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

John Gilbey[1]*, Jamie Coughlan[2], Vidar Wennevik[3], Paulo Prodöhl[4], Jamie R. Stevens[5], Carlos Garcia de Leaniz[6], Dennis Ensing[7], Eef Cauwelier[1], Corrine Cherbonnel[8], Sofia Consuegra[9¶], Mark W. Coulson[10ŧ], Tom F. Cross[2], Walter Crozier[7], Eileen Dillane[2], Jonathan S. Ellis[5†], Eva García-Vázquez[11], Andrew M. Griffiths[5], Sigurdur Gudjonsson[12], Kjetil Hindar[13], Sten Karlsson[13], David Knox[1], Gonzalo Machado-Schiaffino[11‡], Dorte Meldrup[14], Einar Eg Nielsen[14], Kristinn Ólafsson[15], Craig R. Primmer[16ɟ], Sergey Prusov[17], Lee Stradmeyer[1], Juha-Pekka Vähä[16ŧ], Alexei Je. Veselov[18], Lucy M.I. Webster[10+], Philip McGinnity[2Δ] and Eric Verspoor[19Δ]

[1]*Marine Scotland, Freshwater Fisheries Laboratory, Faskally, Pitlochry PH16 4LB, UK*

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

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

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

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

[6]*Department of Biosciences, Swansea University, Swansea, United Kingdom*

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

[8]*GENINDEXE, 6 rue des Sports, 17000 La Rochelle, France*

[9]*IBERS, Aberystwyth University, Aberystwyth, United Kingdom*

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

[11]*Departament of Functional Biology, Genetics, Universidad de Oviedo, C/Julian Claveria s/n, 33006 Oviedo, Spain*

[12]*Marine and Freshwater Research Institute, Skúlagata 4, 101 Reykjavík, Iceland*

[13]*Norwegian Institute for Nature Research (NINA), PO Box 5685 Sluppen, 7485 Trondheim, Norway*

[14]*DTU Aqua, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøve 39, 8600 Silkeborg, Denmark*

[15]*Matis ohf., Vinlandsleid 12, 113 Reykjavik, Iceland*

[16]*Department of Biology, University of Turku, 20014 Turku, Finland*

[17]*Knipovich Polar Research Institute of Marine Fisheries and Oceanography, 6 Knipovich Street, Murmansk, 183763, Russia*

[18]*Institute of Biology, Karelian Research Institute, Pushkinskaya 11, 10 185610 Petrozavodsk, Russia*

[19]*Marine Scotland, Freshwater Fisheries Laboratory, Faskally, Pitlochry PH16 4LB, UK and Rivers and Lochs Institute, University of the Highlands and Islands, Inverness College, 1 Inverness Campus, Inverness, IV2 5NA, UK*

**Corresponding Author: tel*: +44 1796 472945; *fax:* 01796 473523; *email: John.Gilbey@gov.scot*

[Δ]*Philip McGinnity and Eric Verspoor equally share senior co-authorship*

[¶]*Present address: Department of Biosciences, Swansea University, Swansea, United Kingdom*

[ŧ]*Present address: Rivers and Lochs Institute, University of the Highlands and Islands, Inverness College, 1 Inverness Campus, Inverness, IV2 5NA, UK*

[†]*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: Science and Advice for Scottish Agriculture, Roddinglaw Road, Edinburgh, EH12 9FJ, UK*

[ɔ]*Present address: Department of Biosciences and, Biotechnology Institute, University of Helsinki, 00014, Finland*

[ŧ]*Present address: Water and Environment of Western Uusimaa, POB 51, FI-08101 Lohja, Finland*

Running title: Stock identification of European Atlantic salmon

Jamie Stevens 31/7/2017 13:32
Deleted: *y*

Abstract

Atlantic salmon (*Salmo salar* L.) populations of different river origins mix in the North Atlantic during the marine life stage. To facilitate stock identification in this environment, we developed a genetic baseline covering the European component of the species' range from the Russian River Megra in the north-east, the Icelandic Ellidaar in the west, and the Spanish Ulla in the south. Coverage extends over 3700 km North to South and over 2700 km East to West. The baseline encompasses 26,822 fish from 13 countries, 282 rivers and 467 sampling sites screened for 14 microsatellites. A hierarchical subdivision of regional genetic assignment units was defined using a combination of distance based and Bayesian clustering. A top level assignment level of three units was identified comprising Northern, Southern and Icelandic regions. A second assignment level was also defined composed of 18 regional units where individual assignments could be accurately performed and 29 units where accurate mixed stock estimates were possible. This baseline represents the most comprehensive population coverage for an Atlantic salmon genetic data-set, and constitutes a unique resource in the European marine fisheries context and is freely available to researchers to facilitate identification of the natal origin of European salmon.

