

Fauna crime: elucidating the potential source and introduction history of European smelt (*Osmerus eperlanus* L.) into Lake Storsjøen, Norway

Mari Hagenlund¹ · Kjartan Østbye^{1,2} · Kjell Langdal¹ · Marius Hassve¹ · Ruben A. Pettersen² · Eric Anderson³ · Finn Gregersen⁴ · Kim Præbel⁵

Received: 23 September 2014 / Accepted: 20 April 2015 / Published online: 1 May 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract The ability to accurately determine the original source of invading species offers several powerful applications in invasive species ecology and management and may enable important information on the invading species in its native habitat. Lake Storsjøen in South-Central Norway was recently found to have been subjected to an illegal translocation of the European smelt (*Osmerus eperlanus*). The main aim of this study was to infer the most likely source (s) of the invading smelt by using microsatellite markers, and subsequently to infer its introduction history. The results indicated that the smelt is most likely a result of introduction from the large Lake Mjøsa, and that the translocated smelt comprise a large number of individuals. The smelt in Lake Storsjøen showed no significant genetic

bottleneck effect. However, a corresponding significant test for a recent population expansion indicates that the smelt has had a high reproductive success and population growth in its new environment. The results from this study illustrate the usefulness of applying multilocus genetic markers for inferring origin of translocated populations, demographic events and introduction histories comprising an effective tool for assessment of invasive species.

Keywords Bayesian inference · Bottleneck effects · Fauna crime · Microsatellites · Population expansion · Source populations · Translocation

Introduction

Identifying the source of invaders is a valuable tool in fauna crime and wildlife forensics, e.g. knowledge of possible introduction routes to reduce further introductions (Geller et al. 2010). Wildlife DNA forensic methods have primarily been used as a means to identify the species of collected evidence in wildlife crime (e.g. Linacre (2009)). However, the expanding field of genetic methods and genetic markers (e.g. microsatellites and SNP's) offer a wide array of related applications in fauna crime related questions (Ogden et al. 2009; Alacs et al. 2010; Geller et al. 2010; Ogden et al. 2013; Johnson et al. 2014).

There has also been an advance in the field of statistical inference with regards to interpreting patterns from genetic markers (Hansen et al. 2001; Beaumont and Rannala 2004; Drummond et al. 2005). Bayesian inference methods utilizing e.g. microsatellites provide an effective tool for natural scientists (Beaumont et al. 2002; Stauffer, 2008; Stephens and Balding 2009) allowing for statistical genetic assignment and identification of a given individual to

Electronic supplementary material The online version of this article (doi:10.1007/s10592-015-0724-2) contains supplementary material, which is available to authorized users.

✉ Mari Hagenlund
hagenlundmari@gmail.com; kjartan.ostbye@hihm.no
Kim Præbel
kim.praebel@uit.no

- ¹ Faculty of Applied Ecology and Agricultural Science, Hedmark University College, Evenstad, 2418 Elverum, Norway
- ² Center for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway
- ³ Fisheries Ecology Division, Southwest Fisheries Science Center, Santa Cruz, CA, USA
- ⁴ Aage Håkonsens vei 1D, 3610 Kongsberg, Norway
- ⁵ Department of Arctic and Marine Biology, Faculty for Bioscience, Fisheries and Economy, UIT, The Arctic University of Norway, 9037 Tromsø, Norway

putative source populations (Pearse and Crandall 2004). Such methods are useful in identifying indigenous and introduced individuals (Primmer et al. 2000), and have been extensively used in a number of convictions, e.g. regarding illegal salmon fishing and trade (Withler et al. 2004). Furthermore, genetic software have been developed to infer the past demographic history (Pybus et al. 2000; Heled and Drummond 2008; Guillemaud et al. 2009), making it theoretically possible to infer the most likely number of translocated individuals from one source population to a new locality (Anderson and Slatkin 2007).

Lake Storsjøen in Rendalen municipality, South-Central Norway, was recently discovered to have been exposed to a translocation event of the European smelt (*Osmerus eperlanus* L.) from an unknown source (County Governor of Hedmark 2011). European smelt (hereafter, smelt) is an osmerid species native to Norway, but has not previously been observed in Lake Storsjøen (Museth et al. 2008). It was first discovered by local fishermen in Lake Storsjøen in 2008 (Strømsoen 2008), but the exact time of translocation is unknown.

Norwegian law prohibits the translocation of any freshwater-species, both alive and as baitfish, to localities where they have not previously been known to inhabit (Innlandsfiskeloven [law relating to salmonids and freshwater fish 2014]: Omsetnings- og sykdomsforskriften for akvatiske dyr [Sale and disease regulation for aquatic animals, 2008]). This study has the main objective of identifying the most likely source of the introduced smelt, and to get an insight into the introduction history to Lake Storsjøen. To achieve this, genetic samples from several potential source locations were compared at 15 microsatellite loci against samples from the introduced smelt in Lake Storsjøen. By testing microsatellite data using multiple inference programs, we aim to pinpoint the most likely source population(s) and illustrate the value of population genetics as a tool in wildlife forensics.

Based on the likely assumption that the smelt in Lake Storsjøen was illegally translocated by humans either intentionally, or by accident when using smelt as bait, the following hypotheses were tested, that; (1) *translocation of smelt occurred from a locality in geographic proximity to Lake Storsjøen*, (2) *the translocation of smelt to Lake Storsjøen occurred from only one source location*, (3) *the translocated population will exhibit a reduced genetic diversity compared to the source populations due to a limited number of individuals translocated, i.e. founder effects*.

Material and methods

The European smelt is a small fish species of the family Osmeridae. It is widely distributed in the north east Arctic coastal waters, from the White- and Barents seas in the

north to Garonne estuary in France (Kottelat and Freyhof 2007). In Norway the smelt is naturally distributed in the South-Eastern part, mainly in large lakes (Sandlund and Næsje 2000).

As the main aim of this study was to identify the most likely origin of the translocated smelt in Lake Storsjøen a sub-set of smelt populations were selected from the complete number of existing populations in order to test the specific hypotheses. A priori prediction suggests that the most closely situated smelt population is the most likely founder, where the source population is probably large and publicly well-known, corresponding well with Lake Mjøsa that has a large population of smelt with several well-known spawning locations. Thus, two selected spawning locations from Lake Mjøsa were considered to be a likely source. Secondly, a set of six more southerly distributed smelt populations at an increasing distance from Lake Storsjøen were selected (Fig. 1; Table 1). A smelt population from Lake Vänern in Southeastern Sweden was selected to be used as an outgroup for polarizing genetic assignments geographically. The large Lake Vänern and the ancient Lake Ancylus have likely been important with regard to colonization of the freshwater fishes in Norway (Borgstrøm 2000). In addition, one locality in Lake Mjøsa was sampled twice in different years (Mjøsa Lågen 2009, 2011, Table 1) offering an opportunity to compare temporal samples and thus test the assignment ability of the different softwares when using temporal samples from the same locality (see methods and details below). In total, 416 smelt from 10 localities were collected between 2009 and 2012 in Norway and Sweden and used in the analyses (Table 1; Fig. 1). Sampling was performed by both gill-net fishing and by dip-netting at spawning sites until a minimum number of 20–30 smelt were obtained from each location, a number of individuals assumed to be sufficient for population genetic assignment analyses when assuming a high genetic differentiation between populations (Cornuet et al. 1999; Hansen et al. 2001). The sampled smelt were immediately either frozen, or preserved in containers with 96 % ethanol (EtOH). In the laboratory, fin clips from pectoral fins were taken, and fin clips were preserved in individually marked 2 ml Eppendorf tubes.

