The same set of 2682 randomly selected fish used for the individual assignments was used and mixture proportions estimated in ONCOR using conditional maximum likelihood (Millar, 1987) with confidence intervals calculated based on 1000 bootstraps.

Equal proportions

Mixed stock proportions were calculated for each assignment unit using simulated fishery mixtures with equal proportions of fish at each assignment unit in ONCOR. One hundred fish were simulated for each unit and confidence intervals of the estimates calculated using 1000 bootstraps.

Baseline coverage analysis—river removal

A baseline rarely covers all possible source populations completely, and so some fish in real fishery mixtures may be from populations not included in the baseline. Hence, simulation analysis may overestimate the success rates of assignments of fish in an actual fishery due to being based only on samples from sites and rivers contained in the baseline (Anderson *et al.*, 2008). This issue was addressed using a further test panel and associated test baseline. A random 10% of the rivers in each assignment unit were removed from the baseline and used as test mixtures that were then assigned back to the reconstructed baseline. All assignment units comprising more than one river had at least one river randomly removed (see Supplementary Data S1 for details of sites and rivers removed). Fish in these “unrepresented” mixture panels were thus from sites and rivers not included in the reconstructed baseline. In this way, we tested the capability of the baseline to reflect the regional signal of each assignment unit and to assign fish from sites and rivers not included in the baseline but from the assignment unit. This procedure was repeated at both assignment unit levels, again using ONCOR, with confidence intervals calculated based on 1000 bootstraps.

Results

Baseline QC

From a total of 551 sites sampled, 84 sites were removed, leaving 467 sites containing 26 822 fish representing 282 rivers in the final baseline (Table 1). From those removed, 17 sites were not in H-W proportions, 51 had <70% of fish screened at all loci, and 15 had <30 individuals representing the site after correction for full-siblings and individual fish for which <10 loci could be reliably genotyped. A further site (a sample of adult rod-caught fish from the Norwegian River Flekkeelva in 2007) was removed due to extreme outlier behaviour in the STRUCTURE analysis (data not shown). Full site details are contained in Figure 1, Table 1, and Supplementary Data S1 and S2. Across sites most loci were highly variable, with allele numbers ranging from 10 for *Ssa14* to 46 for

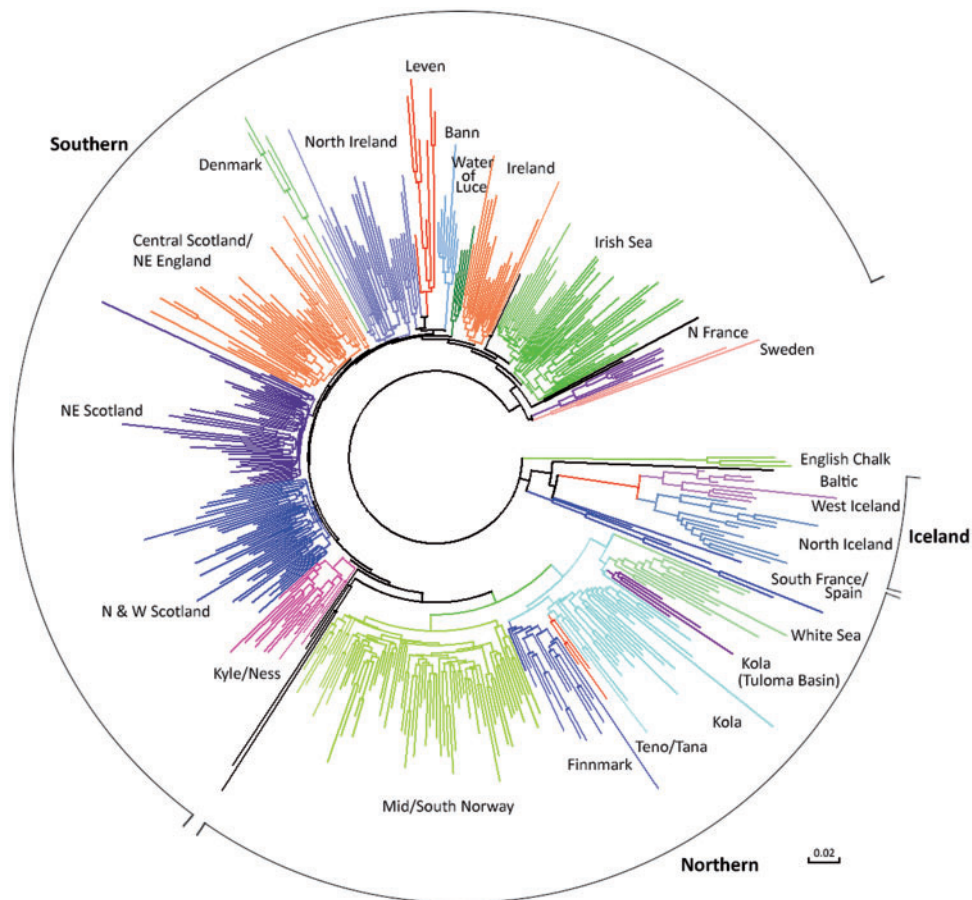


Figure 2. Neighbour-joining phylogenetic tree of sample sites based on D_A with major clusters coloured and named. Expanded tree with all sites identified is detailed in [Supplementary Data S4](#).

SsaD157 (mean 29.9). Additional descriptive and diversity estimates for each locus and site are presented in [Supplementary Data S3](#).

Definition of initial assignment regions

A neighbour-joining tree of Nei's D_A is summarized in [Figure 2](#) with an expanded version detailed in [Supplementary Data S4](#) and full site level D_A matrix in [Supplementary Data S5](#). A plot of ΔK , and a map showing the geographic positioning of the clusters at each hierarchical STRUCTURE level are shown in [Figure 3](#). Assignment units as defined by POPTREE and STRUCTURE are compared in [Supplementary Data S6](#).

