Range-wide regional assignment of Atlantic salmon (*Salmo salar*) using genome wide single-nucleotide polymorphisms

Nicholas W. Jeffery[a,*], Brendan F. Wringe[a], Meghan C. McBride[b], Lorraine C. Hamilton[b], Ryan R.E. Stanley[c], Louis Bernatchez[d], Matthew Kent[e], Marie Clément[f,g], John Gilbey[h], Timothy F. Sheehan[i], Paul Bentzen[j], Ian R. Bradbury[a,j]

[a] *Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's, NL, A1C 5X1, Canada*
[b] *Aquatic Biotechnology Lab, Bedford Institute of Oceanography, Fisheries and Oceans Canada, Dartmouth, NS, B2Y 4A2, Canada*
[c] *Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS, B2Y 4A2, Canada*
[d] *Department of Biology, Université Laval, Québec, QC, G1V 0A6, Canada*
[e] *Centre for Integrative Genetics, Norwegian University of Life Sciences, Ås, Norway*
[f] *Centre for Fisheries Ecosystems Research, Fisheries and Marine Institute, Memorial University of Newfoundland, P.O. Box 4920, St. John's, NL, A1C 5R3, Canada*
[g] *Labrador Institute, Memorial University of Newfoundland, Happy Valley-Goose Bay, NL, A0P 1E0, Canada*
[h] *Marine Scotland, Freshwater Fisheries Laboratory, Faskally, Pitlochry, PH16 5LB, Scotland, UK*
[i] *NOAA Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA, 02543, USA*
[j] *Department of Biology, Dalhousie University, Halifax, NS, B3H 4R2, Canada*

ABSTRACT

The estimation of stock specific exploitation is imperative to fisheries management and the conservation of biodiversity, particularly in instances where fisheries simultaneously exploit mixtures of stocks. Mixed stock harvests are particularly common in species that have extensive marine migrations, such as Atlantic and Pacific salmon. Here we develop a range-wide genetic baseline for Atlantic salmon (*Salmo salar*) from North American and European rivers to allow regional assignment of individuals targeted in international mixed stock fisheries. A combination of published data and additional genotyping was used to assemble a dataset of 96 SNPs for 285 range-wide Atlantic salmon populations for regional assignment. Clustering of baseline samples identified 20 North American and eight European reporting groups with mean individual assignment accuracy of 90% (range 70–100%). This baseline was applied to disentangle the stock composition of individuals in a subset of individuals from the West Greenland Atlantic salmon fishery. Genetic mixture analysis revealed that both European and North American individuals originated from multiple regions, with 92% of European individuals originating from the United Kingdom and Ireland, and North American individuals originating from three regions; Gulf of St. Lawrence (28%), Gaspé Peninsula (23%), and coastal Labrador (21%). The baseline represents a significant resource for the management of Atlantic salmon fisheries and the quantification of salmon migration patterns at sea.

1. Introduction

Mixed stock fisheries target a mixture of individuals that originate from several independent stocks thereby adding a layer of complexity to the management strategies of highly migratory marine and anadromous fish species (Begg et al., 1999; Hilborn, 1985; Seeb and Crane, 1999). As the conservation of stock diversity of exploited species is important to species and fisheries stability and persistence (Hilborn et al., 2003), the quantification of stock specific exploitation is central to successful management of mixed-stock fisheries (Begg et al., 1999;

Carvalho and Hauser, 1994). Traditionally, approaches such as physical tagging (Reddin et al., 2012), morphometrics (Reddin and Friedland, 1999), and otolith chemistry or morphology (Friedland and Reddin, 1994) have been used to explore the composition of mixed-stock fisheries. More recently, genetic stock identification (GSI) has been utilized to identify fishery composition in several species including Atlantic herring (Bekkevold et al., 2011), Atlantic cod (Ruzzante et al., 2000), and various Pacific and Atlantic salmonids (Ackerman et al., 2011; Beacham et al., 2004; Bradbury et al., 2015b; Bradbury et al., 2016a, 2016b; Gauthier-Ouellet et al., 2009; Gilbey et al., 2017). Genetic stock
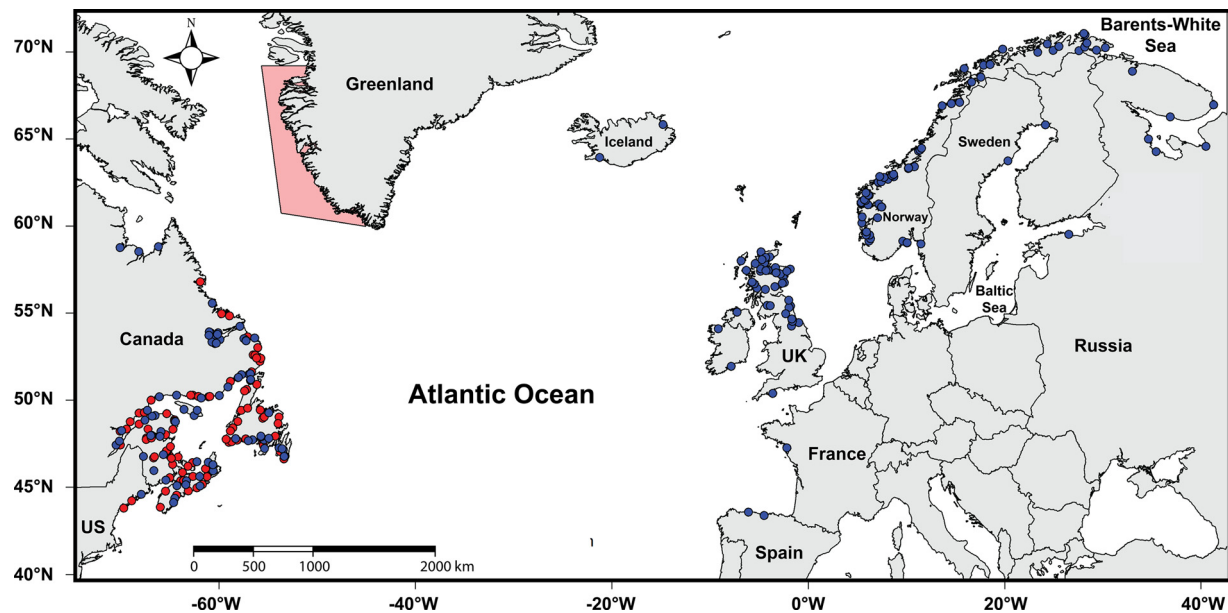
Fig. 1. Range-wide map of 285 rivers across North America and Europe which were refined into regional reporting groups for individual assignment of Atlantic salmon (*Salmo salar*). These rivers are in Canada, the USA, Spain, France, the United Kingdom and Ireland, Norway, Iceland, and the Barents-White and Baltic Seas. Blue circles represent previously published data, and red circles represent additional North American rivers genotyped at the top 96 SNPs based on high F_{ST} and low linkage disequilibrium. The red polygon represents the approximate extent of the West Greenland mixed-stock fishery (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