DNA extraction and PCR

DNA extraction was performed at the fish genetics lab at Tromsø University using the E-Z 96 Tissue DNA kit following the manufacturers procedures (OMEGA Bio-tek) followed by Nanodrop quantification of DNA quality. A total of 15 microsatellite loci (see Supp. Table 5) were optimized. Subsequent polymerase chain reaction (PCR) was conducted in two multiplex panels. The PCR products were separated on an ABI 3130 XL Automated Genetic

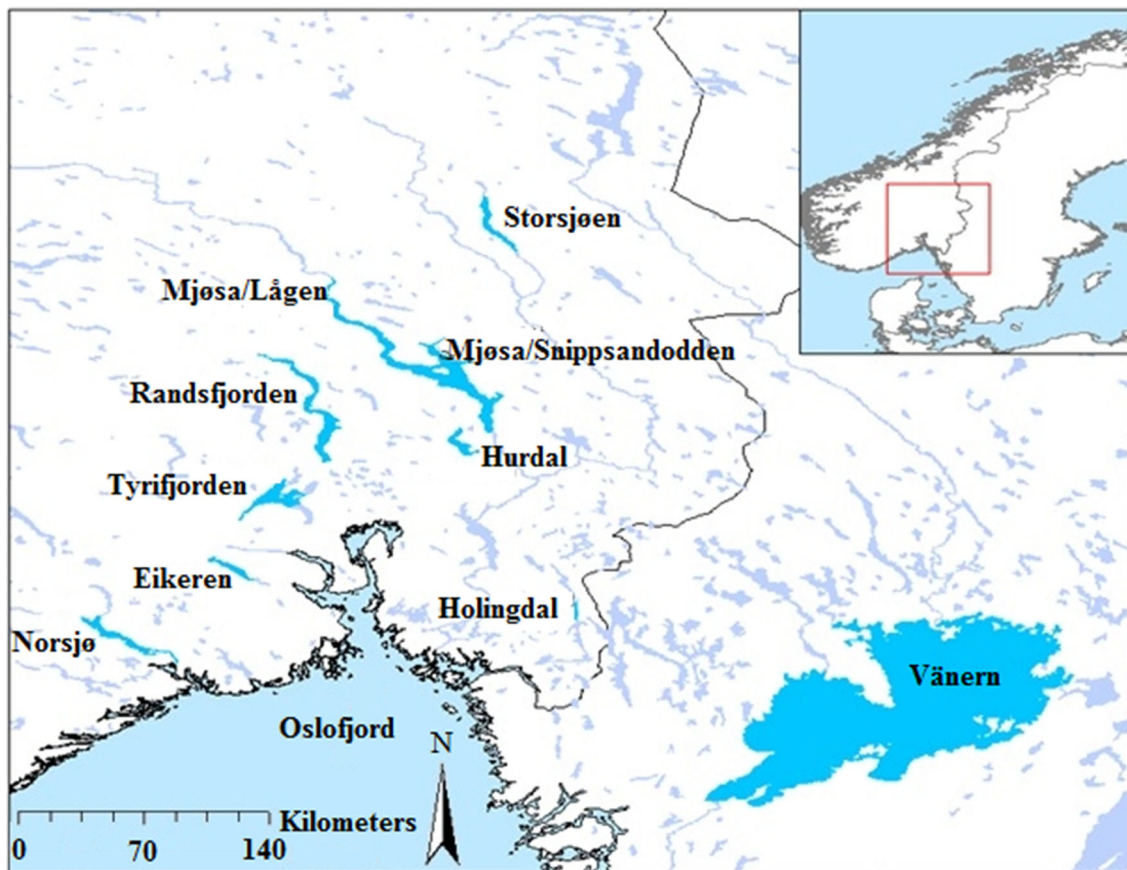


Fig. 1 The ten sampling locations of smelt (*Osmerus eperlanus*) in Norway and Sweden. Lake Storsjøen is the translocated smelt population. The map was created in ArcGIS Version10.1 (ESRI 2012)

Analyzer (Applied Biosystems) and alleles scored in the GENEMAPPER 3.7 software (Applied Biosystems). The scoring was verified twice by visual inspection. Replicate samples (5–10 samples per population) were also included as a part of the quality assessment of the dataset. One locus (*M-Omo4*) was excluded due to poor amplification, leaving 14 microsatellites for further analyses.

Data analysis

The software MICRO-CHECKER 2.2.3 (Van Oosterhout et al. 2004) was used to check for genotyping errors, followed by the program FREENA (Chapuis and Estoup 2007; Chapuis et al. 2008) that corrects for allele-frequency bias. There was no systematic occurrence of homozygote excess within loci across populations, and no significant differentiation in F_{st} when comparing uncorrected and corrected loci.

LOSITAN (Beaumont and Nichols 1996; Antao et al. 2008) was run to test if loci were candidates for directional or balancing selection. Analyses were run with 100 000

simulations under the “Force mean F_{st} ”, and “Neutral mean F_{st} ” alternatives. One locus (*Oep539*) was candidate to directional selection. Putative effects of directional selection at *Oep539* were tested for influence on the genetic assignment tests by both including and excluding the locus in the following analyses. However, no difference in assignment was detected when removing the candidate locus, thus the locus *Oep539* was included in the subsequent analyses.

GENEPOP 4.0 (Raymond and Rousset 1995; Rousset 2008) was used to check for deviations from Hardy–Weinberg equilibrium and linkage disequilibrium (LD (Guo and Thompson 1992)). False discovery rate (FDR) corrections (Pike 2011) was used to adjust p-values for multiple tests. The results showed that significant deviations were found in only one locus (*Oep384*, $p < 0.0005$) in one population (Lake Mjøsa/Lågen-11) after FDR corrections (threshold, $p < 0.0005$). This was in concordance with the results from MICRO-CHECKER, and the deviation was possibly due to null alleles. Significant LD was not discovered in any tests following FDR

Table 1 Description of the sampled smelt (*Osmerus eperlanus*) locations including: locality, population codes, latitude and longitude, area of the lake (km²), mean depth of the lake (m), year sampled, size range of the smelt (cm), mean length of the smelt (cm), sample size (N) and method of collection

Locality	Code	N	E	Area (km ²)	Mean depth (m)	Year sampled	Smelt size range (cm)	Mean length (cm)	Sample size (N)	Collection method
Introduced population:										
1. Lake Storsjön	Sto	61°40'32	11°12'48	47.55	309	2011/2012	9.3–18.8	12.2	47	Gill-nets
Natural populations:										
2a. Lake Mjøsa, Lågen 2011	Lag11	61°04'49	10°16'06	365	453	2011	8.5–22.7	12.6	60	Gill-nets and dip-nets
2b. Lake Mjøsa, Lågen 2009	Lag09	61°04'49	10°16'06	365	454	2009	8.6–23	12.4	26	Gill-nets and dip-nets
3. Lake Mjøsa, Snippsandodden	MjN	60°29'10	10°35'03	365	455	2009	8.6–25	13	40	Gill-nets and dip-nets
4. Lake Eiker	Eik	59°25'34	09°32'58	27.72	156	2012	8.6–17.5	10	40	Gill-nets
5. Lake Hurdalssjøen	Hur	60°21'04	11°03'47	33.16	60	2012	10.8–15.3	12	32	Gill-nets
6. Lake Øymarksjøen, Hølingdal	Hol	59°17'16	11°24'25	14.13	37.6	2011	8.8–18.5	10.7	40	Dip-nets
7. Lake Norsjø, Gvarvelva	Nor	59°23'27	09°12'09	55.48	176	2012	7.9–9.3	8.5	40	Dip-nets
8. Lake Randsfjorden, Odnes	Ran	60°48'54	10°11'15	140.06	131	2011	8.5–12.7	10.8	40	Gill-nets and trawl
9. Lake Tyrifjorden, Breitenlandet	Tyr	60°04'08	10°05'21	138.56	330	2010	9.3–12.7	10.8	40	Gill-nets
10. Lake Vänern: N. Dalbosjön	Van	59°02'20	12°51'54	5648	106	2012	9.4–22.2	14.1	40	Gill-nets

Localities 2 through 9 are the putative natural founder populations sampled in Norway, while locality 10, Lake Vänern was sampled in Sweden. Locality 1, Lake Storsjön is the introduced smelt population

correction. Thus, only one locus, *Oep384* was removed, and a total of 13 loci were used in the following genetic analyses.

Genetic diversity and population differentiation

Genetic diversity estimates were calculated by the use of several different software; Expected heterozygosity (H_{exp}), and genetic divergence between populations (pairwise F_{st}) was calculated by the use of GENEPOP 4.0 (Raymond and Rousset 1995; Rousset 2008; Kalinowski 2004), GENALEX 6.5 (Peakall and Smouse 2012), and Fstat 2.9.3.2 (Goudet 1995 Weir and Cockerham 1984). Standardized private allelic richness (A_p) and standardized allelic richness (A_r) accounting for differences in sample size, was calculated with HP-RARE 1.0 (Kalinowski 2005) with rarefaction using 36 genes (i.e. the minimum gene number across samples). Pairwise Wilcoxon rank sum tests were performed in R (R Core Team 2014) to check for significant differences in A_p , A_r and H_{exp} between the introduced population and the potential source populations.

