UNIVERSITETET I TROMSØ UIT

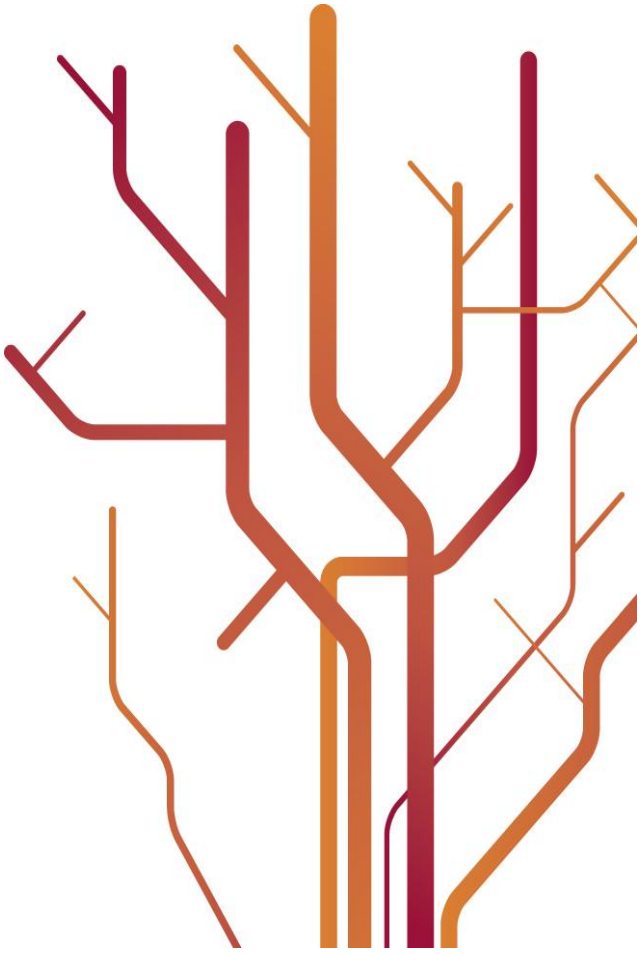
FACULTY OF BIOSCIENCES, FISHERIES AND ECONOMICS
DEPARTMENT OF ARCTIC AND MARINE BIOLOGY

Adaptive radiation of the European whitefish, *Coregonus lavaretus* (L.), in the Pasvik watercourse: the genetic description of a new morph

Marjorie Couton

BIO - 3950 Master's thesis in biology

November 2012



UNIVERSITETET I TROMSØ UIT

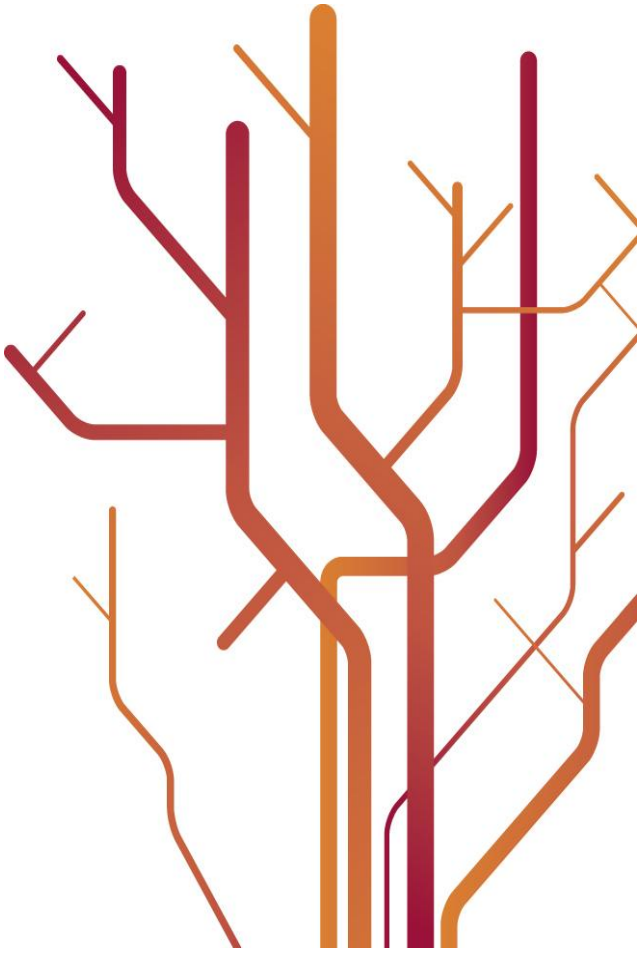
FACULTY OF BIOSCIENCES, FISHERIES AND ECONOMICS
DEPARTMENT OF ARCTIC AND MARINE BIOLOGY

Adaptive radiation of the European whitefish, *Coregonus lavaretus* (L.), in the Pasvik watercourse: the genetic description of a new morph

Marjorie Couton

BIO - 3950 Master's thesis in biology

November 2012



ACKNOWLEDGMENTS

First of all, I would like to thank my supervisors Rune Knudsen, Per-Arne Amundsen and Kim Præbel for their never failing support. Thank you Rune for your enthusiasm and your sense of humor. Thank you Per-Arne for your encouragements and your ecological expertise. And thank you to Kim for always being so enthusiastic about my work, even when I was not, and for being so available, even on the other side of the earth.

I am grateful to Kimmo Kahilainen for gathering the samples in Finland as well as all the people who took part on the field work. Thank you also for the helpful comments you gave me on this thesis. Thank you to Chris Harrod for the stable isotope data.

Then I would like to thank the freshwater ecology group for being so friendly and especially Laina for teaching me how to cook typical Norwegian meals. Thank you to the other master students as well for the now famous “Tuesday coffee” and this amazing trip in Scotland.

Thank you to all the people working in genetic lab for keeping me company during the countless hours I spend there and especially to Tanja.

