Prosjektrapport

Rapport:	Nr. – År:	Dato:
RAPPORT FRA HAVFORSKNINGEN	[X-XXXX]	11.09.2017

Tittel (norsk og engelsk):
Utvandringsruter for smolt i norske fjorder – genetiske analyser

Migration routes of smolt in Norwegian fjords – genetic analysis

Forfattere:
Alison C. Harvey, Maria Quintela, Kevin A. Glover & Vidar Wennevik

Distribusjon: Åpen

Havforskningsprosjektnr.:
14891-07

Oppdragsgiver(e):
Nærings- og fiskeridepartementet

Oppdragsgivers referanse:
[xxxxx xxxxxx]

Program:
Akvakultur

Forskningsgruppe:
Populasjonsgenetikk

Antall sider totalt:
29

Sammendrag (norsk):
[Tekst]

Summary (English):
This report describes the genetic analyses used to assign salmon smolts caught in the 2017 sea lice monitoring trawls in three fjord systems in Norway back to their river of origin. Samples were gathered from various rivers in Boknafjord, Hardangerfjord and Sognefjord in order to create genetic baselines for each fjord to be able to genetically assign the trawl smolts. The present results are to be treated as preliminary as the completeness of the baseline samples varied across the fjord systems.

Emneord (norsk):
1. Laks
2. Genetikk
3. Utvandering

Subject heading (English):
1. Altantic salmon
2. Genetics
3. Migration routes

Vidar Wennevik
prosjektleder

faggruppeleder

HAVFORSKNINGSINSTITUTTET

Background

Each year the salmon louse monitoring program conducts trawling surveys for wild salmon within a number of Norway's fjord systems in order to estimate the dispersion of sea lice larvae, and the resulting potential infection pressure and mortality on wild salmonids. The spread of larvae is estimated through the use of a hydrodynamic model, and a smolt migration model has been developed to estimate temporal and spatial overlap of larvae and smolts. However, the smolt migration model is parameterised using only estimates of smolt movement. To improve and verify the model, genetic tools may be used to assign smolts caught in trawl surveys in these fjords back to their river of origin. Through such assignments our knowledge of timing and migration routes of smolts from different rivers can be increased and applied to improve the model estimates about the spatial and temporal distribution of migrating salmon. These data will also strengthen our knowledge of potential areas in the fjord where infection pressure and the risk for mortality of wild smolts is high. In the current pilot study, we have explored the feasibility of using a microsatellite based genetic baseline of rivers in three fjord systems to identify the origin of post-smolts caught in trawl surveys in these fjords.

Methods

Sample collection & sampling

Trawl samples

Trawl samples were collected over the summer (May – July) of 2017 for each fjord system. Trawling in the fjords was conducted over 4 weeks, covering different parts of the fjords. Salmon post-smolts caught in the trawl were examined for sea lice, biologically measured (wet weight, total and fork length) and then humanely killed, labelled with a unique identifier and frozen for storage. Any evidence of the presence of passive integrated transponder (PIT) tags was recorded. Trawl samples were then transported back to the Institute of Marine Research (IMR), Bergen, for further sampling. Further sampling consisted of collecting fin clips and scale tissue from each sample for genetic and age analysis. Where possible, PIT tags or external tags were recovered from fish. The fin clips were conserved in 100% ethanol. The number of samples per trawl varied among the fjord systems (Table 1).

Baseline samples

The baseline samples for each of the fjord systems originated from 4 main sources: (i) historical scale or tissue samples provided to IMR by Rådgivende Biologer, (ii) existing genetic

data from previously-genotyped river samples, (iii) samples stored as frozen whole fish at IMR, and (iv) samples collected by electrofishing during the summer of 2017. The number and source of baseline samples varied among the fjord systems (Table 1). In Boknafjord, 2045 samples were obtained from 19 rivers covering a period from 2006 to 2017 (Fig. 1). The number of samples per river ranged from 13 (Storelva i Sauda) to 234 (Espedalselva), with a mean of 114 samples per river. In Hardangerfjord, 884 samples were obtained from 10 rivers covering a period from 2011 to 2017 (Fig. 2). The number of samples per river ranged from 34 (Rosendalelvane) to 150 (Tysseelva), with a mean of 88 fish per river. In Sognefjord, 753 samples were obtained from 10 rivers covering a period from 2001 to 2017 (Fig. 3). The number of samples per river ranged from 57 (Nærøydalselva) to 119 (Mørkridselva), with an average of 75 fish per river. Whole fish were biologically measured (frozen weight) and fin clips and scale tissue were taken for genetic and age analysis.

Genotyping

All samples (trawl and baseline) were registered in a database and given a unique ID number. DNA analysis took place in the Molecular Biology laboratory at IMR in Bergen during the period 1 May 2017 – 1 September 2017. DNA was extracted from either fin clips or scale tissue in 96-well plates using the Qiagen DNeasy 96 Blood & Tissue Kit with two negative controls. In total, 31 microsatellite markers were amplified in five PCR multiplexes (details available on request). PCR products were resolved on an ABI 3730 Genetic Analyser and sized using a 500LIZ size standard (Applied Biosystems). Genemapper version 5.0 was used to score alleles manually. Scoring was quality checked before exporting the data for statistical analysis.

Existing genetic data consisted of samples that were previously genotyped at 18 microsatellite loci, and these individuals were further genotyped at 13 microsatellite loci as above. The monomorphic locus SsD486 was removed from the dataset, therefore, each individual was analysed at 30 loci.

Statistical Analysis

The total number of alleles and allelic richness of each river and trawl were calculated with Microsatellite Analyser (MSA) (Dieringer and Schlötterer, 2003). Pairwise F_{ST} and its significance were tested among baseline rivers within each fjord system using ARLEQUIN v.3.5.1.2 (Excoffier et al., 2005).

The evaluation of how well individual fish could be assigned back to each baseline was conducted with the Self-Assignment test in GeneClass 2 (Piry et al., 2004) using the Rannala

& Mountain (1997) method of computation with a significance threshold of $\alpha = 0.05$. The test sequentially removes one fish from the baseline and attempts to assign the fish to its most likely river of origin using the rest of the baseline (sampling without replacement). Fish are given an assignment score based on the likelihood of their belonging to the assigned river. Assignment accuracy was investigated visually by plotting the percentage of individuals successfully assigned to their river of origin and the percentage of individuals included at different assignment score thresholds (Fig. 4). This was performed for each fjord system. ONCOR (Kalinowski et al., 2007) was also used to conduct a mixture analysis using the baseline samples from each fjord system to estimate the stock composition of the trawl samples using conditional maximum likelihood (Fig. 5).

The assignment of individual trawl samples to their potential rivers of origin was conducted with GeneClass 2 (Piry et al., 2004) using the Rannala & Mountain (1997) method. Individuals were assigned using a combination of direct genetic assignment to a potential source river and exclusion from all potential source rivers within a fjord system with significance thresholds of $\alpha = 0.05$ and $\alpha = 0.001$. Individuals that were assigned to a river with an assignment score of above 70 were deemed to be correctly assigned. In order to ensure maximum accuracy, individual assignment was also carried out with ONCOR. It was decided to accept assignments as correct when both GeneClass and ONCOR agreed on the potential river of origin, and it is recommended that only fish with a Geneclass assignment score above the cut-off of 70 are to be accepted.

STRUCTURE v.2.3.4 (Pritchard et al., 2000) was used to identify possible genetic groups among rivers and trawls in each fjord system under a model assuming admixture and correlated allele frequencies without using population information. Ten runs with a burn-in period of 100 000 replicates and a run length of 1 000 000 Markov Chain Monte Carlo (MCMC) iterations were performed for clusters ranging from 1-13 (Boknafjord), 1-10 (Hardanger) and 1-10 (Sognefjord). STRUCTURE Harvester was then used to calculate the Evanno et al. (2005) ad hoc summary statistic ΔK, based on the rate of change of the estimated likelihood between successive K values, allowing the determination of the uppermost hierarchical level of stricture in the data. Runs were averaged with CLUMPP v.1.1.1 using the LargeK-Greedy algorithm and the G' pairwise matrix similarity statistic, and graphically displayed using bar plots. The structure analyses, together with the F_{ST} estimates among rivers and the level of miss-assignment among baseline rivers within each fjord system would allow for the inference of potential area groupings of certain rivers which may be genetically and geographically similar, and could thus be treated as a single assignment unit in further analyses.

Results

Genotyping

Individuals with more than 30% (9 loci) missing were excluded from the baseline data prior to any statistical analysis. The Boknafjord baseline consisted of 1883 individuals from 19 rivers, ranging from 13 (Storelva i Sauda) to 231 (Espedalselva) individuals per river with an average of 105 individuals per river. Within the Boknafjord trawl samples, 213 were genotyped and 201 were entered into Geneclass and ONCOR to be assigned back to their river of origin. The Hardangerfjord baseline consisted of 780 individuals from 10 rivers, ranging from 33 (Rosendalselvane) to 149 (Etneelva) individuals per river with an average of 88 individuals per river. Within the Hardangerfjord trawl samples, 291 were genotyped and 272 were entered into Geneclass and ONCOR to be assigned back to their river of origin. The Sognefjord baseline consisted of 580 individuals from 10 rivers, ranging from 23 (Mørkridselva) to 95 (Ytredalselva) individuals per river with an average of 58 individuals per river. The large drop in number of samples within the Sognefjord baseline was due to a high number of trout present in some river samples. Within the Sognefjord trawl samples, 226 were genotyped and 183 were entered into Geneclass and ONCOR to be assigned back to their river of origin For an overview of sample numbers per baseline and trawl see Table 1.

