Using genetic markers to reveal the source and introduction history of the translocated European smelt (*Osmerus eperlanus* L.) in Lake Storsjøen

Mari Hagenlund

Master in applied ecology

Hedmark University College

Faculty of Applied Ecology and Agricultural Sciences

2013

Applied ecology

Evenstad, May 14, 2013

Mari Hagenlund

Contents

ABSTRACT **4**
SAMMENDRAG **5**
1. INTRODUCTION **6**
2. METHODS **9**
2.1 STUDY SPECIES 9
2.2 STUDY LOCATIONS AND FIELD-WORK 11
2.3 GENETIC ANALYSIS 14
2.4 DATA ANALYSIS 15
3. RESULTS **22**
3.1 GENETIC DIVERSITY 22
3.2 POPULATION DIFFERENTIATION 23
3.3 POPULATION ASSIGNMENT 26
3.4 FOUNDER NUMBERS AND DEMOGRAPHIC HISTORY 30
4. DISCUSSION **33**
4.1 SMELT INTRODUCTION FROM LAKE MJØSA TO LAKE STORSJØEN 33
4.2 COMPARISON OF GENETIC ASSIGNMENT METHODS 34
4.3 GEOGRAPHIC- AND GENETIC DISTANCE OF SMELT POPULATIONS 35
4.4 FOUNDER NUMBERS AND GENETIC DIVERSITY OF THE LAKE STORSJØEN SMELT 36
4.5 BOTTLENECK EVENTS AND POPULATION EXPANSION IN THE LAKE STORSJØEN SMELT 37
4.6 MANAGEMENT IMPLICATIONS 39
4.7 FUTURE DIRECTIONS 40
5. CONCLUSION **41**
ACKNOWLEDGEMENTS **42**
6. REFERENCES **43**
APPENDIX ..... **I**
APPENDIX 1: GENECLASS ASSIGNMENT TABLE I
APPENDIX 2: BAPS ASSIGNMENT TABLE II
APPENDIX 3: LNP(K) AND ΔK-VALUES FROM STRUCTURE III
APPENDIX 4: ASSIGNMENT TABLE WITH REMOVED LOCUS OEP539 IV
APPENDIX 5: PHYLOGENETIC NEIGHBOR JOINING TREES V
APPENDIX 6: MICROSATELLITE PRIMER & MULTIPLEX DETAILS VI
APPENDIX 7: LOCUS DETAILS PER POPULATION (MEAN & SE) VII
APPENDIX 8: LOCUS DETAIL PER POPULATION VIII

Words: 15,030

Abstract

Hagenlund, M. (2013). Using genetic markers to reveal the source and introduction history of the translocated European smelt (*Osmerus eperlanus* L.) in Lake Storsjøen. 57 pages including appendix.

Key words: *Bayesian inference, bottleneck effects, fauna crime, microsatellites, population expansion, source populations, translocation*

1. Introduced species are one of the major threats to freshwater systems worldwide. The ability to accurately determine the original source of invading species offers several powerful applications in invasive species ecology and may enable vital information on the invading species in its native habitat.

2. Lake Storsjøen in Rendalen municipality was recently found to have been subjected to translocation of the European smelt. The smelt is naturally distributed in Southern Norway, but is not native to Lake Storsjøen. The main aim of this study was to infer the most likely source population(s) of the invading smelt in Lake Storsjøen by utilization of neutral microsatellite markers from several potential source populations. Subsequently, I attempted to infer the introduction history of the smelt in Lake Storsjøen.

3. The results indicated that the smelt in Lake Storsjøen is most likely a result of introductions from the spawning locality Lågen, in the closely situated Lake Mjøsa, and that the number of translocated individuals was substantial (>100 individuals). The smelt in Lake Storsjøen showed no significant bottleneck effects, supported by having roughly the same level of genetic diversity as its putative source population. A corresponding significant test for a recent population expansion indicates that the Lake Storsjøen smelt has had a high reproductive success and population growth in their new environment.

4. The results from this study illustrate the usefulness of applying multilocus genetic markers for inferring origin of translocated populations, demographic events and introduction histories. Thus, the methods used comprise an effective tool for assessment of invasive species.

Sammendrag

Hagenlund, M. (2013). Bruken av genetiske markører for å avdekke kildepopulasjonen og introduksjonshistorien til den innførte krøkla (*Osmerus eperlanus* L.) i Storsjøen. 57 sider inkludert vedlegg.

Nøkkelord: *Bayesiansk slutning, flaskehalseffekt, faunakriminalitet, mikrosatellitter, kildepopulasjon, overflytting, populasjons ekspansjon*

1. Introduserte arter er en av de største truslene for ferskvannssystemer på verdensbasis. Avdekking av kildepopulasjonen til innførte arter kan være nyttig innen studier av fremmede arter, og kan bidra med viktig informasjon om den innførte arten i dens opprinnelige habitat.
2. Krøkle, en liten laksefisk ble nylig påvist i Storsjøen i Rendalen kommune. Krøkla er naturlig tilhørende i Sør-Norge, men er ikke stedegen i Storsjøen. Hovedmålet med dette studiet var å identifisere den innførte krøklas kildepopulasjon(er) ved bruk av mikrosatellitter fra flere potensielle kildepopulasjoner. I tillegg forsøkte jeg å kartlegge introduksjonshistorien til krøkla i Storsjøen.
3. Resultatene mine indikerte at krøkla i Storsjøen mest sannsynlig er introdusert fra gytelokaliteten Lågen i den nærliggende innsjøen Mjøsa, og at det ble overført ett betydelig antall individer (>100 individer). Krøkla i Storsjøen viste ingen tegn på å ha vært utsatt for flaskehalseffekter, videre støttet ved at den hadde tilnærmet likt nivå av genetisk diversitet som dens kildepopulasjon. En indikasjon på en nylig populasjonsvekst tyder på at krøkla i Storsjøen har hatt en høy reproduksjonssuksess, og en rask populasjonsvekst i sitt nye miljø.
4. Dette studiet illustrerte nytten av å bruke multilocus genetiske markører for å identifisere innførte arters kildepopulasjoner, demografiske hendelser og introduksjonshistorie. Metoden i dette studiet demonstrerer dermed ett effektivt verktøy for studier av innførte arter.

1. Introduction

Introduction of non-indigenous species is considered one of the major threats to biodiversity loss worldwide (Chapin et al., 2000; Vitousek, D'Antonio, Loope, Rejmanek, & Westbrooks, 1997). Increased connectivity related to human activity has resulted in an increased spread of exotic species beyond their natural borders and an accelerated translocation of native species to previously unoccupied areas (Kahilainen et al., 2011; Mooney & Cleland, 2001). This may lead to unforeseen consequences through ecological interactions (e.g. interspecific competition, predation, and disease transmission), biotic homogenization (reduction of regional and global diversity) and a potential alteration of the local dynamics of species (McKinney & Lockwood, 1999; Rahel, 2002; Seehausen, Takimoto, Roy, & Jokela, 2008). In a worst case scenario, introductions may even result in extinction of native species or populations (see e.g. Kaufman, 1992; Sato et al., 2010).

The effect of introduced species and populations on local communities varies greatly (Gurevitch & Padilla, 2004), thus adding complexity to consequence predictions. The vulnerability of a community to invasions is governed by characteristics of both the invading- and native species, as well as the properties of the community (Lonsdale, 1999; Sakai et al., 2001). In a study of the impacts from introduction of the rainbow smelt (*Osmerus mordax*), Hrabik, Magnuson, and McLain (1998) stated that knowledge of the native fish community along with information on the invaders' spatial distribution are important when predicting future ecological interactions. Insight into the population dynamics, genetic structure, parasite abundance and niche utilization of the source of an introduced species, is vital knowledge for better predicting possible long term consequences and changes in the affected community (Hrabik et al., 1998).

Identifying the specific source of invaders is a valuable tool in environmental/fauna crime and wildlife forensics, e.g. knowledge of possible introduction routes to reduce further introductions (Geller, Darling, & Carlton, 2010). Wildlife DNA forensic methods have primarily been used as a means to identify the species of collected evidence in wildlife crime (e.g. poached animals, or illegally harvested wood; Linacre, 2009). However, the expanding field of genetic methods and useful genetic markers (e.g. single nucleotide polymorphisms (SNP's) and microsatellites) offer a wide array of related applications in fauna crime related questions (Alacs, Georges, FitzSimmons, & Robertson, 2010; Geller et al., 2010; Ogden,

2009). The success of assigning individuals back to their most likely source populations will be dependent on the genetic variation between source and invader populations (Huffman & Wallace, 2012), where the selection of high resolution genetic markers for identification of geographical origin is crucial (Wan, Wu, Fujihara, & Fang, 2004). Microsatellites, (short sequence repeats; SSRs), have a high degree of allelic diversity, high rates of mutation, and are easily amplified in the polymerase chain reaction (PCR; Chistiakov, Hellemans, & Volckaert, 2006; Powell, Machray, & Provan, 1996; Selkoe & Toonen, 2006). These markers often display differences even in recently diverged populations (due to their high mutation rate) which make them very useful in order to assess inference of geneflow and contemporary genetic events (Wan et al., 2004).

Concurrent with the development of genetic methods, there has been an advance in the field of statistical inference with regards to interpreting patterns from genetic markers (Beaumont & Rannala, 2004; Drummond, Rambaut, Shapiro, & Pybus, 2005; Hansen, Kenchington, & Nielsen, 2001). Bayesian inference methods utilizing e.g. microsatellites provide an effective tool for natural scientists (Beaumont, Zhang, & Balding, 2002; Stauffer, 2008; Stephens & Balding, 2009). These approaches allow for statistical genetic assignment and identification of a given individual to putative source populations (Pearse & Crandall, 2004). Such methods are useful in e.g. identifying indigenous and introduced individuals (Primmer, Koskinen, & Piironen, 2000), and has been extensively used in a number of convictions e.g. regarding illegal salmon fishing and trade (Withler, Candy, Beacham, & Miller, 2004). Furthermore, genetic software have been developed where one can infer the past demographic history (Guillemaud, Beaumont, Ciosi, Cornuet, & Estoup, 2009; Heled & Drummond, 2008; Pybus, Rambaut, & Harvey, 2000), making it theoretically possible to infer the most likely number of translocated individuals from one source population to a new locality (Anderson & 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., hereafter smelt) from an unknown source population (Fylkesmannen i Hedmark, 2011). My study is a part of a larger project that intends to evaluate the long term effects of the introduced smelt on the native fish community in Lake Storsjøen. It is thus a vital first step, with 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 populations were compared at 15 microsatellite loci with the

introduced smelt in Lake Storsjøen. Several genetic programs were used, both Bayesian and other analyses, to contrast the collected smelt populations, and to compare the results with regard to the ability of the assignment softwares. Pinpointing the most likely source population enables testing of microsatellites in conjunction with different inference programs intending to illustrate a tool for assessing invasive species in the framework of fauna crime.

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, I test the following hypotheses that;

i. *Translocation of smelt occurred from a locality in geographic proximity to Lake Storsjøen*

   I expect a higher degree of genetic similarity with decreased geographical distances of compared locations to Lake Storsjøen.

ii. *The translocation of smelt to Lake Storsjøen occurred from only one source location*

   I predict that the Lake Storsjøen smelt have the genetic signature of only one of the putative source locations.

iii. *The stocking of smelt was intentional, illustrated through a relatively high number of translocated individuals*

   An estimated number of founders <10-20 may be accounted to an accidental release, while an estimated >100-200 smelt will be interpreted as an intentional stocking event of smelt into Lake Storsjøen.

iv. *The smelt has experienced a demographic bottleneck and a rapid population expansion after translocation*

   I predict that the smelt has adapted quickly to their new environment and will exhibit a signal of a demographic bottleneck and a recent demographic expansion.

Finally, I evaluate the ability and the similarity of the different genetic softwares applied in assigning the Lake Storsjøen smelt individuals back to the most likely source population(s).

2. Methods

2.1 Study species

The European smelt is a small, salmonid species in the family Osmeridae (Kottelat & Freyhof, 2007). It is tolerant to varying salinity levels, existing as estuary populations as well as anadromous and stationary freshwater populations (Doherty & McCarthy, 2004). It is widely distributed in the north east Atlantic coastal waters, from the White- and Barents seas in the north to Garonne estuary in France (Fig. 1; Kottelat and Freyhof, 2007). In Norway the smelt is naturally distributed in the south-eastern part, mainly in large lakes (Sandlund & Næsje, 2000). The smelt exhibits variation in adult-length (Fig. 2) with the normal size ranging from 10 to 20 cm, although body lengths up to 30 cm are reported (Maitland & Lyle, 1996). The rare large-bodied smelt is considered piscivorous predators (Sandlund & Næsje, 2000) often interpreted as being cannibals (Northcote & Hammar, 2006).

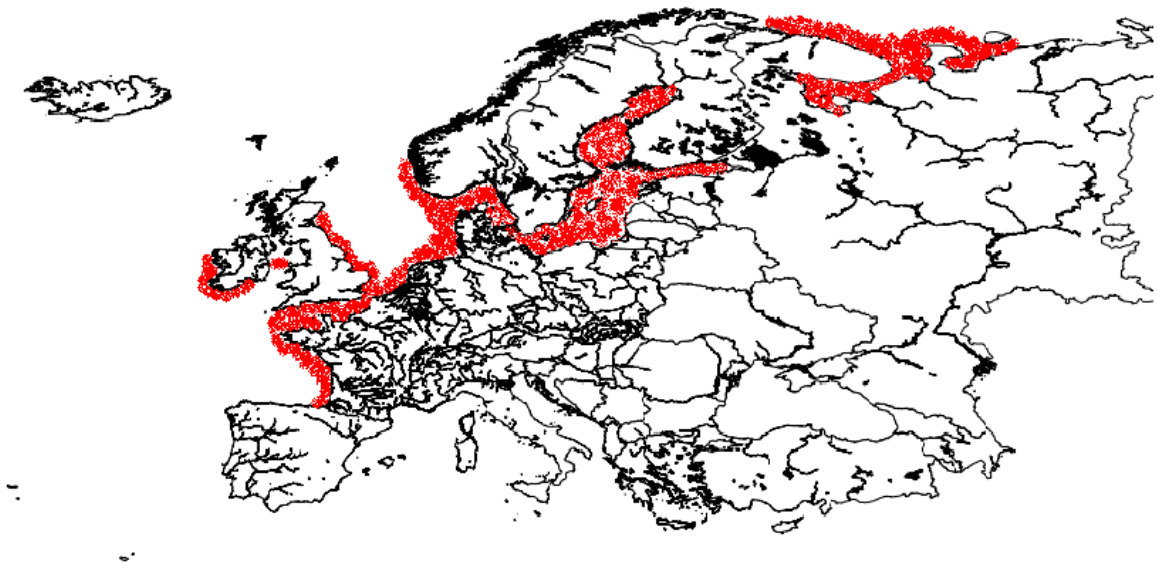

Figure 1. The distribution of smelt in European brackish- marine environments (red coloration).The distributional map has been modified from Kottelat & Freyhof (2007).

The smelt spawning run takes place in spring, the timing of initiation most likely determined by water temperature (Quigley, Igoe, & O`Connor, 2004). Smelt usually return to the same spawning grounds annually at approximately the same date, where their large abundance in shallow water make them highly accessible for both predatory fish and recreational fishing (Krause & Palm, 2008; Sandlund, Stand, Kjellberg, Næsje, & Hambo, 2005). In general, the growth rate of smelt is rapid, and the age at sexual maturity is usually low (between 2-3 years) with landlocked populations often spawning as early as at one years of age (Klyve, 1985; Shpilev, 2005).The demographic size of smelt populations can be very large and is often subject to large, temporal fluctuations in densities (Kottelat & Freyhof, 2007). The smelt is considered a key-species in the pelagic environment in large lakes. It is a motor of nutrient turnover in the ecosystem, as it is a pelagic forager, but feeds on different trophic levels during ontogeny, and is eaten by many different fish-species (Krause & Palm, 2008; Sandlund et al., 2005). Smelt has historically had a high value as food for domesticated animal consumption (and humans in certain parts of Europe) and is considered an ideal baitfish for anglers (Lyle & Maitland, 1997; Quigley et al., 2004).

Figure 2. Gill-net caught smelt from Lake Storsjøen showing size-ranges from small (common) to large sized (rare) individuals, photo: by author, 2012).

2.2 Study locations and field-work

European smelt is a species native to Norway, but has not previously been observed in Lake Storsjøen (Museth, Sandlund, Johnsen, Rognerud, & Saksgård, 2008). It was first discovered by local fishermen in Lake Storsjøen in May 2008 (Strømsmoen, 2008), but the exact time of translocation is unknown.

As the main aim of this thesis was to find the most likely founding source of the translocated smelt population in Lake Storsjøen a set of smelt populations were selected in order to test my specific hypotheses. My *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 the closely situated Lake Mjøsa that has a large population of smelt with several well-known spawning locations. Thus, Lake Mjøsa was considered to be a likely source. Secondly, a set of more southerly distributed smelt populations at an increasing distance from Lake Storsjøen were selected (Fig. 3, Table 1).

