Work Package 1

Deliverable - D 1.3

Web-based Trans-European genotype database for Atlantic salmon GSI

(Month 24)

Sub-task 1.2.1 Develop DB structure

The database strategy will be translated into an electronic database format implemented using standard database software, to be identified in the strategy developed in subtask 1.1.1. Entry of data into the base, and data access and review, will be web-enabled to allow remote, rapid and universal access by consortium members. A consortium agreement will be developed and agreed with consortium partners to ensure IPR but maximize availability of individual existing data sets to other researchers and to the science community generally.

Participants: Lead – 1; other

Operational Web-based Database

For the term of the SALSEA-Merge project, a web-based database was developed by IMR (Participant 1) in collaboration with MS (Participant 3), and hosted by IMR, with operational access restricted to these two partners. The data base was populated with information by data sets being sent by the contribution partners to the database manager (Participant 3) for data conversion according to the agreed inter-laboratory calibration and nomenclature standardization protocol. The data was quality control checked and loaded on the database by the web portal

http://www2.imr.no/salsea/admin/

Public Access Web-based Database

For the end of the SALSEA-Merge project, an operational web-based meta-database has been set up by IMR (Participant 1) in collaboration with MS (Participant 3), and hosted by IMR. The data base is populated with meta-information on the integrated quality controlled and standardised SALSEA-Merge baseline, including rivers covered, sample sizes. The database will be web accessible for interrogation and provide information on contacting the database manager (Participant 3) for access to the actual information. Based on the agreed IPR arrangement, the use of the data by the public at large will be restricted to purposes of assignment of fish to region/river of origin. Use of the data set beyond this, will not be allowed for a period anticipated to be for two years, until the additional uses of the data base for regional phylogeographic analysis have been completed by the database contributors. A paper describing the database has been drafted and submitted for publication and the meta-data on the database can be accessed at

http://appl2.imr.no/salsea/devel/

The SALSEA-Merge GRAASP database – a resource for the assignment of natal origin Atlantic salmon from Eastern Atlantic rivers

John Gilbey*, Philip McGinnity†, E. Verspoor*, Anja Armstrong*, Torvieg Balstad‡, Eef Cauwelier*, Corrine Cherbonnel¶, Jamie Coughlan†, Tom F. Cross†, Walter Crozier**, Eillane Dillane†, Jonathan S. Ellis ††, Denis Ensing, Eva Garciá-Vázquez‡‡, Andrew M. Griffiths ††, Sigurdur Gudjonsson†††, Kjetil Hindar‡, David Knox*, Gonzalo Machado-Schiaffino‡‡, Dorte Meldrup¶¶, Einar Eg Nielsen¶¶, Kristinn Olafsson***, Craig R. Primmer‡‡‡, Paulo Prodohl¶¶¶I, Lee Stradmeyer*, Jamie R. Stevens††, Juha-Pekka Vähä‡‡‡, and Vidar Wennevik*****

*Marine Scotland, Freshwater Laboratory,Faskally, Pitlochry PH16 4LB, UK, CDepartment of Zoology, Ecology & Plant Science/Aquaculture& Fisheries Development Centre, University College,

Cork, Ireland, ‡Norwegian Institute for Nature Research (NINA),7485 Trondheim, Norway, ¶GENINDEXE, 6 rue des Sports, 17000 La Rochelle, France,**Agri-Food and Biosciences Institute Northern Ireland,Fisheries and Aquatic Ecosystems Branch, Newforge Lane,Belfast BT9 5PX, UK, †† School of Biosciences, Hatherly Laboratories,University of Exeter, Prince of Wales Rd,Exeter EX4 4PS, UK, ‡‡ Departament of Functional Biology, Genetics, Universidad de Oviedo, C/Julian Claveria s/n, 33006 Oviedo, Spain,,¶¶ DTU Aqua, National Institute of Aquatic Resources,Technical University of Denmark, Vejlsøve 39,8600 Silkeborg, Denmark, *** Matis ohf., Vinlandsleid 12, 113 Reykjavik, ††† Iceland Icelandic Fisheries Laboratories. Skulagata 4, 101 Reykjavik, Iceland, ‡‡‡Department of Biology, University of Turku,20014 Turku, Finland, ¶¶¶I School of Biological Sciences,Queens University, Belfast BT9 7BL, UK, *****I nstitute of Marine Research, Nordnesgaten 50,5005 Bergen, Norway,

Correspondence: John Gilbey; Marine Scotland Science, Freshwater Laboratory, Faskally, Pitlochry, Scotland, U.K. PH16 5LB; Fax: 44 (0)1796 473523; E-mail: gilbeyj@marlab.ac.uk