Population assignment

Different Bayesian inference programs may give deviating results (Frantz et al. 2009). Thus, the comparison of several softwares may lead to a stronger support for assignment of individuals, and minimize the risk of bias. Here, two different Bayesian assignment software were applied; STRUCTURE 2.3.2 (Pritchard et al. 2000), and GENECLASS2 (Piry et al. 2004). STRUCTURE was run with an admixture model using 100,000 burn-in steps, and 100,000 Markov Chain Monte Carlo repetitions with 10 iterations using the LOCPRIOR function which incorporates geographic sampling information as recommended by (Hubisz et al. 2009), as well as a hierarchical approach following the recommendation by (Evanno et al. 2005). The most likely number of clusters (based on $\text{LnP}(K)$ and ΔK) was determined using STRUCTURE HARVESTER (Earl and Vonholdt 2012). The software GENECLASS2 (Piry et al. 2004) was used to exclude or assign reference groups as possible sources, i.e. to determine which groups are likely source populations, and to significantly exclude unlikely sources (Pearse and Crandall 2004). This was done by using all the different criteria available for calculation; Bayesian, allele frequency, and distance based. Bayesian and frequency based approaches in this software have the advantage that they do not assume that the source population is among the sampled populations. This gives the possibility of asking if the “true” source population is among the sampled populations, rather than asking which population has the highest likelihood as a potential source,

and to significantly exclude unlikely sources (Pearse and Crandall 2004). Initial tests, using all criteria, were performed with GENECLASS2 by testing all spawning populations with known origin against all the potential source populations to determine the power, or consistency of this analysis. All spawning populations were consistently assigned back to their known origin (100 % for assignment of groups, 77.8–99.8 % for assignment of individuals), indicating a high power of the analysis. All computations were executed with an assignment threshold of $p < 0.01$.

Phylogenetic analysis

The program POPULATIONS 1.2.30 (Langella 1999) was used to create phylogenetic rooted neighbor-joining trees. The trees were created with bootstrap values from 100 permutations using the Nei’s standard distance (Nei 1972), Nei’s DA distance ((Nei et al. 1983) Supp. Fig. 7)) and Cavalli-Sforza and Edwards distance method. Results are shown with the Cavalli-Sforza and Edwards distance as this method assume that genetic differentiation occurs due to genetic drift, and do not assume that population size remains constant (Cavalli-Sforza and Edwards 1967). As the translocated smelt in Lake Storsjøen most likely consisted of a limited number of individuals (where random genetic drift may be influential), this method seemed to be the most appropriate. The tree was visualized using TreeView32 (Page 1996) using the Swedish population Vänern as a geographical outgroup/root.

Genetic analyses for demographic events

BOTTLENECK (Piry et al. 1999) was used to evaluate if the translocated smelt individuals in Lake Storsjøen have gone through a bottleneck at the time of release. BOTTLENECK was run with 1000 iterations for all the three mutation models (stepwise mutation, infinite alleles, and two-phased model (TPM (recommended by the authors))), and with all statistical tests. Evaluation of the Wilcoxon sign-rank test under the TPM-model was given most emphasis as this test was recommended by the authors.

In order to get an estimate of the approximate number of individuals that was transferred from the most likely source population into Lake Storsjøen two different packages were used; COLONIZE (Mergeay et al. 2007), and COALIT/NFCONE (Anderson and Slatkin, 2007). COLONIZE was run 10 independent times with rare allele correction, maximum 10,000 colonizers, 10 batches, and 100 randomizations. The program calculates probabilities for maximum and minimum, as well as a joint probability value (joint probability for minimum and maximum

colonizers), for potential number of colonizers. The COLONIZE program estimate the probability that a simulated founder event of a certain size would result in at least as many alleles as observed amongst the actual colonized population. The COALIT/NFCONE package uses a Monte Carlo approximation to the likelihood that allows for estimation of number of founding individuals (or chromosomes) by calculating maximum likelihood estimates, and upper and lower support limits that correspond to a confidence interval (Anderson and Slatkin 2007). The software was run using the input of the source and translocated populations, under a wide range of values of intrinsic growth rate (r) and carrying capacity (K) to establish how the analysis was affected by the different assumptions of r and K , which are not well known for smelt. At r values above 2.0, further increase in r produced only negligible changes in the estimated number of founders. The intrinsic growth rate parameters; 0.5, 1.0, 2.0, 3.0, and the levels of carrying capacities; 50,000, 250,000, 500,000, 1000,000, 5000,000 diploid individuals, were therefore used for the final analysis. These scenarios are assumed to capture the range of likely demographic scenarios of the smelt during invasion. The scenario of intrinsic rate of increase values of 3, was only run under values of $K = 50,000$ and $250,000$ due to extensive run-times.

To test for a demographic population expansion event of the smelt in Lake Storsjøen, the k -, and g -test of (Reich et al. 1999) implemented in KG-TEST was applied. The g -test significance level was checked according to the recommended cutoff values in Table 1 (p. 455) reported by (Reich et al. 1999). Most emphasis will be put on the k -test as the k -test has a maximum sensitivity for detecting expansions that happened within a few generations, while the

g -test is more suitable for detecting expansions that happened further in the past (Donnelly et al. 2001).

Results

Genetic diversity and population differentiation

A total of 155 alleles were observed in the 11 populations, across the 13 loci. Standardized private allelic richness (A_p) ranged from 0.04 to 0.92, but there was no statistical significant difference between any of the lakes (Wilcoxon rank test: $p > 0.05$). Lake Eikeren, Hurdal, Norsjø, Randsfjorden and Storsjøen exhibited the lowest private allelic richness and Lake Vänern exhibited the highest A_p . Lake Storsjøen had a slightly lower A_p (0.15) than the Lake Mjøsa populations; Lake Mjøsa/Snippsandodden (0.24), Lake Mjøsa/Lågen-11 (0.26), and Lake Mjøsa/Lågen-09 (0.29, Fig. 2a).

Allelic richness varied between 2.92 and 5.98 across populations, but was not significantly different between any of the populations (Wilcoxon rank test: $p > 0.05$). Lake Storsjøen exhibited an allelic richness of 5.00, relatively similar to Lake Mjøsa/Lågen-09 (5.06), Lake Mjøsa/Lågen-11 (5.20) and Lake Mjøsa/Snippsandodden (4.97, Fig. 2b). The populations exhibiting the lowest allelic richness were the westerly distributed populations; Lake Eikeren (2.92), Lake Norsjø (3.08) and Lake Tyrifjorden (3.22), while Lake Vänern exhibited the highest allelic richness of 5.98 (Fig. 2b).

The mean expected heterozygosity (H_{exp}) ranged from 0.28 to 0.51, with statistically significant differences between some population pairs (Wilcoxon rank test, Table 2);

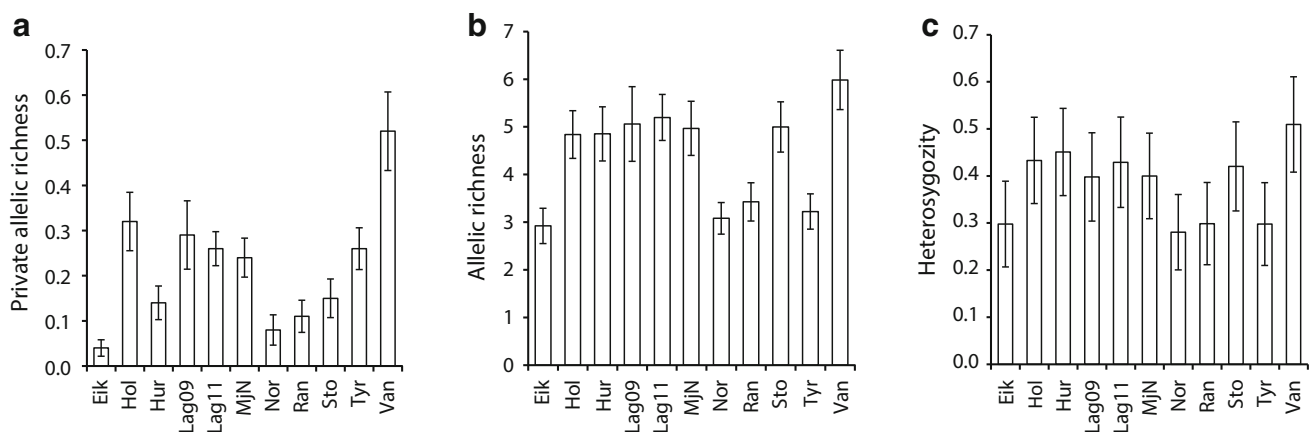


Fig. 2 **a** Private allelic richness (A_p), **b** allelic richness (A_r), and **c** expected heterozygosity (H_{exp}), in the 11 smelt populations; Lake Eikeren (*Eik*), Holingdal (*Hol*), Hurdal (*Hur*), Mjøsa/Lågen-09 (*Lag09*), Mjøsa/Snippsandodden (*MjN*), Mjøsa/Lågen-11 (*Lag11*),

Norsjø (*Nor*), Randsfjorden (*Ran*), Storsjøen (*Sto*), Tyrifjorden (*Tyr*), Vänern (*Van*). Values are given with mean \pm standard error of the mean (SEM)