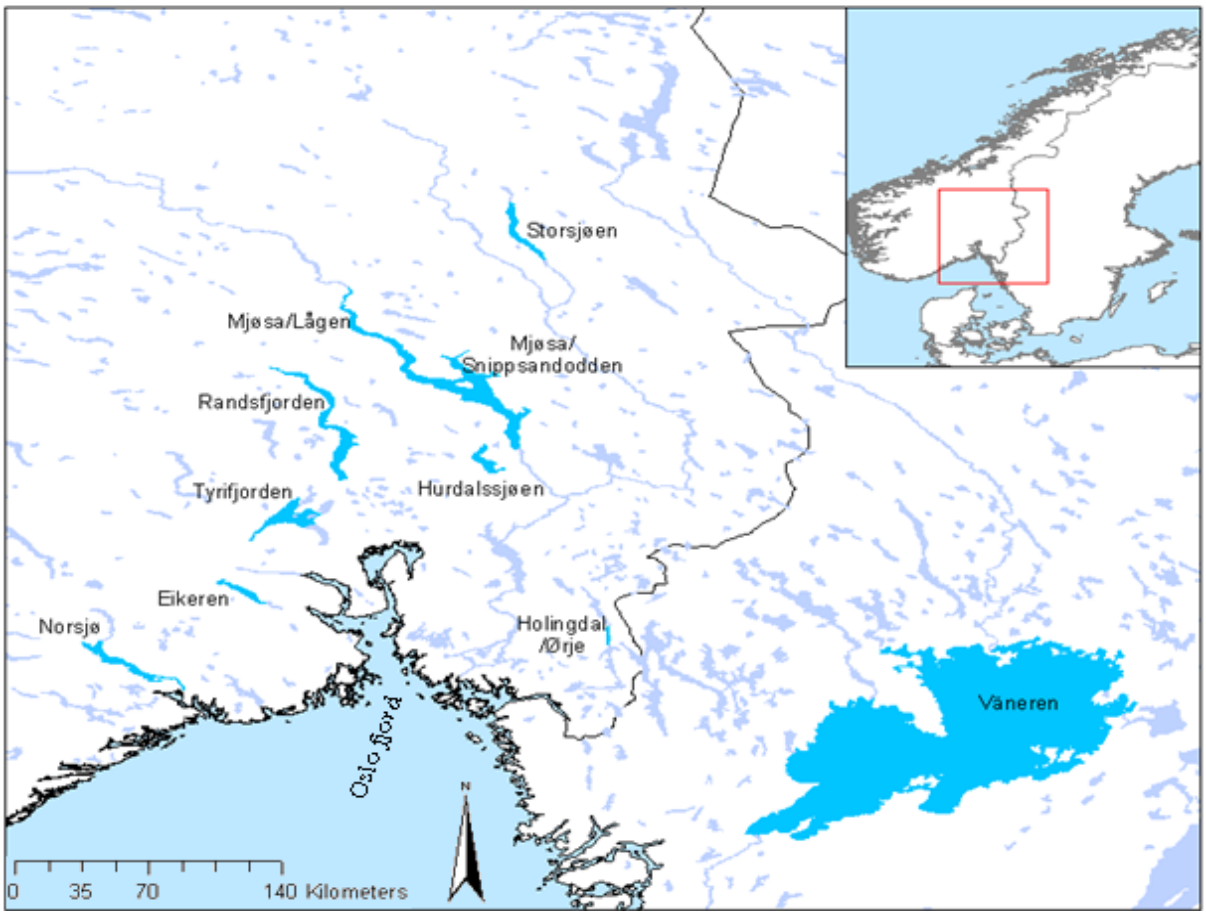

Figure 3. The ten sampling locations of smelt in Norway and Sweden. Lake Storsjøen is the translocated smelt population. The map was created in ArcGIS Version10.1 (ESRI, 2012).

A set of six smelt populations, four distributed on the western side of the Oslo fjord; Lake Randsfjorden, Tyrifjorden, Eikeren and Norsjø, and two populations distributed on the eastern side of the Oslo fjord; Lake Hurdal and Holingdal were sampled (Fig. 3). These populations could also be potential founders of the Lake Storsjøen smelt populations given that oxygen was provided during transport. Finally, a smelt population from Lake Vänern in southeastern Sweden was selected. This population was not expected as the most likely founder populations for the Lake Storsjøen smelt, but was used as an outgroup for polarizing genetic assignments geographically. Moreover, the large Lake Vänern and the vast, ancient, freshwater lake Ancylus have likely been important with regard to colonization of the freshwater fishes in Norway (Borgstrøm, 2000). Since one locality in Lake Mjøsa was sampled twice in different years (Mjøsa Lågen 2009, 2011, Table 1) it offered an opportunity to compare temporal samples from the same locality. This comparison was used to 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 in 8 lakes and 1 river was collected between 2009 and 2012 in Norway and Sweden and used in the analyses (Table 1, Fig. 3).

Table 1. Description of the sampled smelt locations including: locality, population codes, latitude and longitude, year sampled, sample size (N) and collectors of different locations. Localities 2 through 9 are the putative natural founder populations sampled in Norway, while locality Lake Väneren was sampled in Sweden. Locality 1, Lake Storsjøen has the introduced population.

Locality	Code	N 00'00"000'	E 00'00"000'	Year	N	Collector
Introduced population:						
1. Lake Storsjøen	*Sto*	61 67 577	11 19 675	2011/12	47	[1,2]
Natural populations*:*						
2a. Lake Mjøsa, Lågen 2011	*Lag11*	61 06 381	10 26 841	2011	60	[3,4,5]
2b. Lake Mjøsa, Lågen 2009	*Lag09*	61 06 381	10 26 841	2009	26	[3,4,5]
3. Lake Mjøsa, Snippsandodden	*MjN*	60 48 611	10 58 424	2009	40	[5]
4. Lake Eikeren	*Eik*	59 40 950	09 53 265	2012	40	[1]
5. Lake Hurdalssjøen	*Hur*	60 35 117	11 04 647	2012	32	[1,3]
6. Lake Øymarksjøen,Holingdal	*Hol*	59 28 776	11 39 028	2011	40	[4]
7. Lake Norsjø, Gvarvelva	*Nor*	59 37 418	09 20 239	2012	40	[7]
8. Lake Randsfjorden, Odnes	*Ran*	60 79 828	10 18 737	2011	40	[6]
9. Lake Tyrifjorden, Breienlandet	*Tyr*	60 06 883	10 08 932	2010	40	[4,6]
10. Lake Vänern: N. Dalbosjøn	*Van*	59 03 877	12 84 825	2012	40	[8]

[1]Own fieldwork (Mari Hagenlund and Marius Hassve), [2]students from Hedmark University College, [3]Kjartan Østbye, [4]Ruben Pettersen, [5]Finn Gregersen, [6]Finn A. Grøhndal, Geir Høitomt and Monica Trondhjem, [7]Jan R. Kristiansen, [8]Martina Blass and Magnus Andersson.

Smelt was collected either by hand-nets at spawning grounds, or by gill-net fishing. Gill-nets were used with mesh sizes 8, 10, 13.5 and 16 mm (Fig. 4). The gill nets were set at depths ranging from 5 to 20 meters, and left to fish for 12 hours. Lake Storsjøen was sampled during two different time periods (Table 1).

The main aim was to collect a sufficient number of smelt individuals for population genetic assignment, not to obtain a representative material of smelt in different localities/populations with regard to population dynamics. Sampling was thus performed 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, Piry, Luikart, Estoup, & Solignac, 1999; Hansen et al., 2001). The sampled smelt were immediately either frozen, or preserved in plastic containers with 96% ethanol (EtOH) during field collection prior to transportation to the laboratory. In the laboratory the frozen fish were placed in 96 % EtOH immediately following tissue sampling. The ethanol in all containers was replaced the following day after sampling, in order to optimize the preservation of the DNA.

In the laboratory, fin clips from pectoral fins were taken from a random subset of 40-60 sampled smelt when more than 40 smelt were available, whereas all smelt were fin-clipped from populations where only a smaller number than 40 were obtained (Table 1). Fin-clips were preserved in individually marked 2 ml eppendorf tubes in 96% EtOH.

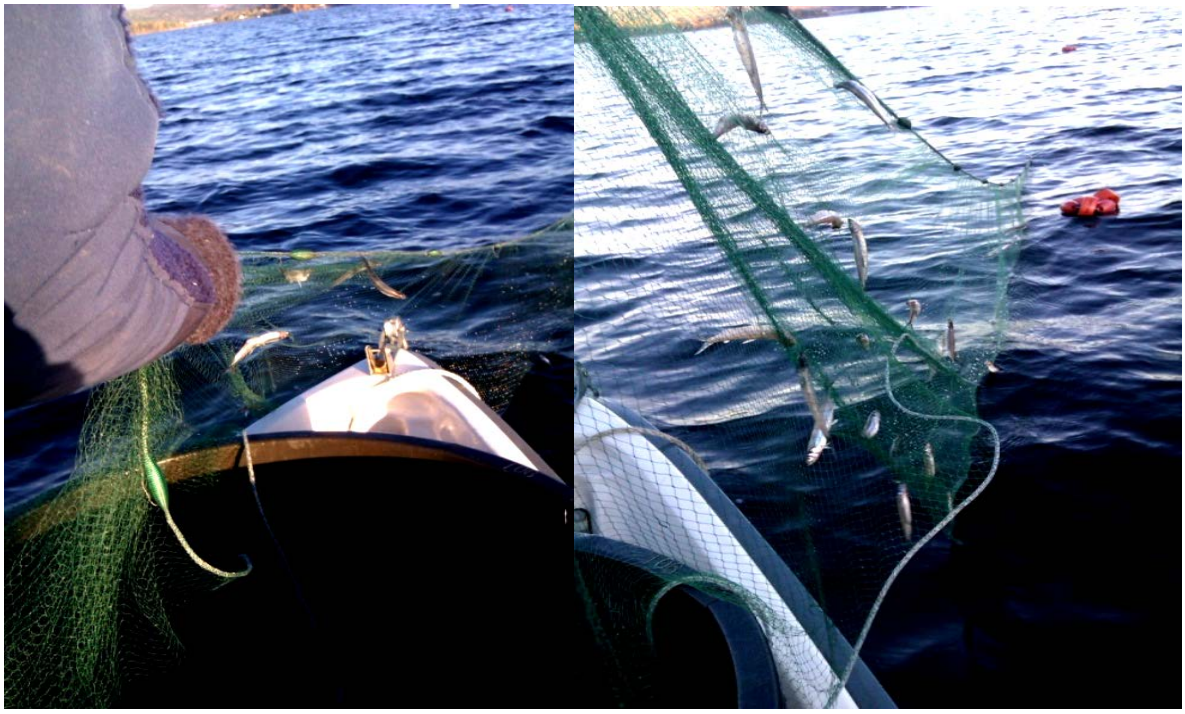

Figure 4. Field sampling with pelagic gill-nets in Lake Eikeren (photo: by author, 2012).

2.3 Genetic analysis

DNA extraction and PCR:

I performed DNA extraction (Fig. 5) at the fish genetics lab at Tromsø University using the E-Z 96 Tissue DNA kit (OMEGA Bio-tek). Briefly, 5 to 10 mg of tissue was placed in round-well plates with 250 µl of OB protease/Buffer TL mix, spun in a centrifuge at 3,000 xg and left to incubate over night at 60 ºC. Plates were mixed with 2x225 µl of Buffer BL/EtOH mix and spun at 3,000 xg. Subsequently E-Z 96 DNA plates were activated by 100 µl equilibrium buffer, incubated for 4 minutes and spun at 4,000 xg for three minutes. As much of the lysate as possible from the round-well plates was then transferred to the E-Z 96 DNA plates, sealed and spun at 6,000 xg for 15 minutes. 2x250 µl of HB buffer was then added and the plates were spun at 5,000 xg for 5 minutes. Following centrifugation, the DNA was cleaned with two rounds of 3x200 µl DNA wash buffer with subsequent centrifugation at 5,000xg for 5 minutes and 6,000 xg for 15 minutes. The plates were then dried at 60 ºC for 20 minutes and subsequently transferred to corresponding Racked Microtubes 1.2 ml plates before addition of 100 µl pre-heated eluation buffer. Plates were then incubated for 2-5 minutes and spun at 6,000 xg for 5 minutes. 30 µl of the extracted DNA was then transferred to ABgene PCR plates, and both Racked Microtubes and aliqouts were sealed and frozen at -20 ºC following Nanodrop quantification of DNA quality of 6 random samples from each plate.

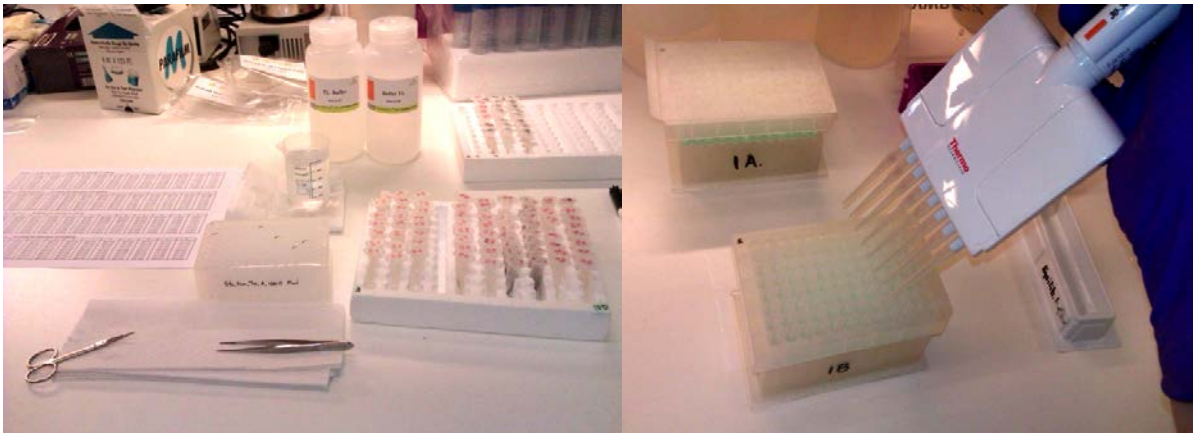

Fig. 5.DNA extraction in the laboratory using E-Z 96 Tissue DNA kit (OMEGA Bio-tek, photo: by author, 2012).

A total of 15 microsatellite loci (see Supplementary Table S4) were optimized and arranged in two multiplex panels with subsequent polymerase chain reaction (PCR).Microsatellite optimization and PCR's were performed by Kim Præbel at the fish genetics lab in Tromsø

University. Amplifications were performed in 2.5 µl PCR reactions containing 5-10 ng template DNA, 1.25 µl Multiplex Master Mix (Qiagen), 0.25 µl of primer mix (see Supp. Table S4), and 0.5 µl ddH_2O. The PCR conditions for the multiplex assays consisted of an initial denaturation step at 95 ºC for 15 min, 25 cycles of 95 ºC for 30 s, 59 ºC for 3 min, and 72 ºC for 1 min, and a final elongation step of 30 min at 60 ºC. The PCR products were separated on an ABI 3130 XL Automated Genetic Analyzer (Applied Biosystems) and alleles scored in the GENEMAPPER 3.7 software (Applied Biosystems). I conducted the genotyping evaluations and binning of alleles following the raw data output from GeneMapper 3.7. together with Kim Præbel. GENEMAPPER uses predefined allelic bins where each genotype is placed. This was later verified by visual inspection. Bins were modified manually if needed, and genotyping scores were checked twice by visual inspection. One locus (*M-Omo4*) was excluded due to poor amplification.

2.4 Data analysis

Microsatellite quality analyses:

A visual inspection of genotypes was performed twice to ensure accurate scoring and full amplification of alleles, as well as a methodological comparison of replicate samples (a minimum of three individuals from each population). The results showed that both scoring inspections were corresponding, and that the replicate samples were identical. Thus, the scored microsatellites and the methods applied are evaluated as being of satisfactory quality.