Summary Statistics and Self Assignment

Total number of alleles and allelic richness among the rivers in each baseline ranged from 192 and 6.43 to 410 and 8.11 in Boknafjord, 306 and 9.62 to 410 and 10.51 in Hardangerfjord, and from 265 and 8.03 to 386 and 9.87 in Sognefjord. Pairwise F_{st} among the rivers within the three fjord systems revealed significant differentiation between most rivers (Table 2), apart from between Jørpelandelva and several rivers that are located nearby in Boknafjord (Table 2A, Fig. 1) and between Daleelva and Vikja in Sognefjord (Table 2C).

The number of individuals correctly self-assigned back to river varied among the fjord systems. In Boknafjord, self-assignment accuracy averaged 53.3% overall, and varied between rivers from 16.13% (Jørpelandelva) to 61.54% (Suldalslågen) (Table 3A). For some rivers, there was a visible trend of frequent miss-assignment to rivers that were located nearby. For example, many rivers located in the south-eastern part of Boknafjord were incorrectly assigned to Espedalselva at levels above 10% (Table 3A). In Hardangerfjord, 519 out of 789 individuals were correctly assigned back to the baseline (65.8%), and correct self-assignment ranged from

24.24% (Rosendal) to 85.04% (Tysse). In general, rivers located further into Hardangerfjord miss-assigned more often to each other, and most of the south-eastern rivers were miss-assigned to Etneelva at levels above 10% (Table 3B). In Sognefjord, 349 out of 580 individuals were correctly assigned back to the baseline (60.2%), and correct self-assignment ranged from 34.78% (Mørkridselva) to 77.89% (Ytredalselva). Many individuals were incorrectly assigned to Daleelva and Nærøydalselva at levels above 10%, with no apparent trend among rivers located near to each other (Table 3C).

STRUCTURE, direct assignment and mixed stock analysis

Based on the STRUCTURE results, Evanno's test showed that ΔK was highest when K = 2 (28.1) and K = 5 (10.2) for Boknafjord (Fig 6A), indicating that two and five genetic clusters would fit the data best. Bar plots for 2 and 5 clusters are shown in Figure 7. At two clusters, there appears to be a weak gradient from Figgjo in the south-west moving north-east, with the exceptions of Høleåna, Førre and Suldalselva. In general, there is a similar mix of the two clusters within most of the rivers in Boknafjord. At five clusters, Høleåna, Førre and Suldalselva appear more distinct from the other rivers, while the rivers in the south-eastern part of Boknafjord display a similar mix of clusters, with some exceptions (Fig. 7B). The STRUCTURE results and the Evanno's test for Hardangerfjord indicated that ΔK was highest when K = 2 (81.21). At two clusters, all rivers display a similar mix apart from Tysse and Oselva (Fig 7C). The Evanno's table for Sognefjord indicated that ΔK was highest when K = 2 (8.17), K = 4 (3.82) and K = 9 (4.09). At K = 2, Ytredalselva and Aurlandselva displayed a more distinct mix of the clusters than the other rivers, with a weak geographical gradient visible from west to east (Fig. 6D). At K = 4, once again Ytredal and Aurland appear most different to the other rivers, and rivers on the northern bank of the fjord display a similar mix of clusters (Fig. 6E). At K = 9, the trends are not very clear, apart from the rivers located on the northern bank displaying more similar levels of clustering than the other rivers (Fig. 6F).

In Boknafjord, of the 201 trawl samples to be assigned back to their river of origin, the direct assignment for 158 individuals were in agreement between Geneclass and ONCOR (78.6%) and of those, 105 had a Geneclass assignment score above 70 (66.5%) (Table 4A). For Hardangerfjord, of the 272 trawl samples, the direct assignment between Geneclass and ONCOR were in agreement for 205 individuals (75.4%), of which 167 individuals had a Geneclass assignment score above 70 (81.5%) (Table 4B). In Sognefjord, of the 214 trawl samples assigned back to their river of origin, 192 (89.71%) were in agreement between

Geneclass and ONCOR; of which 157 (81.77%) had a Geneclass assignment score above 70 (Table 4C).

The mixture analysis for Boknafjord estimated that the stock composition of Boknafjord was dominated by Espedalselva and Suldalselva, while Hjelmeland, Håland, Storelva i Sauda and Vikedal were estimated to have little to no individuals within the stock (Fig. 5A). In Hardanger, the stock composition was estimated to be made up of 70% Etne fish, while Tysse and Oselva had little to no estimated contribution to the stock (Fig. 5B). The mixture analysis for Sognefjord estimated that the stock composition was made up of Daleelva Høyanger, Nærøydalselva and Lærdalselva fish, while the other rivers contributed far less to the stock (Fig. 5C).

In the Boknafjord trawl samples, 10 of the 213 fish had external tags. All the tags were from fish released from the River Imsa, which was not present in the baseline. Geneclass assigned 7 of these fish to Høleåna, while ONCOR assigned 8 out of 10 to Høleåna, which is located very close to Imsa (Table 5). In the Hardangerfjord trawl samples, PIT tags were found inside 27 of the 291 fish. One PIT tag was from a fish released from Eidfjord, 5 were from fish released from Guddal, and 21 were from Etne releases. Neither Geneclass nor ONCOR assigned the Eidjford fish correctly, and Guddal was not present in the baseline. Of the PIT tagged Etne fish, Geneclass assigned 13 correctly, while ONCOR assigned 20 correctly (Table 5). In the Sognefjord trawl samples, PIT tags were found inside 9 of the 226 fish., all releases from Årøyelva. Both Geneclass and ONCOR assigned 3 of the PIT tagged fish correctly back to Årøyelva (Table 5).

Discussion

Boknafjord

In Boknafjord, the non-significant F_{ST} values and high levels of self miss-assignment between rivers located close to each other highlights the potential of grouping the fjord system into areas for assignment rather than basing assignment on rivers alone. For certain rivers within Boknafjord, the STRUCTURE results and bar plots further support a grouping of rivers into areas (Fig 6). Rivers within the south-eastern part of Boknafjord exhibited a similar level of cluster mixing and could be treated as a single area unit in future genetic analyses. However, this trend was not apparent among all rivers located near to each other. When K = 5, Høleåna appeared to consist of individuals that were distinct from the nearest rivers (Fig. 6B), although it is possible that these individuals were fish originating from the River Imsa, located close to Høleåna, which was not represented in our baseline. This assumption is reinforced by the fact

that external tags recovered from several trawl fish showed that these fish were from the River Imsa, and 7 of the 10 tagged fish were assigned back to Høleåna by both Geneclass and ONCOR. Anecdotal accounts also confirm that the level of straying between Høleåna and Imsa is high. A siblingship analysis using COLONY 2.06.4 (Jones and Wang, 2010) found several related individuals, including a family of 11 members, therefore relatedness between individuals may also explain the unique clustering observed in Høleåna. When K = 5, three other rivers in Boknafjord displayed distinct clustering compared to the rest: Førre, Suldal and Sauda (Fig. 6B). It is possible that the difference observed in Suldal is caused by stocking, which is carried out annually, while it is unknown whether stocking occurs in Førre or Sauda (Anonymous, 2017). Siblingship analysis for Suldalselva revealed several families of related individuals, although families were not larger than 4 individuals. Siblingship analysis for Førre revealed several related individuals, including a family of 12 members. The sample size for Sauda was very small (n = 13), and a siblingship analysis found that several of the individuals were related, potentially inflating their apparent differences compared to other rivers.

Hardangerfjord

In Hardangerfjord, although all rivers exhibited significantly different F_{ST} values, there were high levels of miss-assignment between rivers located in the inner part of the fjord, and between rivers located nearby each other in the south-eastern part of the fjord (Table 3B). There was also a trend of many rivers miss-assigning to Etne, and the level of direct assignment agreement between Geneclass and ONCOR was lowest (75.4%) in the Hardangerfjord, potentially due to that ONCOR directly assigned over 20% more fish to Etne than Geneclass (Table 4B). This potential bias for/against assignment to Etne is shown in the results for the assignment of the PIT tagged Etne fish, which differed largely between Geneclass and ONCOR (Table 5). Etne was also estimated to be the largest contributing river to the simulated stock mixture (Fig. 5B), which is not unexpected as the estimated spawning biomass of the River Etne is the largest of the Hardangerfjord rivers (Anonymous, 2017). The STRUCTURE results and bar plots indicated that Tysseelva and Oselva displayed distinct clustering compared to the other rivers (Fig. 6C) and were estimated to contribute almost nothing to the stock mixture (Fig. 5B), which is intuitive based on their locations within the fjord compared to the other baseline rivers (Fig. 2). There has been no reported stocking within these two rivers (Anonymous, 2017).