Table 2 Upper diagonal: Pairwise comparison of expected heterozygosity (H_{exp}) among the 11 smelt populations from Wilcoxon rank test

	Eik	Hol	Hur	Lag09	MjN	Lag11	Nor	Ran	Sto	Tyr	Van
Eik		0.0122*	0.0047*	0.0015*	0.0060*	0.0060*	0.7507	0.1441	0.0052*	0.9579	0.0011*
Hol	0.1862 ^{HS}		0.4412	0.4837	0.9327	0.9579	0.0182	0.1370	0.4532	0.0682	0.0060*
Hur	0.1648 ^{HS}	0.1019 ^{HS}		0.9579	0.5207	0.4532	0.0060*	0.0252*	0.9579	0.0128*	0.0972
Lag09	0.1716 ^{HS}	0.0900 ^{HS}	0.0658 ^{HS}		0.5439	0.4837	0.0060*	0.0107*	0.9579	0.0128*	0.0128*
MjN	0.1856 ^{HS}	0.1027 ^{HS}	0.0647 ^{HS}	0.0045		0.9579	0.0107*	0.0630	0.5439	0.0380*	0.0060*
Lag11	0.1627 ^{HS}	0.0861 ^{HS}	0.0518 ^{HS}	0.0020	0.0058*		0.0111*	0.0780	0.4902	0.0380*	0.0060*
Nor	0.4778 ^{HS}	0.3044 ^{HS}	0.2899 ^{HS}	0.3643 ^{HS}	0.3630 ^{HS}	0.3290 ^{HS}		0.1441	0.0060*	0.6422	0.0012*
Ran	0.0696 ^{HS}	0.1915 ^{HS}	0.1680 ^{HS}	0.1496 ^{HS}	0.1702 ^{HS}	0.1484 ^{HS}	0.4734 ^{HS}		0.0319*	0.3862	0.0012*
Sto	0.1498 ^{HS}	0.0888 ^{HS}	0.0555 ^{HS}	-0.0030	0.0107**	0.0030	0.3372 ^{HS}	0.1316 ^{HS}		0.0153*	0.0613
Tyr	0.0219*	0.1877 ^{HS}	0.1835 ^{HS}	0.1694 ^{HS}	0.1838 ^{HS}	0.1658 ^{HS}	0.4975 ^{HS}	0.0693 ^{HS}	0.1462 ^{HS}		0.0021*
Van	0.1828 ^{HS}	0.0670 ^{HS}	0.0341 ^{HS}	0.0677 ^{HS}	0.0725 ^{HS}	0.0558 ^{HS}	0.2039 ^{HS}	0.1853 ^{HS}	0.0580 ^{HS}	0.1897 ^{HS}	

Lower diagonal: Pairwise comparisons of F_{st} among the 11 smelt populations. The p values for the presented F_{st} and H_{exp} values were adjusted for multiple comparisons following the False Discovery Rate corrections described by Pike (2011). Significant population differentiation after FDR correction is marked with asterisk, non-significant differentiation is marked in bold. Localities: Eikeren (Eik), Holingdal (Hol), Hurdal (Hur), Lake Mjøsa/Lågen-09 (Lag09), Lake Mjøsa/Snippsandodden (MjN), Lake Mjøsa/Lågen-11 (Lag11), Norsjø (Nor), Randsfjorden (Ran), Lake Storsjøen (Sto), Tyrifjorden (Tyr), Lake Vänern (Van)

Significance levels: <0.0000001^{HS} <0.001** <0.05 *

the most distant population from Lake Storsjøen, Lake Norsjø had a significantly lower H_{exp} than Lake Hurdal, the three Lake Mjøsa populations, Lake Storsjøen and Lake Vänern. Lake Eikeren exhibited a significantly lower H_{exp} than all populations except Lake Norsjø, Randsfjorden and Tyrifjorden. Lake Holingdal exhibited a significantly lower H_{exp} than Lake Vänern and Lake Hurdal exhibited a significantly higher H_{exp} than Lake Norsjø, Randsfjorden and Tyrifjorden. The three Lake Mjøsa populations exhibited a significantly higher H_{exp} than Lake Norsjø (as well as between Lake Randsfjorden and Lake Mjøsa/Lågen-09) and a significantly lower H_{exp} than Lake Vänern. Lake Storsjøen had a significantly higher H_{exp} than Lake Eikeren, Lake Norsjø and Lake Randsfjorden, but no difference in H_{exp} to the Lake Mjøsa populations (Lake Mjøsa/Snippsandodden, Lake Mjøsa/Lågen-09, and Lake Mjøsa/Lågen-11 (Fig. 2c; Table 2)).

Pairwise comparisons of population differentiation (F_{st}) showed highly significant differentiation ($p < 0.001$) among most of the population pairs after adjustment of alpha ($\alpha < 0.0005$, Table 2). The only non-significant F_{st} values were between two locations within Lake Mjøsa ($p = 0.67$), and the temporal samples of Lake Mjøsa ($p = 0.36$), and between Lake Storsjøen and the two temporal samples of Lake Mjøsa. Lake Storsjøen was highly genetically divergent from all other populations except Lake Mjøsa. This indicates that Lake Storsjøen was most genetically similar to the two temporal samples from the same locality in Lake Mjøsa, making this location candidate as the likely source of the smelt in Lake Storsjøen (Table 2).

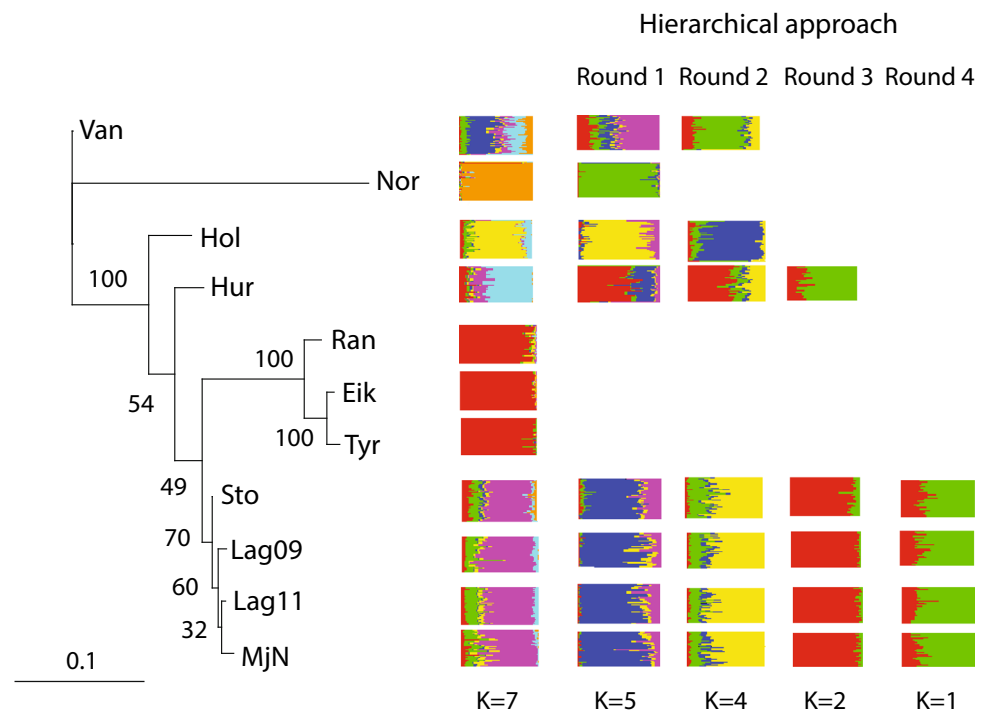
Phylogenetic analysis

The phylogenetic neighbor-joining tree with Lake Vänern as the root, showed a pattern where Lake Storsjøen was only moderately separated from the three Lake Mjøsa samples with a bootstrap support of only 70 %. Even less bootstrap support (60 %) differentiated Lake Mjøsa/Lågen-09 from Lake Mjøsa/Lågen-11 and Lake Mjøsa/Snippsandodden. Finally, only a very low bootstrap support (32 %) differentiated Lake Mjøsa/Lågen-11 from Lake Mjøsa/Snippsandodden (Fig. 3).