Finally I would like to show a special appreciation to my office mates: Conny, Kristin, Sigrid, Birgitte for supporting me during this never-ending month of constant writing. All the coffe breaks, lunch and dinner would not have been the same without you. Thank you Emma for sharing the couch with me and cheering me up when I needed it. Our breakfast were memorable. And last but not least, thank you to Thomas, Roger, Bill and Stacy for their motivational stares. Your silence was worth a thousand words.

The study was financially supported by the Norwegian Research Council (grants no. 186320/V40 and 183984/S30) and the Finnish Ministry of Agriculture and Forestry for the field work.

Tromsø, November 2012

Marjorie Couton

TABLE OF CONTENTS

SUMMARY

Sympatric occurring fish morphs in postglacial lakes usually exhibit differences in morphology and physiology driven by adaptations to differential trophic niches. The European whitefish (*Coregonus lavaretus* (L.)) is a highly variable fish species, with more than 200 intraspecific forms described in Europe. The morphs usually differ in their number of gill-rakers, therefore this trait has been traditionally used for whitefish classification. According to this taxonomy, three different morphs can be distinguished in northern Fennoscandia: the large sparsely rakered (LSR), the densely rakered (DR), and the small sparsely rakerd (SSR) morphs. They all exhibit differences in morphology, diet, habitat, and physiology. Recently, a new morph has been discovered in several lakes of the Pasvik watercourse which displays densely rakered gills and an external morphology similar to the LSR whitefish, and was called large densely rakered (LDR). In this study, genetic data from 18 microsatellites markers were used to evaluate the genetic differentiation and the possible origin of this new morph in three Finnish lakes of the Pasvik watercourse. The LDR morph in each lake was found to be genetically different from the three other morphs. Several possible origins were suggested, but the sympatric speciation from either the LSR or the DR morphs was the hypothesis that gained the most support from the results. Moreover, the three different LDR populations were found to have a common origin, suggesting that the divergence occurred only once after the last ice retreat and the same population divided into three when the different lakes were formed.

DEFINITIONS

Adaptive radiation: evolution of ecological and phenotypic diversity within a rapidly multiplying lineage (Schluter 2000).

Allelic richness (N_{AR}): mean number of alleles across loci per population normalized for differences in sample size.

Allopatry: geographical separation between two or more populations rendering encounter between their members impossible (Glossary 2001).

Assortative mating: non-random choice of a mate because individuals select a partner on phenotypic characteristics (e.g. morphology, coloration, behavior) (Glossary 2001).

Bottleneck: short-term drastic reduction in population size, resulting in a loss of genetic diversity and heterozygosity (Hartl & Clark 2007).

Divergent natural selection: selection arising from environmental forces acting on different phenotypic traits (morphology, physiology, behavior) resulting in two different phenotypes. Intermediates are never favored (Gillespie 2009).

Drift (random genetic drift): stochastic fluctuation in allele frequency from generation to generation (Hartl & Clark 2007).

Ecological speciation: process by which barriers to gene flow evolve between populations as a result of ecologically based divergent natural selection (Gillespie 2009).

Founder effect: occurs when a small group of emigrants from an established subpopulation founds a new subpopulation. It is often accompanied by a severe population bottleneck (Hartl & Clark 2007).

F_{ST}: is a value measuring the genetic differentiation between populations and varies from 0 (no divergence between populations) to 1 (different alleles are fixed in the populations) (Hartl & Clark 2007). In this study the F_{ST} values will be estimated according to Weir and Cockerham (1984) formula (see p.15).

The Hardy-Weinberg equilibrium (HWE): This principle assumes that the organism is diploid with a sexual reproduction under random mating (panmixie), and nonoverlapping generations. It also assumes that no evolution forces are applied to the population, i.e. no mutation, migration or selection, and infinite population size to cancel the effect of drift. If a population is in HWE, the genotypic frequencies are stable over generations. If a population is not found to be in HWE, one or several of the previous conditions are not fulfilled (Hartl & Clark 2007).

Heterozygosity: proportion of heterozygotes in the population. The observed heterozygosity (H_o) is the proportion actually observed of heterozygotes in the population as opposed to the expected heterozygosity (H_e) which is the proportion of heterozygotes deduced from the Hardy-Weinberg principle from the sampled set of alleles.

Linkage disequilibrium (LD): genes are said in a state of linkage equilibrium if their alleles are in random association with one another. However, some situations can produce a state of linkage disequilibrium (LD) between genes, such as natural selection, substructuring of the population or a non-random reproduction (Hartl & Clark 2007).

Parallel evolution: evolution of similar phenotypic traits repeatedly in independent populations responding to a similar environmental pressure (Rundle *et al.* 2000).

Parapatry: partial geographical isolation such that only the members from the populations at the contact edges encounter one another (Glossary 2001).

Private alleles: alleles specific to a population.

Private allelic richness (N_{PAR}): mean number of private alleles across loci per population normalized for differences in sample size.

Reinforcement: adaptive strengthening of prezygotic isolating mechanisms in a zone of secondary contact between two distinct populations by reducing hybrid fitness (Glossary 2001).

Reproductive isolation: absence (or severe restriction) of gene flow between populations in contact with each other (Glossary 2001).

Resource polymorphism: occurrence of different phenotypes associated with segregation in habitat and diet (Skulason & Smith 1995).

Secondary contact: co-occurrence in one area of two populations that were previously geographically isolated and had accumulated genetic differences (Glossary 2001).

Sympatry: absence of geographical separation, such that all individuals have the same chance of meeting each other (Glossary 2001).

Transgressive hybridization: hybrids between two populations, who form a new population and occupy a different niche from their parents (Seehausen 2004).

INTRODUCTION

Transformation of living organisms over time have been suggested on multiple occasions for almost 2500 years (Edelstein 1944), but it was not until Darwin published "The Origin of Species" in 1859 that the evolution concept as we know it today brought a new vision of life history. Understanding evolution and the emergence of new species is presently one of the major challenges in evolutionary biology and generates many passionate debates (Sobel *et al.* 2010). Although the general idea of species changing over time and the role of natural selection behind these changes have now been widely accepted by scientists (Schluter 2000), how new species are formed is still under active research (Nosil & Rundle 2009).