The software MICRO-CHECKER 2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004) was used to check for: null alleles (false homozygotes, i.e. one of the alleles is not amplified for example due to a mutation in the primer seat), stutter-errors (stutter, i.e. false peaks, creates difficulties in distinguishing between homo- and heterozygotes), large allele dropout (small alleles are amplified more easily than large alleles due to competition in the PCR multiplex) and size-independent allelic dropout (some alleles not amplified due to poor DNA quality or concentration). Null alleles in the dataset may in some cases lead to overestimation of nuclear genetic differentiation (Fst) and genetic distances (Chapuis & Estoup, 2007). Of the fourteen loci, MICRO-CHECKER found five loci to exhibit homozygote excess, potentially due to null alleles, in one or more populations (see Table 1 for population codes), namely; Locus *Oep750 (MjN), Oep538 (Sto), Oep610 (Hol), Oep380 (Hur), Oep384 (Lag11, Nor),* and *Oep135 (Tyr, Van).* Of these, only two loci exhibited homozygote excess in two population; Locus *Oep384*

in populations *Lag11* and *Nor* and Locus *Oep135* in populations *Tyr* and *Van.* Since null alleles were present, the program FREENA (Chapuis & Estoup, 2007; Chapuis et al., 2008) was run to correct for allele-frequency bias with the ENA method (*Excluding Null Alleles*) as described by Chapuis & Estoup (2007). Here, the FREENA software was run with 5,000 replicates, and corrected/uncorrected Fst values were compared using a one-factor ANOVA to test for significant differences, but the corrected values did not exhibit a significant difference from the uncorrected values (ANOVA: $F_{1.108}$=3.93, p=0.90). Thus, given the lack of a systematic occurrence of homozygote excess within loci across populations, the lack of a significant differentiation in Fst when comparing uncorrected and corrected null-allele loci, and that homozygote excess may have other sources than null-alleles (e.g. sampling more than one population in a sample that are assumed to consist only one population), I evaluate that the presence of homozygote excess in some loci will have minor effect on my genetic analyses.

In order to test if loci were candidates for directional or balancing selection, which may affect genetic structure and Fst values (Beaumont & Nichols, 1996), all microsatellites were run in the software LOSITAN (Antao, Lopes, Lopes, Beja-Pereira, and Luikart, 2008; Beaumont and Nichols, 1996) under both the stepwise mutation model (SMM) and the infinite alleles model (IAM). Analyses were run with 100,000 simulations under the "*Force mean Fst*", and "*Neutral mean Fst*" alternatives. Both models found that one locus (*Oep539*) was candidate to directional selection, five loci were candidates for balancing selection under the IAM model (*Oep135, Oep711, M-Omo1* and *M-Omo6*), and four loci (*Oep135, Oep711, Oep380* and *M-*Omo6) were candidates for balancing selection under the SMM model. Putative effects of directional selection at *Oep539* were tested for influence on the genetic assignment tests by both including and excluding the loci in the analyses (details of the analyses are given below). However, no difference in assignment was detected when removing the candidate loci (see Supp. Fig. S3), and the loci *Oep539* was thus included in the subsequent analyses.

GENEPOP 4.0 (Raymond & Rousset, 1995; Rousset, 2008) was used to check for deviations from Hardy Weinberg equilibrium (HWE, may indicate genotyping errors or natural causes of deviation such as sampling more than one location at a given geographical site) and linkage disequilibrium (LD, i.e. locus pairs being transferred on the same chromosome due to geographical proximity on the genome (Guo & Thompson, 1992)) using an exact test. Linkage disequilibrium may lead to false interpretation of the data in some cases of population genetic analyses as loci will not have independent evolutionary histories. Thus, loci exhibiting significant LD should be excluded from the following analyses. A way to test for significance

of HWE deviation and LD is to use the false discovery rate (FDR) corrections (Pike, 2011), which is a less stringent test than the Bonferroni method. Here, FDR was used to adjust p-values for multiple tests. The results showed that out of the 154 tests of departures from HWE, significant deviations were found in only one locus (*Oep384*, p=0.0000) in one population (Lag11) after FDR corrections (threshold, p<0.0005). This was in concordance with the results from MICRO-CHECKER, and the deviation is possibly due to null alleles. The remaining loci exhibiting homozygote excess did not deviate from HWE after FDR correction. Significant LD was not discovered in any of the 178 pairwise tests following FDR correction. Thus, only one locus, *Oep384* was removed, and a total of 14 loci were used in the following genetic analyses.

Genetic diversity estimates:

Number of alleles (N_a), expected (H_{exp}) and observed heterozygosity (H_{obs}) per loci and population, as well as genetic divergence between populations (Fst) was estimated with the software GENEPOP 4.0 (Raymond & Rousset, 1995; Rousset, 2008) using log-likelihood based exact tests (Kalinowski, 2004), as well as the software Fstat 2.9.3.2 (Goudet, 1995) using the Weir and Cockerham estimators (Weir & 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). Mean and standard error of number of alleles, number of effective alleles, observed heterozygosity and expected and unbiased expected heterozygosity over loci and populations, as well as percentage of polymorphic loci, was calculated using GENALEX 6.5 (Peakall & Smouse, 2012). A one-factor ANOVA was used to test for significant differences in standardized private allelic richness ($A_{p)}$, allelic richness A_r), and number of alleles (N_a) among populations. ANOVA was not computable for differences in expected heterozygosity among populations as only mean values were available. A post-hoc ANOVA was performed between Lake Storsjøen and the most likely source population to test for significant difference in A_p, A_r and N_a.

Population assignment methods:

Recent studies have shown that different Bayesian inference programs may give deviating results which may in some circumstances lead to erroneous conclusions (Frantz, Cellina, Krier, Schley, & Burke, 2009). Thus, the use of different softwares that apply alternative methods may lead to a stronger support for assignment of individuals, and minimize the risk

of bias. Thus, a set of three different Bayesian clustering software was applied to assign the Lake Storsjøen smelt; STRUCTURE 2.3.2 (Pritchard, Stephens, & Donnelly, 2000), GENECLASS2 (Piry et al., 2004), and BAPS 5.4 (Corander, Marttinen, Siren, & Tang, 2008).

The software STRUCTURE (Pritchard et al., 2000), determines how many genetic clusters (K) are most likely, and separates individuals based on q-values (i.e. ancestry or membership values) into populations where they are most likely to belong. STRUCTURE was run with an admixture model using 300,000 burn-in steps, and 300,000 Markov Chain Monte Carlo (MCMC) repetitions with 10 iterations. This number was sufficient to reach convergence (results not given). The number of genetic clusters was estimated by calculating the logarithmic probability (LnP(K)) and ΔK which is based on changes in K (Evanno, Regnaut, & Goudet, 2005). The most likely number of clusters (based on LnP(K) and ΔK) was determined using STRUCTURE HARVESTER (Earl & Vonholdt, 2012). STRUCTURE was also run with the LOCPRIOR function wich incorporates geographic sampling information as recommended by Hubisz, Falush, Stephens, and Pritchard (2009). This approach was run assuming default values of migration/translocation two generations past, and a prior migration rate of 0.05 (using 100,000 burn-ins and 100,000 MCMC). Evanno et al. (2005) recommend utilizing the conservative estimate of ΔK to determine number of clusters, and subsequently running STRUCTURE to reanalyze each cluster to find further sub-structuring, i.e. hierarchical approach. Thus, all populations were run a first time where ΔK found 2 clusters, and LnP(K) suggested further structuring into seven clusters. Populations that grouped separately were removed from further analyses, and remaining populations were run a second time. Populations were distinguished from the others by visual inspection of the plots, as well as inspection of highest amount of membership (q) to the different clusters. This was repeated until no further sub-structuring could be observed. The number of iterations were increased as number of populations decreased to keep as high number of iterations as possible (the maximum number of replications possible in the program i.e. 100=10 populations*10 iterations, 5 populations*20 iterations).

The software BAPS 5.4 (Corander et al., 2008) was used with both the method of clustering of groups of individuals without prior information, and the option of trained clustering, where it is possible to “ask” the program to determine from which of the populations with a known origin the population with an unknown origin comes from. The program calculates log (marginal likelihood) values of the most likely clusters, and assigns the introduced population. BAPS produces both a color partition plot of clusters, as well as a model based investigation

of the most likely cluster for a specified group based on changes in log (marginal likelihood) if the group or individuals are moved to another group. Lowest values indicate most likely cluster (Corander, Siren, & Arjas, 2008).

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 which are highly unlikely. This was done by using all the different criterion available for calculation; Bayesian, allele frequency, and distance based. The Bayesian and frequency based approaches in this software has 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 & Crandall, 2004). Individual probability of assignment to the different reference groups was calculated and reference groups were ranked according to their probability as potential sources of smelt individuals in Lake Storsjøen. Both bayesian approaches (Baudouin & Lebrun, 2001; Rannala & Mountain, 1997), and the allele frequency based method (Paetkau, Calvert, Stirling, & Strobeck, 1995) was used to calculate log likelihood (log(l)) values for assignment to the different reference groups. In addition all 5 distance based methods; Nei's Standard (Nei, 1972), Nei's Minimum (Nei, 1973), Nei's DA (Nei, Tajima, & Tateno, 1983), Cavalli-Sforza and Edwards (Cavalli-Sforza & Edwards, 1967) and Goldstein et.al. (Goldstein, Linares, Cavallisforza, & Feldman, 1995) was used to calculate rank with accompanying scores (%) of the different reference groups as potential sources, as well as distance values (Piry et al., 2004). All computations were executed with an assignment threshold of $p<0.01$.

The program POPULATIONS 1.2.30 (Langella, 1999) was used to create phylogenetic rooted neighbour joining trees that group the different populations based on their genetic distances. The trees were created with bootstrap values from 100 repetitions using the Nei's standard distance (Nei, 1972), Nei's DA distance (Nei et al., 1983) Supp. Fig. S2)) and Cavalli-Sforza and Edwards distance method (Fig. 8). Results are shown with the Cavalli-Sforza and Edwards distance as they assume that genetic differentiation occurs due to genetic drift, and do not assume that populations size remains constant (Cavalli-Sforza & Edwards, 1967). As the translocated smelt in Lake Storsjøen most likely consisted of a limited number of individuals (where random genetic drift may be influental), this method seemed most

appropriate. The tree was visualized using TreeView32 (Page, 1996) using the Swedish population Väneren as a geographical outgroup/root.

GENALEX 6.5 (Peakall & Smouse, 2012) was used to create principal coordinate analysis (PCoA) for all populations using the covariance-standardized method. This multivariate technique uses distance estimates (Nei et al., 1983) and Fst to discover patterns of genetic variation in multiple samples across loci, where patterns are proportioned to different axes based on their variation. Groups who share similar genetic patterns will thus group more or less together along the axes. The first axis has the highest explanatory power, with successive axes explaining proportionally less (Peakall & Smouse, 2012).

Genetic analyses for demographic events:

BOTTLENECK (Piry, Luikart, & Cornuet, 1999) was used to evaluate if the translocated smelt individuals in Lake Storsjøen have gone through a bottleneck (i.e. decreased genetic diversity due to a large reduction in effective population size) at the time of release. BOTTLENECK assumes a faster reduction in allelic diversity compared to heterozygosity in bottlenecked populations, thus exhibiting heterozygosity excess. BOTTLENECK calculates the expected and observed heterozygosity relative to the observed allele number and sample size under the assumption of mutation-drift equilibrium using three different mutation models; Stepwise Mutation (SMM), Infinite Alleles (IAM), and Two-Phased Model (TPM, recommended by the authors). It also incorporates three tests; the significance test, a standardized difference test and a Wilcoxon sign-rank test, as well as a "mode-shift" indicator, that discriminates bottlenecked populations from demographically stable populations (Cornuet & Luikart, 1996). BOTTLENECK was run with 1,000 iterations for all the three mutation models, 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 I used two different softwares. First, I used the program COLONIZE (Mergeay, Vanoverbeke, Verschuren, & De Meester, 2007). The program 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. It assumes no linkage disequilibrium (LD) and no substantial genetic drift since the time of release.

The other programs used were COALIT and NFCONE (Anderson & Slatkin, 2007). These programs use a Monte Carlo simulation that allows for estimation of number of founding individuals (or chromosomes) by calculating maximum likelihood estimates, and upper and lower support limits that roughly corresponds to a confidence interval (Anderson & Slatkin, 2007). Here, Eric Anderson (author of the program) kindly ran the softwares using my input of the source and translocated populations, under different scenarios of varying intrinsic growth rate (r) parameters (0.5, 1.0, 2.0, and 3.0), and under varying levels of carrying capacities (50,000, 250,000, 500,000, 1000,000 and 5000,000 diploid individuals). The scenario of intrinsic rate of increase (r) values of 3, was only run under values of 50 000 and 250,000 for carrying capacities due to extensive run-times. Assumed time of translocation was set to four generations ago since this roughly corresponds with suspected time of translocation in 2007/2008.

To test for a population expansion event of the smelt in Lake Storsjøen, the k-, and g-test of Reich, Feldman, and Goldstein (1999) implemented in the KG-TEST was applied. The intralocus k-test explores differences between allele distributions in demographic stable populations compared to a population with a demographic population expansion. The interlocus g-test compares the loci variance in number of repeats to what is expected in a stable population (Reich et al., 1999; Reich & Goldstein, 1998). Both tests assume a stepwise mutation model (SMM). Significance levels for the k-test is based on the number of loci with negative k-values, where a significant number of negative k-values indicate a signature of a demographic population expansion. The g-test significance level was checked according to the recommended cutoff values in table 1 (p.455) reported by Reich et al. (1999).

3. Results

3.1 Genetic diversity

A total of 155 alleles were observed in the 11 populations, across the 14 applied loci. The one-factor ANOVA indicated a tendency, but no significant difference in the number of alleles between populations, (ANOVA: $F_{10.153}$=1.81, p=0.06). The highest number of alleles (N_a=106) were found in the Swedish population, Lake Väneren, and the lowest number of alleles (N_a=46) were observed in Lake Eikeren. Lake Storsjøen exhibited a slightly higher mean number of alleles (6.2) than both Lake Mjøsa/Snippsandodden (6.0), and Lake Mjøsa/Lågen-09 (5.2), but slightly lower than the Lake Mjøsa/Lågen-11 (6.9, Fig. 6b). Percentage of polymorphic loci in the populations ranged from 64.2% (Eikeren) to 92.9% (Holingdal, Mjøsa/Lågen-09, and Mjøsa/Snippsandodden).

The mean expected heterozygosity (H_e) ranged from 0.28 to 0.51, with the most distant population from Lake Storsjøen, Lake Norsjø having the lowest mean, and Lake Väneren exhibiting the highest. Lake Storsjøen had a relatively similar H_e (0.42) to the Lake Mjøsa populations (Lake Mjøsa/Snippsandodden (0.40), Lake Mjøsa/Lågen-09 (3.98), Lake Mjøsa/Lågen-11 (0.43)). The four westerly distributed populations; Lake Norsjø, Lake Eikeren, Lake Tyrifjorden, and Lake Randsfjorden exhibited the overall lowest H_e, with values of 0.30 (Fig. 6a)

There was no significant difference in standardized allelic richness (A_r) between populations (ANOVA: $F_{10.153}$=1.60, p=0.112). Allelic richness varied between 2.92 and 5.98 across populations. 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). 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 Väneren exhibited the highest allelic richness of 5.98 (Fig. 6d). Standardized private allelic richness (A_p) was significantly different among populations (ANOVA: $F_{10.153}$ =2.70, p=0.005), which ranged from 0.04 to 0.92, with Lake Eikeren, Hurdal, Norsjø, Randsfjorden and Storsjøen exhibiting the lowest private allelic richness and Lake Väneren exhibiting the highest. 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), Mjøsa/Lågen-09 (0.29; Fig. 6c).

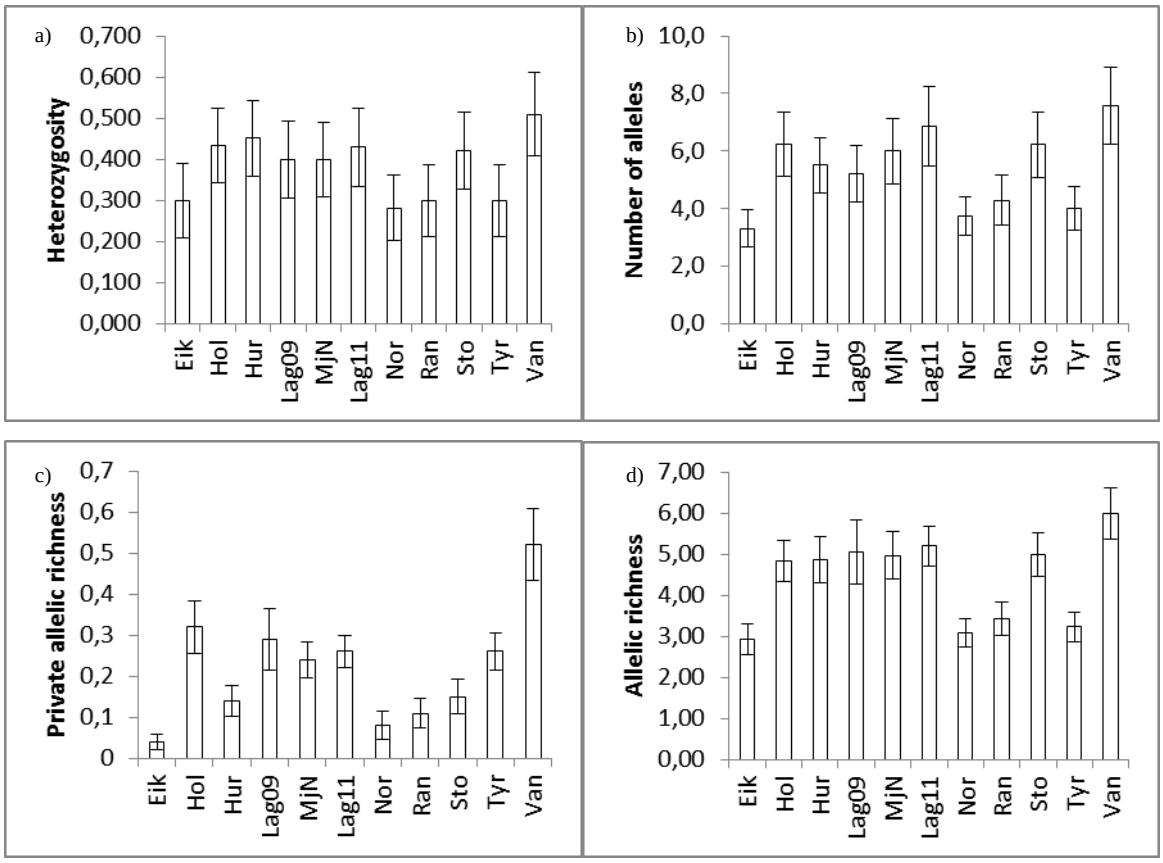

Figure 6. Expected heterozygosity (a), number of alleles (b), private allelic richness (c) and allelic richness (d) in the 11 smelt populations; Eikeren, Holingdal/Ørje, Hurdal, Mjøsa/Lågen-09, Mjøsa/Snippsandodden, Mjøsa/Lågen-11, Norsjø, Randsfjorden, Storsjøen, Tyrifjorden, Vänern. Values are given with mean ± standard error of the mean (SEM).