Population assignment

The first STRUCTURE analysis resulted in two clusters according to the ΔK value ($\Delta K = 855.687$, mean $\text{LnP}(K) = -11693.26$). However, the $\text{LnP}(K)$ value suggested further structuring into seven different clusters (mean $\text{LnP}(K) = -10911.67$, $\Delta K = 24.71$, Fig. 3, Supp. Fig 6). Clustering all populations into two clusters resulted in one cluster containing Lake Eikeren, Tyrifjorden and Randsfjorden, while the remaining populations were assigned to the other cluster. Round 1 of the hierarchical approach resulted in further sub-structuring into $\Delta K = 2$, and $\text{LnP}(K) = 5$, where ΔK grouped the Lake Norsjø population into a single cluster (Fig. 3, Round 1). Round 2 resulted in a $\Delta K = 2$, and $\text{LnP}(K) = 4$, where ΔK separated Lake Holingdal, Vänern and Hurdal (Fig. 3, Round 2). However, closer inspection of the q-values of the Lake Hurdal population revealed only a 0.036 higher q value to the opposite cluster. Round 3 is thus shown with Lake Mjøsa/

Fig. 3 Plots from hierarchical approach in STRUCTURE (*right side*) with corresponding phylogenetic neighbor-joining tree from Cavalli-Sforza chord measure (*left side*). First plot presents all populations, round 1 without populations *Eik*, *Ran* and *Tyr*, round 2 without population *Nor*, round 3 without populations *Hol* and *Van*, and round 4 without population *Hur*, i.e. only populations *Lag09*, *Lag11*, *MjN* and *Sto*



Lågen-09, Mjøsa/Lågen-11, Mjøsa/Snippsandodden, Storsjøen & Hurdal ($\Delta K = 2$, $\text{LnP}(K) = 1$, Fig. 3, Round 3), and without Lake Hurdal (Fig. 3, Round 4). The three Lake Mjøsa populations also had the highest proportion of membership in the same cluster as Lake Storsjøen (Table 3). The most likely partition was thus a cluster containing all the three Lake Mjøsa populations, together with Lake Storsjøen.

All approaches in the software GENECLASS (Bayesian, frequency based and distance), including the various simulation criterion, assigned Lake Mjøsa as the most likely source of the smelt in Lake Storsjøen. Seven of the eight tests ranked Lake Mjøsa/Lågen-11 as the most likely source while one distance based method (Goldstein et al. 1995) suggested Lake Mjøsa/Snippsandodden as the most likely source (Table 3, Supp. Table 4).

Genetic analyses for demographic events

The simulations done by the program COLONIZE showed that a minimum number of 70 translocated smelt was necessary to have at least a 90 % chance of obtaining as many alleles in the Lake Storsjøen population as were observed. 100 or more translocated smelt were required to have more than a 95 % chance of observing as many alleles (Fig. 4). This was supported by the similar results obtained for the ten replicate runs, thus indicating that the original founding population in Lake Storsjøen likely consisted of at least 70–100 translocated smelt individuals. It was not possible to produce a reliable estimate for maximum number of colonizers to Lake Storsjøen, probably due to low sample size.

The COALIT/NFCONE softwares yielded a maximum likelihood estimate of between 531 and 1053 founders with a minimum support limit between 76 and 149, varying with demographic assumptions (Fig. 5). The maximum support limit reached a peak at approximately 4000 founding individuals, but as the complete limit could not be calculated, only the estimates of maximum likelihood and lower support limits are depicted in Fig. 5.

The Wilcoxon sign-rank test from BOTTLENECK did not detect significant heterozygote excess ($p > 0.05$) under any of the three mutation model scenarios, suggesting no sign of a recent bottleneck event in Lake Storsjøen. The mode-shift indicator from BOTTLENECK suggested a normal L-shaped mode distribution, indicating a demographically stable population.

The intralocus k-test from KG-TEST for detecting population expansions revealed a significant signal for a recent population expansion in Lake Storsjøen, with 12 of 13 loci exhibiting negative k-values ($p = 0.005$). The inter-locus g-test on the other hand did not reveal significant signs of a population expansion in Lake Storsjøen ($p = 1.3$) with a cutoff value of 0.22 from Reich et al. (1999).

Discussion

The results suggested that Lake Mjøsa was the most likely source of the introduced smelt in Lake Storsjøen, supporting the initial hypothesis that the translocation of smelt occurred from a locality in geographic proximity to Lake

Table 3 Assignment of Lake Storsjøen smelt to potential sources using Bayesian clustering in STRUCTURE with prior population information (i.e. trained clustering), and three different approaches (Bayesian, frequency and distance based) in GENECLASS 2 with eight different tests (Bayesian; Rannala & Mountain 1997; Baudouin

& Lebrun 2001, Frequency based; Paetkau et al. 0.1995, Distance based; Nei's standard distance (Nei's $_{SD}$; Nei, 1972), Nei's minimum distance (Nei's $_{MD}$; Nei, 1973), Nei's DA distance (Nei's $_{DA}$; Nei et al.1983), Cavalli-Sforza and Edwards distance (Cavalli-Sforza and Edwards, 1967) and Goldstein's et al. distance, (Goldstein et al.1995)

Locality	STRUCTURE Proportion of membership to the same cluster	GENECLASS							
		Bayesian approach: score (%) Rannala & Mountain	Frequency: score (%)		Distance based: score (%)				
			Baudouin & Lebrun	Paetkau et al.	Nei's $_{SD}$	Nei's $_{MD}$	Nei's $_{DA}$	Cavalli-Sforza	Goldstein
Lag11	0.725	100.0*	100.0*	100.0*	30.334	30.879	23.51	15.849	26.052
Lag09	0.732	0.0	0.0	0.0	27.731	27.148	15.702	12.696	20.148
MjN	0.739	0.0	0.0	0.0	18.223	18.053	15.348	12.033	26.682
Van	0.201	0.0	0.0	0.0	5.228	5.542	10.938	11.298	4.568
Hur	0.215	0.0	0.0	0.0	4.877	5.222	9.545	10.017	10.19
Hol	0.049	0.0	0.0	0.0	3.588	3.815	6.928	9.092	6.322
Ran	0.010	0.0	0.0	0.0	3.320	3.035	5.515	8.170	1.375
Eik	0.002	0.0	0.0	0.0	2.945	2.719	4.774	7.419	1.518
Tyr	0.005	0.0	0.0	0.0	2.824	2.617	4.492	7.329	1.380
Nor	0.023	0.0	0.0	0.0	0.930	0.971	3.247	6.098	1.766

STRUCTURE assignment is presented with the proportion of membership values to the same cluster, GENECLASS 2 results are presented with percent score of most likely source (significantly different scores marked with asterisks, threshold $p < 0.01$) for the Bayesian and frequency based methods, and with percent score for the distance based methods. Highest likelihood is marked in bold. See Table 1 for smelt population name abbreviations

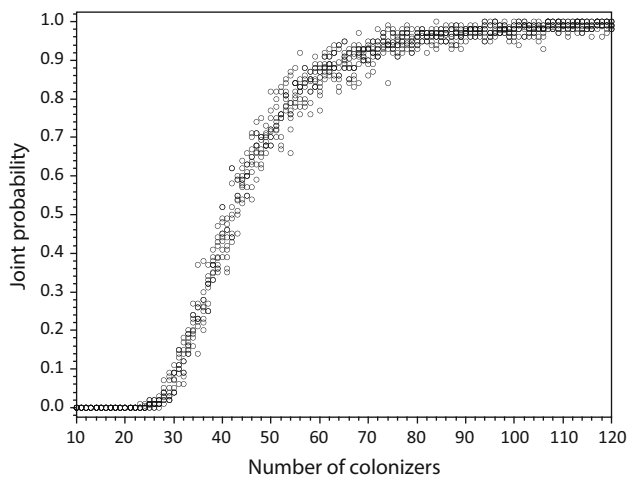


Fig. 4 Result from 10 independent runs (not separated) in COLONIZE estimating joint probability (0–1) for potential number of colonizers from Lake Mjøsa/Lågen-11 into Lake Storsjøen

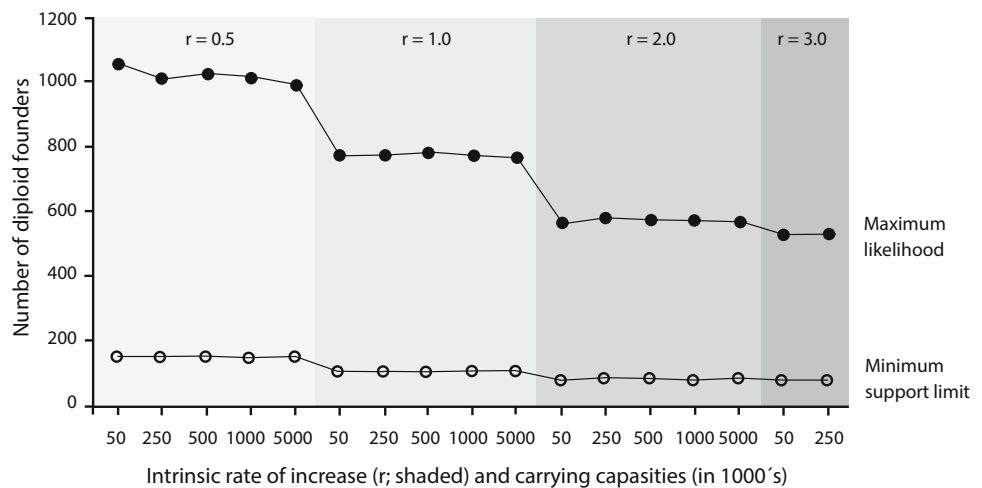
Storsjøen. Thus, based on the findings, the most likely introduction history is that the translocation of smelt to Lake Storsjøen occurred from only one source location. The Lake Storsjøen smelt exhibited no reduction in heterozygosity levels compared to the putative source population, and no difference in private allelic richness, and allelic richness compared to the remainder of the sampled populations.

