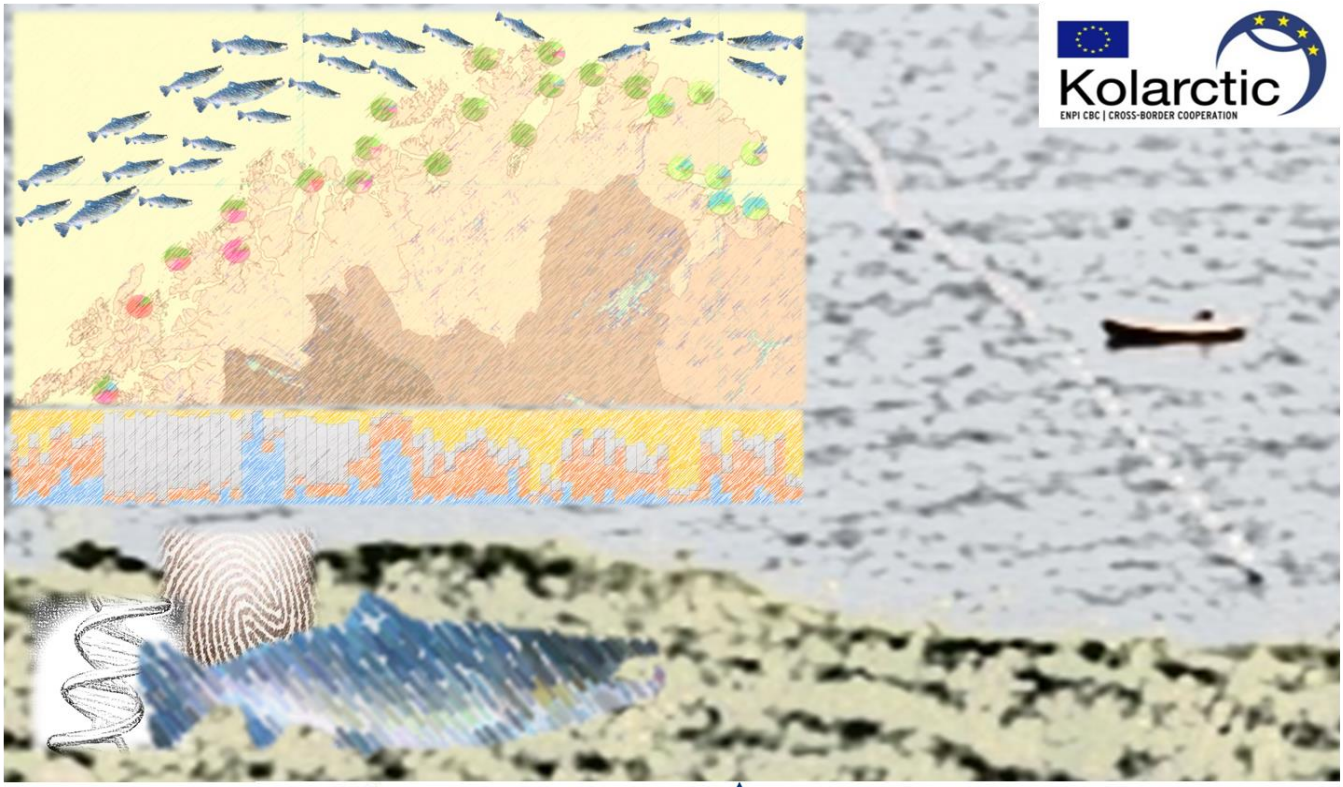
Kolarctic ENPI CBC – Kolarctic salmon project (KO197)

Genetic structure of Atlantic salmon in the Barents region and genetic stock identification of coastal fishery catches from Norway and Russia

Juha-Pekka Vähä[1, 2], Vidar Wennevik[3], Mikhail Ozerov[1, 2], Rogelio Diaz Fernandez[1], Laila Unneland[3], Kristiina Haapanen[1], Eero Niemelä [4], Martin S. Svenning [5], Morten Falkegård[5], Sergey Prusov[6], Ivan Lyzhov[6], Kira Rysakova[6], Tiia Kalske[7], Bente Christiansen[7], and Gennadiy Ustyuzhinsky8

[1]Kevo Subarctic Research Institute, University of Turku, Finland
[2]Department of Biology, University of Turku, Finland
[3]Institute of Marine Research (IMR), Tromsø, Norway
[4]Finnish Game and Fisheries Research Institute, Teno River Research Station Utsjoki, Finland
[5]Norwegian Institute of Nature Research (NINA), Tromsø, Norway
[6]Knipovitch Polar Research Institute of Marine Fisheries and Oceanography (PINRO), Murmansk, Russia
[7]County Governor of Finnmark (FMFI), Vadsø, Norway
[8]Knipovitch Polar Research Institute of Marine Fisheries and Oceanography (PINRO), Archangelsk, Russia

Corresponding author: Juha-Pekka Vähä, Kevo Subarctic Research Institute, 20014 University of Turku, Finland, fax +358 2 333 5960, tel. +358 2 333 6664, e-mail: juha-pekka.vaha@utu.fi



Kolarctic ENPI CBC | CROSS-BORDER COOPERATION

Turun yliopisto University of Turku

Kevo Subarctic Research Institution

HAVFORSKNINGSINSTITUTTET INSTITUTE OF MARINE RESEARCH

GAME AND FISHERIES RESEARCH

NINA

Fylkesmannen i Finnmark Finnmárkku Fylkkamánni

Abstract

The Kolarctic Atlantic Salmon project has generated one of the most comprehensive and detailed genetic datasets for any fish species. More than 13 000 individuals from over 200 samples collected from over 180 rivers in the Kolarctic area have been analysed for 31 DNA markers displaying well over 600 alleles. This dataset not only gives the first complete overview of genetic structure of salmon from northern Europe, but also serves as a genetic baseline data for identifying stock of origin for individuals caught in mixed stock fisheries. Hence, stock of origin was estimated for more than 23 000 Atlantic salmon sampled from coastal fishery catches of Northern Norway and White Sea. Results of genetic stock identification provide first and comprehensive overview to spatial and temporal variation in stock compositions in coastal fisheries of Northern Norway.

In genetic structure analyses, major genetic divisions were found at different geographical scales; the main genetic barrier appearing between the eastern populations of Russia, including the White Sea populations, and populations from northern Kola and northern Norway. Genetic barriers/shifts are also observed at finer geographic scales. Genetic differences between populations, overall and within a region, are greatest for the eastern populations of Russia. Genetic structuring within major river systems is observed in the Pechora, Ponoi and Teno rivers. In these river systems multiple populations exist and they should be managed as separate units. The genetic baseline developed for this project allows for precise identification of salmon caught at sea to individual rivers/reporting groups, providing opportunities for more adaptive and informed management of coastal salmon fisheries.

Power tests of genetic stock identification using test samples from the baseline data revealed large differences among rivers and regions in the expected level of stock identification. On average, 69% of samples assigned to a river were correct, but more than 70 stocks were distinguished and identified with high (>80%) assignment success to their river of origin. Highest correct assignment was observed for rivers in the Eastern Barents and White Sea and Teno River system salmon stocks (90%), while the lowest was observed for the Troms and Nordland stocks (54%).

Nine reporting groups, roughly following genetic boundaries, were delineated for identifying the geographical region of origin of salmon from coastal catches. Individuals from Russian rivers and Teno River system were correctly assigned to their respective four reporting groups with 94-99% accuracy, while slightly lower assignment success was obtained for the samples from rivers in eastern and western Finnmark; 86%. When northern Troms and southern Troms reporting groups were combined, 80% of Troms salmon were correctly identified while salmon from rivers in Nordland had correct assignment of 72%. In some cases, correct stock assignment to rivers and reporting

groups could be increased substantially but with cost of having to discard identified samples.

Genetic stock identification analyses confirmed that coastal fisheries exploit multiple stocks. Altogether, 145 rivers were found to contribute to fishery samples. Fisheries generally exploited salmon from wide geographical areas with catch localities on the open coast showing greater stock diversity than catch localities within fjords. Fishery samples from May and June were composed of salmon from wider geographical areas, whereas samples from July and August were composed of more local populations. In the fjords, fisheries target mostly local populations throughout the year. Neither Norwegian or western Kola salmon were encountered within the White Sea basin.

Genetic baseline developed in this project allows for further studies of the marine distribution and exploitation of salmon from the Kolarctic area, such as mapping of migration of post-smolts and adults in the open sea, as well as identification of important genetic biodiversity units for conservation. Assignment accuracy and precision can be further increased by supplementing the baseline population data with more samples. With accumulating baseline data, genetic stock assignments assessed here can be refined, but the current data already now provides valuable information on the stock compositions, harvest rates and migration patterns of salmon of the Barents Sea Region..

This report has been produced with financial support from the European Union, but the contents can in no way be taken to reflect the views of the European Union.

NINA arranged for special permission from Norwegian Directorate of Nature Management for fishing to be conducted outside the ordinary season.

Contents

Introduction

Rivers in northernmost Norway, Finland and Russia support the world's largest wild Atlantic salmon (*Salmo salar*) stocks and resources. Fisheries catches in this area account up to 50% of the total harvest of the wild Atlantic salmon in the world (ICES 2009). The status of salmon stocks in rivers in the Barents region however vary considerably and many suffer from reduced numbers of spawning salmon. Large variation in the population size has negative impacts on the stocks e.g. through decrease of the genetic diversity. Currently the reasons for the annual fluctuations are largely unknown, but potential factors include overfishing, diseases, pollution and changing climate. From the perspective of sustainable commercial and recreational fisheries and the significant cultural heritage, conservation and management of these unique and important stocks is of extreme importance. To achieve this goal, we need a better understanding of the migratory behaviour of salmon, which imposes a great challenge.

Migration is a key element in the life history of Atlantic salmon. The spawning migrations of Atlantic salmon cover thousands of kilometres from the feeding areas at sea to their natal rivers where they reproduce. The migratory phase of the life cycle exposes salmon to diverse risks, including intensive human exploitation. Fisheries in the coastal areas exploit salmon from a large number of rivers and over a wide geographical range. Such fisheries are referred to as mixed-stock fisheries (NASCO 2009) and they have been associated with Atlantic salmon stock declines occurring throughout the North Atlantic area since the 1970s (ICES 2013). However, despite vigorous limits to coastal fisheries and great reductions in the catches during the past decades, the abundance of many populations continues to fall (Parrish et al. 1998; Dempson et al. 2004). The challenge to fisheries management is to develop regulations that balance between protecting stocks at risk while still allowing fisheries access to healthy stocks.

Understanding stock specific migrations and exploitation patterns of Atlantic salmon is essential to management of this important resource. In areas, where exploitation is restricted to fisheries within the river, management decisions have a direct effect on a single stock. However, coastal fisheries are largely mixed stock fisheries and management decisions affect several stocks with varying status. The vital element for management of mixed stock fisheries is the knowledge about the stocks contributing to the catches. The ability to differentiate between stocks allows for estimation of relative stock contributions and in some cases the absolute numbers of salmon being harvested. Such knowledge can then allow for modelling and increased understanding of stock specific migrations – essential elements in developing sustainable management plans.

Genetic stock identification (GSI) has been used in salmon research and management for three decades now (Milner et al 1983) allowing assessment of origin of the stocks being harvested. With the advent of powerful genetic markers, reduced costs of analysing large numbers of samples accompanied with the development of tailored

statistical methods, genetic stock identification is one of the most successful biological tools available for assessing stock compositions in mixed stock fisheries. During the last decade it has become an indispensable and powerful tool to understand fishery dynamics, especially of salmonid fishes (e.g. Oregon Salmon Commission 2008, Beacham et al. 2008, Hess et al. 2011).

There are a number of advantages to modern GSI over other traditional stock identification techniques such as coded wire tags or other external markers. In genetic stock identification; i) conceptually all stocks and individuals sampled are marked internally by specific sequences in their DNA, ii) fish can be sampled non-lethally with minimal handling, iii) fast processing speed provide close to real-time information on stock identity and iv) estimates are based on solid statistical framework. While GSI is not constrained by the number of stocks or samples, there are several prerequisites to successful identification of the stock origin. The feasibility of applying genetic stock identification depends on the relative genetic distinctiveness of stocks and the key to successful stock identification is adequate baseline data capturing the genetic structure and diversity of all the potential stocks in the mixture. Incomplete baseline data or unrepresentative sampling of fish to generate population baseline data deteriorate/degrade the accuracy of stock estimation potentially resulting in biased estimates.

This report describes the GSI methods and the underlying baseline population data used to estimate the stock of origin for the salmon sampled from the mixed stock fisheries in Norway and Russia operated in 2008-2012. Generally the report is divided in four sections where we first i) provide details of genetic markers and describe methods of their analyses, then ii) we describe the baseline population data and the underlying genetic structure of the salmon stocks in the Barents and the White Sea regions. In the third section iii) we describe the statistical methods of GSI and present results from the assessment of accuracy of stock identification. In the final chapter iv) we present results from the application of the GSI on over 23 000 Atlantic salmon samples collected in coastal fisheries of northernmost Norway in 2008, 2009, 2011 and 2012, and Russia in 2010-2012.

Section I – Material and Methods

1.1 Sample collection

1.1.1 Baseline sample collection

More than 14 000 samples for genetic analysis were collected from 201 sampling locations within 185 rivers along the Norwegian and Russian northern coasts between 14°E and 60°E (Fig. 1, Appendix Table 1). In total, Salmon juveniles were sampled using electrofishing, sacrificed by decapitation and tissue sample of each individual was immediately stored in 96% ethanol. The permits for samples collection were issued by: 1) Federal Agency for Fisheries (Russia), 2) County Governor of Finnmark, Troms and Nordland (Norway) and 3) Center for Economic Development, Transport and the Environment (Finland).

Some populations were sampled in different years to enable assessment of temporal stability in the baseline samples. In total, temporal samples were analysed for 5 Russian rivers and 8 Norwegian rivers. For detailed description of the juvenile sampling activities, see Niemelä et al. 2014 for details.

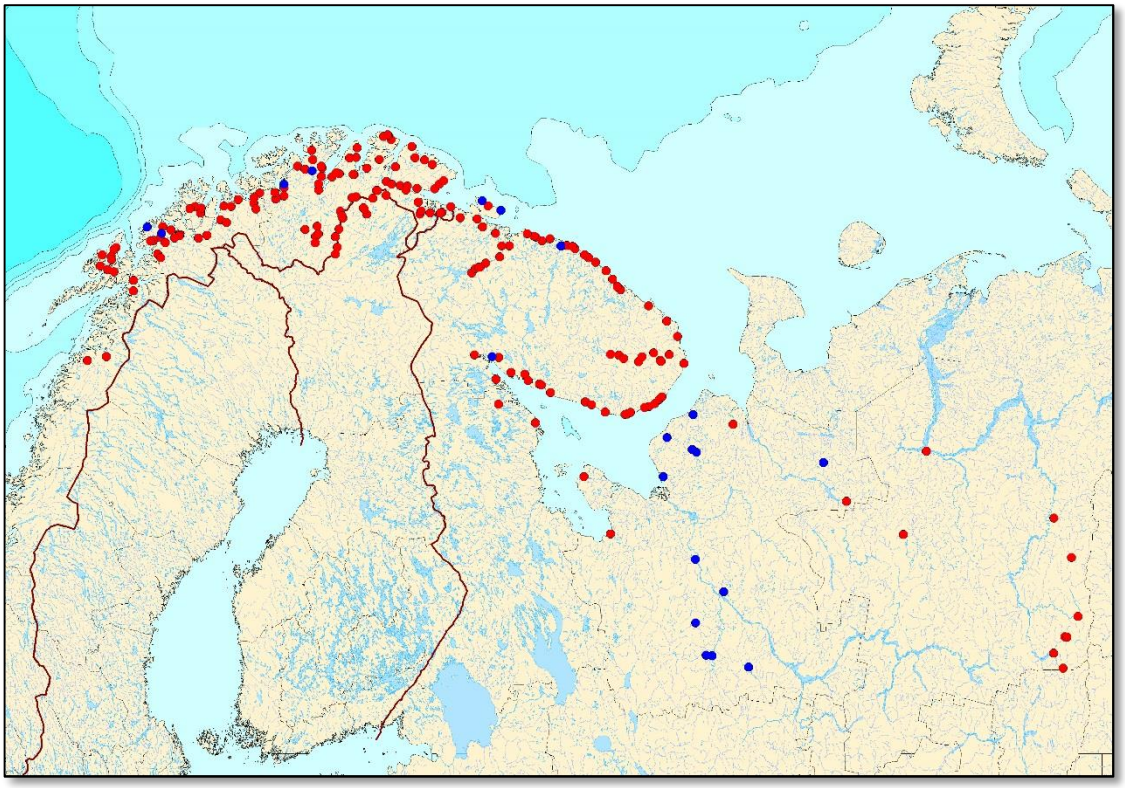


Figure 1. Map showing the baseline river samples. Red dots represent river samples included in the final baseline, blue dots represent samples not included for various reasons (see text).

1.1.2 Coastal sample collection

More than 25000 scale samples were collected in 2008, 2009, 2011 and 2012 from adult individuals captured by local professional fishermen using commercial fishing gear. Commercial coastal fishery in Norway (Finnmark, Troms and Nordland counties) spanned from 13°E to 31°E, including 72 localities in total (see Niemelä et al 2014 for details). In addition, more than 2600 scale samples from 19 locations of coastal and scientific fishery in Russia extending from 31°E to 57°E (Murmansk and Arkhangelsk region, Komi Republic and Nenets Autonomous Okrug) were collected in 2010, 2011 and 2012 (see Prusov et al. 2014 for details). In order to detect the spatial and temporal use of the different Barents salmon populations along the North-Norwegian coastal areas during their spawning migration, the sampling was conducted during the whole summer season from May till September (i.e. before, during and after the ordinary fishing season; see Niemelä et al. 2014 for details).

1.2 Genetic and statistical analyses

Baseline and coastal samples were analyzed for genetic variation at microsatellite DNA markers at the laboratories of the University of Turku, Finland (UTU), and the Molecular Biology laboratory at the Institute of Marine Research, Norway (IMR). Below we have described the laboratory procedures used in the two laboratories, where they differ, and the procedures for calibrating the scoring of alleles between the laboratories.

1.2.1 DNA extraction at University of Turku

Total genomic DNA was extracted from scale or fin tissue specimen by digesting with proteinase K, followed by purification of nucleic acids on silica fines in 96-well filter plates (0.45 μm GHP, Pall Life Sciences) using a protocol slightly modified from Elphinstone et al. (2003).

1.2.2 DNA extraction at IMR

Total genomic DNA was extracted from 15-20 μg fin tissue, or in some cases 15-20 scales from parr, in 96-well format using Qiagen DNeasy blood and tissue kits, following the manufacturer's protocol. Each 96well-plate contained two or more negative control wells. DNA concentration of the extracts was measured for 15 samples on each plate, averaged, and a working dilution for PCR was prepared with a DNA concentration of approximately 15ng/μl.

1.2.3 Genetic markers and amplification procedures at University of Turku

All microsatellite loci were amplified by multiplex PCR (Table 1). Simultaneous amplification of up to nine loci were carried out in 7.8 µl reaction volume, which consisted of 0.8µl of extracted DNA elute, 0.5x KAPA2G™ Fast HotStart ready mix (2x) from KAPABIOSYSTEMS (www.kapabiosystems.com) and varying concentrations of primers (details available upon request). Thermal cycling profiles for the multiplex protocols were as follows: 15 minutes at 95 °C, followed by 36 cycles of 30 seconds at 94 °C, 90 seconds touchdown with increment of -1°C (TD62-55) or -0.5°C (TD58-53) and 60 seconds at 72 °C, followed by final extension step of 10 minutes at 72 °C (Table 1). Varying volumes of the PCR amplified products were pooled (Table 1) and 0.09 µl of GS600LIZ size standard (Applied Biosystems) was added as an internal size standard to each sample. Electrophoresis was then performed on ABI 3130*xl* (Applied Biosystems). Allele scoring was performed with GENEMAPPER V3.7 (Applied Biosystems) followed by manual corrections.

1.2.4 Genetic markers and amplification procedures at IMR

The procedures for amplification of DNA markers at IMR differed from those applied at UTU. Through earlier projects, a baseline containing around 50 of the river populations analysed in this project had already been analysed for 18 of the 31 microsatellite markers, and two multiplexes, containing the extra 13 markers, that could be combined in one run on the sequencing machine were developed at UTU and employed at the IMR laboratory. It was also deemed more efficient to continue using this procedure for analysis of further samples, thus avoiding an extra level of calibration of allele scores at the IMR lab. The loci were combined into a total of 5 multiplexes:

- Multiplex1: SSsp2210, SSspG7, SsaD144, Ssa202 and SSsp2201
- Multiplex2: Ssa289, Ssa14, SSsp1605, Ssa171, SSsp2216
- Multiplex3: SsaF43, Ssa197, SsaD486, MHC1, MHC2 and SsOls85
- Multiplex 4: Ssa405, Ssa412, Ssa98, SsOSL25, SSsp2215, EST107, EST68
- Multiplex 5: EST28, EST19, Ssa407, Ssleer15.1, Sleen82, Sleer53

For multiplexes 1-3 amplifications were carried out in 15 µl volumes, including 3ul DNA, 3 µl buffer, 1.2 µl $MgCl_2$, 2.4 µl dNTPs, 0.5 U *Taq* DNA Polymerase (Promega), and 0.12 to 0.345 µl of forward and reverse primers. Reactions were carried out on a ABI9700 thermocycler, and consisted of an initial denaturation step of 4 min at 94°C, followed by 25-27 cycles of denaturation at 94°C for 50 s, annealing at 55°C for 50 s and extension at 72°C for 80 s. The same PCR protocol was used for multiplexes 1-3.

Table 1. Details of PCR amplification set-ups at UTU, primer sequences and allele numbers of microsatellite loci.

panel	multiplex	PCR program	pooling	marker name	label	primer sequence F	primer sequence R	Reference	A #
1	1	TD62-65	2						
				Ssa412	VIC	GTGGAGATACACAGCACTTA	GTTTCTTGGTTAGTACCGGACATG	Cairney et al. 2000	11
				Ssa171	FAM	ATTATCCAAAGGGGTCAAAA	GTTTGAGGTCGCTGGGGTTTACTAT	O'Reilly et al. 1996	29
				Ssa197	VIC	TGGCAGGGATTTGACATAAC	GGGTTGAGTAGGGAGGCTTG	O'Reilly et al. 1996	31
				Ssa202	NED	TTCATGTGTTAATGTTGCGTG	GTTTCTTGGAATATCTAGAATATGGC	O'Reilly et al. 1996	21
				SSsp3016	PET	GACAGGGCTAAGTCAGGTCA	GATTCTTATATACTCTTATCCCCAT	Paterson et al. 2004	24
				Ssa14	VIC	CCTTTTGACAGATTTAGGATTTC	GTTTCAAACCAAACATACCTAAAGCC	McConnell et al. 1995	7
				Ssa289	VIC	GTCATACAGTCACTATCATC	GTTTCTTTACAAATAGACAGACT	McConnell et al. 1995	12
				EST68	PET	TGACACTGTGGCCTGTCTCT	GTTTGAGTTCTGGGTTATTTATTCACA	Vasemägi et al. 2005	9
	2	TD62-65	1.5						
				EST28	VIC	CACAGGCACACACTCCTCAT	GTTTCAGGTGAAGAGCATGACCAA	Vasemägi et al. 2005	16
				MHCII	PET	GATGGCAAAGAGGAAAGTGAG	GTTTGTTATGCTCTACCTCTGAA	Stet et al. 2002	19
				SSsp2216	FAM	CTCCTCCTGGGATTTCCTGTCA	GTTTCTGGAGCAGAGGATTGCTG	Paterson et al. 2004	26
				SSA405	NED	CTGAGTGGGAATGGACCAGACA	GTTTACTCGGGAGGCCCAGACTTGAT	Cairney et al. 2000	34
				Ssa98	NED	GCAGTCCTTACCTGTGTGATTA	GTTTGGTAGTGATCTGGAGAGTGC	O'Reilly et al. 1996	19
				Ssosl25	NED	ATCTACACAGCTCCTGGTGGCAG	GTTTCATGTAATGGGTCGAGAGAAGTG	Slettan et al. 1995	21
				SSsp2215	FAM	GGTCAGTCAGTCACACCATGC	GTTTAGGTGTCCTGCCGGTCAAT	Paterson et al. 2004	31
				SsaD486	PET	ACTCGGATAACACTCACAGGTC	GTT(T)CGCTGTGTATCAGTATTTTGG	King et al. 2005	6
				SSf43	NED	GAGTCACTCAAAGTGAGGCC	GTTTAGCGGCATAACGTGCTGTGT	Sánchez et al. 1996	13
2	1	TD58-53	2						
				Ssosl85	NED	TGTGGATTTTTGTATTATGTTA	GTTTATACATTTCCTCCTCATTCAG	Slettan et al. 1995	29
				SsaD144	PET	TCAATTGTTGGGTGCACATAG	GTTTGTGAAGGGGCTGACTAAC	King et al. 2005	43
				EST19	VIC	CGCTTCCTGGACAAAAATTA	GTTTCATCTCTGTCATTCTCTTGC	Vasemägi et al. 2005	44
				Sleel53	FAM	TGATTTGTTGCCTGCTGCTTCC	GTTTCCTGCTGCCCACATCATCC	U86704	11
				SSspG7	VIC	CTTGGTCCCGTTCTTACGACAACC	GTTTGCACGCTGCTTGGTCCTTG	Paterson et al. 2004	26
	2	TD58-53,	2.2						
				Sleen82	FAM	CATGGAGAATCCCACTTTCTTA	GTTTCAGGGAGTGATATGGGACATAA	U86706	13
				SSsp2210	NED	CCTTTTTCCAATGGGATTCA	GTTTCATGCACACACATTCACTGC	Paterson et al. 2004	17
				Ssa407	NED	TCGTGACTACTAAGTCTTTGACCA	GTTTGTGTAGGCAGGTGTGGAC	Cairney et al. 2000	43
				EST107	NED	AGCGTTACGTCGAATCCAA	GTTTCTCATGGAGGGTGGAAGTGT	Vasemägi et al. 2005	11
				SSsp2201	FAM	TTAGATGGTGGGATACTGGGAGGC	GTTTCGGGAGCCCCATAACCCTACTAATAAC	Paterson et al. 2004	43
				SsaD157	PET	GCTTAGGGCTGAGAGAGGAATAC	GTTTATCGAAATGGAACTTTTGAATG	King et al. 2005	44
				SSsp1605	VIC	TCTGAGGCTCCTTCTACACTGA	GTTTGGTAGGTGCAAGAAAAAAGGAC	Paterson et al. 2004	13
				Ssleer15.1	FAM	CATGTGCGTGTGCTTTTACAG	GTTTTCTGCATGTAGAACCCTGACC	U86708	8
				MHC I	FAM	GAAGGTGCTGAAGAGGAACGTC	GTTTCAATTACCACAAGCCCGCTC	Grimholt et al. 2002	22

For multiplexes 4-5 the following protocol was used: amplifications were carried out in 7.8 µl reaction volume, which consisted of 0.8 µl of extracted DNA elute, 0.5x KAPA2G™ Fast HotStart ready mix (2x) from KAPABIOSYSTEMS (www.kapabiosystems.com) and varying concentrations of primers (details available upon request). Reactions were carried out on a ABI9700 thermocycler, and consisted of an initial denaturation step of 150 s at 94°C, followed by 9 cycles of denaturation at 95°C for 25 s, annealing at 58°C for 30 s and extension at 72°C for 25 s, then followed by 28 cycles of denaturation at 95°C for 25 s, annealing at 53°C for 30 s and extension at 72°C for 25 s, and finally an extension step of 10 minutes at 72°C.

PCR products were analysed on an ABI 3730XL Genetic Analyser and sized by a 500LIZ™ size standard (Fig. 2). Multiplexes were loaded into the machine in separate runs, except multiplexes 4 and 5 that were physically mixed and combined before fragment analysis. Size estimation and scoring of alleles was conducted in GENEMAPPER 4.0, by two persons evaluating the results independently.

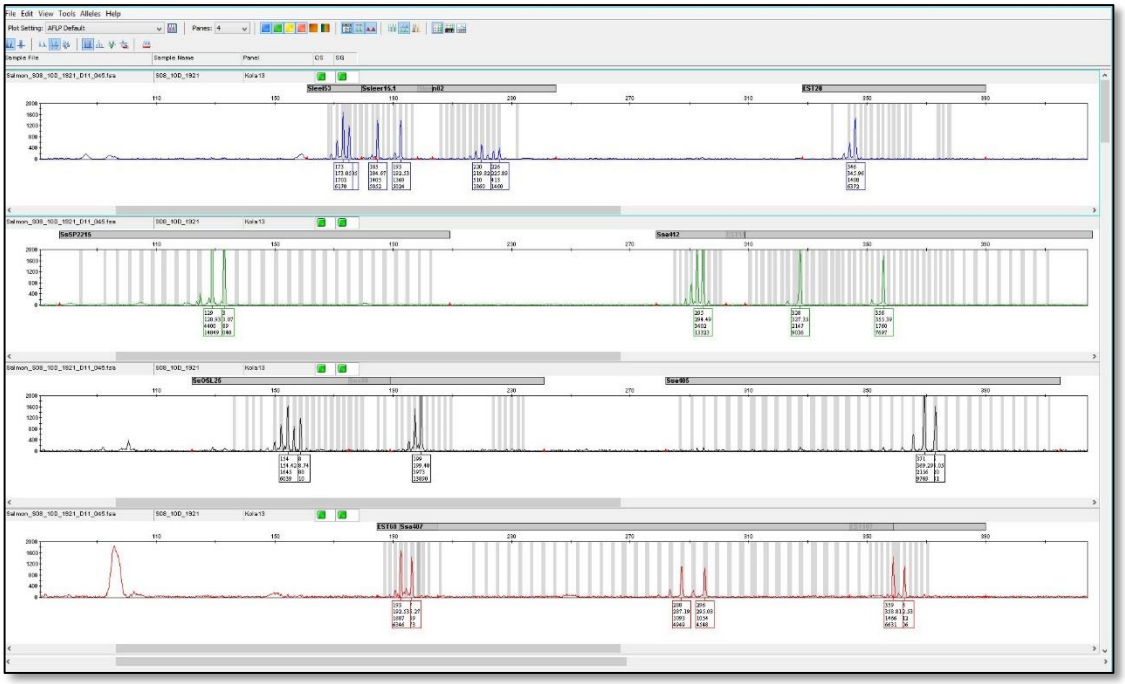


Figure 2. Scoring of microsatellite loci in GeneMapper 4.0.

1.2.5 Genotyping quality at University of Turku

The amplification procedure was only slightly modified from Vähä et al. (2008), where the genotyping error rate for the original procedure was estimated to be low (<0.4 %). All individuals genotyped at the University of Turku were subject to manual checking by two persons independently. This was done to minimize genotyping errors.

Genotyping quality threshold was initially set to having 29 out of 31 loci producing unambiguous data and if failed, sample was re-analysed from either DNA extraction or PCR amplification step forward.

1.2.6 Treatment of contaminated samples

The cases of contamination were classified into two main types: a) particle contamination and b) cross contamination. Particle contamination occurs when DNA fragments from two individuals are amplified in a single reaction most often producing a multiploid instead of normal diploid genotype. Cross-contamination occurs when two individuals (samples) have identical DNA fingerprint, but non-identical phenotype information. Both contamination types may arise during field or laboratory handling.

Particle contaminated individuals were re-extracted from single scale samples. If contamination persisted in the second extract, the individual was discarded from further analyses.

Samples were screened for cross-contamination using allelic match scores: the percent of alleles that match between pairs of multilocus genotypes. The chance of two individuals having a >95% matching genotype in 27-31 highly variable microsatellite loci is extremely small and thus such a pair of samples were identified as cross-contaminated. If two individuals shared more than 95% of alleles in 27-31 loci, they were determined as cross-contaminated. If both samples of the pair were provided by the same fisherman, the sample with less complete information or later reported sampling time was discarded. In all other cases, both samples were discarded. Allelic match scores were estimated using MICROSATELLITE TOOLKIT FOR EXCEL (Park 2001).

1.2.7 Microsatellite loci: standardization between laboratories

Cross-laboratory standardization of 31 microsatellite loci was performed according to Ellis et al. (2011) using a set of 144 individuals representing 13 populations genotyped in both laboratories. The populations were chosen to cover the largest distribution range of Atlantic salmon (*S. salar*) in the Barents and the White Sea areas (Table 2). Correction indices were determined for each microsatellite locus by comparing amplicons and their respective peak and bin positions. Allele bins were compared and remaining ambiguities were resolved case by case by comparing peak profiles of the individual genotypes from the two laboratories. Correction indices were then applied to standardize the allele scores between the two laboratories.

Table 2. Populations used for calibration of microsatellite loci, their geographical locations and number of individuals (n).

Population	***Year***	***Location***		***n***
Russia				
Varzuga	2009	66.29°N	36.86°E	8
Kitsa	2009	66.29°N	36.88°E	8
Ponoi (mainstem)	2008/2009	66.98°N	41.29°E	16
Drozdovka	2009	68.30°N	38.44°E	8
Vostochnaya Litsa	2009	68.63°N	37.79°E	8
Rynda	2009	68.92°N	36.83°E	16
Kola	2009	68.88°N	33.03°E	8
Ulita	2009	68.69°N	32.10°E	8
Ura	2009	69.30°N	32.85°E	8
Titovka	2009	69.56°N	32.02°E	8
Norway				
Børselv	2006	70.31°N	25.52°E	16
Lakselva Porsanger	2006	70.08°N	24.93°E	16
Alta	2006	69.97°N	23.38°E	16
Total				**144**

1.3 Statistical analysis

1.3.1 Quality control - analyses of kinship in the baseline samples

When sampling juveniles from a river to establish a genetic profile for the population, it is important to try to collect a sample that is representative for the population as a whole, and that the sample captures as much as possible of the genetic variation present in the population. This is not always easy to accomplish, as the resources for both collection of samples, and for analysis of these samples for variation in genetic markers are usually limited. There may be genetic variation within a river between different stretches of the river due to selective processes operating over time, and also local variation over short time scales due to stochastic processes and family structure. Sampling in the field may be constrained in a number of ways, not least by the accessibility of sampling areas. Not all stretches of a river are suitable for collection of juveniles by electrofishing, and often collection will be restricted to sites that are easily accessible, and where water flow and depth is suitable for electrofishing. Thus there is a risk that samples may be collected from a limited area, and that only a limited number of families may be represented among the samples collected. It has been demonstrated, that sampling of closely related individuals may have an effect on clustering algorithms applied to analyse genetic structure among populations (Rodriguez-Ramilo & Wang 2012).