Key words: Atlantic salmon, genetic stock identification, individual assignment, marine ecology, microsatellites

Introduction

The homing behaviour of Atlantic salmon together with other factors such as phylogeography, founder effects, isolation, selection and genetic drift, has resulted in local adaptations (Garcia de Leaniz et al., 2007) and significant population structuring at a hierarchy of levels from intra-river to inter-continental (King et al., 2001). Variation in marine migratory patterns of these various Atlantic salmon populations from different parts of the species range are known to occur but the full extent of differences among

93 populations and how this may be changing in response to shifting environmental conditions
94 remains to be resolved (Jonsson et al., 2016). The study of population and stock-specific
95 migration, distribution and feeding patterns, and their implications for marine mortality
96 rates, and the impact of climate change, are hampered by a lack of information relating to
97 the marine-phase of the lifecycle (Crozier et al., 2004). This makes it extremely difficult to
98 mitigate anthropogenic influences on different stock components in mixed-stock fisheries or
99 bycatches. Effective marine ecosystem management of Atlantic salmon could greatly benefit
100 from accurate identification of genetically distinct populations and regional entities
101 (MacKenzie et al., 2011) and the ability to discriminate the stock origins of fish in mixed
102 feeding aggregations or during migratory phases.
103 Methods for identifying the region or river/tributary of origin of salmonids using DNA
104 profiling have advanced significantly over recent decades and are now widely applied to
105 stock management of Pacific salmon (*Oncorhynchus* spp.) (e.g. Shaklee et al., 1999;
106 Beacham et al., 2004; Beacham et al., 2006; Shedd et al., 2016). Their application to Atlantic
107 salmon stock management has been less extensive but has provided valuable insights into
108 stock mixing at several scales including intercontinental (e.g. North American and European
109 stocks in the West Greenland fishery: Gauthier-Ouellet et al., 2009), regional (e.g. stock
110 composition in Canadian gill-net fisheries: Bradbury et al., 2016) and river level (e.g.
111 population structuring in the river Teno: Vähä et al., 2016).
112 The assignment of salmon to continent of origin is now routine and increasingly cost-
113 effective (Gauthier-Ouellet et al., 2009). On the Western side of the Atlantic, several genetic
114 baselines are available (e.g. Bradbury et al., 2015; Sheehan et al., 2010) including a recently
115 developed fine scale range-wide North American microsatellite baseline (Bradbury et al.,
116 2016). Together, these allow the within-region identification of fish originating from

Western Atlantic populations at high geographic resolution. Although partial baselines have been developed on 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) there is no high resolution resource covering the entire species' (non-Baltic) eastern Atlantic range. A DNA-based approach to the GSI of marine samples from the Eastern Atlantic would, in conjunction with ecological studies, provides a basis for advancing understanding of the migration and distribution patterns of Atlantic salmon. This would help to improve our knowledge of factors influencing marine mortality, and facilitate the implementation of more effective management programmes (Crozier et al., 2004).

Advances in DNA profiling over recent years have allowed for the development of genetic stock identification (GSI) using various genetic markers. Allozymes (Koljonen and McKinnell, 1996) and mitochondrial DNA (Moriya et al., 2007) have both been successfully used for stock identification in salmonid species including Atlantic salmon. However, the levels of resolution achieved with such markers have been insufficient for intra-regional discrimination. Historically microsatellites have been the genetic marker most widely used with Atlantic salmon. Various studies have screened numerous populations of salmon over many years resulting in high resolution coverage of many parts of the species' range. As such, even with the development of other markers such as Single Nucleotide Polymorphisms, the large body of microsatellite data available provides a powerful resource for GSI. The use of microsatellite data does, however, come with certain problems (reviewed in Moran et al., 2006) which include: 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; whether the forward or reverse primer is labelled; and differences in primer sizes. All of these can result in

141 inconsistent allele-size designations. Nevertheless, evidence from large-scale
142 standardisation projects with these marker types among Pacific salmonid species such as
143 *Oncorhynchus mykiss* (Stephenson et al., 2009) and *Oncorhynchus tshawytscha* (Seeb et al.,
144 2007), together with previous studies in Atlantic salmon (e.g. Ellis et al., 2011), suggest that
145 these issues can be overcome to construct comprehensive integrated genetic baselines
146 (Moran et al., 2006).

