1. Article type: Original article

Ice sheets and genetics: insights into the phylogeography of Scottish Atlantic salmon, *Salmo salar* L.

Eef Cauwelier*#1, Eric Verspoor#1,2, Mark W. Coulson2,3, Anja Armstrong3, David Knox1, Lee Stradmeyer1, Lucy M. I. Webster3‡, and John Gilbey1

1Marine Scotland Science, Freshwater Fisheries Laboratory, Faskally, Pitlochry, PH16 5LB, UK, 2Rivers and Lochs Institute, University of Highlands and Islands Inverness College, Inverness Campus, Inverness IV2 5NA, 3Rivers and Fisheries Trusts of Scotland, Capital Business Centre, 24 Canning Street, Edinburgh, EH3 8EG, UK.

*Correspondence: Eef Cauwelier, Marine Scotland Science, Freshwater Fisheries Laboratory, Faskally, Pitlochry, PH16 5LB, UK.

E-mail: E.Cauwelier@marlab.ac.uk

#Joint first authors

‡Present address: Wildlife DNA Forensics – Diagnostics, Wildlife & Molecular Biology Section, Science and Advice for Scottish Agriculture, Roddinglaw Road, Edinburgh, EH12 9FJ, UK

Running head Phylogeography of Scottish Atlantic salmon
ABSTRACT

Aim We constructed an independent phylogeographic hypothesis (IPH) for Atlantic salmon (Salmo salar L.) in Scotland and adjacent northern England and tested if the distribution of contemporary microsatellite variation accorded with the IPH.

Location NW Europe

Methods Knowledge of post-Last Glacial Maximum (LGM) landscape development and salmon biology were used to define a temporal sequence of late Quaternary ecological niche models and, combined with the central-marginal (CM) model of diversity evolution, to formulate an IPH. Observed microsatellite locus variation for salmon from 102 rivers in Scotland and northern England was tested for agreement with the IPH.

Results The IPH postulated 1) initial colonization (c.17 ka BP) of multiple isolated deglaciating coastal regions by long distance dispersal, predominantly from SW Europe via the Atlantic, 2) intra-region colonization dominated by local range expansion, 3) differentiation by founder effects/genetic drift, in line with CM, until deglaciation complete and regions merged (c. 10 ka BP). IPH regions were weakly but significantly differentiated, showed broad agreement with genetically defined regions, and genetic diversity within them positively correlated with time post-deglaciation, conforming to the CM model. The greatest differentiation was seen for the most recently diverged populations, with evidence that accumulating homoplastic mutations muted older regional phylogeographic signatures.

Main conclusions Microsatellite differentiation in Scottish Atlantic salmon is significantly but weakly conditioned by post-LGM deglaciation, with a mosaic of regional groups and a temporal cascade of within-region CM effects caused by local range expansion as patches became deglaciated. Colonization was predominantly or exclusively from the south-western European LGM refugium. Accumulating homoplastic mutations have increasingly reduced regional differentiation.
Keywords
Atlantic salmon, colonization, dispersal, drift, ecological niche model, glaciation, mutation, phylogenetics, phylogeography

INTRODUCTION
Understanding of a species’ phylogeography is crucial to defining and conserving contemporary intraspecific biodiversity (Moritz & Potter, 2013). In glaciated regions, population genetics theory suggest phylogenetic biodiversity has been conditioned by Quaternary ice sheet dynamics and associated climatic change (e.g. Hewitt, 2004; Ursenbacher et al., 2015) but specific understanding for most species is limited and poorly supported. Proposed causal links usually represent inferences drawn directly from genetic data and post-hoc associations with selected aspects of deglaciation and involve unstated assumptions of historical species distributions. This risks giving incorrect and shallow perceptions of the specifics of historical change associated with glacial cycles and of other conditioning factors (e.g. drift, selection or mutation). Patterns of genetic variation can potentially arise for multiple reasons; given variation can be influenced by a complex historical interplay of multiple genetic and demographic processes (Richards et al., 2007; Arroyo-Santos et al., 2014).

To clarify causal historical factors and avoid “just so” stories, the testing of independent phylogeographic hypotheses (IPHs) is essential (Nielsen & Beaumont, 2009). This provides a more robust and specific understanding and can be developed from ecological niche models (ENMs) which can define expected changes in a species’ historical biogeography (e.g. Richards et al., 2007). IPHs can be generated from reconstructions of Quaternary landscape and climate change, consideration of a species’ habitat needs and dispersal biology and
population genetic models of recolonization processes (e.g. Eckert et al., 2008; Excoffier & Ray, 2008). The IPHs can then be tested for their capacity to explain observed patterns of diversity distribution (Richards et al., 2007; Ye et al., 2014).