Therefore, to evaluate the degree of kinship within the samples collected for the baseline, all baseline samples were analysed in the program COLONY (Jones & Wang 2010; Wang 2004, 2012), which estimates full- and halfsib relationships between individuals of

a sample. Due to computing time constraints, the analyses were made using a dataset including only 18 of the 31 DNA markers analysed, which was considered to be sufficient for precise estimates of kinship between individuals. Based on the results from this analysis, all full sibs except one was removed from the baseline samples, so that no family contained more than two siblings. Further statistical analyses were performed on the baseline containing these reduced samples with fullsibs removed.

1.3.2 Genetic structure of the baseline

The genetic structuring of the populations in the Kolarctic area was explored by cluster analysis in the program STRUCTURE 2.3.4 (Pritchard et al. 2000, Falush et al. 2003). The size of the dataset, and the different levels of genetic differentiation within the study area, made an exhaustive analysis of the complete dataset in Structure impractical, and an hierarchical approach where the dataset was explored sequentially, by identifying major genetic shifts/major clusterings, dividing the dataset between these divisions, and then repeating the analysis on these smaller datasets. This approach is similar to the one applied on analysis of salmon from Teno River by Vähä et al. (2007), and is illustrated in Fig. 3. The datasets were initially analysed using standard settings allowing for admixed individuals and correlated allele frequencies, and not using sampling localities as a prior in the analysis. Burn-in runs were set to 125.000 before conducting 250.000 runs. Five replicate runs were conducted at each of the K-values explored, and the results were further analysed in STRUCTURE HARVESTER (Earl & von Holdt 2012), before replicate runs were combined in the CLUMPP program (Jakobsson & Rosenberg 2007). Microsoft Excel was used to construct plots of the merged replicates of population clustering from CLUMPP.

Population descriptive statistics (e.g. observed and expected heterozygosity and allelic richness) were calculated in the R 3.0.3 statistical package (R Core Team 2013), using the package diveRsity (Keenan et al. 2013). For estimates of allelic richness, the calculations were based on 1000 resamples with n = smallest sample size. GENEPOP 4.2.1 (Raymond & Rousset 1995) was used to test samples from all locations and all loci for conformance with Hardy– Weinberg expectations and gametic disequilibrium and to estimate F_{ST} for all loci, and pairwise F_{ST} between populations. GENEPOP was also used to test for genic differentiation between all pairs of populations samples.

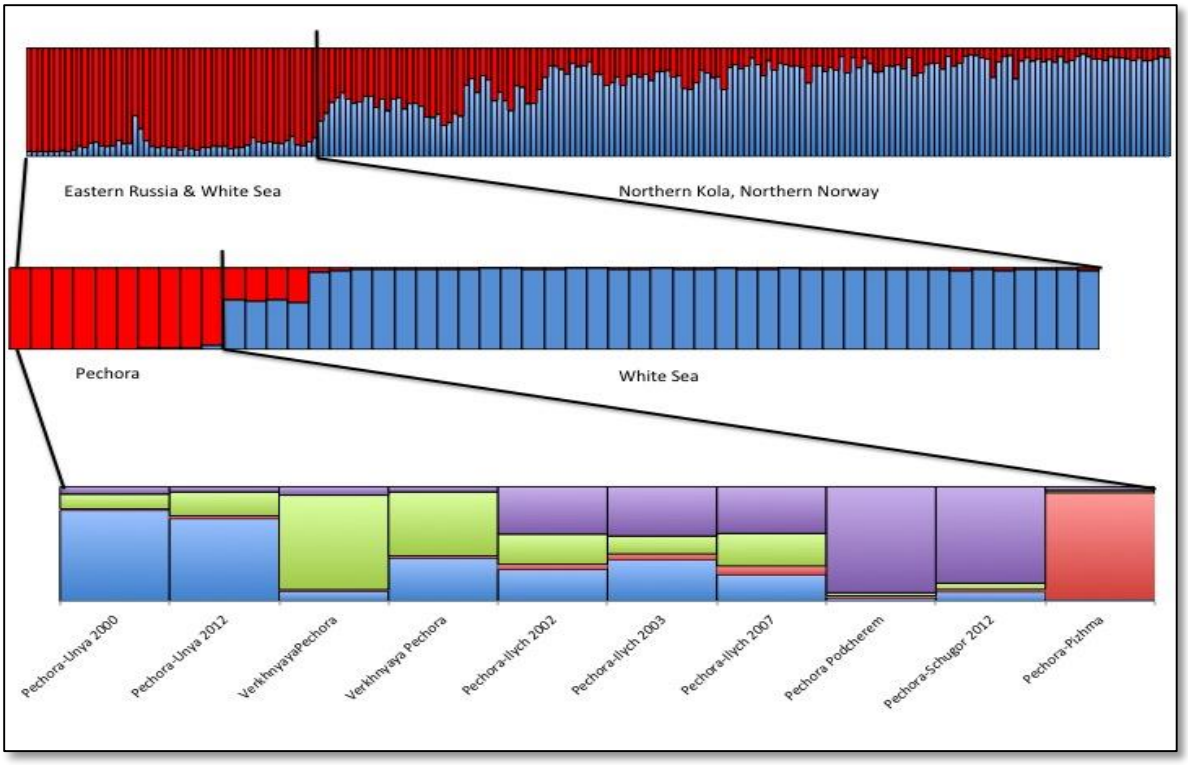


Figure 3. Hierarchical analysis of genetic clusters in Structure. The upper plot shows the full baseline analysed for K = 2, the middle plot shows the segment on the left of the upper plot, the eastern rivers, reanalyzed at K = 2, showing a clear division between Pechora and other White Sea populations. When analysed alone, the Pechora samples shows internal river structure at K = 4.

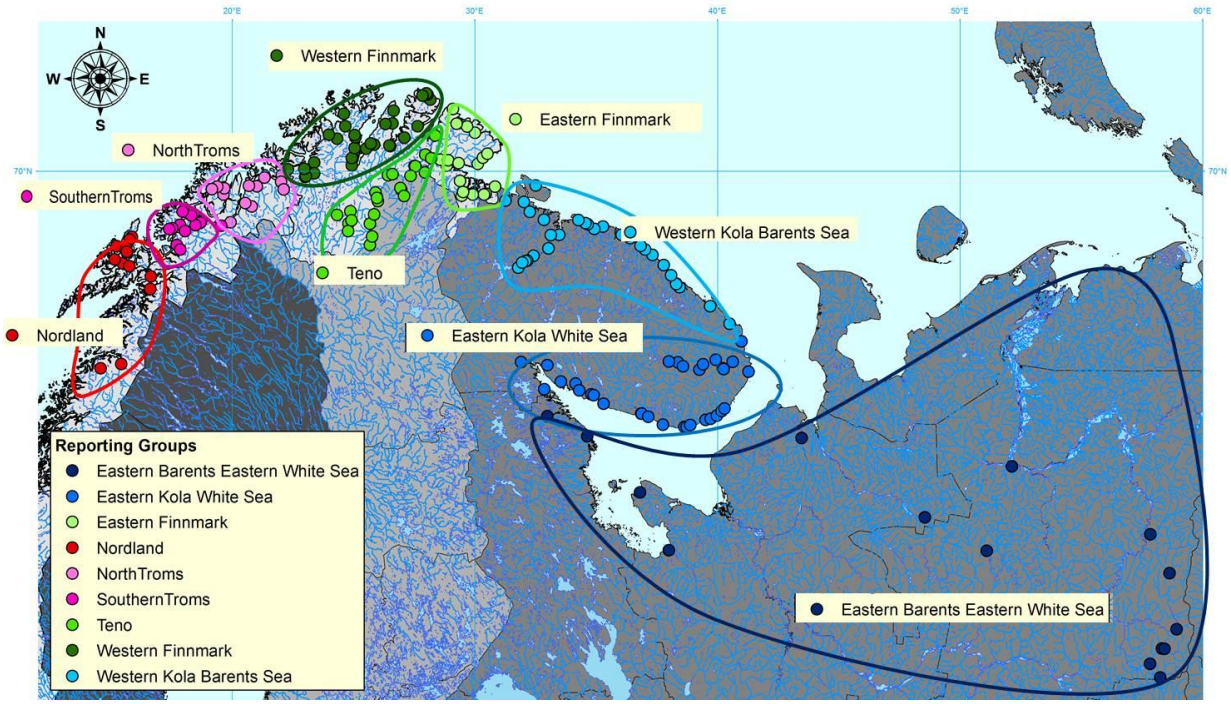


Figure 4. Map showing the reporting groups used for genetic stock identification (see 1.4).

1.3.3 Reporting groups

Baseline populations were grouped into 155 river-based and 9 regional reporting groups that were determined primarily by the relative genetic similarity among populations according to phylogenetic and genetic structure analyses and according to management priorities (Appendix Table 2, Fig. 4). The following abbreviations were used when referring to the reporting groups: I) 01 Eastern B and WS; II) 02 Eastern Kola; III) 03 Western Kola; IV) 04 Eastern FM; V) 05 Teno River; VI) 06 Western FM; VII) 07 Northern Troms; VIII) 08 SouthernTroms; and IX) 09 Nordland.

1.4 Genetic stock identification

1.4.1 Assessing the power of different methods for GSI

The performance of tree approaches for individual assignment were initially tested using a simulated dataset; GENECLASS implementing the method of Rannala and Mountain (1997), tailored to mixture modelling in ONCOR (Kalinowski et al. 2007) and CBAYES (Neaves et al. 2005) implementing the pseudo-Bayesian method of Pella and Masuda (2001). There are many software packages available for performing GSI, but ONCOR and CBAYES are most commonly used.

Rannala and Mountain (1997) use an equal probability Dirichlet density as the prior for the allele frequencies at a locus. RM assigns a frequency of 1/(n+1) to an allele not found in a population. The prior densities updated with the observed baseline data give the posterior probability densities of allele frequencies.

In Pella and Masuda (2001) the prior distribution of alleles at a locus follows the mean of the allele frequencies over all stocks. Then, posterior distributions of the baseline allele frequencies are the product of priors and the observed AFs. Shrinking the observed values toward central values takes advantage of the genetic similarity of populations is thought to minimize estimation error in allele frequencies. Further, the allele frequencies of mixture individuals assigned to a baseline population, at each MCMC step, are used to update the baseline allele frequencies (Michele Masuda, personal communication).

Initially, to test the three methods, we simulated 25 multilocus genotypes of 30 populations including 800 individuals each using hierarchical island model assuming 10 archipelagos (2-4 populations in each) with EASYPOP software (Balloux 2001). Further, a subset of 40 individuals per population was used as the baseline (1200 individuals in total) and another subset of 100 individuals per population was used as the mixed fishery sample (3000 individuals in total).

1.4.2 GSI: power analyses

The accuracy of individual assignments was tested by analyzing fishery test samples (= control samples) built from the baseline data. Each fishery test sample included 10 individuals from each baseline population sample from 3 neighboring reporting groups at a time. Each such compilation was repeated five times with different baseline individuals if the sample size allowed. Fishery test samples were analyzed as compiled and together with a subset of real fishery samples (unknown stock compositions).

1.4.3 Genetic stock identification of the Norwegian coastal fishery samples

Fishery catch samples from 74 fishermen were grouped into 24 analysis regions and two time periods within sampling year period 1 (May-June), and period 2 (July-August) (Table 3). Altogether, 20976 samples were divided to 129 subsets for analysis after seven subsets were combined within a year due to low number of samples.

Table 3. Analysis regions, Norwegian coastal fishery samples.

Region	Year	2008		2009		2011		2012		
	Period	1	2	1	2	1	2	1	2	TOTAL
1 - Sør-Varanger - East		189	47			296	96	407	117	1152
2 - Sør-Varanger - West		722	290	114	34	666	567	555	652	3600
3 - Nesseby - Fjord		146	94	137	33	152	146	164	234	1106
4 - Vadsø						135	249	181	276	841
5 - Vardø		46	43					47	81	217
6 - Båtsfjord-Berlevåg		116	34	21	15	51	91	105	87	520
7 - Teno		340	119	43	26	90	51	78	179	926
8 - Gamvik		49	48	36	22	80	6*	6*	13	260
9 - Lebesby		331	178			133	253	129	552	1576
10 - Nordkapp – Outer		149	71			122	85	236	197	860
11 - Nordkapp – Inner		23	119	29	69	178	230	117	432	1197
12 – Porsanger						15	8*	26	327	376
13 – Måsøy		30	1*			9*	46	154	122	362
14 - Kvalsund-Hammerfest						85	73	38	59	255
15 – Hasvik		199		10	6*	108	45	123	31	522
16 – Loppa		176	292	19	84	84	126	139	234	1154
17 – Alta			39	1*	95	45	255	123	202	760
18 - N.-Troms – Inner						191	287	238	423	1139
19 - N.-Troms – Outer								86	144	230
20 - S.-Troms – North out						297	169	172	158	796
21 - S.-Troms – Middle out						314	483	482	628	1907
22 - S.-Troms – Middle in						81	86	168	115	450
23 - Nordland - North						59	63	109	44	275

24 - Nordland - South					37	35	296	127	495
TOTAL	**2516**	**1375**	**410**	**384**	**3228**	**3450**	**4179**	**5434**	**20976**

1.4.4 Genetic stock identification of the Russian coastal fishery samples

Fishery catch samples from 19 locations of coastal and scientific fishery in Russia in 2010-2012 were grouped into 7 analysis regions (Table 4). The subsets were combined across the years due to low number of samples.

Table 4. Analysis regions, Russian coastal fishery samples.

Region	**Year**			**Total**
	2010	**2011**	**2012**	
Big Eina_estuary		55	1	56
Kola Bay			15	15
Ponoi estuary			116	116
Tersky Bereg	178	442	860	1480
West of Umba		45	92	137
Severnaya Dvina		111	101	212
Zinmiy Bereg		59		59
Pechora		94	90	184
Total	**178**	**806**	**1275**	**2259**

All power and genetic stock identification analyses were performed using five independent chains of 100K iterations starting from 3 random stocks in CBAYES software , although population-wise Raftery-Lewis diagnostics (Lewis & Raftery 1997) of the preliminary test runs suggested 40K-80K iterations for adequate chain length.

Section II – Genetic population structure

2.1 Samples, and scoring of microsatellite loci in baseline samples

In total, over 14 000 tissue samples were collected from individuals in 201 populations spawning in 185 rivers along the Norwegian and Russian north-west coasts and these were analysed for the suite of 31 microsatellites. Most samples analysed yielded usable results, but a few samples were degraded to such an extent that reliable genotypes could not be established for some or all loci. This was the case for some of the Russian samples collected in the Arkhangelsk region, and one sample from the Kola Peninsula. These samples were excluded from further analysis. Also, through quality control procedures, individual samples that did not amplify for a number of loci, or individuals that after analyses turned out to be trout, or salmon-trout hybrids were excluded from further analysis.

2.2 Analysis of kinship and family structure in baseline samples

The analyses of kinship in the samples collected demonstrated that full siblings were present in most samples, but also that the proportion varied greatly, with some samples having no full siblings, and other samples containing up to 68% (Storelva Båtsfjord). On average, the baseline samples contained 16% full siblings. The proportion of siblings, as well as the number of samples retained for further analysis after sibling removal are listed in Appendix Table 1, and an example of the graphical representation of full and half sib relationships within river samples is presented in Fig. 5. The results from other statistical analyses presented below are based on the baseline with all full siblings except one removed.

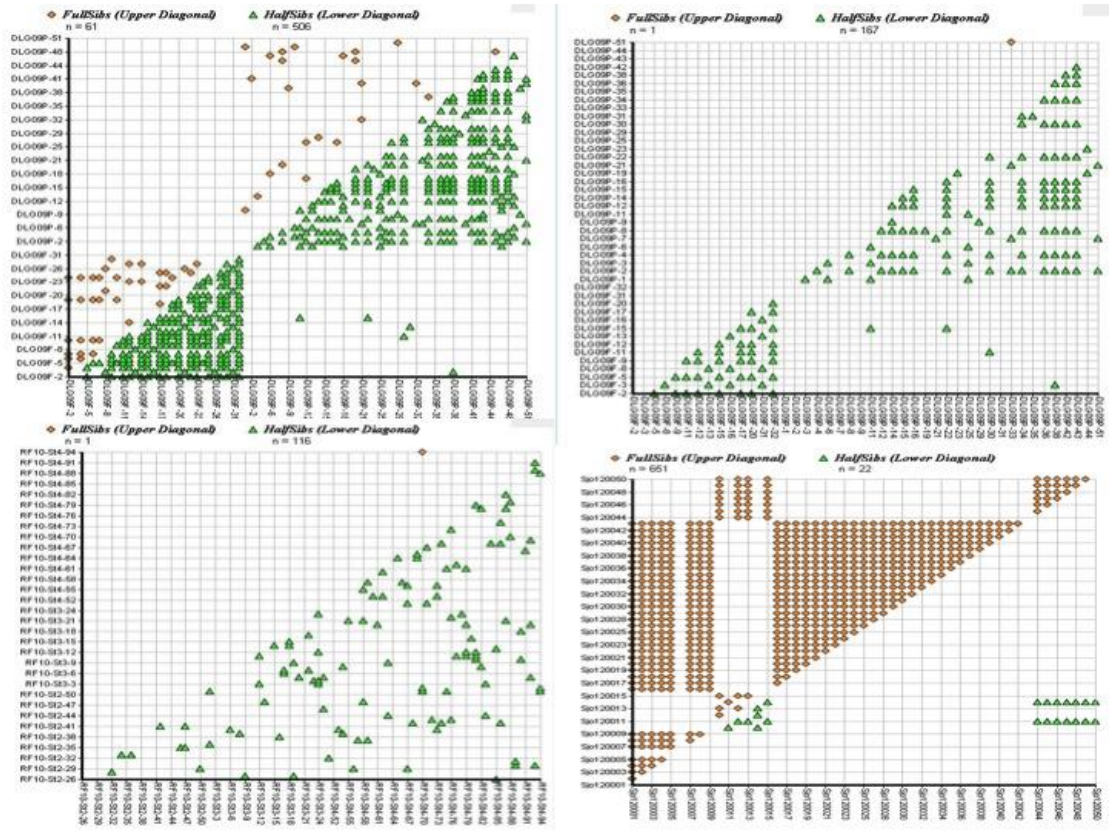


Figure 5. Colony plots showing the number of full siblings and half siblings. The two upper panels shows plots for river Dolgaya before and after removal of all but one of the full sibs within a family. The lower left plot shows a sample with little kinship between individuals (Repparfjordelva), and the lower right an extreme situation where almost all individuals are full siblings. This sample could not be used in the baseline.

2.3 Microsatellite locus variability and diversity in baseline and coastal samples

In the baseline samples, the observed heterozygosity varied from 0.012 in SsaD486 to 0.926 in SsaD144 with mean overall loci of 0.704, the same loci demonstrated minimum and maximum heterozygosity in coastal samples with mean of 0.703 (Table 5). Similarly, the number of alleles per locus varied from 6 to 43 with locus SsaD486 being least variable and locus EST19 – most variable among baseline populations. The same loci demonstrated minimum (5) and maximum (44) number of alleles among coastal samples. In total, 660 and 673 alleles were observed in the baseline and coastal samples, respectively. The highest overall F_{ST} was observed at locus Ssa412, suggesting that this locus may be very informative in describing structure, and for assignment of individuals. The lowest overall F_{ST} was observed in SsaD486, which was monomorphic in many samples (Table 5).

Table 5. Statistical properties of the loci in baseline and coastal samples.

Locus	F_{ST}	G'_{ST}	D	H_O		H_E		Number of alleles		Alleles **NOT observed** in:	
		Baseline		*Baseline*	*Coastal*	*Baseline*	*Coastal*	*Baseline*	*Coastal*	*Baseline*	*Coastal*
EST28	0.054	0.081	0.029	0.333	0.310	0.333	0.325	15	16	372	
EST68	0.068	0.178	0.118	0.611	0.629	0.618	0.666	8	9	187	
MHCII	0.093	0.331	0.262	0.709	0.697	0.718	0.786	18	19	413	
SSA405	0.033	0.410	0.389	0.920	0.921	0.918	0.949	33	34	419	
SSf43	0.042	0.069	0.028	0.390	0.382	0.391	0.401	12	13	129	
SSsp2215	0.038	0.348	0.322	0.888	0.898	0.890	0.927	31	30		197
SSsp2216	0.036	0.278	0.251	0.877	0.882	0.871	0.905	24	25	346; 366	354
SSsp3016	0.049	0.233	0.194	0.797	0.794	0.791	0.830	22	24	80; 164	
Ssa14	0.095	0.175	0.088	0.458	0.468	0.453	0.493	7	7		
Ssa171	0.052	0.276	0.236	0.813	0.825	0.811	0.855	29	29		
Ssa197	0.044	0.400	0.372	0.889	0.898	0.889	0.929	31	30		147
Ssa202	0.059	0.337	0.295	0.827	0.829	0.825	0.862	20	19	296	226; 330
Ssa289KA	0.082	0.232	0.163	0.634	0.624	0.646	0.678	12	9		101; 105; 109
Ssa412	0.131	0.289	0.182	0.548	0.534	0.546	0.588	10	11	298	
Ssa98	0.055	0.092	0.039	0.397	0.348	0.400	0.378	17	18	209; 217	233
SsaD486	0.025	0.025	0.000	0.012	0.009	0.012	0.009	6	5		163
Ssosl25	0.062	0.245	0.195	0.751	0.749	0.746	0.787	20	21	181	
EST107	0.049	0.137	0.092	0.643	0.635	0.638	0.663	10	11	354	
EST19	0.042	0.357	0.328	0.879	0.878	0.881	0.911	43	44	340	
MHC I	0.058	0.296	0.252	0.797	0.794	0.803	0.849	20	22	126; 130	
SSsp1605	0.063	0.251	0.200	0.750	0.782	0.747	0.823	13	13		
SSsp2201	0.030	0.375	0.356	0.920	0.925	0.921	0.950	40	41	385; 391; 399	235; 249
SSsp2210	0.069	0.315	0.265	0.783	0.779	0.782	0.820	17	16		101
SSspG7	0.045	0.283	0.250	0.842	0.835	0.841	0.862	25	25	213	205
Sleel53	0.090	0.182	0.101	0.500	0.535	0.505	0.575	11	10		175
Sleen82	0.057	0.215	0.168	0.729	0.729	0.735	0.769	12	13	228	
Ssa407	0.030	0.248	0.225	0.872	0.870	0.878	0.900	41	42	215; 222	281
SsaD144	0.032	0.436	0.417	0.926	0.922	0.926	0.956	39	42	104; 260; 268; 288	264
SsaD157	0.028	0.315	0.295	0.913	0.906	0.911	0.937	41	39	337; 345; 423	287; 349; 427; 431; 435
Ssleer15.1	0.093	0.249	0.171	0.618	0.628	0.624	0.680	7	8	183	
Ssosl85	0.057	0.272	0.228	0.793	0.792	0.789	0.833	26	28	172; 180; 228	212
Total	0.055	0.188	0.164	0.704	0.703	0.704	0.738	**660**	**673**		

2.4 Estimates of genetic diversity within and between the baseline samples

Estimates of genetic diversity in the baseline population samples are given in Appendix Table 2. The average number of alleles observed across loci was 306.8, with the highest value (411) being observed in Børselva, western Finnmark, and the lowest number of alleles (127) was observed in Kovda, western White Sea basin.

The mean level of genetic diversity across 31 microsatellite loci per population in the baseline also varied from relatively low in Kovda (H_E = 0.55; H_O = 0.58; A_R = 3.7) to relatively high in Tønsvikelva population in Troms (mean H_E = 0.76; H_O = 0.73; A_R =7.63) (Appendix Table 2). The average across all populations was H_E = 0.70; H_O = 0.70 and A_R = 7.0.

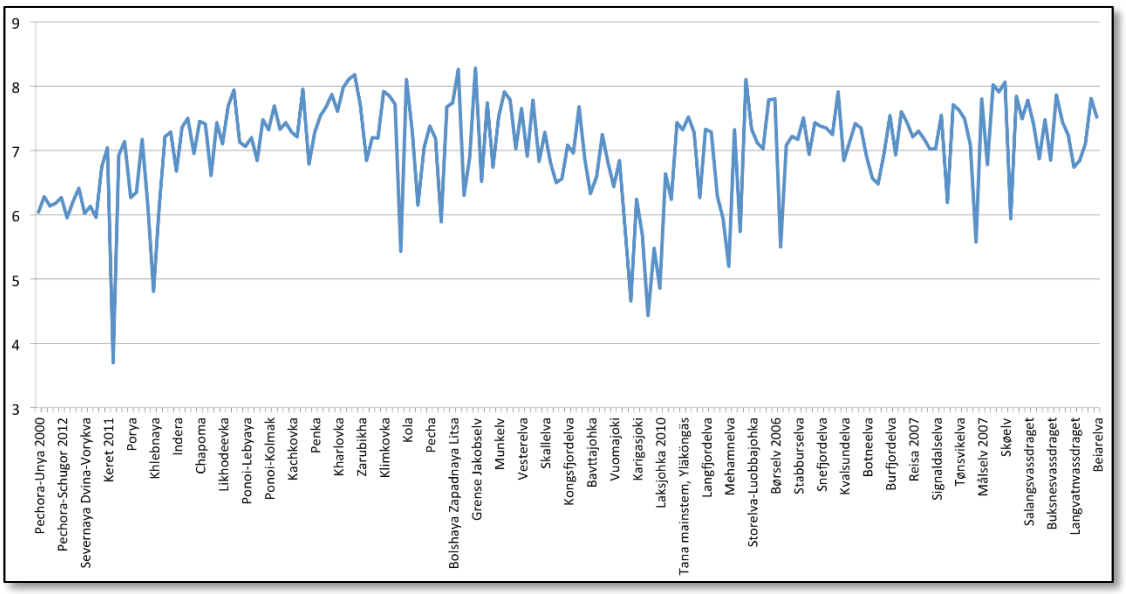


Figure 6. Allelic richness in the population samples, from Pechora to Beiarelva.

The measure allelic richness, which reflects the average number of alleles in a population using a relative sample size equal to the smallest sample, showed some interesting patterns (Fig. 6). In general, genetic diversity, estimated from allelic richness declined from east to west, and was higher on the northern coast of Kola Peninsula than in the White Sea. Also, the diversity in the Teno populations was markedly lower than populations to the east and west of Tanafjord.

Tests for linkage disequilibrium, which indicates non-random association of alleles at two loci, showed that of the 86494 pairwise comparisons of loci within the 185 population samples, $P < 0.01$ for 3650 of the comparisons. The presence of linkage disequilibrium in samples can be caused by several factors. Among those are admixture

of different populations within one sample and reduced population sizes with genetic drift.

In Fig. 7 we show that the occurrences of LD was most prominent in the samples from Maskejohka in Teno, and Smørfjordelva, indicating that these samples could have been collected from small or reduced populations. In Fig. 8 we show the distribution of LD among the 465 different locus pair combinations. For most loci combinations the number of occurrences was low, but LD was most frequent in the locus pairs Ssa407&Sleel53 and Ssa407&SsOSL25, possibly indicating physical proximity of these loci in the genome.

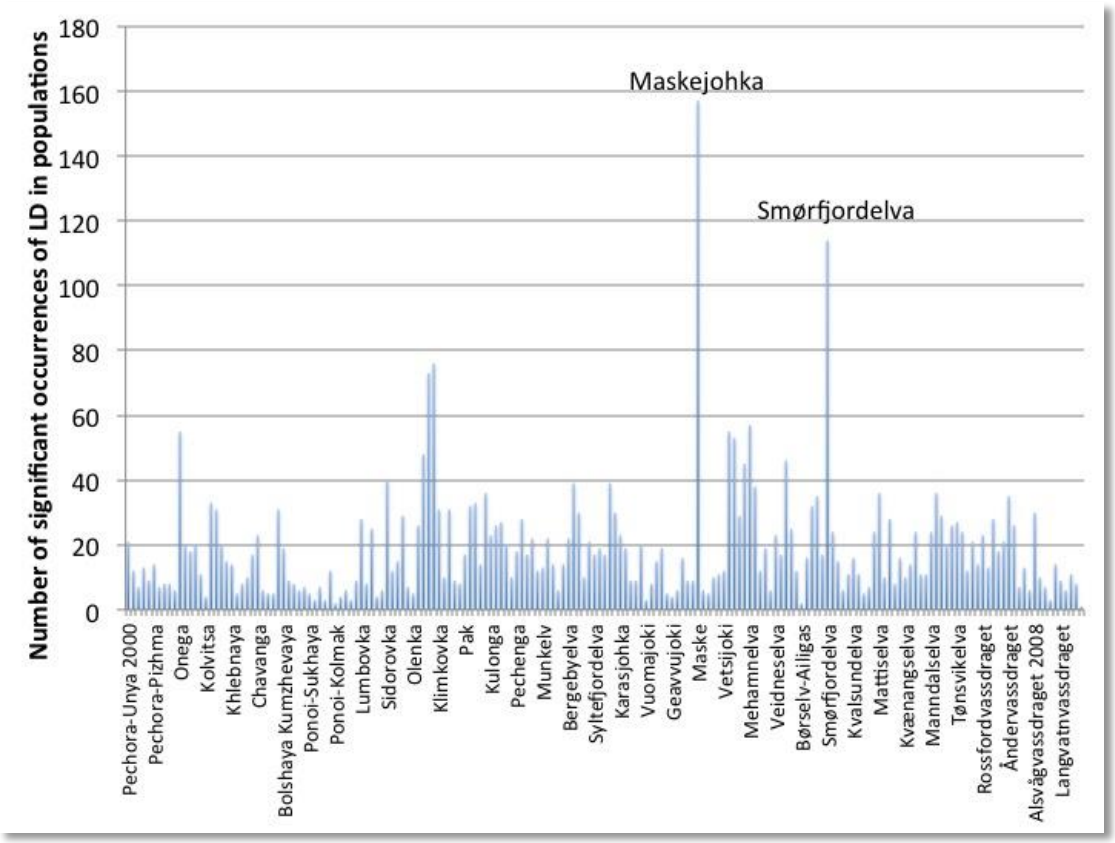


Figure 7. Occurrences of significant linkage disequilibrium in population samples.

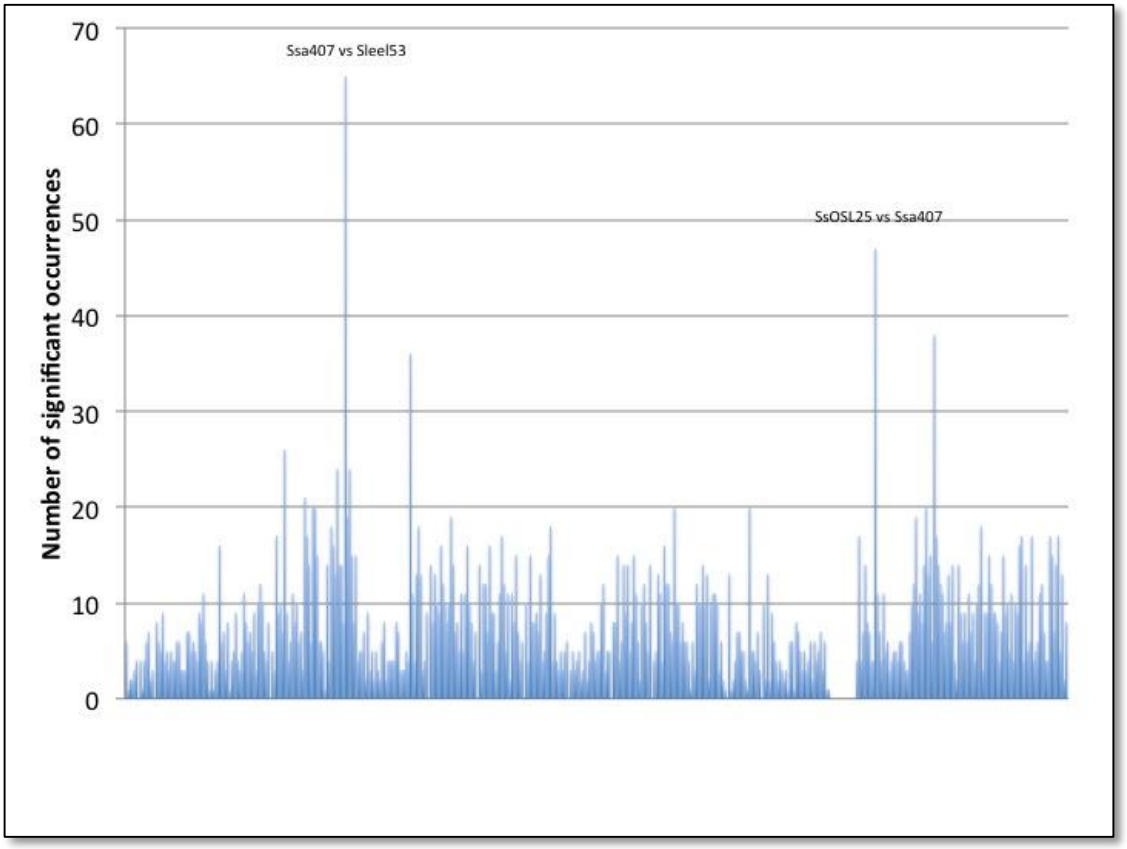


Figure 8. Occurrences of significant linkage disequilibrium in locus pairs.

Tests for deviance from Hardy-Weinberg equilibrium revealed that in 16 of the 185 populations, there was significant heterozygote deficiency at significance level 0.01 (Suppl. Table 1). However, after sequential Bonferroni correction for multiple tests (Holm 1979), only three of these remained significant. These were the sample from Onega, the sample from Umba and the sample from Kvalsundelva. In the case of Onega, the heterozygote deficiency may be an effect of the sample containing individuals from several sub-populations within the river as this sample was collected by a smolt trap. Also, when testing for heterozygote excess, only one river sample remained significant at $P < 0.01$ after sequential Boferroni correction; Manndalselva. It is ulnlikely that the relatively small deviations from Hardy-Weinberg equilibrium, and Linkage disequilibrium observed in the samples should have any major effect on our results.

2.5 Genetic relationships among the populations

Exact tests for genic differentiation in GENEPOP, using all loci combined, revealed that almost all population pairs were significantly different at significance level $P < 0.001$ (Suppl. Table 2). The exceptions to this general pattern were Børselva and its tributary

Ailigas, Åseelva and Gårdselva, Ponoi mainstem and the Ponoi tributary Tomba, Ponoi tributaries Ryoboga and Tomba, Pacha and Purnach, Purnach and Ryoboga, and Pacha and Tomba. Åseelva is a small population, and allele frequencies in this river are probably subject to fluctuations over years, as occasional strayers from other rivers successfully spawning would have a significant genetic impact in such a small population. The high genetic diversity observed in this small river (see section 2.4) is also an indication of strayers of multiple origins contributing to the population. Low genetic differentiation between the Ponoi tributaries is discussed further in section 2.7.