147 This study builds on existing national and international microsatellite screening
148 programmes to develop a comprehensive database of microsatellite variation. It contains
149 data for a common set of 14 microsatellite loci for a geographically extensive range of rivers,
150 spanning the species' Eastern Atlantic European range from the Russian river Megra in the
151 north-east (66.151 N, 41.484 W), to the Icelandic Ellidaar in the west (64.117 N, 21.833 E)
152 and the Spanish Ulla river in the south (42.639 N, 8.761 E). Samples encompass rivers
153 responsible for about ~85% of wild-salmon production in the eastern Atlantic (estimate
154 based on rod-catch data from numerous sources). Baltic salmon populations are excluded
155 from the baseline, as they do not migrate outside the Baltic Sea (Karlsson and Karlstrom,
156 1994; Torniainen et al., 2013), though one Baltic sample was included as a genetic out-group
157 to represent this region. Existing and new data supplied by partners in a multi-laboratory
158 trans-European consortium were calibrated (Ellis et al., 2011), subjected to stringent quality
159 control and integrated to form the baseline.

160 The baseline was constructed to identify the region of origin of marine-phase salmon
161 in the Eastern Atlantic, and a hierarchical approach was used to partition the baseline into
162 genetically distinctive regional assignment units. The power and accuracy of assignment to
163 these units were assessed using both simulations and test samples constructed by removing

fish from the dataset, and the utility of the baseline for regional assignment of salmon from the marine environment was evaluated.

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) (Fig. 1, Table 1, Supplementary data S1 & S2), including one Baltic River acting as a genetic out-group. Sampled sites spanned the entire eastern Atlantic range of the species covering a range of 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. In general, they were from juvenile fish, mostly parr and fry, but in some cases from smolts or mature salmon, sampled when returning to fresh water to spawn. Numbers sampled at a site ranged from 11 to 300 with a mean of 58, and rivers were characterised by 1 to 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 material (S1 & 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*

188 (Sanchez et al., 1996), *Ssa14*, *Ssa289* (McConnell et al., 1995), *Ssa171*, *Ssa197*, *Ssa202*
189 (O'Reilly et al., 1996) *SSsp1605*, *SSsp2201*, *SSsp2210*, *SSsp2216*, *SSspG7* (Paterson et al.,
190 2004), *SsaD144*, *SsaD157* (King et al., 2005) and *SSsp3016* (unpublished, GenBank number
191 AY37820).

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

210 Based on the standardisation rules, all data generated for baseline sites was converted
211 to the standard size ranges and stored in a single bespoke database for further analysis (see

212 Ellis et al., 2011 for full details). Sib-ship analysis among individuals in each sample was
213 investigated using COLONY (Jones and Wang, 2010) and used to exclude all but one fish
214 from each full-sib family in each sample prior to inclusion in the database. Fish with less
215 than 10 microsatellite loci genotyped were removed from further analysis due to concerns
216 with DNA and genotype quality. Sites with more than half of the loci out of Hardy-Weinberg
217 equilibrium (examined in GENEPOP 4.2.2; Rousset, 2008) (potentially not representative of a
218 single population), or had less than 70% of fish scored at all loci (potentially poor quality
219 DNA and genotypes), or consisted of less than 30 individuals after quality control checks
220 listed above (potential failure to provide accurate estimates of allele frequencies), were also
221 removed. We estimated descriptive statistics with GenAlEx 6 (Peakall and Smouse, 2006).
222 **Assignment units**
223 Assignment units were defined in an iterative way similar to that employed by Gilbey et al.
224 (2016a). Initial units were first defined by a combination of distance based and Bayesian
225 clustering. Individual assignment accuracies using these units were then examined and units
226 where accuracies did not meet a predefined threshold were combined with units which saw
227 reciprocal misassignments until all units had accuracies at or above the threshold level.
228 The distance based approach was based on a neighbour-joining tree (Saitou and Nei,
229 1987) constructed using Nei's genetic distance D_A (Nei et al., 1983) calculated in POPTREE2
230 (Takezaki et al., 2010) and visualised in MEGA7 (Kumar et al., 2016). The clustering approach
231 was carried out in STRUCTURE (Pritchard et al., 2000), using a burn-in of 100,000 and a run
232 phase of 300,000 iterations during each application. Three replicates for each cluster
233 number (*K*) were run with values of *K* from 1 to 10. *K* = 10 was chosen as an upper limit after
234 examination of the results of the runs while they were underway which showed in each case
235 estimates 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 Δ*K* 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.