The smelt introduction from Lake Mjøsa to Lake Storsjøen

Even though the assignment tests indicated that the smelt in Lake Storsjøen most likely originates from only one source location, it was not possible to deduce if the translocation to Lake Storsjøen was a single introduction event, or a results from several introductions from Lake Mjøsa. To address these unknowns, one option is to apply a larger set of higher-resolution genetic markers that can firmly differentiate between founders from the two Lake Mjøsa localities and the two temporal samples. However, resolving the question if the Lake Storsjøen smelt stems from multiple translocations from the very same population within Lake Mjøsa will be very hard, or even impossible, to reveal with any genetic marker, no matter the degree of resolution.

Interestingly, most tests were able to distinguish between the two temporal samples in Lake Mjøsa (Mjøsa/Lågen-09 and Mjøsa/Lågen-11), and the second sampling location in Lake Mjøsa; Mjøsa/Snippsandodden, with the majority of the tests assigning Lake Mjøsa/Lågen-11 as the most likely source. The 2011 sample from Lake Mjøsa/Lågen exhibited a higher similarity to Lake Storsjøen than the sample from 2009. This is possibly an artifact of the limited samples from 2009 ($n = 26$), compared to 2011 ($n = 60$), reflecting only a part of the genetic diversity of

Fig. 5 Maximum likelihood estimates from COALIT/NFCONE runs for the number of founders in Lake Storsjøen under different scenarios of intrinsic rate of increase (r) with corresponding lower support limits. Vertical bars represent carrying capacities ranging for different values of r



the population. This further illustrates that sampling effects may be an important issue in genetic assignment analyses. Thus, all of the performed analyses revealed a high genetic similarity between the Lake Mjøsa/Lågen population and Lake Storsjøen, and most analyses revealed a high differentiation of this assemblage to all of the other populations.

Population assignment programs use genotypes to calculate probabilistic inference of possible source populations (Piry et al. 2004). However, if the applied genetic markers do not have a high enough power to distinguish between putative sources with a similar genetic composition, they may not be able to reveal the real source (Huffman and Wallace 2012). Alternatively, there is a possibility that the Lake Storsjøen smelt may have originated from an un-sampled population that holds a genetic composition similar to that of the Lake Mjøsa/Lågen populations. However, the existence of a second population, identical in genetic composition to Lake Mjøsa/Lågen seems highly unlikely, especially because we included samples from the majority of the neighboring lakes. In addition, the combination of the high resolution of microsatellite markers, in conjunction with the ability of the majority of the analyses performed, to consistently distinguish between populations (even temporal and spatial samples from the same lake) makes this an unlikely scenario.

The origin of smelt in Lake Storsjøen

For all analyses the Lake Storsjøen population consistently had the highest likelihood of origin from Lake Mjøsa, and the majority of the tests assigned the spawning locality Lake Mjøsa/Lågen as the most likely source.

In this study, the Bayesian and frequency based approaches implemented in GENECLASS gave the most detailed interpretation through the ability to significantly

exclude all other populations than Lake Mjøsa/Lågen as potential sources at a significance threshold of $p < 0.01$. This is in correspondence with previous simulation studies indicating a higher assignment success through Bayesian and frequency based methods compared to distance based approaches (Cornuet et al. 1999). Nevertheless, in this study, all analyses reached the same general conclusion making it very likely that Lake Mjøsa is indeed the true source population.

Introduction history of the Lake Storsjøen smelt

Founder populations will often consist of a small proportion of individuals of the original population, comprising only a part of the original genetic diversity (Nei et al. 1975; Dlugosch and Parker 2008). Interestingly, there was no statistical difference in heterozygosity, and no difference in the level of allelic richness between the invaders and putative source population. Similar results were found by (Clegg et al. 2002) who argued that the inability to detect strong founder effects in their study was due to large founder numbers (>100) increasing the likelihood of the founders being genetically representative of the original population. Accordingly, (Nei et al. 1975) stated that the amount of genetic loss is dependent on the number of founding individuals. The lack of any genetic reductions in the Lake Storsjøen smelt may thus have been caused by a substantial number of founders. Indeed, this is supported both by the COLONIZE and COALIT/NFCONE tests that estimated an initial translocation of a substantial number of smelt individuals (70–1000) from Lake Mjøsa/Lågen to Lake Storsjøen. The COLONIZE program provides the probability that a simulated founder event of a certain size would give us at least as many distinct alleles as observed amongst the actual colonized population (Mergeay et al. 2007). As such, it provides a rather ad-hoc method of

estimating the number of founders. By contrast, COALIT/NFCONE uses more information than that available in the number of distinct alleles: since it is based on a sufficient statistic. COALIT/NFCONE uses all the information available in the sample for estimating the number of founders under the model (Anderson and Slatkin 2007). Accordingly we place most weight on the COALIT/NFCONE results which gave a maximum likelihood estimate for number of founders in Lake Storsjøen between 531 and 1053 individuals. Even if one assumes a very high intrinsic rate of increase ($r < 3$), the maximum likelihood estimate remains near 531. Thus, the best estimate for the number of founders is 531 with a lower support limit around 75. In contrast, Kinziger et al. (2011) studying the speckled dace (*Rhinichthys osculus*), an introduced fish species, discovered a reduction in allelic richness relative to the source population. However, the estimated number of founding individuals in that study was much smaller ($n = 7\text{--}17$). In general it seems that a potential explanation for the lack of reduced genetic variation in reported translocated populations as compared with source populations may be that a large number of founder individuals preserve the main composition of the genetic diversity in the original population. Indeed, in our study, only a marginal difference in private allelic richness was discovered between the different lakes, and a post hoc Kruskal–Wallis test revealed no statistical difference between the Lake Storsjøen smelt, and the most likely source population.

Lake Storsjøen is an attractive lake for fishing large-sized brown trout (*Salmo trutta*). There is extensive sport-fishing and annually a competition is held where the winner that catches the largest trout is awarded 150,000 NOK. Last year, 303 fishermen competed in this competition with a total of 170 kg brown trout caught in 48 h (Storsjøen Fiskeforening 2014). These estimates of a substantial number of founders suggest that the translocation to Lake Storsjøen is unlikely to have happened as an accident e.g. by tipping over a bucket of live bait. Smelt is an important forage fish for brown trout (Krause and Palm 2008), and its potential to facilitate a population of the highly desired, large-sized trout (Sandlund and Næsje 2000), may be a possible explanation for the translocation to Lake Storsjøen. Another possibility is that the smelt may have been released repeatedly in small numbers from the same founder population over a long time period through accidental release while being used as bait. This however, seems less likely because smelt was not discovered during an extensive survey fishing in Lake Storsjøen in 2007 (Museth et al. 2008), and was first discovered by local fishermen in 2008 (Strømsmoen 2008), leaving a small time period for numerous releases of a small number of fish. In addition, it seems unlikely that smelt would be systematically collected to be used as bait from Lake

Mjøsa, and not from any other lakes. Regardless, translocation of freshwater fish, both alive and as baitfish, to new localities where they do not already exist, is prohibited by Norwegian law (Innlandsfiskeoven [law relating to salmonids and freshwater fish 2014]; Omsetnings- og sykdomsforskriften for akvatiske dyr (Sale and disease regulation for aquatic animals, 2008).

Population expansion of the smelt in Lake Storsjøen

The smelt in Lake Storsjøen may have had an initial advantage in establishment due to the moderate to large number of translocated individuals and the related high amount of genetic variation. The signal for a recent population expansion after translocation to Lake Storsjøen strongly supports this and indicates that the smelt has had a high reproductive success in its new environment. A similar scenario have been shown for vendace (*Coregonus albula*), that appeared as a highly successful invader associated with limited signatures of founder effects (Amundsen et al. 2012; Præbel et al. 2013). However, although the smelt seems to increase rapidly in population size in Lake Storsjøen, only few generations have passed as the colonization was likely recent in time. No smelt were caught during an extensive survey of the Lake Storsjøen fish community in 2007 (Museth et al. 2008), and the first registered observation was made in 2008 by local fishermen (Strømsmoen 2008). In contrast, during field sampling in 2011 and 2012, smelt were caught at several different localities in the lake, which indicates that the smelt in Lake Storsjøen has undergone a recent population expansion.