Genetic differentiation measured as F_{ST} was calculated between all pairs of populations (Suppl. Table 3). On average, the F_{ST} between populations was 0.055. This is high compared to studies of Atlantic salmon in other regions, where F_{ST} values in the range 0.01 – 0.04 have often been reported for Atlantic salmon (Skaala et al. 2004, Griffiths et al. 2010). The highest value observed in this dataset was 0.254, between the samples Laksjohka 2010 and Kovda, which is on the same level as those reported between populations from the eastern and western side of Atlantic (Wennevik et al. 2004). The lowest F_{ST} value of 0.001 was observed between the sample from Ponoi mainstem, and the Ponoi tributary Lebyazhya. In general, the F_{ST} values were highest between the Pechora tributaries and rivers in the western range of the Kolarctic area. For instance, the overall average F_{ST} between the upper Pechora tributary Unya and all other populations was 0.108, which can be considered very high. Some of the lowest F_{ST} values was observed between populations in the western range of the Kolarctic area.

2.6 Cluster analyses in Structure

As explained in section 1.3.2, the baseline dataset was analysed in STRUCTURE using a hierarchical approach. This approach was chosen for several reasons; cluster analysis of large datasets with many samples and many markers is computationally intensive and runs take a long time to complete (in some cases days), and also, different levels of genetic differentiation within geographic regions in the dataset could result in greater differences in the dataset masking smaller differences. By splitting the dataset and analyzing such regions separately, significant structure at lower levels of genetic differentiation can be resolved.

We initially analysed the complete dataset using low numbers of inferred clusters (K = 2, K = 3) as input to the program, identifying the major genetic transitions and then splitting the dataset. This approach was shown in Fig. 3 in the previous section. At the uppermost level, we divided the dataset into two main parts; division 1 and division 2 between the rivers Ponoi and Kachkovka in the Kola peninsula. This was the most apparent division at low K when looking at the complete dataset. It is also interesting to observe, that when the dataset of European populations, containing some of the same rivers as our Kolarctic dataset analysed for a subset of the markers, was analysed in the EU-project SALSEA-Merge the main genetic division in the northern populations was

observed in Troms county (V. Wennevik, unpublished data). At lower hierarchical levels, more runs could be conducted with different values for K, and also by applying various measures such as those suggested by Evanno et al. (2005) the "true K", or numbers of clusters in the dataset, could be estimated.

Below we explain how the cluster analysis in Structure classified groups of populations into genetic clusters, and illustrate this with plots of the different regions.

2.6.1 Division 1 – Pechora to Ponoi

In Fig. 10 we present how the clustering of populations in the eastern part of the Kolarctic area, division 1 (see Fig. 9), resolved through analysis at different levels of clustering/increasing K. At the lowest level, K = 2, the major split can be seen between the samples from tributaries of the Pechora river system, and the other populations in the area (red vs. blue). Some populations in the Arkhangelsk region appear to be a mixture of the two main clusters. With increasing K, one of these two main clusters (the blue cluster) resolves into several clusters, mainly corresponding to the geographic distribution of populations, while the red cluster comprising the samples from Pechora remain a distinct unit also at higher levels of K, demonstrating the large genetic differentiation of the Pechora populations relative to the other populations in the region. At K = 3 we observe how the blue cluster is split into two different clusters. One (the green cluster) is composed of the populations in the Arkhangelsk, Karelia and the inner part of the White Sea (Mezen to Olenitsa), wile the other cluster contains populations from Varzuga to Ponoi. At higher levels of K populations in the Arkhangelsk area break out to form their own cluster, and Ponoi is separated from the other populations at the southern coast of Kola Peninsula. At K = 7, K = 8, the patterns begin to break down a little, introducing noise into the clusters, indicating that within this dataset, we are exceeding the number of clusters that can be reliably identified. One interesting observation is that the river L. Zolotitsa on the eastern side of the White Sea seems to be more closely related to geographically close populations on the western side.

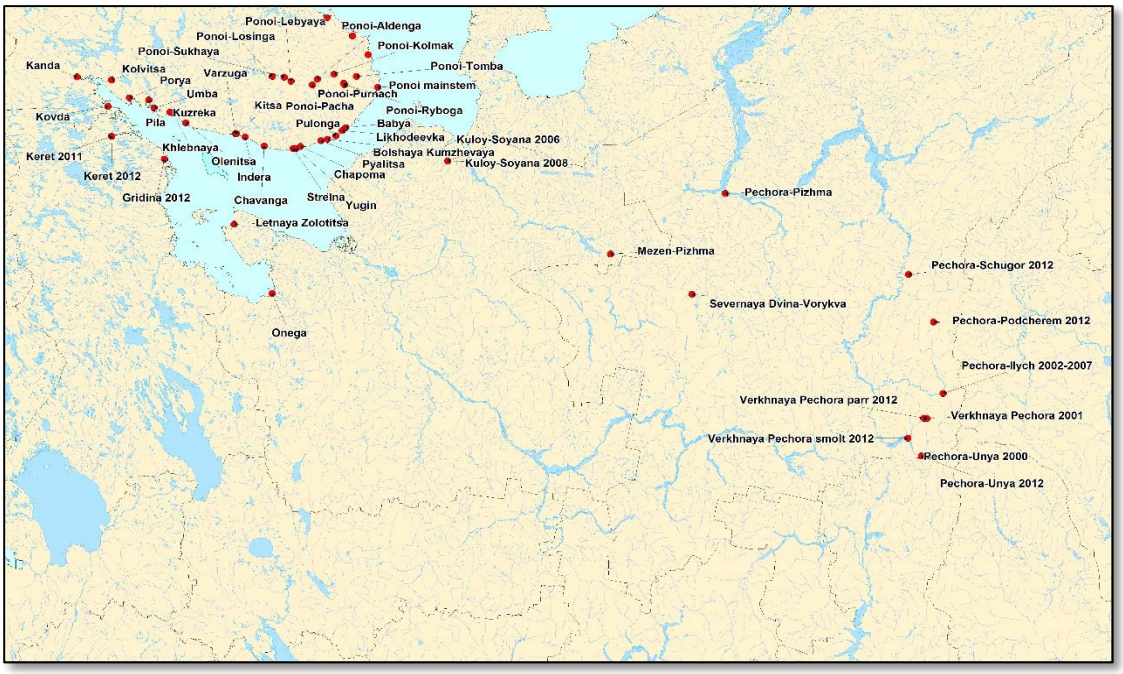


Figure 9. Map showing the rivers in division 1.

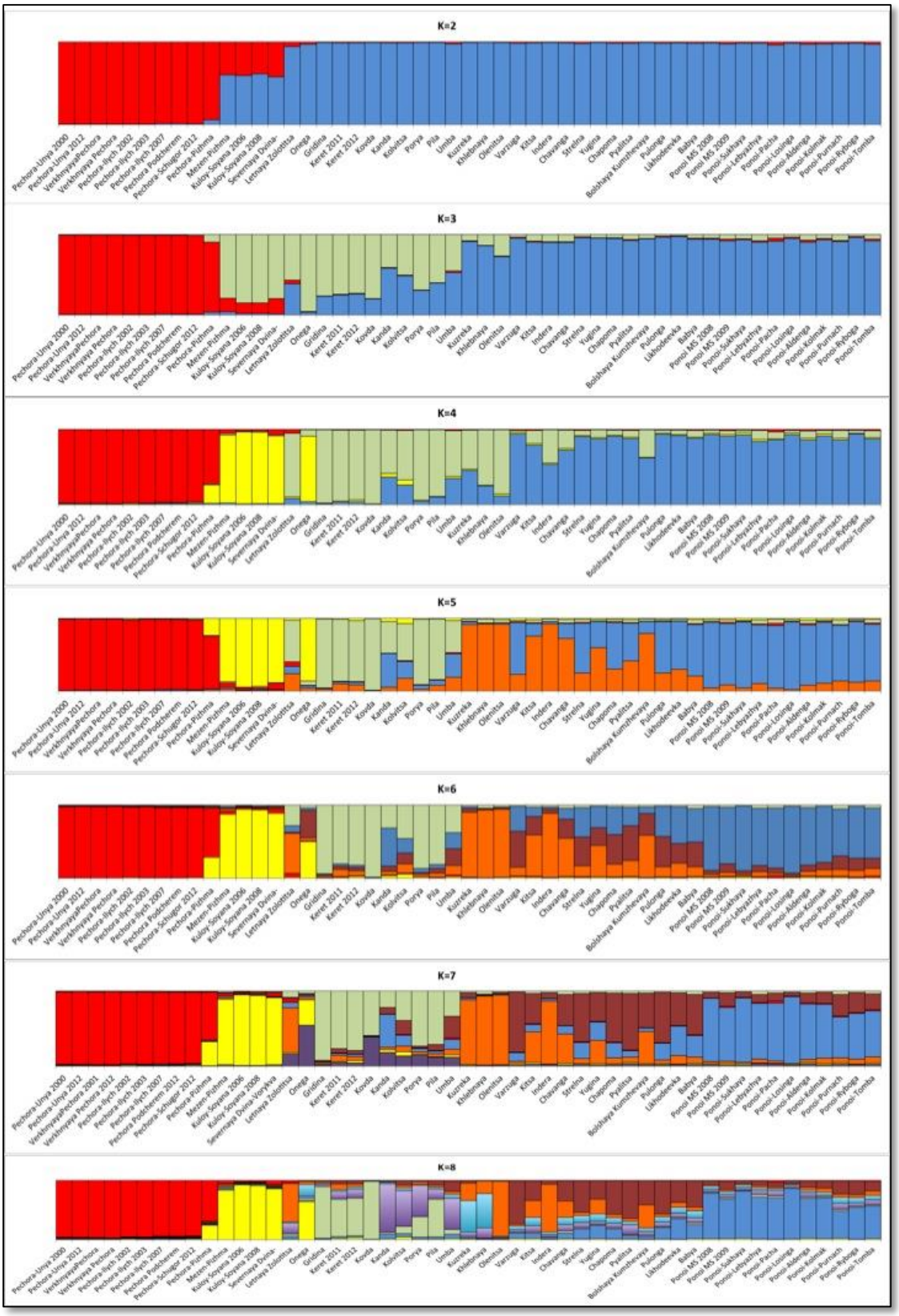


Figure 10. Results of the cluster analysis in Structure for division 1, for K = 2 – K = 8.

2.6.2 Division 2 – Kachkovka to Beiarelva

This division is much larger in terms of number of rivers and geographic range, compared to division 1, and the initial STRUCTURE analysis at low K values revealed a more complicated structure than what was observed in division 1. The results from analysis at K = 2 to K = 4 are presented in Fig. 11. Apparently in this division there are several substructures on various geographic scales, and the genetic differences between rivers is very variable as demonstrated by the pairwise F_{ST} values (Suppl. Table 3). For further analysis we therefore split this division into three subdivisions, as illustrated in Fig. 11.

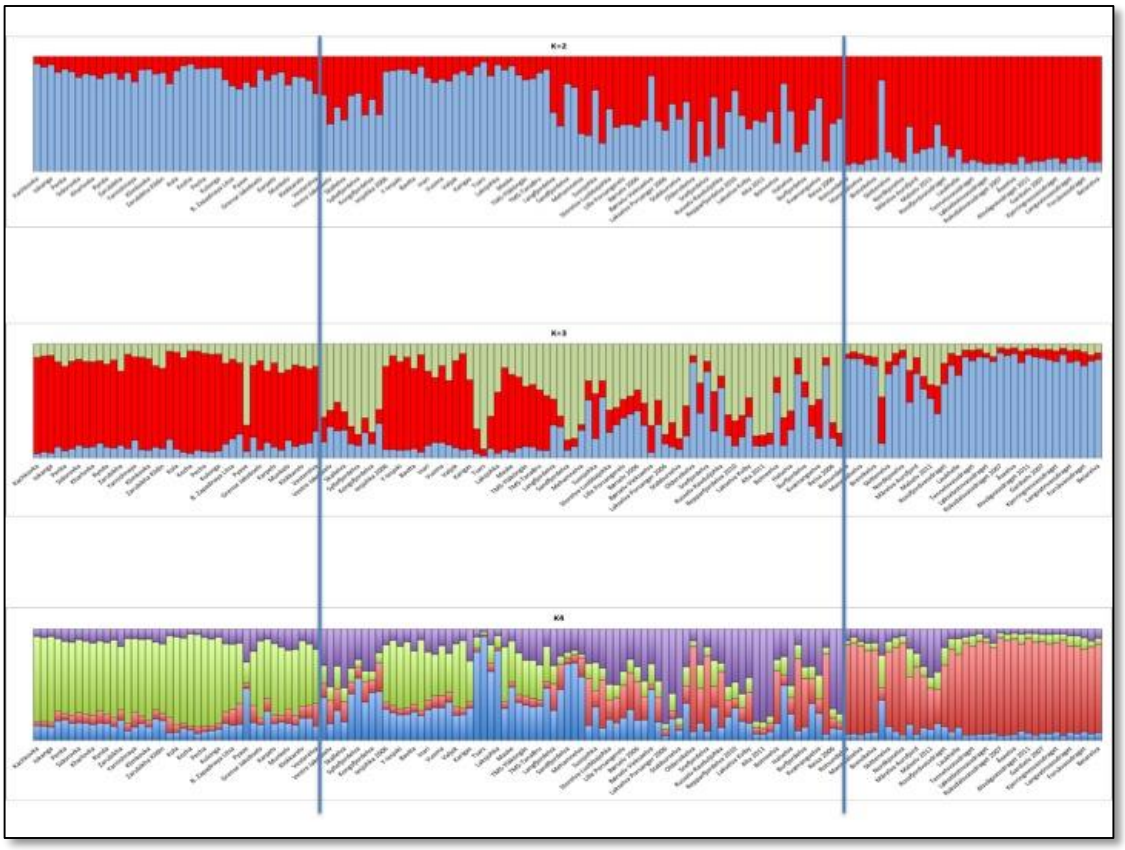


Figure 11. Results of the cluster analysis in Structure for K = 2 to K = 4 in division 2 – Kachkovka to Beiarelva - with vertical lines denoting divisions into three subdivisions.

2.6.3 Division 2.1 Kachkovka to Vesterelva

This subdivision comprises the rivers of the northern coast of the Kola Peninsula to the inner part of Varangerfjord in Norway (Fig. 12). These rivers drain into the Barents Sea, or into fjords that open into the Barents Sea. The exception to this is the rivers draining into the Tuloma basin in the inner end of the Murmansk fjord. These rivers (Pak to Shovna) drain into a freshwater basin, created by a dam, with access from the fjord

through the Tuloma fish ladder. These rivers, and the other rivers in the Murmansk fjord, form a distinct cluster (in the middle of the plots) at all values of K. At K = 4 and K = 5 it also appears that the populations on both sides of the Murmansk fjord form separate clusters, where Kachkovka to Zarubikha Kildin constitute an eastern Kola cluster, and Ura to Vesterelva form the western Kola/eastern Finnmark cluster (Fig. 13).

Figure 12. Map showing the rivers in division 2.1.

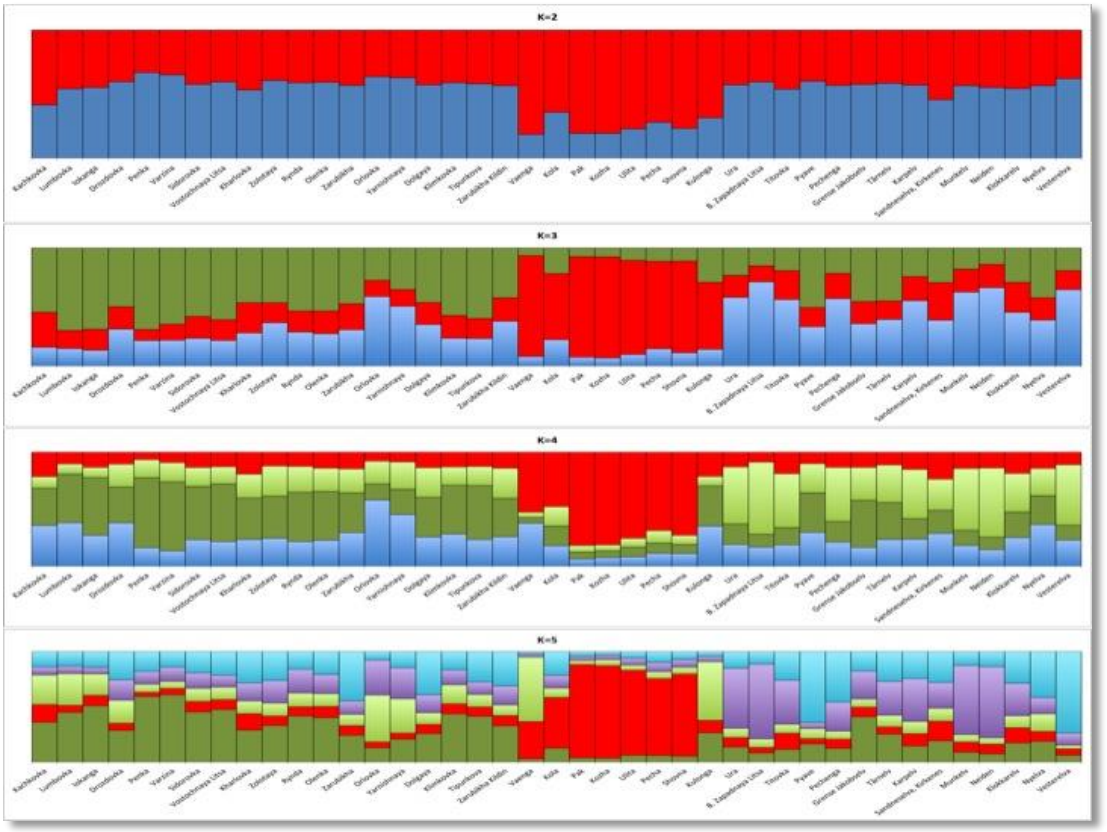


Figure 13. Results of the cluster analysis in Structure for K = 2 to K = 5 for division 2.1.

2.6.4 Division 2.2 Bergebyelva to Reisa

As was also evident in the analysis of division 2 as a whole, this middle section containing mainly Finnmark rivers is a very diverse region (Fig. 14). The Teno river system within this region is in its own a structured system with ~14 different genetic groups as have been identified in earlier work by Vähä et *al.* (2008). At K = 2 the Teno already appears as a distinct and coherent unit separated from populations to the east and west of it (Fig. 15). As K increases, more structure is revealed, also within the Teno, where some of the tributaries are quite distinct. Some of the greatest F_{ST} values were also observed between these Teno tributaries and populations in Russia. At these levels of K, no clear divisions was observed between other groups of rivers in this division, but at K = 4, it appears that the populations to the west of Teno constitute a different group relative to the populations to the east, though there seems to be some overlap. The rivers within this region represent a variety of population sizes, ranging from large river systems such as Teno, Alta, Målselv and others, to small rivers with relatively low numbers of spawners. Observing that samples collected in different years and different parts of the rivers system in some of these rivers cluster close together (see for instance different samples from Alta and Reisa) gives confidence and validates the results. Also, looking at the results it is apparent that some smaller rivers show deviant cluster composition

relative to neighbouring rivers in the same region. This could be an effect of small population sizes in these rivers, where random genetic drift, and straying from other populations may lead to year-on-year fluctuations in allele frequencies.

Figure 14. Map of rivers in division 2.2.

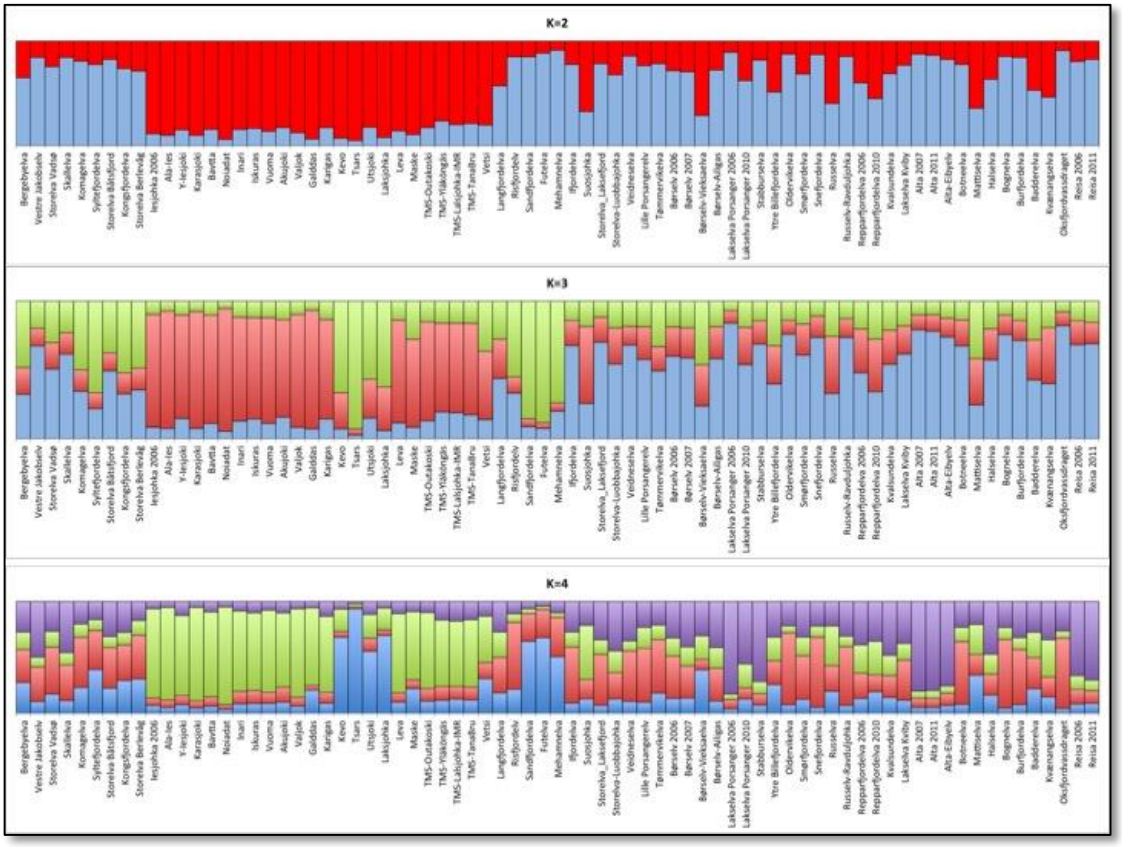


Figure 15. Results from clustering analysis in Structure of rivers in division 2.2 (Bergebyelva to Reisa) for K = 2 to K = 4.

2.6.5 Division 2.3 – Rotsundelva to Beiarelva

This division, as the previous one, is relatively diverse. Many/most rivers are small in terms of numbers of spawners, but the division contains Målselva, which is a relatively large river (Fig. 16). Målselva is represented by three samples, from different years and stretches of the river, which appear similar in the cluster plot (Fig. 17). Apparent already at K = 2 is the shift between Salangselva and the samples from Roksdalselva and onwards. This river, and the next rivers in the plot are situated on the outer coast on islands (see Fig. 16). As in the previous division, there are some small rivers in this division where the admixture of clusters within populations do not resemble that of their neighboring populations, and this could be an effect of small population size and temporal fluctuations in allele frequencies.

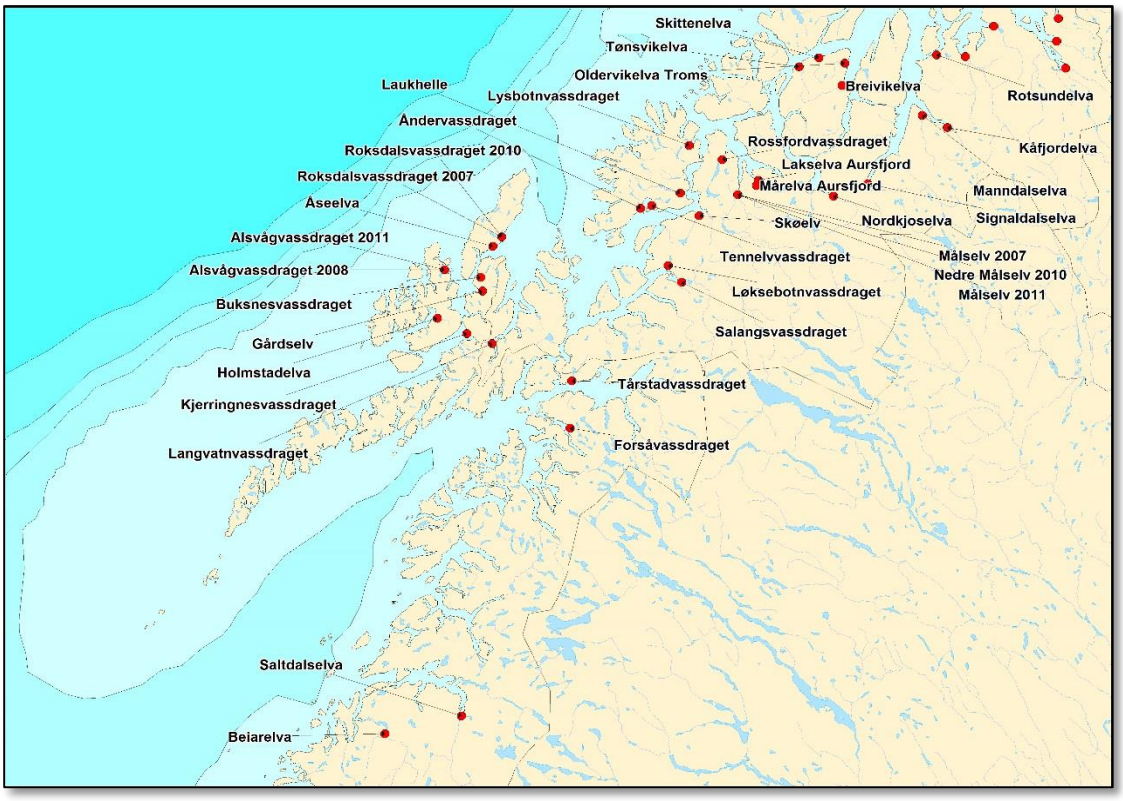


Figure 16. Map of rivers in division 2.3.

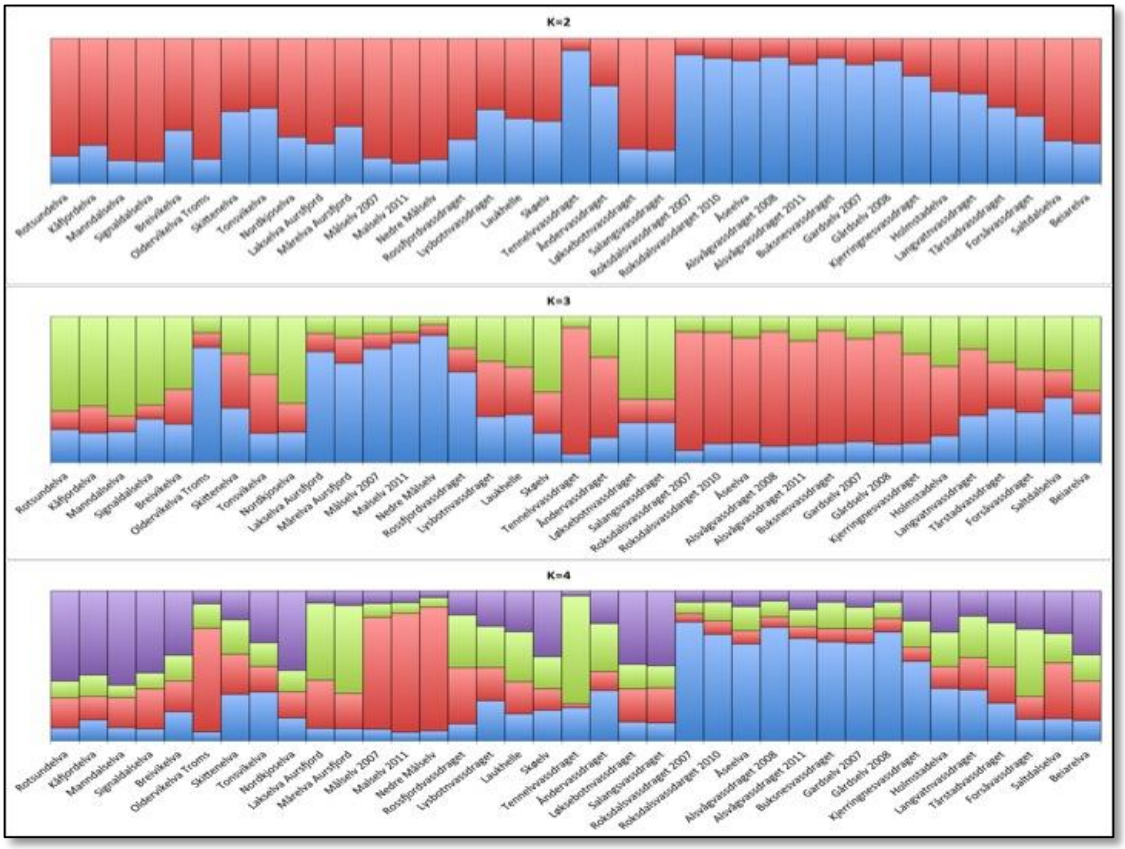


Figure 17. Results from the Structure analysis of division 2.3 - Rotsundelva to Beiarelva for K = 2 to K = 4.

2.7 Differentiation within large river systems

Earlier studies have documented intra-river structure in three of the rivers included in this baseline. Using allozymes as genetic markers, Heggberget et *al.* (1986) demonstrated genetic structuring between different parts of the Alta river system. Primmer et *al.* (2006) examined genetic variation in microsatellites in the Varzuga river and demonstrated patterns of isolation by distance within the river. The river Teno has been extensively investigated with genetic methods in later years, demonstrating high and temporally stable genetic complexity within this huge river system (Vähä et *al.* 2008). Physical complexity such as waterfalls or lakes within river systems can contribute towards development of genetic structure (Dillane et *al.* 2008), and also quality and patchiness of spawning habitat, and the size of the river system itself may have an effect.

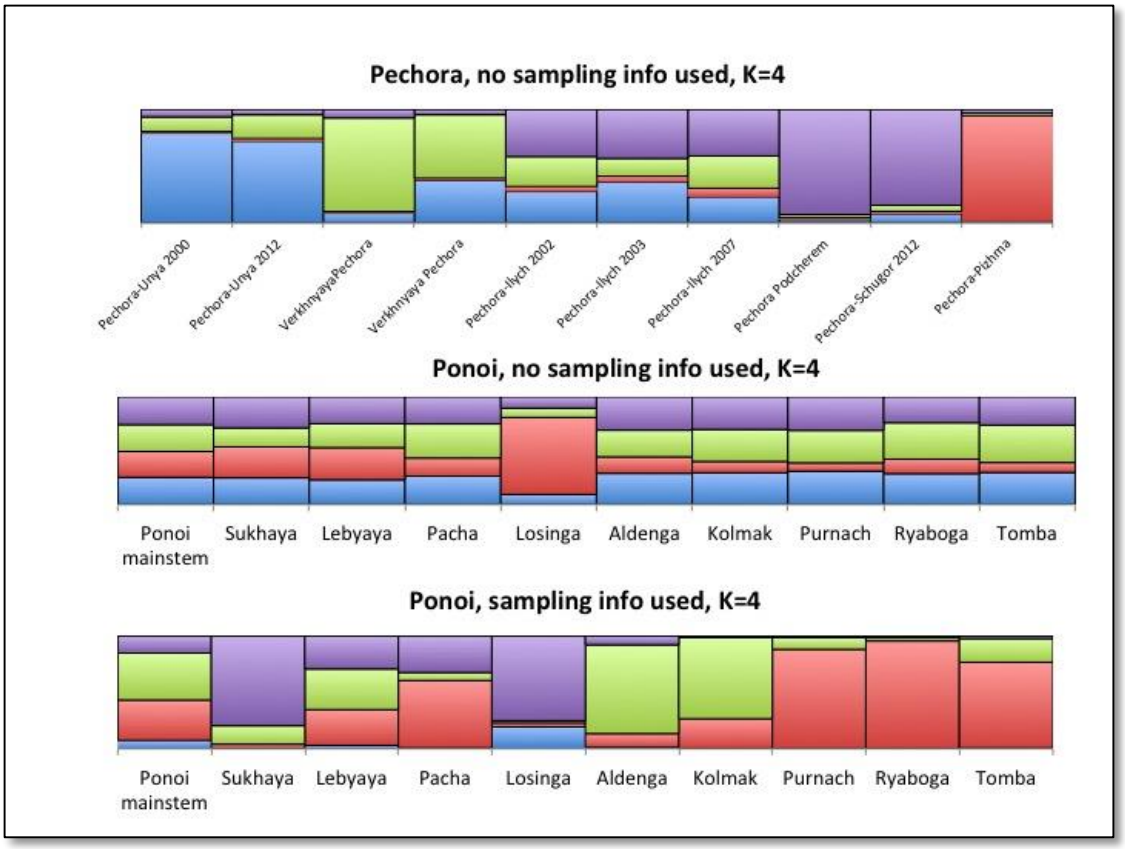


Figure 18. Structure plots showing clustering at K = 4 in Pechora and Ponoi rivers. For Ponoi to plots are shown; the upper plot shows the clustering when no information on sampling localities is incorporated in the analysis, the lower plot shows results when this info is used in the clustering.

In this project, there is available data from multiple tributaries from additional river systems; Pechora and Ponoi. Both these rivers are long river systems, with large drainage

areas, and the salmon are migrating around 2000 km to the spawning grounds of the upper Pechora tributaries, and probably undertake the longest in-river spawning migration of any Atlantic salmon population. The structure among the different tributary samples collected and analysed in the Pechora river system was presented in Fig. 3, and also here in Fig. 18.

Analyses of genetic differentiation measured as F_{ST} demonstrated that the Pechora tributaries were highly differentiated from other rivers in division 1, and even more differentiated to rivers in division 2, with many F_{ST} values exceeding 0.1, even between the Pechora and rivers within division 1 (Suppl. Table 3). Among Pechora tributaries, differences were on level with that observed between rivers in the Kolarctic area, with an average F_{ST} of 0.025. This differentiation is also evident in the STRUCTURE plots, where it also can be observed that samples from tributaries, collected over a time period of 10-12 years show temporal stability. It also seems there may be some isolation by distance patterns. Looking at the pairwise F_{ST} values for instance for the upper tributary Unya, we see that the smallest values are relative to Verkhnaya Pechora, and increasing towards the Pechorskaya Pizhma, the lowest tributary sampled. The two tributaries Podcherem and Schugor appear to be relatively little differentiated (pairwise F_{ST} = 0.018), and they are also geographically close.