The Atlantic salmon (Salmo salar L.) is potentially a model species for this approach and for testing spatial models of diversity evolution, such as the central-marginal (CM) model (e.g. Micheletti & Storfer, 2015). A detailed understanding of the species’ general biology (Aas et al., 2011) and genetics (e.g. Verspoor et al., 2007; King et al., 2007; Griffiths et al., 2010; Bourret et al., 2013; Bradbury et al., 2015) exists. They show fine scale spatial population structuring, with deep to shallow phylogenetic divergence on trans-Atlantic to within-river geographical scales across their North Atlantic distribution and, while divergence generally scales with geographic distance, there are notable exceptions (King et al., 2007), including some sharp boundaries in regional differentiation (Makhrov et al., 2005; Tonteri et al., 2005; Bradbury et al., 2015).

Existing studies of species’ phylogeography (e.g. Säissä et al., 2005; Nilsson et al., 2001; Tonteri et al., 2005; Finnegan et al., 2013) suggest Quaternary ice sheet dynamics underlie some population differentiation. However, most draw inferences directly from genetic data and post-hoc associations with post-glacial landscape development. None independently test differing possible phylogeographic hypotheses. Thus, the actual effect of glaciers remains vague and equivocal, even on basic questions, such as the number of LGM refugial areas for salmon (Finnegan et al., 2013; Olafsson et al., 2014).

Here the IPH approach is applied to a phylogeographic analysis of salmon diversity in Scotland and northern England and tested using a new database of microsatellite variation. No prior studies of molecular genetic diversity in this part of the species’ range (Verspoor & Jordan, 1989; Verspoor et al., 1991; Jordan et al., 2005; Griffiths et al., 2010; Gilbey et al., 2012, 2016) address the species’ phylogeography of this area. The objectives were to 1)
develop a regional IPH from post-LGM landscape development for Scotland and adjacent NE England and 2) test if the IPH accounted for a significant component of the regional microsatellite differentiation among populations. Additionally, the study objective was to explore the general usefulness of the IPH approach for wider application to salmon and other species.

MATERIALS AND METHODS

IPH development

Maps summarising basic landscape change in Britain were constructed from the latest reviews of deglaciation and sea-level dynamics (Carr et al., 2006; Shennan et al., 2006; Clark et al., 2012; Brooks et al., 2011) and transformed into ENMs using understanding of contemporary Atlantic salmon marine and freshwater distribution and biology. The species is fundamentally anadromous and a temperate-subarctic aquatic generalist (Webb et al., 2007; Aas et al., 2011), able to live in close proximity to ice sheets (Nielsen, 1961; MacCrimmon & Gots 1979). It shows strong but imperfect philopatry (Webb et al., 2007). When rivers or tributaries are newly habitable, colonization by local range expansion (LRE) occurs (e.g. Nielsen, 1961; Perrier et al., 2011, 2014) where local populations exist but it can colonize by long distance dispersal (LDD) over 1,000s of kilometres (Nielsen, 1961; King et al., 2007).

Dispersal will likely be greater from newly established populations where genes promoting straying will increase in frequency (Quinn et al., 2001; Phillips et al., 2010). Given salmon show structuring into small local populations (King et al., 2007; Webb et al., 2007) genetic drift, including founder effects, will be a major driver of population divergence as in the central-marginal (CM) model (Eckert et al., 2008). This model predicts decreasing diversity from refugial areas to the margins of the expanding range through an increasing sequential loss of diversity from a temporal cascade of drift events.
Initial LDD could have been from the main LGM refugial area in south-western Europe, south of the ice sheet (Consuegra et al., 2002; Kettle et al., 2011) spanning rivers in the Bay of Biscay, Channel/Manche and southern North Sea (Kettle et al., 2011; Ménat et al., 2006). A second refugial area can be inferred to have existed in North America (King et al., 2007). A further refugial area has been proposed in north-east Russia to explain genetic patterns (Säisä et al., 2005; Tonteri et al., 2005) but lacks independent support.

**Genetic data and analysis**

Genetic data was derived from 12,724 juvenile salmon electrofished from 357 sites in 95 rivers across Scotland and seven rivers in northern England (Fig. 1, Table S1 in Appendix S1). Fin clips were taken from anaesthetised fish, the fish released and clips stored in 99% ethanol. DNA was extracted (Knox et al., 2002) and amplified for three multiplexes encompassing fourteen microsatellites (Ellis et al., 2011 – excluding SsaD486) plus SsaD48, SsaD71 (King et al., 2005) and SP1608 (Paterson et al., 2004). PCR used 10-150 ng DNA and Qiagen TypeIt kit (Qiagen); for final primer concentrations and PCR conditions see Table S2 in Appendix S1. Fragments were separated by MegaBACE 1000 DNA analyser (GE Healthcare) and genotyped using Fragment Profiler (GE Healthcare).