Running title: SALSEA-Merge Atlantic salmon database

Abstract

We present a genetic database of Atlantic salmon (*Salmo salar* L.) for sample sites from rivers throughout the species' Eastern Atlantic range assembled for the EU SALSEA-Merge project to assign marine fish to natal region or river of origin using GRAASP (Genetically-based Regional Assignment of Atlantic Salmon Protocol). The database contains genotypes for 26,822 salmon scored at 14 microsatellites from 467 sample sites representing 285 rivers in 12 countries encompassing standardised and calibrated data from 12 laboratories. The rivers include those responsible for the of 85% of salmon production (based on rod catch statistics) in the Eastern Atlantic. It provides a unique baseline resource to be exploited for genetic assignment techniques of migrating salmon caught in the marine environment.

Introduction

Recent years have seen significant decreases in abundance of many commercially exploited fish species (Myers & Worm 2003) with these signals being especially apparent in diadromous fish (Jonsson *et al.* 1999). Over the last few decades stocks of diadromous Atlantic salmon in both the Eastern and Western Atlantic have shown sometimes drastic declines (ICES 2010) resulting in fears for their long-term persistence in many rivers (Jonsson *et al.* 1999; Potter & Crozier 2000; WWF 2001). Experimental investigations and monitoring programs suggest that increased ocean mortality is a major factor driving the observed fall in adult spawner numbers and associated productivity (Friedland 1998; Klemetsen *et al.* 2003; Potter & Crozier 2000). Marine mortality has increased from ~70% in the 1970's to 90% and above in recent years (ICES 2010; MacKenzie *et al.* 2011).

Differences in the migratory patterns of stocks from different parts of the range in the marine environment are known to occur but the full extent of differences among stocks remains to be resolved (Webb et al., 2007). The study of stock specific migration and feeding patterns, and its implications for marine mortality rates, are hampered by a lack of information relating to the marine phase of the salmon's lifecycle (Crozier *et al.* 2004); this also makes it difficult to u investigate and regulate anthropomorphic influences such as mixed stock fisheries on endangered stock components. Effective management in the marine ecosystem is reliant on accurate identification of individual population identities (MacKenzie *et al.* 2011).

Physical tagging techniques widely used in terrestrial or freshwater contexts, can address the problem only in part in respect of marine fish such as salmon. They have provided unambiguously accurate information over many years for certain stocks (Hansen & Jacobsen 2003; ICES 2008, 2010). However, only a small proportion of river stocks have been studies for relatively short periods, and the high mortality and low recapture rates of salmon mean that very large numbers of fish need to be tagged to get statistically meaningful results. Further, recaptures are biased towards areas where fish may be caught and reported as part of commercial by-catches (MacKenzie *et al.* 2011). Other techniques such as stable isotope analysis (MacKenzie *et al.* 2011), otolith morphology and microchemistry (Campana 2005; Friedland & Reddin 1994), and parasite tracking (Williams *et al.* 1992) have also been used to identify stock origins but with various but limited levels of success.

The development of highly polymorphic genetic markers in recent years has enabled application of genetic stock identification (GSI) techniques (e.g. Pella & Masuda 2001). These have been extensively used with both Pacific and Atlantic salmon species (see Griffiths *et al.* 2010; Shaklee *et al.* 1999 and references within these) to assign fish back to stock of origin at a variety of geographic scales, from within catchments to between continents (Griffiths *et al.* 2010). The use of GSI means that potentially, all fish can be thought of as genetically 'tagged' and so all might be useful for determining stock specific patterns of marine utilisation or missed stock fishery exploitation levels.

The ability to determine the origin of salmon using GSI depends on the availability of a high quality, representative baseline to which to assign. Atlantic salmon from all countries in the Eastern Atlantic part of their range potentially utilise similar parts the marine environment depending on life-stage (Dadswell *et al.* 2010; MacKenzie *et al.* 2011) and so the most useful baseline will extend across all potential sites of origin. Presented here is a baseline that encompasses the full Eastern Atlantic part of the salmon range, and is available for use in assigning Eastern North Atlantic fish to origin. The Western North Atlantic is excluded but, while fish from the two sides are known to mix (e.g. in West Greenland: refs needed), the two groups are highly divergent genetically (King et al. 2007) and can be easily distinguished using a small number of diagnostic markers (Gilbey et al., 2005?). Thus where the latter are present this can be e asily ascertained.