Speciation mechanisms could not be properly investigated without a common agreement on what is a species. The popularization of Mayr (1942) biological species concept, which defines species as members of populations that actually or potentially interbreed in nature, allows us to view speciation as the process by which barriers to genetic exchange evolve between populations. Speciation can thus be seen as ecological or non-ecological depending on the nature of the process leading to reproductive isolation (Nosil & Rundle 2009), but there is a broad agreement that adaptation to the environment (i.e. ecological speciation) plays a significant role in most cases (McKinnon *et al.* 2004; Schluter 2009; Sobel *et al.* 2010).

The speciation process can be seen as a continuum, going through four main states: first a situation of continuous variation within the population, then a state of discontinuous variation with minor reproductive isolation followed by the appearance of a strong but reversible reproductive isolation, finally leading to a situation of strong and irreversible reproductive isolation (Hendry *et al.* 2009). In ecological speciation, the build-up of reproductive isolation requires ecologically based divergent natural selection on adaptive traits and may be facilitated by a geographic separation of the populations (allopatry or parapatry) (Hendry *et al.* 2009). In an allopatric scenario, the reproductive isolation starts to build up as populations accumulate adaptations to unique aspects of their environments but a shift back to sympatry (secondary contact) might be necessary to strengthen the separation by reinforcement (Schluter 2001). In a case of sympatric speciation, the populations are not separated by a geographical barrier, thus a strong disruptive selection is required to overcome the homogenizing effect of gene flow (Nosil & Rundle 2009). Reproductive isolation evolves as a by-product of adaptation to different environments and

resources, and develops more quickly when the divergent selection is strong (Schluter 2001), suggesting a more rapid divergence in sympatry.

Adaptive radiation is generally initiated by colonization of a newly available environment with expansion of the ecological possibilities of the invading taxon. Then, a process of specialization leads to the creation of new species from a common ancestor, generally diverging in morphology, physiology and behavior (Gillespie 2009). Resource polymorphism is a common feature in adaptive radiation (Skulason & Smith 1995) as seen in the famous example of Darwin's finches (Grant & Grant 1994), and the best examples are from isolated archipelagoes or similar island-like settings such as subarctic lakes (Gillespie 2009).

Few animal species can be seen as truly Arctic in the sense that they live and spawn in the polar zone and most or all of their distributional range is within the Arctic. Amongst freshwater fish, most of the typical Arctic species belong to the families Coregonidae and Salmonidae (Reshetnikov 2004), which are well known to the scientists for their divergence in polymorphic forms (Hudson *et al.* 2007; Klemetsen 2010). The European whitefish (*Coregonus lavaretus* (L.)) is distributed in the central and northern part of Europe and Russia (Bernatchez & Dodson 1994) and is the most divergent of all the coregonid fishes (Svärdson 1957; Østbye *et al.* 2005a; Hudson *et al.* 2007). This species can be considered as a species complex to avoid any taxonomical problems as more than 200 intraspecific forms have been described in Europe (Reshetnikov 2004; Østbye *et al.* 2005a). The morphs usually differ in the number of gill-rakers (series of cartilaginous projections at the inner part of the branchial gill arch) which is a highly heritable and stable character even when the fishes are transferred to new environments (Svärdson 1970; Hermida *et al.* 2002; Rogers & Bernatchez 2007). Therefore, this trait has been traditionally used for whitefish classification (Himberg 1970; Svärdson 1979).

The European whitefish is widely distributed in northern Fennoscandia and often dominates the fish biomass in the lakes where it is present (Reshetnikov 2004; Siwertsson *et al.* 2010; Kahilainen *et al.* 2011). According to the gill-raker classification, three different morphs may be distinguished within this region (Kahilainen & Østbye 2006; Siwertsson *et al.* 2010). The first, the large sparsely rakered morph (LSR), lives in the littoral zone and feeds on benthic macroinvertebrates. It is the largest of the three morphs, its gill-rakers are short and widely spaced, its mouth is subterminal and positioned low on the head, and it exhibits the classical color of whitefish (silvery sides, dark fins and back) (Kahilainen & Østbye 2006; Harrod *et al.* 2010). The second, the densely rakered morph (DR), lives in the pelagic zone and feeds on zooplankton. Its long and numerous gill-rakers act as cross flow filter directing the small prey particles towards the esophagus (Sanderson *et al.* 2001). It exhibits a slender body form and a smaller body size, more adapted to continuous swimming. Its eyes have a

lower positioning, the mouth is superior and its coloration is darker than the LSR morph (Kahilainen & Østbye 2006; Harrod *et al.* 2010). The last morph is called small sparsely rakered (SSR) and occupies the deep benthic (profundal) habitat. Its body is adapted to maneuvering with large pectoral fins since it feeds on benthic macroinvertebrates buried in soft sediments. Its eyes are also larger to overcome the very low levels of light in the profundal area. Its coloration is different with a brownish back and reddish fins (Kahilainen & Østbye 2006; Harrod *et al.* 2010). A representation of each morph with details of their gill-rakers is given in Figure 1.