Full sibs in samples were identified by COLONY2 (Jones & Wang, 2010) using the pedigree likelihood approach, assuming biparental polygamy and no inbreeding, and only one full sib from each family retained to avoid inflating genetic differences among samples by family effects (Hansen et al., 1997). The number of breeders (Nb) was also estimated and the Nb/Ns ratio calculated; Ns being the initial sample size. This ratio is an indicator of the effective genetic size of sampled populations standardized for sample size, which varied among sites.
Hardy-Weinberg proportions were tested using FSTAT (Goudet, 1995) and a Bonferroni correction applied for multiple tests. GENETIX (Peakall & Smouse, 2012) was used to assess linkage disequilibrium, based on 100 randomisations (Belkhir et al., 2004). Average number of alleles ($A$), observed ($H_o$) and expected ($H_e$) heterozygosity, pairwise $F_{ST}$ and $D_A$ (Nei et al., 1983) were calculated using GENALEX (Peakall & Smouse, 2012). Allelic richness ($A_r$) was determined with HP-RARE (Kalinowski, 2005), standardized to $N = 20$.

Potential neutrality of the microsatellite variation was evaluated using a hierarchical island model $F_{ST}$ outlier test (ARLEQUIN - Excoffier & Lischer, 2010), based on genetic groups derived from the ENM, neighbour-joining (NJ) tree and clustering analysis (details below). CODIDI (Wang, 2015) was used to assess the effects of mutation on inter-population differentiation.

A $D_A$–based bootstrapped (500 replicates) NJ tree was constructed (POPULATIONS - Langella, 2002) and drawn (MEGA6 - Tamura et al., 2013). Population relatedness was also assessed with STRUCTURE 2.3.4 (Pritchard et al., 2000) using the admixture model assuming correlated allele frequencies (Falush et al., 2003). STRUCTURE was run with a burn-in and run phase of 100,000 and 150,000 iterations, respectively, for five replicates for each number of clusters ($K$) and increasing $K$ values until $\text{Ln } P(K)$ plateaued. Location priors were used to improve resolution (Hubisz et al., 2009) and the log-likelihood probability and $\Delta K$ (Evanno et al., 2005) calculated with STRUCTURE HARVESTER (Earl & von Holdt, 2012) to identify the smallest $K$ that captured the main structure in the data (Pritchard & Wen, 2004). Individual membership coefficients were combined across replicates with CLUMPP (Jakobsson & Rosenberg, 2007), employing the Greedy method with random order input (1,000 repeats), and repeated hierarchically on identified sub-clusters until either within-river structuring was observed or $K = 1$ was most likely. Graphical output was generated with DISTRUCT (Rosenberg, 2004).
Hierarchical analysis of molecular variance (AMOVA) was carried out (ARLEQUIN – Excoffier & Lischer, 2010) for the groups defined by the ENM, NJ Tree and STRUCTURE analyses. Kruskal-Wallis tests were employed to assess regional heterogeneity in $A_r$, $H_e$ and $N_d/N_s$. The time of deglaciation of sample locations was determined from the maps in Clark et al. (2012). Spearman rank correlations of diversity ($A_r$ and $H_e$) as a function of timing of deglaciation were tested in R (R Core team, 2013) for ENM, NJ tree, and STRUCTURE defined regions, as well as the whole study area. The probability values across regions were combined by Fisher’s method (Sokal & Rohlf, 1995) to give an overall test of within-region association.

RESULTS

Phylogeographic model

Study area deglaciation started in eastern Scotland after c. 19,000 years (19 ka) BP (Fig. 2), encompassing an area with two small coastal rivers (Fig. 1: 91-92) and from 17-16 ka BP, further isolated deglaciated regions emerged, all with small habitable coastal rivers: the SW (6-8), Argyll (17-20) on the west, the southern Outer Hebrides (44-49) and opposite west coast (29-56), the northern Outer Hebrides (50-53), the most NW tip of the mainland (57-59), the most NE tip of the mainland (66-68), the east (80-94), coastal SE (95-98) and SE rivers (99-101) that were part of larger central North Sea rivers. By 12 ka BP, further SW rivers (1-2) had emerged. Over time, these deglaciating regions enlarged and variously merged with the last regions deglaciated being the rivers Leven (15) and Rhiconich (57) and the headwaters of some large eastern rivers (74, 79, 93). Deglaciation was complete by 10 ka BP (Fig. 2).