Sognefjord

In Sognefjord, although non-significant F_{ST} values were observed between some rivers located near to each other (Table 2C), in general there appeared to be no real pattern to the level of miss-assignment among rivers (Table 3C). All rivers displayed some levels of miss-assignment to Daleelva i Høyanger. Rivers that are located further into the fjord had less individuals miss-assigned to them, apart from Lærdalsevla (Table 3C). The level of agreement of direct assignment between Geneclass and ONCOR was highest in Sognefjord (89.7%), although the number of PIT tagged individuals that were correctly assigned back to their river was low (Table 5). The STRUCTURE results highlighted distinct clustering for Ytredalselva and Aurlandselva (Fig.6). Aurlandselva is stocked yearly, while no stocking takes place in Ytredalselva (Anonymous, 2017), and an investigation into putative sibling relationships showed that both rivers contained some related individuals, potentially contributing to the difference observed in clustering between these rivers and the other baseline rivers in Sognefjord.

Caveats and Recommendations

Accurate genetic assignment depends on the reliability and completeness of the baseline, and the level of genetic differences between populations. The completeness of the baselines in the present data varied across fjord systems, therefore it is recommended that these results be treated as preliminary. Further sampling of existing and absent baseline rivers is required within each of the fjord systems in order to increase the accuracy of the genetic assignment. The potential of the present marker set to accurately assign individuals back to their natal rivers, can only be fully assessed when the baseline sample sets have been completed beyond what was available for this pilot study. Such analyses may reveal that a higher number of markers may be required, and that the marker sets may have to be adapted to different fjord systems. Markers that differentiate well in one fjord system, may not necessarily work as well in others. Further work to complement the baseline, and the sampling strategy, is planned for the next year. In this context, it will also be important to assess the temporal stability of marker frequencies in small rivers where a small number of effective breeders may result in genetic drift and fluctuating frequencies. Further analysis may also show that assigning individuals to clusters of populations may be a viable approach where inter-river genetic differences are too small.

The discrepancy in direct assignment between the two chosen assignment programs used in this pilot study highlights a level of uncertainty within the results, and it was chosen to report the river of origin of those individuals where the two programs were in agreement and

where the Geneclass assignment score was above 70. For the baseline rivers where siblings were detected, it is recommended that future analyses conduct putative sib-ship tests before any other analyses, and to remove all but a chosen number of full siblings from each river sample. Studies recommend that either one (Ozerov et al., 2017) or two (Olafsson et al., 2014) full sibs per family are retained with baseline river samples (but see Waples and Anderson, 2017).

Acknowledgements

This project is funded by the Norwegian Ministry of Trade, Industry and Fisheries. We would like to thank everyone who contributed to the project, including those who collected the trawl samples (coordinated by Ørjan Karlsen), contributed baseline samples (Rådgivende Biologer & IMR), helped to sample the fish at IMR and contributed to the genotyping in the laboratory.

References

ANONYMOUS 2017. Vedleggsrapport med vurdering av måloppnåelse for de enkelte bestandene. *Rapport fra Vitenskapelig råd for lakseforvaltning.* Trondheim.

DIERINGER, D. & SCHLÖTTERER, C. 2003. MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes,* 3**,** 167-169.

EVANNO, G., REGNAUT, S. & GOUDET, J. 2005. Detecting the number of clusters of individuals usinf the software structure: a simulation study. *Molecular Ecology,* 14**,** 2611 - 2620.

EXCOFFIER, L., LAVAL, G. & SCHNEIDER, S. 2005. Arlequin (version 3.0): An integrated software package for populatino genetics data anlaysis. *Evolutionary Bioinformatics Online,* 1**,** 47-50.

JONES, O. R. & WANG, J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources,* 10**,** 551 - 555.

KALINOWSKI, S. T., MANLOVE, K. R. & TAPER, M. L. 2007. ONCOR: a computer program for genetic stock identificatino. Department of Ecology, Montana State Unversity.

OLAFSSON, K., PANPOULIE, C., HJORLEIFSDOTTIR, S., GUDJONSSON, S. & HREGGVIDSSON, G. O. 2014. Present-day genetic structure of Atlantic salmon (*Salmo salar*) in Icelandic rivers and ice-cap retreat models. *PLOS One,* 9**,** 1 - 12.

OZEROV, M., VÄHÄ, J. P., WENNEVIK, V., NIEMELA, E., SVENNING, M.-A., PRUSOV, S., FERNANDEZ, R. D., UNNELAND, L., VASEMAGI, A., FALKEGÅRD, M., KALSKE, T. & CHRISTENSEN, B. 2017. Comprehensive microsatellite baseline for genetic stock identification of Atlantic salmon (*Salmo salar* L.) in northermost Europe. *ICES Journal of Marine Science***,** 1 - 11.

PIRY, S., ALAPETITE, A., CORNUET, J.-M., PAETKAU, D., BAUDOUIN, L. & ESTOUP, A. 2004. GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity,* 95**,** 536 - 539.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155**,** 945-959.

RANNALA, B. & MOUNTAIN, J. L. 1997. Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA,* 94**,** 536 - 539.

WAPLES, R. S. & ANDERSON, E. C. 2017. Purging putative siblings from population genetic data sets: a cautionary view. *Molecular Ecology,* 26.

Tables

Table 1: Summary data for the three fjord systems.

Fjord system	River Name	No. Genotyped	No. Samples	No. alleles	Allelic richness	Year sampled	Sample type	Stage	Source	Stocking
Boknafjord	Figgjo	89	89	388	7.96	2006	DNA	juvenile	IMR	Yes
	Høleåna	103	92	309	6.96	2017	fin clip	juvenile	IMR	No
	Forsandåna	108	98	343	7.45	2017	fin clip	juvenile	IMR	No
	Espedalselva	234	231	435	8.01	2009, 2016	DNA & scale	juvenile & adult	IMR & RB	Historical
	Dirdalselva	115	110	410	8.11	2015, 2016	scale	adult	RB	No
	Frafjordelva	76	76	375	8.01	2009	DNA	juvenile	IMR	Historical
	Lyseelva	67	66	351	7.73	2009	DNA	juvenile	IMR	No
	Jørpeland	33	31	296	7.76	2016	scale	adult	RB	Histprical
	Årdalselva	204	196	403	7.88	2009, 2017	DNA & fin clip	juvenile	IMR & RB	Yes
	Hjelmeland	105	105	357	7.56	2017	fin clip	juvenile	IMR	No
	Vormo	194	176	406	7.89	2009, 2014	DNA & scale	juvenile & adult	IMR & RB	No
	Førre	132	106	339	7.41	2014, 2015	scale	juvenile	RB	No
	Ulla	121	100	372	7.85	2014, 2015	scale	juvenile	RB	Yes
	Hålandselva	47	45	322	7.78	2016	scale	adult	RB	No
	Suldalslågen	213	171	381	7.53	2010, 2007	DNA & scale	juvenile	IMR & RB	Yes
	Storelva i Sauda	13	13	192	6.43	2017	fin clip	juvenile	IMR	No
	Vikedalselva	82	82	341	7.58	2006	DNA	juvenile	IMR	No
	Rødneelva	109	96	371	7.87	2017	fin clip	juvenile	IMR	No
Hardangerfjord	Oselva	98	73	332	9.64	2015, 2016	scale & fin clip	juvenile	IMR & RB	Historical
	Tysse	150	127	326	9.19	2014, 2015	scale	juvenile	RB	No
	Steindal	60	60	346	10.39	2017	fin clip	juvenile	IMR	Yes
	Eidfjord	118	99	358	9.88	2014, 2017	fin clip	juvenile	IMR & RB	Yes
	Opo	58	56	328	9.91	2014	fin clip	juvenile	RB	Yes
	Æneselv	43	35	306	9.62	2014	scale	juvenile	RB	No
	Rosendal	34	33	320	10.33	2017	fin clip	juvenile	IMR	Yes
	Omvik	76	65	335	9.90	2011, 2017	fin clip	juvenile	IMR & RB	No
	Uskedal	98	92	366	10.10	2017	fin clip	juvenile	IMR	No
	Etne	149	149	410	10.51	2013	DNA	adult	IMR	Yes
Sognefjord	Ytredalselva	106	95	332	8.74	2017	fin clip	juvenile	IMR	No
	Daleelva	98	87	386	9.73	2001/2002	DNA	adult	IMR & RB	Yes
	Vikja	64	61	375	9.87	2006/2008	DNA	juvenile	IMR & RB	Yes
	Sogndalselva	75	45	297	8.72	2017	fin clip	juvenile	IMR	No
	Årøyelva	81	34	295	9.10	2007	DNA	adult	IMR & RB	Yes
	Nærøydal	57	56	317	8.83	2003/2004	DNA	adult	IMR & RB	No
	Flåmselva	74	68	336	8.86	2007	DNA	juvenile	IMR & RB	No
	Aurland	100	61	290	8.03	2007	fin clip	juvenile	IMR	Yes

	Lærdalselva	51	50	313	8.78	2017	fin clip	juvenile	IMR	Yes
	Mørkridselva	119	23	265	8.16	2017	fin clip	juvenile	IMR	No
Boknafjord	Trawl	213	201	436	8.06	2017	fin clip	smolt	IMR	
Hardangerfjord	Trawl	291	272	451	10.78	2017	fin clip	smolt	IMR	
Sognefjord	Trawl	226	183	427	9.56	2017	fin clip	smolt	IMR	

Table 2: Pairwise FST (below diagonal) and P-values after 10000 permutations (above diagonal) for each fjord system. A: Boknafjord; B: Hardangerfjord; C: Sognefjord.