Both distance-based N-J tree and Bayesian STRUCTURE approaches identified three large regional groupings of sites covering the northern, southern and Icelandic regions and these will henceforth be referred to as the Level 1 assignment units. There was in general a good agreement between the two population structuring techniques at the lowest level units identified. Indeed, of the 26 and 22 units defined by the NJ Tree and Bayesian clustering methods, respectively, 17 units were identical ([Supplementary Data S6](#)). Using the lowest level divisions produced from each technique resulted in a total of 29 units identified for the initial Level 2 assignment accuracy testing (column 1 in [Table 2](#), [Supplementary Data S6](#)). The assignment units at both initial levels are mapped in [Figure 1](#), with D_A matrixes detailed in [Supplementary Data S8](#).

Assignment analysis

Initial assignment accuracy

Using the 2682 fish removed from the baseline, individual assignments were performed at Level 1 and at the initially defined Level 2 assignment units. At Level 1, the assignment accuracy of all fish to the northern, southern, and Icelandic unit respectively was 90.8, 92.7, and 99.5% respectively. Using a probability cut-off score ≥ 0.8 this increased to 94.2, 95.5, and 100% with 86.8, 90.2, and 99.5% of fish in the mixture being assigned.

Assignment accuracy of fish with probability scores ≥ 0.8 to the Level 2 units was $\geq 80\%$ in 19 of the 29 units ([Table 2](#); for full breakdown of assignments at each Level 2 iterative level see [Supplementary Data S7](#)). After combining assignment units based on reciprocal misassignments, 21 assignment units remained with recalculated accuracies $\geq 80\%$. A final round of assignment unit combination resulted in 18 assignment units for which assignment accuracies were all $\geq 80\%$ ([Table 2](#), [Supplementary Data S7](#)).

100% simulations

The 100% simulations for each assignment unit showed robust estimates of stock proportions at both assignment levels ([Figure 4](#)). At Level 1, the mean estimates matched the actual proportions extremely well with a maximum difference of just 0.3% between the actual and estimated values and all upper CI at 100%. The initial Level 2 assignment units again showed relatively

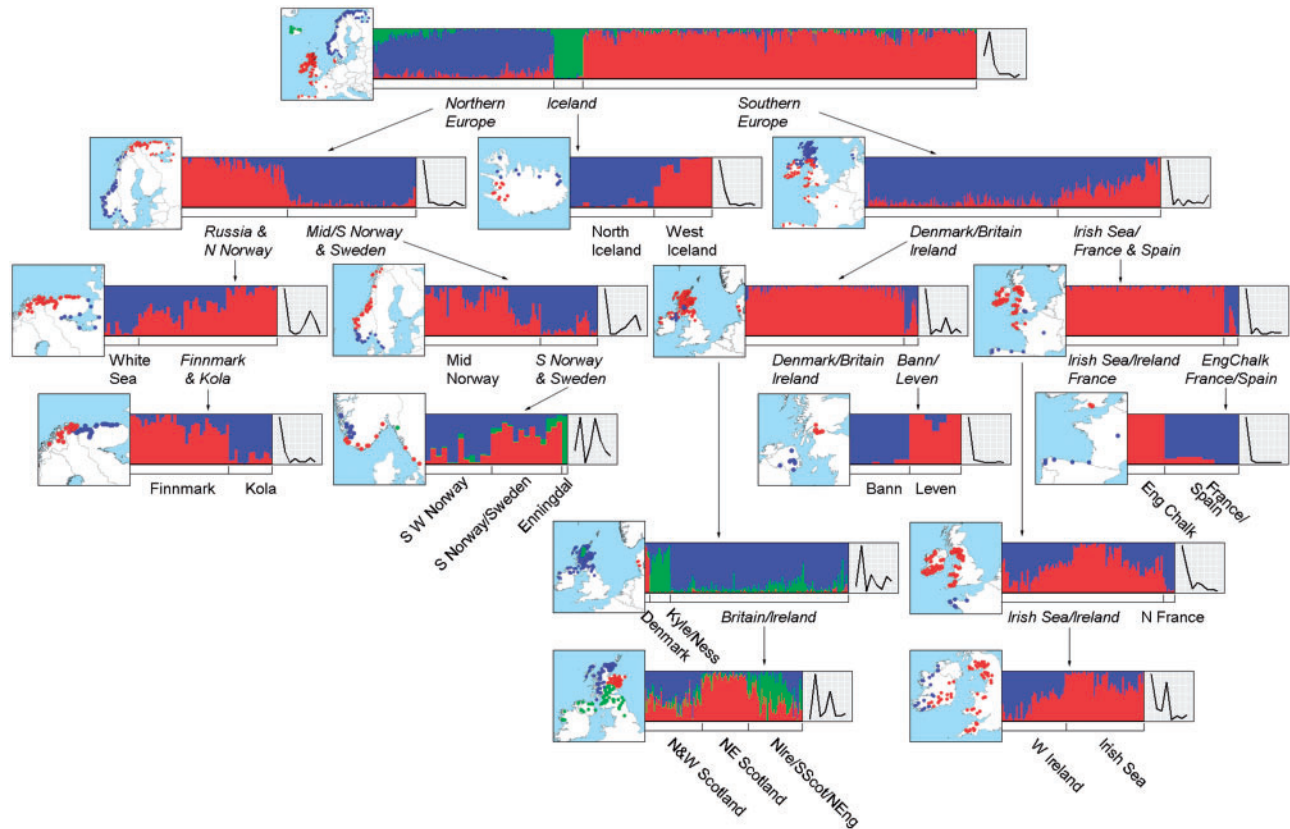


Figure 3. Hierarchical STRUCTURE based clustering analysis of sites. Each cluster analysis is described using three components. Firstly the results of the STRUCTURE analysis are shown with vertical bars representing individual sites and colours relating to cluster membership of that site. A plot of the ΔK values (Earl and vonHoldt, 2012) associated with the analysis is also shown defining the K identified in that cluster analyses. Finally, a map is shown detailing the geographic location of the clusters identified. Cluster names in italics refer to clusters for which further hierarchical analysis was performed. Cluster names in regular text refer to final cluster assignment groups.