Ice sheet decay reconstruction indicated a potential for seven large and three small (SW Scotland bordering England, the northern Outer Hebrides and the NW tip of the mainland)
phylogeographic groups to have evolved and diverged. Initial LDD is suggested to be predominantly or exclusively from the western part of the SW European refugial area via the Atlantic, with possible LDD from North America and the eastern part of the SW refugial area via the North Sea, given the uncertainty of the North Sea deglaciation timing (Clark et al. 2012). With this spatial deglaciation mosaic and LDD, regional differentiation would primarily be driven by founder effects and initial genetic drift within regions would be expected. Due to sequential LRE colonization and further drift, CM dynamics are expected to cause genetic diversity to decrease from the first to the last river colonized within a region and, within rivers, from lower to upper reaches, if this variation is effectively neutral and drift effects dominate those of mutation.

**Genetic clustering of sites**

Excepting the NW tip of the mainland, sampling covered all ENM regions and only one sample site deviated from HW proportions after Bonferroni correction (Table S1 in Appendix S1). LD was significant in 8.7% sites and involved all microsatellites while allele number ranged from 9 in Ssa14 to 77 for SsaD48 (mean ± SD = 34.4 ± 19.9). Diversity and pairwise differentiation of sites are detailed in Appendix S1 (Tables S1 and S3, respectively).

The constructed NJ tree based on sample $D_A$’s (Fig. 3) had 39 nodes (10.5%) with bootstrap values > 50, most occurring for samples from the rivers Kyle (74), Leven (15), Tay and Forth (93, 94), in Galloway (3 – 8) and Ayrshire (9 – 13). However, at a higher level, eighteen reasonably coherent geographical groupings of rivers resolved. For the AMOVA analysis and Kruskal-Wallis tests, those sites in the geographic sub-trees exhibiting no spatial affinity were excluded.

The individual relatedness-based STRUCTURE analysis initially showed $\Delta K$ peaks at $K = 2, 3$ and 7, with log-likelihood probabilities then plateauing (Fig. S1a in Appendix S1). At $K = 7$, the single rivers Leven (15; Fig. 4: light blue) and Ouse (101; Fig. 4: brown)
formed distinct clusters with six further coherent sub-clusters: Eden/Nith/Annan/Galloway
(1-8, 102; Fig. 4: orange), Ayrshire/Clyde (9-14; Fig. 4: red), West and North (16-72; Fig. 4: dark blue), Kyle/Cromarty/Ness (73-79; Fig. 4: purple), NE (80-92; Fig. 4: green) and the Tay/Forth/SE (93-101; Fig. 4: red/green mix).

Further sub-cluster analysis (Fig. S1b-I in Appendix S1) gave a total of 26 groups (Fig 4), three within the Eden/Annan/Nith/Galloway region (Fig. 4: oranges): Luce (8), the other Galloway rivers (3-7) and Eden/Nith/Annan (Figs. 4 and 5). The Clyde and Ayr (12) were distinct (Fig. 4: reds) and nine groups identified in the West and North (Fig. 4: dark blues): the Awe (22), four groups in the Outer Hebrides (44-53), two in NE corner (66-71), the Creran (24) and the remainder of the West and North rivers. The Kyle/Cromarty/Ness area (Fig. 4: purples) presented six groups: two groups in the Kyle (74) system, the Ness (79) and three in the Cromarty area. The NE coast samples showed no further sub-structuring while the Tay/Forth/SE cluster (Fig. 4: yellows) had within-river sub-structuring, with samples from the upper Tay (93) and Forth (94) being different. These 26 clusters were used in subsequent analysis.

Tests for selection and mutation

Independent of how the regions were defined, no loci were outliers when examining and simulating $F_{ST}$ versus $H_e$, consistent with being selectively neutral. Analysis across loci of the association between $G_{ST}$ and $H_S$ showed a significant negative correlation ($r = -0.56$, permutation test: $p = 0.038$), suggesting significant homoplasy (Wang, 2015).

Tests of IPH

AMOVA analysis showed most variation (~97.8%) resided within sites (Table 2) but significant differentiation was found, both among sites, as well as regions, for all methods.
Regional differences accounted for < 1% of variation, half of that among sites within regions. NJ Tree and STRUCTURE defined regions accounted for twice the amount of the overall variation as did ENM regions and had less variation within regions. Significant regional differences were manifested for $A_r$, $H_e$ and $N_i/N_s$ (Table 2) and a significant negative relationship was also found of time of deglaciation with $A_r$ and $H_e$ (Table 3), across regions (Fig. 6), as well as within regions, except for regions defined by clustering analysis.