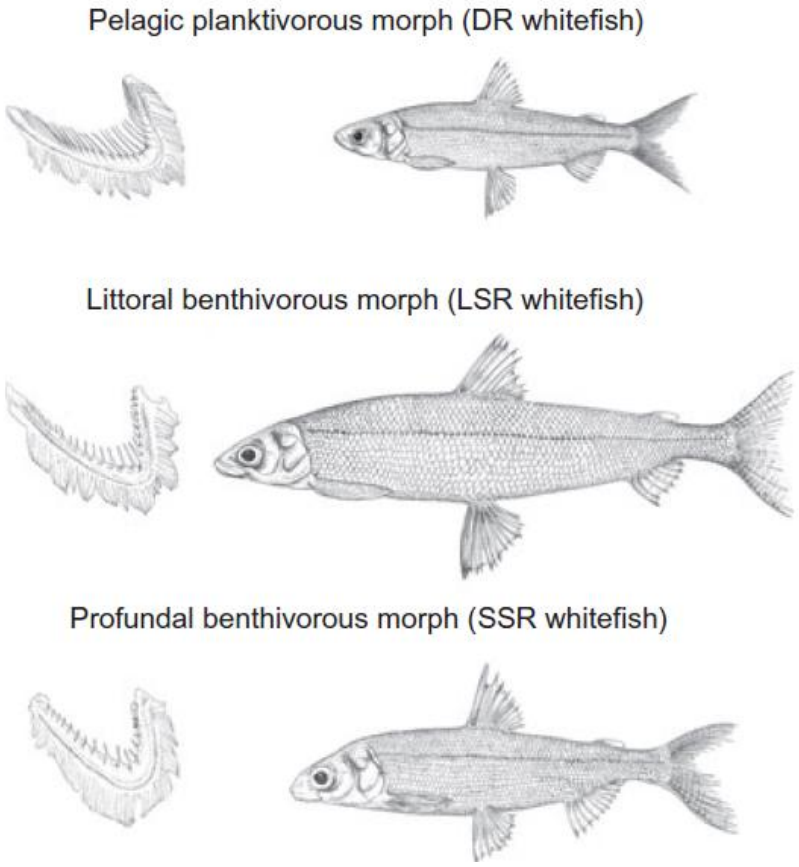


Figure 1 Morphology and gill raker characteristics of three sympatric whitefish morphs in the Pasvik watercourse: densely rakered (DR), large sparsely rakered (LSR), and small sparsely rakered (SSR) whitefish. Figure from Harrod *et al.* (2010)

A study conducted by Siwertsson *et al.* (2010) in the three main northbound watercourses in northern Fennoscandia (Alta/Kautokeino, Tana, and Pasvik watercourses) showed that all three whitefish morphs are not always found together in the lakes. These lakes can be sorted into three different categories. The monomorphic lakes contain only LSR whitefish, which are in this case generalists in diet and habitat utilization. The dimorphic lakes contain both LSR and DR morphs, each one of them utilizing more specialized food

resources and occupying either a benthic or a pelagic niche, respectively. Finally, the trimorphic lakes contain all three morphs and the LSR and SSR whitefish share the benthic zone with the LSR morph becoming more littoral and the SSR morph specializing on the profundal habitat. All morphs have been found to be genetically different (Præbel *et al.* in press).

The present distribution of coregonids in northern Fennoscandia was influenced by the development of drainage networks during the last glacial advance and retreat (10,000 to 25,000 years ago). A phylogeographic study based on mitochondrial DNA (mtDNA) by Østbye *et al.* (2005a) suggested that the northern Fennoscandian lakes have been colonized by whitefish coming from a glacial refugium on the western side of the Ural mountain ridge. Moreover, the genetic diversity in the three main northbound watercourses of Fennoscandia have been shown to decrease from East to West (from Pasvik to Alta/Kautokeino) and the alleles found in the Alta lakes represents only a subset of what is found within the Pasvik lakes (Østbye *et al.* 2006; Præbel *et al.* 2011). This supports the previous statement of an eastern colonization. Furthermore, all northern European populations appear to belong to the same mtDNA clade which indicates that the different morphs diverged after the colonization from a common ancestor by adaptive radiation into unoccupied niches (Østbye *et al.* 2005a).

The emergence of the different whitefish morphs can have three different origins: colonization by three different ancestors, divergence after secondary contact, or parallel sympatric speciation. The mtDNA analyses described above ruled out the first hypothesis, whereas Østbye *et al.* (2006) provided evidence for parallel speciation as the different populations within their phylogenetic study were grouped by lakes and not by morph. Moreover, LSR whitefish is the only morph found in allopatry and they are more generalists in diet and habitat choice than the other two (Kahilainen *et al.* 2007). Thus, the LSR morph is most likely the closest phenotype to the ancestral form, which could have then diverged into the two other morphs (Østbye *et al.* 2006; Præbel *et al.* submitted). Besides, no lakes are found with a monomorphic population of DR or SSR whitefish or with a dimorphic population of LSR and SSR, suggesting that the diversification process may follow a sequential pattern with first a separation along the benthic/pelagic axis, and then, a separation among the benthic habitats (Siwertsson *et al.* 2010). However, a more recent study by Præbel *et al.* (submitted) suggests that this radiation could be a combination of scenarios where the LSR and DR morphs have colonized the lakes in parallel whereas the SSR morph may have evolved from the LSR morph by sympatric divergence.

Recently, some individuals that could not be morphologically classified into one of the three existing morphs have been discovered in several lakes of the Pasvik watercourse (Figure 2). Their body size and external morphology are closely related to the LSR individuals but their gill-rakers are long and dense, similar to those of a DR whitefish (K. K. Kahilainen pers. comm.). This new type of individuals was called large densely rakered (LDR) whitefish.

The aims of this study can be reported as three main questions: 1) Are LDR individuals genetically different from the other three existing morphs? 2) If so, what mechanism(s) led to the divergence of this new morph? Four hypotheses can be suggested in this respect: a sympatric speciation from the LSR or DR population, a divergence after secondary contact, a transgressive hybridization, or the result of a stocked population. 3) Finally, are the LDR populations of the different lakes similar (suggesting a common origin) or different (suggesting parallel evolution or different origins)?