accurate estimates with an average difference between the estimated and actual mean proportions of 4.5%. The West and Central Scotland level, however, showed a difference of 17.6% between estimated and actual proportions. At the first round of assignment unit combinations accuracies were seen to improve, as expected, with average and maximum differences between the estimated and actual mean proportions of 4.5 and 9.0%. These levels reduced to 1.9 and 8.0% respectively at the final Level 2 assignment unit combination round.

Mixed Stock Analysis

The results of the MSA using the 2682 fish removed from the baseline and used as a fishery mixture are shown in Figure 5a. For all assignment units, within both assignment levels, apart from a single unit in Level 2, South France/Spain, where the upper *CI* was 0.19 below the actual value, the estimated proportions of fish in the unit mixtures matched actual proportions (i.e. were within the *CI* bands). The estimates were also very precise with average *CI* bands of just 2.2 and a maximum of 4.7. Considering the high accuracy of the mixed stock estimates at this initial assignment unit composition, no further assignment unit amalgamations were deemed necessary for MSA.

Equal proportions

As with the previous analysis the equal proportion simulation showed excellent agreement between the actual and estimated

proportions in the mixture (Figure 5b). At Level 1, there was an average difference between actual and estimated of just 0.06% and a maximum of 0.09% (southern unit) and at Level 2 these two differences only rise to a mean difference of 0.4 and a maximum of 1.1% (North Ireland unit).

Baseline coverage analysis—river removal

The most demanding test of assignment capabilities of the baseline was the “river removal” test in which entire river systems were removed from the baseline and their fish assigned to region of origin using the remainder of the rivers in the reference baseline. However, even here relatively high levels of assignment accuracy were obtained (Figure 5c). Average differences between actual and estimated mixture proportions were 1.9% with a maximum of 2.3% (southern unit) at Level 1 and 1.3 and 2.9% (Central Scotland/North England) respectively at Level 2. At no time were significant proportions assigned to any of the six single-river assignment units which were not represented in the mixture file (lower *CI* at zero in these units).

Discussion

The study, encompassing the largest analysis of Atlantic salmon population structure in the Eastern Atlantic, for the first time, provides a genetic framework to exploit the power of microsatellite variation to assign Atlantic salmon from this part of the species’ range to smaller scale regional stock groups. As such, the

Table 2. Individual assignment accuracy using fish removed from the reference baseline.

Assignment unit	Sample size	Assigned %	Correct %	Assignment unit	Assigned %	Correct %	Assignment unit	Assigned %	Correct %
White Sea	758/86	68.6	90.2	White Sea	70.9	90.3	White Sea	72.1	90.3
Kola	1561/160	50	82.1	Kola	51.9	82.1	Kola	53.1	82.1
Kola (Tuloma Basin)	287/39	61.5	100	Kola (Tuloma Basin)	66.7	96	Kola (Tuloma Basin)	66.7	96
Finnmark	1109/107	54.2	84.7	Finnmark	59.3	82.9	Finnmark	59.3	82.9
Teno/Tana	271/28	42.9	10						
Mid Norway	3195/369	54.5	84.1	Mid & SW Norway	68.3	84.4	Mid & SW Norway	69.2	84.4
South West Norway	816/95	42.1	73.8						
South Norway	693/83	32.5	81.25	South Norway	45.8	82.4	South Norway	47.0	82.4
Enningdalselva	86/8	87.5	100	Enningdalselva	87.5	100	Enningdalselva	87.5	100
Sweden	108/12	33.3	100	Sweden	41.7	100	Sweden	41.7	100
Baltic	47/5	60	100	Baltic	80.0	100	Baltic	80.0	100
Denmark	176/13	61.5	100	Denmark	76.9	100	Denmark	76.9	100
Central Scotland/North England	1711/200	32	73.5	Scotland/North East England	66.3	80.4	Scotland/NE England/Irish Sea	76.4	87.2
North East Scotland	2183/233	42.5	56.5						
Kyle/Ness	814/99	42.4	78.7						
North & West Scotland	2005/255	35.7	72						
Water of Luce	225/20	30	40						
West Central Scotland	242/28	46.4	83.3						
Irish Sea	1992/214	39.7	77.3	Irish Sea	52.3	76.3			
Leven	324/41	75.6	100	Leven	82.9	96.9	Leven	85.4	96.9
English Chalk	134/9	88.9	100	English Chalk	88.9	100	English Chalk	100	100
North France	283/35	45.7	78.9	North France	51.4	78.9	France/Spain	68.0	91.7
South France/Spain	282/40	70	100	South France/Spain	72.5	100			
North Ireland	1519/161	50.3	87.0	North Ireland	59.6	85.9	Ireland	64.0	87.0
South West Ireland	34/35	54.3	85.7	South West Ireland	52.2	77.4			
South Ireland	572/57	29.8	58.8						
Bann	619/51	66.7	93.9	Bann	66.7	93.9	Bann	66.7	93.9
North Iceland	976/110	95.5	96.3	North Iceland	95.5	96.3	North Iceland	95.5	96.3
West Iceland	811/89	91.0	98.7	West Iceland	92.1	98.7	West Iceland	93.3	98.7

Initial assignment units in first column defined by distance and STRUCTURE based analysis. Remaining assignments represent amalgamations of units where assignment accuracy is < 80%. Assignment accuracy was calculated using only fish with individual assignment probabilities ≥ 0.8. Values in bold represent accuracy of at least 80% to assignment units. Sample sizes represent baseline/mixture size.