A	Figgjo	Høleåna	Forsand	Espedal	Dirdal	Frafjord	Lyse	Jørpeland	Årdal	Hjelmeland	Vormo	Førre	Ulla	Håland	Suldals	Sauda	Vikedal	Rødne
Figgjo	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Høleåna	0.017	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Forsand	0.016	0.022	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Espedal	0.014	0.017	0.009	-	**0.016**	**0.000**	**0.000**	0.763	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Dirdal	0.012	0.019	0.010	0.001	-	**0.000**	**0.000**	0.986	**0.010**	**0.000**	**0.000**	**0.000**	**0.000**	**0.039**	**0.000**	**0.000**	**0.000**	**0.000**
Frafjord	0.011	0.017	0.010	0.003	0.003	-	**0.000**	0.378	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Lyse	0.017	0.022	0.012	0.005	0.006	0.006	-	0.083	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Jørpeland	0.007	0.010	0.007	0.000	0.000	0.000	0.002	-	0.907	**0.000**	0.136	**0.001**	0.836	**0.000**	**0.000**	**0.000**	**0.003**	**0.031**
Årdal	0.009	0.016	0.009	0.001	0.001	0.002	0.004	-0.002	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.006**	**0.000**	**0.000**	**0.000**	**0.000**
Hjelmeland	0.015	0.026	0.016	0.010	0.012	0.012	0.019	0.009	0.012	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Vormo	0.010	0.019	0.011	0.006	0.005	0.006	0.010	0.001	0.005	0.011	-	**0.000**	**0.000**	**0.009**	**0.000**	**0.000**	**0.000**	**0.000**
Førre	0.025	0.032	0.018	0.009	0.010	0.013	0.011	0.009	0.010	0.024	0.012	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Ulla	0.009	0.017	0.008	0.003	0.002	0.004	0.006	-0.002	0.002	0.013	0.002	0.009	-	0.063	**0.000**	**0.000**	**0.000**	**0.000**
Håland	0.008	0.019	0.012	0.005	0.002	0.006	0.009	0.011	0.002	0.005	0.002	0.013	0.002	-	**0.000**	**0.000**	**0.000**	**0.000**
Suldals	0.026	0.034	0.018	0.012	0.012	0.014	0.014	0.012	0.012	0.026	0.016	0.019	0.012	0.015	-	**0.000**	**0.000**	**0.000**
Sauda	0.029	0.048	0.032	0.031	0.029	0.031	0.031	0.028	0.027	0.042	0.031	0.045	0.026	0.032	0.043	-	**0.000**	**0.000**
Vikedal	0.017	0.027	0.013	0.007	0.008	0.007	0.011	0.005	0.008	0.017	0.009	0.015	0.008	0.013	0.022	0.036	-	**0.000**
Rødne	0.012	0.018	0.012	0.006	0.007	0.006	0.010	0.003	0.007	0.014	0.008	0.018	0.008	0.009	0.019	0.035	0.012	-

B	Oselva	Tysse	Steindal	Eidfjord	Opo	Aneselv	Rosendal	Omvik	Uskedal	Etne
Oselva	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Tysse	0.017	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Steindal	0.019	0.016	-	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**
Eidfjord	0.017	0.015	0.004	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Opo	0.019	0.018	0.007	0.012	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Aneselv	0.015	0.014	0.006	0.012	0.010	-	**0.005**	**0.001**	**0.000**	**0.009**
Rosendal	0.018	0.015	0.008	0.008	0.011	0.007	-	**0.000**	**0.000**	**0.001**
Omvik	0.017	0.018	0.010	0.015	0.013	0.007	0.008	-	**0.000**	**0.000**
Uskedal	0.014	0.013	0.007	0.010	0.011	0.007	0.008	0.010	-	**0.000**
Etne	0.008	0.013	0.008	0.011	0.009	0.003	0.005	0.008	0.005	-

C	Ytredal	Daleelva	Vikja	Sogndal	Årøyelva	Nærøydal	Flåmselva	Aurland	Lærdal	Mørkrids
Ytredal	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Daleelva	0.011	-	0.067	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Vikja	0.010	0.002	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**	**0.017**
Sogndal	0.027	0.012	0.010	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Årøyelva	0.030	0.021	0.013	0.023	-	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Nærøydal	0.021	0.008	0.007	0.014	0.025	-	**0.000**	**0.000**	**0.001**	**0.002**
Flåmselva	0.030	0.017	0.015	0.017	0.027	0.011	-	**0.000**	**0.000**	**0.000**
Aurland	0.049	0.039	0.039	0.054	0.071	0.047	0.048	-	**0.000**	**0.000**
Lærdal	0.021	0.010	0.005	0.014	0.018	0.007	0.015	0.050	-	0.105
Mørkrids	0.018	0.013	0.005	0.020	0.023	0.009	0.019	0.054	0.003	-

Table 3: Percentage of the baseline samples that were assigned to each river within each fjord system by the Self-Assignment test in Geneclass in order to evaluate the accuracy of the baselines for each fjord system. The diagonal (in bold) represents the percentage of individuals that were correctly self-assigned by the program. The tables should be read from left to right. A: Boknafjord; B: Hardangerfjord; C: Sognefjord.

A	Figgjo	Høleåna	Forsand	Espedals	Dirdals	Frafjord	Lyse	Jørpeland	Årdalselva	Hjelmeland	Vormo	Førre	Ulla	Håland	Suldals	Sauda	Vikedal	Rødne
Figgjo	**61.80**	3.37	1.12	3.37	4.49	3.37	0.00	0.00	6.74	1.12	6.74	0.00	5.62	0.00	0.00	0.00	0.00	2.25
Høleåna	3.26	**71.74**	1.09	5.43	1.09	1.09	3.26	2.17	4.35	0.00	2.17	0.00	0.00	0.00	1.09	0.00	0.00	3.26
Forsand	0.00	0.00	**76.53**	4.08	3.06	1.02	1.02	0.00	3.06	1.02	6.12	0.00	1.02	0.00	0.00	0.00	1.02	2.04
Espedals	1.30	1.30	1.73	**36.36**	10.39	9.52	4.33	1.30	13.85	2.16	4.76	1.30	4.76	1.30	1.73	0.00	3.46	0.43
Dirdals	0.00	0.00	0.00	21.82	**37.27**	4.55	2.73	0.00	12.73	0.91	4.55	0.91	4.55	1.82	2.73	0.00	2.73	2.73
Frafjord	3.95	1.32	1.32	17.11	9.21	**28.95**	3.95	0.00	10.53	1.32	6.58	0.00	5.26	1.32	2.63	0.00	6.58	0.00
Lyseelva	3.03	0.00	4.55	15.15	4.55	7.58	**39.39**	0.00	6.06	4.55	6.06	0.00	0.00	1.52	3.03	0.00	0.00	4.55
Jørpeland	0.00	6.45	0.00	12.90	12.90	9.68	3.23	**16.13**	6.45	0.00	12.90	0.00	6.45	3.23	0.00	0.00	3.23	6.45
Årdalselva	1.53	0.00	0.51	16.84	10.71	5.61	3.06	0.00	**39.80**	1.53	6.12	1.02	5.10	2.04	0.51	0.00	2.55	3.06
Hjelmeland	4.76	0.00	0.95	6.67	2.86	0.95	1.90	1.90	2.86	**62.86**	6.67	0.00	1.90	1.90	0.00	0.00	0.00	3.81
Vormo	2.27	0.57	1.14	8.52	8.52	2.27	1.14	1.14	5.11	2.27	**51.70**	0.57	5.68	2.27	0.57	0.00	2.27	3.98
Førre	0.94	0.00	0.00	3.77	1.89	1.89	0.94	0.00	3.77	0.00	7.55	**71.70**	2.83	2.83	0.94	0.00	0.00	0.94
Ulla	9.00	0.00	1.00	7.00	5.00	2.00	0.00	4.00	7.00	0.00	8.00	0.00	**49.00**	1.00	3.00	0.00	2.00	2.00
Håland	6.67	0.00	0.00	6.67	2.22	6.67	2.22	0.00	6.67	6.67	11.11	2.22	6.67	**35.56**	2.22	0.00	0.00	4.44
Suldals	0.58	0.00	0.58	4.09	4.09	2.34	2.34	0.58	2.34	0.00	1.75	0.00	1.75	0.58	**76.61**	0.00	1.75	0.58
Sauda	0.00	0.00	0.00	7.69	0.00	0.00	0.00	0.00	7.69	7.69	7.69	0.00	7.69	0.00	0.00	**61.54**	0.00	0.00
Vikedal	1.22	0.00	1.22	4.88	4.88	6.10	2.44	2.44	4.88	0.00	4.88	0.00	0.00	0.00	1.22	0.00	**63.41**	2.44
Rødneelva	3.13	0.00	1.04	5.21	3.13	6.25	2.08	1.04	3.13	1.04	5.21	1.04	4.17	2.08	2.08	0.00	0.00	**59.38**