MATERIAL AND METHODS

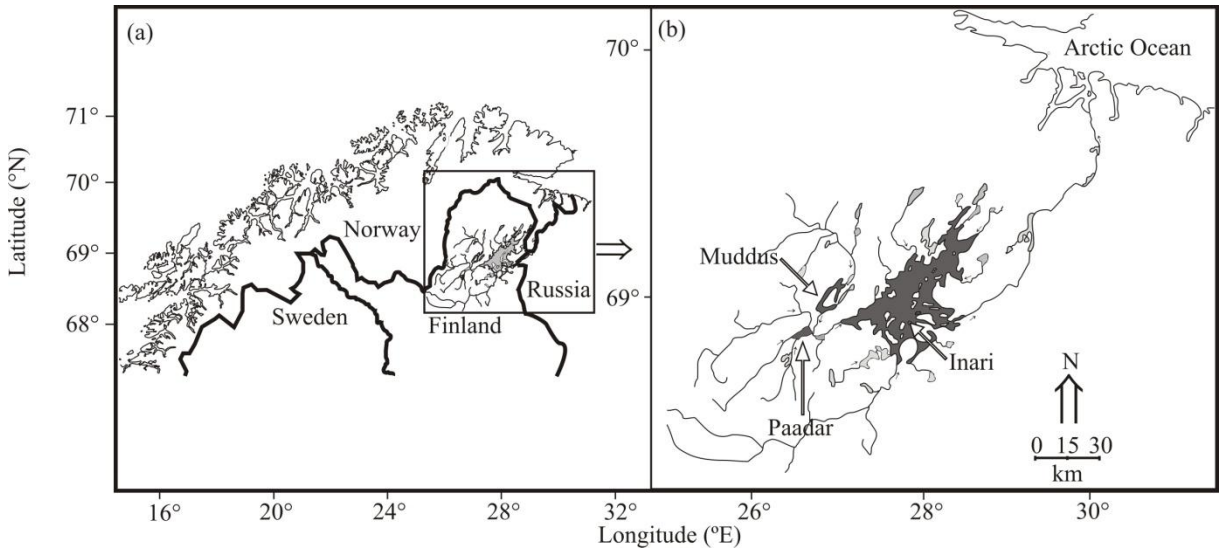


Figure 2. Map of (a) northern Fennoscandia and (b) Paatsjoki/Pasvik watercourse with the three study lakes: Muddusjärvi, Paadarjärvi, and Inarijärvi (provided by K. Kahilainen).

STUDY AREA

The samples were collected from three subarctic lakes in the Finnish Lapland situated in the large Paatsjoki/Pasvik watercourse. The headwaters of this watershed discharge into Lake Inarijärvi in northern Finland, run throughout Russia for 30 km, and finally form the border between Norway and Russia for a distance of approximately 120 km before entering the Arctic Ocean (Figure 2a). The total catchment area of the watershed is 18 403 km². The three lakes studied are Lake Muddusjärvi (L. Muddus), Lake Paadarjärvi (L. Paadar) and Lake Inarijärvi (L. Inari) (Figure 2b). The surface area of the lakes varies from 21 to 1043 km², with maximum and mean depths of 56-92 m and 8.5-14.3 m, respectively (Table 1). All the lakes are oligotrophic, neutral and with some humic impacts (Secchi depths of either 3 or 6 m). The rivers Kaamasjoki and Terstojoki, which are rich in humic substances, discharge their waters into L. Muddus, making the water brownish. All three lakes are connected via rivers with L. Muddus being the highest positioned and L. Inari the lowest. The ice-free period lasts from May-June to November-December (Jensen *et al.* 2008; Kahilainen *et al.* 2011).

Table 1 Background data on location, morphometry, and secchi depth of the study lakes.

Parameter	***Lake Muddusjärvi***	***Lake Paadarjärvi***	***Lake Inari***
Latitude (°N)	69°00’	68°52’	69°58’
Longitude (°E)	26°50’	26°35’	27°40’
Surface area (km²)	48	21	1043
Altitude (m above sea level)	146	144	118
Maximum depth (m)	73	56	92
Mean depth (m)	8.5	11.7	14.3
Secchi depth (m)	3	6	6

The natural fish fauna within these lakes consists of ten species: whitefish, Arctic charr (*Salvelinus alpinus* (L.)) except for L. Paadar, brown trout (*Salmo trutta* L.), grayling (*Thymallus thymallus* (L.)), perch (*Perca fluviatilis* L.), pike (*Esox lucius* L.), burbot (*Lota lota* (L.)), minnow (*Phoxinus phoxinus* (L.)), three-spined stickleback (*Gasterosteus aculeatus* (L.)), and nine-spined stickleback (*Pungitius pungitius* (L.))(Kahilainen *et al.* 2011). Besides these, other fish species have been introduced into the area and particularly in L. Inari, such as land-locked salmon (*Salmo salar* m. *sebago* (Girard)) and lake trout (*Salvelinus namaycush,* (Walbaum)); vendace (*Coregonus albula* (L.)) being the one with the most significant influence on the catches and the whole ecosystem of L. Inari (K. Kahilainen pers. comm.). Nevertheless, whitefish is the most numerous species in each lake.

The water level in L. Inari has been regulated since the 1940s (max. amplitude 2.36 m) for the production of hydroelectric power in Russia and Norway in the Paatsjoki/Pasvik system. Long-term regulation has caused considerable erosion of the shoreline in the littoral zone of the lake, which has reduced the availability of littoral benthic invertebrate prey and the number of spawning sites (Mutenia & Ahonen 1990). These effects were followed by a decline in the total annual catch within the lake, especially for the whitefish (Mutenia & Salonen 1994). As a result, the Finnish Water Court made the decision that one million one summer-old whitefish must be stocked into the lake each year. The compensation was initiated in the mid 1970s and the mean number of whitefish stocked in the late 1970s and 1980s was nearly 1.5 million per year. The stocked fish were mainly LSR whitefish from the river Ivalojoki (Salojärvi & Mutenia 1994).