The within-river structure in Ponoi appears to be less pronounced. Initial analysis in STRUCTURE, not using information on sampling location as a prior for the clustering, revealed almost no discernable structure, with the exception of Losinga (Fig. 18, middle plot). Losinga is the uppermost tributary in the river, and is also located above a lake. As mentioned previously, lakes in rivers systems may facilitate genetic structuring. When sampling location was given as prior to the clustering analysis, more structuring became discernable within the river. For instance, the lower tributaries Purnach, Ryaboga and Tomba appear to form a genetic cluster, and are also geographically close, as are Kolmak and Aldenga which also form a cluster. The average pairwise F_{ST} among the Ponoi samples was only 0.008. The highest pairwise F_{ST} (0.019) was observed between Purnach and Losinga. The lower genetic structuring observed in Ponoi may be a result of several factors. It could be a sampling issue, as several of the tributary samples were collected at locations fairly close to the mainstem, and mainstem parr may have migrated into these tributaries. Also, the river appears fairly homogenous, with suitable spawning habitat in most reaches of the river. With the exception of the lake separating the upper tributary Losinga, there are few physical barriers in the river that would facilitate structuring.

2.8 Temporal stability among baseline samples collected in different years

In some rivers and tributaries, samples were collected in different years. In total 13 rivers/tributaries were sampled at different time points, with a time span between samplings from 3-12 years. Exact tests for genic differentiation (across all loci) showed

that 9 of the 13 samples were significantly different at significance level 0.01. Samples from three sampling years from the Pechora tributary Ilych were not significantly different, as was the case for samples from Soyana, Alta and Målselv. Though some samples were different in the exact test, the F_{ST} values between samples were mostly low, with an average of only 0.007 (Table 6). This is low compared to the average between-river F_{ST} of 0.055, and temporal samples were therefore merged in the baseline used for assignments. This was also done because some of the temporal samples were small (smallest sample 12 individuals).

Table 6. Temporal river samples in the baseline. Genic differentiation at $P > 0.01$ and F_{ST} values between temporal samples.

River	Sampling years	Significantly different	F_{ST}
Unya	2000/2012	Yes	0.0253
Verkhnaya Pechora	2001/2012	Yes	0.0095
Ilych	2002/2003/2007	No	-0.0042
Soyana	2006/2008	No	0.0003
Keret	2011/2012	Yes	0.0194
Børselv	2006/2010	Yes	0.0027
Lakselva	2006/2010	Yes	0.0193
Repparfjordelva	2006/2010	Yes	0.0034
Alta	2007/2010	No	-0.0007
Reisa	2007/2011	Yes	0.0009
Målselv	2007/2011	No	0.0003
Roksdalsvassdraget	2007/2010	Yes	0.0058
Alvsvågvassdraget	2008/2011	Yes	0.0068

Section III – Genetic stock identification

3.1 Assessment of GSI accuracy

3.1.1 Choice of the method

The assignment success for simulated data varied depending on the applied approach (Table 7). The lowest proportion of correct assignments was observed for Rannala & Mountain (1997) and Kalinowski et al. (2007) methods, whereas the Bayesian approach (Pella & Masuda 2001) provided the highest proportion of correctly assigned individuals. Due to significant outperformance of the Bayesian approach (Pella & Masuda 2001), cBayes was chosen for subsequent analyses.

Table 7. Performance of three different methods on the simulated dataset including 30 populations.

Method	Correct assignments
GENECLASS (Rannala & Mountain 1997)	78%
ONCOR (Kalinowski et al. 2007)	78%
CBAYES (Neaves et al. 2005)	91%

3.1.2 Estimated vs. expected source contributions

Expected accuracy to identify the river of origin for each unknown individual in the fishery sample was evaluated through analysing fishery test samples built from the baseline data. It should be noted that such an approach decreases the number of samples in the baseline data and as such is expected to have a negative effect on the observed level of accuracy. However, such an approach was necessary since true blind samples were not available except for the Alta stock.

During the course of the project 30 externally tagged salmon of Alta origin were caught in the coastal fisheries. These samples provide a true blind sample to test the performance of the genetic assignments.

In the following sections we will explore the expected level of accuracy in assigning individuals to the reporting groups and to rivers.

3.1.3 Testing samples from the baseline data. Accuracy in assigning individuals to the reporting groups

Throughout the report, the 9 groups of populations were used for assigning individuals to their geographical region of origin. Different groups can be applied

depending on the application. These groups were compiled based on preliminary results from the genetic structure analyses and taking into account geographical location of the rivers as well as the management regions.

3.1.3.1 Assignment success of test samples

The mean correct assignment was 85% over all reporting groups. As suspected, the results demonstrated large variation in observed accuracy of genetic stock identification. Generally, individuals from rivers in Nordland and Troms counties appeared to be more difficult to identify than those from Finnmark and Russia (Table 8). Individuals from Russian rivers and Teno River system were correctly identified with 94-99% accuracy, while slightly lower success rate was obtained for the samples from rivers in Finnmark county; 86%. Individuals from rivers in Southern Troms (75%) and Nordland (72%) had higher correct identification success than northern Troms, where the reporting group was correctly identified for only 59% of the individuals.

Incorrect assignments were generally to the neighbouring reporting groups. For example, mis-assigned individuals of northern Troms were incorrectly assigned to western Finnmark rivers (47%) and southern Troms rivers (38%).

On the other hand, 94% of the samples assigned to five reporting groups from 01 Eastern Barents and White Sea to 05 Teno river were correct. However only 65% of the samples assigned to northern Troms, southern Troms and Nordland were correct. The largest contributions to the miss-assignments were again from the neighbouring areas. For example, 58% of the samples incorrectly assigned to northern Troms reporting group were actually of southern Troms origin. While this result highlights the uncertainty related to identifying Troms area salmon, it also reflects the unequal sizes of the reporting groups and the number of baseline rivers within the groups. When northern Troms and southern Troms reporting groups were combined, the proportion of correctly identified samples increased to 80%.

Analysing test samples from three reporting groups separately illustrated that the high level of correct assignment to reporting groups 01 to 06 was not significantly affected by the composition of the fishery test sample; the proportion of correctly assigned samples in reporting groups 01 to 05 was always more than 90%. For northern and southern Troms the proportion of correctly assigned samples varied from 64% to 74% depending on the test fishery sample composition.

3.1.3.2 Assignment success of the test samples in the real mixtures

Overall level of accuracy and pattern of assignments did not change significantly when fishery test sample compilations were analysed together with real fishery samples. Small changes did occur of which the most notable were observed for the eastern Finnmark and Northern Troms test samples where the proportion of correct assignment

decreased by four percent units and for southern Troms samples, where the proportion of correct assignments increased by five percent units (Table 9).

Interestingly, mis-assignment of individuals from rivers in eastern Finnmark to western Kola increased from 6% to 10% possibly reflecting the high proportion of Russian origin fish in the real mixture sample accompanying the test samples (see results section 3.1.4.4).

Notwithstanding, the level of correct assignment of the samples assigned to reporting groups from 01 to 05 remained above 90%. Again, lowest correct assignment levels were observed for the northern Troms and southern Troms reporting groups. Combining the assignments to two Troms reporting groups increased the accuracy to 78% for the salmon of Troms origin.

Table 8. Assignment success of the test samples.

a) correctly identified

	Reporting Group to which they were assigned to								
Reporting Group of origin	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 Southern Troms	09 Nordland
01 Eastern B and WS	**99 %**	1 %							
02 Eastern Kola		**97 %**	3 %						
03 Western Kola		3 %	**94 %**	2 %	1 %				
04 Eastern FM		1 %	6 %	**86 %**	2 %	4 %			
05 Teno River				1 %	**98 %**	1 %			
06 Western FM				2 %	2 %	**86 %**	6 %	3 %	1 %
07 Northern Troms					2 %	19 %	**59 %**	16 %	4 %
08 SouthernTroms						5 %	14 %	**75 %**	5 %
09 Nordland							11 %	17 %	**72 %**

b) correctly assigned

	Reporting Group to which they were assigned to								
Reporting Group of origin	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 SouthernTroms	09 Nordland
01 Eastern B and WS	**98 %**								
02 Eastern Kola	2 %	**95 %**	2 %						
03 Western Kola		4 %	**94 %**	4 %	1 %		1 %		
04 Eastern FM		1 %	3 %	**91 %**	2 %	2 %	1 %		
05 Teno River				1 %	**94 %**	1 %			
06 Western FM				3 %	3 %	**88 %**	18 %	9 %	9 %
07 Northern Troms					1 %	8 %	**67 %**	22 %	14 %
08 SouthernTroms						1 %	10 %	**62 %**	11 %
09 Nordland							3 %	6 %	**66 %**

Table 9. Assignment success of the test samples in the real mixtures.

a) correctly identified

	Reporting Group to which they were assigned to								
Reporting Group of origin	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 Southern Troms	09 Nordland
01 Eastern B and WS	**99 %**	1 %							
02 Eastern Kola		**95 %**	4 %						
03 Western Kola		2 %	**94 %**	2 %	1 %				
04 Eastern FM		1 %	10 %	**82 %**	2 %	4 %	1 %		
05 Teno River			1 %	0 %	**96 %**	1 %			
06 Western FM			1 %	2 %	3 %	**85 %**	5 %	3 %	1 %
07 Northern Troms			1 %	1 %	2 %	22 %	**54 %**	18 %	3 %
08 SouthernTroms						6 %	10 %	**80 %**	4 %
09 Nordland						1 %	8 %	20 %	**70 %**

correctly assigned

	Reporting Group to which they were assigned to								
Reporting Group of origin	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 Southern Troms	09 Nordland
01 Eastern B and WS	**98 %**								
02 Eastern Kola	2 %	**96 %**	2 %						
03 Western Kola		3 %	**90 %**	5 %	1 %		1 %		
04 Eastern FM		1 %	5 %	**90 %**	2 %	2 %	1 %	2 %	
05 Teno River			1 %	1 %	**93 %**	1 %	1 %		
06 Western FM			1 %	4 %	4 %	**87 %**	18 %	10 %	6 %
07 Northern Troms				1 %	1 %	9 %	**68 %**	23 %	12 %
08 SouthernTroms						1 %	7 %	**58 %**	9 %
09 Nordland							3 %	7 %	**72 %**

3.1.3.3 Applying cut-off values. Reporting groups

In some applications of individual assignment it is important to maximise the level of assignment accuracy. Since each individual is provided with a probability value for the assignment to river stocks we can define a cut-off value when to accept it. For example, using 90% as a cut-off probability, resulted in nearly 30% increase in the proportion of correctly assigned samples for Troms and Nordland reporting groups (Table 10). However this would come with cost of having to discard 44% of the samples assigned to these reporting groups. In some applications this might still be acceptable. On the other hand, for reporting groups or rivers with initially high correct assignment level, there is very little benefit in setting stringent rules for accepting an assignment. For example, using 90% cut-off for Western Kola reporting group would result only 5% increase in the correct assignment but with the cost of losing 19% of the samples assigned. Thus, there is a trade-off associated with using cut-off values which has to be considered.

Table 10. Assignments of test samples to reporting groups with varying cut-off values. A) test samples only; upper panel shows proportion of assigned samples and lower panel shows proportion of correctly assigned. B) Test samples analyzed with real mixture samples; upper panel shows proportion of assigned samples and lower panel shows proportion of correctly assigned.

test samples only	Reporting Group									
Cut-off	01 Eastern B&WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 SouthernTroms	Troms combined	09 Nordland
0	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
30	100 %	100 %	100 %	99 %	100 %	99 %	98 %	99 %	98 %	98 %
50	99 %	97 %	98 %	96 %	98 %	96 %	91 %	90 %	91 %	88 %
70	98 %	95 %	94 %	89 %	95 %	90 %	76 %	76 %	76 %	73 %
90	96 %	86 %	81 %	78 %	88 %	78 %	54 %	58 %	56 %	57 %

test samples only	Reporting Group									
Cut-off	01 Eastern B&WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 SouthernTroms	Troms combined	09 Nordland
0	98 %	95 %	94 %	91 %	94 %	88 %	67 %	62 %	80 %	66 %
30	98 %	95 %	94 %	92 %	94 %	88 %	68 %	62 %	80 %	67 %
50	98 %	96 %	95 %	93 %	95 %	89 %	70 %	66 %	82 %	73 %
70	99 %	98 %	96 %	96 %	97 %	92 %	75 %	72 %	85 %	80 %
90	100 %	98 %	98 %	98 %	98 %	95 %	81 %	79 %	89 %	91 %

b

with mixture	Reporting Group									
Cut-off	01 Eastern B&WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 SouthernTroms	Troms combined	09 Nordland
0	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	99 %	100 %
30	100 %	88 %	92 %	93 %	99 %	99 %	98 %	99 %	98 %	98 %
50	99 %	87 %	90 %	88 %	97 %	96 %	89 %	91 %	89 %	89 %
70	98 %	84 %	85 %	81 %	92 %	89 %	73 %	75 %	74 %	76 %
90	97 %	74 %	72 %	69 %	85 %	77 %	53 %	54 %	53 %	61 %

with mixture	Reporting Group									
Cut-off	01 Eastern B&WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 SouthernTroms	Troms combined	09 Nordland
0	98 %	96 %	90 %	90 %	93 %	87 %	68 %	58 %	78 %	72 %
30	98 %	96 %	91 %	90 %	94 %	87 %	69 %	59 %	79 %	73 %
50	98 %	97 %	92 %	92 %	95 %	88 %	72 %	62 %	81 %	78 %
70	99 %	98 %	94 %	94 %	96 %	91 %	76 %	67 %	84 %	84 %
90	99 %	99 %	96 %	98 %	98 %	94 %	83 %	78 %	89 %	93 %

3.1.3.4 River specific assessment of assignments to reporting groups

Proportions of correctly identified test samples from rivers belonging to reporting groups from 01 Eastern Barents and White Sea to 05 Teno River were high; only four out of 91 rivers had lower than 80% correct assignment to the reporting group of origin (Table 11). In the western Finnmark reporting group, there was a pattern of decreasing correct assignment from east to west: rivers in the eastern part had higher correct assignment than those closer to Troms reporting group. In Troms and Nordland reporting groups, the proportions of correct assignment were substantially lower and only for eight out of 37 rivers more than 80% of test samples were identified to correct reporting group of origin. In Northern Troms, populations close to Western Finnmark had lower correct assignments increasing towards south. While average correct assignment of northern Troms test samples was only 58% to the reporting group, 75% of samples were identified as being of Troms origin. Similarly, 90% of southern Troms test samples were identified as being of Troms origin.

Table 11. Assignments of test samples to reporting groups.

	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 Southern Troms	09 Nordland
01 Eastern B and WS	**99 %**								
001Pechora	100 %								
002Mezen-Pizhma	100 %								
003Kuloy-Soyana	100 %								
004Severnaya D.	100 %								
005Letnaya Zolotitsa	100 %								
006Onega	94 %	4 %	2 %						
007Gridina	98 %		2 %						
008Keret	96 %	4 %							
02 Eastern Kola		**97 %**							
009Kovda		100 %							
010Kanda	1 %	93 %	6 %						
011Kolvitsa	1 %	89 %	9 %	1 %					
012Porya		98 %	2 %						
013Pila	5 %	94 %	1 %						
014Umba	2 %	96 %	2 %						
015Kuzreka	1 %	99 %							
016Khlebnaya		100 %							
017Olenitsa		100 %							
018Varzuga		100 %							
019Kitsa		100 %							
020Indera		100 %							
021Chavanga		98 %	2 %						
022Strelna		100 %							
023Yugin		100 %							
024Chapoma		99 %		1 %					
025Pyalitsa	1 %	99 %							
026Bolshaya K.		100 %							
027Pulonga		100 %							
028Likhodeevka		100 %							
029Babya		98 %	2 %						
030Ponoi		96 %	4 %	0 %					
031Kachkovka		78 %	21 %	1 %					
03 Western Kola			**93 %**						
032Lumbovka		14 %	85 %		1 %				
033Iokanga		3 %	97 %						
034Drozdovka		1 %	95 %	1 %			2 %		
035Penka		1 %	98 %	1 %					
036Varzina			92 %	7 %	1 %				
037Sidorovka		7 %	87 %	3 %			2 %		
038Vostochnaya		3 %	93 %	4 %	1 %				
039Kharlovka		5 %	89 %	5 %					
040Zolotaya		1 %	97 %		1 %	1 %			
041Rynda		1 %	96 %	1 %	1 %				
042Olenka		10 %	85 %	3 %	1 %				
043Zarubikha		7 %	89 %	3 %	1 %				
044Orlovka			99 %	1 %					
045Yarnishnaya		1 %	97 %	1 %		1 %			
046Dolgaya		3 %	91 %	6 %	1 %				
047Klimkovka		3 %	92 %	3 %	2 %				
048Tipunkova		6 %	90 %	3 %					
049Zarbikha Kildin		3 %	95 %	1 %	1 %				
050Vaenga			100 %						
051Kola		2 %	95 %	3 %	1 %				
052Tuloma		1 %	98 %						
053Kulonga			100 %						
054Ura			93 %	6 %	1 %				
055Bolshaya Z. Litsa		1 %	95 %	3 %	1 %				
056Titovka		3 %	85 %	6 %	2 %		4 %		
057Pyave		2 %	94 %	4 %					
058Pechenga		2 %	92 %	2 %	1 %		3 %		

Table 11. (cont.)

	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 Southern Troms	09 Nordland
04 Eastern FM				**86 %**					
059Grense Jakobselv		1 %	24 %	72 %	2 %	1 %			
060Tårnelv		1 %	6 %	87 %	4 %	1 %			
061Karpelv		1 %	7 %	85 %	3 %	3 %			
062Sandneselva K.		2 %	8 %	89 %	2 %				
063Munkelv			9 %	87 %	1 %	3 %			
064Neiden			12 %	81 %	6 %	1 %			
065Klokkarelv		4 %	4 %	84 %	4 %	4 %			
066Nyelva		3 %	10 %	80 %	2 %	4 %			
067Vesterelva		3 %	5 %	85 %	4 %	1 %			
068Bergebyelva			1 %	95 %	1 %	3 %			
069Vestre Jakobselv		1 %	3 %	78 %	1 %	14 %			
070Storelva Vadso			10 %	88 %		2 %			
071Skallelva				91 %		7 %			
072Komagelva			2 %	93 %	1 %	4 %			
073Syltefjordelva			2 %	91 %	2 %	4 %			
074Storelva Batsfj.			6 %	79 %	2 %	11 %			1 %
075Kongsfjordelva			1 %	92 %	1 %	4 %			
076Storelva Ber		1 %		89 %	1 %	9 %			
05 Teno River					**98 %**				
077Iesjohka			1 %	1 %	98 %		1 %		
078Karasjoki					99 %	1 %			
079Inari			2 %		96 %	1 %			
080Aku			2 %		98 %				
081Valjok					100 %				
082Galddas					100 %				
083Karigas					99 %	1 %			
084Kevo					100 %				
085Tsars					100 %				
086Utsjoki			1 %		98 %	1 %			
087laksjohka					100 %				
088Leva			1 %	3 %	94 %	1 %		1 %	
089Maske					100 %				
090TenoMS				2 %	94 %	4 %			
091Vetsi					99 %	1 %			
06 Western FM						**86 %**			
092Langfjordelva			1 %	2 %	3 %	84 %	8 %	2 %	
093Risfjordelv			1 %	2 %		93 %	2 %	1 %	2 %
094Sandfjordelva				4 %		95 %		1 %	
095Futelva						100 %			
096Mehamnelva						99 %	1 %		
097Ifjordelva						90 %	7 %	3 %	
098Suosjohka					1 %	96 %	2 %	1 %	
099Storelva Laksefj.			1 %	2 %	1 %	74 %	11 %	8 %	2 %
100Veidneselva			1 %	1 %	1 %	86 %	10 %	2 %	
101Lille Porsangere.						96 %		4 %	
102Tømmervikelva			1 %	5 %	1 %	85 %	5 %	3 %	1 %
103Børselva				1 %	2 %	83 %	10 %	4 %	1 %
104Lakselva Pors				1 %		98 %	1 %		
105Stabburselva				1 %		98 %	1 %		
106Ytre Billefjord				3 %	3 %	87 %	3 %	3 %	1 %
107Oldervikelva						86 %	7 %	5 %	3 %
108Smørfjordelva		2 %	1 %	5 %	1 %	81 %	5 %	1 %	3 %
109Snefjordelva				3 %		72 %	9 %	9 %	7 %
110Russelv				4 %	11 %	64 %	19 %	1 %	
111Repparfjordelva				3 %	9 %	80 %	3 %	3 %	1 %
112Kvalsundelva		1 %		2 %	1 %	92 %	2 %	1 %	1 %
113Lakselva_Kviby				4 %	10 %	68 %	13 %	5 %	
114Alta				1 %		97 %	2 %		
115Botneelva				1 %	3 %	76 %	19 %	2 %	
116Mattiselva					10 %	86 %	2 %	2 %	
117Halselva					8 %	85 %	6 %		
118Bognelva						64 %	25 %	9 %	2 %

Table 11. (cont.)

	01 Eastern B and WS	02 Eastern Kola	03 Western Kola	04 Eastern FM	05 Teno River	06 Western FM	07 Northern Troms	08 Southern Troms	09 Nordland
07 Northern Troms							**58 %**		
119Burfjordelva					2 %	31 %	50 %	16 %	1 %
120Badderelva					7 %	51 %	24 %	18 %	1 %
121Kvænangselva			1 %		11 %	54 %	28 %	6 %	
122Oksfjordvassdr.						10 %	65 %	19 %	6 %
123Reisa			1 %	1 %	3 %	8 %	85 %	3 %	
124Rotsundelva						10 %	66 %	22 %	2 %
125Kåfjordelva						3 %	61 %	27 %	9 %
126Manndalselva						5 %	70 %	23 %	1 %
127Signaldalselva						8 %	65 %	20 %	6 %
128Breivikelva						15 %	59 %	18 %	8 %
129Oldervikelva						57 %	41 %	2 %	
130Skittenelva			1 %			19 %	60 %	11 %	9 %
131Tønsvikelva						8 %	76 %	11 %	5 %
132Nordkjoselva						7 %	62 %	24 %	7 %
08 SouthernTroms								**76 %**	
133Lakselva						4 %	4 %	87 %	5 %
134Mårelva							6 %	89 %	6 %
135Målselva						8 %	10 %	78 %	4 %
136Rossfordvassdr.						2 %	2 %	95 %	1 %
137Lysbotnvassdr.						5 %	18 %	66 %	11 %
138Laukhelle						9 %	13 %	74 %	4 %
139Skøelv						5 %	28 %	59 %	8 %
140Tennelvvassdr.								100 %	
141Åndervassdraget							14 %	81 %	5 %
142Løksebotnvassdr.						6 %	18 %	67 %	9 %
143Salangsvassdr.						10 %	39 %	44 %	7 %
09 Nordland									**70 %**
144Roksdalsvassdr.							2 %	2 %	96 %
145Åseelva							10 %	19 %	71 %
146Alsvågvassdr.							4 %	4 %	92 %
147Buksnesvassdr.t							10 %	17 %	73 %
148Gårdselv						2 %	4 %	12 %	82 %
149Kjerringnesvassdr.							10 %	21 %	69 %
150Holmstadelva							15 %	15 %	70 %
151Langvatnvassdr.							19 %	19 %	63 %
152Tårstadvassd.							23 %		77 %
153Forsåvassdraget							9 %	24 %	68 %
154Saltdalselva						2 %	6 %	32 %	60 %
155Beiarelva							37 %	46 %	17 %

3.1.3.5 Assessment of assignments of test samples to river

On average, 64% of test samples were correctly assigned to the river of origin. However, there was large variation in the assignment success among rivers. While for 32 rivers more than 90% of test samples were correctly identified, all test samples of Storelva (Båtsfjord) and Åseelva river were assigned incorrectly. These two rivers are small, and as mentioned above the allele frequencies in such rivers may be subject to random fluctuations between years. The Storelva sample also contained the highest number of siblings among the baseline samples (68%). It is unlikely that the observed lack of correct assignment of individuals from these rivers will have any significant effect on the final GSI results of coastal fishery samples as the expected contribution to the fishery is very small.

In general, the highest assignment success was found for the eastern Barents and White Sea populations where correct assignment rate to the river of origin was 97% on average (Fig. 19 a). Test samples from rivers in Teno river reporting group also had high assignment success to their population of origin; 89% on average were identified. Again, northern Troms had lowest assignment success rate with only 36% of samples correctly identified from each river. However, for the large Reisa River the assignment success was relatively high, 82%.

From the point of interpreting assignment results of the coastal fishery samples, the proportion of correct assignments of all assigned samples to river is more important. On average 69% of the samples assigned to each river were correct and for nearly 50% of the rivers (73/155) correct assignment was 80% or higher (Fig. 19 b).

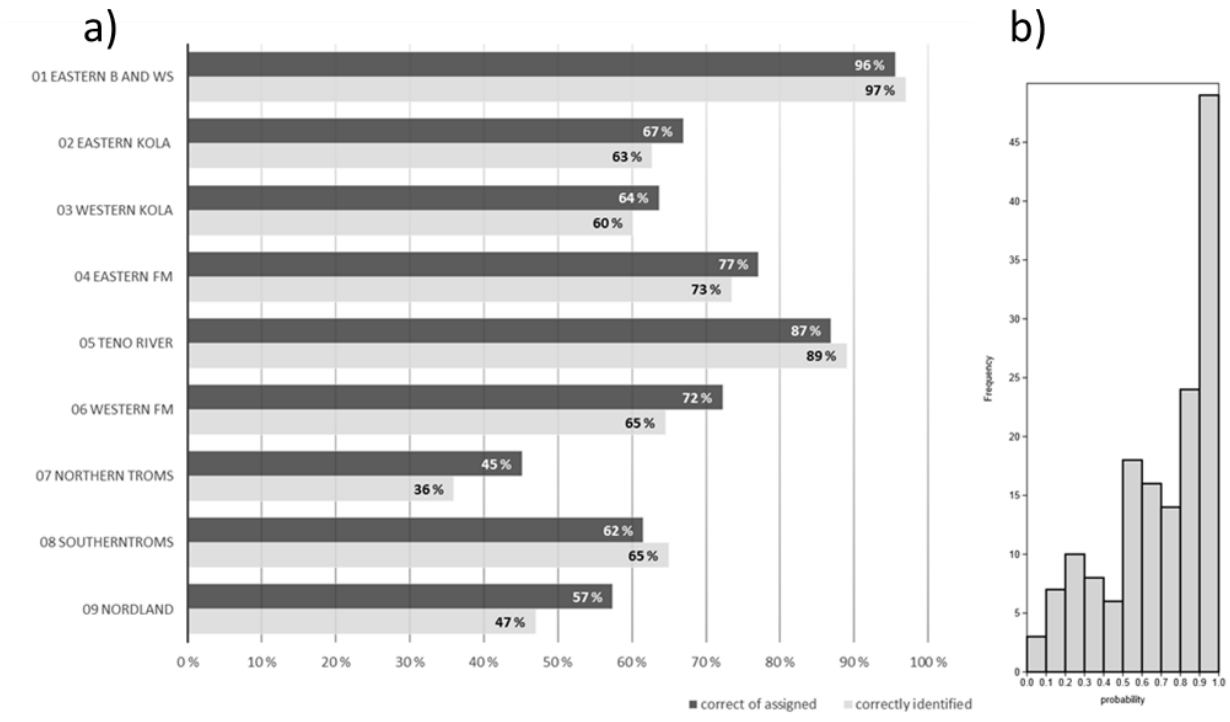


Figure 19. a) Overall assignment success of test samples to the river of origin within reporting groups. Dark grey – proportion correct of all samples assigned to a river. Light grey – proportion of samples correctly assigned to river from which they originated. **b)** Number of rivers correctly assigned.

3.1.4 Assignments to River

3.1.4.1 Rivers of Eastern Barents and White Sea reporting Group

Assignment success to rivers was generally excellent for the test samples in the eastern Barents and White Sea reporting group. Only Keret and Severnaya-Dvina rivers showed small levels of upward bias in the total assignments when test samples were analysed with real fishery samples (Fig. 20).

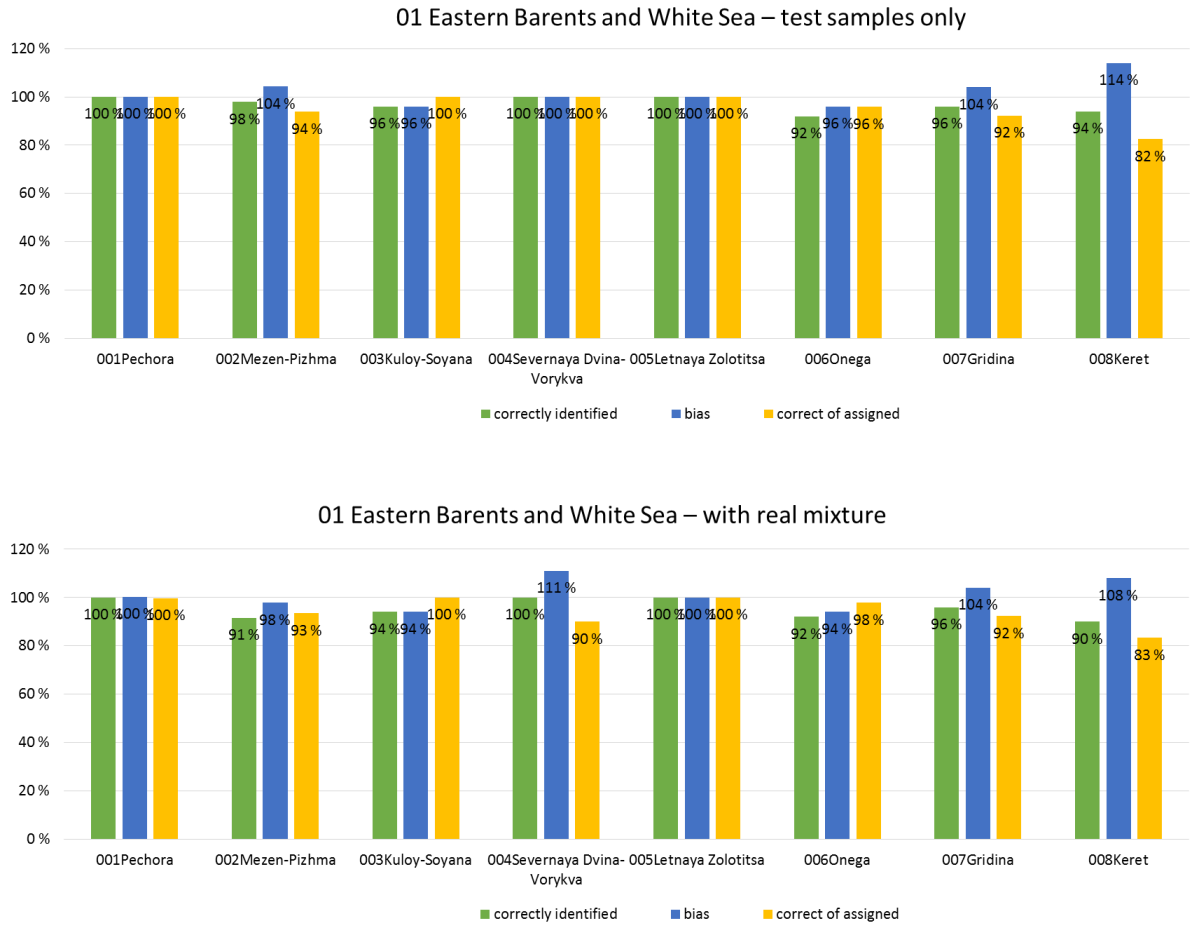


Figure 20. Assignment success of test samples to rivers of the Eastern Barents and White Sea reporting group.

3.1.4.2 Rivers of Eastern Kola reporting group

Assignment success for the rivers in the eastern Kola reporting group was binary. Rivers from Kovda to Olenitsa showed excellent assignment success while rivers from Varzuga to Kachkovka were generally poorly distinguished. The only exceptions were Ponoi and Indera rivers for which there was no bias in the total assignment and where test samples were moderately well identified. Assignments to Strelna river were largely incorrect and inflated. The largest contributions to mis-assignments to Strelna were from neighbouring rivers (23-30), but not from 26 (Bolshaya Kumzhevaya). In addition to Strelna, Kitsa, Chapoma Pulonga and Babya assignments were inflated and their contribution to total catch should be treated with caution.

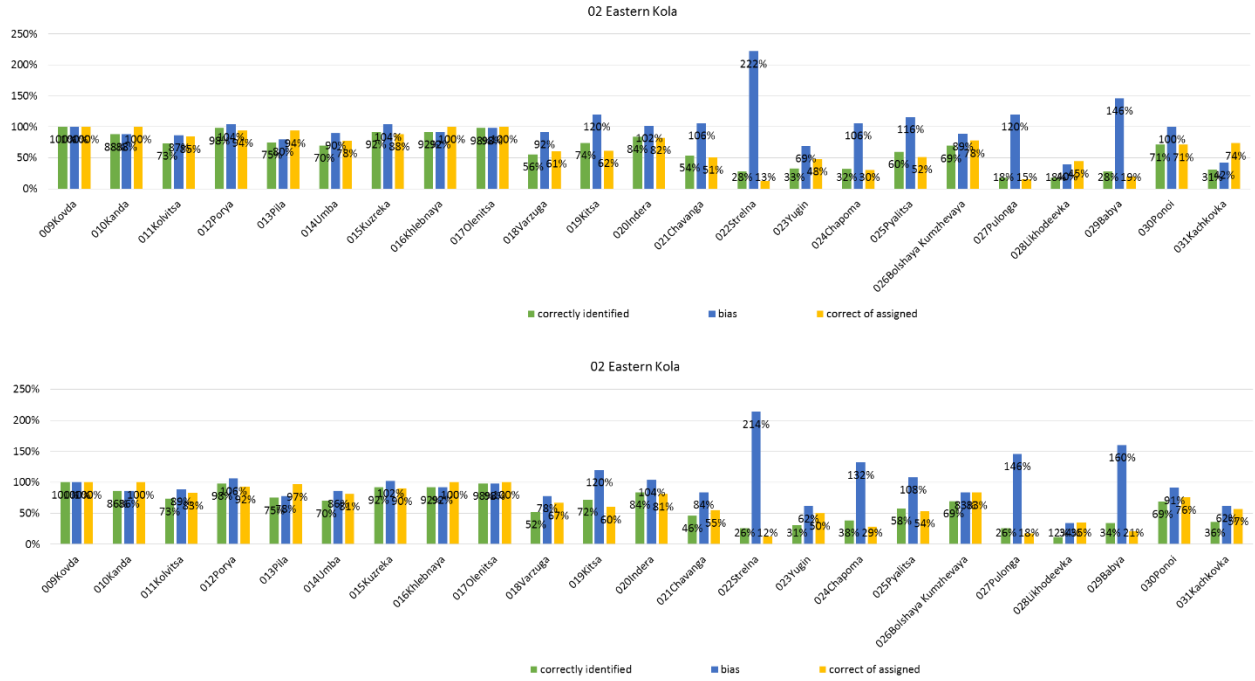


Figure 21. Assignment success of test samples to rivers of the Eastern Kola reporting group.