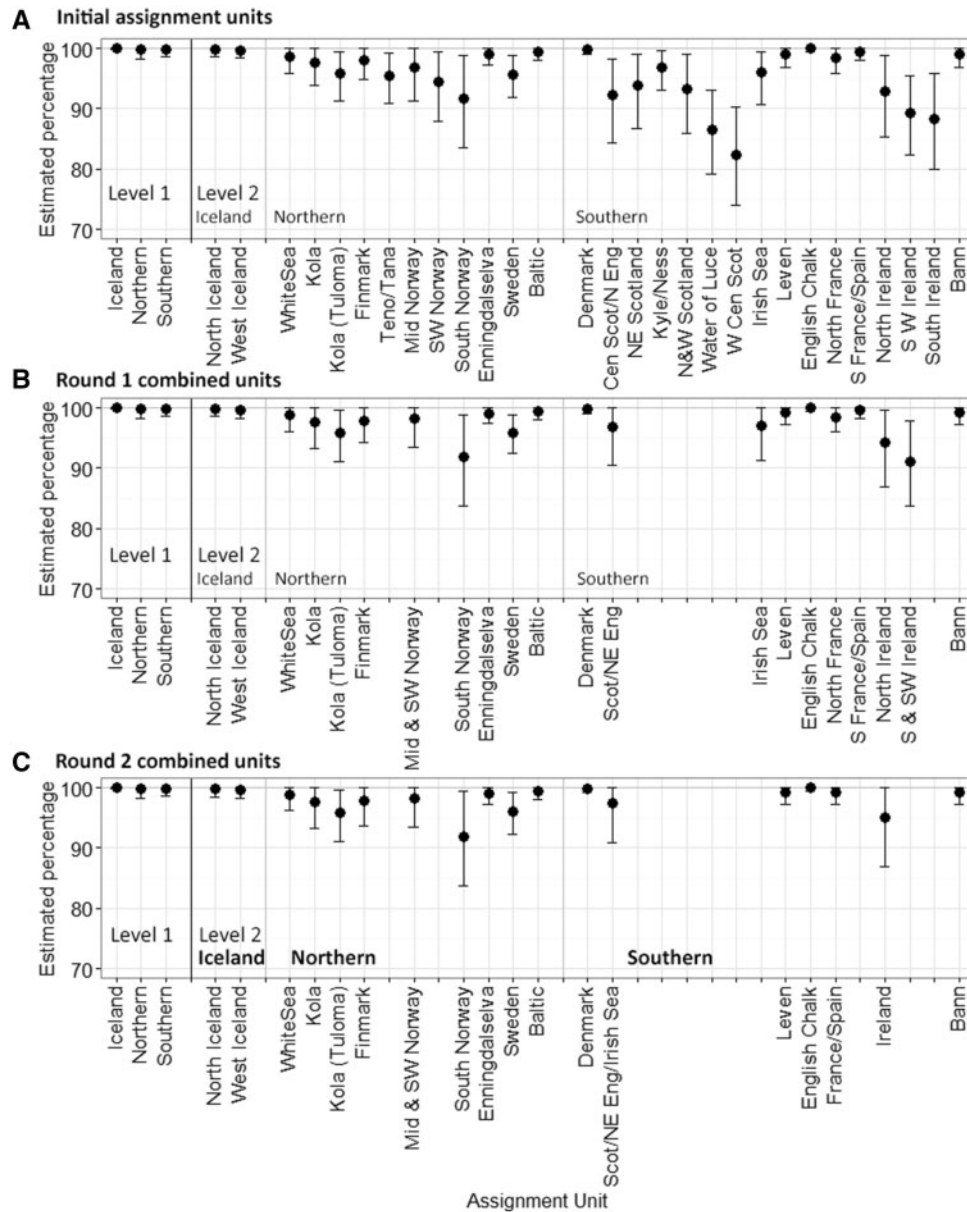


Figure 4. Proportion estimates from independent 100% simulation studies of the genetic baseline at Level 1 and all stages of the iterative formation of the Level 2 assignment unit levels. Points represent mean estimates with bars showing 95% confidence intervals.

reported genetic baseline provides a powerful resource that can be used to increase understanding of the biology of European Atlantic salmon stocks in the North Atlantic marine environment. Enhanced understanding of stock-specific marine migration, distribution, feeding patterns, exploitation, and mortality rates, will help to provide guidance towards a more efficient management of Atlantic salmon in a changing environment (Crozier *et al.*, 2004).

Distance-based and Bayesian cluster based analyses both reveal hierarchical structuring of river populations of European and Icelandic salmon into regional groups. At the highest level, this structure encompasses large-scale geographical discontinuities between northern (Scandinavia-Russia), Icelandic, and southern regions (Britain-Ireland-France-Denmark-Spain). Such differences have been identified in previous analyses of Atlantic salmon

population structure. For example, King *et al.* (2001) showed with microsatellites an unambiguous separation of Iceland, Norway, and Scotland-Ireland-Spain (their Figure 3), and Verspoor *et al.* (2005) identified an Icelandic group together with a southern British Isles-Northern France group using allozymes, although a more complex pattern was apparent in their analysis amongst the more central range groups.