We used 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, as above. 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 if one technique had identified a single unit where the other had identified sub-units with this then the sub-units were added to the initial 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 then calculated for each of these units (see below). If 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 combined with other units to which reciprocal misassignments were taking place. 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 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). Accuracies were based on fish randomly removed from the reference baseline and combined into a mixture file. A random 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 means that fish whose origin is difficult to determine (low probability) are removed from the analysis and so potential accuracy can be increased (Gilbey et al., 2016a; Bekkevold et al., 2015). However, the application of cut-off scores also increases the proportion of unassigned fish (Gilbey et al., 2016a) and will thus influence apparent stock proportions if calculated from the individual assignments, and so this should never be performed. In order to estimate accurate stock proportions a Mixed Stock Analysis approach was therefore utilised (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 removed fish 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. 100 fish were simulated for each unit and confidence intervals of the estimates calculated using 1000 bootstraps.

Baseline coverage analysis – River removal

A baseline rarely completely covers all possible source populations, and so some fish in fishery mixtures may be from populations either not sampled or 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 (Waples 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 which 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 material 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 the 551 sites, 17 sites were removed as genotypes 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 reliable 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 details of sites are contained in Fig. 1, Table 1 and Supplementary data S1 & S2. Most loci across sites were highly variable with allele numbers ranging from 10 for *Ssa14* to 46 for *SsaD157* (mean 29.9). Additional descriptive and diversity estimates for each locus and site appear in Supplementary material S3.

Definition of initial assignment regions

A neighbour-joining tree of Nei's D_A is summarised in Fig. 2 with an expanded version with all nodes labelled detailed in Supplementary data S4 and full site level D_A matrix in Supplementary data S5. A plot of Δ*K*, and a map showing the geographic positioning of the clusters at each hierarchical STRUCTURE level are shown in Fig. 3. Assignment units as defined by POPTREE and STRUCTURE are compared in Supplementary data S6.

Both POPTREE and STRUCTURE 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. In general there was also good agreement between the two techniques in the lowest level units identified. POPTREE identified 26 distinct units (coloured differently in Fig. 2) and STRUCTURE identified 22 (lowest level splits on Fig. 3) and in 17 cases the same units were identified by both techniques (Supplementary data S6). Using the lowest level split in from each technique a total of 29 units were identified for the initial Level 2 assignment accuracy tests (column 1 in Table 2, Supplementary data S6).

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 accuracy of all fish to the Northern, Southern and Icelandic unit respectively was 90.8%, 92.7% and 99.5%. 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 being assigned using such a cut-off.

Assignment accuracy of fish with probability scores ≥ 0.8 to the Level 2 units was ≥ 80% in 19 of the 29 units (Table 2; for full breakdown of assignments at each Level 2 iterative level see Supplementary data S7). Assignment units which contained reciprocal misassignments were then combined resulting in 21 assignment units and accuracy

355 recalculated resulting in 18 of the 21 achieving accuracies ≥ 80%. A final round of
356 assignment unit combination resulted in 18 assignment units at which assignment
357 accuracies were all ≥ 80% (Table 2, Supplementary data S7). Initial assignment units at both
358 levels are mapped in Fig. 1 with D_A matrixes detailed in Supplementary data S8.
359 *100% simulations*
360 The 100% simulations for each assignment unit showed robust estimates of stocks
361 proportions at both assignment levels (Fig. 4). At Level 1, the mean estimates matched the
362 estimated proportions extremely well with a maximum difference of just 0.3% between the
363 actual and estimated values and all upper CI at 100%. At the initial Level 2 assignment units
364 again showed relatively accurate estimates with an average difference between the
365 estimated and actual mean proportions of 4.5%. However individual units did perform
366 below this with a maximum difference with the West and Central Scotland level of 17.6%. At
367 the first round of assignment unit combinations accuracies are seen to improve as expected
368 with average and maximum differences between the estimated and actual mean
369 proportions of 4.5% and 9.0%. These levels reduced to 1.9% and 8.0% respectively at the
370 final assignment unit combination round.
371 *Mixed stock analysis*
372 The results of the MSA using the 2682 fish removed from the baseline and used as a fishery
373 mixture are shown in Fig. 5A. At all assignment units within both assignment levels
374 estimated proportions matched actual proportions (were within the CI bands) apart from a
375 single unit in Level 2, South France/Spain where the upper CI was just 0.19 below the actual
376 value. The estimates are also seen to be very precise with average CI bands of just 2.2 and a
377 maximum of 4.7. Considering the high accuracy of the mixed stock estimates at this initial

assignment unit composition, no further assignment unit combinations are presented for mixed stock analysis.

Equal proportions

As with the previous analysis the equal proportion simulation shows excellent agreement between the actual and estimated proportions in the mixture (Fig. 5B). At Level 1 there is an average difference between actual and estimated of just 0.06% and a maximum of 0.09% and at Level 2 these two differences only rises to a mean difference of 0.4 and a maximum of 1.1%.