Database description

The database contains genetic information for 14 microsatellite loci from rivers in 12 countries across Europe (Denmark, England, Scotland, France, Iceland, Ireland, Northern Ireland, Norway, Russia, Spain, Sweden and Wales) produced by separate laboratories 12 laboratories (Table 1). The microsatellites are 14 of the 15 described by Olafsson (Olafsson et al. 2010) and identified at meeting in Leetown, Virginia of researchers working of Atlantic salmon microsatellites hosted a United States Geological Service. These were identified as the optimal suite of microsatellites for new work, from those available, that best facilitated multiplexing, minimised mis-typing and null allele problems, and best built on existing population work. A single loci (*SsaD486*, King et al. 2005) was removed from the dataset due to its lack of variation over much of the range, resulting in a baseline consisting of 14 microsatellite loci: *SsaF43* (Sanchez *et al.* 1996), *Ssa14*, *Ssa289* (McConnell *et al.* 1995), *Ssa171*, *Ssa197*, *Ssa202* (O'Reilly *et al.* 1996) *SSsp1605*, *SSsp2201*, *SSsp2210*, *SSsp2216*, *SsspG7* (Paterson *et al.* 2004), *SsaD144*, *SsaD486*, *SsaD157* (King *et al.* 2005) and *SSsp3016* (unpublished, GenBank number AY37820). Intra-laboratory calibration and standardisation was carried out as detailed in Ellis et al. (2011) as part of the EU SALSEA-Merge Project.

The database contains data relating to 26,822 salmon from 467 sample sites representing 285 rivers in the 12 countries (Table 2). Together with the 14 microsatellite loci genetic data the database contains information on each fish which describes the fish's country, river and site of origin. Each river in the dataset also has a latitude and longitude associated with its mouth to allow geo-location of rivers/sites. The geographical coverage of the baseline stretches from North Norway in the North to Spain in the South, and from Iceland in the West to Russia in the East (Fig. 1).

The information represents samples collected over the period of 199? To 2010. In general the samples are from juvenile fish, mostly parr, but in some case are from smolts and adults. Sample sizes range from X to Y with a mean sample size of Z. Rivers are characterized by from x to y samples with a modal number of samples per river of Z.

Discussion

The 467 rivers included in the data base are only XX% of all salmon rivers in Western Europe. However, the database encompasses ~85% of the Atlantic salmon production in the Eastern Atlantic based on reported rod catches and has been exploited as part of the EU SALSEA-Merge project for the assignment of marine caught post-smolt salmon to phylogeographic region of origin (Verspoor et al., submitted; Gilbey et al., submitted). Based on blind samples from both data base and non-data base rivers, the data base allows assignment to defined phylogeographic regions with >XX% success (Gilbey et al., submitted).

An exploration of the stability of the data base with regard to different spatial and temporal sampling is underway (Nielsen et al., in prep) and the capacity of the data base for river assignment remains to be assessed; the issue of temporal stability of river specific genetic differences will be particularly important in this regard. However, the dataset already provides a valuable resource for assignment of regional origin of Atlantic salmon from Western Europe, in particular for marine specimens to add to a growing data base on the North Atlantic distribution of different regional stock groups.

The large collaborative project which resulted in this dataset, is unique to Atlantic salmon and to the European fisheries context. It represents a valuable resource assembled by a rigorous calibration, standardisation and quality control procedures utilised has resulted in a unique which is freely available resource to facilitate identification of the natal origin of European origin salmon. It also provides a baseline for the analysis of temporal changes in the genetic character of salmon stocks in Europe. Integration of this dataset with information from the Western Atlantic would provide significant enhancements for examination of areas where he two continental stocks mix such as the seas around West Greenland. For those wishing to exploit the database, it is available on request from the corresponding author in Microsoft Access, Microsoft Excel or CSV formats, or as a GENEPOP file (Raymond & Rousset 1995).

Analysis of Pacific Salmonids (GAPS) consortium for very helpful advice when planning was underway for this project and database.

References

Campana SE (2005) Otolith elemental composition as a natural marker of fish stocks. In: *Stock Identification Methods: Applications in Fishery Science.* (eds. Cadrin SX, Friedland KD, Waldman JR), pp. 227-245. Elsevier Academic Press, Oxford.

Crozier WW, Schön P-J, Chaput G*, et al.* (2004) Managing Atlantic salmon (Salmo salar L.) in the mixed stock environment: challenges and considerations. *ICES Journal of Marine Science: Journal du Conseil* **61**, 1344-1358.

Dadswell MJ, Spares AD, Reader JM, Stokesbury MJW (2010) The North Atlantic subpolar gyre and the marine migration of Atlantic salmon Salmo salar : the 'Merry-Go-Round' hypothesis. *Journal of Fish Biology* **77**, 435-467.

Ellis JS, Gilbey J, Armstrong A*, et al.* (2011) Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (Salmo salar L.). *Genetica* **139**, 353-367.

Friedland KD (1998) Ocean climate influences on critical Atlantic salmon (Salmo salar) life history events. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 119-130.

Friedland KD, Reddin D (1994) Use of Otolith morphology in stock discriminations of Atlantic salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 91-98.