3.2 Population differentiation

Pairwise comparisons of population differentiation, using genetic differentiation (Fst) showed highly significant differentiation ($p<0.001$) among most of the population pairs after adjustment of alpha (p) following FDR correction. The only non-significant Fst values were between Lake Mjøsa/Lågen-09 and Lake Mjøsa/Snippsandodden ($p=0.67$), Lake Mjøsa/Lågen-09 and Lake Mjøsa/Lågen-11 ($p=0.36$), and between Lake Storsjøen and Lake Mjøsa/Lågen-09 ($p=0.43$) as well as between Lake Storsjøen and Lake Mjøsa/Lågen-11 ($p=0.16$). Lake Storsjøen was highly genetically divergent from all other populations (except Mjøsa). This indicates that Lake Storsjøen is most genetically similar to the two temporal samples from the same locality in Lake Mjøsa (Mjøsa/Lågen-09 and Mjøsa/Lågen-11), making them candidates as the likely source of the smelt in Lake Storsjøen (Table 2). The only other population pairs that did not exhibit highly significant divergence were; Lake Mjøsa/Snippsandodden and Mjøsa/Lågen-11 ($p=0.03$), Mjøsa/Snippsandodden and Storsjøen ($p<0.001$), and Lake Eikeren and Tyrifjorden ($p=0.012$).

Table 2. Pairwise comparisons of Fst among smelt populations. Significant population differentiation after FDR correction is marked with superscript, non-significant differentiation is marked in bold. Populations: Eikeren, Holingdal, Hurdal, Lake Mjøsa/Lågen-09, Lake Mjøsa/Snippsandodden, Lake Mjøsa/Lågen-11, Norsjø, Randsfjorden, Lake Storsjøen, Tyrifjorden, Lake Vänern.

	Eik	*Hol*	*Hur*	*Lag09*	*MjN*	*Lag11*	*Nor*	*Ran*	*Sto*	*Tyr*
Hol	0.1862^{HS}									
Hur	0.1648^{HS}	0.1019^{HS}								
Lag09	0.1716^{HS}	0.0900^{HS}	0.0658^{HS}							
MjN	0.1856^{HS}	0.1027^{HS}	0.0647^{HS}	**0.0045**						
Lag11	0.1627^{HS}	0.0861^{HS}	0.0518^{HS}	**0.0020**	0.0058*					
Nor	0.4778^{HS}	0.3044^{HS}	0.2899^{HS}	0.3643^{HS}	0.3630^{HS}	0.3290^{HS}				
Ran	0.0696^{HS}	0.1915^{HS}	0.1680^{HS}	0.1496^{HS}	0.1702^{HS}	0.1484^{HS}	0.4734^{HS}			
Sto	0.1498^{HS}	0.0888^{HS}	0.0555^{HS}	**-0.0030**	0.0107**	**0.0030**	0.3372^{HS}	0.1316^{HS}		
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}	
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}

<0.0000001^{HS} <0.000001 ***** <0.00001 **** <0.0001*** <0.001 ** <0.05 *

For the specific comparison among the likely source populations (the three Lake Mjøsa populations) and Lake Storsjøen, the post-hoc ANOVA revealed no significant difference in standardized private allelic richness (A_p, ANOVA: $F_{3.55}$=0.603, p=0.616), standardized allelic richness (A_r, ANOVA: $F_{3.55}$=0.012, p=0.998), or number of alleles (N_a, ANOVA: $F_{3.55}$=0.338, p=0.798).

The phylogenetic neighbor joining tree with Lake Väneren as the root, showed a pattern where Lake Norsjø was split into its own branch separated from all other populations (although with no reported bootstrap support, (Fig. 8)). The population of Holingdal was strongly differentiated into a separate branch (100%), being different from other populations, while only low bootstrap support (54%) separated Lake Hurdal from other populations. Low bootstrap support (49%) separated two clusters; one being Lake Randsfjorden, Eikeren and Tyrifjorden, and the other being the three Lake Mjøsa samples (Lågen-09, Lågen-11 and Snippsandodden) and Lake Storsjøen. Within the first cluster, Lake Randsfjorden, Eikeren and Tyrifjorden were highly differentiated from each other (two splits with each of 100%). Within the second cluster, Lake Storsjøen was moderately separated from the three Lake Mjøsa samples with a bootstrap support of only 70%. Even less bootstrap support (60%) differentiated Lågen-09 from Lågen-11 and Snippsandodden. Finally, only a very low bootstrap support (32%) differentiated Lågen-11 from Snippsandodden.

The Principal coordinate analysis (PCoA) revealed four major population groups with regard to allelic differentiation. Here, the first axis (PC1) explained 41.5% of the total variance (eigenvalue 26.6), while the second axis (PC2) explained 19.1% (eigenvalue 12.26), with a total explained variance of 60.6%. Here, evaluating the two axes jointly, the Lake Norsjø population was a single group, while the three populations, Lake Eikeren, Randsfjorden, and Tyrifjorden were pooled into a second group. Lake Väneren, Hurdal and Holingdal comprised a third group, which overlapped somewhat with the fourth group, consisting of the three Lake Mjøsa populations (Mjøsa/Lågen-09, Mjøsa/Lågen-11 and Mjøsa/Snippsandodden) as well as the introduced Lake Storsjøen (Fig. 7).

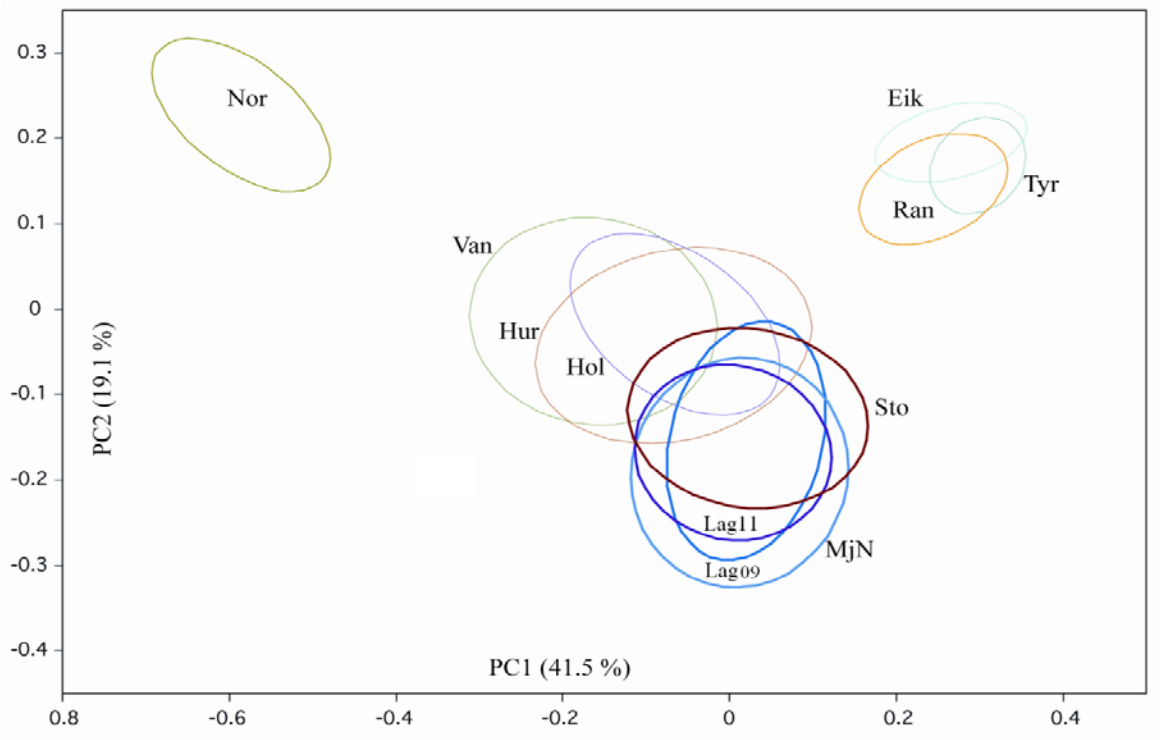

Figure 7. PCoA plot of genetic patterns among smelt populations. Ellipses encompass 50% of the observations for each population. PC1 explains 41.5% of the variation, while PC2 explains 19.1% of the variation. Lake Mjøsa populations and Lake Storsjøen are shown with bold lines.

3.3 Population assignment

The first STRUCTURE analysis resulted in two clusters according to the ΔK value (ΔK=855.687, mean LnP(K) = -11693.26). However, the LnP(K) value suggested further structuring into seven different clusters (mean LnP(K)= -10911.67, ΔK=24.71, Fig. 8, Supp. Fig. S1). Clustering all population into two clusters (2K) 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 (after removing Lake Eikeren, Tyrifjorden, & Randsfjorden) resulted in further sub-structuring into ΔK=2, and LnP(K)=5, where ΔK grouped the Lake Norsjø population into a single cluster (Fig. 8, Round 1). Round 2 resulted in a ΔK=2, and LnP(K)=4, where ΔK separated Lake Holingdal, Väneren and Hurdal (Fig. 8, 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/Lågen-09, Mjøsa/Lågen-11, Mjøsa/Snippsandodden, Storsjøen & Hurdal (ΔK=2, LnP(K)=1, Fig. 8, Round 3), and without Lake Hurdal (Fig. 8, 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.

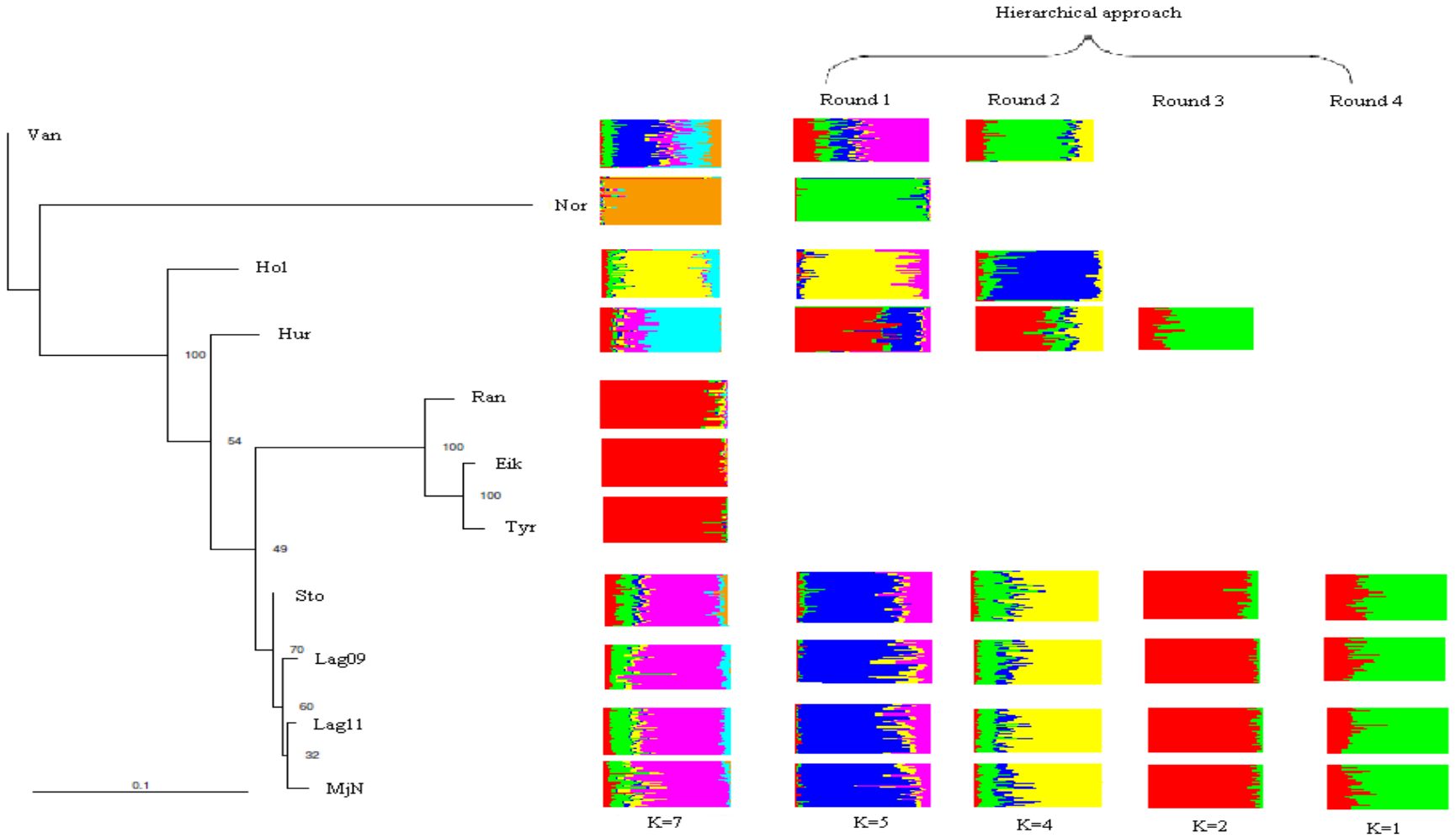

Figure 8. Plots from hierarchical approach in STRUCTURE (right side) with corresponding phylogenetic neighbor-joining tree from Cavalli-Sforza chord measure with bootstrap values obtained from 100 permutations (left side). First plot presents all populations, round 1 without populations Eik, Ran and 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.

Group level mixture analysis in BAPS revealed an optimum partition of the eleven populations into four clusters (Log (marginal likelihood) of optimal partition = 11656.69, Fig. 9, Table 3). Cluster 1 consisted of the Lake Holingdal population, whereas cluster 2 contained Lake Hurdal, Lake Väneren, Lake Mjøsa/Lågen-09, Lake Mjøsa/Lågen-11, Lake Mjøsa/Snippsandodden and the introduced Lake Storsjøen group. Clusters 3 consisted of the Lake Norsjø population, and cluster 4 contained Lake Eikeren, Lake Randsfjorden and Lake Tyrifjorden (Supp. Table S2). A BAPS color plot of partition to the various clusters (similar colors indicate partition to the same cluster) grouped Lake Storsjøen with the three Lake Mjøsa populations (Lake Mjøsa/Lågen-09, Lake Mjøsa/Lågen-11 and Lake Mjøsa/Snippsandodden, Fig. 9a). The option of trained clustering in BAPS incorporating sampling information to cluster the population with "unknown" origin, grouped Lake Storsjøen in the same cluster with Lake Mjøsa/Lågen-11 (Fig. 9b, Table 3).