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 its fish assigned to region using the remaining rivers. However, even here relatively high levels of accuracy were obtained (Fig. 5C). There is an average difference between actual and estimated mixture proportions of just 1.9% and a maximum of 2.3% at Level 1 and 1.3% and 2.9% respectively at Level 2. At no time were significant proportions assigned to any of the six assignment units consisting of a single river which did not therefore have representatives in the mixture file (lower CI at zero in these units).

Discussion

The study presented here represents the largest analysis of Atlantic salmon population structure. The results demonstrate the utility of microsatellites to successfully assign Atlantic salmon from the NE Atlantic to regions of origin. The genetic baseline developed represents a powerful resource to better understand the biology of Atlantic salmon in the marine environment. The use of this resource may help to understand the causes of

differential mortality among salmon stocks and inform a more efficient management of Atlantic salmon fisheries (Crozier et al., 2004).

Distance based and cluster analysis both revealed substantial hierarchical sub-structuring of river populations of European and Icelandic salmon. At the highest level, structure is related to large-scale geographical discontinuities between Scandinavia-Russia, Iceland, and the southern region (Britain-Ireland-France-Denmark-Spain) populations. These units are similar to those identified in previous analyses of population structure in salmon. For example, King et al. (2001) showed with microsatellites an unambiguous separation of Iceland, Norway and Scotland-Ireland-Spain (their Fig. 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 among the more central range groups).

At the second level, two assignment units shared the largest average degree of distinctiveness to other units and this is reflected in the fact that both are on the extreme end of the neighbour-joining tree (Fig. 2). The Baltic unit had a mean D_A of 0.236 to other units (Supplementary data S8) and this significant differentiation to other European rivers has been seen in previous studies (Bourret et al., 2013). This divergence is consistent with the restricted migration of Baltic salmon (Karlsson and Karlstrom, 1994) and the Baltic Sea's long history of geographical isolation (Bourret et al., 2013). Interestingly, a second assignment unit, the English Chalk streams also shared this same very high mean D_A of 0.236. Griffiths et al. (2010) and Ikediashi et al. (2018) also found these rivers highly differentiated in comparison to others in the southern part of the European range but it is perhaps unexpected that the degree of differentiation matches that of the Baltic when the entire European and Icelandic range is examined.

426 Salmon populations in the Icelandic region segregated into two distinct units. This
427 division into Northern and Western Icelandic units was also reported by Olafsson et al.
428 (2014) and may reflect the patterns of recolonisation after the Last Glacial Maximum.
429 Initially the Northern Level 2 unit was subdivided into eleven geographically
430 coherent second-level genetic clusters that match well with previously reported structure in
431 this region. Bourret et al. (2013), using SNP markers, found separation of northern Norway
432 and Russian rivers from the Norwegian and Swedish Atlantic coast rivers, and Kjærner-Semb
433 et al. (2016) found separation of northern and southern Norwegian groupings. Within the
434 northern Norway-Russian complex, Vähä et al. (2016) also found the same North Kola,
435 Northern Norway and Russia-White sea units, as reported here. Their use of 33
436 microsatellites and a more comprehensive baseline coverage allowed them to define
437 structure at further hierarchical levels within these groups, which was not apparent with the
438 14 microsatellites and site coverage used here.
439 The population structuring of rivers from across the part of the range covered by the
440 Southern Level 1 unit into an initial sixteen Level 2 units is coherent with that reported by
441 Griffiths et al. (2010). They used 12 microsatellites, 11 of which were also contained in the
442 panel used here, on fish sampled from 57 rivers across the Southern region, but excluding
443 rivers from the East coast of Scotland and Northern Ireland. They reported similar
444 geographic patterns of genetic structure as in this study (their Fig. 2). Similar assignment
445 units in France and Northern Spain appeared in both analyses and also broadly reflected
446 allozyme-based regional differentiation (Verspoor et al., 2005).
447 Despite these similarities, differences were seen with some of the units and between
448 the two methods identifying the assignment units. Griffiths et al. (2010) identified groupings
449 stretching across both Scotland and Ireland (see their Fig. 2) and similar groups were

identified here using the STRUCTURE based approach (Fig. 3). However in the distance based approach the various Scottish and Irish units were clearly separated (Fig. 2) and this was also reflected in the final assignment units where accuracy of assignment could be made between the two geographic areas. Nevertheless, misassignments were still evident between the Irish and Scottish units (Supplementary data S7) suggesting a degree of homology between the units.

Accurate assignments to the initial 32 Level 2 units was not possible using individual assignments but was achieved when using a mixed stock fishery approach. Acceptable levels of individual assignments could be made to some units using the initial split but some areas proved problematic at this scale particularly within the UK/Ireland areas. This observation reflects the differing power of the two 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* and this will depend on the specific questions being examined and the tools being utilised.