B	Oselva	Tysse	Steindal	Eidfjord	Opo	Aneselv	Rosendal	Omvik	Uskedal	Etne
Oselva	**71.23**	2.74	2.74	2.74	0.00	1.37	0.00	1.37	2.74	15.07
Tysse	2.36	**85.04**	2.36	0.79	0.79	2.36	0.79	0.00	3.15	2.36
Steindal	1.67	3.33	**48.33**	21.67	6.67	3.33	1.67	5.00	3.33	5.00
Eidfjord	0.00	1.01	15.15	**72.73**	1.01	1.01	0.00	2.02	4.04	3.03
Opo	1.79	0.00	16.07	7.14	**44.64**	3.57	7.14	1.79	5.36	12.50
Æneselv	2.86	2.86	11.43	2.86	2.86	**28.57**	2.86	5.71	14.29	25.71
Rosendal	0.00	6.06	3.03	3.03	3.03	6.06	**24.24**	6.06	12.12	36.36
Omvik	0.00	1.54	1.54	1.54	0.00	10.77	1.54	**67.69**	7.69	7.69
Uskedal	1.09	1.09	2.17	5.43	1.09	3.26	6.52	2.17	**65.22**	11.96
Etne	2.68	0.67	2.01	3.36	4.03	4.03	0.67	2.01	6.04	**74.50**

C	Ytredal	Daleelva	Vikja	Sogndal	Årøyelva	Nærøydal	Flåmselva	Aurland	Lærdal	Mørkrid
Ytredalselva	**77.89**	12.63	7.37	0.00	1.05	0.00	0.00	0.00	1.05	0.00
Daleelv	5.75	**55.17**	24.14	2.30	2.30	6.90	0.00	1.15	1.15	1.15
Vikja	3.28	31.15	**39.34**	1.64	3.28	8.20	8.20	0.00	4.92	0.00
Sogndalselva	2.22	8.89	6.67	**68.89**	4.44	4.44	0.00	0.00	4.44	0.00
Årøyelva	0.00	11.76	5.88	5.88	**61.76**	11.76	2.94	0.00	0.00	0.00
Nærøydalselva	0.00	10.71	7.14	0.00	3.57	**55.36**	1.79	0.00	21.43	0.00
Flåmselva	0.00	8.82	4.41	0.00	0.00	11.76	**69.12**	2.94	2.94	0.00
Aurland	0.00	18.03	4.92	0.00	3.28	1.64	3.28	**67.21**	1.64	0.00
Lærdalselva	0.00	8.00	10.00	2.00	4.00	18.00	8.00	0.00	**48.00**	2.00
Mørkridselva	0.00	17.39	4.35	0.00	0.00	21.74	8.70	0.00	13.04	**34.78**

Table 4: Direct assignment comparisons of the trawl samples between Geneclass and ONCOR for the three fjord systems. A: Boknafjord; B: Hardangerfjord; C: Sognefjord. In agreement; the number of individuals directly assigned to each river by both Geneclass and ONCOR, >70; the number of individuals in agreement and where Geneclass had an assignment score above 70.

	Geneclass		ONCOR		In Agreement		>70	
A	Number	%	Number	%	Number	%	Number	%
Figgjo	15	7.46	17	8.46	14	8.86	11	10.38
Høleåna	11	5.47	12	5.97	11	6.96	10	9.43
Forsand	6	2.99	3	1.49	3	1.90	3	2.83
Espedal	31	15.42	55	27.36	31	19.62	13	12.26
Dirdal	14	6.97	13	6.47	10	6.33	4	3.77
Frafjord	13	6.47	6	2.99	5	3.16	3	2.83
Lyse	6	2.99	5	2.49	5	3.16	3	2.83
Jørpeland	3	1.49	3	1.49	3	1.90	3	2.83
Årdal	16	7.96	15	7.46	11	6.96	4	3.77
Hjelmeland	4	1.99	0	0.00	0	0.00	0	0.00
Vormo	14	6.97	16	7.96	14	8.86	7	6.60
Førre	2	1.00	1	0.50	1	0.63	1	0.94
Ulla	10	4.98	11	5.47	7	4.43	5	4.72
Håland	6	2.99	0	0.00	0	0.00	0	0.00
Suldals	42	20.90	43	21.39	42	26.58	37	34.91
Sauda	0	0.00	0	0.00	0	0.00	0	0.00
Vikedal	3	1.49	0	0.00	0	0.00	0	0.00
Rødne	5	2.49	1	0.50	1	0.63	1	0.94
Total	201		201		158		105	
B								
Oselva	5	1.84	0	0.00	0	0.00	0	0.00
Tysse	3	1.10	0	0.00	0	0.00	0	0.00
Steindal	14	5.15	6	2.21	6	2.93	5	2.99
Eidfjord	22	8.09	16	5.88	15	7.32	14	8.38
Opo	15	5.51	7	2.57	6	2.93	4	2.40
Aneselv	14	5.15	6	2.21	6	2.93	6	3.59
Rosendal	9	3.31	2	0.74	2	0.98	2	1.20
Omvik	16	5.88	6	2.21	6	2.93	5	2.99
Uskedal	34	12.50	28	10.29	24	11.71	19	11.38
Etne	140	51.47	201	73.90	140	68.29	112	67.07
Total	272		272		205		167	
C								
Ytredal	10	4.67	9	4.21	9	4.69	9	5.73
Daleelva	48	22.43	57	26.64	48	25.00	37	23.57
Vikja	28	13.08	28	13.08	26	13.54	21	13.38
Sogndal	11	5.14	6	2.80	6	3.13	6	3.82
Årøyelva	16	7.48	14	6.54	13	6.77	12	7.64
Nærøydal	39	18.22	46	21.50	39	20.31	32	20.38
Flåmselva	13	6.07	7	3.27	7	3.65	7	4.46
Aurland	2	0.93	0	0.00	0	0.00	0	0.00
Lærdal	43	20.09	46	21.50	43	22.40	32	20.38
Mørkrids	4	1.87	1	0.47	1	0.52	1	0.64
Total	214		214		192		157	

Table 5: Summary of the comparison of direct assignment for trawl fish with external or PIT tags between Geneclass and ONCOR.

				Number assigned correctly			
Fjord system	**Number of tags**	**Tag type**	**Origin of tags**	**Geneclass**	**ONCOR**	**Number in agreement**	**Comments**
Boknafjord	10	External tags	Imsa	0 (7 assigned to Høleåna)	0 (8 assigned to Høleåna)	7	Imsa was not in the baseline
Hardangerfjord	1	PIT tags	Eidfjord	0	0	1	Assigned to Æneselva
Hardangerfjord	5	PIT tags	Guddal	0	0	4	Guddal was not in the baseline
Hardangerfjord	21	PIT tags	Etne	13	20	14	
Sognefjord	9	PIT tags	Årøyelva	3	3	3	

Some of the fish captured in the trawls had external or PIT tags, and thus had known origins therefore we were able to test the accuracy of the genetic assignment by comparing the direct assignment results to the tags.

Figures

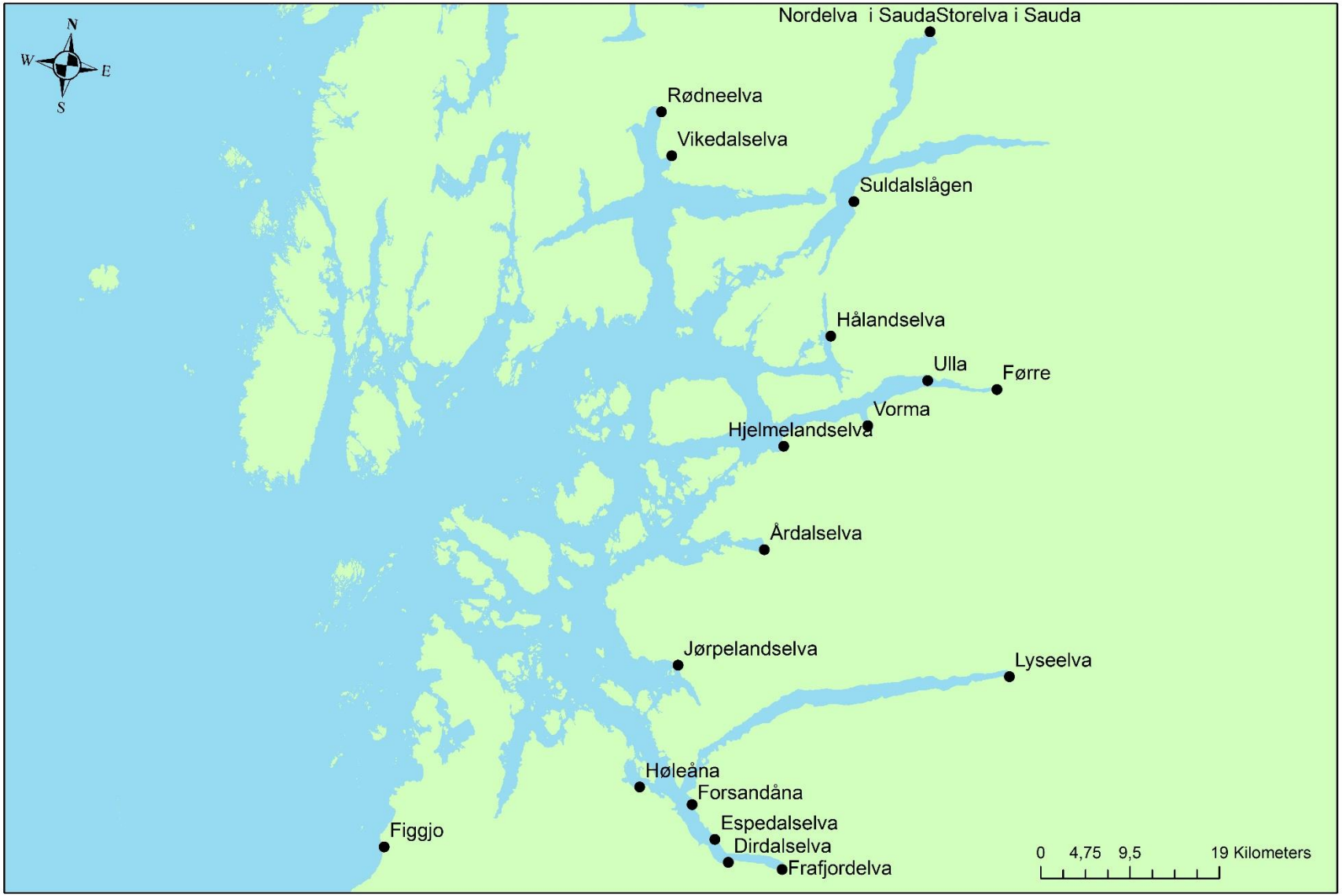


Figure 1. Map of Boknafjord indicating the rivers included in the baseline.