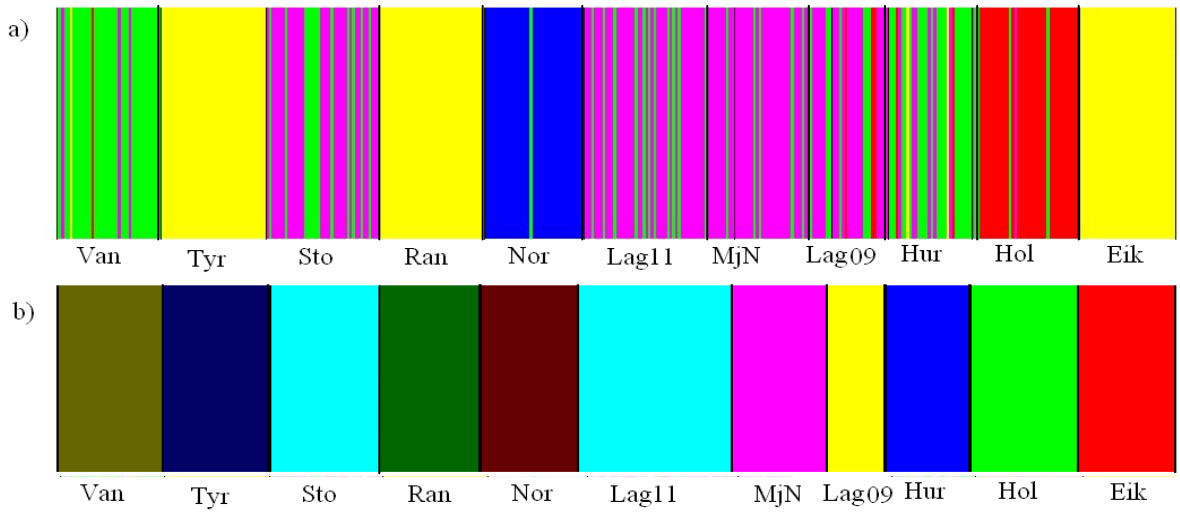

Figure 9. BAPS partition plots of populations to clusters. Clustering of groups (a), and trained clustering (b) Color similarities indicate equal partition to a cluster.

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 S1).

Table 3. Assignment of Lake Storsjøen smelt to potential sources using Bayesian clustering in STRUCTURE with prior population information(i.e. trained clustering), trained clustering in BAPS and three different approaches (Bayesian, frequency and distance based) in GENECLASS 2 with eight different tests (Bayesian, Rannala & Mountain(Rannala & Mountain, 1997), Baudouin & Lebrun(Baudouin & Lebrun, 2001), Frequency based; Paetkau et al (Paetkau et al., 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 & Edwards, 1967) and Goldstein's et al. distance, (Goldstein et al., 1995). STRUCTURE assignment is presented with the proportion of membership values to the same cluster, BAPS assignment presented with changes in Log (marginal likelihood) if Lake Storsjøen is moved to a different cluster (the value zero represents most likely cluster). GENECLASS 2 results are presented with percent score of most likely source (significantly different scores marked with superscript stars, threshold p< 0.01) for the Bayesian and frequency based methods, and with percent score for the distance based methods. Highest likelihood marked in bold.

Locality	STRUCTURE	BAPS	GENECLASS							
			Bayesian approach: Score %		Frequency: Score %	Distance based: Score %				
	Proportion of membership to the same cluster	Changes in Log (marginal likelihood) if sample is moved to cluster :j	*Rannala & Mountain*	*Baudouin & Lebrun*	*Paetkau et al*	*Nei's $_{SD}$*	*Nei's $_{MD}$*	*Nei's $_{DA}$*	*Cavalli-Sforza*	*Goldstein*
Lag11	0.725	**0**	**100.0***	**100.0***	**100.0***	**30.334**	**30.879**	**23.51**	**15.849**	26.052
Lag09	0.732	-31.2	0.0	0.0	0.0	27.731	27.148	15.702	12.696	20.148
MjN	**0.739**	-38.0	0.0	0.0	0.0	18.223	18.053	15.348	12.033	**26.682**
Van	0.201	-70.6	0.0	0.0	0.0	5.228	5.542	10.938	11.298	4.568
Hur	0.215	-97.3	0.0	0.0	0.0	4.877	5.222	9.545	10.017	10.19
Hol	0.049	-171.8	0.0	0.0	0.0	3.588	3.815	6.928	9.092	6.322
Ran	0.010	-235.2	0.0	0.0	0.0	3.320	3.035	5.515	8.170	1.375
Eik	0.002	-280.7	0.0	0.0	0.0	2.945	2.719	4.774	7.419	1.518
Tyr	0.005	-290.7	0.0	0.0	0.0	2.824	2.617	4.492	7.329	1.380
Nor	0.023	-430.3	0.0	0.0	0.0	0.930	0.971	3.247	6.098	1.766

3.4 Founder numbers and demographic history

The program COLONIZE, through joint probabilities for number of colonizers, estimated that approximately 100 individuals was the most likely minimum number of potential colonizers (Fig. 10). Most likely due to low sample size, it was unable to produce a reliable estimate for maximum number of colonizers to Lake Storsjøen. The ten replicate runs produced highly similar results, thus indicating that the original founding population in Lake Storsjøen consisted of a minimum number of 100 translocated smelt individuals.

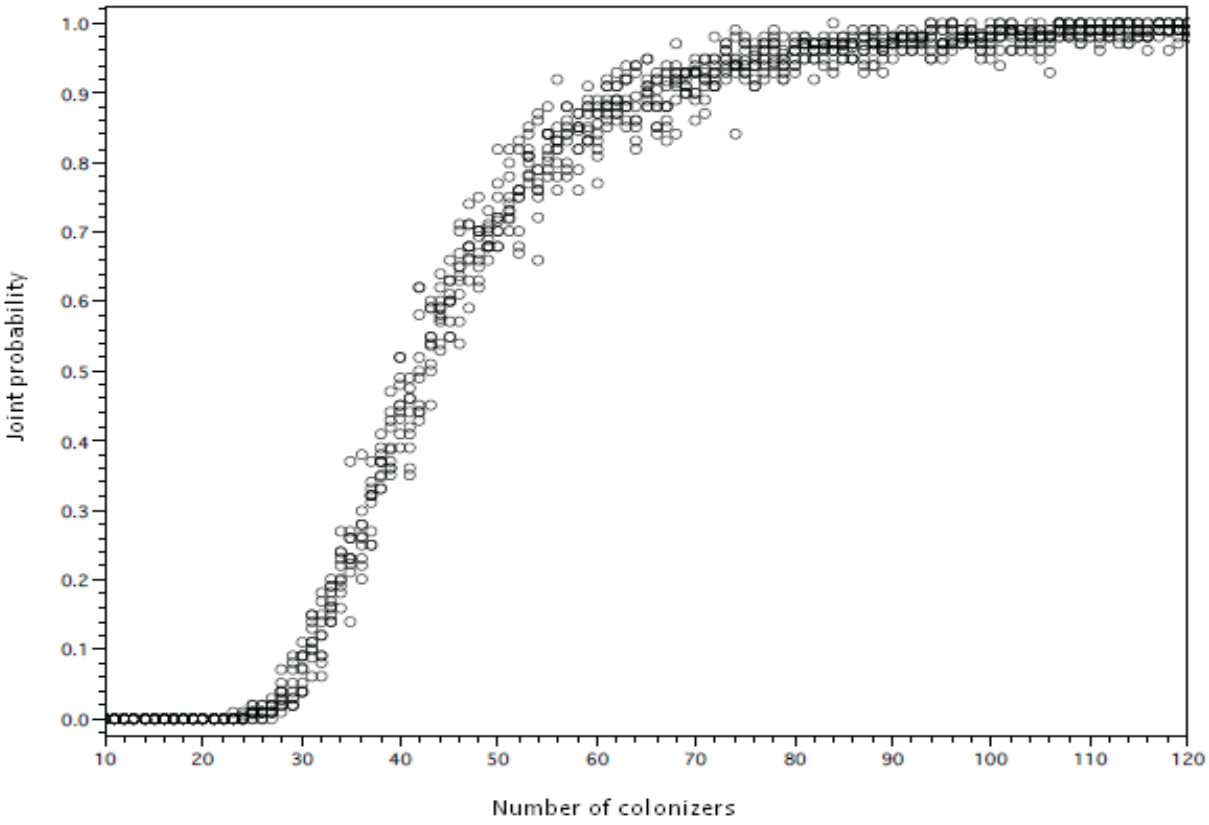

Figure 10. Result from 10 independent runs (not separated) in COLONIZE estimating joint probability (0-1) for potential number of colonizers from Lake Mjøsa/Lågen-11 into Lake Storsjøen.

The COALIT/NFCONE softwares estimated the most likely number of founders in Lake Storsjøen for each of maximum and minimum number of colonizing individuals under variable scenarios of intrinsic rate of increase (r 0.5-3) and carrying capacities (50,000-5000,000). Here, the results suggested a minimum support limit for number of founders to be between 76-149 individuals (r of 3 and 0.5) and the maximum likelihood estimates for

number of founders to be between 531-1053 (r of 3 and 0.5) under the tested scenarios of varying carrying capacities (Fig. 11). The maximum support limit reached a peak at approximately 4000 founding individuals, but as the complete limit could not be calculated, only the conservative estimates of maximum likelihood and lower support limits is depicted in the figure.

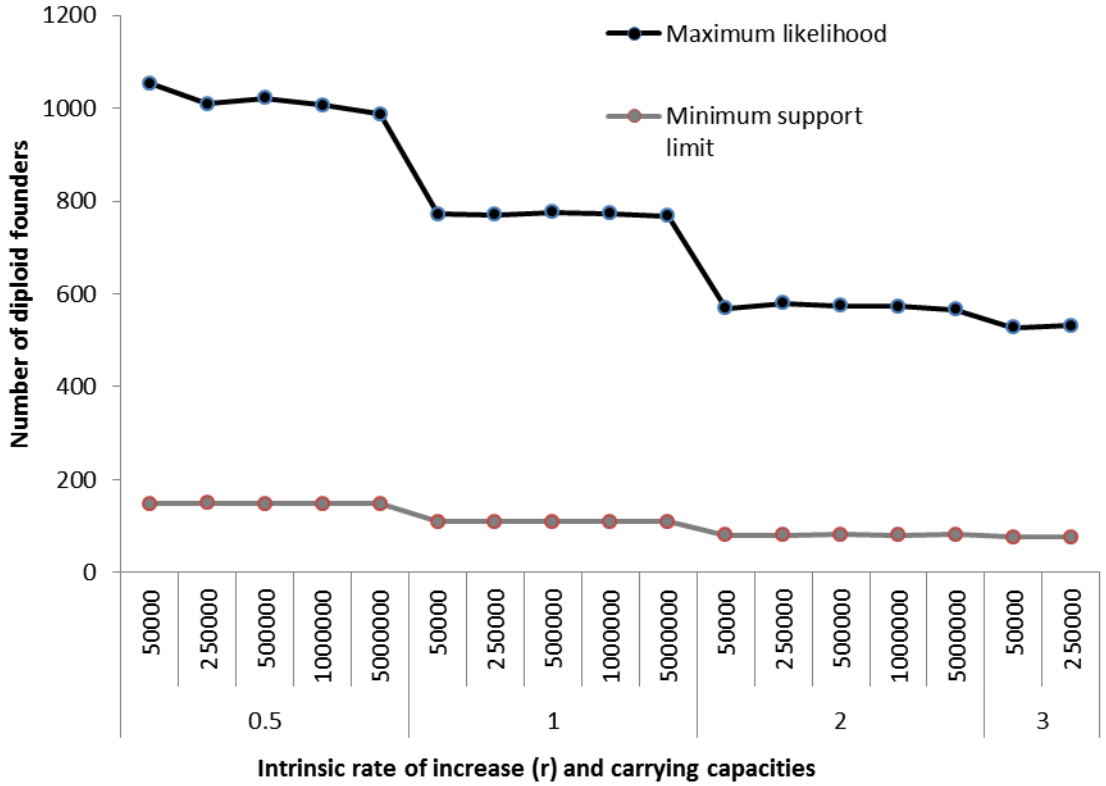

Figure 11. Maximum likelihood estimates from COALIT/NFCONE runs for the number of founders in Lake Storsjøen under different scenarios of intrinsic rate of increase (r) with corresponding lower support limits. Vertical bars represent carrying capacities ranging from 50,000 to 5000,000 for different values of r, r=3 is only presented with carrying capacities 50,000 and 250,000.

The Wilcoxon sign-rank test from BOTTLENECK revealed no sign of a recent bottleneck event in Lake Storsjøen since no significant heterozygote excess was detected under any of the three mutation model scenarios (Table 4). Significant heterozygote deficiency was found under both the SMM and the TPM mutation models. Furthermore, the mode-shift indicator from BOTTLENECK suggested a normal L-shaped mode distribution, indicating a demographic 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 14 loci

exhibiting negative k-values (Table 4). The interlocus g-test on the other hand did not reveal significant signs of a population expansion in Lake Storsjøen (Table 4) with a cutoff value of 0.22 from Reich et al. (1999).

Table 4. Results from Wilcoxon sign-rank test in BOTTLENECK for detecting recent population bottleneck events in the Lake Storsjøen smelt under the IAM, SMM and TPM microsatellite mutation models, and results from k-, and g-test from KG-TEST for detecting population expansion events in the smelt in Lake Storsjøen. Significant results are marked with superscript stars.

Assumptions: all loci fit I.A.M., mutation-drift equilibrium.	
Probability (one tail for H deficiency)	0.133
Probability (one tail for H excess)	0.883
Probability (two tails for H excess and deficiency)	0.266
Assumptions: all loci fit S.M.M., mutation-drift equilibrium.	
Probability (one tail for H deficiency)	0.000^{***}
Probability (one tail for H excess)	1.000
Probability (two tails for H excess or deficiency)	0.000^{***}
Assumptions: all loci fit T.P.M., mutation-drift equilibrium.	
Probability (one tail for H deficiency)	0.002^{**}
Probability (one tail for H excess)	0.998
Probability (two tails for H excess or deficiency)	0.003^{**}
k-test for population expansion	0.005^{**}
g-test for population expansion	1.312

$P<0.05$ * $P<0.01$ ** $P<0.001^{***}$

4. Discussion

My results suggested Lake Mjøsa as the most likely source of the introduced smelt in Lake Storsjøen, supporting my initial hypothesis that the translocation of smelt occurred from a locality in geographic proximity to Lake Storsjøen. Thus, based on my findings, the most likely introduction history is that the translocation of smelt to Lake Storsjøen occurred from only one source location, strengthening my second hypothesis. In support with my third hypothesis, the smelt in Lake Storsjøen is most likely a result of an intentional stocking event as the substantial estimated number of founding individuals (>100) is unlikely to have been translocated accidentally. With regard to demographics, the Lake Storsjøen population did not exhibit a sign of a bottleneck event. However, the Lake Storsjøen population showed a significant sign of demographic expansion. Hence, the results partially support my fourth hypothesis that the smelt experienced a rapid population expansion after translocation to Lake Storsjøen.

4.1 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 illuminate these questions, one option is to apply a larger set of higher-resolution genetic markers that can firmly differentiate between founders from the three Lake Mjøsa populations. 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 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. That the 2011 sample from Lake Mjøsa/Lågen exhibited a higher similarity to Lake Storsjøen than the sample from 2009 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 the population. This further illustrates that sampling effects may be an

important issue in genetic assignment analyses. Thus, all 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 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 & Wallace, 2012). In such, there is a possibility that the Lake Storsjøen smelt may have originated from an unsampled 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. 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.

Due to the limited number of individuals sampled, it is likely that I have not achieved a representative sample of the entire genetic composition in Lake Mjøsa, which may potentially affect the pattern of genetic differentiation. The same applies to some degree among the other sampled populations. However, previous studies have argued that a sample number of 30 individuals and ten loci are sufficient to obtain a high assignment success when genetic differentiation is high (Cornuet et al., 1999; Hansen et al., 2001), which was the case among most of the sampled locations.

4.2 Comparison of genetic assignment methods

For all analyses the Lake Storsjøen group consistently had the highest likelihood of origin from Lake Mjøsa, and the majority of the tests assigned the spawning locality Lågen as the most likely source. Results from STRUCTURE and Goldstein's distance based tests inferred Lake Mjøsa Snippsandodden as the most likely source, a result that deviated from all other tests. The somewhat deviating performance of Goldstein's test has previously been noted by Eldridge, Kinnear, & Onus (2001) in their study of dispersing rock-wallabies (*Petrogale lateralis*). However, considering Goldstein's distance values, the assignment of Lake Storsjøen to Mjøsa-Snippsandodden is only marginally larger (26.682) than to Lågen-11 (26.052), and Lågen-09 (20.148). The next most likely population assignment has the value

of only 10.190 (Lake Hurdal). Similarly, by scrutinizing the proportion of membership values of STRUCTURE, it is evident that the assignment of Lake Storsjøen to Mjøsa-Snippsandodden has only a marginally larger proportion of membership to the same cluster (0.739) than to Lågen-11 (0.725) or Lågen-09 (0.732). The next most likely population assignment has the value of only 0.215 (Lake Hurdal). Thus, even though STRUCTURE and Goldstein's distance test are slightly deviating from the other applied assignment analyses they are in general very similar in their performance as they both suggest that the founders of the Lake Storsjøen smelt indeed originated from within the large Lake Mjøsa populations.

All tests systematically ranked Lake Randsfjorden, Eikeren, Tyrifjorden and Norsjø as the least likely source. Lake Norsjø consistently had the lowest likelihood of being a putative source, except for the STRUCTURE- & Goldstein's distance based test where Lake Eikeren & Randsfjorden were grouped as the least likely source.