SAMPLING

Whitefish were sampled from the three main habitat types (littoral, pelagic, and profundal) during ice-free periods between 2004 and 2009 using multimesh gillnets. The number of fish sampled by population and the population codes used are listed in Table 2. A set of eight nets was used, each having a length of 30 m and height of 1.8 m. The gillnet was randomly combined from eight nets having mesh sizes 12, 15, 20, 25, 30, 35, 45 and 60 mm (from knot to knot). The fishing sites were selected randomly in each depth zone (littoral 0-10 m, profundal 10-20 m, and pelagic 0-10 m). Nets were set in the evening at around 20.00 and collected in the morning around 08.00 (Kahilainen & Lehtonen 2003).

Whitefish were chilled in ice after sampling. The catch was field identified according to the well established characteristics of the morphs, such as fish morphology, and number of gill-rakers and their morphometry (Amundsen *et al.* 2004; Kahilainen & Østbye 2006). The LDR whitefish morph was identified by having the morphological characteristics of the LSR morph and the gill-rakers of the DR morph (K. Kahilainen pers. comm.). After determining the morph, each individual was given an identification code number, and total length (± 1 mm) and blotted mass (± 0.1 g) were recorded. The first left gill arch was stored in 96% EtOH for counting of gill-rakers and DNA extraction.

Table 2 Populations details for each lake with the codes used, morph abbreviation, the number of individuals used for the genetic analyses, and the year the fish were captured.

Lake	***Population code***	***Morph***	***Nb. Individuals***	***Year***
L. Inari	ID	DR	35	2009
	ILD	LDR	38	2009
	IL	LSR	35	2009
	IS	SSR	29	2009
L. Muddus	MD	DR	38	2006
	MLD	LDR	47	2006
	ML	LSR	48	2006
	MS	SSR	42	2006
L. Paadar	PD	DR	40	2004
	PLD	LDR	43	2004
	PL	LSR	35	2004
	PS	SSR	35	2004

GILL-RAKER COUNT AND STABLE ISOTOPE ANALYSES

The number of gill-rakers was determined from the first left branchial gill arch under a preparation microscope. The average number of gill-rakers between the whitefish morphs was compared using an ANOVA to test for general differences. Then, a Tukey's honest significant differences test was performed to detect pair-wise differences. These statistical analyses were implemented in R (R Development Core Team 2011).

A small sample of dorsal muscle tissue was excised from each whitefish for stable isotope analyses. Samples were dried, ground, weighed into tin cups and analyzed for carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotope ratios and elemental C and N composition. These analyses were not performed by the author and the values were provided by K. Kahilainen (see Harrod *et al.* (2010) for more details). Stable isotopes give a long-term integrated signal of food intake. In this case, the carbon and nitrogen ratios within fish muscle tissue reflect the assimilated food during the summer growth period (Perga & Gerdeaux 2005). Stable isotopes can distinguish between resources from the three main habitats within the lakes (Harrod *et al.* 2010). Resources from the littoral and pelagic habitats exhibit the same $\delta^{15}N$ enrichment but differ in carbon values since pelagic phytoplankton is depleted in $\delta^{13}C$ compared to benthic algae (Vander Zanden & Rasmussen 1999). Meanwhile, resources from the benthic habitat exhibit the same carbon values as those from the pelagic zone but are more enriched in $\delta^{15}N$ due to the presence of the detritus food chain (Post 2002). General differences in carbon enrichment and nitrogen enrichment between populations were tested separately using an ANOVA, followed by a pair-wise Tukey's honest significant differences test in R (R Development Core Team 2011).

GENETIC ANALYSES

DNA extraction

DNA was extracted from a piece of dorsal muscle or gill with the DNA extraction kit DNeasy Blood and Tissue Kit (QIAGEN®) using the supplied manufacturer protocol. PCR were run using the QIAGEN® Multiplex PCR Kit following the manufacturer's instructions and a GeneAmp 9700 thermal cycler (Applied Biosystems). Nineteen microsatellite loci were analyzed (Table 3). For amplification the loci were arranged into four PCR multiplex panels (Præbel *et al.* in press) and each reaction consisted of 1.25 µl QIAGEN® Multiplex PCR Master Mix, 0.25 µl primer mix (multiplex 1 to 4), 0.5 µl water and 5-10 ng template DNA. The general PCR profile was: 95°C for 15 min followed by 25 or 26 cycles of 94°C for 30 s,

57°C/60°C/61°C for 3 min and 72°C for 1 min, with a final 60°C extension for 30 min. The number of cycles and the alignment temperature for each multiplex panel are summarized in Table 3. The PCR products were separated on an ABI 3130XL Automated Genetic Analyzer (Applied Biosystems) and the alleles were scored using the GeneMapper 3.7 software (Applied Biosystems). Each genotype was automatically binned in predefined allelic bins by the GeneMapper software and verified by visual inspection. All private alleles found in the subsequent data analysis were manually verified.

Table 3 Details of the 19 microsatellite loci used for designing four PCR multiplexes for *Coregonus lavaretus*. The locus ID, PCR multiplex assignment (Mplx), concentration (Conc.), fluorophor (Fph), alignment temperature (Ta), number of cycles, repeat motif (RM), and source are given. Equal concentration was used for the forward and reverse primers. Source 1: (Patton *et al.* 1997), 2: (Rogers *et al.* 2004), 3: (Winkler & Weiss 2008), 4: (Susnik *et al.* 1999), 5: (Turgeon *et al.* 1999)

Mplx	***Locus ID***	***Conc. (µM)***	***Fph***	***Ta (°C)***	***Nb. Cycles***	***RM***	***Source***
1	BWF1	2.50	PET	57	25	GA	1
	BWF2	1.00	PET	57	25	CA	1
	Cocl-Lav4	0.75	6-FAM	57	25	CA	2
	Cocl-Lav6	1.50	NED	57	25	GT	2
	Cocl-Lav10	0.50	NED	57	25	GT	2
	Cocl-Lav27	0.40	VIC	57	25	GT	2
	ClaTet3	1.25	PET	57	25	TGTC	3
	ClaTet13	1.50	6-FAM	57	25	GACA	3
	ClaTet18	2.00	VIC	57	25	GACA	3
2	Cocl-Lav18	0.80	PET	60	25	GA	2
	Cocl-Lav49	0.80	NED	60	25	GT	2
	Cocl-Lav52	2.50	6-FAM	60	25	GT	2
	BFRO018	0.80	PET	60	25	GT	4
3	ClaTet6	1.50	6-FAM	61	25	TGTC	3
	ClaTet9	1.00	VIC	61	25	TGTC	3
	ClaTet15	1.00	PET	61	25	GATA	3
4	ClaTet1	1.50	PET	60	26	GACA	3
	ClaTet17	1.50	6-FAM	60	26	GATA	3
	C2-157	1.00	VIC	60	26	GT	5