Gilbey J, + others Marine distribution of regional Atlantic salmon post-smolt stocks in the NE Atlantic ascertained by microsatellite DNA based assignment. ICES Journal of Marine Research (submitted)

Griffiths AM, Machado-Schiaffino G, Dillane E*, et al.* (2010) Genetic stock identification of Atlantic salmon (Salmo salar) populations in the southern part of the European range. *BMC Genetics* **11:31**.

Hansen LP, Jacobsen JA (2003) Origin and migration of wild and escaped farmed Atlantic salmon, Salmo salar L., in oceanic areas north of the Faroe Islands. *ICES Journal of Marine Science: Journal du Conseil* **60**, 110-119.

ICES (2008) Report of the workshop on salmon historical information - new investigations from old tagging data (WKSHINI). ICES, Halifax, Canada.

ICES (2010) Report of the Working Group on North Atlantic Salmon, p. 302, 22-31 March 2010 Copenhagen, Denmark.

Jonsson B, Waples RS, Friedland KD (1999) Extinction considerations for diadromous fishes. *ICES Journal of Marine Science* **56**, 405-409.

King T, Eackles M, 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.

Klemetsen A, Amundsen PA, Dempson JB*, et al.* (2003) Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.): a review of aspects of their life histories. *Ecology of Freshwater Fish* **12**, 1-59.

MacKenzie KM, Palmer MR, Moore A*, et al.* (2011) Locations of marine animals revealed by carbon isotopes. *Sci. Rep.* **1**.

McConnell SK, Hamilton L, Morris D*, et al.* (1995) Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian stocks of Atlantic salmon. *Aquaculture* **137**, 19-30.

Myers RA, Worm B (2003) Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280-283.

O'Reilly PT, Hamilton LC, McConnell SK, Wright JM (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 GO, 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.

Paterson S, Piertney SB, Knox D, Gilbey J, Verspoor E (2004) Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (Salmo salar L.) microsatellites. *Molecular Ecology Notes* **4**, 160-162.

Pella J, Masuda M (2001) Bayesian method for analysis of stock mixtures from genetic characters. *Fisheries Bulletin* **99**, 151-167.

Potter ECE, Crozier WW (2000) A perspective on the marine survival of Atlantic salmon. In: *The ocean life of Atlantic salmon: environmental and biological factors influencing survival.* (ed. Mills D), pp. 19-36. Fishing News Books, Blackwell Science, Oxford.

Raymond M, Rousset R (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Heredity* **86**, 248-249.

Sanchez JA, Clabby C, Ramos D*, et al.* (1996) Protein and microsatellite single locus variability in Salmo salar L. (Atlantic salmon). *Heredity* **77**, 423-432.

Shaklee JB, Beacham TD, Seeb L, White BA (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research* **43**, 45-78.#

Verspoor E, McGinnity P, Gilbey J + others GRAASP: a genetic tool for advancing understanding of the marine ecology of Atlantic salmon in the Eastern Atlantic Ocean. ICES Journal of Marine Research (submitted)

Williams HH, MacKenzie K, McCarthy AM (1992) Parasites as biological indicators of the population biology, migrations, diet, and phylogenetics of fish. *Reviews in Fish Biology and Fisheries* **2**, 144-176.

WWF (2001) The Status of Wild Atlantic Salmon: A river by river assessment. WWF at

http://www.panda.org/news_facts/publications/general/index.cfm

Table 1. Institutions contributing genetic data.

Institute	Country
Danish Institute for Fisheries Research (DIFRES)	Denmark
University of Exeter (UE)	England, UK
University of Turku (UT)	Finland
Geneindex (GENI)	France
Institute of Freshwater Fisheries (IFL)	Iceland
National University of Ireland, Cork (NUIC)	Ireland
Queen's University Belfast (QUB)	Northern Ireland, UK
Institute of Marine Research (IMR)	Norway
Norwegian Institute for Nature Research (NINA)	Norway
Marine Scotland Science (MSS)	Scotland, UK
University of Oviedo (UO)	Spain
University of Wales, Swansea (UWS)	Wales, UK

Table 2. Origin and numbers of fish with genetic data represented in the database.

Country	Rivers	Sample sites	Fish
Denmark	3	4	189
England	22	33	1409
France	9	9	450
Iceland	16	22	1986
Ireland	29	40	2053
Northern Ireland	7	18	1302
Norway	84	104	7401
Russia	30	33	2350
Scotland	71	187	8973
Spain	4	4	190
Sweden	4	4	172
Wales	6	9	347
Total	285	467	26822

Fig. 1 Map of rivers which have genetic data represented in the database. Points represent river mouths (multiple sample sites may lay within a river).