3.1.4.3 Rivers of Western Kola reporting group

Assignment results from analysis with test samples only indicated that rivers in the eastern part (49 -58) of the reporting group region were generally well identified. However, results from analysis with real fishery samples with unknown stock composition revealed potential for significant overestimation of Kola, Ura and Bolshaya Zapadnaya Litsa river contribution (Fig. 22). Kola River assignments were inflated by a factor of three. While the majority of the mis-assignmetns were from the neighbouring Tuloma river system, 16 out of 27 rivers in the reporting group contributed to the mis-assignments. Furthermore, 9% of test samples of Karpelva, Munkelva and Klokkerelva in eastern Finnmark were also assigned to Kola River. Patterns of mis-assignments to Ura and Bolshaya Zapadnaya Litsa rivers were similar. Test samples from 21 and 19 rivers from western Kola reporting group contributed to mis-assignmetns to Ura and Bolshaya Zapadnaya Litsa, respectively. Furthermore, 14% of test samples from rivers 59 Grense Jakobselva to 65 Klokkarelva in eastern Finnmark reporting group were mis-assigned to either Ura or Bolshaya Zapadnaya Litsa. The marked difference between the results of test samples only, and results with mixture samples analysis indicates that in cases where these rivers significantly contribute to fishery sample their contribution may be significantly over-estimated and should be treated with caution.

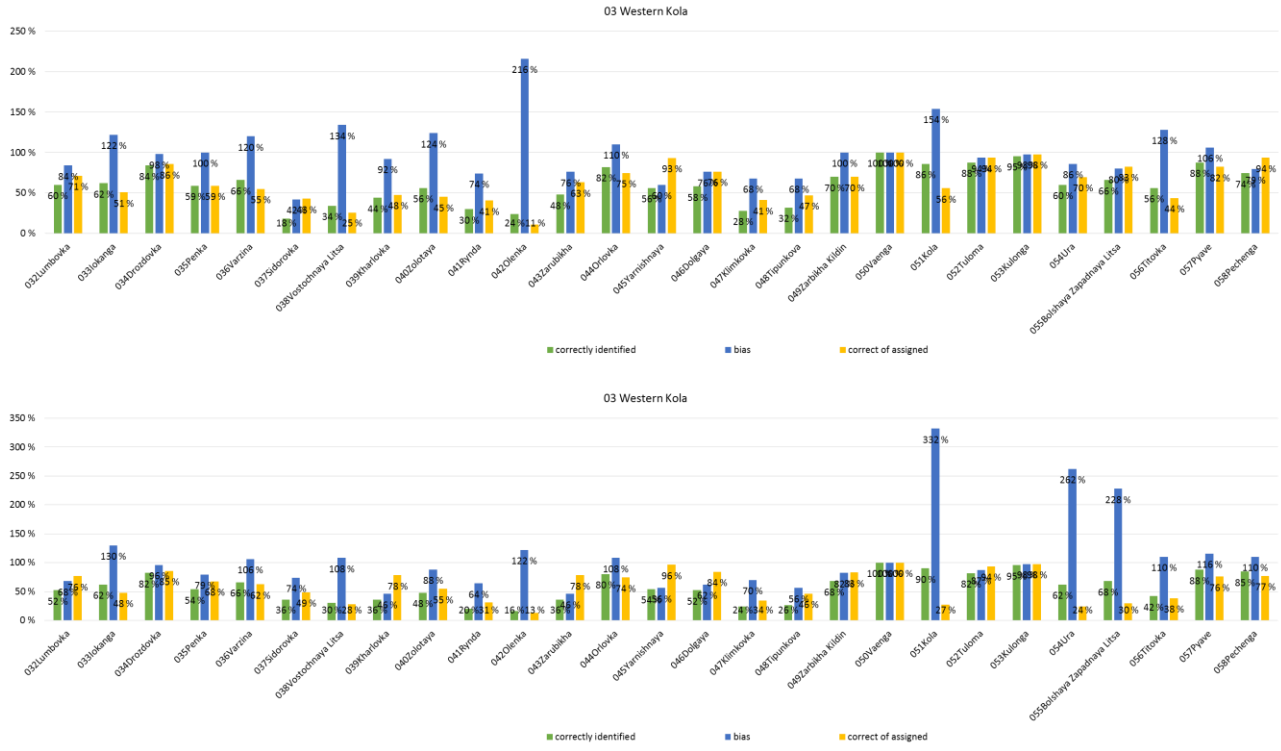


Figure 22. Assignment success of test samples to rivers of the Western Kola reporting group.

3.1.4.4. Rivers of Eastern Finnmark reporting group

Assignment success of test samples from rivers of Eastern Finnmark were generally very good. The only exception was Storelva in Båtsjord for which all test samples were incorrectly assigned to either Vestre Jakobselva or Syltefjordelva (Fig. 23). Again, this is a small population where precise definition of a genetic profile is difficult. Accordingly, total assignments of the two rivers were over-estimated. There were no other major patterns of mis-assignments within the reporting group. However, on average 9.6% of test samples from rivers 59 to 74 were assigned to Western Kola reporting group. The largest mis-assignment to western Kola was from Grense Jakobselva (34%). Nevertheless, the majority of rivers in the eastern Finnmark reporting group could be well identified.

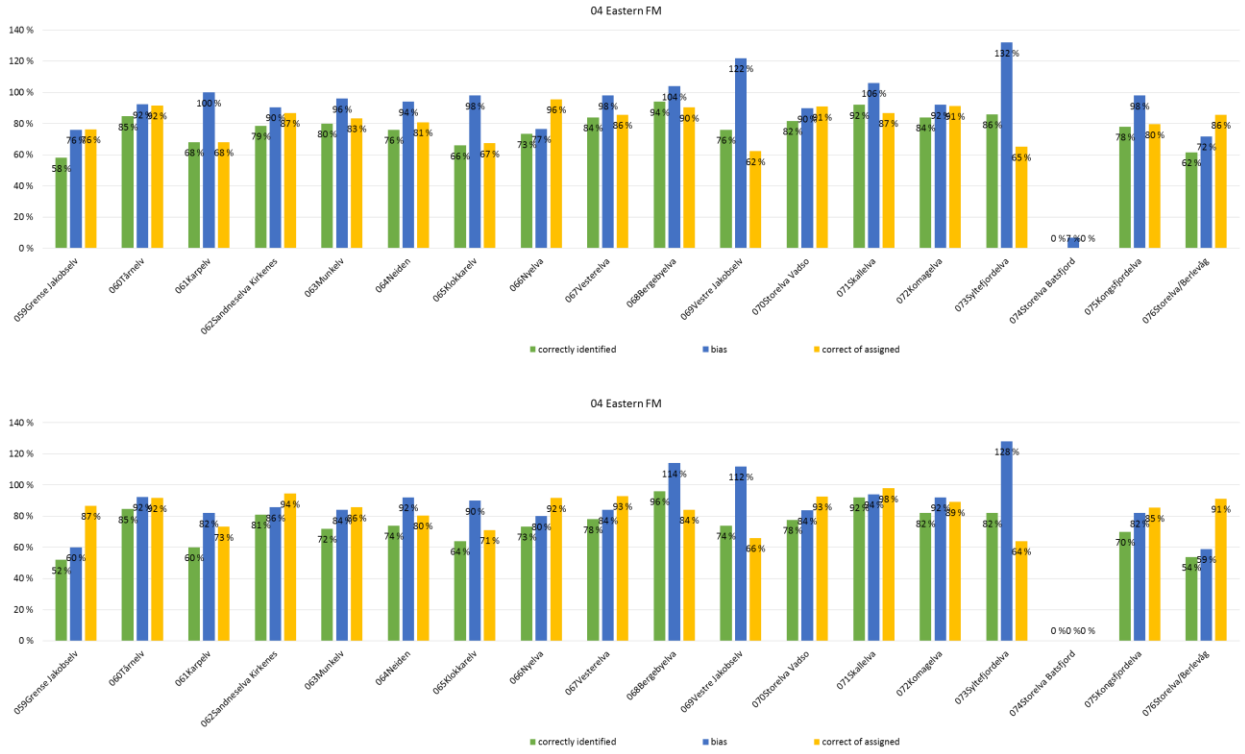


Figure 23. Assignment success of test samples to rivers of the Eastern Finnmark reporting group.

3.1.4.5 Rivers of Teno reporting group

Generally, the assignment success of test samples from rivers in Teno reporting group was very good (Fig. 24). The lowest assignment success was observed for Aku river for which test samples were mis-assigned to river Karigasjoki (25%, 27%) and Teno mainstem (16%, 11%). Albeit river Iesjohka test samples were nearly all identified (96%, 98%) the total assignment was over-estimated. In addition to the contribution of Karasjohka (19%) and Teno mainstem test samples (19%) to mis-assignments of river Iesjohka, nearly 62% originated from western and eastern Finnmark. In addition to river Iesjohka, Teno mainstem and river Inarijoki were also over-estimated, partly due to mis-assignments from western and eastern Finnmark river test samples.

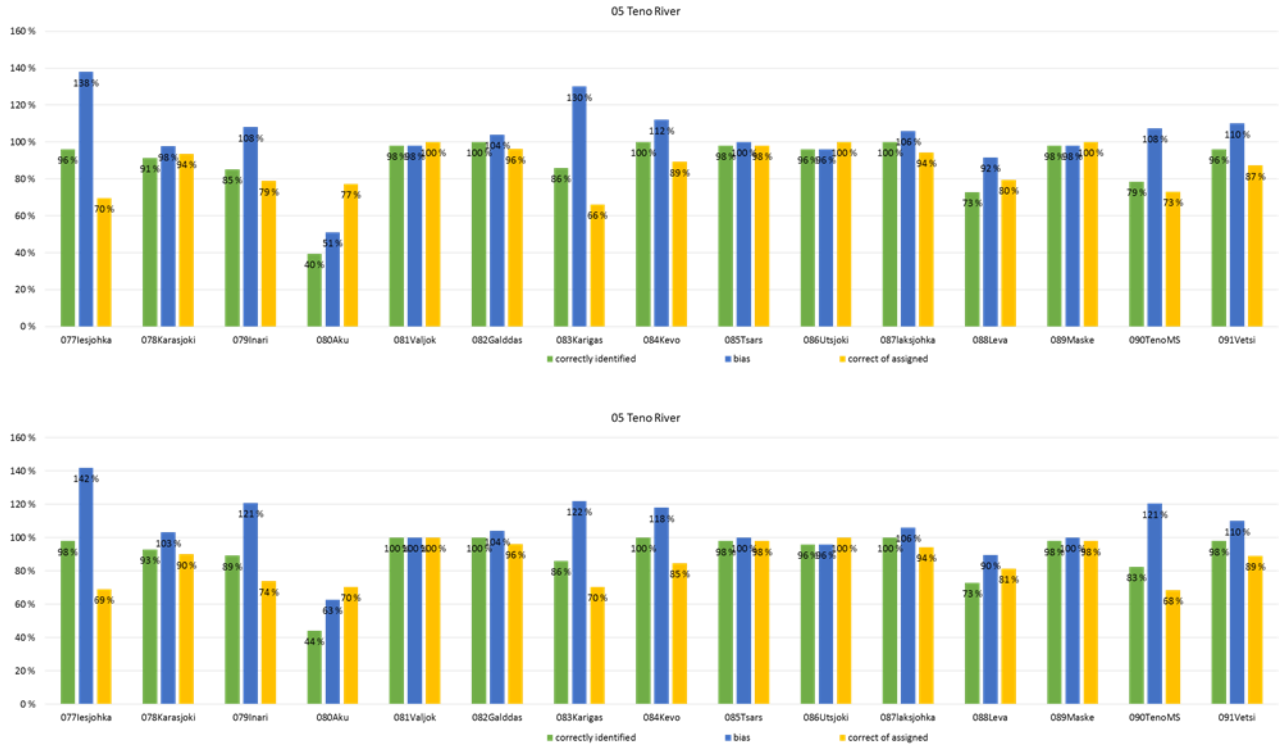


Figure 24. Assignment success of test samples to rivers of the Teno reporting group.

3.1.4.6 Rivers of Western Finnmark reporting group

Apart from Langfjordelva, rivers in the eastern part of the western Finnmark reporting group were identified and the proportions estimated without bias (Fig. 25). Test samples from Storelva of Laksefjord and Repparfjordelva were identified with poor success (49%-58%) yet their total assignments were over-estimated. In analysis with real mixture samples only 16% percent of samples assigned to Repparfjordelva were correct while the total number of samples assigned was 3.6 times the expected. Such pattern of assignment resulted in extensive over-estimation of Repparfjordelva contribution to the mixed stock fishery sample.

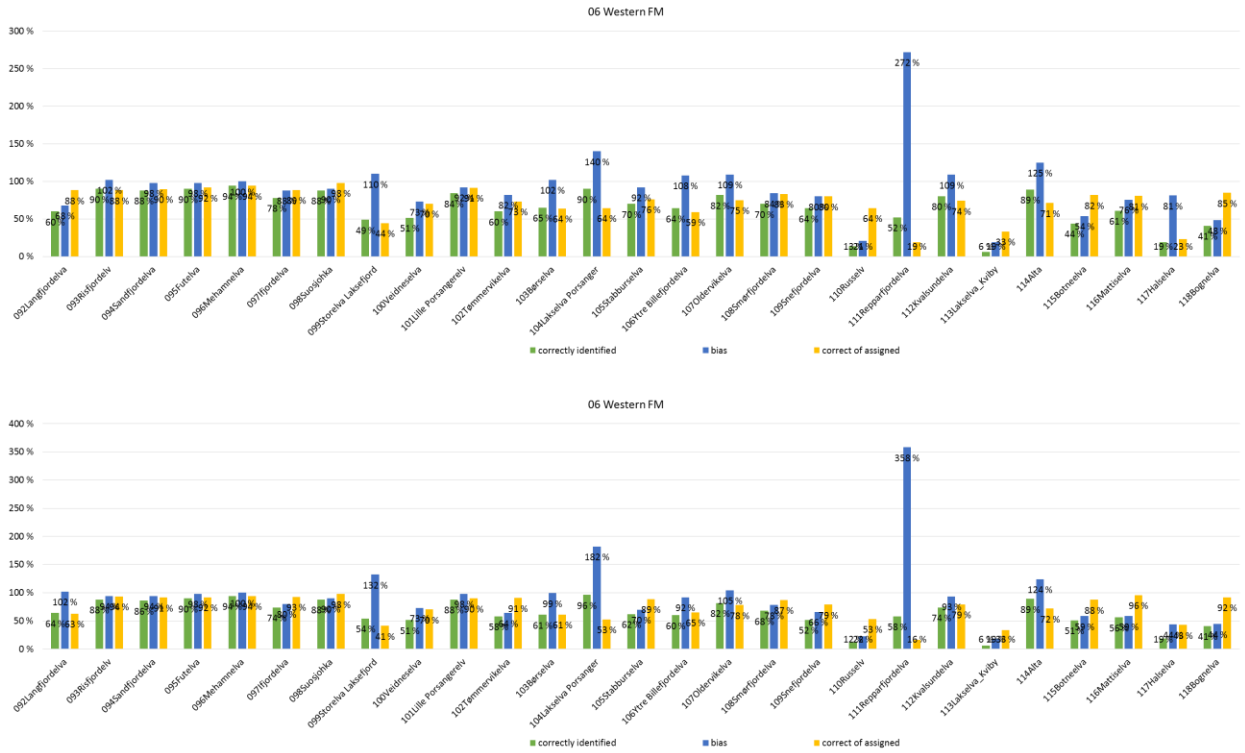


Figure 25. Assignment success of test samples to rivers of the Western Finnmark reporting group.

3.1.4.7 Rivers of Northern Troms reporting group

In general the assignments at the river level within Northern Troms were poor (Fig. 26). The only exception was the largest stock of region, River Reisa, where up to 85% of samples were correctly identified and contribution estimated without much bias. In addition to Reisa River, test samples assigned to Manndalselva and Oldervikelva were also largely correct. Test samples from Badderelva were mostly assigned incorrectly to Repparfjordelva in Western Finnmark contributing to its large over-estimation. Nearly 17% of northern Troms test samples were mis-assigned to Målselva, Skoelva and Salangvassdraget rivers in Southern Troms.

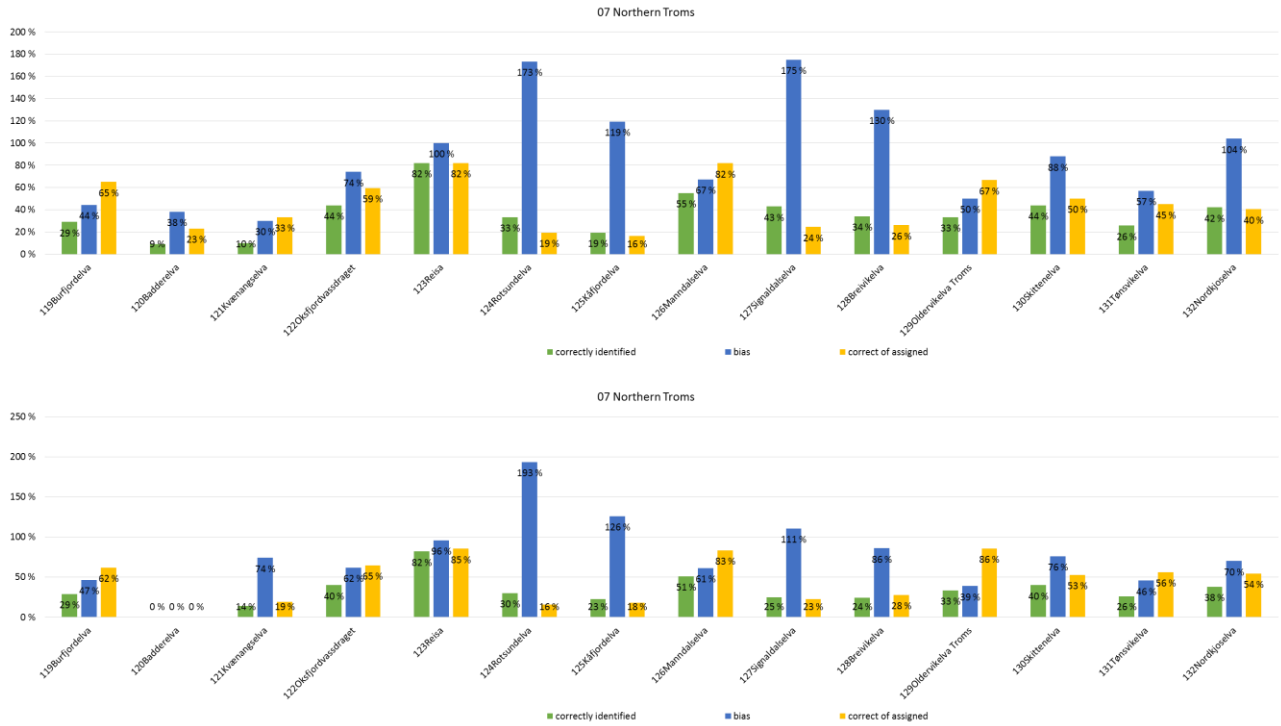


Figure 26. Assignment success of test samples to rivers of the Northern Troms reporting group.

3.1.4.8 Rivers in Southern Troms

Four out of 11 rivers in southern Troms were identified with very high assignment succcess and their conrtributions estimated without large bias (Fig. 27). Samples assigned to rivers Lysbotnvassdraget, Skoelva and Salangvasdraget were largely incorrect. While 74% of test samples from the large Målselva stock were correctly identified its contribution to the mixed stock fishery was over-estimated due to small proportion of mis-assignments from 10 out of 14 northern Troms river test samples.

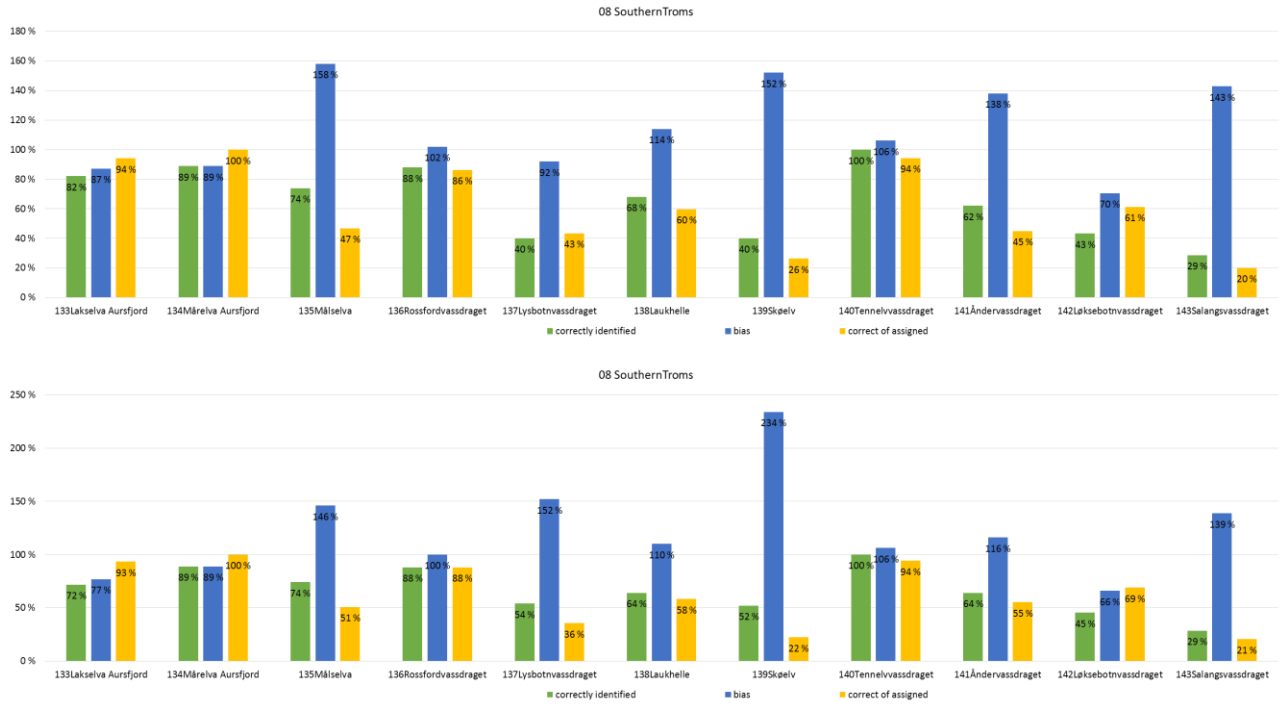


Figure 27. Assignment success of test samples to rivers of the Southern Troms reporting group.

3.1.4.9 Rivers of Nordland

Test samples from Åselva, Kjerringnesvassdraget and Beiarelva rivers could not be identified (Fig. 28). While more than 50% of test samples from Gårdselva river were identified, the majority of samples assigned were incorrect resulting in large over-estimation of its contribution to mixed stock sample. Excluding the four rivers, test samples from other Nordland rivers were relatively well identified to their stock of with 75% mean correct assignment rate.

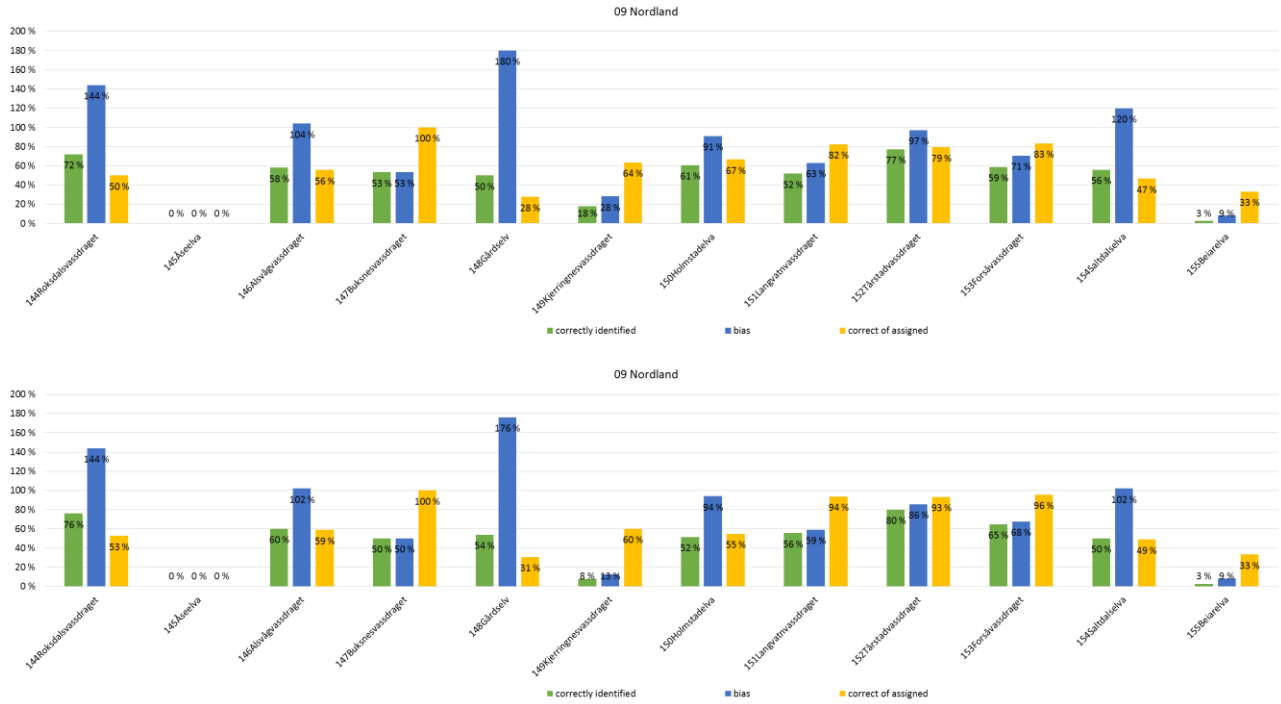


Figure 28. Assignment success of test samples to rivers of the Southern Troms reporting group.

3.1.4.9 Applying cut-off values. Rivers

Using cut-off values to filter out potential mis-assignments at the river level appeared largely effective, but with cost of also filtering out correct assignments. However, trade-off between gain in the proportion of correctly assigned and and excluding true positives varied among stocks, and no single cut-off value appeared superior. For example, in Pechora no increase in the proportion of correctly assigned could be achieved since all were correct, but using a stringent cut-off value would result in excluding up to 25% of the samples. On the other hand, when there is a strong genetic signal, high genetic divergence, in the assigned samples, e.g. River Kovda, setting a cut-off value did not have any effect on the assignments. However, for the majority of the rivers, there was a trade-off. For example in 011Kolvitsa, the proportion of correctly assigned samples increased from 82.5% to 100% using a value 0.7 as cut-off. Such significant increase was achieved with the cost of 15% decrease in the number of identified samples.

Despite the fact that the proportion of correct assingments to river increased by applying more stringent cut-off value, relative stock proportions in the mixture samples became more biased (Fig. 29). Over all stocks, there was no significant difference in mean bias (28% ± 32%) when cut-off was 0.5 or less. Using cut-off value 0.9 to assign samples to populations followed by estimation of their relative stock contribution to mixed stock fishery would result on average 45% (± 29%) bias. In general, the relative contribution of populations showing high assignment success would be biased upwards. For example, test samples from populations of the Eastern Barents and White Sea were always identified with high probability of belonging, and even very stringent cut-off value did not result in filtering them out. At the same time, test samples from other populations in the mixture were excluded and the relative contribution of Eastern Barents and White Sea rivers to mixture appeared elevated.

Finally, stock specific cut-off values were determined by maximising the relationship between the proportion of correctly identified and the proportion of correctly assigned test samples. Applying such cut-off values resulted in an increase of 11.1% in the mean of correctly assigned test samples (75.5% vs 68.8%; test samples with mixture) with the cost of excluding 12.2% of test samples. Assignment success with stock specific cut-off values was higher than with applying uniform cut-off 0.5 (73.8% correctly assigned; test samples with mixture) and allowing 5% more samples to be identified (87.8% vs. 83.6%). Alternative stock specific cut-off values can be applied balancing between the proportion of identified samples and the proportion of correcty assigned samples.

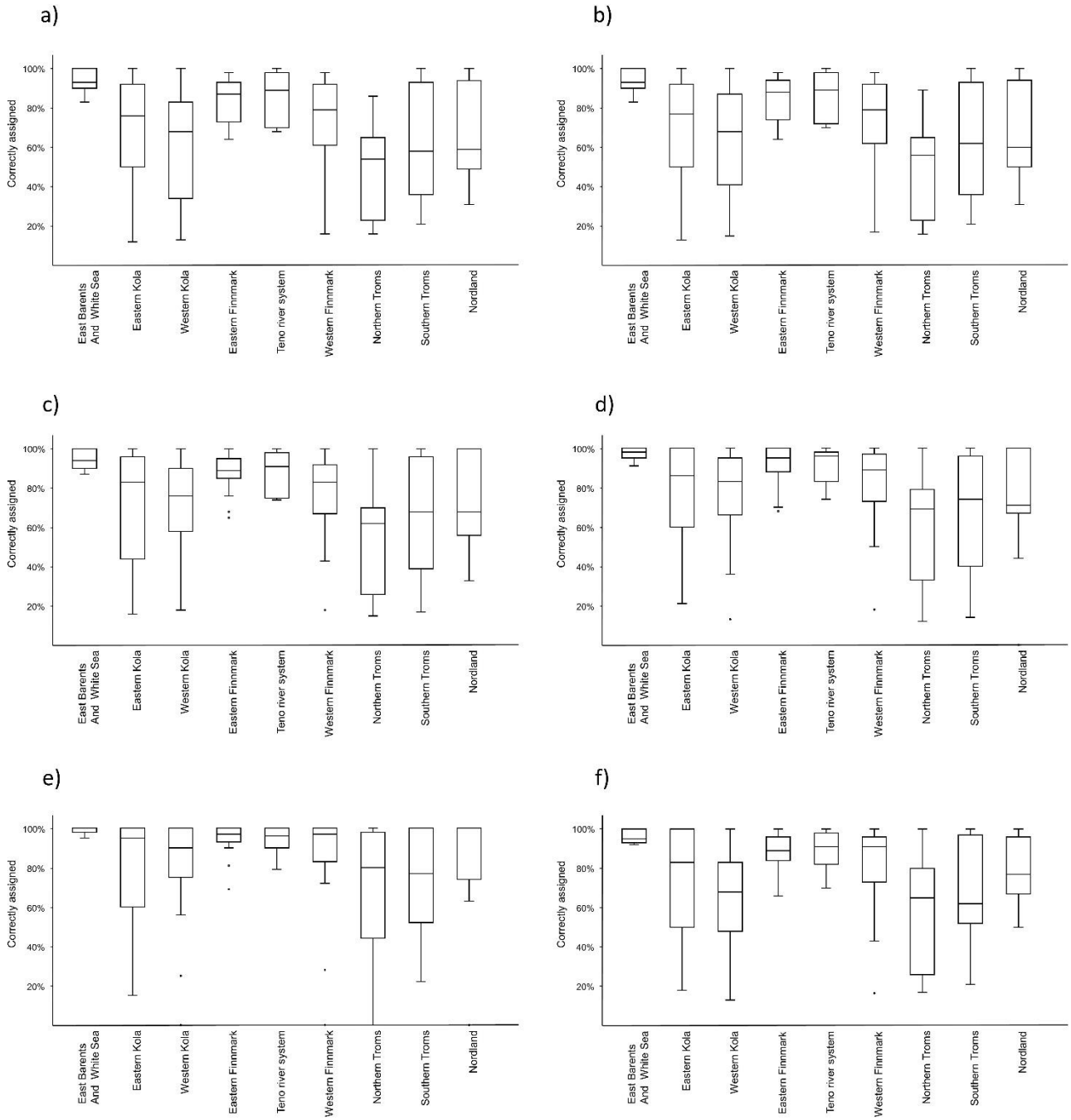


Figure 29. Correct assignments to populations for each reporting group using different cut-off values to accept or reject an individual. a) No cut-off, all data included, b) cut-off 0.30 c) cut-off 0.50 d) cut-off 0.70 e) cut-off 0.90 f) stock specific cut-off values. See text for details.

There was a positive, but non-linear relationship between the assignment success and the genetic divergence of rivers. Test samples from genetically highly differentiated rivers (population specific F_{ST} > 0.06) were always identified with high success (> 78%) When rivers with high divergence (pairwise Fst within RG > 0.06) and large enough sample sizes (>100) were excluded from the data, genetic divergence ($F = 125.3$; $P < 0.0001$) and sample size ($F = 30.7$; $P < 0.0001$) both had a significant contribution to variation in assingment success ($r^2 = 0.57$, $F = 78$, $P < 0.0001$; d.f. 2,122) (see also Fig 31, rivers where population specific F_{ST} > 0.045 and sample size > 100 not included).