Hardy-Weinberg equilibrium and linkage disequilibrium

Genotyping errors are common and can bias further genetic analyses. To detect these errors, we used MICRO-CHECKER (Van Oosterhout *et al.* 2004). Four types of errors can be seen: allelic dropout which refers to the failure to amplify some alleles due to poor DNA quality or low DNA concentration, large allele dropout which is the preferential amplification of small alleles, scoring problem due to the presence of too many stutters around the real peaks, and the presence of null alleles which refers to an homozygote excess due to a mutation on the primer site leading to the non-amplification of some alleles. All these errors cause deviations from Hardy-Weinberg proportions and can be misinterpreted as inbreeding or assortative mating for instance. The software checks for deviations in Hardy-Weinberg equilibrium (HWE) and, if any, suggests a possible cause. However the results should be interpreted with caution since several natural causes can result in the same signature.

The software GENEPOP 4.0 (Rousset 2007) was used to estimate whether each locus is in HWE for each population, using an exact test which is a better approach for multiple alleles loci than large-sample goodness-of-fit tests (Guo & Thompson 1992). If a locus is not in HWE for several populations, it could indicate an amplification problem during the PCR or a scoring problem during genotyping. However, several loci not in HWE for the same population could indicate that this population does not fulfill one of the assumptions of the Hardy-Weinberg principle.

The state of linkage disequilibrium (LD) was tested for each locus-pair in each population using GENEPOP 4.0 (Rousset 2007). If a locus-pair exhibits LD for several populations, the loci could be situated on the same chromosome and should therefore be removed from the subsequent analyses to avoid misinterpretation. However, a lot of different locus-pairs found to be in LD in the same population could be a sign of inbreeding or population substructure.

All the p-values for HWE and LD tests were adjusted for multiple comparisons using sequential Bonferroni corrections (Rice 1989). As the number of tests increases, the probability of making one or more type I error (rejecting the null hypothesis if it is actually true) among the collection of tests increases. To circumvent this problem, we can adjust the significance levels for each test using the sequential Bonferroni method. The p-values are ranked from the largest to the smallest and the smallest p-value is tested at α/c, the next at $\alpha/(c-1)$, the next at $\alpha/(c-2)$, etc... where α is the nominated significance level and c is the number of comparisons. Testing stops when a non-significant result occurs (Quinn & Keough 2002). In this study, the p-values were tested for $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.005$, and $\alpha = 0.001$ successively.

Genetic diversity

The number of alleles (N_A) per locus per population, and the expected (H_e) heterozygosity, were estimated using GENALEX 6 (Peakall & Smouse 2006). The allelic richness (N_{AR}) and the private allelic richness (N_{PAR}) were estimated with HP-RARE 1.0 (Kalinowski 2005), using the rarefaction function to compensate for differences in sample size. N_{PAR} was estimated within lakes among populations first, and then between the three LDR populations.

Population structure

The F_{ST} value is a common tool for examining the overall level of genetic divergence among populations. The pair-wise F_{ST} values for each population-pair were calculated following Weir and Cockerham (1984) formula implemented by Slatkin (1991): $F_{ST} = \frac{f_0 - f_1}{1 - f_1}$ where f_0 is the probability of identity by descent of two different genes drawn from the same population and f_1 is the probability of identity by descent of two genes drawn from two different populations. The calculation was done using ARLEQUIN 3.5.1.2 (Excoffier & Lischer 2010) and the values were tested for statistical significance (10,000 permutations).

The number of potential genetic clusters (K) was estimated via the Bayesian clustering algorithm in STRUCTURE 2.3.2 (Pritchard *et al.* 2000) by identifying the highest mean log likelihood value for the data. The populations were analyzed over 15 independent runs with 50 000 burn-ins and 100 000 Markov Chain Monte Carlo (MCMC) repetitions using a model including admixture and correlated allele frequencies. In addition, the LOCPRIOR function, which uses sampling location information to improve clustering, was included in the model as recommended by Hubisz *et al.* (2009). Each of the 12 populations was coded as a sampling location. The STRUCTURE output depicts an individual as a horizontal bar divided into the proportion of membership (q) to the inferred number of clusters.

Recently, Evanno *et al.* (2005) assessed the ability of STRUCTURE to detect population structure according to a hierarchical model. They found that the value ΔK (derived as the global maximum of the curve when plotting the likelihood probabilities versus K), provided a stronger estimate of the true number of populations in a hierarchical system. Thus, each cluster must be reanalyzed separately to find further population structuring. First, all 12 populations were run and 2 clusters were found. One of these clusters was run another time (50,000 burn-ins, 100,000 MCMC) to be divided into 2 clusters and 3 groups: one group for each cluster and one with an unresolved pattern. This last group was then run a third time (100,000 burn-ins, 200,000 MCMC). See the Results section for further details.