In my study, the Bayesian and frequency based approaches implemented in GENECLASS gave the most confident 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.

4.3 Geographic- and genetic distance of smelt populations

In support with my first hypothesis, the pairwise test of genetic differentiation indicated a decreased genetic similarity of the potential source populations to Lake Storsjøen with increasing geographical distance (Table 2, Fig. 3). This was especially apparent when considering the Norsjø population which is located at the greatest distance from Lake Storsjøen. Interestingly, this was not a consistent pattern, as the more proximate, westerly distributed populations (Randsfjorden, Tyrifjorden and Eikeren) were genetically less similar than the more remote populations distributed to the east (Holingdal and Väneren). Väneren, which is the second most distant population from Lake Storsjøen, exhibited approximately the same genetic differentiation to Lake Storsjøen as the more closely situated Hurdal

population. This pattern of increased genetic similarities of Lake Mjøsa and Storsjøen to the easterly distributed populations was also visible on the PCoA, STRUCTURE and BAPS plots, thus indicating that the smelt in these lakes may have a historical origin from Lake Väneren. This seems likely, as smelt is considered to be part of the freshwater fish group commonly believed to have immigrated to Norway from the east through the large glacial-lake Ancylus, which previously connected Norway to Sweden, and Lake Väneren (Borgstrøm, 2000; Hesthagen & Østborg, 2004).

The genetic dissimilarities between the eastern and western populations may thus indicate two phyleogeographic scenarios; [1]that the western populations have been isolated for some time, and genetic drift, bottleneck effects, and/or mutation events may have changed their genetic composition (Guillemaud et al., 2009), or [2]that the western populations have immigrated into Norway through a second route (e.g. through the sea) or during a second time period. The western populations exhibited a relatively low genetic variation (though not significantly different), compared to the eastern populations. This might indicate isolated populations founded by a limited number of individuals that has been influenced by e.g. drift and/or inbreeding (Dlugosch & Parker, 2008), thus supporting the first scenario. Alternatively, the western populations may be the result of a secondary introduction route or introduction period. Disentangling the most likely cause of these geographical patterns is however, outside the scope of this study.

4.4 Founder numbers and genetic diversity of the Lake Storsjøen smelt

Founder populations will often consist of only a small proportion of individuals of the original population, comprising only a part of the original genetic diversity (Dlugosch & Parker, 2008; Nei, Maruyama, & Chakraborty, 1975). Interestingly, the post-hoc ANOVA test of differences in genetic diversity between Lake Storsjøen and Lake Mjøsa revealed no statistical difference in allelic variance estimates, and no visual difference in the level of heterozygosity between the invaders and putative source population. Similar results were found by Clegg et al. (2002) where they 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), states that the amount of genetic loss is dependent on number of founding

individuals, and/or on the severity of the bottleneck. The lack of any 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 (100-1,000) from Lake Mjøsa/Lågen to Lake Storsjøen. In contrast, Kinziger, Nakamoto, Anderson, and Harvey (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=7-17). 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.

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 (*Salmo trutta*; Krause & Palm, 2008), and its potential to facilitate a population of the highly desired, large-sized trout (Sandlund & Næsje, 2000), may be a possible explanation for the translocation to Lake Storsjøen. Recent studies of the trout population in Lake Storsjøen revealed that moderate growth and high mortality, likely due to a lack of appropriate sized prey-fish, results in only a small number of trout able to attain larger sizes (>1.5 kg; Museth et al., 2008). The translocation of smelt may thus have been done in an attempt to provide the trout population with prey-fish to induce a population of larger trout.

4.5 Bottleneck events and population expansion in the Lake Storsjøen smelt

The failure to detect any recent bottleneck event in Lake Storsjøen may be due to a low sample size reducing statistical power as recent reviews shows that bottleneck tests may have limited statistical power and can be heavily affected in their ability to detect recent bottleneck events as a function of low sample size and loci number, as well as time since bottleneck, and severity and duration of the bottleneck (Peery et al., 2012; Spencer, Neigel, & Leberg, 2000). Alternatively, Nei et al. (1975) argued that the effects of bottlenecks on heterozygosity can be masked if founding populations show a rapid population increase

shortly after the bottleneck. This might be in concordance with my findings as the intralocus k-test exhibited signals for a population expansion of the smelt in Lake Storsjøen. The interlocus g-test, on the other hand, did not detect signals of a population expansion, but as discussed by Donnelly, Licht, and Lehmann (2001), this test has a higher power for detecting expansions that happened further in the past, while the k-test has a maximum sensitivity for detecting expansions that happened within a few generations. Thus, the k-test seems more appropriate than the g-test for the introduction of smelt into Lake Storsjøen as the colonization was likely recent in time. Indeed, 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 different localities in the Lake. This indicates that the smelt in Lake Storsjøen has undergone a recent population expansion, and a bottleneck event in Lake Storsjøen may thus have been masked by a short bottleneck duration (Peery et al., 2012) followed by a rapid population expansion.

The smelt in Lake Storsjøen may have had an initial advantage in establishment due to the 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. However, although the smelt seems to increase rapidly in population size in Lake Storsjøen, only few generations have passed. Thus, the putative association between genetic diversity and likelihood of reproductive success and population growth must be followed over several generations.

4.6 Management implications

My 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, 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 & Loftus, 1987; Mercado-Silva, Sass, Roth, Gilbert, & Zanden, 2007). 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 & Næsje, 2000; Sandlund et al., 2005), the whitefish population may be affected, subsequently leading to economic and socioeconomic consequences for the local human population. On the other hand, the smelt may increase the size of the local trout through provision of a new food source. Studies of the trophic interactions of smelt within Lake Mjøsa as well as identifying what mechanism limits or regulates the population (e.g. competition, predation), may help to determine if, and what kind of management actions are needed in Lake Storsjøen. Since the smelt has probably already reached high densities eradication may be impossible. Future studies to monitor what effect the smelt may have on the Lake Storsjøen ecosystem with focus on early damage control and compensatory actions may thus be more effective.

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 fauna crime and invasive species management. In addition, estimation of how many founders are sufficient to attain minimal losses of genetic diversity with regard to establishment of a viable population may be useful for e.g. stocking- or re-introduction programs of endangered populations/species and for fish aimed at human consumption. Finally, the ability to confidently ascertain from where, and how an introduction happened, may illustrate that illegal introductions can theoretically be exposed, thus acting as a cautionary note for the future.

4.7 Future directions

Future studies to monitor the introduced smelt in Lake Storsjøen should ideally include;

- Continued monitoring of the smelt population development with microsatellites to determine if the smelt population continues to grow or if it declines due to e.g. genetic drift, competition and predation.
- Estimating the effective population size (Ne) of the introduced smelt population using microsatellite markers and softwares such as CoNe (Anderson, 2005).
- Inferring time of translocation by using microsatellite markers and softwares such as Bayesian Skyline and Approximate Bayesian Computation (ABC; Athrey, Barr, Lance, and Leberg, 2012; Gignoux, Henn, and Mountain, 2011) for identifying a possible lag phase (phase of low population size after translocation where eradication is most likely to succeed).
- Assessment of genetic differentiation in the smelt population to determine how quickly the smelt population adapts, and diverges e.g. into different spawning localities.
- In addition, inferring time of translocation gives a unique opportunity to compare how, and in what time scale the smelt population in Lake Storsjøen will diverge genetically from Lake Mjøsa through genetic drift or adaptations.
- In addition, ecological studies should be implemented in conjunction with the genetic studies to evaluate ecological impacts of the introduced smelt in Lake Storsjøen such as; stable isotope studies on the smelt to determine niche use and trophic position and isotope studies using pre- and post smelt translocation data to assess the impact of the introduced smelt on the local fish community and food-web dynamics.

5. Conclusion

This study demonstrates the applicability of multilocus genetic markers as an effective tool for inference of source population and assessment of introduction history of an introduced population. The methods used were successful in deducing that the source of the introduced smelt in Lake Storsjøen is most likely Lake Mjøsa, and that the smelt were most likely translocated from the spawning location Lågen. The substantial number of founding individuals suggests that the translocation from Lake Mjøsa/Lågen is unlikely to have happened accidentally. Most likely due to these large founder numbers, the smelt population in Lake Storsjøen has probably had a high reproductive success and a rapid population growth in their new environment. The question now is; how, and to what degree the introduced smelt will affect the ecosystem in Lake Storsjøen, and if these effects will entail positive or negative consequences for the fish-community and economic and socioeconomic consequences for the local human population. Continued genetic and demographic monitoring of the smelt population, as well as ecosystem monitoring via the trophic web is highly recommended. This combination of methods will provide management authorities and local managers with a strong scientific platform to evaluate the most appropriate actions.

Acknowledgements

This project was funded by Hedmark University College and Hedmark county governor.

First of all, I would really like to thank my supervisors Kjartan Østbye, Kjell Langdal and Kim Præbel for invaluable help throughout this entire process. Thank you so much for guidance, advice and help with everything from planning, instruction in genetic analyses and subsequent data analysis.

Second, to Marius Hassve; Thank you for all your invaluable help during the fieldwork and patience through the following process of writing this thesis. It could not have been done without your patience, guidance and tireless pursuit of the ever elusive smelt with me through perilous expeditions in rain, flood and wind.

Thanks to all the other people who has had the endurance to review and help me improve my thesis throughout the process; Lars R. Eriksen, Linn Hagenlund, Kim M. Bærum, Peter Kiffney, Eivind Østbye, Katrine Kongshavn and Trond Andersen.

A big thanks to; Jan Roar Kristiansen, Finn Audun Grøhndal, Ruben Pettersen, Finn Gregersen, Geir Høitomt, Monica Trondhjem, Aare Werlin, Thomas Wanke, Thomas Mehner, Martina Blass, Magnus Andersson, Espen Lydersen, Knut M. Wold, who has participated in obtaining smelt for this thesis.

Thank you so much to all the people who has helped with advice, arrangements, or offered up their boats, time and accommodations to help us through the fieldwork; Christian Juel, Max Wahlberg with family, and Morten Kristoffersen (*Eikeren Fiskevernforening*).

Thanks to Helene H. Lødemel, Marjorie Couton and all the other people who gave me a warm welcome and helped me in the lab in Tromsø, to Ruben Pettersen for teaching me about parasites on smelt, to Asbjørn Vøllestad for allowing me use of the laboratory facilities in Oslo, and finally thanks to Eric Andersson for executing the NFCONE analysis.

6. References

Alacs, E. A., Georges, A., FitzSimmons, N. N., & Robertson, J. (2010). DNA detective: a review of molecular approaches to wildlife forensics. *Forensic Science, Medicine, and Pathology, 6*(3), 180-194. doi: 10.1007/s12024-009-9131-7

Anderson, E. C. (2005). An efficient Monte Carlo method for estimating N-e from temporally spaced samples using a coalescent-based likelihood. *Genetics, 170*(2), 955-967. doi: 10.1534/genetics.104.038349

Anderson, E. C., & 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.

Athrey, G., Barr, K. R., Lance, R. F., & Leberg, P. L. (2012). Birds in space and time: genetic changes accompanying anthropogenic habitat fragmentation in the endangered black-capped vireo (*Vireo atricapilla*). *Evolutionary Applications, 5*(6), 540-552. doi: 10.1111/j.1752-4571.2011.00233.x

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, M. A., & Nichols, R. A. (1996). Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 263*(1377), 1619-1626. doi: 10.1098/rspb.1996.0237

Beaumont, M. A., & Rannala, B. (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics, 5*(4), 251-261. doi: 10.1038/Nrg1318

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Computation in Population Genetics. *Genetics, 162*(4), 2025-2035.

Borgstrøm, R. (2000). Ferskvannsfisk-en viktig ressurs i fortid og nåtid. In R. Borgstrøm & L. P. Hansen (Eds.), *Fisk i Ferskvann: Et samspill mellom bestander, miljø og forvaltning* (2 ed.). Ås: Landbruksforlaget.

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis: Models and estimation procedures. *American Journal of Human Genetics, 19*, 233-257.

Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., . . . Diaz, S. (2000). Consequences of changing biodiversity. *Nature, 405*(6783), 234-242. doi: 10.1038/35012241

Chapuis, M. P., & Estoup, A. (2007). Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution, 24*(3), 621-631. doi: 10.1093/molbev/msl191

Chapuis, M. P., Lecoq, M., Michalakis, Y., Loiseau, A., Sword, G. A., Piry, S., & Estoup, A. (2008). Do outbreaks affect genetic population structure? A worldwide survey in Locusta migratoria, a pest plagued by microsatellite null alleles. *Molecular Ecology, 17*(16), 3640-3653. doi: 10.1111/j.1365-294X.2008.03869.x

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture, 255*(1-4), 1-29. doi: 10.1016/j.aquaculture.2005.11.031

Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 8127-8132. doi: 10.1073/pnas.102583399

Corander, J., Marttinen, P., Siren, J., & Tang, J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics, 9*. doi: 10.1186/1471-2105-9-539

Corander, J., Siren, J., & Arjas, E. (2008). Bayesian spatial modeling of genetic population structure. *Computational Statistics, 23*(1), 111-129. doi: 10.1007/s00180-007-0072-x

Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics, 144*(4), 2001-2014.

Cornuet, J. M., 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.

Coulson, M. W., Paterson, I. G., Green, A., Kepkay, R., & Bentzen, P. (2006). Characterization of di- and tetranucleotide microsatellite markers in rainbow smelt (*Osmerus mordax*). *Molecular Ecology Notes, 6*(3), 942-944. doi: 10.1111/j.1471-8286.2006.01409.x

Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology, 17*(1), 431-449. doi: 10.1111/j.1365-294X.2007.03538.x

Doherty, D., & McCarthy, T. (2004). The Ecology and Conservation of European Smelt (*Osmerus eperlanus* L.) from Waterford Estuary, in Southeastern Ireland. *Biology & Environment: Proceedings of the Royal Irish Academy, 104*(2), 125-130. doi: 10.3318/BIOE.2004.104.2.125

Donnelly, M. J., Licht, M. C., & Lehmann, T. (2001). Evidence for recent population expansion in the evolutionary history of the malaria vectors Anopheles arabiensis and Anopheles gambiae. *Molecular Biology and Evolution, 18*(7), 1353-1364.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution, 22*(5), 1185-1192. doi: 10.1093/molbev/msi103

Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*(2), 359-361. doi: 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. *Molecular Ecology, 14*(8), 2611-2620. doi: 10.1111/j.1365-294X.2005.02553.x

Evans, D. O., & Loftus, D. H. (1987). Colonization of Inland Lakes in the Great-Lakes Region by Rainbow Smelt, *Osmerus-Mordax* - Their Fresh-Water Niche and Effects on Indigenous Fishes. *Canadian Journal of Fisheries and Aquatic Sciences, 44*, 249-266.

Frantz, A. C., 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? *Journal of Applied Ecology, 46*(2), 493-505. doi: 10.1111/j.1365-2664.2008.01606.x

Fylkesmannen I Hedmark. (2011). *Forslag til ny fiskeforskrift for Rendalen kommune*. Fylkesmannen i Hedmark, Miljøverndepartementet.

Geller, J. B., Darling, J. A., & Carlton, J. T. (2010). Genetic Perspectives on Marine Biological Invasions *Annual Review of Marine Science* (Vol. 2, pp. 367-393). Palo Alto: Annual Reviews.

Gignoux, C. R., Henn, B. M., & Mountain, J. L. (2011). Rapid, global demographic expansions after the origins of agriculture. *Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6044-6049. doi: 10.1073/pnas.0914274108

Goldstein, D. B., Linares, A. R., Cavallisforza, L. L., & Feldman, M. W. (1995). Genetic Absolute Dating Based on Microsatellites and the Origin of Modern Humans. *Proceedings of the National Academy of Sciences of the United States of America, 92*(15), 6723-6727. doi: 10.1073/pnas.92.15.6723

Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity, 86*(6), 485-486.

Guillemaud, T., Beaumont, M. A., Ciosi, M., Cornuet, J. M., & Estoup, A. (2009). Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity, 104*(1), 88-99. Retrieved from http://www.nature.com/hdy/journal/v104/n1/suppinfo/hdy200992s1.html

Guo, S. W., & Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics, 48*(2), 361-372. doi: 10.2307/2532296

Gurevitch, J., & Padilla, D. K. (2004). Are invasive species a major cause of extinctions? *Trends in Ecology & Evolution, 19*(9), 470-474. doi: 10.1016/j.tree.2004.07.005

Hansen, M. M., Kenchington, E., & Nielsen, E. E. (2001). Assigning individual fish to populations using microsatellite DNA markers. *Fish and Fisheries, 2*(2), 93-112. doi: 10.1046/j.1467-2960.2001.00043.x

Heled, J., & Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *Bmc Evolutionary Biology, 8*. doi: 10.1186/1471-2148-8-289

Hesthagen, T., & Østborg, G. (2004). Utbredelse av ferskvannsfisk, naturlige fiskesamfunn og fisketomme vatn i Troms og Finnmark. *NINA Oppdragsmelding* (Vol. 805, pp. 1-30). Trondheim: Norsk Institutt For Naturforskning.