In the NJ tree analysis, each defined ENM region was broadly represented within the tree topography with 46/102 rivers (45.1%) clustering with other rivers in their respective regions (Fig. 3). Just over half of the sites (54.2%) corresponded when comparing ENM regions to STRUCTURE regions (Fig. 5), in part due to sub-structuring within some ENM regions (e.g. regions 2, 3, 6 and 8 – Fig. 5). Correspondence between NJ tree and STRUCTURE defined regions was low (28.8%).

**DISCUSSION**

**Robustness of the IPH**

Multiple reconstructions of post-LGM deglaciation of Scotland and surrounding areas exist but that of Clark et al. (2012), which underpins the IPH, is both the latest and most comprehensive reconstruction, encompassing a broad synthesis of data across multiple studies. With studies of sea level change (Shennan et al., 2006; Brooks et al., 2011), there is arguably a good consensus on the basic deglaciation dynamic of landscape in the study area. The only substantive uncertainty is the pattern and timing of northern North Sea deglaciation (Clark et al., 2012; Carr et al., 2006), where two scenarios remain possible. However, both share the same pattern and timing of terrestrial deglaciation in the study area and suggest the same basic IPH. They differ only in that one scenario sees an earlier North Sea deglaciation
allowing for a possibility of earlier colonization by migrants from rivers in the south-eastern North Sea, as well as from the Atlantic from SW Europe.

As for the marine ENM developed by Kettle et al. (2011), the IPH assumes that the salmon’s ecological niche is unchanged. While niches can change (e.g. Takahashi et al., 2014), the critical aspects of the ecology of anadromous salmon are likely to be the same e.g. temperature needs, fluvial habitat and philopatry; these being typical of all northern salmonids, including contemporary northern salmon populations (Webb et al., 2007).

Combined with a typically fragmented spawning habitat, both within and among rivers (Webb et al., 2007; King et al., 2007; Aas et al., 2011), most likely, once established, pre-Holocene stocks in deglaciating areas rapidly formed multiple small increasingly genetically independent and locally adapted spawning populations with little inter-population gene flow.

During the establishment phase, the success of initial strayers would be expected to initially selectively elevate straying tendencies in new populations, given the trait is heritable (e.g. Quinn et al., 2001; Phillips et al., 2010). However, once at carrying capacity, straying rates would decline, with selection in established populations favouring philopatry and local adaptation (Garcia de Leaniz et al., 2007), a process requiring only a few generations (Quinn et al., 2001; Phillips et al., 2010). This dynamic allows for colonization, strong founder effects and drift and the development of phylogeographic structure.

**Genetic support for IPH – regional differentiation**

Significant differentiation of IPH regions supports the conditioning of contemporary microsatellite diversity by post-LGM deglaciation, as does their broad agreement with genetically defined clusters. It accords with there being multiple independent LDD events and explains the lack of a south-north pattern of differentiation predicted by the CM model (Eckert et al., 2008; Excoffier & Ray, 2008). In contrast, within the regions, the CM model predictions are manifested in the negative correlation of genetic diversity with timing of
deglaciation, supporting initial LDD colonization being then dominated by within region LRE, as deglaciating regional patches expanded.

The existence of differentiation of IPH regions is consistent with strong founder effects and initial genetic drift and small numbers long distance strayers, as expected given the dispersal distances involved. While time required for LLD to establish the first populations is unclear, based on contemporary observations on salmon colonization (Nielsen, 1961; Perrier et al., 2010, 2014), subsequent LRE into newly available adjacent rivers and tributaries within regions probably took just decades, once rivers or tributaries became habitable, and probably involved relatively greater numbers of founders. Yet, the observed population differentiation is greater within regions.

The IPH accounts for only a small (0.4%) though significant portion of inter-sample regional differentiation in the study area, certainly compared to that seen in some other areas (Tonteri et al., 2009; Bradbury et al., 2015). Also, regional boundaries are poorly defined, possibly reflecting the mis-assignment of populations to ENM regions due to the unavoidable crudeness of their definition, given that fine-scale accounts of post-LGM landscape development exist for only a few areas (e.g. the Lomond region - MacLeod et al., 2011, Kintyre – Finlayson et al. 2014). Certainly, in some boundary areas, some populations are more closely related to populations in the adjacent region (Fig. 5). Boundary regions may also encompass populations of mixed regional ancestry and mismatches may partly reflect inaccurate reconstruction of landscape development or peculiarities of local historical marine straying. In SW Scotland, two highly distinct small-scale regional groups resolved, something that probably reflects the eastern part being colonized by LRE from NW England and the rest colonized from the west.

Some inconsistencies may also arise from contemporary factors, such as genetic drift caused by population reductions associated with over-exploitation, habitat degradation
(Verspoor et al., 2007) or genetic changes resulting from stocking cultured fish (Perrier et al. 2013; Oserov et al. 2016) or introgression of farm genes (Glover et al. 2013). Thus, though there is broad support for the IPH, further work is clearly needed to understand local deviations from the IPH, ideally encompassing more extensive sampling, and other types of markers.