At the next highest level, two assignment units shared the largest average degree of distinctiveness from other units, the two also being on opposite extremes of the neighbour-joining tree (Figure 2). The Baltic unit had a mean D_A of 0.236 to other units (Supplementary Data S8), a level of differentiation to other European rivers seen in previous studies (Bourret *et al.*, 2013) and consistent with the restricted migration of Baltic stocks (Karlsson and Karlstrom, 1994) and their long history of geographical

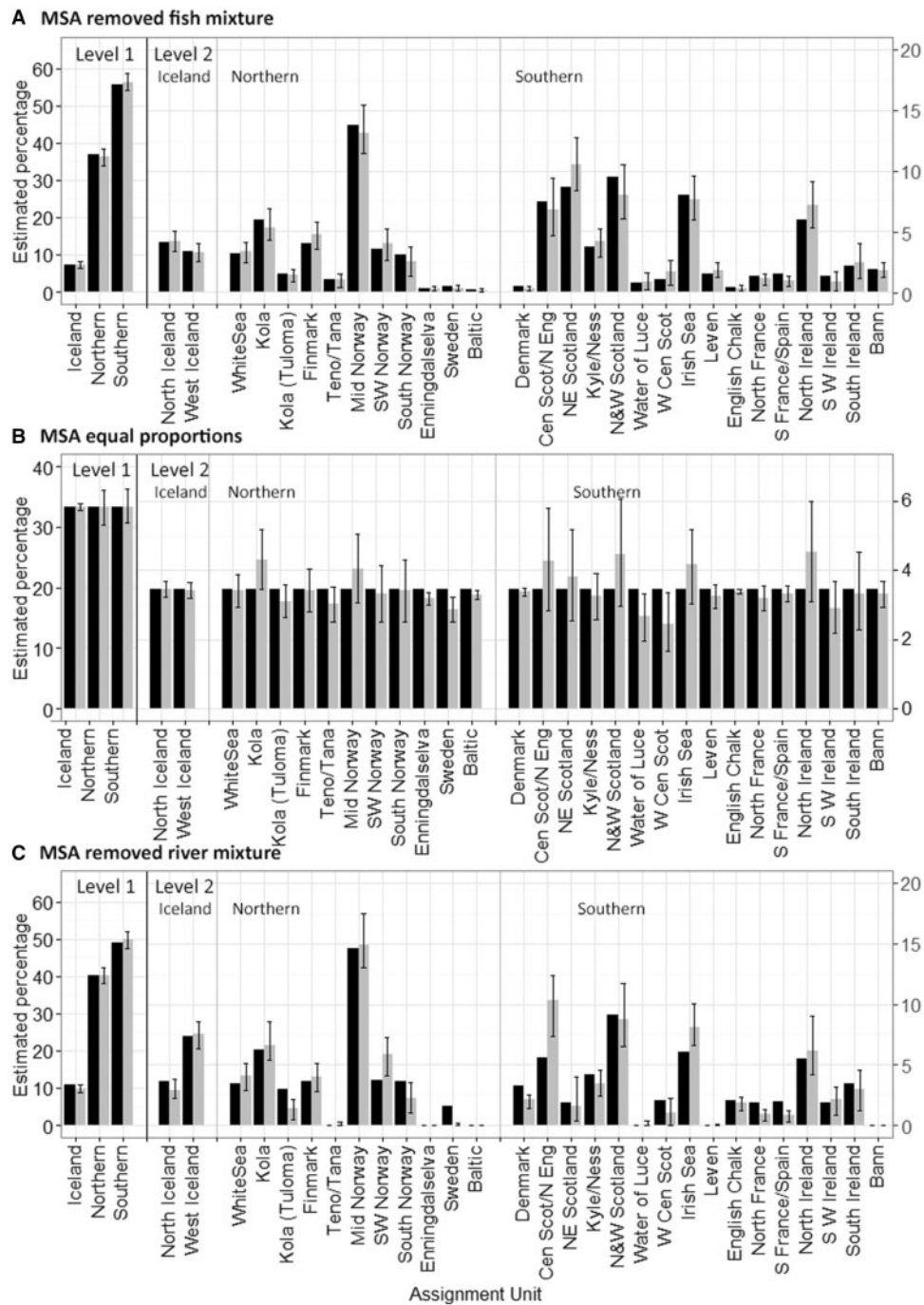


Figure 5. (a) Mixed stock fishery estimates using fish removed from the baseline and used as fishery mixtures. (b) Mixed stock fishery estimates using simulated equal proportions of fish from each assignment unit in the mixture. (c) Mixed stock fishery estimates using entire rivers removed from the baseline and used as fishery mixtures. Dark bars represent actual proportions in the mixture files and grey bars ONCOR estimates. Bars represent mean estimates with 95% confidence intervals around these estimates. NOTE change of Y-axis scale for the Levels 1 and 2 assignment levels.

isolation (Bourret *et al.*, 2013). A second assignment unit, the English Chalk streams, also shared a similarly high mean D_A of 0.236. Griffiths *et al.* (2010) and Ikediashi *et al.* (2018) also reported these rivers in southern England to be highly differentiated from others in the southern part of the European range. However, it is unexpected in the context of the entire European and Icelandic range, that the degree of differentiation matches that of the Baltic.

Within Iceland the salmon populations segregate into northern and western Icelandic units as was also reported by Olafsson *et al.* (2014) which is thought may reflect the patterns of recolonization after the Last Glacial Maximum.

Initially, the northern Level 2 unit subdivided into eleven geographically coherent genetic clusters that matched well with previously reported structure in this region. Bourret *et al.* (2013),

using SNP markers, found separation of northern Norway and Russian rivers from the Norwegian and Swedish Atlantic coast rivers, and Kjærner-Semb *et al.* (2016) found separation of northern and southern Norwegian groupings. Within the northern Norway-Russian complex, Ozerov *et al.* (2017) also found the same North Kola, northern Norway, and Russia-White sea units as reported here. However, their use of 33 microsatellites and a more comprehensive geographical coverage allowed them to define structure at further hierarchical levels within these groups unresolved in the present study using only 14 microsatellites and more limited population coverage.