identification has been dramatically facilitated with the development of large genomic datasets from which highly informative panels of genetic markers can be designed. Tens to hundreds of thousands of genome-wide markers can be screened, and targeted panels of informative loci can be designed for application to mixed-stock fishery analysis which may provide unprecedented resolution of stocks.

Atlantic salmon (*Salmo salar*) are an anadromous fish of high socioeconomic value that have been extensively studied across its native range in North America and Europe. Salmon show fine-scale population structure among rivers and regions due to their strong homing behaviour and local adaptation across a wide latitudinal and thermal range (e.g. Bourret et al., 2013b; King et al., 2001). Large-scale fisheries for Atlantic salmon in North America have largely ceased following the drastic decline of wild populations (COSEWIC, 2010), although several fisheries still exist in the Northwest Atlantic (Bradbury et al., 2016a, 2015b; King et al., 2001; Reddin and Friedland, 1999; Sheehan et al., 2010). In particular, the fishery off West Greenland is the largest intercept fishery in the Northwest Atlantic and is comprised of multi sea-winter (MSW) fish from both North American and European populations, the relative proportions of which vary over time. The ability to assign individuals to their region of origin using genetic markers, both across continents (generally 100% success) and within a continent (70–90%), has been previously demonstrated (e.g. Bourret et al., 2013b; Bradbury et al., 2016a, 2015a; Gilbey et al., 2016, 2017). Most of these studies are regional in scope, and have relied on microsatellite loci, which require extensive standardization across laboratories for large-scale analyses (Ellis et al., 2011; Moran et al., 2006). Recently, Moore et al. (2014) demonstrated an improved regional and population-level assignment with SNPs for Canadian rivers, especially when using > 3000 SNPs, therefore supporting the hypothesis that large scale assignment is possible using an extensive and highly targeted SNP database.

The objective of our study was to develop a highly informative and cost-effective panel of SNPs for accurate genetic assignment and analysis of mixed stocks across the native range of Atlantic salmon. We demonstrate the utility of this panel using a cross range reference baseline to identify regional contributions to a subset sample of the mixed stock fishery from West Greenland. We build upon previously

developed Atlantic salmon baselines based on microsatellites (Bradbury et al., 2015b; Moore et al., 2014) and SNPs (Bourret et al., 2013b; Moore et al., 2014) to extend our North American coverage and further refine our fine-scale regional clustering of rivers and compare the assignment power of our SNP baseline to a microsatellite baseline from Bradbury et al. (2015b). The development of a small panel (< 100) of SNPs will allow for range-wide, rapid regional assignment of salmon to inform fisheries management and aid in the conservation of at-risk populations in both North America and Europe.

2. Methods

2.1. Outline of methods

Panel development and application was divided into the following eight steps: 1) We used genotype data from Moore et al. (2014), Bradbury et al. (2015b) and Sylvester et al. (2018) for SNP panel identification. 2) Samples from preliminary reporting groups based on those from (Bradbury et al., 2015b) were split into training and hold-out groups for SNP panel selection and testing, each consisting of 50% of the individuals. 3) Using the training set of individuals, we tested six methods for panel design of 288 SNPs to further refine into a final panel of 96 SNPs. While we aimed to produce a panel of 96 SNPs for mixture analyses, 288 were initially chosen in case a second panel of 96 was required for desired levels of accuracy, and to provide redundant SNPs for assay design and testing since this would not be possible for each of the top 96 chosen. 4) Additional genotype data were added to our baseline, including individuals from populations from Bourret et al. (2013b), Mäkinen et al. (2015), Barson et al. (2015), and Gilbey et al. (2016). 5) Fish from 100 additional North American rivers were genotyped using our panel of 96 SNPs for assignment and incorporated into our baseline. 6) Final reporting groups were evaluated using clustering analyses and geographic proximity of 285 rivers from North America and Europe (Fig. 1). 7) Baseline power was assessed using self-assignment and a leave-one-out approach, as well as simulations in the R package *rubias* (Anderson et al., 2008; Hasselman et al., 2015). This resulted in a final baseline of 28 reporting groups. 8) 280 individuals from a mixed-stock fishery in West Greenland were genotyped using