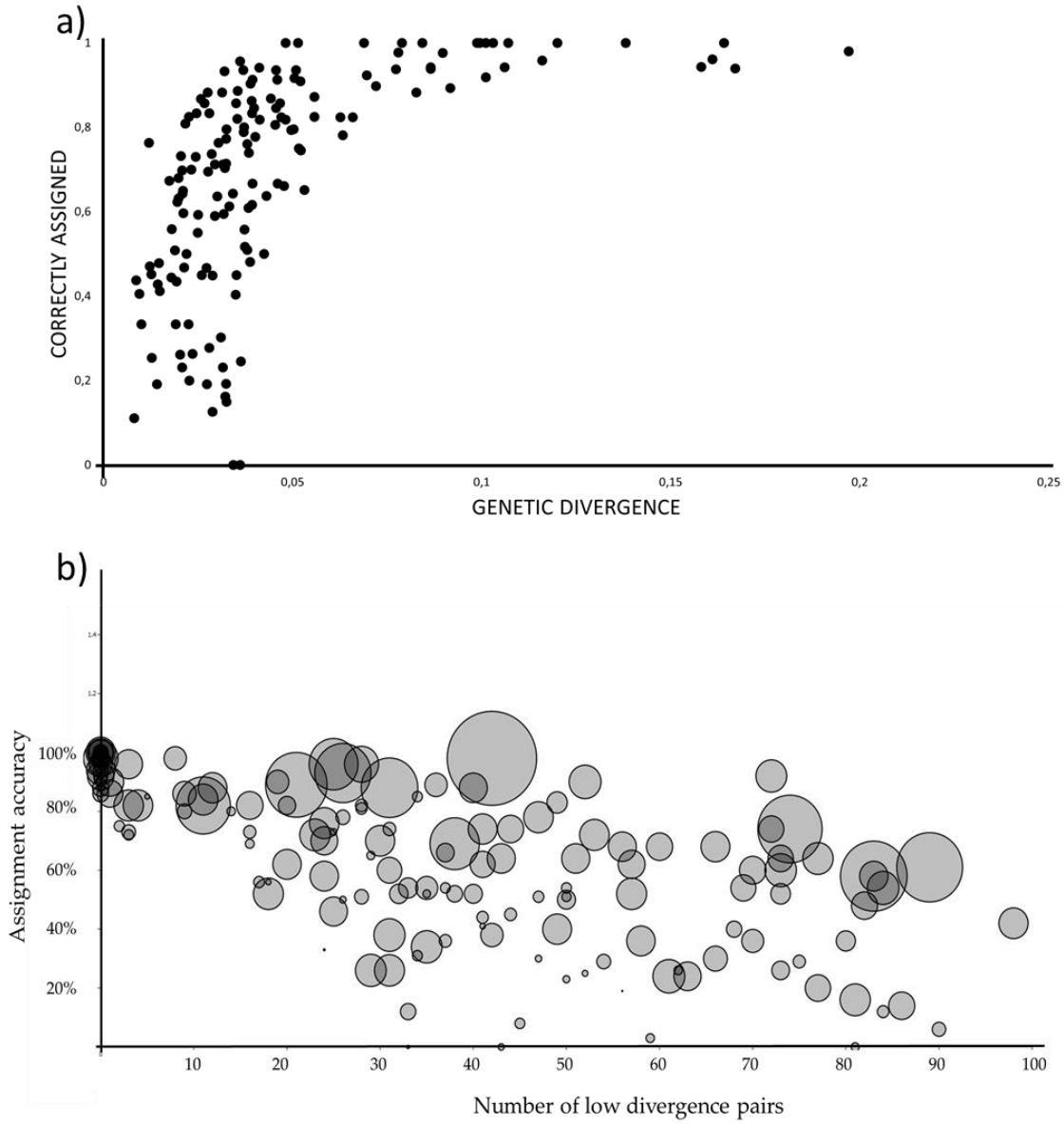


Figure 30. Relationships between **a)** assignment success and genetic divergence and **b)**. sample size (bubble size), genetic divergence and assignment accuracy.

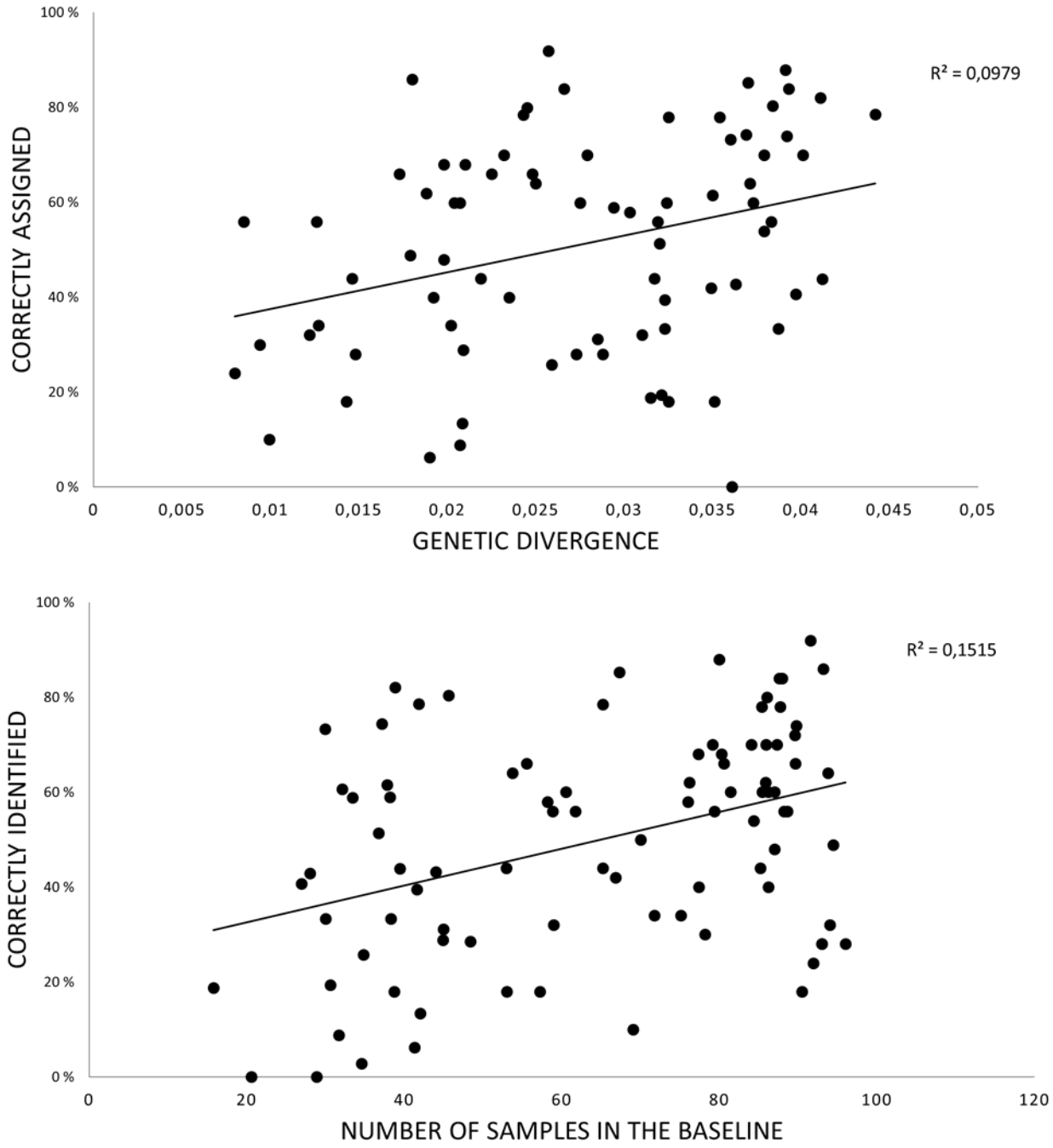


Figure 31. Relationships between assignmnet success and genetic divergence (upper plot), and number of samples in the baseline (lower plot). Rivers where population specific F_{ST} > 0.045 and sample size > 100 not were excluded.

3.2 GSI of Carlin tagged salmon included in the coastal mixed stock catch samples

Among the 20976 analysed coastal fishery samples, 33 salmon were found to be marked with external tags carrying history of their release site and potential stock of origin. According to recovered tag information 27 salmon were of Alta river origin released as smolt in the estuary of river Alta or river Halselva (5), and one salmon was tagged as a maiden adult fish (age of 3+. 1+) below the lower Tuloma fish passage in July 2010. Records were not available for five tagged salmon.

Externally tagged salmon serve as blind test samples to allow assessment of the accuracy of GSI to river Alta. Altogether, 23 out of 27 were genetically identified as originating from the river Alta (85%, which is in good agreement with the proportion of correctly identified test samples (89%, no cut-off; Fig. 32). Out of the 4 miss-assignments, two had low probability of belonging to the assigned river (Klokkarelva 0.37; Skoelva 0.42), while the mean probability of belonging was 0.97. Salmon tagged as adult at lower Tuloma fish passage was genetically identified as River Kola stock. This assignment is likely correct, since the Kola river drains into the Kola bay just 10 km seaward from the catch and release site of the fish at the Lower Tuloma dam, below which there is only brackish waters of the Kola bay.

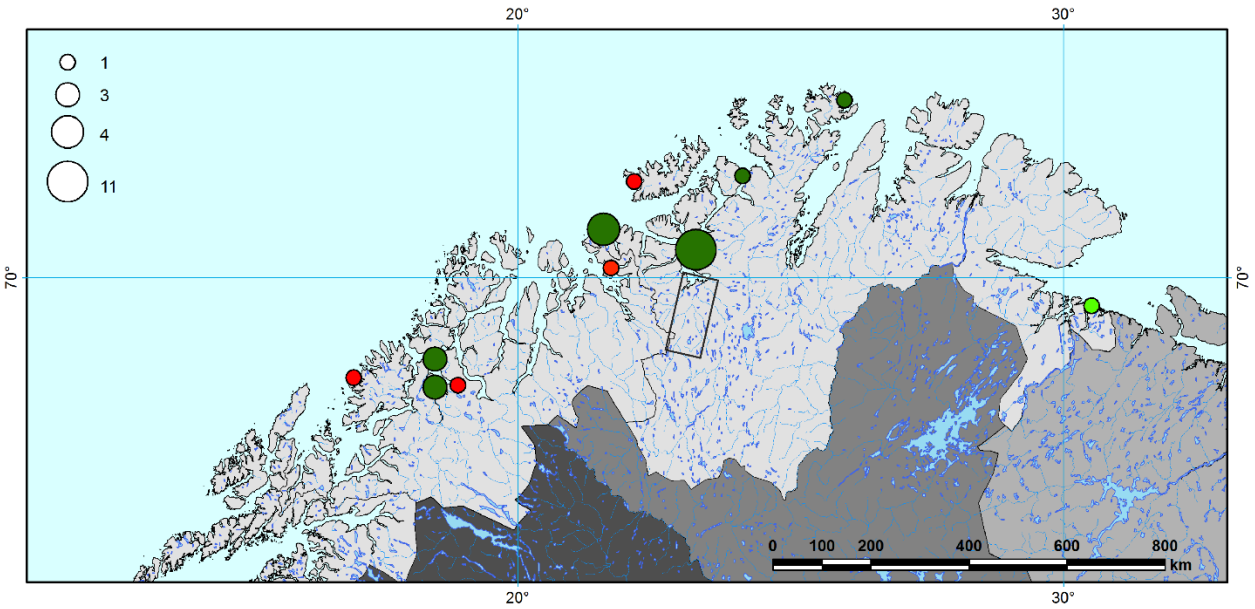


Figure 32. Distribution of externally tagged Alta fish. Green circles indicate fish genetically identified to have Alta origin, red circles – other origin. Light green circle indicate one externally tagged Tuloma fish, genetically identified as River Kola stock.

3.3 GSI of salmon from the Russian coastal mixed stock fisheries

According to our expectations, none of the salmon sampled from the White Sea fisheries were assigned to Norwegian rivers (Fig. 33). Furthermore, none of the salmon sampled in the White Sea had origins in the western Kola region. These results not only give strong support to reliability of the genetic stock identification to easternmost reporting groups, but also strongly indicate that salmon from the rivers draining to Barents Sea do not migrate into the White Sea.

While all salmon sampled in the Pechora estuary were assigned to Pechora (Fig. 34), salmon sampled from estuary of the River Big Eina originated equally from the rivers in the western Kola (46%) and eastern Finnmark (43%). In addition to salmon from nearby rivers, three had origins in the Iesjohka population in Teno River and unexpectedly two salmon had origins in the Nordkjoselva in northern Troms. Individual probability scores to rivers for three out the five non-local individuals were poor, but one of salmon had high (> 0.9) probability of belonging to Nordkjoselva and northern Troms reporting group (100%) giving strong support to the unexpected observation.

Nevertheless, river specific assignment patterns for the salmon sampled in the White Sea fisheries indicated a strong contribution of local rivers (Fig. 34). However, several large rivers contributed significantly to Tersky Bereg and Zimniy Bereg fisheries which should be considered in the fisheries management. More detailed analyses and results are available in accompanying project report by Prusov et al (2014).

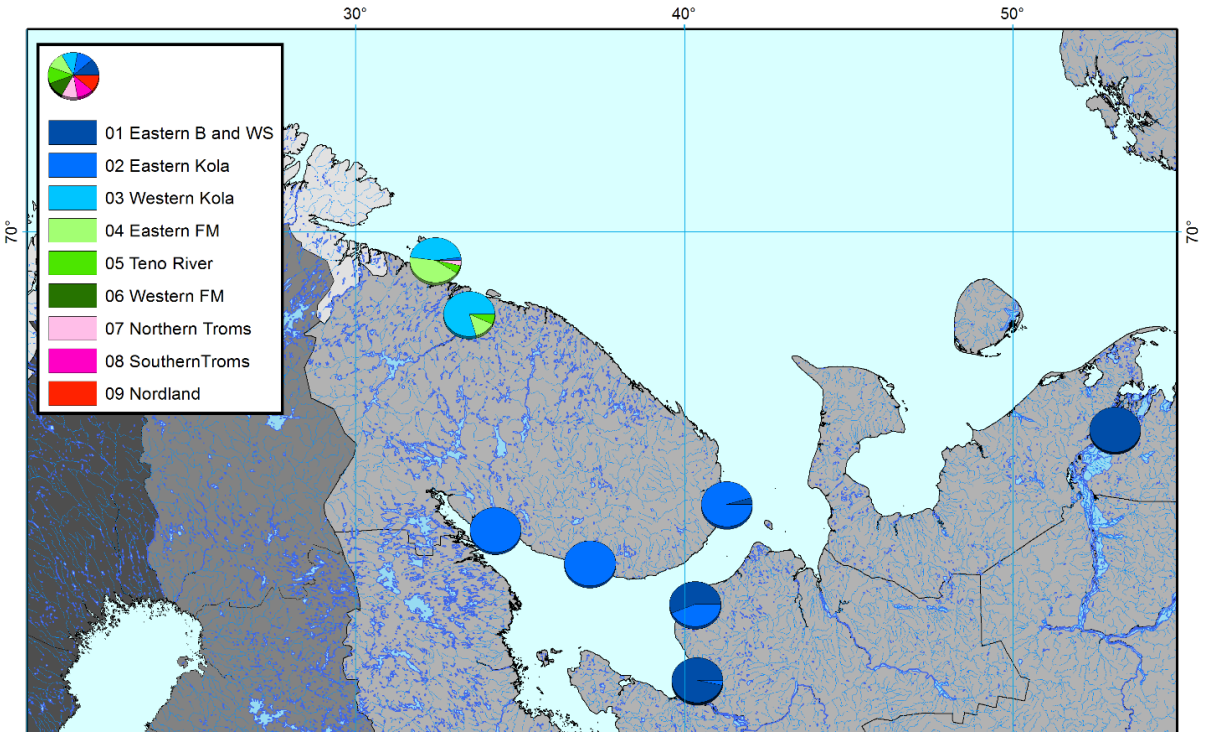


Figure 33. Proportions of reporting group contributions to fishery samples in the White Sea, Big Eina, Kola bay and Pechora estuary.

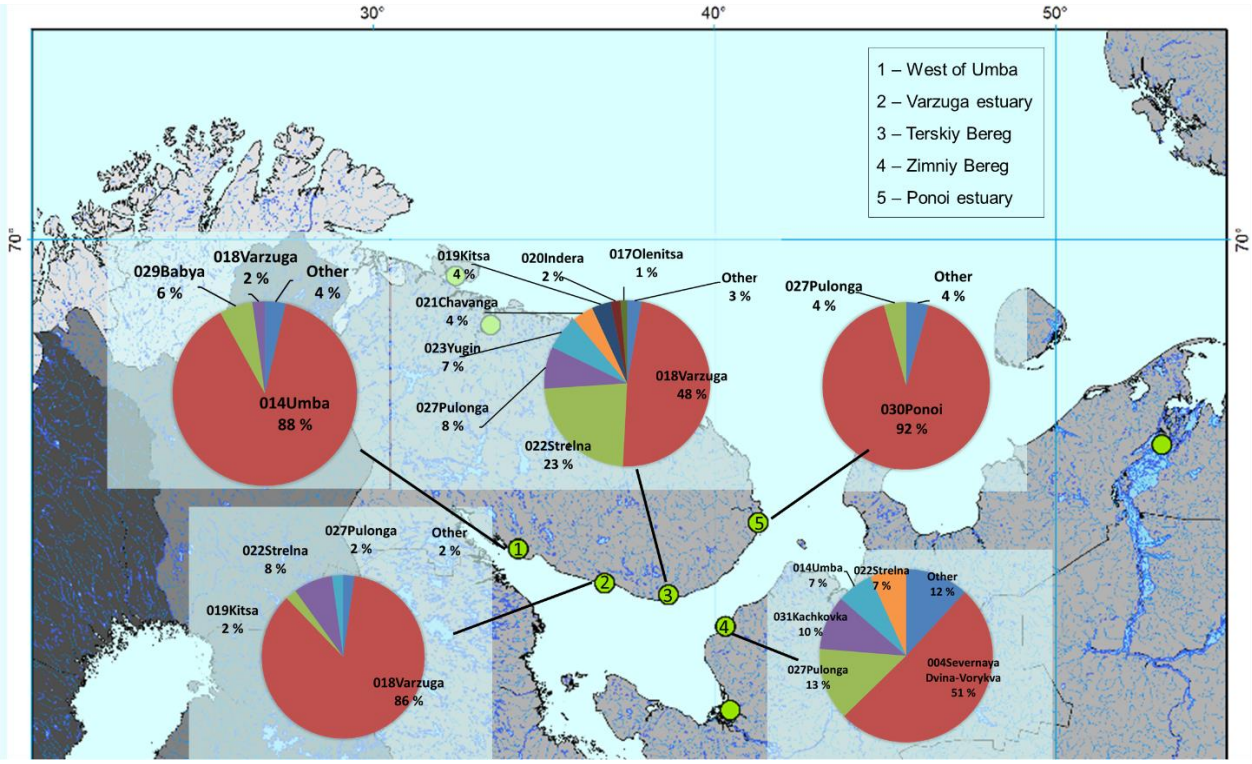


Figure 34. Contributions of the rivers to the catches in each of 5 groups of the White Sea: (1) West of Umba, (2) Varzuga estuary, (3) Terskiy Bereg, (4) Zimniy Bereg, and (5) Ponoi estuary.

3.4 Genetic stock identification of the Norwegian coastal fishery samples

3.4.1 Individual assignments and probabilities of belonging

Generally, the probabilities of belonging to specific river stocks were high; 11128 samples had probability of belonging to single river stock higher than 90% (Fig. 35). Individual assignments to reporting groups were classified to six subjective categories based on probability values (Table 12). Assignments to reporting groups largely followed the pattern revealed in the power tests with Eastern Barents and White Sea reporting group displaying highest proportion of samples falling the 'A –Excellent' –group and Northern Troms assignments having the lowest (Fig. 36). Samples where the probability of belonging to reporting group was higher than 50% (19863 samples) were considered suitable for further analyses.

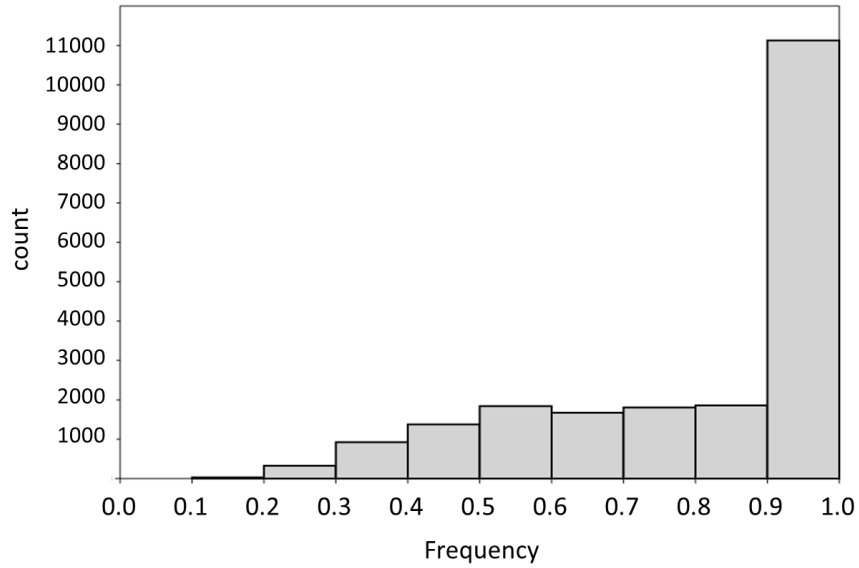


Figure 35. Number of individuals (count) and the probability (Frequency) of their belonging to single river stock.

Table 12. Criteria and the number of samples.

Class	cut-off	n
A - Excellent	> 0.90	14486
B - Very Good	> 0.80	1784
C - Good	> 0.70	1375
D - Fair	> 0.50	2218
E - Poor	> 0.40	642
F - Fail	< 0.40	471
Total		20976

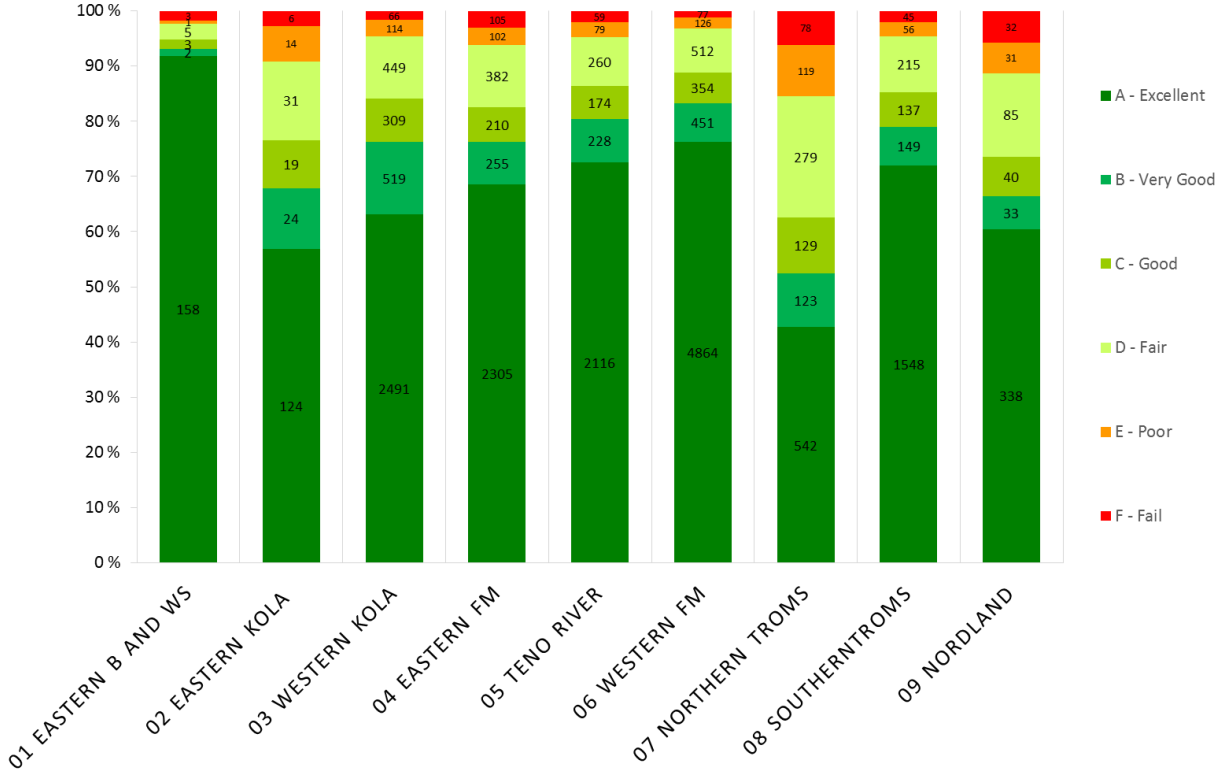


Figure 36. Proportions and number of samples assigned to reporting groups classified in 6 categories representing reliability of assignment.

For the subsequent analyses in this report, if not stated otherwise, data includes samples falling in reporting group categories A to D. For individual assignments to river stocks we then applied river specific cut-off values (Table 13). Thus, river of origin was determined for 18115 out of 20976 Atlantic salmon samples.

Altogether, 145 rivers were found to contribute to fishery samples, but there was large variation among rivers, as well as among reporting groups in contribution to fishery samples. The largest river in terms of assigned samples was river Alta accounting for nearly 10% of fishery samples. The 10 largest river stocks accounted for 46% of the fishery samples. However, River Teno, a single large river system fostering several genetically distinct salmon stocks was treated as a reporting group and as such accounted for 14.2% of the assigned salmon sampled from the coastal fisheries. Regarding Teno as a single stock, the 10 largest stocks accounted for 54% of the coastal fishery samples.

It should be noted however, that power tests (section 3.1.4) indicated that the contribution of certain river stocks may be significantly over-estimated due to individuals from neighbouring, genetically most similar rivers being incorrectly assigned to them. Among the 10 largest contributing rivers, for Alta, Teno mainstem and Bergebyelva relatively unbiased estimates were expected (upward bias 10% - 24%) while for Repparfjordelva (250%), Lakselva Porsanger (30%), Kola river (80%), and Målselva (40%) upward bias was large even after applying river stock specific cut-off values. Power tests indicated relatively unbiased estimates of contribution also for Titovka (+10%) and Bolshaya Zapadnaya Litsa (-20%) stocks after applying stock specific cut-off values, but Gelman & Rubin diagnostics indicated that the discrimination among Titovka, Bolshaya Zapadnaya Litsa and Ura river stocks was not satisfactory, implying that rather than considering the three as separate stocks in subsequent analyses, they should be considered as a group. Nevertheless, genetic stock identification allows the partitioning of samples from the mixed stock fishery catches in to smaller units and if not always representing single rivers, representing regions with genetically similar stocks.

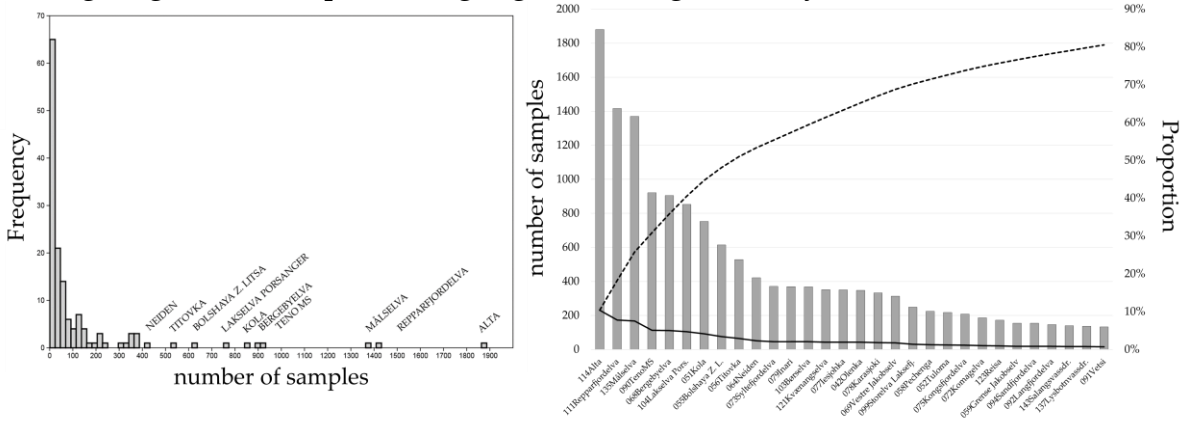


Figure 37. Contribution of stocks to fishery samples. Dashed line in figure on the right shows cumulative proportion of stock contributions (solid line).

Table 13. River contributions to coastal fishery samples. Columns from left –population specific cut-off values, number of salmon identified, mean and range of probability values of assigned samples and contribution to total sample.

Reporting Group / River	cut-off P_R	count	mean P_R	range P_R	contribution
01 Eastern B and WS					
001Pechora	0	67	0.88	(0.53-1)	0.32 %
002Mezen-Pizhma	0	24	0.88	(0.48-1)	0.11 %
003Kuloy-Soyana	0	9	0.87	(0.48-1)	0.04 %
004Severnaya Dv.	0.7	8	0.96	(0.75-1)	0.04 %
005Letnaya Zol.	0	1	1	(1-1)	0.00 %
006Onega	0.9	21	0.99	(0.98-1)	0.10 %
007Gridina	0	8	0.88	(0.55-1)	0.04 %
008Keret	0.9	22	0.98	(0.91-1)	0.10 %
uncertain	-	8	0.64	(0.49-0.88)	0.04 %
02 Eastern Kola					
010Kanda	0	8	0.62	(0.30-1)	0.04 %
011Kolvitsa	0.5	7	0.75	(0.59-0.97)	0.03 %
012Porya	0.7	1	1	(1-1)	0.00 %
014Umba	0.3	41	0.87	(0.30-1)	0.20 %
015Kuzreka	0.7	3	0.86	(0.70-0.98)	0.01 %
016Khlebnaya	0	1	0.92	(0.92-0.92)	0.00 %
017Olenitsa	0	9	0.99	(0.96-1)	0.04 %
018Varzuga	0.3	20	0.79	(0.34-0.99)	0.10 %
019Kitsa	0	6	0.60	(0.37-0.99)	0.03 %
021Chavanga	0	6	0.78	(0.55-0.95)	0.03 %
022Strelna	0.7	4	0.83	(0.72-0.93)	0.02 %
024Chapoma	0	11	0.56	(0.25-0.81)	0.05 %
027Pulonga	0	5	0.36	(0.27-0.46)	0.02 %
028Likhodeevka	0.3	4	0.50	(0.38-0.61)	0.02 %
029Babya	0.7	3	0.77	(0.76-0.78)	0.01 %
030Ponoi	0	53	0.58	(0.19-1)	0.25 %
031Kachkovka	0.7	1	0.73	(0.73-0.73)	0.00 %
uncertain	-	15	0.50	(0.29-0.69)	0.07 %
03 Western Kola					
033Iokanga	0	101	0.77	(0.27-0.99)	0.48 %
034Drozdovka	0	10	0.77	(0.35-1)	0.05 %
035Penka	0	9	0.78	(0.46-0.99)	0.04 %
036Varzina	0	8	0.79	(0.33-1)	0.04 %
037Sidorovka	0	13	0.68	(0.50-0.96)	0.06 %
038Vostochnaya L.	0	50	0.63	(0.28-1)	0.24 %
039Kharlovka	0	24	0.69	(0.32-1)	0.11 %
040Zolotaya	0	44	0.59	(0.22-0.91)	0.21 %
041Rynda	0	64	0.73	(0.34-0.99)	0.31 %
042Olenka	0	347	0.68	(0.21-0.99)	1.65 %
043Zarubikha	0	13	0.73	(0.22-0.99)	0.06 %
044Orlovka	0	3	0.87	(0.64-1)	0.01 %
046Dolgaya	0	1	0.30	(0.30-0.30)	0.00 %
047Klimkovka	0.5	50	0.73	(0.51-0.99)	0.24 %
048Tipunkova	0.5	24	0.69	(0.50-0.88)	0.11 %
049Zarbikha Kildin	0	10	0.71	(0.50-0.97)	0.05 %
050Vaenga	0	29	0.99	(0.89-1)	0.14 %
051Kola	0.7	752	0.94	(0.70-1)	3.59 %
052Tuloma	0	218	0.85	(0.33-1)	1.04 %
053Kulonga	0	13	0.94	(0.63-1)	0.06 %
054Ura	0.7	111	0.87	(0.70-0.99)	0.53 %
055Bolshaya Z. L.	0.5	615	0.83	(0.50-1)	2.93 %
056Titovka	0	528	0.64	(0.27-0.99)	2.52 %
057Pyaye	0	102	0.91	(0.27-1)	0.49 %

Reporting Group / River	cut-off P_R	count	mean P_R	range P_R	contribution
058Pechenga	0	225	0.85	(0.23-1)	1.07 %
uncertain	-	404	0.49	(0.18-0.69)	1.93 %
04 Eastern FM					
059Grense Jakobselv	0	154	0.70	(0.31-0.99)	0.73 %
060Tårnelv	0	12	0.90	(0.72-1)	0.06 %
061Karpelv	0.5	15	0.75	(0.51-0.99)	0.07 %
062Sandneselva K.	0	37	0.86	(0.49-1)	0.18 %
063Munkelv	0	40	0.79	(0.31-1)	0.19 %
064Neiden	0.5	422	0.82	(0.50-1)	2.01 %
065Klokkarelv	0	67	0.75	(0.35-0.99)	0.32 %
066Nyelva	0.3	17	0.81	(0.31-1)	0.08 %
067Vesterelva	0	122	0.88	(0.33-1)	0.58 %
068Bergebyelva	0	905	0.95	(0.25-1)	4.31 %
069Vestre Jakobselv	0	313	0.82	(0.29-1)	1.49 %
070Storelva Vadso	0.5	19	0.88	(0.53-0.99)	0.09 %
071Skallelva	0	117	0.83	(0.33-1)	0.56 %
072Komagelva	0.5	186	0.92	(0.52-1)	0.89 %
073Syltefjordelva	0.9	371	0.99	(0.90-1)	1.77 %
074Storelva Batsfj.	0	12	0.71	(0.34-0.99)	0.06 %
075Kongsfjordelva	0.3	208	0.85	(0.40-1)	0.99 %
076Storelva Ber.	0.5	13	0.73	(0.55-1)	0.06 %
uncertain	-	122	0.59	(0.24-0.89)	0.58 %
05 Teno River					
077Iesjohka	0.7	350	0.94	(0.70-1)	1.67 %
078Karasjoki	0.3	332	0.82	(0.31-1)	1.58 %
079Inari	0.5	369	0.81	(0.50-1)	1.76 %
080Aku	0	37	0.76	(0.35-0.99)	0.18 %
081Valjok	0	75	0.94	(0.43-1)	0.36 %
082Galddas	0.9	16	0.99	(0.97-1)	0.08 %
083Karigas	0.9	9	0.97	(0.90-1)	0.04 %
084Kevo	0.9	55	0.99	(0.90-1)	0.26 %
085Tsars	0	41	0.96	(0.64-1)	0.20 %
086Utsjoki	0	47	0.97	(0.33-1)	0.22 %
087laksjohka	0.5	40	0.98	(0.64-1)	0.19 %
088Leva	0.7	60	0.96	(0.70-1)	0.29 %
089Maske	0.5	82	0.93	(0.52-1)	0.39 %
090TenoMS	0.3	921	0.77	(0.31-1)	4.39 %
091Vetsi	0.7	132	0.96	(0.73-1)	0.63 %
uncertain	-	212	0.51	(0.18-0.89)	1.01 %
06 Western FM					
092Langfjordelva	0.7	147	0.91	(0.70-1)	0.70 %
093Risfjordelv	0.5	63	0.94	(0.52-1)	0.30 %
094Sandfjordelva	0	154	0.93	(0.35-1)	0.73 %
095Futelva	0	41	0.89	(0.31-1)	0.20 %
096Mehamnelva	0.7	7	0.98	(0.96-1)	0.03 %
097Ifjordelva	0	13	0.82	(0.32-1)	0.06 %
098Suosjohka	0	16	0.95	(0.59-1)	0.08 %
099Storelva Laksefj.	0.5	249	0.76	(0.50-1)	1.19 %
100Veidneselva	0.9	88	0.98	(0.90-1)	0.42 %
101Lille Porsangerelv	0.5	98	0.93	(0.51-1)	0.47 %
102Tømmervikelva	0.5	39	0.81	(0.50-0.99)	0.19 %
103Børselva	0	368	0.72	(0.22-1)	1.75 %
104Lakselva Pors.	0.9	853	0.98	(0.90-1)	4.07 %
105Stabburselva	0	118	0.82	(0.26-1)	0.56 %
106Ytre Billefjordelva	0.7	33	0.90	(0.70-1)	0.16 %
108Smørfjordelva	0	3	0.79	(0.38-0.99)	0.01 %
109Snefjordelva	0.7	5	0.92	(0.76-1)	0.02 %
110Russelv	0.7	83	0.89	(0.70-1)	0.40 %
111Repparfjordelva	0.3	1416	0.82	(0.30-1)	6.75 %
112Kvalsundelva	0.9	16	0.98	(0.90-1)	0.08 %
113Lakselva Kviby	0.9	4	0.97	(0.90-0.99)	0.02 %
114Alta	0	1880	0.91	(0.22-1)	8.96 %
115Botneelva	0	1	0.64	(0.64-0.64)	0.00 %
116Mattiselva	0	7	0.74	(0.42-1)	0.03 %
117Halselva	0.3	1	0.51	(0.51-0.51)	0.00 %
118Bognelva	0	9	0.64	(0.32-1)	0.04 %
uncertain	-	469	0.57	(0.21-0.89)	2.24 %
07 Northern Troms					
119Burfjordelva	0.3	13	0.72	(0.34-1)	0.06 %
120Badderelva	0	26	0.76	(0.41-1)	0.12 %
121Kvænangselva	0	351	0.76	(0.32-1)	1.67 %
122Oksfjordvassdr.	0.7	9	0.96	(0.80-1)	0.04 %
123Reisa	0.9	173	0.99	(0.90-1)	0.82 %
124Rotsundelva	0.5	115	0.86	(0.50-1)	0.55 %

Reporting Group / River	cut-off P_R	count	mean P_R	range P_R	contribution
125Kåfjordelva	0.3	115	0.78	(0.32-1)	0.55 %
126Manndalselva	0.3	7	0.63	(0.53-0.92)	0.03 %
127Signaldalselva	0.5	10	0.74	(0.55-1)	0.05 %
128Breivikelva	0.7	50	0.87	(0.70-0.99)	0.24 %
129Oldervikelva	0.5	2	0.93	(0.89-0.98)	0.01 %
130Skittenelva	0.5	24	0.75	(0.51-0.99)	0.11 %
132Nordkjoselva	0.7	37	0.87	(0.72-1)	0.18 %
uncertain	-	141	0.56	(0.22-0.89)	0.67 %
08 SouthernTroms					
133Lakselva Aursfjord	0.5	55	0.95	(0.57-1)	0.26 %
134Målrelva Aursfjord	0	1	1	(1-1)	0.00 %
135Målselva	0.3	1370	0.94	(0.32-1)	6.53 %
136Rossfordvassdr.	0.5	38	0.91	(0.50-1)	0.18 %
137Lysbotnvassdr.	0.3	137	0.74	(0.30-1)	0.65 %
138Laukhelle	0.3	74	0.73	(0.31-1)	0.35 %
139Skøelv	0.9	53	0.96	(0.90-0.99)	0.25 %
140Tennelvvassdr.	0.9	4	0.99	(0.98-1)	0.02 %
141Åndervassdr.	0.5	26	0.82	(0.51-1)	0.12 %
142Løksebotnvassdr.	0	8	0.60	(0.25-0.99)	0.04 %
143Salangsvassdr.	0.3	140	0.72	(0.34-0.99)	0.67 %
uncertain	-	143	0.62	(0.21-0.89)	0.68 %
09 Nordland					
144Roksdalsvassdr.	0.9	51	0.97	(0.90-1)	0.24 %
145Åseelva	0	11	0.58	(0.34-0.99)	0.05 %
146Alsvågvassdr.	0.7	19	0.87	(0.71-0.99)	0.09 %
147Buksnesvassdr.	0	5	0.53	(0.46-0.73)	0.02 %
148Gårdselv	0.9	86	0.96	(0.90-1)	0.41 %
149Kjerringnesvassdr.	0.5	35	0.81	(0.52-0.99)	0.17 %
150Holmstadelva	0.9	2	0.92	(0.91-0.93)	0.01 %
151Langvatnvassdr.	0	5	0.83	(0.61-0.99)	0.02 %
152Tårstadvassdr.	0	2	0.95	(0.93-0.97)	0.01 %
153Forsåvassdr.	0	8	0.84	(0.44-0.99)	0.04 %
154Saltdalselva	0.7	27	0.85	(0.70-1)	0.13 %
155Beiarelva	0.5	11	0.70	(0.50-0.91)	0.05 %
uncertain	-	234	0.63	(0.30-0.89)	1.12 %
Uncertain					
uncertain	-	1113	0.36	(0.10-0.49)	5.31 %

3.4.2 Stock identification of fishery samples at reporting group level

Generally, the reporting group compositions of fishery samples were more similar between years than between periods within years (Fig. 38, 39, 40).