Overall, the two levels of genetic structure are geographically coherent and in basic agreement with regional groups previously reported. This agreement over many studies and across marker types suggests the higher level regional structuring identified is likely to be geographically and temporally robust. However, differentiation between the identified regional units at the finer geographic scales may be influenced 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.

Potential genetic changes resulting from these contemporary influences mean that the temporal stability of contributing populations will require future monitoring. In a

474 previous examination of temporal stability on assignment of Atlantic salmon in the species'
475 southern European range (Griffiths et al., 2010), test samples collected 20 years before the
476 baseline samples still showed predominant allocation back to region of origin. This finding
477 suggests at least at the larger regional levels temporal stability may be temporarily stable,
478 however this should not be assumed to be the case for all units, and a program of
479 resampling should be incorporated into future developments using these reference baseline
480 populations.

481 At Level 1 and the final Level 2 assignment units all tests of assignment power
482 suggest high accuracies can be achieved with both individual assignments and mixed stock
483 analysis. The use of an assignment probability cut-off of 0.8 for individual assignments will
484 always improve assignment accuracies; however, this comes at the cost of the proportion of
485 fish assigned. The actual cut-off used will thus depend on the situation under investigation
486 and will be a decision for the investigator/manager at the time. Further the actual
487 assignment units may also be varied. If reduced accuracies to some of the combined units
488 are acceptable these may also be used in specific circumstances.

489 Based on the various assignment tests, the baseline described can be exploited to
490 investigate patterns of ocean utilisation and associated differences in marine mortality at
491 the regional stock level, however important quantitative variation linked to how individual
492 population components use the ocean and which may affect mortality rates, also exists at
493 the level of individual rivers within regions and among river tributaries (Barson et al., 2015).
494 Evaluation of river-specific problems in some contexts will require information at the
495 individual river level, for which the current baseline may have limited usefulness. However,
496 even if river level identification is problematic, identification of region of origin may allow
497 further analysis using region specific baselines of higher resolution.

The identification of intra-regional population contributions in mixed samples will be facilitated by further increases in the coverage and resolution of the baseline. This is an on-going process; for example, higher resolution is already being achieved in selected areas covered by the baseline reported here using other markers (Gilbey et al., 2016a; Ozerov et al., 2017; Vähä et al., 2016). Future baseline development will likely increase the coverage of the baseline reported here towards the estimated 2000 rivers in the North-East Atlantic Commission area. However, diminishing returns will apply given that the rivers currently in the baseline represent an estimated ~85% of the non-Baltic European adult salmon production.

Considerable value could be added by combining the European baseline reported here with North American information to provide a trans-ocean baseline to enable oceanic scale investigations. This has already begun in a limited way 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). A trans-Atlantic baseline is likely to benefit from identification of strategic, level specific, diagnostic markers for continental, regional and intra-river groupings.

Supplementary data

Supplementary material

is available at the ICESJMS online version of the manuscript.

Acknowledgments

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 7th 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