Genetic drift might be expected to increase differentiation of populations over time, both among and within regions, given very low levels of gene flow between populations once established (Webb et al., 2007). Yet, regional differentiation, while significant, is surprisingly small, possibly due to the cumulative effect of low levels of historical gene flow among populations. However, the negative correlation between $G_{ST}$ and $H_S$ found suggests that historical differentiation may be increasingly muted by accumulating homoplastic mutations. Such mutations have been reported for microsatellites (Estoup et al., 2002) and may lead to underestimation of general phylogenetic differentiation using this marker type (Wang, 2015). These mutations would most erode the signature of older drift events, such as those underlying regional differentiation and diversity clines and would explain the dearth of significant clustering nodes in the higher levels of the phylogenetic tree (Fig. 3).

This dearth could also explain why the greatest divergence in the study area is seen for the most recently established salmon stocks, e.g. the recently bottlenecked and re-established River Ouse stock (101; Axford, 1991) and for the Kyle (74) and Leven (15) river systems, the last of the large catchments to be deglaciated. These three groups of stocks also show the lowest genetic diversity (Table S1 in Appendix S1), in line with less time for homoplasy to increase diversity. If so, this suggests both caution in interpreting highly divergent groups in recently deglaciated areas as ancient lineages (e.g. Baltic stocks – Sääsä et al., 2005), at least based on microsatellite data, and that other marker types, such as mitochondrial or nuclear SNPs, may be more informative (Bradbury et al. 2013).
Genetic Support for IPH – origin of colonists

Given regional deglaciation conditions, the IPH sees LLD as predominantly from the south-western part of the southern European glacial refugial area, the only evidenced refugial area for salmon in Europe and which probably encompassed the southern North Sea (Consuegra et al., 2002; Kettle et al., 2011). While some parts of the species’ range show evidence of colonization from both European and North American refugial areas, such as the Kola Peninsula rivers (Mahkrov et al. 2005; King et al., 2007; Bradbury et al., 2015), no evidence was found that this was substantively the case in the study area.

The current study is unclear on the question of whether the study area was colonized from both the western and eastern sides of the European refugial area. The IPH is that it was dominated by migrants from the western Atlantic side rather than the southern North Sea. That contemporary stocks are significantly differentiated between the two sides of this refugial area (Griffiths et al. 2010; Finnegan et al., 2013) suggests that, if both did contribute, some degree of consistent differentiation of eastern and western populations might be present. However, while not found, this can only be robustly assessed in an analysis encompassing data from populations from both the western and eastern parts of the SW refugial area.

Alternative causes for population differentiation

No known contemporary environmental factors, e.g. temperature, ocean bathymetry or current patterns, show spatial correlation with the landscape regions defined. Furthermore, the spatially unstructured differentiation among regions and the distribution of locus-specific $F_{ST}$ values fitting neutral expectations support a non-adaptive explanation for the observed patterns. While selection acting on some of the variation cannot be ruled out (e.g. Haasl & Payseur, 2012), the contemporary temporal stability of patterns of microsatellite variation (King et al., 2007) and correlation of levels of diversity with time since deglaciation observed
here, accord with a non-adaptive, phylogeographic basis for the regional differentiation found.

Observed spatial clines in genetic diversity could reflect clines in the contemporary genetically effective size of populations, given the high degree of population structuring seen within rivers in this species (Youngson et al., 2003; King et al., 2007), as population size decreases with decreasing stream order, as available habitat reduces. Smaller coastal rivers would deglaciate and become habitable first and, in larger rivers, higher tributaries would deglaciate last. In both cases, samples would derive from smaller, less diverse populations.

Indeed, there was a weak but significant association among populations within ($\chi^2 = 31.14$, df = 16, $p = 0.013$) and across (Spearman rank correlation: $r_s = 0.137$, $p = 0.01$) the ENM-defined regions of timing of deglaciation with $N_e/N_s$. However, this association of this contemporary measure with a historical parameter is likely an artefact of contemporary spatial influences on population size, as described. However, the relationship of population size to genetic diversity is complex (e.g. Ardren & Kapuscinski, 2003).