Management implications

The results suggest that the smelt in Lake Storsjøen has experienced a rapid population growth following the translocation from Lake Mjøsa. The population is thus likely to expand rapidly and proliferate into available niches in Lake Storsjøen in the future. Studies of introduced rainbow smelt (*Osmerus mordax*), a close relative of the European smelt, has revealed diverse effects on the local community in North American lakes and rivers (Hrabik et al. 1998), through e.g. predation and interspecific resource competition (Evans and Loftus 1987; Mercado-Silva et al. 2007). Similar rapid effects has been observed for other systems and species, such as the intentional introduction of *Coregonus albula* in the Pasvik watercourse (Northern Norway) (Mutenia and Salonen 1992; Bøhn et al. 2008; Præbel et al. 2013). Long-term population genetic and demographic monitoring of the smelt and the ecosystem in Lake Storsjøen is thus crucial since the introduction of smelt is likely to have implications for the food web. Common whitefish (*Coregonus*

lavaretus), the most abundant fish species in Lake Storsjøen (Museth et al. 2008), is an important resource with traditions for domestic use, as well as for commercial- and recreational purposes (H. B. Sundet, advisor for Hedmark County Governor, pers. comm., May, 2013). As whitefish and smelt may have overlapping niches (Sandlund and Næsje 2000; Sandlund et al. 2005), the whitefish population may be affected, subsequently leading to socioeconomic consequences for the local community. On the other hand, the smelt may increase the size of the local trout through provision of a new food source. The question now is, “how, and to what degree, will the introduced smelt affect the ecosystem in Lake Storsjøen, and will these effects impart negative impacts on the fish community, or have positive or negative economic and socioeconomic consequences on the local human population.

In regards to fauna crime, this study has given a unique opportunity to study an introduction event at an early stage, and to monitor the future course in the affected ecosystem, potentially illustrating alternative applications in the framework of invasive species management and fauna crime. It further demonstrates the applicability of multilocus genetic markers as an effective tool for inference of source population and assessment of introduction history of an invasive population. The methods used were effective in assigning Lake Mjøsa as the most likely source of the introduced smelt in Lake Storsjøen, and that the smelt were most likely translocated from the spawning location Lågen. Thus, this study demonstrates an efficient tool to discover and evaluate illegal introductions, which can be used in law enforcement when addressing fauna crime. These methods may be especially useful as a means to stop further introductions in cases when the introduction route is unknown, regardless of the translocation being intentional or unintentional. In Norway, illegal fish translocations seem to be widespread and our genetic methods can be used in such cases addressing various aspects. Thus, the application of these methods can help authorities and law enforcement regulate the spread of the invasive organism through the detected route of transmission. These methods also have the potential to aid in wildlife forensic prosecution such as uncovering illegal poaching, cheating in e.g. fishing competitions by unveiling the true origin of the organism in question, and stop escape of farmed fish by discovering from which fish pen the farmed fish are escaping. The ability to confidently ascertain from where and how an introduction happened, may also illustrate that illegal introductions can theoretically be exposed, thus acting as a cautionary note for the future.

Acknowledgments This project was mainly funded by Hedmark University College and partly by Hedmark county governor to Kjartan Østbye. Thanks to; Lars R. Eriksen, Linn Hagenlund, Kim M. Bærum, Peter Kiffney, Eivind Østbye, Katrine Kongshavn, Trond Andersen, Jan Roar Kristiansen, Martina Blass, Magnus Andersson, Geir

Høitomt, Monica Trondhjem, Aare Werlin, Thomas Wanke, Thomas Mehner, Espen Lydersen, Håkon Hansen, Knut M. Wold, Christian Juel, Max Wahlberg with family, Morten Kristoffersen (*Eikeren Fiskevernforening*), Helene H. Lødemel, Olivier Devineau and Marjorie Couton for help in the field, with analyses, and for useful comments on the study, and to Asbjørn Vøllestad for access to fish lab facilities.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alacs EA, Georges A, FitzSimmons NN, Robertson J (2010) DNA detective: a review of molecular approaches to wildlife forensics. *Forensic Sci Med Pathol* 6(3):180–194. doi:10.1007/s12024-009-9131-7
- Amundsen PA, Salonen E, Niva T, Gjelland KO, Præbel K, Sandlund OT, Knudsen R, Bøhn T (2012) Invader population speeds up life history during colonization. *Biol Invasions* 14:1501–1513. doi:10.1007/s10530-012-0175-3
- Anderson EC, Slatkin M (2007) Estimation of the number of individuals founding colonized populations. *Evolution* 61(4):972–983. doi:10.1111/j.1558-5646.2007.00080.x
- Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9(1):323
- Baudouin L, Lebrun P (2001) An operational Bayesian approach for the identification of sexually reproduced cross-fertilized populations using molecular markers. *Acta Hort* 546:81–94
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B Biol Sci* 263(1377):1619–1626. doi:10.1098/rspb.1996.0237
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nat Rev Genet* 5(4):251–261. doi:10.1038/Nrg1318
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in Population Genetics. *Genetics* 162(4):2025–2035
- Borgstrøm R. (2000). *Fisk i Ferskvann: et samspill mellom bestander, miljø og forvaltning* (2 ed.): landbruksforlaget
- Bøhn T, Amundsen P-A, Sparrow A (2008) Competitive exclusion after invasion? *Biol Invasions* 10(3):359–368
- Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 19:233–257
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24(3):621–631. doi:10.1093/molbev/msl191
- Chapuis MP, Lecoq M, Michalakakis Y, Loiseau A, Sword GA, Piry S, Estoup A (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Mol Ecol* 17(16):3640–3653. doi:10.1111/j.1365-294X.2008.03869.x
- Clegg SM, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proc Nat Acad Sci USA* 99(12):8127–8132. doi:10.1073/pnas.102583399
- Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude