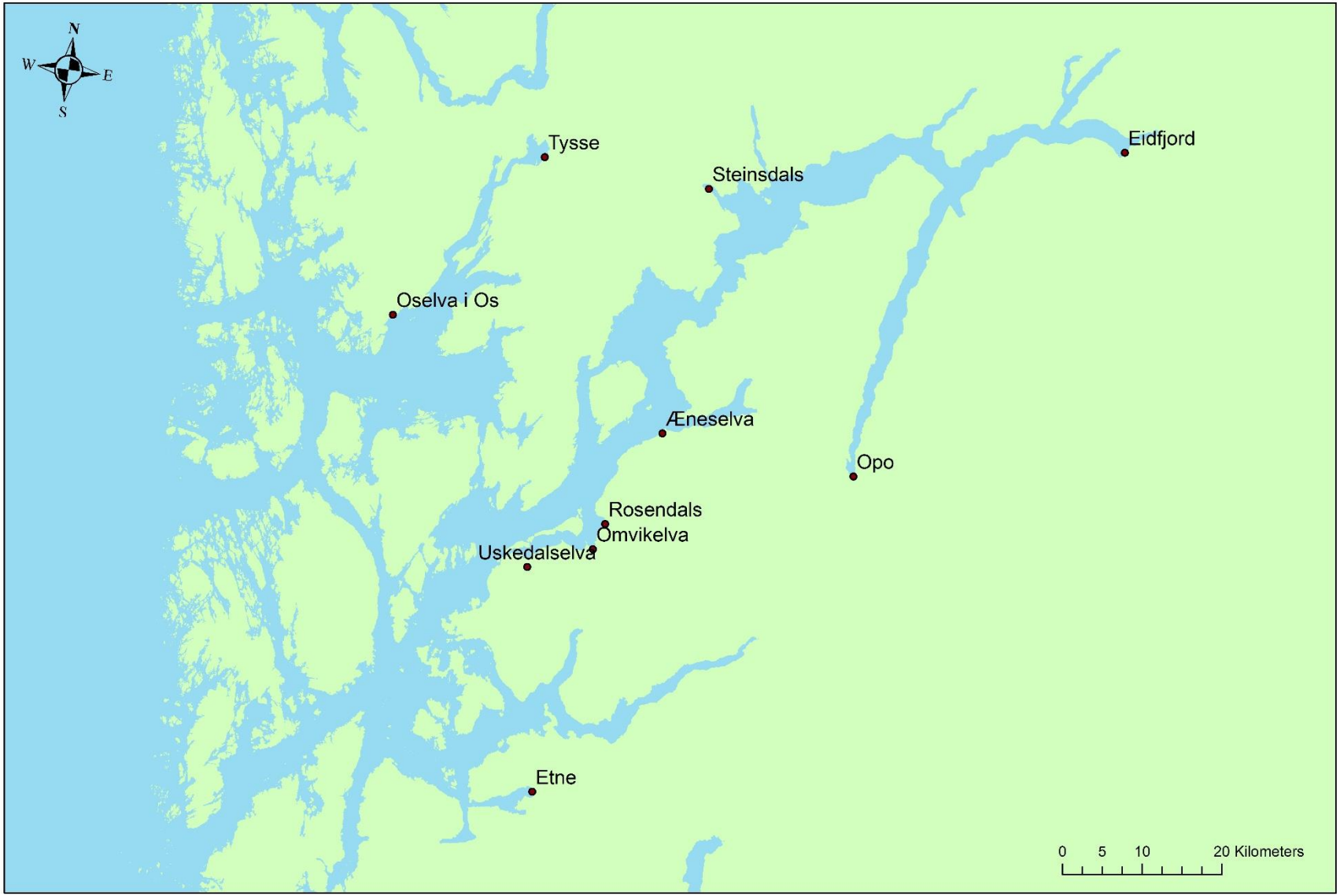


Figure 2. Map of Hardangerfjord indicating the rivers included in the baseline.

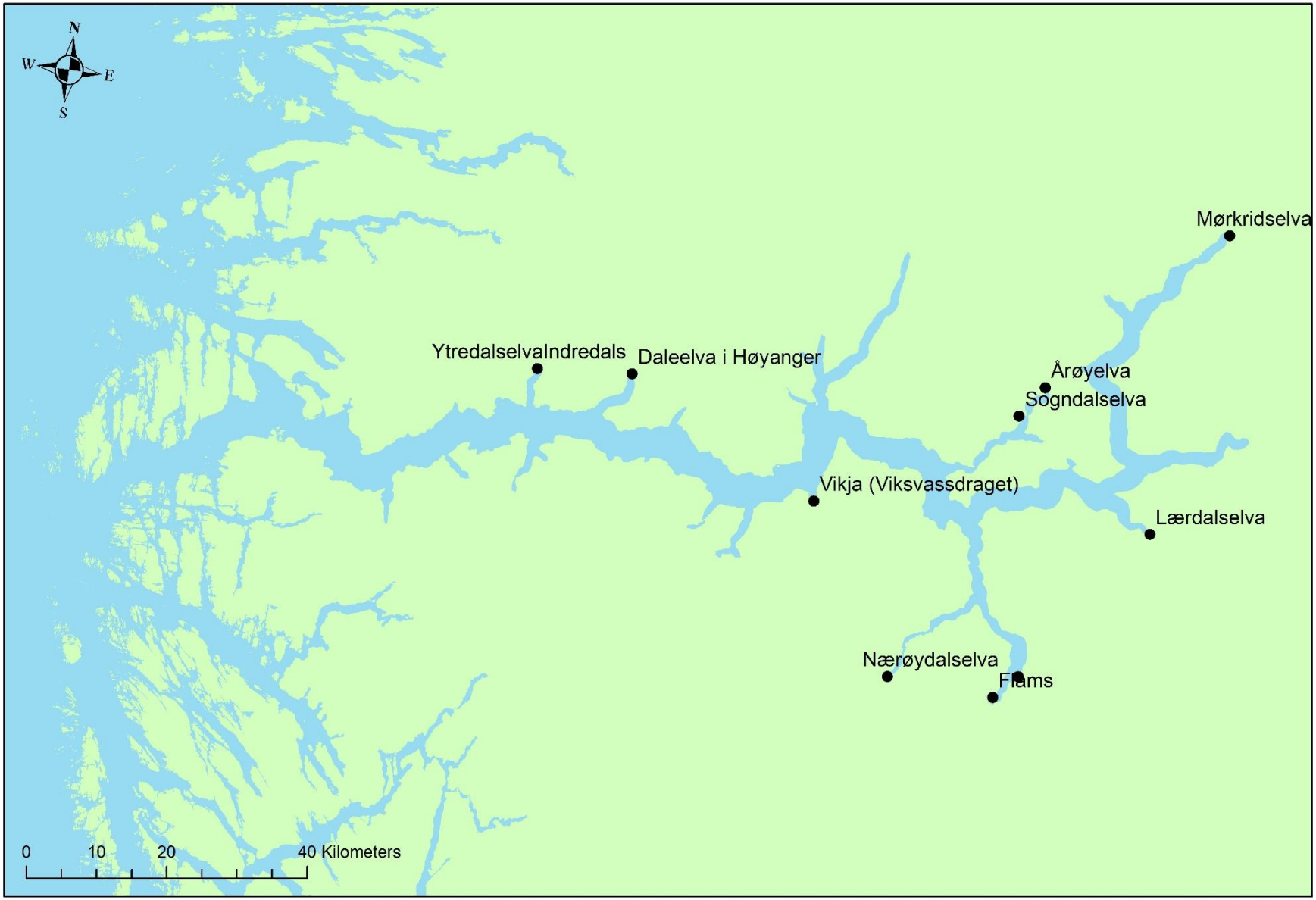


Figure 3. Map of Sognefjord indicating the rivers included in the baseline.