**Usefulness of IPH approach**

Many questions remain regarding the impact of post-LGM landscape development on contemporary genetic diversity in the Atlantic salmon. However, the IPH analysis, the first for this species, points to the potential for the approach to deliver new and more robust insights into the effect of Quaternary landscape and climate change on the distribution of molecular genetic variation in salmon, as well as other species. Arguably, insights gained from testing IPHs will be more robust than those drawn directly from genetic data alone and arbitrary post hoc inferences of associations with landscape change (Arroyo-Santos et al., 2014). Certainly, they provide an independent perspective.
The potential for exploiting the IPH approach is facilitated by the rapidly expanding extent and robustness of reconstructions of Quaternary landscape change (e.g. Shaw et al. 2006; Clark et al., 2012; Stroeven et al., 2016). These can be particularly informative for species such as the Atlantic salmon for which there exist both a robust knowledge of their relevant biology as well as large genetic data sets from across the species’ range (e.g. mtDNA - Nilsson et al., 2001; Verspoor et al., 2012; nuclear SNPs - Bourret et al., 2013; Bradbury et al., 2015; Gilbey et al., 2016; microsatellites - Ellis et al., 2011; Bradbury et al., 2016). This makes the Atlantic salmon an ideal model species for exploring the IPH approach and advancing general understanding of how post-glacial landscape and climate change has specifically conditioned contemporary patterns of phylogenetic diversity.

ACKNOWLEDGEMENTS

Genetic data was collected and analysed as part of the European Union (7th Framework) SALSEA-MERGE and a cooperative Rivers and Fisheries Trusts Scotland and Marine Scotland Science FASMOP research project. We are grateful to fisheries biologists and field assistants from Trusts, District Salmon Fisheries Boards in Scotland and the Environment Agency in northern England, who collected many of the samples, and are too numerous to mention. Without their input this study would not have been possible.

REFERENCES


Goudet, J. (2001) *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3).*

salmon (Salmo salar) populations in the southern part of the European range. *BMC Genetics*, 11, 31.


populations from the Barents and White seas, northwest Russia. *Canadian Journal of Fisheries and Aquatic Science, 66*, 717-735.


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix 1 Table S1: Diversity parameters for each site

Appendix 1 Table S2: Microsatellite details and PCR conditions

Appendix 1 Table S3: Pairwise estimates of $F_{ST}/D_A$.

Appendix 1 Figure S1: LnP(K)/ΔK values of STRUCTURE.
BIOSKETCHES

**Eef Cauwelier** is a molecular genetics analyst. Her interests are focused on fish ecology and population genetics as it relates to fisheries management, and in the evolutionary processes driving the distribution of contemporary biodiversity, including the effects of historical landscape development.

**Eric Verspoor** is an ecological geneticist whose focus is the application of molecular population genetics to advancing understanding of the nature of contemporary aquatic biodiversity in salmonids and its application to biodiversity management, with a particular interest in the role of local adaptation and historical factors in defining intraspecific diversity.

**Author contributions** EC, EV and MC developed the conceptual framework for the paper. EC and EV lead the writing of the manuscript, with input from MC. EV generated the historical landscape maps and developed the associated ENMs and colonization models. EC led the data QC and analysis with contributions from MC and EV. EV coordinated the field sampling for the database with JG, LS, MC and LW, and DK carried out standardization and optimisation of microsatellite screening protocols. AA, EC, MC, LW and LS carried out the genotyping and dealt with related QC issues, assisted by DK. JG and LS constructed and populated the database, respectively, and JG contributed to standardization of genotyping and to data analysis.

Editor: Jim Provan
Table 1 Results from AMOVA analysis, with regions determined by a) ENM, b) NJ Tree and c) clustering analysis. SS: Sum of Squares.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SS</th>
<th>Variance</th>
<th>% explained</th>
<th>Fixation indices</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) ENM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>among regions</td>
<td>695.3</td>
<td>0.03</td>
<td>0.4</td>
<td>$F_{CT}$: 0.004</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>among sites within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regions</td>
<td>5176.3</td>
<td>0.12</td>
<td>1.8</td>
<td>$F_{SC}$: 0.018</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>within sites</td>
<td>158361.8</td>
<td>6.35</td>
<td>97.8</td>
<td>$F_{ST}$: 0.022</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>b) NJ Tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>among regions</td>
<td>1315.7</td>
<td>0.05</td>
<td>0.8</td>
<td>$F_{CT}$: 0.008</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>among sites within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regions</td>
<td>4276.8</td>
<td>0.09</td>
<td>1.5</td>
<td>$F_{SC}$: 0.015</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>within sites</td>
<td>149475.3</td>
<td>6.35</td>
<td>97.8</td>
<td>$F_{ST}$: 0.022</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>c) Clustering analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>among regions</td>
<td>1706.9</td>
<td>0.06</td>
<td>0.9</td>
<td>$F_{CT}$: 0.009</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>among sites within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regions</td>
<td>4164.7</td>
<td>0.08</td>
<td>1.3</td>
<td>$F_{SC}$: 0.013</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>within sites</td>
<td>158361.8</td>
<td>6.35</td>
<td>97.8</td>
<td>$F_{ST}$: 0.022</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Table 2 Results of Kruskall-Wallis tests for regional heterogeneity of site-specific genetic diversity ($N_b/N_s$, $A_r$ and $H_e$) for regions defined by a) ENM b) NJ Tree and c) clustering analysis.