The population structuring of rivers from across the part of the range covered by the Level 1 southern unit into an initial sixteen Level 2 units accords well with that reported by Griffiths *et al.* (2010) based on 12 microsatellites, 11 of which form part of in the panel used in the present study. Their study encompassed fish from 57 rivers across the southern region but excluded rivers from the East coast of Scotland and northern Ireland and showed similar geographic patterns of genetic structure (their Figure 2). Similar assignment units in France and northern Spain appeared in both analyses and also broadly reflected allozyme-based regional differentiation (Verspoor *et al.*, 2005). However, some differences were seen with some of the units between the two methods used to resolve assignment units. Griffiths *et al.* (2010) identified groupings stretching across both Scotland and Ireland (see their Figure 2) and similar groups were identified here using the STRUCTURE based approach (Figure 3). In contrast, using the distance-based approach the various Scottish and Irish units were clearly separated (Figure 2) to which generally good assignments of fish could be made. Nevertheless, some reciprocal misassignment was still evident (Supplementary Data S7) suggesting a degree of homology between the units. Further, finer-scale investigation is perhaps required to disentangle completely the complex patterns of population grouping within these regions.

Accurate assignments to the initial Level 2 units was not possible at the individual level but was achieved for mixed stock fishery estimates. Acceptable levels of individual assignments could be made to some defined units using the initial split but some areas proved problematic at this scale particularly for Britain and Ireland. This difference reflects the differing power of the two IA and MSA techniques (Manel *et al.*, 2005) and suggests that, when using the baseline for a particular purpose, the required levels of both accuracy and resolution should be defined *a priori*. In turn, this will depend on the specific questions being examined and the tools being utilized.

Overall, the two levels of genetic structure are geographically consistent and in basic agreement with major regional phylogeographic groups previously reported using a variety of markers, suggesting the higher level regional structuring is geographically and temporally robust. In contrast, differentiation between regional units identified at the finer geographic scales may in part be conditioned by human activities, such as the transport and escape of fish from aquaculture facilities, stocking, habitat alteration, fisheries-induced evolution, and indirect genetic changes from disease and ecological disturbances. Such genetic structuring, if defined by such contemporary influences, may not have temporal stability and such lower level units thus will need to be monitored to determine if they are stable. Encouragingly, in a previous assessment of temporal stability on assignment of Atlantic salmon in the species' southern European range (Griffiths *et al.*, 2010), test samples collected 20 years before the baseline samples still showed pre-

dominant allocation back to region of origin. This finding suggests, at least at the larger scale, regional level units are likely to be temporarily stable. However, this should not be assumed to always be the case and a program of resampling should be incorporated if the baseline is exploited in the future.

For the Level 1 and the final Level 2 regional units, all tests of power suggest high accuracies can be achieved with both individual assignments and MSA. Accuracies are improved by use of a probability cut-off of 0.8 for individual assignments, which may be useful in some contexts. However, this will reduce the proportion of fish assigned. Thus in application, the best cut-off will depend on the question address and will need to be decided by each individual user. This will also apply to the assignment units used; if reduced accuracies to some of the combined units are acceptable these may also be used in specific circumstances.

The assignment tests carried out indicate that the described baseline can be exploited to help investigate patterns of ocean utilization and associated differences in marine mortality operating at the regional stock level. However, important quantitative variation linked to how individual population components use the ocean, which may affect mortality rates, also exists at the level of individual rivers within regions and amongst river tributaries (Barson *et al.*, 2015). Evaluation of river-specific problems, likely to exist in some contexts, will require assignments at the individual river level, for which the current baseline appears to have limited usefulness. Nevertheless, even if river-level identification is problematic, identification of region of origin may allow finer scale analysis using higher resolution region-specific baselines.

Resolution of intra-regional population contributions in mixed oceanic samples, including within-river contribution assessments, would be facilitated by further increases in the coverage and resolution of the baseline. Higher resolution is already being achieved in selected areas covered by the baseline reported here (Gilbey *et al.*, 2016a; Vähä *et al.*, 2016; Ozerov *et al.*, 2017). Ideally, future work will likely increase baseline coverage to include most of the estimated 2000 rivers in the North-East Atlantic Commission area. However, this will involve diminishing returns given that the rivers currently in the baseline represent an estimated $\approx 85\%$ of the non-Baltic European adult salmon production. Nevertheless, genetic characterization of as many populations as possible will be important for biodiversity inventory and assessment. Considerable value could also be added by combining the European baseline reported here with North American information to provide a trans-ocean baseline and thus enable oceanic scale investigations. This has already started using a reduced set of microsatellite markers and shows promise in the ability to assign fish from the entire species' range (Gilbey *et al.*, 2016b).

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Acknowledgements

This work forms part of the SALSEA-Merge research project (Project No. 212529) and was funded by the European Union under theme six of the Seventh Framework programme. It was also co-sponsored by the Atlantic Salmon Trust and the Total Foundation, who we thank for financial support. PMcG and JC were partly supported by the Beaufort Marine Research Award in Fish Population Genetics funded by the Irish Government under the Sea Change Programme. The work was also supported under

financial support of the program of fundamental research of Presidium of RAS “Searching fundamental scientific investigations in the interests of development of the Arctic zone of Russian Federation.” The authors also thank the numerous people responsible across Europe whom helped collect samples and assisted with the laboratory analyses. The manuscript benefited greatly from editorial and reviewer comments on an earlier draft and the authors wish to express thanks for their time and comprehensive inputs.