- populations as origins of individuals. *Genetics* 153(4):1989–2000
- County Governor of Hedmark (2011). Forslag til ny fiskeforskrift for Rendalen kommune
- Coulson MW, Paterson IG, Green A, Kepkay R, Bentzen P (2006) Characterization of di- and tetranucleotide microsatellite markers in rainbow smelt (*Osmerus mordax*). *Mol Ecol Notes* 6(3):942–944. doi:[10.1111/j.1471-8286.2006.01409.x](https://doi.org/10.1111/j.1471-8286.2006.01409.x)
- Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol* 17(1):431–449. doi:[10.1111/j.1365-294X.2007.03538.x](https://doi.org/10.1111/j.1365-294X.2007.03538.x)
- Donnelly MJ, Licht MC, Lehmann T (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Mol Biol Evol* 18(7):1353–1364
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185–1192. doi:[10.1093/molbev/msi103](https://doi.org/10.1093/molbev/msi103)
- Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361. doi:[10.1007/s12686-011-9548-7](https://doi.org/10.1007/s12686-011-9548-7)
- ESRI (2012) ArcGIS Desktop: version 10.1 Redlands: Environmental Systems Research Institute
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14(8):2611–2620. doi:[10.1111/j.1365-294X.2005.02553.x](https://doi.org/10.1111/j.1365-294X.2005.02553.x)
- Evans DO, Loftus DH (1987) Colonization of inland lakes in the Great-Lakes region by rainbow smelt, *Osmerus mordax*—their fresh-water niche and effects on indigenous fishes. *Can J Fish Aquat Sci* 44:249–266
- Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J Appl Ecol* 46(2):493–505. doi:[10.1111/j.1365-2664.2008.01606.x](https://doi.org/10.1111/j.1365-2664.2008.01606.x)
- Geller JB, Darling JA, Carlton JT (2010) Genetic perspectives on marine biological invasions. *Ann Rev Mar Sci* 2:367–393. Palo Alto: Annual Reviews
- Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Nat Acad Sci USA* 92(15):6723–6727. doi:[10.1073/pnas.92.15.6723](https://doi.org/10.1073/pnas.92.15.6723)
- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* 86(6):485–486
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2009) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* 104(1):88–99
- Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48(2):361–372. doi:[10.2307/2532296](https://doi.org/10.2307/2532296)
- Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers. *Fish and Fish* 2(2):93–112. doi:[10.1046/j.1467-2960.2001.00043.x](https://doi.org/10.1046/j.1467-2960.2001.00043.x)
- Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evol Biol*. doi:[10.1186/1471-2148-8-289](https://doi.org/10.1186/1471-2148-8-289)
- Hrabik TR, Magnuson JJ, McLain AS (1998) Predicting the effects of rainbow smelt on native fishes in small lakes: evidence from long-term research on two lakes. *Can J Fish Aquat Sci* 55(6):1364–1371. doi:[10.1139/cjfas-55-6-1364](https://doi.org/10.1139/cjfas-55-6-1364)
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9(5):1322–1332. doi:[10.1111/j.1755-0998.2009.02591.x](https://doi.org/10.1111/j.1755-0998.2009.02591.x)
- Huffman JE, Wallace JR (2012) *Wildlife forensics: methods and applications*, vol 6. Wiley-Blackwell, Hoboken
- Johnson RN, Wilson-Wilde L, Linacre A (2014) Current and future directions of DNA in wildlife forensic science. *Forensic Sci Int: Genet* 10:1–11. doi:[10.1016/j.fsigen.2013.12.007](https://doi.org/10.1016/j.fsigen.2013.12.007)
- Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conserv Genet* 5(4):539–543. doi:[10.1023/B:Coge.0000041021.91777.1a](https://doi.org/10.1023/B:Coge.0000041021.91777.1a)
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5(1):187–189. doi:[10.1111/j.1471-8286.2004.00845.x](https://doi.org/10.1111/j.1471-8286.2004.00845.x)
- Kinziger AP, Nakamoto RJ, Anderson EC, Harvey BC (2011) Small founding number and low genetic diversity in an introduced species exhibiting limited invasion success (speckled dace, *Rhinichthys osculus*). *Ecol Evol* 1(1):73–84. doi:[10.1002/Eec3.8](https://doi.org/10.1002/Eec3.8)
- Kottelat M, and Freyhof, J. (2007). *Handbook of European freshwater fishes*. vol 13. Publications Kottelat, Berlin, Germany
- Krause T, Palm A (2008) Dynamics of smelt (*Osmerus eperlanus*) numbers in Lake Peipsi over a decade. *Est J Ecol* 57:111–118
- Langella, O. (1999). *Populations 1.2*. 30. French National Center for Scientific Research (CNRS)
- Law relating to salmonids and freshwater fish, § 9. (2014). Retrieved Dec 20, 2014, from https://lovdata.no/dokument/NL/lov/1992-05-15-47#KAPITTEL_8
- Linacre A (2009) *Forensic science in wildlife investigations*. CRC Press, Boca Raton
- Mercado-Silva N, Sass GG, Roth BM, Gilbert S, Zanden MJV (2007) Impact of rainbow smelt (*Osmerus mordax*) invasion on walleye (*Sander vitreus*) recruitment in Wisconsin lakes. *Can J Fish Aquat Sci* 64(11):1543–1550. doi:[10.1139/f07-112](https://doi.org/10.1139/f07-112)
- Mergeay J, Vanoverbeke J, Verschuren D, De Meester L (2007) Extinction, recolonization, and dispersal through time in a planktonic crustacean. *Ecology* 88(12):3032–3043. doi:[10.1890/06-1538.1](https://doi.org/10.1890/06-1538.1)
- Museth J, Sandlund O T, Johnsen S I, Rognerud S and Saksgård R. (2008). The fish community in Lake Storsjøen. The significance of hydropower development, changed exploitation, and mitigation measures. *NINA Report* (Vol. 388, pp. 61)
- Mutenia A and Salonen E (1992). The vendace [*Coregonus albula* L.], a new species in the fish community and fisheries of Lake Inari. *Polskie Archiwum Hydrobiologii* 39:583–591
- Nei M (1972) Genetic distance between populations. *Am Nat* 106:283–291
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29(1):1–10. doi:[10.2307/2407137](https://doi.org/10.2307/2407137)
- Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular-data. 2. Gene-frequency data. *J Mol Evol* 19(2):153–170. doi:[10.1007/Bf02300753](https://doi.org/10.1007/Bf02300753)
- Ogden R, Gharbi K, Mugue N, Martinsohn J, Senn H, Davey JW, Congiu L (2013) Sturgeon conservation genomics: sNP discovery and validation using RAD sequencing. *Mol Ecol* 22(11):3112–3123. doi:[10.1111/mec.12234](https://doi.org/10.1111/mec.12234)
- Ogden R, Dawnay N, McEwing R (2009) Wildlife DNA forensics; bridging the gap between conservation genetics and law enforcement. *Endanger Species Res* 9(3):179–195. doi:[10.3354/esr00144](https://doi.org/10.3354/esr00144)
- Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population-structure in Canadian polar bears. *Mol Ecol* 4(3):347–354. doi:[10.1111/j.1365-294X.1995.tb00227.x](https://doi.org/10.1111/j.1365-294X.1995.tb00227.x)
- Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12(4):357–358

- Peakall R, Smouse PE (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537–2539. doi:10.1093/bioinformatics/bts460
- Pearse D, Crandall K (2004) Beyond FST: analysis of population genetic data for conservation. *Conserv Genet* 5(5):585–602. doi:10.1007/s10592-003-1863-4
- Pike N (2011) Using false discovery rates for multiple comparisons in ecology and evolution. *Methods Ecol Evol* 2(3):278–282. doi:10.1111/j.2041-210X.2010.00061.x
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Heredity* 95(6):536–539. doi:10.1093/jhered/esh074
- Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Heredity* 90(4):502–503. doi:10.1093/jhered/90.4.502
- Primmer CR, Koskinen MT, Piironen J (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proc R Soc B-Biol Sci* 267(1453):1699–1704
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959
- Præbel K, Gjelland KØ, Salonen E, Amundsen PA (2013) Invasion genetics of vendace (*Coregonus albula* (L.)) in the Inari-Pasvik watercourse: revealing the origin and expansion pattern of a rapid colonization event. *Ecol Evol* 3:1400–1412. doi:10.1002/ece3.552
- Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155(3):1429–1437
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proc Nat Acad Sci USA* 94(17):9197–9201. doi:10.1073/pnas.94.17.9197
- Raymond M, Rousset F (1995) Genepop (Version-1.2)—population-genetics software for exact tests and ecumenicism. *J Heredity* 86(3):248–249
- R Core Team. (2014) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Retrieved Dec 20, 2014, from <http://www.R-project.org/>
- Reich DE, Feldman MW, Goldstein DB (1999) Statistical properties of two tests that use multilocus data sets to detect population expansions. *Mol Biol Evol* 16(4):453–466
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8(1):103–106. doi:10.1111/j.1471-8286.2007.01931.x
- Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchell). *Mol Ecol* 12(2):315–330. doi:10.1046/j.1365-294X.2003.01735.x
- Sale and disease regulation for aquatic animals, FOR-2008-06-17-819. § 44. (2008). <https://lovdata.no/dokument/SF/forskrift/2008-06-17-819>
- Sandlund OT, Næsje TF (2000) Komplekse, laksefiskdominerte fiskesamfunn på Østlandet. In: Borgstrøm R, Hansen P (eds) Fisk i Ferskvann: et samspill mellom bestander, miljø og forvaltning : landbruksforlaget, vol 2., pp 109–129
- Sandlund OT, Stand YG, Kjellberg G, Næsje TF and Hambo MU (2005). European smelt (*Osmerus eperlanus*) eats all; eaten by all: Is it a key species in lakes? *Verhandlungen internationale Vereinigung für theoretische und angewandte Limnologie*, 29, 432–436
- Stauffer HB (2008) Contemporary Bayesian and frequentist statistical research methods for natural resource scientists. Wiley, Hoboken
- Stephens M and Balding DJ (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10(10), 681–690. doi: http://www.nature.com/nrg/journal/v10/n10/suppinfo/nrg2615_S1.html
- Storsjoen Fiskeforening (2014) *Storsjødreggen*. Retrieved from <http://storsjoenfiskeforening.blogspot.no/>
- Strømsoen P I (2008). Krøkle i Storsjøen, *Østlendingen*. Retrieved from <http://www.ostlendingen.no/nyheter/krokle-i-storsjoen-1.4607488>
- Taylor AN, Taylor MI, McCarthy ID, Beaumont AR (2008) Isolation and characterization of 12 dinucleotide microsatellite loci in the anadromous fish *Osmerus eperlanus* (L.). *Mol Ecol Resour* 8(3):581–583. doi:10.1111/j.1471-8286.2007.02001.x
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4(3):535–538. doi:10.1111/j.1471-8286.2004.00684.x
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* 38(6):1358–1370. doi:10.2307/2408641
- Withler R, Candy J, Beacham T, Miller K (2004) Forensic DNA analysis of Pacific salmonid samples for species and stock identification. *Environ Biol Fishes* 69(1–4):275–285. doi:10.1023/B:EBFI.0000022901.26754.0b