In general, stock fishery samples from the first period were composed of salmon from wide geographical areas whereas samples from the second period were composed largely of more local populations. For example, in July and August salmon originating from the western Finnmark rivers contributed most to fishery catch samples from west of Gamvik region to southern Troms.

In addition to fishery samples from eastern Finnmark, where Russian salmon was found to be present through the sampling season, a high proportion of Russian salmon was found in western Finnmark fishery samples, especially in May. Of all the fishery samples collected in May, 48% were genetically identified as to originate from Russian salmon stocks. Even excluding the two regions from the Varanger area, the proportion of Russian salmon in the fishery samples was high in May (32%) decreasing to 9.2% in June, 5.4% in July and 3.5% in August.

Similarly, the proportion of Teno salmon in samples from analysis regions was found to be highest in May and June, but contributing more than 10% to the fishery in eight of the regions in July (Fig. 38, 39, 40).

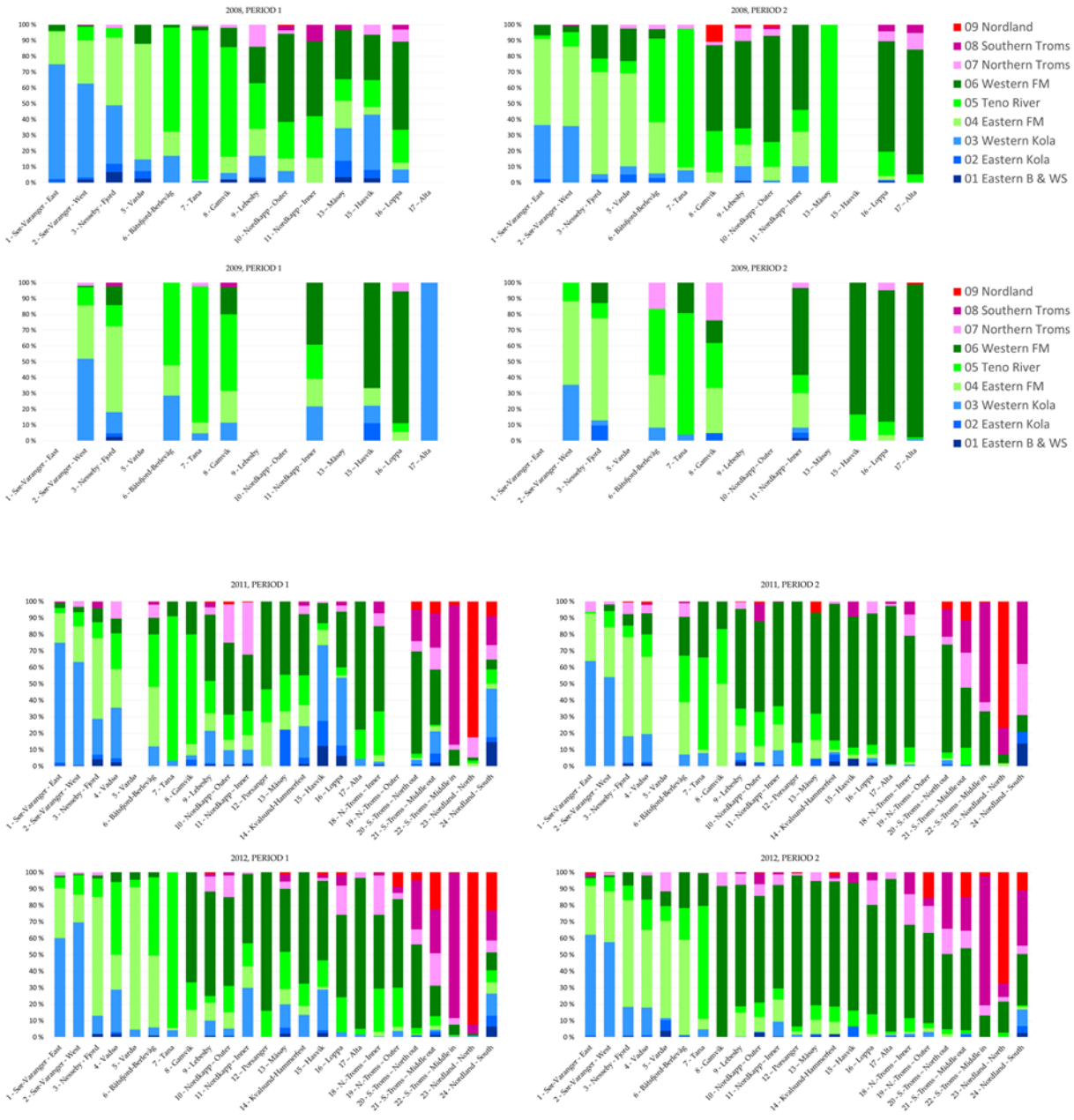


Figure 38. Proportions of reporting group contributions to fishery samples in 24 analysis regions in four years and two periods each. Period 1 includes samples until end of June and period 2 from July onwards.

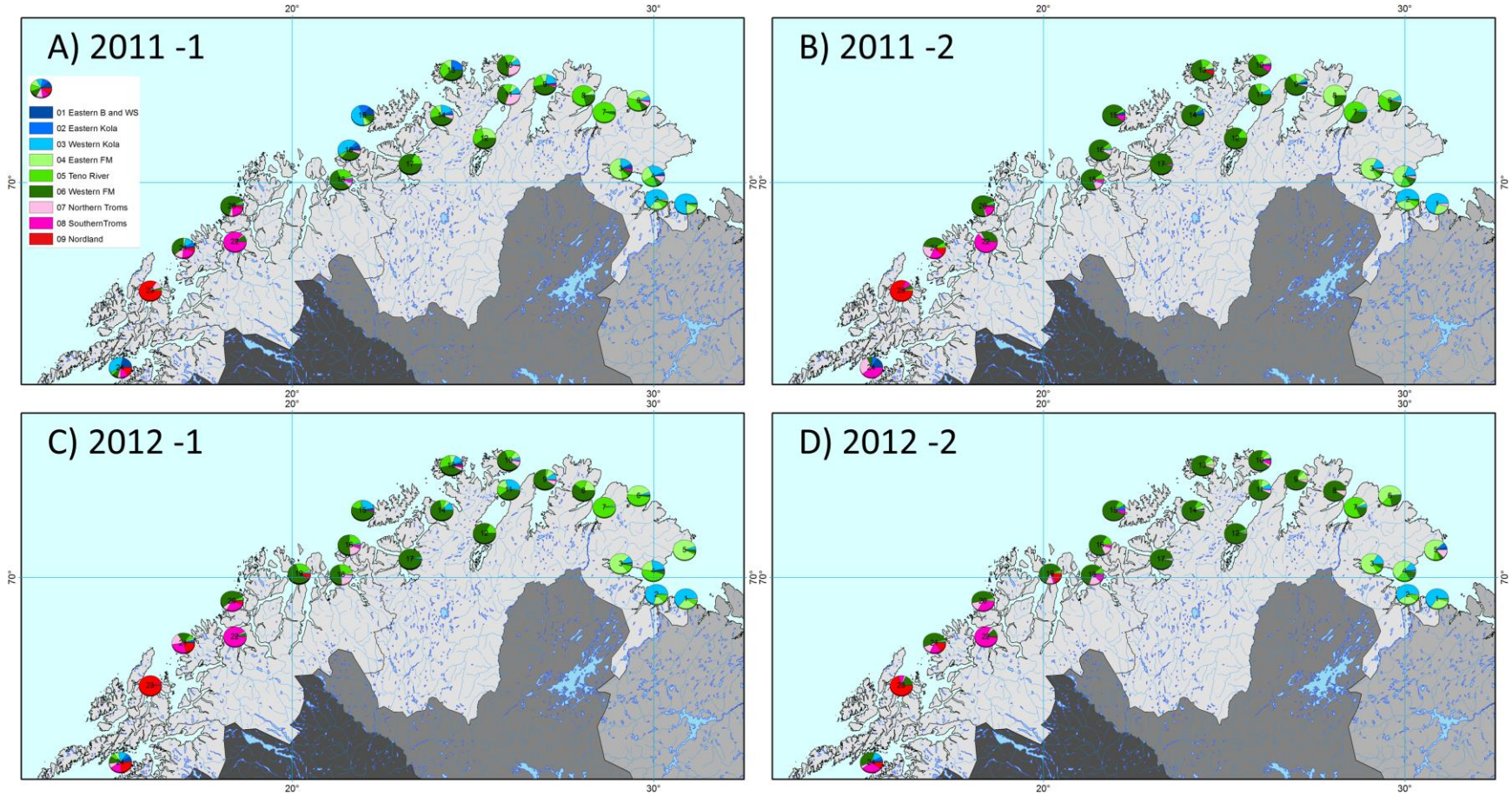


Figure 39. Proportions of reporting group contributions to fishery samples in 24 analysis regions in four years and two periods each. Period 1 includes samples until end of June and period 2 from July onwards.

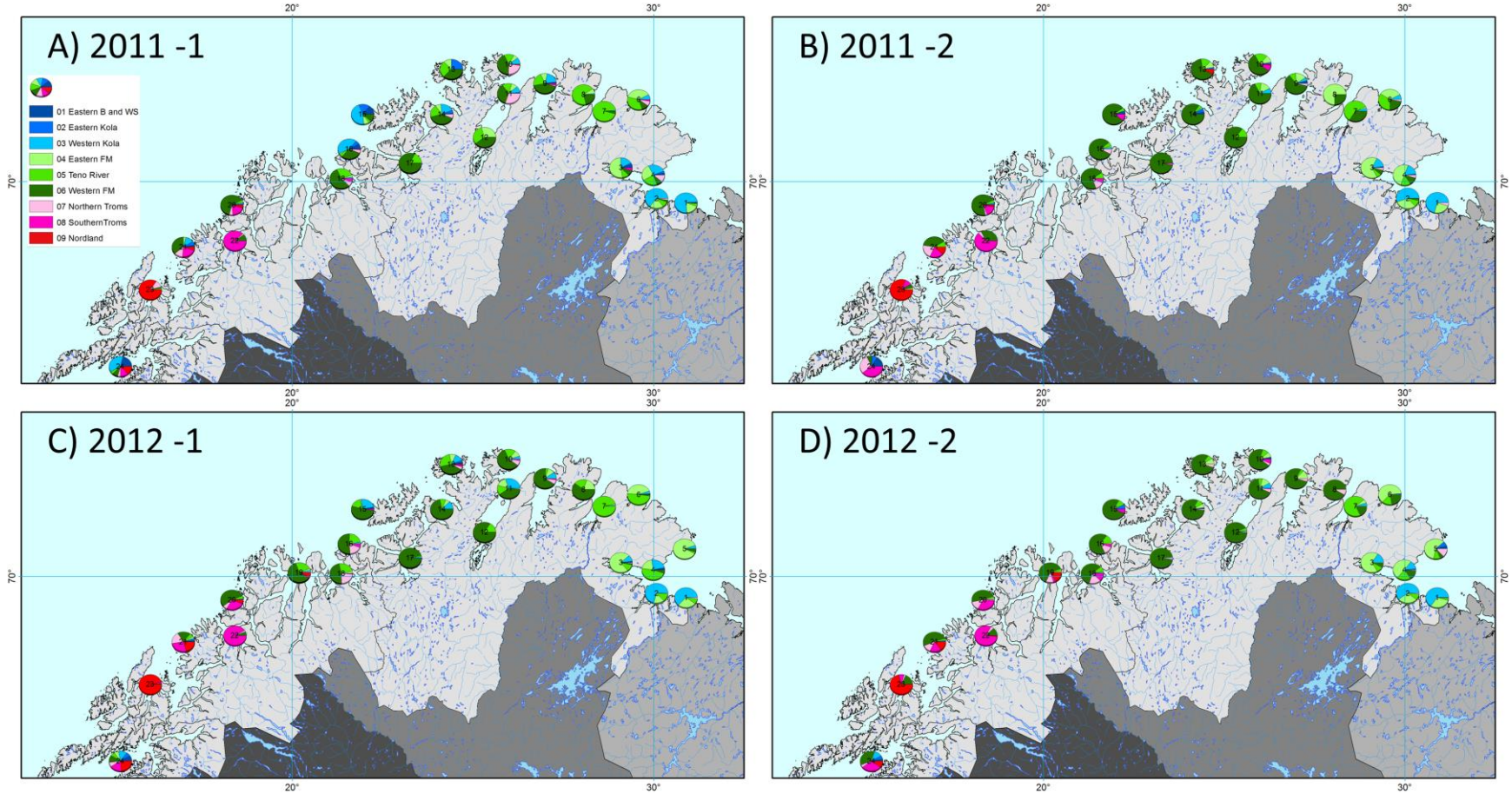


Figure 40. Proportions of reporting group contributions to fishery samples in 24 analysis regions in four years and two periods each. Period 1 includes samples until end of June and period 2 from July onwards.

3.4.4 Stock compositions in the samples from outer coastal areas

In samples from the outer coastal areas, the proportions of Russian and Teno salmon decrease through the season while the proportions of Eastern and Western Finnmark salmon increase (Fig. 41). No trends were apparent for Troms and Nordland salmon which were found largely in the two most southern areas. In Troms, Russian salmon were identified mostly during weeks 17 to 23 when the total number of salmon sampled was low. However, in Hasvik region salmon originating from Russian rivers were abundant until the end of June, week 26. In eastern Finnmark, Russian salmon were abundant throughout the sampling season.

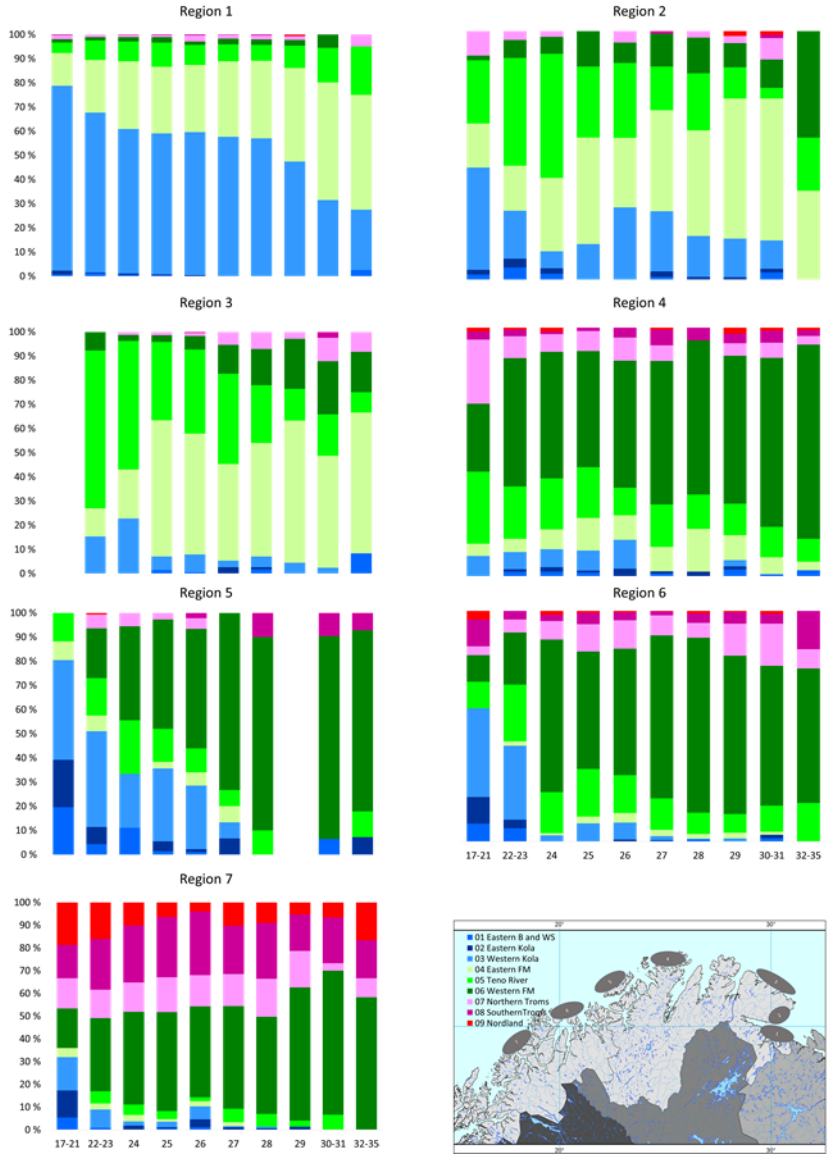


Figure 41. Weekly contribution of reporting groups to outer coast fishery samples. All years combined.

3.4.3 Stock compositions in the samples from fjords

In the fjords, fisheries target mostly local populations throughout the sampling season (Fig. 42). For example, in the Porsangerfjord, 70% of the salmon during both periods were from the local rivers (Børselva, Lakselva and Stabburselva) while in Altafjord, 80% and 89% originated from river Alta, in periods one and two respectively. Interestingly, in Tanafjord and Aursfjord, the proportion of local stocks in the fishery samples was smaller in the second period. For example in Tanafjord 93% of salmon in the May-June originated from Teno River, decreasing to 76% in July and August (all years combined).

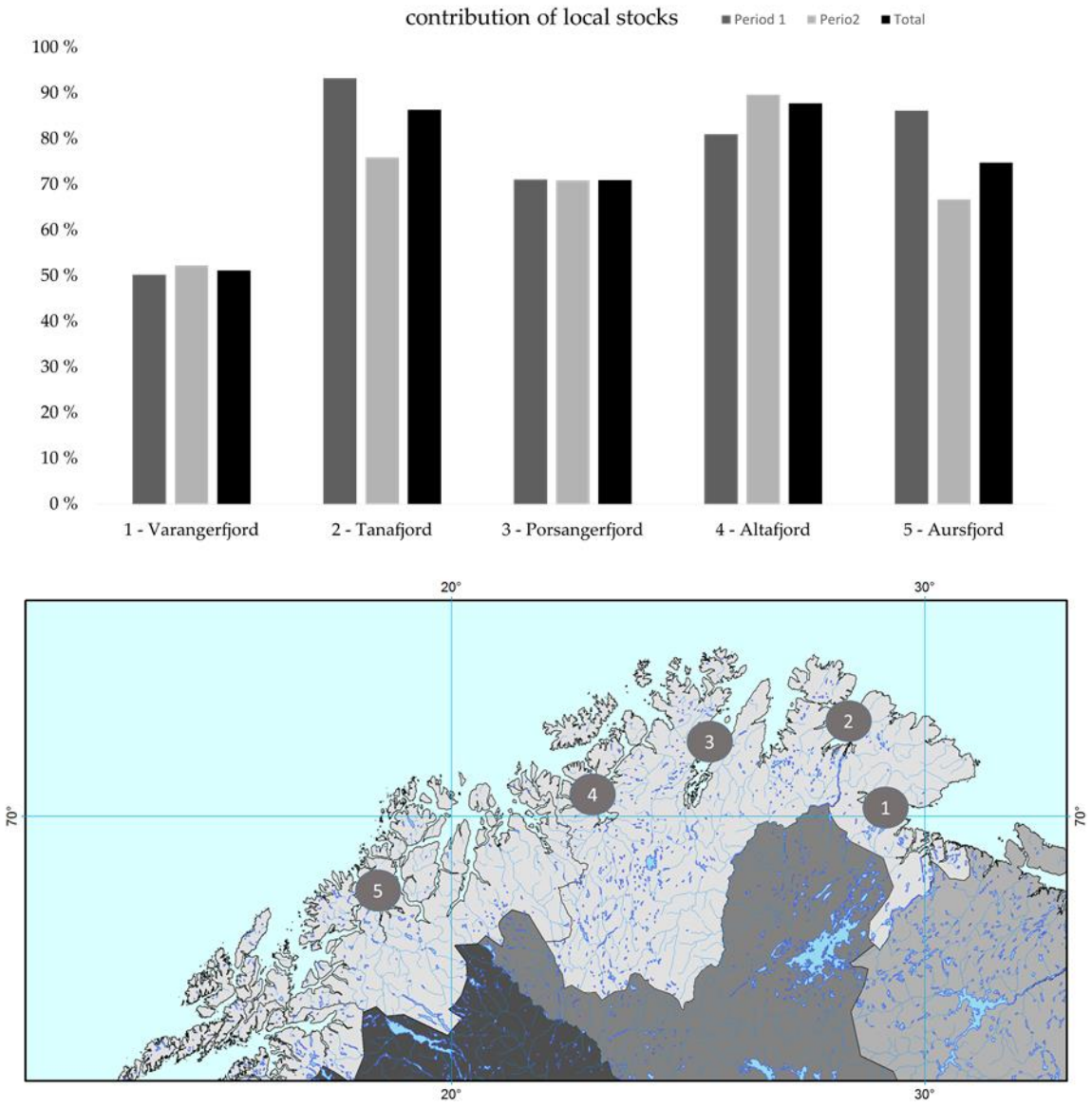


Figure 42. Contribution of local stocks to inner fjord fishery samples in May-June (dark grey), July-August (light grey) and total (black). Data from all years was combined.

Discussion

Genetic stock identification (GSI) allows assessment of stock origin of the fish being harvested, and during the last decade it has become an indispensable and powerful tool to understand fishery dynamics, especially of salmonid fishes (e.g. Oregon Salmon Commission 2008, Beacham et al. 2008, Hess et al. 2011). In this report we present stock contributions to Norwegian and Russian coastal fishery catches based on genetic stock identification of more than 23 000 sampled Atlantic salmon individuals.

Genetic stock identification analyses confirmed that coastal fisheries exploit multiple stocks. Altogether, 145 rivers were found to contribute to the fishery samples. Fisheries generally exploited salmon from wide geographical areas. For example, salmon from the Pechora River in the eastern frontier of the salmon distribution range were caught in the most western sampling sites in Nordland. Gamvik, Kvalsund-Hammerfest and the outer region of Northern Troms were the most stock rich analysis regions, while the inner parts of Sør-Varanger and Aursfjord analysis regions had the lowest diversity of stocks.

The largest river in terms of assigned samples was river Alta accounting for nearly 10% of the fishery samples. However, River Teno, a large river system fostering several genetically distinct salmon stocks accounted for 14.2% of the assigned salmon sampled from the coastal fisheries. When considering the Teno river system as a single stock, the 10 largest stocks accounted for 54% of the coastal fishery samples. Power tests indicated, however, that assignments to some of the river stocks such as Repparfjordelva may be substantially over-estimated by mis-assigned samples from genetically similar stocks in the region.

Nevertheless, fishery samples from May and June were composed of salmon from wider geographical areas than samples from July and August, which were composed of more local populations. In the fjords, fisheries target mostly local populations throughout the season.

More detailed analyses of coastal fishery samples applying the results from this report, the genetic stock identification data are available in accompanying reports which make use of the stock estimate results of this report together with e.g. total catch statistics (Niemelä et al. 2014) or fishing effort (Svenning et al. 2014). Nevertheless, stock and reporting group contributions to samples in this report are the first to provide scientific information on the compositions of mixed stock fisheries in northern Norway and Russia.

Studies applying external tags such as CWT generally suffer from low number of tagged stocks and require a large number of individuals to be tagged to obtain information as only a limited number of the tagged fish will be recaptured. Genetic "tagging" or DNA fingerprinting enables estimation of the stock of origin for every fish

sampled in the mixed stock fisheries. Thus, one of the key benefits of GSI is the opportunity to gain information on a large number of individuals from a large number of stocks. However, when interpreting results of GSI and results of subsequent analyses relying on genetic stock identification it is important to acknowledge that determination of stock of origin is probabilistic and is associated with varying levels of uncertainty.

The feasibility of applying genetic stock identification depends on the accuracy and precision of the assignments of individuals. There are five main factors that have an effect on the power of genetic stock identification. Essentially accuracy depends on (1) the number of potential stocks of origin and (2) the level of genetic differentiation among them. These are the factors which we have no control over. However, the latter of the two is related to (3) the number and features of applied genetic markers. In addition, (4) the number of reference samples collected from each population to describe their relative allele frequencies as well as (5) the statistical methods are key factors that contribute to the obtainable level of accuracy of individual assignments. Below we discuss the main points relevant to accuracy of the genetic stock identification in relation to this project.

Genetic structure of baseline populations

Number of potential stocks of origin for the Atlantic salmon harvested in the coastal areas of northern Norway and Russia is 178 at minimum i.e. the number of genetically distinct units we have described in this report. However, true number is likely higher for two reasons: 1) there are populations in the area we have not sampled or analysed 2) there may be transient salmon from rivers beyond our study area even though our baseline sampling spanned from Nordland County in Norway to Pechora river in Russia, well covering the main area of our coastal sampling (Troms and Finnmark counties). However, the magnitude of genetic difference among stocks is more important to reliability or precision of genetic stock identification than the number of stocks.

The various analyses conducted on this dataset have demonstrated large variations in genetic differentiation and diversity within the Kolarctic region. These variations, and the patterns observed are likely the results of several factors that may shape the distribution of genetic variation, such as the recolonization of salmon rivers since the last glaciation from various refugia, oceanic and environmental factors contributing to reproductive isolation and development of alternative life-histories. The genetic differences observed between the rivers in the region is generally higher than those reported from other areas, and the differences between the eastern populations in the Pechora river system and some of the Norwegian rivers are on the same level as previously observed in comparisons across the Atlantic. Along the axis from east to west in the Kolarctic area, there are large variations in environmental factors. Also, rivers in the area have been subjected to varying levels of human influence. Some of the populations sampled here are among the most pristine and undisturbed in the whole

distribution range of Atlantic salmon, while others have experienced negative effects from human activities. Also, in parts of the study area, the genetic structure of the populations may be influenced by spawning of escaped aquaculture salmon in the rivers, and such changes have been documented for one of the rivers included in this study; Vestre Jakobselv (Glover et al. 2013). On the whole, the Kolarctic area can still be viewed as a region with mostly intact and relatively undisturbed salmon populations and the genetic structure observed is the result of natural processes.

The genetic structure of the northern populations of Atlantic salmon was assessed in a number of earlier studies, using different classes of markers such as allozymes (Kazakov & Titov 1991, 1993; Bourke et al. 1997; Skaala et al. 1998), mitochondrial DNA (Nilsson et al. 2001; Asplund et al. 2004; Makhrov et al. 2005), microsatellites (Wennevik et al. 2004; Tonteri et al. 2005; Saisa et al. 2005; Ozerov et al. 2012, 2013c) and SNPs (Bourret et al. 2013, Ozerov et al. 2013a,b). Several of these studies have, based on analyses of varying numbers of populations, made inferences about the historical and present factors shaping the genetic structure and variation of Atlantic salmon in general, and of the northern populations in particular. No studies published to date however, have included populations from the whole northern periphery of the European distribution range of Atlantic salmon. The number of populations, and the number of markers included in this study, makes it the most detailed dataset ever assembled for Atlantic salmon.

The dataset presented here demonstrated that the patterns of genetic diversity within this area, measured as heterozygosity and allelic richness, showed trends that most likely reflects the colonization history in different parts of the region. Genetic diversity was generally lower in the eastern populations, and showed an increasing trend when moving from the White Sea to the populations on the Barents Sea coast of the Kola Peninsula. This is also consistent with the pattern observed by Ozerov et al. (2012) who examined the factors shaping diversity and differentiation in populations from the Karelian coast in the White Sea to the river Titovka close to the Norwegian border. Based on the results in their study they suggested that there was a general trend of increased genetic diversity with an increased carrying capacity of the river, i.e. an association between available habitat for spawning and production of juveniles would result in increased diversity. Both contemporary processes, and history have had an influence on the patterns of diversity. The lowest diversities observed in this dataset were found in some of the Teno populations, and in Kovda in the White Sea. Kovda, according to Kazakov & Titov (1991), has at one point in time suffered a severe reduction in population size, which may explain the low diversity in this river. Similarly, heavy exploitation on some of the Tana populations has reduced their size below the conservation limit, and the observed low diversity may be a reflection of this.

Looking at our dataset as a whole, the most apparent genetic shift observed was between the eastern Barents and White Sea populations of Komi, Arkhangelsk, Karelia

and Murmansk, and populations from northern Kola and westwards. This shift was observed around/between the rivers Ponoi and Kachkovka. A genetic shift between population groups in this area was also observed by Asplund et al. (2004) exploring distribution of mtDNA haplotypes. Tonteri et al. (2005), including a limited number of anadromous populations from the Barents Sea and White Sea areas, also found support for grouping Barents and White Sea/Eastern Barents populations into different phylogenetic groups when combining data from microsatellites and allozymes.

Within the eastern group of rivers, Pechora remained distinct with increasing number of clusters in the STRUCTURE analysis, demonstrating that this river system is highly differentiated from other rivers in the area. Also, the STRUCTURE analysis revealed structuring and grouping of other populations in this area (see section 2.6.1). Kazakov & Titov (1991), using allozymes as genetic markers, noted the distinctness of Pechora, and also Onega river, to other rivers in the area, and found these rivers to be related to rivers from the Baltic drainage, hypothesizing that these two rivers were in part or wholly colonized from south through what is today their upper drainages. In our data the river Onega groups with rivers from the Arkhangelsk region, possibly indicating common influences on these populations from early colonisations after the last glaciation. Tonteri et al. (2005) suggested the existence of a glacial refugium in the eastern Barents Sea as a source of the salmon recolonizing the eastern rivers and the White Sea rivers, while populations from northern Kola also could have been colonised from the Atlantic.