References

- Anderson, E. C., Waples, R. S., and Kalinowski, S. T. 2008. An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 65: 1475–1486.
- Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., *et al.* 2015. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, 528: 405–408.
- Beacham, T., Lapointe, M., Candy, J., Miller, K., and Withler, R. 2004. DNA in action: rapid application of DNA variation to sockeye salmon fisheries management. *Conservation Genetics*, 5: 411–416.
- Beacham, T. D., Candy, J. R., Jonsen, K. L., Supernault, J., Wetklo, M., Deng, L., Miller, K. M., *et al.* 2006. Estimation of stock composition and individual identification of Chinook Salmon across the Pacific Rim by Use of Microsatellite Variation. *Transactions of the American Fisheries Society*, 135: 861–888.
- Bekkevold, D., Helyar, S. J., Limborg, M. T., Nielsen, E. E., Hemmer-Hansen, J., Clausen, L. A. W., and Carvalho, G. R. 2015. Gene-associated markers can assign origin in a weakly structured fish, Atlantic herring. *ICES Journal of Marine Science*, 72: 1790–1801.
- Bourret, V., Kent, M. P., Primmer, C. R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P., *et al.* 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 22: 19.
- Bradbury, I. R., Hamilton, L. C., Chaput, G., Robertson, M. J., Goraguer, H., Walsh, A., Morris, V., *et al.* 2016. Genetic mixed stock analysis of an intercoastal Atlantic salmon fishery in the Northwest Atlantic. *Fisheries Research*, 174: 234–244.
- Bradbury, I. R., Hamilton, L. C., Rafferty, S., Meerburg, D., Poole, R., Dempson, J. B., Robertson, M. J., *et al.* 2015. Genetic evidence of local exploitation of Atlantic salmon in a coastal subsistence fishery in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, 72: 83–95.
- Crozier, W. W., Schön, P.-J., Chaput, G., Potter, E. C. E., Maoiléidigh, N. Ó., and MacLean, J. C. 2004. Managing Atlantic salmon (*Salmo salar* L.) in the mixed stock environment: challenges and considerations. *ICES Journal of Marine Science*, 61: 1344–1358.
- Earl, D., and vonHoldt, B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4: 359–361.
- Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., *et al.* 2011. Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). *Genetica*, 139: 353–367.
- Ensing, D., Crozier, W. W., Boylan, P., O’Maoiléidigh, N., and McGinnity, P. 2013. An analysis of genetic stock identification on a small geographical scale using microsatellite markers, and its application in the management of a mixed-stock fishery for Atlantic salmon *Salmo salar* in Ireland. *Journal of Fish Biology*, 82: 2080–2094.
- Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14: 2611–2620.
- García de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., Aubin-Horth, N., *et al.* 2007. A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews*, 82: 173–211.
- Gauthier-Ouellet, M., Dionne, M., Caron, F., King, T. L., and Bernatchez, L. 2009. Spatiotemporal dynamics of the Atlantic salmon (*Salmo salar*) Greenland fishery inferred from mixed-stock analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 66: 2040–2051.
- Gilbey, J., Cauwelier, E., Coulson, M. W., Stradmeyer, L., Sampayo, J. N., Armstrong, A., Verspoor, E., *et al.* 2016a. Accuracy of assignment of Atlantic Salmon (*Salmo salar* L.) to rivers and regions in Scotland and Northeast England based on single nucleotide polymorphism (SNP) markers. *PLoS ONE*, 11: e0164327.
- Gilbey, J., Wennevik, V., Bradbury, I. R., Fiske, P., Hansen, L. P., Jacobsen, J. A., and Potter, T. 2016b. Genetic stock identification of Atlantic salmon caught in the Faroes fishery. *Fisheries Research*, 187: 110–119.
- Griffiths, A. M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J. L., Bowkett, A. E., 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.
- Ikediashi, C. I., Paris, J. R., King, R. A., Ibbotson, A., and Stevens, J. R. 2018. Atlantic salmon (*Salmo salar* L.) in the chalk streams of England are genetically unique. *Journal of Fish Biology*, 92, in press.
- Jakobsson, M., and Rosenberg, N. A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23: 1801–1806.
- Jones, O. R., and Wang, J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10: 551–555.
- Jonsson, B., Jonsson, N., and Albretsen, J. 2016. Environmental change influences the life history of salmon *Salmo salar* in the North Atlantic Ocean. *Journal of Fish Biology*, 88: 618–637.
- Kalinowski, S. T., Manlove, K. R., and Taper, M. L. 2007. ONCOR: a computer program for genetic stock identification. Department of Ecology, Montana State University. Available from: <http://www.montana.edu/kalinowski/Software/ONCOR.htm>.
- Karlsson, L., and Karlstrom, O. 1994. The Baltic salmon (*Salmo salar* L.): its history, present situation and future. *DANA*, 10: 24.
- King, T., Eackles, M., and Letcher, B. 2005. Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. *Molecular Ecology Notes*, 5: 130–132.
- King, T. L., Kalinowski, S. T., Schill, W. B., Spidle, A. P., and Lubinski, B. A. 2001. Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, 10: 807–821.
- Kjaerner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemela, E., Ozerov, M., *et al.* 2016. Atlantic salmon populations reveal adaptive divergence of immune related genes – a duplicated genome under selection. *BMC Genomics*, 17: 610.
- Koljonen, M. L., and McKinnell, S. 1996. Assessing seasonal changes in stock composition of Atlantic salmon catches in the Baltic Sea with genetic stock identification. *Journal of Fish Biology*, 49: 998–1018.
- Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33: 1870–1874.
- Langella, O. 1999. Populations 1.2.32: a population genetic software. CNRS UPR9034. Available from: <http://bioinformatics.org/~tryphon/populations/>.