553 divergence across the natural range of Atlantic salmon (*Salmo salar*). Molecular Ecology, 22:
554 19.
555 Bradbury, I. R., Hamilton, L. C., Chaput, G., Robertson, M. J., Goraguer, H., Walsh, A., Morris, V., et al.
556 2016. Genetic mixed stock analysis of an interceptory Atlantic salmon fishery in the
557 Northwest Atlantic. Fisheries Research, 174: 234-244.
558 Bradbury, I. R., Hamilton, L. C., Rafferty, S., Meerburg, D., Poole, R., Dempson, J. B., Robertson, M. J.,
559 et al. 2015. Genetic evidence of local exploitation of Atlantic salmon in a coastal subsistence
560 fishery in the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences, 72: 83-
561 95.
562 Crozier, W. W., Schön, P.-J., Chaput, G., Potter, E. C. E., Maoiléidigh, N. Ó., and MacLean, J. C. 2004.
563 Managing Atlantic salmon (*Salmo salar* L.) in the mixed stock environment: challenges and
564 considerations. ICES Journal of Marine Science, 61: 1344-1358.
565 Earl, D., and vonHoldt, B. 2012. STRUCTURE HARVESTER: a website and program for visualizing
566 STRUCTURE output and implementing the Evanno method. Conservation Genetics
567 Resources, 4: 359-361.
568 Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., et al.
569 2011. Microsatellite standardization and evaluation of genotyping error in a large multi-
570 partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). Genetica,
571 139: 353-367.
572 Ensing, D., Crozier, W. W., Boylan, P., O'Maoiléidigh, N., and McGinnity, P. 2013. An analysis of
573 genetic stock identification on a small geographical scale using microsatellite markers, and
574 its application in the management of a mixed-stock fishery for Atlantic salmon *Salmo salar* in
575 Ireland. Journal of Fish Biology, 82: 2080-2094.
576 Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using
577 the software structure: a simulation study. Molecular Ecology, 14: 2611-2620.
578 Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., Aubin-Horth,
579 N., et al. 2007. A critical review of adaptive genetic variation in Atlantic salmon: implications
580 for conservation. Biological Reviews, 82: 173-211.
581 Gauthier-Ouellet, M., Dionne, M., lanie, Caron, F., ois, King, T. L., and Bernatchez, L. 2009.
582 Spatiotemporal dynamics of the Atlantic salmon (*Salmo salar*) Greenland fishery inferred
583 from mixed-stock analysis. Canadian Journal of Fisheries and Aquatic Sciences, 66: 2040-
584 2051.
585 Gilbey, J., Cauwelier, E., Coulson, M. W., Stradmeyer, L., Sampayo, J. N., Armstrong, A., Verspoor, E.,
586 et al. 2016a. Accuracy of Assignment of Atlantic Salmon (*Salmo salar* L.) to Rivers and
587 Regions in Scotland and Northeast England Based on Single Nucleotide Polymorphism (SNP)
588 Markers. PLoS ONE, 11: e0164327.
589 Gilbey, J., Wennevik, V., Bradbury, I. R., Fiske, P., P., H. L., Jacobsen, J. A., and Potter, T. 2016b.
590 Genetic stock identification of Atlantic salmon caught in the Faroes fishery. Fisheries
591 Research, 187: 110-119.
592 Griffiths, A. M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J. L., Bowkett, A. E.,
593 Minting, P., et al. 2010. Genetic stock identification of Atlantic salmon (*Salmo salar*)
594 populations in the southern part of the European range. BMC Genetics, 11:31.
595 Ikediashi, C. I., Paris, J. R., Ibbotson, A., and Stevens, J. R. (2018) Atlantic salmon (*Salmo salar* L.) in
596 the chalk streams of England are genetically unique. Journal of Fish Biology, 92: in press.
597 Jakobsson, M., and Rosenberg, N. A. 2007. CLUMPP: a cluster matching and permutation program
598 for dealing with label switching and multimodality in analysis of population structure.
599 Bioinformatics, 23: 1801-1806.
600 Jones, O. R., and Wang, J. 2010. COLONY: a program for parentage and sibship inference from
601 multilocus genotype data. Molecular Ecology Resources, 10: 551-555.
602 Jonsson, B., Jonsson, N., and Albretsen, J. 2016. Environmental change influences the life history of
603 salmon *Salmo salar* in the North Atlantic Ocean. Journal of Fish Biology, 88: 618-637.

Jamie Stevens 31/7/2017 13:37
Formatted: Font:Italic

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 at 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, 1: DOI: 10.1038/srep00021.

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 Tateno, 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 abalysis of 15 microsatellite loci of comtemporary 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.