The populations of the northern coast of the Kola Peninsula, and westwards, differs from the eastern rivers and those in the White Sea. These rivers are also more genetically diverse, indicating that recolonisation of these rivers may have occured from several sources and a gradual northward expansion of the range of Atlantic salmon from more southern refugia as the main contributor to recolonisation of this area has been proposed (Verspoor et al. 1999). Also, it is interesting to note that Verspoor et al. (1999), Asplund et al. (2004) and Makhrov et al. (2005) observed the occurrence of an mtDNA haplotype in this area that is otherwise almost exclusively observed in the Western Atlantic, highlighting the possibility of multiple sources for recolonisation of these areas. The main shifts in the Barents and Atlantic Seas populations were observed in the Troms area, around the large populations of Reisa and Målselv, and in the inner Varangerfjord. The shift in the Troms region coincides also with changes in the coastal current and benthic communities in this area (Svein Sundby, IMR, pers. comm.), suggesting that oceanographic conditions may contribute to maintaining reproductive isolation between groups of populations. Numerous other groupings and substructures are observed in the data, such as the grouping of Teno populations, populations in the Murmansk fjord area, and the "island populations" in Nordland County. These geographically consistent genetic structures allow for the definition of reporting groups for genetic assignment, allowing for precise assignment of individuals to these groups.

Evaluating the genetic baseline developed, we see opportunities for improvement in future studies. The geographic scope of the baseline could be extended to achieve a better coverage of the eastern populations in Arkhangelsk region, especially the large Severnaya Dvina river system is likely to contain more genetic structure and variation than what has been uncovered here. Also, there is scope for better coverage in the Mezen and Pechora river systems. The structure, and possible IBD (isolation by distance) patterns revealed in Pechora could be explored further through additional sampling. We observe that sampling small rivers with low population sizes subject to fluctuations in allele frequencies complicates evaluation of structural patterns of genetic variability. Sampling of such rivers should be conducted over a number of years to approximate an average genetic profile and remove effects of random variation, and they should be considered in a metapopulation context.

This genetic baseline developed here does not only give opportunities for assignment of marine caught salmon to river of origin, but also provides, through the genetic structuring revealed, opportunities for defining important conservation units that will enable managers and fishermen to preserve the diversity and uniqueness of these populations for future generations.

Genetic markers

The utilization of highly polymorphic microsatellite markers in genetic stock identification has a proven track record of success (e.g. Beacham et al. 2006, 2008). Despite the increasing incentives of Single Nucleotide Polymorphism (SNP) markers, microsatellites still remain the marker of choice in fisheries research (Beacham et al. 2011; Hess et al. 2011; Ensing et al. 2013; Hess et al. 2014) much due to their applicability and affordability in large scale studies.

Here we have applied 31 microsatellite markers displaying more than 600 alleles in nearly 37 000 Atlantic salmon from more than 150 rivers in the Barents region. The number of genetic markers utilized (microsatellite markers) is at least double compared to typical projects with similar endeavours and to our knowledge one of the largest ever to have been applied in Atlantic salmon.

However, this was anticipated as necessary to achieve the desired goals considering the large number of populations as potential stocks of origin contributing to the mixed stock fishery. Furthermore, as discussed above, the genetic distinctiveness of populations of the Barents region was moderate low (0.055), especially in the western areas where mean pairwise F_{ST} within reporting groups was low: 0.025 (range 0.015 – 0.057).

In other projects employing genetic stock identification, the level of genetic differentiation among populations has been reported higher. For example, in Pacific salmon species F_{ST} values average: 0.097 (14 microsatellite markers/299 populations) in Sockeye salmon (*O. nerka*) (Beacham et al. 2006a); 0.063 (13/325) in Chinook salmon (*O. tshawytscha*) (Beacham et al. 2006b); 0.058 (17/318) in Coho salmon (*O. kisutch*) (Beacham

et al. 2012) and 0.033 (14/381) in Chum salmon (*O. keta*) (Beacham et al. 2009). In Atlantic salmon, population distinctiveness has been reported generally lower than for the Pacific salmon and comparable to the differentiation we observed within Nordland and Troms regions. For example, in Scotland F_{ST} values average to 0.03 (14 microsatellite markers/65 populations, Gilbey et al. 2012), in Ireland F_{ST} was reported to be 0.024 (7/27 populations, Ensing et al. 2013), and in southern Europe F_{ST} was reported to be 0.04 (12/57, Griffiths et al. 2010).

Our power test results illustrated that the potential for GSI errors among the populations to be dependent on the genetic distinctiveness. While genetic divergence of population is an intrinsic characteristic, there is large variation among the markers in the information content they supply to genetic stock identification. Loci applied in this report were not subject to a priori selection based on their information content, but were chosen among the most widely utilized in the population genetic studies of Atlantic salmon. From the genomic data resources and with the full genome sequence soon becoming available for Atlantic salmon (Davidson et al. 2010) it is possible to design tailored marker panels providing optimal resolution.

Indeed, screening for the most informative sets of markers among hundreds or thousands, may assist in revealing the loci demonstrating the highest discrimination power among the populations of interest. For example, as reported in the peer reviewed article from results obtained within this project (Ozerov et al. 2013a), application of a small set of SNPs of high divergence increase the assignment success in 1.5-2 times in comparison with randomly chosen markers (see also Wilkinson et al. 2011). However, implementation of such screening prior to the actual GSI studies requires more time and resources, in comparison with the application of "ready-to-go" sets of markers. Moreover, the population dataset may have a great influence on the identification of the most informative markers and the most informative loci selected using one set of populations may lack power when applied to another set of populations (e.g. Rosenberg et al. 2003; Lao et al. 2006, Ozerov et al 2013a). Thus, the screening for highly informative markers for GSI should aim to select among the largest number of possible candidate loci and to cover as wide as possible area of the stock distribution.

In numerous studies, the number of alleles observed at a locus or the total number of independent alleles have been shown to have significant effect on GSI accuracy (e.g. Kalinowski 2004; Beacham et al. 2011). The total number of 629 independent alleles in our study was considerably higher in comparison with similar studies in Atlantic (SALSEA_Merge: number of independent alleles = 405) and in Pacific salmon (e.g. 349 in Chum salmon etc., Smith & Seeb 2008). Moreover, more than half of the microsatellite loci were rather polymorphic with the number of independent alleles being above median of 19. On the other hand, the actual allele frequencies in a baseline and mixture affect the quality of GSI, e.g. the locus with two common alleles may provide better resolution than the locus with one common allele and numerous rare alleles (Kalinowski

2004). Thus, with highly polymorphic loci the probably of detecting alleles not observed in a baseline or in a mixture increases, affecting the overall GSI success (Kalinovski 2004, Anderson et al. 2008). In our dataset the total number of alleles in coastal samples not observed in the baseline was 36 (5.5%), whereas the number of baseline alleles not observed in coastal samples was 23 (3.4%). Given the observed level of genetic divergence among our populations (F_{ST} = 0.055), we consider this effect to have little influence on the accuracy of GSI in our dataset. Rare alleles present in a population and in individuals in the mixture sample, but not present in the baseline data sample could affect the accuracy and precision genetic stock identification. However, the method of Pella and Masuda (2001) implemented in the CBAYES used here for genetic stock identification applies a Bayesian method that shrinks allele frequencies toward a central mean is thought to minimize estimation error in allele frequencies. In addition, using the assignment of mixture individuals to populations is expected to improve the estimation of genetic parameters and therefore rare alleles are considered to have very little influence on the accuracy of GSI in our dataset.

Reference samples collected to describe relative allele frequencies

In addition to population divergence, baseline sample size has major effect on the stock assignment accuracy. The effect of sample size on accuracy of genetic stock identification has been evaluated in more detail by Kalinowski (2004) and Beacham et al. (2011). While Kalinowski (2004) reported that large baseline samples generally produce better estimates of stock proportions, Beacham et al. (2011) found that genetically less distinct populations required larger samples sizes compared to more diverged populations. In more detail, Beacham et al. (2011) found that very distinct populations (average pairwise F_{ST} ~0.07) do not benefit much from increasing sample sizes beyond 40 individuals, but an increase in population sample size profoundly influenced the accuracy of estimated stock compositions for those populations that were initially poorly estimated at smaller population sample sizes. For example, with 40 individuals in the baseline, only ~60% and 83% of asymptotic accuracy (theoretical maximum given the set of markers) was achieved for populations with the lowest divergence levels (mean pairwise F_{ST} ~0.004 and F_{ST} ~0.006, respectively). Increasing the baseline sample to 100 allowed reaching 80% and 93% of asymptotic accuracy for the same populations. Results of Beacham et al. (2011) have multiple important implications to our project.

Firstly, assignment accuracy of problematic stocks can be increased by increasing sample size from the baseline. As an example, the baseline data contained 320 population pairs with F_{ST} below 0.01 (Fig. 42). Majority of the 62 populations with genetically similar counterparts were found in Eastern and Western Kola reporting groups as well as in Nordland and Troms reporting groups. There was on average 3.1 population pairs with low divergence for each baseline stock in the 5 reporting groups above. Accordingly, in these reporting groups, stock identification at the river level was also the lowest.

Similarly, in other reporting groups only 0.5 low divergence pairs were found for each baseline stock. Mean number of individuals sampled for the low divergence populations was low; 69. In Nordland and Northern Troms mean sample sizes were the lowest: 47 and 58, respectively.

In accordance with the chosen approach in the project, considerably smaller sample sizes (60%-75%) are required to attain 95% level of asymptotic regional accuracy (Beacham et al. 2011).

Furthermore, baseline sampling procedure was also found to have an effect on the assignment accuracy. This resulted from decreasing sample sizes due to excluding of close relatives from the data as relative allele frequency estimates may be biased due to strong family structure. The analyses of kinship in the samples collected demonstrated that full siblings were present in most samples, but also that the proportion varied greatly, with some samples having no full siblings, and other samples containing up to 68% (Storelva Båtsfjord). On average, the baseline samples contained 16% full siblings. Sib-ships were eliminated by exclusion leaving total of 12860 individuals in the baseline data. However, kinship analysis and subsequent removing excess of close relatives is no substitute for good sampling procedures to obtain random sampling from the population. Exclusion of multiple full-sib family members from the baseline data does not remove family structure arising from other type of close relatives. Sampling relatives is best avoided by following good sampling practises.

In addition, populations are not necessarily static in time. This is especially relevant in small populations where allele frequencies are subject to strong effects of genetic drift and immigration. A sample from population providing relative allele frequency estimates of the contemporary population free of error, may not be a good representative of the next generation. However, small populations contribute relatively little to mixed stock fisheries and are often genetically distinct and therefore do not represent a large problem. Nevertheless, their equal relative occurrence in the data used testing assignment accuracy may have a negative effect on results and manifested levels of genetic stock identification accuracy.

In conclusion, accuracy of the genetic identification can be increased by supplementing the baseline data. Problematic stocks and regions are highlighted in power tests.

Assignment accuracy

Similarly to genetic divergence, power tests of genetic stock identification using test samples from the baseline data revealed large differences among rivers and regions. In general, levels of genetic divergence was reflected in the power tests of genetic stock identification – salmon from the highly diverged populations were identified with higher

success than those from low divergence populations. Inconveniently, large stocks are generally genetically more diverse and less diverged.

On average, 69% of samples were correctly assigned to rivers, but success varied between 0% and 100%. Applying river specific cut-off values the correct assignment rate increased by nearly 10% averaging 76%. The highest correct assignments were observed for rivers in the Eastern Barents and White Seas (no cut-off; 95%/ river specific cut-off; 97%) and Teno River system salmon stocks (86%/90%), while the lowest were observed for the Troms (54%/61%) and Nordland stocks (60%/74%). Of the largest 45 stocks contributing to fishery samples, low assignment accuracy in power tests were observed for 14 stocks (correct assignment rate < 60%, medium accuracy for 11 stocks (60-80%) and high for 20 stocks (correct assignment rate >80%). Assignment results to river stocks with low to medium assignment success should be taken cautiously and evaluated case-by-case. Where applicable, small groups of neighbouring rivers can be delineated for stock identification. For example, 054Ura - 055Bolshaya Zapadnaya Litsa – 056Titovka group as the three were problematic in the analyses. Despite the rate of correct assignments can be increased by applying more stringent cut-off values to filter out potential false positives as shown in section III, this approach does not always result in success as it comes with the cost of reduction of data. Nevertheless, for about half of the river stocks in the Kolarctic project data assignments are reliable enough for subsequent analyses at the river level. At the reporting group level rate of correct assignments were generally high. Northern and southern Troms were most problematic and for some applications should be combined.

Several similar studies to ours have faced difficulties in assigning individuals to single stocks and instead have assigned them to groups of stocks i.e. reporting (or regional) groups (e. g. Seeb et al. 2004, Templin et al. 2011, Gilbey et al. 2012, McCraney et al. 2012). Nevertheless, genetic stock identification allows the partitioning of samples from the mixed stock fishery catches in to smaller units and if not always representing single rivers, representing regions with genetically similar stocks.

Power tests showed that 90%-98% of samples assigned to Russian, Eastern Finnmark and Teno River system reporting groups were correct. Only slightly lower assignment success was obtained for the samples from rivers in Finnmark County; 87%. Northern and southern Troms reporting groups were most problematic showing 68% and 58% correct assignments, but when combined, 80% of Troms salmon were correct. Nordland had correct assignment rate of 72%. Largest mis-assingments were between Western Finnmark and Northern Troms (WF to NT 5%, NT to WF 22%) followed by Nordland and southern Troms (N to ST 20%, ST to N 4%).

Finally, it should be noted that while power tests provide an overview on how reliable an estimate is generally expected, exact proportions of correct as well as mis-

assignments will depend on real fishery samples, stock composition and their relative proportions. In power tests equal proportions of samples from rivers were used albeit analysed with real mixture samples. Furthermore, when conducting power tests, the number of samples in the baseline data are reduced (because individulas are removed from the baseline for the test) and this, as shown, has a negative effect on the assignments. Therefore, assignment success rates in the power tests may be underestimates of the true level of accuracy that can be obtained. Furthermore, the method of Pella and Masuda (2001) implemented in the CBAYES used here for genetic stock identification also makes use of information from the mixture sample. Allele frequencies of the mixture individuals assigned to baseline populations, at each MCMC step, are used to update the baseline allele frequencies. Reporting groups delineated for the GSI in this report did not follow exactly the genetic boundaries inferred. Following genetic structure boundaries in more detail would likely increase the assignment success to reporting groups and/or allow for smaller reporting groups to be delineated. More importantly, supplementing baseline population data with more samples from rivers is expected to significantly improve the accuracy of GSI.

In conclusion, assignment success to reporting groups were high, especially for Finnmark and Russian regions. Nordland and Troms County reporting groups were slightly more problematic due to low genetic divergence accompanied with low number of individuals in some baseline samples analyzed from rivers. Current data provides very reliable results for subsequent analyses relying on reporting group assignments, especially if Troms County is regarded as a single reporting group. Furthermore, for about half of the salmon stocks in the baseline data assignments at the river level can be considered very reliable for subsequent analyses. For some applications, applying more stringent cut-off values to filter out false positives may be useful, but in this case, increase in accuracy have to be weighed against reduction in number of assigned samples. The massive data set compiled in this project will serve as a backbone also for the future GSI analyses of unknown fishery samples. With accumulating baseline data, genetic stock assignments presented here can be refined, but the current data already now provides valuable information on the stock compositions, harvest rates and migration patterns of salmon of the Barents Sea Region.

Supplementary material

Appendix Table 1. River samples collected and analysed for genetic baseline, with name, sample collector, sampling year, geographic position, numbers collected and used in analysis.

Appendix Table 2. Division of rivers into reporting groups, and diversity indices, and evaluation of Hardy-Weinberg equilibrium for the river samples.

Supplementary Table 1. Tests for Hardy-Weinberg equilibrium.

Supplementary Table 2. Genic differentiation for each population pair (exact G-test).

Supplementary Table 3. Pairwise F_{ST} estimates between all sample pairs.

Acknowledgements

This study was funded by the European Union, Kolarctic ENPI CBC project – *Trilateral cooperation on our common resource; the Atlantic salmon in the Barents region* - "Kolarctic salmon" (KO197).; Finnish Academy of Science (J-PV); Finnish Game and Fisheries Research Institute (E.H.); Norwegian Directorate of Nature Management and Office of the Finnmark County Governor/ Norwegian Ministry of the Environment (T.H.K); Office of the Finnmark County Governor (BCH); NINA-Tromsø (MAS, MF), IMR (VW, LU).

Main responsibilities in this report:

Juha-Pekka Vähä – Genetic and statistical analyses of coastal samples and writing of the report.
Vidar Wennevik – Genetic and statistical analyses of the baseline samples and writing of the report.
Mikhail Ozerov - Statistical analyses and writing of the report.
Rogelio Diaz Fernandez – Genetic analyses of coastal and baseline samples
Laila Unneland - Genetic analyses of baseline samples
Kristiina Haapanen – Assisted in genetic analyses of coastal and baseline samples
Eero Niemelä – Collected baseline samples from Norwegian rivers. Conducted the work concerning the salmon scales collected in Nordland, Troms and Finnmark.
Martin S. Svenning– Organized permissions for fishing outside the ordinary fishing season and provided additional funding for the analyses of coastal samples from Troms.
Morten Falkegård – Provided additional funding for the Teno river system baseline sample analyses
Sergey Prusov – Collected baseline samples from Russian rivers. Organized the coastal sampling on the Russian coastal areas.
Ivan Lyzhov and Kira Rysakova – Assisted in genetic analyses of coastal samples from Russia
Tiia Kalske and Bente Christiansen - Organised scale collection from coastal fisheries and provided additional funding for the analyses of coastal samples from Finnmark.
Gennadyi Ustyuzhinsky - Collected baseline samples from Russian rivers

Without the help of professional fishermen we could not have collected the marvellous biological material from the salmon catches. We are thankful for the salmon fishing organizations in Finnmark and Troms.

References

Anderson E.C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources* 10: 701–710.

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

Asplund T., Veselov A., Primmer C.R., Bakhmet I., Potutkin A., Titov, S., Zubchenko, A., Studenov, I., Kaluzchin, S., Lumme, J. 2004. Geographical structure and postglacial history of MtDNA haplotype variation in Atlantic salmon (*Salmo salar* L.) among rivers of the White and Barents Sea basins. *Annales Zoologici Fennici* 41: 465–475.

Balloux F. 2001. EASYPOP (Version1.7): A computer program for population genetics simulations. *Journal of Heredity* 92: 301–302

Beacham T.D., Candy J.R., Jonsen K.L., Supernault J., Wetklo M., Deng L., Miller K. M., Withler R.E., Varnavskaya N. 2006b. 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.

Beacham T.D., Candy J.R., Le K.D., Wetklo M. 2009. Population structure of chum salmon (Oncorynchus keta) across the Pacific Rim determined from microsatellirte analysis. *Fisheries Bulletin* 107:244–260.

Beacham T.D., Candy J.R., Porszt E., Sato S., and Urawa S. 2011. Microsatellite identification of Canadian sockeye salmon rearing in the Bering Sea. *Transactions of the American Fisheries Society* 140: 296–306.

Beacham T.D., Candy J.R., Wallace C., Wetklo M., Deng L., MacConnachie C. 2012. Microsatellite mixed-stock identification of coho salmon (*Oncorhynchus kisutch*) in British Columbia. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* 4: 85-100.

Beacham T.D., McIntosh B., MacConnachie C., Miller K.M., Withler R.E., Varnavskaya N. 2006a. Pacific Rim population structure of sockeye salmon as determined from microsatellite analysis. *Transactions of the American Fisheries Society* 135: 174–187.

Beacham T.D., Winther I., Jonsen K.L., Wetklo M., Deng L., Candy J. R. 2008. The application of rapid microsatellite-based stock identification to management of a Chinook salmon troll fishery off the Queen Charlotte Islands, British Columbia. *North American Journal of Fisheries Management* 28: 849–855.

Bourke A.A., Coughlan J., Jansson H., Galvin P., Cross T.F. 1997. Allozyme variation in populations of Atlantic salmon located throughout Europe: diversity that could be compromised by introductions of reared fish. *ICES Journal of Marine Science* 54: 974–985.

Bourret V., Kent M.P., Primmer C.R., Vasemägi A., Karlsson S., Hindar K., McGinnity P., Verspoor E., Bernatchez L., Lien S. 2013. SNP–array reveals genome–wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (Salmo salar). *Molecular Ecology* 22: 532–551.

Cairney M., Taggart J. B., Hoyheim B. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (Salmo salar L.) and cross-species mplification in other salmonids. *Molecular Ecology* 9: 2175–2178.

Davidson W.S., Koop B.F., Jones S.J.M., Iturra P., Vidal R. Maass A., Johanssen I., Lien S., Omholt S.W. 2010. Sequencing the genome of the Atlantic salmon (Salmo salar). *Genome Biology* **11:**403

Dempson J.B., O'Connell M.F., Schwarz C.J. 2004. Spatial and temporal trends in abundance of Atlantic salmon, *Salmo salar*, in Newfoundland with emphasis on impacts of the 1992 closure of the commercial fishery. *Fisheries Management and Ecology* 11: 387–402.

Dillane E., McGinnity P., Coughlan J.P., Cross M.C., De Eyto E., Kenchington E., Prodöhl P., Cross T.F. 2008. Demographics and landscape features determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. *Molecular Ecology* 17: 4786–4800.

Earl D., vonHoldt B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Ellis J.S., Gilbey J., Armstrong A., Balstad T., Cauwelier E., Cherbonnel C., Consuegra S., Coughlan J., Cross T.F.., Crozier W., Dillane E., Ensing D., de Leaniz, C.G., Garcia-Vazquez E., Griffiths A.M., Hindar K., Hjorleifsdottir S., Knox D., Machado-Schiaffino G., McGinnity P., Meldrup D., Nielsen E.E., Olafsson K., Primmer C. R., Prodöhl P., Stradmeyer L., Vähä J.P., Verspoor E., Wennevik V., Stevens J.R. 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.

Elphinstone M.S., Hinten G.N., Anderson M.J., Nock C.J. 2003. An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes* 3: 317–320.

Ensing D., Crozier W.W., Boylan P., O'Maoiléidigh N., McGinnity P. 2013. An analysis of genetic stock identification on a small geographical scale using microsatellite markers, and its application in the management of a mixed–stock fishery for Atlantic salmon Salmo salar in Ireland. *Journal of Fish Biology* 82:2080–2094.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 8: 2611–2620.

Falush D., Stephens M., Pritchard J.K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.

Gilbey J., Stradmeyer L., Cauwelier E., Middlemas S. 2012. Genetic investigation of the North East English net fisheries. Marine Scotland Science Report 04/12. www.scotland.gov.uk/marinescotland.

Glover K.A., Pertoldi C., Besnier F., Wennevik V., Kent M., Skaala O. 2013. Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC Genetics* 14: 74.

Griffiths A.M., Machado-Schiaffino G., Dillane E., Coughlan J., Horreo J. L., Bowkett A.E., Minting P., Toms S., Roche W., Gargan P., McGinnity P., Cross T., Bright D., Garcia-Vazquez E., Stevens J.R. 2010. Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in the southern part of the European range. *BMC Genetics* 11: 31.

Grimholt U., Drablos F., Jorgensen S.M., Høyheim B., Stet R.J.M. 2002. The major histocompatability class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics* 54: 570–581

Heggberget T.G., Lund R.A., Ryman N., Ståhl G. 1986. Growth and genetic variation of Atlantic salmon (*Salmo salar*) from different sections of the river Alta, North Norway. *Canadian Journal of Fisheries and Aquatic Sciences* 43: 1828–1835.

Hess J.E., Matala, A.P., Narum S.R. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources* (S1): 137–149.

Hess J.E., Whiteaker J.M., Fryer J.K., Narum S.R. 2014. Monitoring stock-specific abundance, run timing, and straying of Chinook salmon in the Columbia River using genetic stock identification (GSI). *North American Journal of Fisheries Management* 34: 184–201.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65–70.

ICES. 2009. Report of the Working Group on North Atlantic Salmon. ICES Headquarters, Copenhagen, 30 March–8 April 2009. ICES CM 2009/ACFM: 06. 283 pp.

ICES. 2013. Report of the Working Group on North Atlantic Salmon (WGNAS). ICES CM 2013/ACOM: 09, 380pp.

Jakobsson M., Rosenberg N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:

1801–1806.
Jones O.R., Wang J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data**.** *Molecular Ecology Resource*s 10: 551–555.
Kalinowsi S.T., Manlove K.R., Taper M.L. 2007. ONCOR A computer program for Genetic Stock Identification. Department of Ecology, Montana State University, Bozeman MT 59717. Available for download from http://www.montana.edu/kalinowski
Kalinowski S.T. 2004. Genetic polymorphism and mixed stock fisheries analysis. *Canadian Journal of Fisheries and Aquatic Sciences* 61: 1075–1082.
Kazakov R., Titov S. 1991. Geographical patterns in the population genetics of Atlantic salmon, *Salmo salar* L., on U.S.S.R. territory, as evidence for colonization routes. *Journal of Fish Biology* 39: 1–6.
Kazakov R., Titov S. 1993. Population genetics of salmon, *Salmo salar* L., in northern Russia. *Aquaculture and Fisheries Management* 24: 495–506.
Keenan K., McGinnity P., Cross T. F., Crozier W. W., Prodöhl, P.A. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* 4: 782–788.
King T.L., Eackles M.S., Letcher B.H. 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.
Lao O., van Duijn K., Kersbergen P., de Knijff P., Kayser M. 2006. Proportioning whole–genome single–nucleotide–polymorphism diversity for the identification of geographic population structure and genetic ancestry. *American Journal of Medical Genetics* 78: 680–690.
Lewis S.M., Raftery A.E. 1997. Estimating Bayes factors via posterior simulation with the Laplace–Metropolis estimator. *Journal of the American Statistical Association* 92: 648–655.
Makhrov A.A., Verspoor E., Artamonova V.S., O'Sullivan M. 2005. Atlantic salmon colonization of the Russian Arctic coast: pioneers from North America. *Journal of Fish Biology* 67 (S1): 68–79.
McConnell S., Hamilton L., Morris D., Cook D., Paquet D., Bentzen P. Wright J. 1995. Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon. *Aquaculture* 137: 19–30.
McCraney W.T., Farley E.V., Kondzela C.M., Naydenko S.V., Starovoytov A.N., Guyon J.R. 2012. Genetic stock identification of overwintering chum salmon in the North Pacific Ocean. *Environmental Biology of Fishes* 94: 663–668.
Milner G.B., Teel D.J., Utter F.M. 1983. Genetic stock identification study. Bonneville Power Administration, Agreement DE–A179–82BP28044MOOl), Coastal Zone and Estuarine Studies Division, Northwest and Alaska Fisheries Center, National Marine Fisheries Service, National Oceanic and atmospheric Administration, 101 electronic pages (BPA Report DOE/BP–I0002). https://pisces.bpa.gov/release/documents/documentviewer.aspx?doc=0002–I
NASCO. 2009. NASCO Guidelines for the Management of Salmon Fisheries. North Atlantic Salmon Conservation Organization, Edinburgh, Scotland, UK. NASCO Council Document CNL(09)43, 12pp.
Neaves P.I., Wallace J.R., Candy J.R., Beacham T.D. 2005. CBayes: computer program for mixed stock analysis of allelic data, version v4.02. Available: pac.dfo-mpo.gc.ca/sci/mlg/Cbayes_e.htm
Niemelä E., Hassinen E., Aro P., Haantie J., Kuusela J., Prusov S., Samoylova E., Kalske T. (red.), Christiansen B., Vähä J.-P., Ozerov M., Diaz Fernandez R., Wennevik V., Svenning M.A., Falkegård M. 2014. Summary results from the coastal and fjord salmon fishery in the years 2011 and 2012 in Nordland, Troms, Finnmark and White Sea (Russia): timing of the salmon catches, wild and escaped salmon, sea- and freshwater ages, sex distributions and other biological parameters. Kolarctic ENPI CBC - Kolarctic salmon project (KO197) - Report III.
Nilsson J., Gross R., Asplund T., Dove O., Jansson H., Kelloniemi J., Kohlmann K., Loytynoja A., Nielsen E.E., Paaver T., Primmer C. R., Titov S.F., Vasemägi A., Veselov A., Ost T., Lumme J. 2001. Matrilinear

phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area. *Molecular Ecology* 10: 89–102.

Oregon Salmon Commission. 2008. Project CROOS collaborative research on Oregon ocean salmon, 2007. Final Report to the Oregon Watershed Enhancement Board, 214 p.

O'Reilly P.T., Hamilton L.C., McConnell S.K., 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.

Ozerov M., Vasemägi A., Wennevik V., Diaz-Fernandez R., Kent M., Gilbey J., Prusov S., Niemelä E., Vähä, J.-P. 2013a. Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for Genetic Stock Identification. *PLoS One* 8: e82434.

Ozerov M., Vasemägi A., Wennevik V., Niemelä E., Prusov S., Kent M., Vähä J.-P. 2013b. Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon (*Salmo salar* L.). *BMC Genomics* 14: 12.

Ozerov M.Y., Veselov A.E., Lumme J., Primmer C.R. 2012. "Riverscape" genetics: river characteristics influence the genetic structure and diversity of anadromous and freshwater Atlantic salmon (*Salmo salar*) populations in northwest Russia. *Canadian Journal of Fisheries and Aquatic Sciences* 69: 1947–1958.

Ozerov M.Y., Veselov A.E., Lumme J., Primmer C.R. 2013c. Temporal variation of genetic composition in Atlantic salmon populations from the Western White Sea Basin: influence of anthropogenic factors? *BMC Genetics* 14: 88.

Park S.D.E. 2001. The Excel microsatellite toolkit. **http://animalgenomics.ucd.ie/sdepark/ms–toolkit/**.

Parrish D.L., Behnke R.J., Gephard S.R., McCormick S.D., Reeves G.H. 1998. Why aren't there more Atlantic salmon (*Salmo salar*)? *Canadian Journal of Fisheries and Aquatic Sciences* 55(S1): 281–287.

Paterson S., Piertney S.B., 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 methods for analysis of stock mixtures from genetic characters. *Fisheries Bulletin* 99: 151–167.

Primmer C.R., Veselov A.E., Zubchenko A., Poututkin A., Bakhmet I., Koskinen M.T. 2006. Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia. *Molecular Ecology* 15: 653–666.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Prusov S., Ustyuzhinsky G., Wennevik V., Vähä J.-P., Ozerov M., Diaz Fernandez R., Niemelä E., Svenning M.A., Falkegård M., Kalske T., Christiansen B., Samoylova E., Chernov V., Potutkin A., Tkachenko A. 2014. Summary results from coastal salmon fisheries in the White Sea: timing and origin of salmon catches Kolarctic ENPI CBC - Kolarctic salmon project (KO197) – the White Sea Report.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL http://www.R–project.org/.

Rannala B., Mountain J.L. 1997 Detecting immigration by using multilocus genotypes. *Proceedings of National Academy of Sciences USA* 94: 9197–9201.

Raymond M., Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Rodriguez-Ramilo S. T., Wang J. 2012. The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources* 12: 873–884.

Rosenberg N.A., Li L.M., Ward R., Pritchard J.K. 2003. Informativeness of genetic markers for inference ancestry. *American Journal of Human Genetics* 73: 1402–1422.

Säisä M., Koljonen M. L., Gross R., Nilsson J., Tähtinen J., Koskiniemi J., Vasemägi, A. 2005. Population genetic structure and postglacial colonization of Atlantic salmon (*Salmo salar*) in the Baltic Sea area based on microsatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences* 62: 1887–1904.

Sánchez J.A., Clabby C., Ramos D., Blanco G., Flavin F., Vazquez E., Powell R. 1996. Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). *Heredity* 77: 423–432.

Seeb L.W., Crane P.A., Kondzela C.M., Wilmot R.L., Urawa S., Varnavskaya N.V., Seeb J.E. 2004. Migration of Pacific Rim chum salmon on the high seas: insights from genetic data. *Environmental Biology of Fishes* 69: 21–36.

Skaala O., Makhrov A.A., Karlsen T., Jorstad K.E., Altukhov Y.P., Politov D.V., Kuzishin K.V., Novikov G.G. 1998. Genetic comparison of salmon from the White Sea and north–western Atlantic Ocean. *Journal of Fish Biology* 53: 569–580.

Slettan A., Olsaker I., Lie O. 1995. Atlantic Salmon, Salmo salar, microsatellites at the Ssosl25, Ssosl85, Ssosl311, Ssosl417 loci. Animal Genetics 26: 281-282.

Smith C.T., Seeb L.W. 2008. Number of alleles as a predictor of the relative assignment accuracy of short tandem repeat (STR) and single–nucleotide–polymorphism (SNP) baselines for chum salmon. *Transactions of the American Fisheries Society* 137: 751–762.

Stet R.J.M., de Vries B., Mudde K., Hermsen G. T., van Heerwaarden J., Shum B.P., and Grimholt U. 2002. Unique haplotypes of major histocompatibility class II A and class II B alleles in Atlantic salmon (Salmo salar) give rise to diverse class II genotypes. *Immunogenetics* 54: 320–331.

Svenning M.A., Falkegård M., Fauchald P., Yoccoz N., Niemelä E., Vähä J.-P., Ozerov M., Wennevik V., Prusov S. 2014. Region- and stock-specific catch and migration models of Barents Sea salmon. NINA Report 1032.

Templin W.D., Decovich N.A., Guyon J.D., Gilk–Baumer S.E., Molyneaux D.B., Seeb L.W. 2011. Genetic stock identification of Kuskokwim Area Chinook salmon. USFWS Office of Subsistence Management, Fisheries Resource Monitoring Program, Final Report for Study No. 05–305, Anchorage, Alaska.

Tonteri A., Titov S., Veselov A., Zubchenko A., Koskinen M.T., Lesbarreres D., Kaluzhin S., Bakhmet I., Lumme J., Primmer C.R. 2005. Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from Northern Europe. *Annales Zoologici Fennici* 42: 1–22.

Vähä J.-P., Erkinaro J., Niemelä E., Primmer C.R. 2008. Temporally stable genetic structure and low migration in an Atlantic salmon population complex: implications for conservation and management. *Evolutionary Applications* 1: 137–154.

Vasemägi A., Nilsson J., Primmer C. 2005. Seventy-five EST-linked Atlantic salmon (Salmo salar L.) microsatellite markers and their cross-amplification in five salmonid species. Molecular Ecology Notes 5: 282–288.

Verspoor E., McCarthy E., Knox D., Bourke E., Cross T. F. 1999. The phylogeography of European Atlantic salmon (Salmo salar L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. *Biological Journal of the Linnaean Society* 68: 129–146.

Wang J. 2004. Sibship reconstruction from genetic data with typing errors. *Genetics* 166: 1963–1979.

Wang J. 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics* 191: 183–194.

Wennevik V., Skaala O., Titov S.F., Studyonov I., Naevdal G. 2004. Microsatellite variation in populations of Atlantic salmon from north Europe. *Environmental Biology of Fishes* 69: 143–152.

Wilkinson S., Wiener P., Archibald A.L., Law A., Schnabel R.D., McKay S.D., Taylor J.F., Ogden R. 2011. Evaluation of approaches for identifying population informative markers from high density SNP chips. *BMC Genetics* 12: 45.