Hrabik, T. R., Magnuson, J. J., & McLain, A. S. (1998, February 11). Predicting the effects of rainbow smelt on native fishes in small lakes: evidence from long-term research on two lakes. *Canadian Journal of Fisheries and Aquatic Science, Vol. 55 (6),* 1364-1371.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources, 9*(5), 1322-1332. doi: 10.1111/j.1755-0998.2009.02591.x

Huffman, J. E., & Wallace, J. R. (2012). *Wildlife forensics: methods and applications.*: Wiley-Blacwell.

Kahilainen, K. K., Østbye, K., Harrod, C., Shikano, T., Malinen, T., & Merilä, J. (2011, 15 June). Species introduction promotes hybridization and introgression in Coregonus: is there sign of selection against hybrids? *Molecular Ecology, Vol. 20,* 3838-3855.

Kalinowski, S. T. (2004). Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics, 5*(4), 539-543. doi: 10.1023/B:Coge.0000041021.91777.1a

Kalinowski, S. T. (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

Kaufman, L. (1992). Catastrophic Change in Species-Rich Fresh-Water Ecosystems. *Bioscience, 42*(11), 846-858. doi: 10.2307/1312084

Kinziger, A. P., Nakamoto, R. J., Anderson, E. C., & Harvey, B. C. (2011). Small founding number and low genetic diversity in an introduced species exhibiting limited invasion success (speckled dace, *Rhinichthys osculus*). *Ecology and Evolution, 1*(1), 73-84. doi: 10.1002/Ece3.8

Klyve, L. (1985). *Krøkla (Osmerus eperlanus) i Mjøsa. Alder, vekst og ernæring.* (Master thesis) Universitetet i Oslo, Oslo. (75NS02082)

Kottelat, M., & Freyhof, J. (2007). *Handbook of European freshwater fishes*. Berlin, Germany.

Krause, T., & Palm, A. (2008, January 10). Dynamics of Smelt (*Osmerus eperlanus*) numbers in Lake Peipsi over a decade. *Estonian Journal of Ecology, Vol. 57,* 111-118.

Langella, O. (1999). Populations 1.2. 30. *French National Center for Scientific Research (CNRS).*

Linacre, A. (2009). *Forensic Science in Wildlife Investigations*. Boca Raton: CRC Press.

Lonsdale, W. M. (1999). Global patterns of plant invasions and the concept of invasibility. *Ecology, 80*(5), 1522-1536. doi: 10.1890/0012-9658(1999)080[1522:gpopia]2.0.co;2

Lyle, A. A., & Maitland, P. S. (1997). The spawning migration and conservation of smelt *Osmerus eperlanus* in the River Cree, southwest Scotland. *Biological Conservation, 80*(3), 303-311. doi: http://dx.doi.org/10.1016/S0006-3207(96)00039-0

Maitland, P. S., & Lyle, A. (1996). The smelt *Osmerus eperlanus* in Scotland. *Freshwater Forum, 6*, 57-68.

McKinney, M. L., & Lockwood, J. L. (1999). Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology & Evolution, 14*(11), 450-453. doi: 10.1016/S0169-5347(99)01679-1

Mercado-Silva, N., Sass, G. G., Roth, B. M., Gilbert, S., & Zanden, M. J. V. (2007). Impact of rainbow smelt (*Osmerus mordax*) invasion on walleye (*Sander vitreus*) recruitment in Wisconsin lakes. *Canadian Journal of Fisheries and Aquatic Sciences, 64*(11), 1543-1550. doi: 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

Mooney, H. A., & Cleland, E. E. (2001). The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of the United States of America, 98*(10), 5446-5451. doi: 10.1073/pnas.091093398

Museth, J., Sandlund, O. T., Johnsen, S. I., Rognerud, S., & 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).

Nei, M. (1972). Genetic distance between populations. *Amer. Nat., 106*, 283-291.

Nei, M. (1973). *The theory and estimation of genetic distance*. Honolulu: University press of Hawaii.

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The Bottleneck Effect and Genetic Variability in Populations. *Evolution, 29*(1), 1-10. doi: 10.2307/2407137

Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of Estimated Phylogenetic Trees from Molecular-Data .2. Gene-Frequency Data. *Journal of Molecular Evolution, 19*(2), 153-170. doi: 10.1007/Bf02300753

Northcote, T. G., & Hammar, J. (2006). Feeding ecology of *Coregonus albula* and *Osmerus eperlanus* in the limnetic waters of Lake Malaren, Sweden. *Boreal Environment Research, 11*(3), 229-246.

Ogden, R. D., N.; McEwing, R. (2009). Wildlife DNA forensics; bridging the gap between conservation genetics and law enforcement. *Endangered Species Research, 9*(3), 179-195. doi: 10.3354/esr00144

Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite Analysis of Population-Structure in Canadian Polar Bears. *Molecular Ecology, 4*(3), 347-354. doi: 10.1111/j.1365-294X.1995.tb00227.x

Page, R. D. M. (1996). TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences, 12*(4), 357-358.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 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. *Conservation Genetics, 5*(5), 585-602. doi: 10.1007/s10592-003-1863-4

Peery, M. Z., Kirby, R., Reid, B. N., Stoelting, R., Doucet-Beer, E., Robinson, S., . . . Palsboll, P. J. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology, 21*(14), 3403-3418. doi: 10.1111/j.1365-294X.2012.05635.x

Pike, N. (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution, 2*(3), 278-282. doi: 10.1111/j.2041-210X.2010.00061.x

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*(6), 536-539. doi: 10.1093/jhered/esh074

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity, 90*(4), 502-503. doi: 10.1093/jhered/90.4.502

Powell, W., Machray, G. C., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science, 1*(7), 215-222. doi: http://dx.doi.org/10.1016/1360-1385(96)86898-1

Primmer, C. R., Koskinen, M. T., & Piironen, J. (2000). The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society B-Biological Sciences, 267*(1453), 1699-1704.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics, 155*(2), 945-959.

Pybus, O. G., Rambaut, A., & Harvey, P. H. (2000). An Integrated Framework for the Inference of Viral Population History From Reconstructed Genealogies. *Genetics, 155*(3), 1429-1437.

Quigley, D. T. G., Igoe, F., & O`Connor, W. (2004, December). The European Smelt *Osmerus eperlanus* L. in Ireland: General biology, ecology, distribution and status with conservation recommendations. *Biology and environment: Proceedings of the royal Irish academy, Vol. 104B,* 57-66.

Rahel, F. J. (2002). Homogenization of freshwater faunas. *Annual Review of Ecology and Systematics, 33,* 291-315. doi: 10.1146/annurev.ecolysis.33.010802.150429

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America, 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. *Journal of Heredity, 86*(3), 248-249.

Reich, D. E., Feldman, M. W., & Goldstein, D. B. (1999). Statistical properties of two tests that use multilocus data sets to detect population expansions. *Molecular Biology and Evolution, 16*(4), 453-466.

Reich, D. E., & Goldstein, D. B. (1998). Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences of the United States of America, 95*(14), 8119-8123. doi: 10.1073/pnas.95.14.8119

Rousset, F. (2008). GENEPOP 007: a complete re-implementation of the GENEPOP software for Windows and Linux. [Article]. *Molecular Ecology Resources, 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). *Molecular Ecology, 12*(2), 315-330. doi: 10.1046/j.1365-294X.2003.01735.x

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., . . . Weller, S. G. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics, 32*, 305-332. doi: 10.1146/annurev.ecolsys.32.081501.114037

Sandlund, O. T., & Næsje, T. F. (2000). Komplekse, laksefiskdominerte fiskesamfunn på Østlandet. In R. Borgstrøm & L. P. Hansen (Eds.), *Fisk i Ferskvann: Et samspill mellom bestander, miljø og forvaltning.* (2 ed., pp. 109-129): Landbruksforlaget.

Sandlund, O. T., Stand, Y. G., Kjellberg, G., Næsje, T. F., & Hambo, M. U. (2005). European smelt (*Osmerus eperlanus*) eats all; eaten by all: Is it a key species in lakes? *Verhandlungen internationale Vereinigung für theoretische und angewandte Limnology, 29*, 432–436.

Sato, M., Kawaguchi, Y., Nakajima, J., Mukai, T., Shimatani, Y., & Onikura, N. (2010). A review of the research on introduced freshwater fishes: new perspectives, the need for research, and management implications. *Landscape and Ecological Engineering, 6*(1), 99-108. doi: 10.1007/s11355-009-0086-3

Seehausen, O. L. E., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology, 17*(1), 30-44. doi: 10.1111/j.1365-294X.2007.03529.x

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters, 9*(5), 615-629. doi: 10.1111/j.1461-0248.2006.00889.x

Shpilev, H. O., E.;Lankov, A. (2005). Smelt (*Osmerus eperlanus*) in the Baltic Sea Vol. 54. H. Aben (Ed.) (pp. 230-241). Retrieved from http://www.google.no/books?hl=no&lr=&id=rN3AayP4NmoC&oi=fnd&pg=PA230&dq=Smelt+(Osmerus+eperlanus)+in+the+Baltic+Sea&ots=VbvZkRNnyB&sig=xeN8z3QqI8BeIB7pasuUkN_nSjo&redir_esc=y#v=onepage&q=Smelt%20(Osmerus%20eperlanus)%20in%20the%20Baltic%20Sea&f=false

Spencer, C. C., Neigel, J. E., & Leberg, P. L. (2000). Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology, 9*(10), 1517-1528. doi: 10.1046/j.1365-294x.2000.01031.x

Stauffer, H. B. (2008). *Contemporary Bayesian and Frequentist Statistical Research Methods for Natural Resource Scientists*. Hoboken,New Jersey: John Wiley and Sons, Inc.

Stephens, M., & Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. [10.1038/nrg2615]. *Nature Reviews Genetics, 10*(10), 681-690. Retrieved from http://www.nature.com/nrg/journal/v10/n10/suppinfo/nrg2615_S1.html

Strømsmoen, P. I. (2008). Krøkle i Storsjøen, *Østlendingen*. Retrieved from http://www.ostlendingen.no/nyheter/krokle-i-storsjoen-1.4607488

Taylor, A. N., Taylor, M. I., McCarthy, I. D., & Beaumont, A. R. (2008). Isolation and characterization of 12 dinucleotide microsatellite loci in the anadromous fish *Osmerus eperlanus* (L.). *Molecular Ecology Resources, 8*(3), 581-583. doi: 10.1111/j.1471-8286.2007.02001.x

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes, 4*(3), 535-538. doi: 10.1111/j.1471-8286.2004.00684.x

Vitousek, P. M., D'Antonio, C. M., Loope, L. L., Rejmanek, M., & Westbrooks, R. (1997). Introduced species: A significant component of human-caused global change. *New Zealand Journal of Ecology, 21*(1), 1-16.

Wan, Q. H., Wu, H., Fujihara, T., & Fang, S. G. (2004). Which genetic marker for which conservation genetics issue? *Electrophoresis, 25*(14), 2165-2176. doi: 10.1002/elps.200305922

Weir, B. S., & Cockerham, C. C. (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. *Environmental Biology of Fishes, 69*(1-4), 275-285. doi: 10.1023/B:EBFI.0000022901.26754.0b

Appendix

Appendix 1: GENECLASS assignment table

Table S1: GENECLASS assignment table of Storsjøen using three different approaches; Bayesian (Rannala& Mountain=R&M, Baudouin & Lebrun = B&D), frequency based (Paetkau et al.) and distance based (Nei's standard distance (SD), minimum distance (MD) Nei's DA distance, Cavalli-Sforza and Edwards, and Goldstein et al.)

Populations	**Bayesian and frequency based**		
	R&M	***B&L***	***Paetkau***
Lag11	96.9	96.1	82.8
Lag10	110.5	110.3	89.4
MjN	113.4	107.6	97.6
Van	127.6	134.0	149.2
Hur	139.2	131.8	150.5
Hol	171.1	159.0	217.0
Ran	199.1	169.8	292.8
Eik	218.8	178.8	321.7
Tyr	223.2	181.6	319.3
Nor	283.8	246.8	505.5

Distance-based:		***Nei's $_{SD}$***		***Nei's $_{MD}$***		***Nei's $_{DA}$***		***Cavalli-Sforza***				***Goldstein***	
Populations	**Rank**	**Score %**	**Distance**	**Score%**	**Distance**	**Score %**	**Distance**	**Score %**	**Distance**	**Populations**	**Rank**	**Score %**	**Distance**
Lag11	1	30.334	0.010	30.879	0.006	23.510	0.033	15.849	0.140	*MjN*	1	26.682	1.193
Lag10	2	27.731	0.011	27.148	0.007	15.702	0.049	12.696	0.175	*Lag11*	2	26.052	1.222
MjN	3	18.223	0.017	18.053	0.010	15.348	0.051	12.033	0.185	*Lag10*	3	20.148	1.580
Hur	4	5.228	0.060	5.542	0.033	10.938	0.071	11.298	0.197	*Hur*	4	10.190	3.124
Van	5	4.877	0.064	5.222	0.035	9.545	0.081	10.017	0.222	*Hol*	5	6.322	5.036
Hol	6	3.588	0.087	3.815	0.048	6.928	0.112	9.092	0.244	*Van*	6	4.568	6.970
Ran	7	3.320	0.094	3.035	0.060	5.515	0.141	8.170	0.272	*Nor*	7	1.766	18.031
Tyr	8	2.945	0.106	2.719	0.067	4.774	0.162	7.419	0.300	*Eik*	8	1.518	20.97
Eik	9	2.824	0.110	2.617	0.070	4.492	0.173	7.329	0.303	*Tyr*	9	1.380	23.064
Nor	10	0.930	0.335	0.971	0.188	3.247	0.239	6.098	0.364	*Ran*	10	1.375	23.16

Appendix 2: BAPS assignment table

Table S2. Results from group level mixture analysis in BAPS with the four different clusters from optimal partition. Values show changes in log (marginal likelihood) if the different groups are moved to a different cluster. The value zero represents highest likelihood of a group belonging to a specific cluster.

Changes in log(marginal likelihood) if group i is moved to cluster j:

Group	**1**	**2**	**3**	**4**
Eik :	-269.8	-349.5	-492.4	0
Hol :	0	-61.4	-306.1	-340.1
Hur :	-82.4	0	-279.2	-323.2
Lag09:	-105.6	0	-304.5	-280.8
MjN :	-149.8	0	-408.3	-413.4
Lag11:	-183.4	0	-520.5	-547.9
Nor :	-306.1	-484.8	0	-762.3
Ran :	-230.5	-300.8	-518.6	0
Sto :	-179.5	0	-437.9	-407.0
Tyr :	-275.8	-395.8	-595.9	0
Van :	-57.9	0	-244.4	-456.4

Appendix 3: LnP(K) and ΔK-values from STRUCTURE

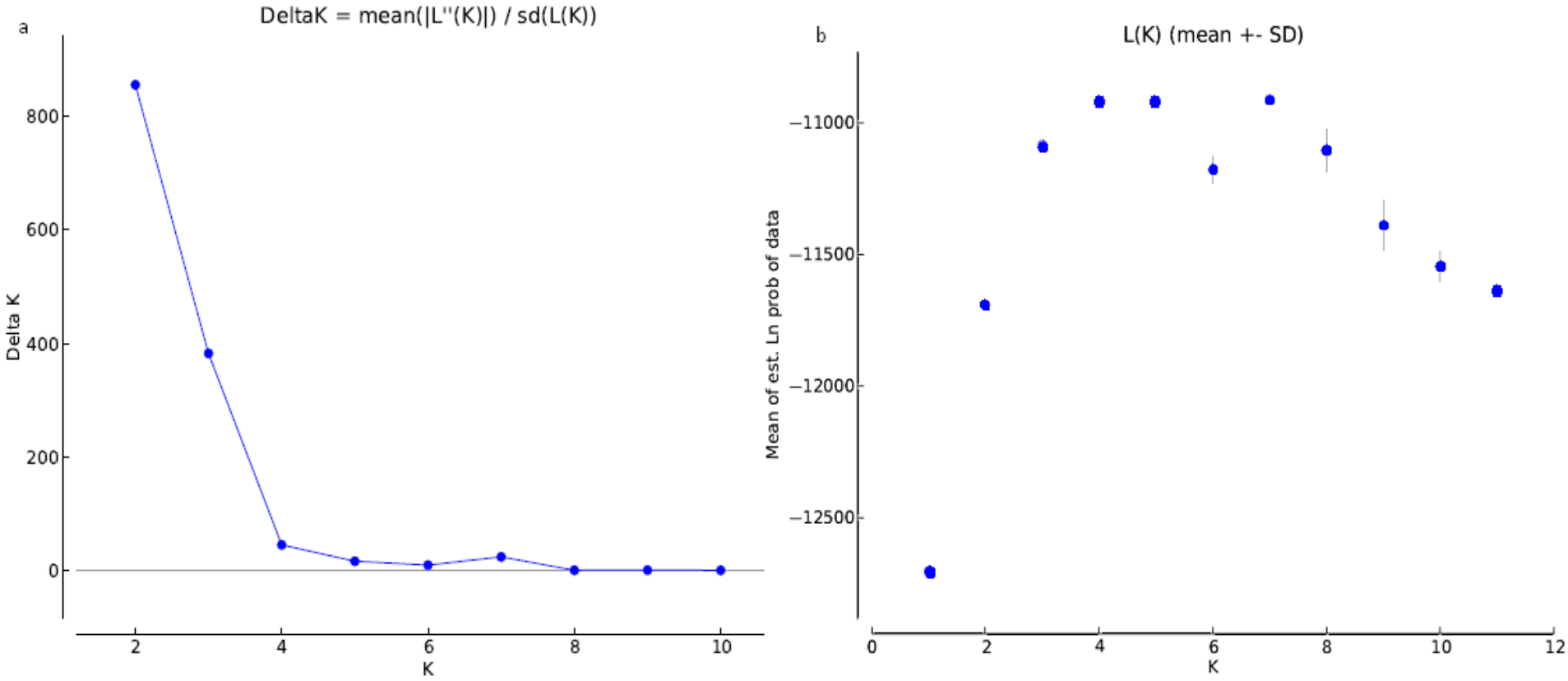

Figure S1. Probability of number of clusters for the 11 smelt populations. a) Most likely number of clusters with highest values of ΔK, b) most likely number of clusters with highest values of LnP(K).