654 Ozerov, M., Vähä, J. P., Wennevik, V., Niemelä, E., Svenning, M., Prusov, S., Diaz Fernandez, R., et al.
655 2017. Comprehensive microsatellite baseline for genetic stock identification of Atlantic
656 salmon (*Salmo salar* L.) in northernmost Europe. ICES Journal of Marine Science, fsx041. doi:
657 10.1093/icesjms/fsx041.
658 Paterson, S., Piertney, S. B., Knox, D., Gilbey, J., and Verspoor, E. 2004. Characterization and PCR
659 multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.)
660 microsatellites. Molecular Ecology Notes, 4: 160-162.
661 Peakall, R. O. D., and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic
662 software for teaching and research. Molecular Ecology Notes, 6: 288-295.
663 Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of Population Structure Using
664 Multilocus Genotype Data. Genetics, 155: 945-959.
665 Rannala, B., and Mountain, J. L. 1997. Detecting immigration by using multilocus genotypes.
666 Proceedings of the National Academy of Science, USA., 94: 9197-9201.
667 Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for
668 Windows and Linux. Molecular Ecology Resources, 8: 103-106.
669 Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing
670 phylogenetic trees. Molecular Biology and Evolution, 4: 406-425.
671 Sanchez, J. A., Clabby, C., Ramos, D., Blanco, G., Flavin, F., Vazquez, E., and Powell 1996. Protein and
672 microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). Heredity, 77: 423-
673 432.
674 Seeb, L. W., Antonovich, A., Banks, M. A., Beacham, T. D., Bellinger, M. R., Blankenship, S. M.,
675 Campbell, M. R., et al. 2007. Development of a standardized DNA database foe Chinook
676 salmon. Fisheries, 32: 540-552.
677 Shaklee, J. B., Beacham, T. D., Seeb, L., and White, B. A. 1999. Managing fisheries using genetic data:
678 case studies from four species of Pacific salmon. Fisheries Research, 43: 45-78.
679 Shedd, K. R., Dann, T. H., Hoyt, H. A., Foster, M. B., and Habicht, C. 2016. Genetic Baseline of North
680 American Sockeye Salmon for Mixed Stock Analyses of Kodiak Management Area
681 Commercial Fisheries, 2014–2016. Fishery Manuscript Series No. 16-03. Alaska Department
682 of Fish and Game. Anchorage, Alaska. 233 pp.
683 Sheehan, T. F., Legault, C. M., King, T. L., and Spidle, A. P. 2010. Probabilistic-based genetic
684 assignment model: assignments to subcontinent of origin of the West Greenland Atlantic
685 salmon harvest. ICES Journal of Marine Science, 67: 537-550.
686 Stephenson, J. J., Campbell, M. R., Hess, J. E., Kozfkay, C., Matala, A. P., McPhee, M. V., Moran, P., et
687 al. 2009. A centralized model for creating shared, standardized, microsatellite data that
688 simplifies inter-laboratory collaboration. Conservation Genetics, 10: 1145-1149.
689 Takezaki, N., Nei, M., and Tamura, K. 2010. POPTREE2: Software for Constructing Population Trees
690 from Allele Frequency Data and Computing Other Population Statistics with Windows
691 Interface. Molecular Biology and Evolution, 27: 747-752.
692 Torniainen, J., Vuorinen, P. J., Jones, R. I., Keinänen, M., Palm, S., Vuori, K. A. M., and Kiljunen, M.
693 2013. Migratory connectivity of two Baltic Sea salmon populations: retrospective analysis
694 using stable isotopes of scales. ICES Journal of Marine Science, 71: 336-344.
695 Vähä, J.-P., Erkinaro, J., Falkegård, M., Orell, P., and Niemelä, E. 2016. Genetic stock identification of
696 Atlantic salmon and its evaluation in a large population complex. Canadian Journal of
697 Fisheries and Aquatic Sciences: 1-12.
698 Verspoor, E., Beardmore, J. A., Consuegra, S., Garcia de Leaniz, C., Hindar, K., Jordan, W. C., Koljonen,
699 M. L., et al. 2005. Population structure in the Atlantic salmon: insights from 40 years of
700 research into genetic protein variation. Journal of Fish Biology, 67: 3-54.
701 Verspoor, E., Consuegra, S., Fridjonsson, O., Hjorleifsdottir, S., Knox, D., Olafsson, K., Tompsett, S., et
702 al. 2012. Regional mtDNA SNP differentiation in European Atlantic salmon (Salmo salar): an
703 assessment of potential utility for determination of natal origin. ICES Journal of Marine
704 Science, 69: 1625-1636.

Waples, R. S., Kalinowski, S. T., and Anderson, E. C. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences, 65: 1475-1486.

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

Country	Pre-QC			Post-QC		
	Rivers	Sites	Fish	Rivers	Sites	Fish
Denmark[1]	3	6	253	2	4	189
England[2,3]	24	38	1652	23	35	1498
Finland[4]	2	5	395	2	5	393
France[2,3,5,6]	13	16	759	9	9	450
Iceland[7]	17	25	2352	16	22	1986
Ireland[8]	29	45	2345	29	40	2053
Northern Ireland[9]	9	20	1469	7	18	1302
Norway[4,10,11]	90	109	7749	81	99	7008
Russia[4,10, 12]	33	36	2506	30	33	2350
Scotland[3]	87	230	11625	69	185	8884
Spain[6]	7	7	342	4	4	190
Sweden[1,4]	4	4	180	4	4	172
Wales[2]	7	10	375	6	9	347
Total	325	551	32002	282	467	26822

714 Institutions contributing data: [1] Danish Institute for Fisheries Research, Denmark; [2] University of Exeter,
715 England; [3] Marine Scotland Science, Scotland; [4] University of Turku, Finland; [5] Geneindex, France; [6] University
716 of Oviedo, Spain; [7] Marine and Freshwater Research Institute, Iceland; [8] University College Cork, Ireland; [9]
717 Queen's University Belfast & Agri-Food and Biosciences Institute Northern Ireland, Northern Ireland; [10]
718 Institute of Marine Research, Norway; [11] Norwegian Institute for Nature Research, Norway, [12] Knipovich Polar
719 Research Institute of Marine Fisheries & Oceanography, Russia.
720
721
722