<table>
<thead>
<tr>
<th></th>
<th>Regional differences</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) ENM regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_b/N_s$</td>
<td>20.95</td>
<td>8</td>
<td></td>
<td>0.007</td>
</tr>
<tr>
<td>$A_r$</td>
<td>95.06</td>
<td>8</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>$H_e$</td>
<td>60.02</td>
<td>8</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>b) NJ Tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_b/N_s$</td>
<td>32.37</td>
<td>17</td>
<td></td>
<td>0.014</td>
</tr>
<tr>
<td>$A_r$</td>
<td>121.05</td>
<td>17</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>$H_e$</td>
<td>130.08</td>
<td>17</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>c) Clustering analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_b/N_s$</td>
<td>44.30</td>
<td>25</td>
<td></td>
<td>0.010</td>
</tr>
<tr>
<td>$A_r$</td>
<td>148.00</td>
<td>25</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>$H_e$</td>
<td>145.65</td>
<td>25</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Table 3 Association of timing of deglaciation with site-specific genetic diversity ($A_r$ and $H_e$) for a) ENM b) NJ Tree and c) clustering analysis, and d) across the whole study area using Spearman rank correlations. For a), b) and c), $\chi^2$ value given is for Fisher’s method for combining probabilities for individual regions, with associations in each region negative. For d), the correlation coefficient (r) and associated p-value are given. P-values in bold are significant.

<table>
<thead>
<tr>
<th>Association with timing of deglaciation</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a) ENM regions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_r$</td>
<td>34</td>
<td>16</td>
<td>0.005</td>
</tr>
<tr>
<td>$H_e$</td>
<td>36.32</td>
<td>16</td>
<td>0.003</td>
</tr>
<tr>
<td><strong>b) NJ Tree</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_r$</td>
<td>64.42</td>
<td>32</td>
<td>0.0006</td>
</tr>
<tr>
<td>$H_e$</td>
<td>48.36</td>
<td>32</td>
<td>0.032</td>
</tr>
<tr>
<td><strong>c) Clustering analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_r$</td>
<td>37.14</td>
<td>34</td>
<td>0.326</td>
</tr>
<tr>
<td>$H_e$</td>
<td>26.79</td>
<td>34</td>
<td>0.806</td>
</tr>
<tr>
<td><strong>d) Across study area</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>-0.351</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>p</td>
<td>-0.364</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
**FIGURE LEGEND**

**Figure 1** Location of rivers sampled for Atlantic salmon in Scotland and northern England. Numbers refer to river names in Table S1.

**Figure 2** General nature of post-LGM landscape change in the study region and adjacent areas from c. 25 ka BP to 10 ka BP, by which time all glaciers were gone and the modern geography of the region was first seen. The top left pane shows the Atlantic Ocean (dark grey), a freshwater ice lake and rivers draining the lake, with the ice sheet in light grey. Dashed line represents the land edge; numbers in squares relate to the potential refugial origin of the migrants; arrows the direction of migration and ellipses the resulting regions defined by the ecological niche model (ENM).

**Figure 3** Linearized neighbour-joining Tree, based on $D_A$. Numbers correspond to sites (Table S1) and colours to subtrees forming the groups. Regionally coherent groups of neighbouring rivers are named by the brackets. Numbers on the nodes relate to bootstrap values and only those > 50% are shown.

**Figure 4** Output from the hierarchical STRUCTURE analysis. The coloured brackets and lines above and below each plot show the hierarchical sub-structuring into the various levels, with the number of likely clusters (K) for each level and sub-cluster given above each plot. Within each plot, horizontal bars represent sites and numbers underneath each plot the final cluster number (1-26).

**Figure 5** Map showing the correspondence between the geographical ENM (numbered 1-9 and shown by black line) and either NJ Tree (left) or hierarchical STRUCTURE (right) regions (coloured circles). Coloured circles correspond to the colours in Figures 3 and 4 for the NT Tree and hierarchical STRUCTURE analysis, respectively, with the exception from
the yellows in Figure 4, which have been replaced by black and grey circles. The white circles on the left are sites from sub-trees that did not show geographical coherence.

**Figure 6** Boxplots of allelic richness, $A_r$ (left) and expected heterozygosity, $H_e$ (right) in relation to the timing of deglaciation (ka BP).
Figure 1
Figure 2
Figure 3
Figure 5

Figure 6