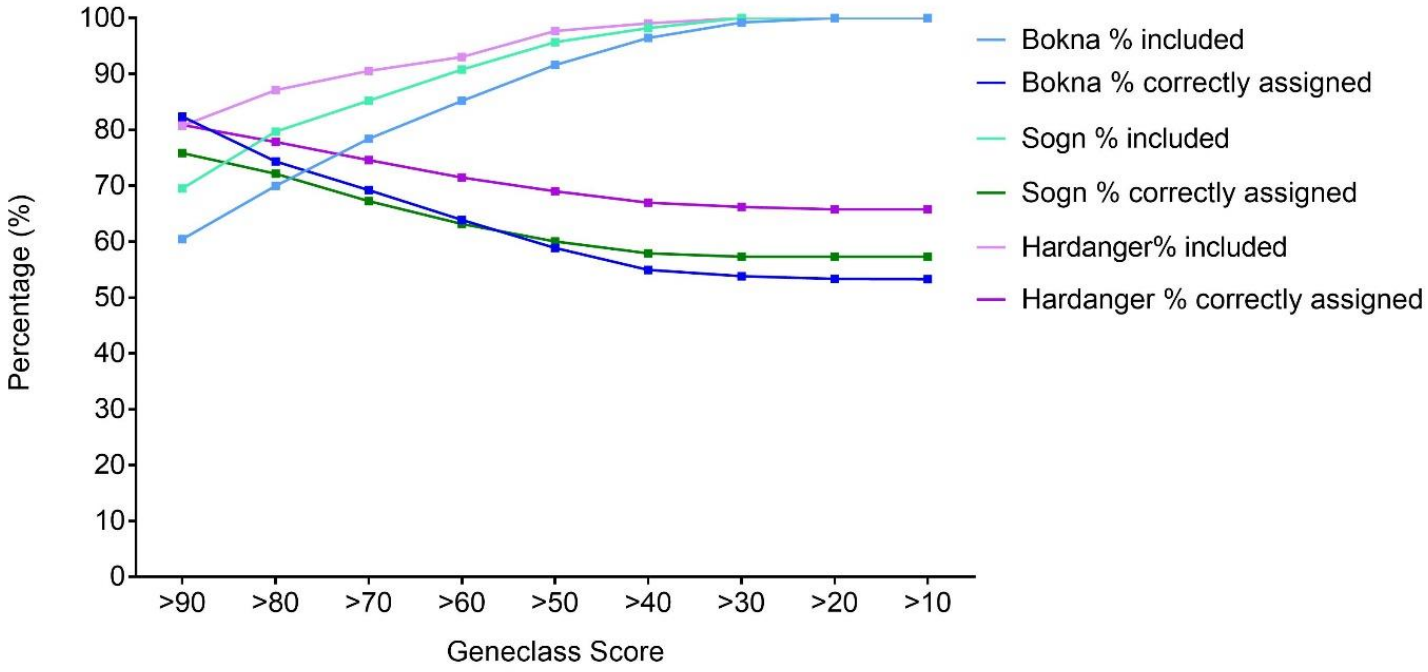


Figure 4: Percentage of individuals included in the baseline and percentage of individuals correctly assigned to the baseline as the cut-off for the assignment score is changed for each fjord system using the results from the Self-Assignment test in Geneclass. A cut-off score of 70 was chosen for the direct assignment of trawl individuals.

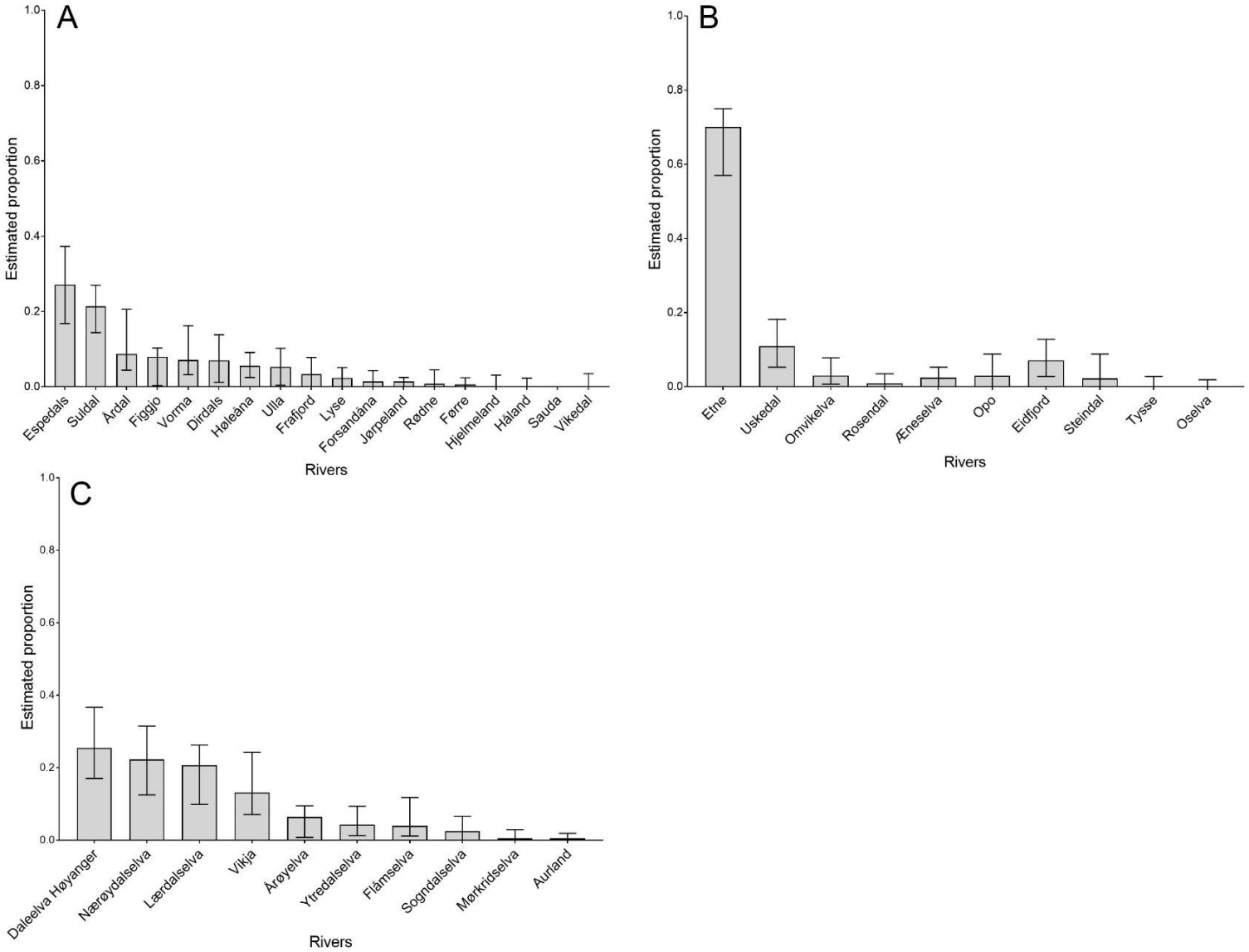
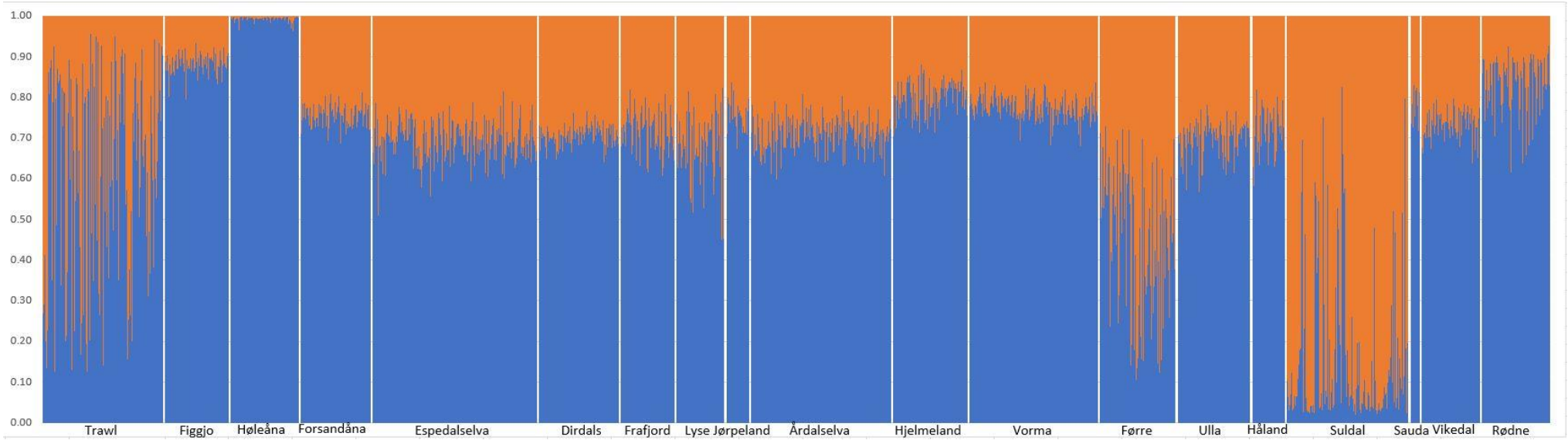
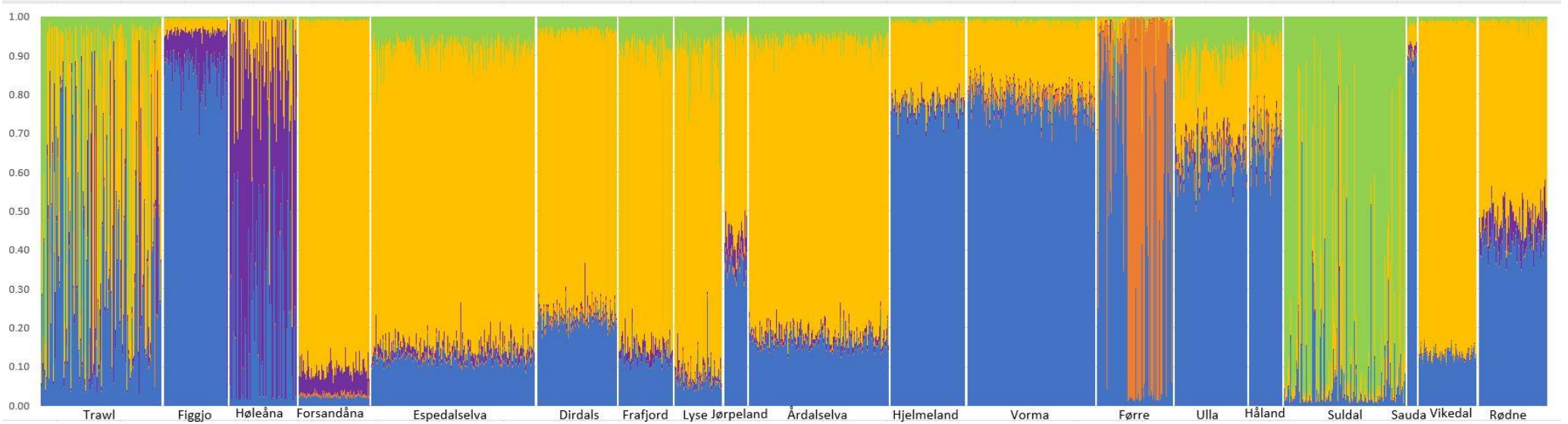