- MacKenzie, K. M., Palmer, M. R., Moore, A., Ibbotson, A. T., Beaumont, W. R. C., Poulter, D. J. S., and Trueman, C. N. 2011. Locations of marine animals revealed by carbon isotopes. *Scientific Reports*, 21: 1–6.
- Manel, S., Gaggiotti, O. E., and Waples, R. S. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, 20: 136–142.
- McConnell, S. K., Hamilton, L., Morris, D., Cook, D., Paquet, D., and Bentzen, P. 1995. Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian stocks of Atlantic salmon. *Aquaculture*, 137: 19–30.
- Millar, R. B. 1987. Maximum likelihood estimation of mixed stock fishery composition. *Canadian Journal of Fisheries and Aquatic Sciences*, 44: 583–590.
- Moran, P., Teel, D. J., LaHood, E. S., Drake, J., and Kalinowski, S. 2006. Standardising multi-laboratory microsatellite data in Pacific salmon: an historical view of the future. *Ecology of Freshwater Fish*, 15: 597–605.
- Moriya, S., Sato, S., Azumaya, T., Suzuki, O., Urawa, S., Urano, A., and Abe, S. 2007. Genetic stock identification of Chum Salmon in the Bering Sea and North Pacific Ocean using mitochondrial DNA microarray. *Marine Biotechnology*, 9: 179–191.
- Nei, M., Tajima, F., and Tatenno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, 19: 153–170.
- 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.
- Olafsson, K., Hjorleifsdottir, S., Pampoulie, C., Hreggvidsson, G. O., and Gudjonsson, S. 2010. Novel set of multiplex assay (SalPrint15) for efficient analysis of 15 microsatellite loci of contemporary samples of the Atlantic salmon (*Salmo salar*). *Molecular Ecology Resources*, 10: 533–537.
- Olafsson, K., Pampoulie, C., Hjorleifsdottir, S., Gudjonsson, S., and Hreggvidsson, G. O. 2014. Present-day genetic structure of Atlantic Salmon (*Salmo salar*) in Icelandic Rivers and ice-cap retreat models. *PLoS ONE*, 9: e86809.
- Ozerov, M., Vähä, J. P., Wennevik, V., Niemelä, E., Svenning, M., Prusov, S., Diaz Fernandez, R., et al. 2017. Comprehensive microsatellite baseline for genetic stock identification of Atlantic salmon (*Salmo salar* L.) in northernmost Europe. *ICES Journal of Marine Science*, 74: 2159–2169.
- Paterson, S., Piertney, S. B., Knox, D., Gilbey, J., and Verspoor, E. 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes*, 4: 160–162.
- Peakall, R. O. D., and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6: 288–295.
- Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.
- Rannala, B., and Mountain, J. L. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Science, USA*, 94: 9197–9201.
- Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8: 103–106.
- Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406–425.
- Sanchez, J. A., Clabby, C., Ramos, D., Blanco, G., Flavin, F., Vazquez, E., and Powell 1996. Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). *Heredity*, 77: 423–432.
- Seeb, L. W., Antonovich, A., Banks, M. A., Beacham, T. D., Bellinger, M. R., Blankenship, S. M., Campbell, M. R., et al. 2007. Development of a standardized DNA database for Chinook salmon. *Fisheries*, 32: 540–552.
- Shaklee, J. B., Beacham, T. D., Seeb, L., and White, B. A. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research*, 43: 45–78.
- Shedd, K. R., Dann, T. H., Hoyt, H. A., Foster, M. B., and Habicht, C. 2016. Genetic Baseline of North American Sockeye Salmon for Mixed Stock Analyses of Kodiak Management Area Commercial Fisheries, 2014–2016. ICES Document Fishery Manuscript Series No. 16-03. 233 pp.
- 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.
- Stephenson, J. J., Campbell, M. R., Hess, J. E., Kozfkay, C., Matala, A. P., McPhee, M. V., Moran, P., et al. 2009. A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. *Conservation Genetics*, 10: 1145–1149.
- Takezaki, N., Nei, M., and Tamura, K. 2010. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Molecular Biology and Evolution*, 27: 747–752.
- Torniainen, J., Vuorinen, P. J., Jones, R. I., Keinänen, M., Palm, S., Vuori, K. A. M., and Kiljunen, M. 2013. Migratory connectivity of two Baltic Sea salmon populations: retrospective analysis using stable isotopes of scales. *ICES Journal of Marine Science*, 71: 336–344.
- Vähä, J.-P., Erkinaro, J., Falkegård, M., Orell, P., and Niemelä, E. 2016. Genetic stock identification of Atlantic salmon and its evaluation in a large population complex. *Canadian Journal of Fisheries and Aquatic Sciences*, 74: 1–12.
- Verspoor, E., Beardmore, J. A., Consuegra, S., Garcia de Leaniz, C., Hindar, K., Jordan, W. C., Koljonen, M. L., et al. 2005. Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *Journal of Fish Biology*, 67: 3–54.
- Verspoor, E., Consuegra, S., Fridjonsson, O., Hjorleifsdottir, S., Knox, D., Olafsson, K., Tompsett, S., et al. 2012. 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*, 69: 1625–1636.
- Webb, J., Verspoor, E., Aubin-Horth, N., Romakkaniemi, A., and Amiro, P. 2007. The Atlantic Salmon. In *The Atlantic Salmon: Genetics, Conservation and Management*, pp. 17–45. Ed. by E. Verspoor, L. Stradmeyer, and J. L. Nielsen. Blackwell Publishing, Oxford, U.K.

Handling editor: W. Stewart Grant