Appendix 4: Assignment table with removed locus Oep539

Table S3. Assignment of Storsjøen to potential sources without locus Oep539. Using Bayesian clustering from STRUCTURE with prior population information(i.e. trained clustering), trained clustering from BAPS and 3 different approaches (Bayesian, frequency and distance based) in GENECLASS 2 with 8 different tests (Bayesian, Rannala & Mountain, Baudouin & Lebrun, Frequency based; Paetkau et al., Distance based; Nei's standard distance (Nei's $_{SD}$), Nei's minimum distance (Nei's $_{MD}$), Nei's DA distance (Nei's $_{DA}$) Cavalli-Sforza and Edwards distance and Goldstein's et al. distance). STRUCTURE assignment presented with proportion of membership values to the same cluster, BAPS assignment presented with changes in Log (marginal likelihood) if Storsjøen is moved to a different cluster (the value zero represents most likely cluster). GENECLASS 2 results are presented with percent score of most likely source (significantly different scores marked with superscript stars, threshold < 0.01) for the Bayesian and frequency based methods, and with percent score for the distance based methods. Highest likelihood marked in bold.

Populations	STRUCTURE	BAPS	GENECLASS							
			Bayesian approach: Score %		Frequency: Score %	Distance based: Score %				
	Proportion of membership to the same cluster	Changes in Log (marginal likelihood) if sample is moved to cluster :j	*Rannala & Mountain*	*Baudouin & Lebrun*	*Paetkau et al*	*Nei's $_{SD}$*	*Nei's $_{MD}$*	*Nei's $_{DA}$*	*Cavalli-Sforza*	*Goldstein*
Lag11	0.714	**0**	**100.0***	**100.0***	**100.0***	**29.411**	**29.875**	**22.883**	**15.378**	25.213
Lag09	0.730	-28.4	0.0	0.0	0.0	26.950	26.385	15.489	12.509	14.202
MjN	**0.763**	-36.4	0.0	0.0	0.0	17.714	17.444	15.178	11.859	**35.118**
Van	0.170	-61.2	0.0	0.0	0.0	6.808	7.107	11.297	11.585	4.567
Hur	0.303	-85.7	0.0	0.0	0.0	5.602	5.963	9.897	10.248	9.808
Hol	0.051	-169.5	0.0	0.0	0.0	3.453	3.670	6.727	8.804	5.881
Ran	0.015	-221.1	0.0	0.0	0.0	3.243	3.033	5.593	8.287	1.584
Eik	0.002	-280.7	0.0	0.0	0.0	2.866	2.705	4.807	7.469	1.656
Tyr	0.006	-268.9	0.0	0.0	0.0	2.743	2.593	4.498	7.355	1.591
Nor	0.029	-366.8	0.0	0.0	0.0	1.209	1.225	3.631	6.505	1.108

Appendix 5: Phylogenetic neighbor joining trees

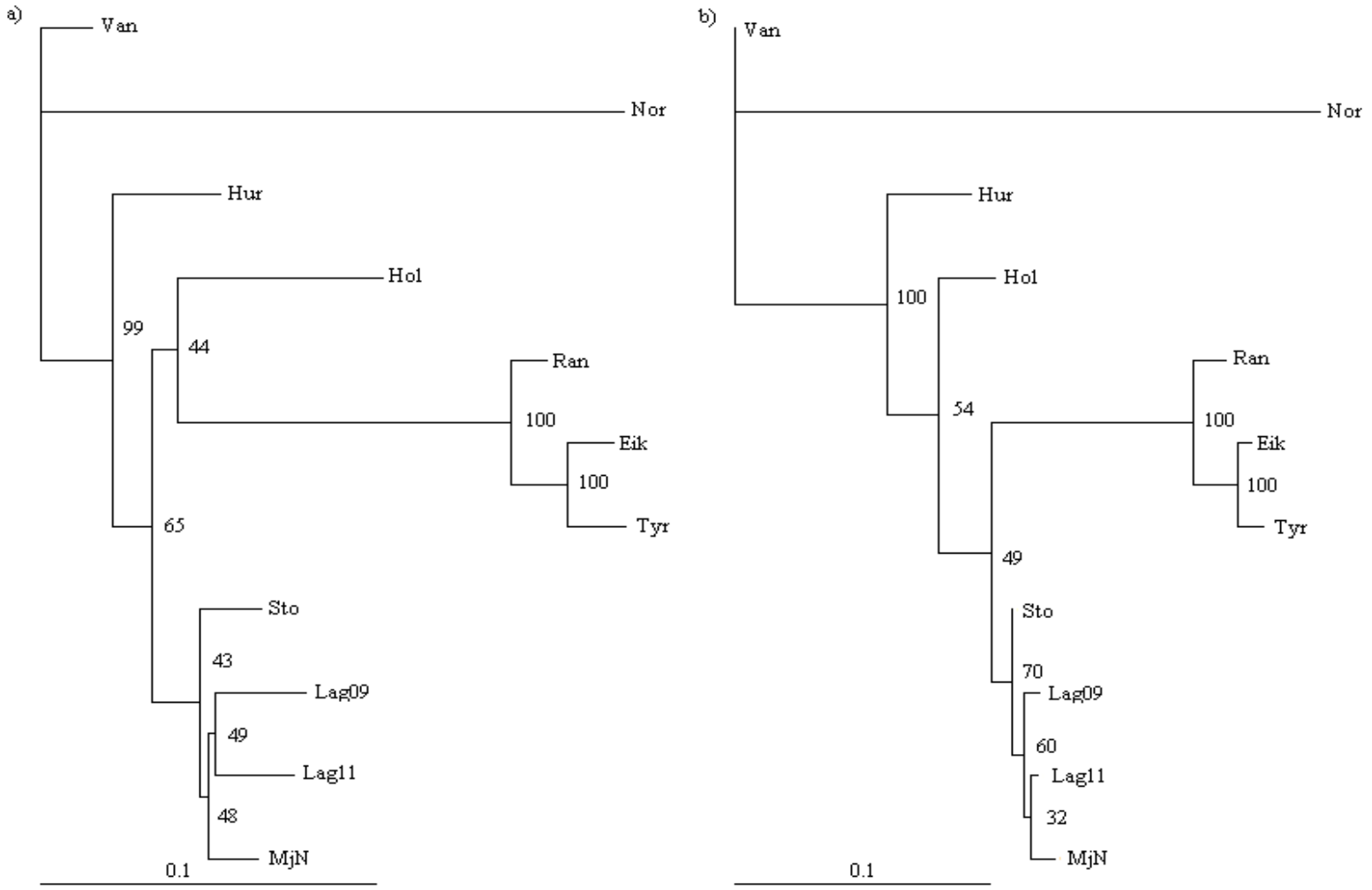

Figure S2. Phylogenetic neighbor joining tree with bootstrap support using Nei's standard distance (a), and Nei's minimum distance (b).

Appendix 6: Microsatellite primer & multiplex details

Table S4. Microsatellite primers and multiplex design for Osmerus eperlanus: Locus ID, fluorophor (F-Co), references (Ref), alignment temperature (Ta), PCR multiplex assignment (mplx), repeats (Rep), concentration µm (Conc).

Locus ID	Ta	Rep	Range	mplx	F-Co	Conc µm	Ref.
Oep5.67	59	2	95-101	Oe1	6FAM	0.4	[1]
Oep5.39	59	2	150-164	Oe1	6FAM	0.8	[1]
Oep7.50	59	2	85-123	Oe1	VIC	0.8	[1]
M-Omo6	59	4	168-206	Oe1	VIC	0.8	[2]
Oep5.38	59	2	111-129	Oe1	NED	0.8	[1]
Oep6.10	59	2	188-202	Oe1	NED	1.2	[1]
Oep7.11	59	2	158-166	Oe1	PET	0.6	[1]
Oep3.80	59	2	236-244	Oe1	PET	1.2	[1]
Oep1.35	59	2	122-133	Oe2	PET	0.4	[1]
Oep5.59	59	2	169-213	Oe2	PET	1.2	[1]
M-Omo1	59	4	104-162	Oe2	NED	0.4	[2]
M-Omo4	59	4	170-222	Oe2	NED	0.6	[2]
M-Omo11	59	4	134-214	Oe2	VIC	0.6	[2]
OSMOLav12	59	2	123-171	Oe2	6FAM	2.4	[3]
Oep3.84	59	2	241-251	Oe2	6FAM	1.2	[1]

Ref. [1] Taylor, Taylor, McCarthy, and Beaumont (2008), [2] Coulson, Paterson, Green, Kepkay, and Bentzen (2006), [3] Saint-Laurent, Legault, and Bernatchez (2003).

Appendix 7: Locus details per population (mean & SE)

Table S5. Mean and standard error over loci for each population; sample size (N), number of alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho), expected (He) and unbiased expected heterozygosity (uHe), percentage of polymorphic loci (Pol).

Pop		N	Na	Ne	I	Ho	He	uHe	Pol
Eik	**Mean**	36	3.3	2.0	0.6	0.3	0.3	0.3	64.3
	SE		0.7	0.4	0.2	0.1	0.1	0.1	
Hol	**Mean**	40	6.2	2.8	1.0	0.4	0.4	0.4	92.9
	SE		1.1	0.5	0.2	0.1	0.1	0.1	
Hur	**Mean**	30	5.5	2.8	1.0	0.4	0.5	0.5	78.6
	SE		1.0	0.5	0.2	0.1	0.1	0.1	
Lag09	**Mean**	21	5.2	2.6	0.9	0.4	0.4	0.4	92.9
	SE		1.0	0.5	0.2	0.1	0.1	0.1	
MjN	**Mean**	35	6.0	2.6	0.9	0.4	0.4	0.4	92.9
	SE		1.1	0.5	0.2	0.1	0.1	0.1	
Lag11	**Mean**	56	6.9	2.8	1.0	0.4	0.4	0.4	71.4
	SE		1.4	0.6	0.2	0.1	0.1	0.1	
Nor	**Mean**	37	3.7	1.8	0.6	0.3	0.3	0.3	64.3
	SE		0.7	0.3	0.2	0.1	0.1	0.1	
Ran	**Mean**	39	4.3	1.9	0.6	0.3	0.3	0.3	71.4
	SE		0.9	0.3	0.2	0.1	0.1	0.1	
Sto	**Mean**	42	6.2	2.7	1.0	0.4	0.4	0.4	85.7
	SE		1.1	0.5	0.2	0.1	0.1	0.1	
Tyr	**Mean**	39	4.0	2.0	0.6	0.3	0.3	0.3	78.6
	SE		0.7	0.4	0.2	0.1	0.1	0.1	
Van	**Mean**	39	7.6	3.6	1.2	0.5	0.5	0.5	85.7
	SE		1.3	0.6	0.2	0.1	0.1	0.1	

Appendix 8: Locus detail per population

Table S6. Locus details per population. Allelic richness (A), standardized- allelic (A_r), and private allelic richness (A_p).

		Eik	*Hol*	*Hur*	*Lag09*	*MjN*	*Lag11*	*Nor*	*Ran*	*Sto*	*Tyr*	*Van*
M-Omo6	A	1.0	2.4	2.4	1.9	1.8	2.3	1.0	1.5	1.9	1.9	1.9
	Ar	1.0	2.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0
	Ap	0.0	0.1	0.2	0.0	0.0	0.5	0.0	0.0	0.0	0.1	0.0
Oep380	A	1.0	1.9	1.8	1.9	1.8	1.0	1.0	1.0	1.0	1.9	1.0
	Ar	1.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0	1.0
	Ap	0.0	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.9	0.0
Oep538	A	6.4	7.0	8.7	9.8	8.5	8.6	4.9	7.5	8.2	6.7	9.8
	Ar	6.0	7.0	9.0	10.0	8.0	9.0	5.0	8.0	8.0	7.0	10.0
	Ap	0.0	0.0	0.5	0.9	0.3	0.7	0.2	0.0	0.0	0.5	0.4
Oep539	A	1.5	2.5	2.6	2.9	3.8	2.9	2.0	1.0	2.4	1.0	4.2
	Ar	2.0	2.0	3.0	3.0	4.0	3.0	2.0	1.0	2.0	1.0	4.0
	Ap	0.0	0.5	0.6	0.4	0.3	0.3	0.0	0.0	0.0	0.0	1.0
Oep567	A	1.0	1.0	1.0	1.9	1.0	1.0	1.0	1.5	1.0	1.0	1.0
	Ar	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Ap	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.5	0.0	0.0	0.0
Oep610	A	4.0	6.7	7.8	7.0	6.2	7.3	2.9	5.3	8.3	4.8	10.0
	Ar	4.0	7.0	8.0	7.0	6.0	7.0	3.0	5.0	8.0	5.0	10.0
	Ap	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0
Oep711	A	1.0	1.5	1.0	1.9	1.5	1.0	1.0	1.0	1.4	1.5	1.5
	Ar	1.0	1.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
	Ap	0.0	0.3	0.0	0.9	0.5	0.0	0.0	0.0	0.2	0.5	0.5
Oep750	A	4.0	4.8	8.2	10.5	10.0	9.2	5.0	5.2	8.8	3.6	10.4
	Ar	4.0	5.0	8.0	11.0	10.0	9.0	5.0	5.0	9.0	4.0	10.0
	Ap	0.1	0.0	0.0	0.4	0.1	0.1	0.0	0.0	0.0	0.6	1.0
M-Omo1	A	6.6	9.6	6.4	9.5	6.4	7.8	6.5	7.0	7.7	8.2	9.2
	Ar	7.0	10.0	6.0	10.0	6.0	8.0	6.0	7.0	8.0	8.0	9.0
	Ap	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1
M-Omo11	A	3.0	7.7	7.8	9.0	9.8	9.3	5.6	4.0	8.2	3.5	8.3
	Ar	3.0	8.0	8.0	9.0	10.0	9.0	6.0	4.0	8.0	3.0	8.0
	Ap	0.0	0.4	0.0	0.0	0.2	0.3	0.0	0.7	0.5	0.0	0.0
OsmoLav1	A	6.6	8.7	7.4	5.0	5.5	9.1	2.6	6.5	8.1	5.7	8.3
	Ar	7.0	9.0	7.0	5.0	5.0	9.0	3.0	7.0	8.0	6.0	8.0
	Ap	0.4	0.5	0.4	0.0	0.0	0.8	0.7	0.0	1.0	0.3	0.6
Oep135	A	1.0	1.5	1.0	1.0	1.5	1.0	1.0	1.0	1.4	1.7	1.7
	Ar	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	2.0	2.0
	Ap	0.0	0.5	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.1	0.1
Oep384	A	1.5	3.7	3.5	1.9	2.6	3.6	3.5	1.5	3.0	1.0	6.5
	Ar	2.0	4.0	4.0	2.0	3.0	4.0	3.0	1.0	3.0	1.0	6.0
	Ap	0.0	0.6	0.1	0.0	0.4	0.3	0.2	0.0	0.1	0.0	1.0
Oep559	A	2.4	9.0	8.3	6.8	9.3	8.8	5.2	4.1	8.4	2.6	10.1
	Ar	2.0	9.0	8.0	7.0	9.0	9.0	5.0	4.0	8.0	3.0	10.0
	Ap	0.1	2.0	0.1	0.4	0.8	0.5	0.0	0.3	0.1	0.5	0.6