Figure 5: The mixed fishery stock composition of the trawl samples from each fjord system estimated by ONCOR using the baseline samples. A: Boknafjord; B: Hardangerfjord; C: Sognefjord.

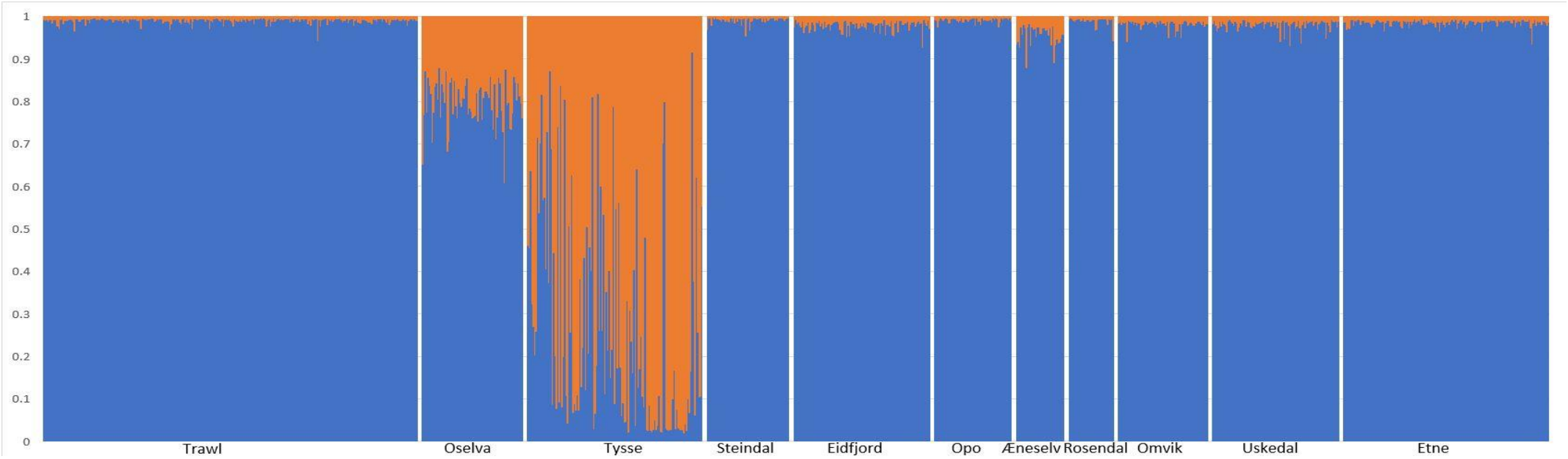
A


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Trawl
Figgjo
Høleåna
Forsandåna
Espedalselva
Dirdals
Frafjord
Lyse Jørpeland
Årdalselva
Hjelmeland
Vorma
Førre
Ulla
Håland
Suldal
Sauda
Vikedal
Rødne

B


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Trawl
Figgjo
Høleåna
Forsandåna
Espedalselva
Dirdals
Frafjord
Lyse Jørpeland
Årdalselva
Hjelmeland
Vorma
Førre
Ulla
Håland
Suldal
Sauda
Vikedal
Rødne

C


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Trawl
Oselva
Tysse
Steindal
Eidfjord
Opo
Æneselv
Rosendal
Omvik
Uskedal
Etne

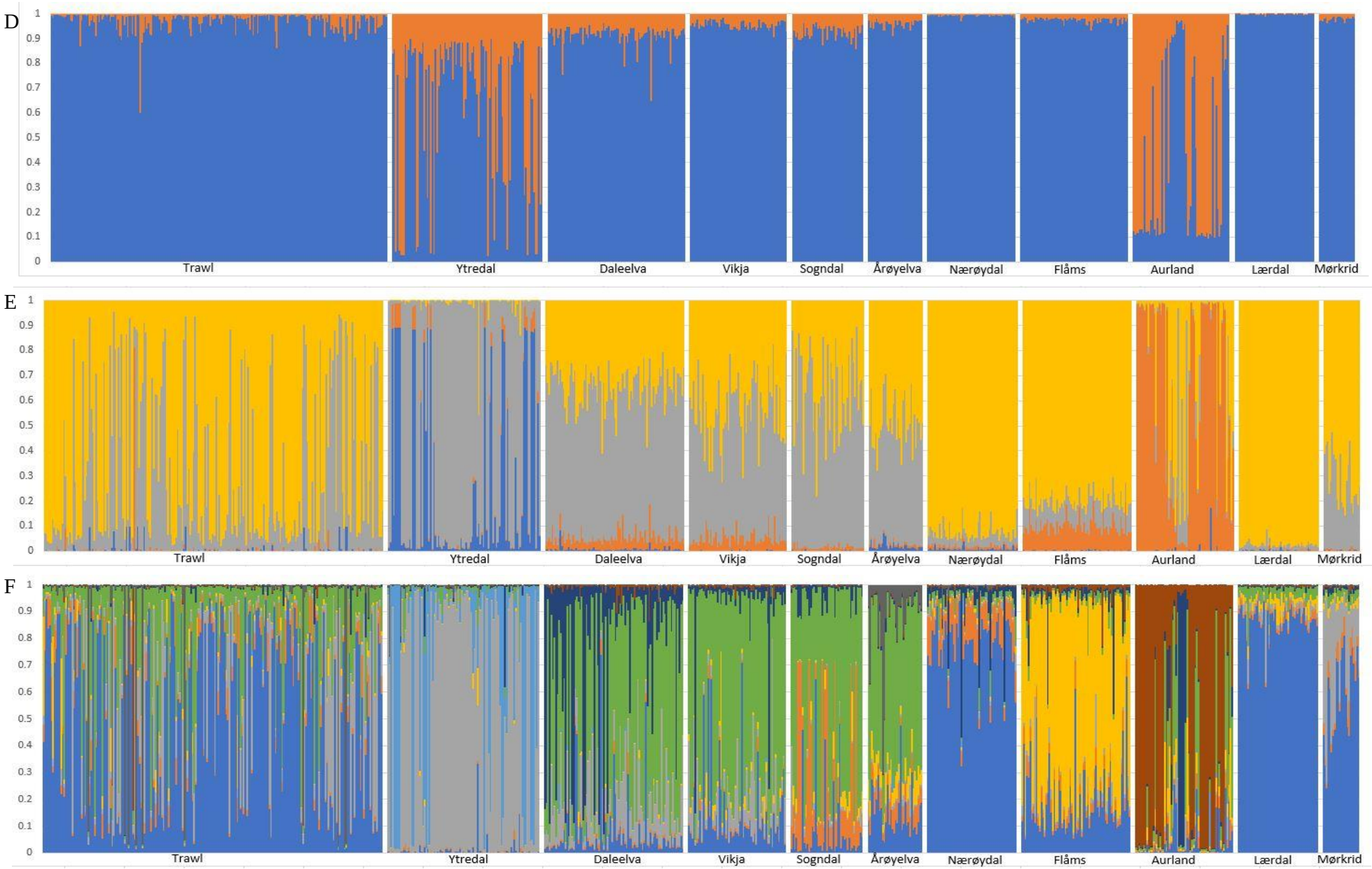


Figure 6: Bar plot clusters for Boknafjord where K = 2 (A), K = 5 (B), Hardangerfjord where K = 2 (C), and Sognefjord where K = 2 (D), K = 4 (E), K = 9 (F).