Phylogeny

A phylogenetic tree was created using the software package PHYLIP 3.65 (Felsenstein 1989). First the data were bootstrapped 1,000 times. Then the data were analyzed using three different genetic distances: Nei's genetic distance (Nei 1972), Cavalli-Sforza chord measure (Cavalli-Sforza & Edwards 1967) and Reynolds *et al.* genetic distance (Reynolds *et al.* 1983). These three distances assume that the differences between populations arise with genetic drift. Nei's genetic distance assumes that all loci have the same rate of neutral mutation, that the populations are at mutation-drift equilibrium and that the effective population size remains constant over generations. The two other distances do not integrate mutations but they are more adapted for populations undergoing size changes since they adapt the effect of drift with the variation of population size (Felsenstein 2009). Since our LDR populations might be the result of a recent divergence, their population size could have varied greatly, therefore the Cavalli-Sforza chord measure and Reynolds *et al.* genetic distance might be more appropriate. The following step is to group the populations that are more closely genetically related. This step was realized using the Neighbor-joining method which is more appropriate for bootstrapped values (Felsenstein 2009). At this point, the software created all possible trees out of our data set. Then, the population groups emerging the largest number of times were summarized in a consensus tree using the setting "majority rule extended". This rule states that only the groups occurring more than 50% of the time are represented (Felsenstein 2009). Two Scottish whitefish populations (LSR morph) were used as out-groups.

Principal component analysis

Principal component analysis (PCA) was used to correlate genotypes and allele frequencies among all individuals in all the populations using no prior information regarding population or sampling identification. The analysis was performed using the program PCAGEN (Goudet 2004) which uses multilocus genotype data to investigate the correlation of allele frequencies and genotypes among all the individuals sampled. From the PCA results, a two-dimensional canonical plot of the first two principal components was produced. Significance was tested for each axis (1,000 permutations).

RESULTS

GILL-RAKER COUNT AND STABLE ISOTOPE ANALYSES

Comparison of the average gill-raker count within lakes indicated general differences between the morphs (L. Inari: $F_{3,130} = 492.6$, $p < 0.001$; L. Muddus: $F_{3,171} = 902.1$, $p < 0.001$; L. Paadar: $F_{3,124} = 276.0$, $p < 0.001$). Later pair-wise comparisons within each lake shows that all four populations were significantly different in L. Inari and L. Paadar whereas MD and MLD were not significantly different ($p = 0.93$) in L. Muddus. Comparisons across lakes indicates that ILD was significantly different to all the other populations ($p < 0.01$) whereas MLD was not significantly different to ID ($p = 0.06$), MD ($p = 0.93$) or PLD ($p = 0.98$) but was significantly different to PD ($p < 0.001$). PLD was significantly different to PD and ID ($p < 0.001$) but not to MD ($p = 0.22$) (Figure 3, Table S1). Despite some differences between populations, MLD and PLD displayed a densely rakered distribution, whereas ILD exhibits significantly less gill-rakers, although being much closer to the DR than the LSR populations.

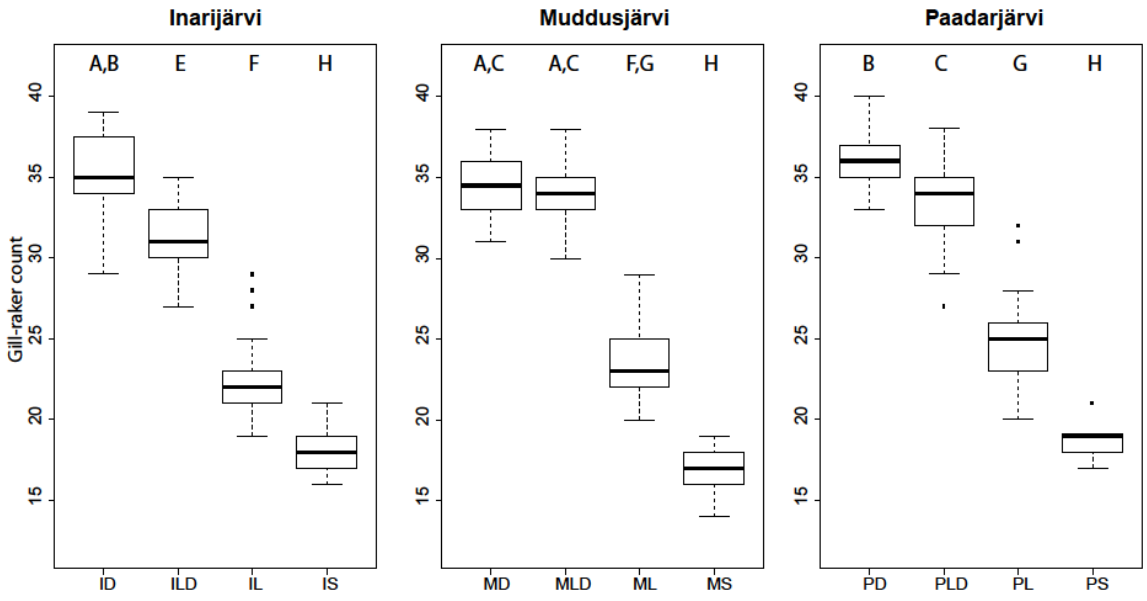


Figure 3 Boxplot of the gill-raker counts for each population in L. Inari, L. Muddus and L. Paadar showing the median (thick line), the interquartile (box), the other values with a maximum of 1.5 times the interquartile (whiskers), and outliers (circles). Populations with the same letters are not significantly different from each other ($p > 0.05$). See Table S1 for exact p-values.

The stable isotope values found in the present study were in accordance with other studies on whitefish morphs from the Pasvik watercourse (Harrod *et al.* 2010; Siwertsson 2012). Three groups can be differentiated: one for the SSR populations, one for the DR populations and the last one for the LSR and LDR populations (Figure 4). ILD and PLD were significantly different from all the other populations within their respective lakes, but MLD was not significantly different from ML. All three LDR populations were significantly different from each other (Figure 4, Table S2). The LDR populations of L. Inari and L. Paadar occupy a different isotopic niche from the other populations of their lake whereas MLD seems to share an isotopic niche with ML. Despite their differences, the ILD and PLD morphs seemed closer to the LSR morph of their respective lake. Moreover, the LDR and LSR morphs exhibited a higher variability in $\delta^{13}C$ enrichment than the DR and SSR morphs indicating a broader diet (Figure 4).

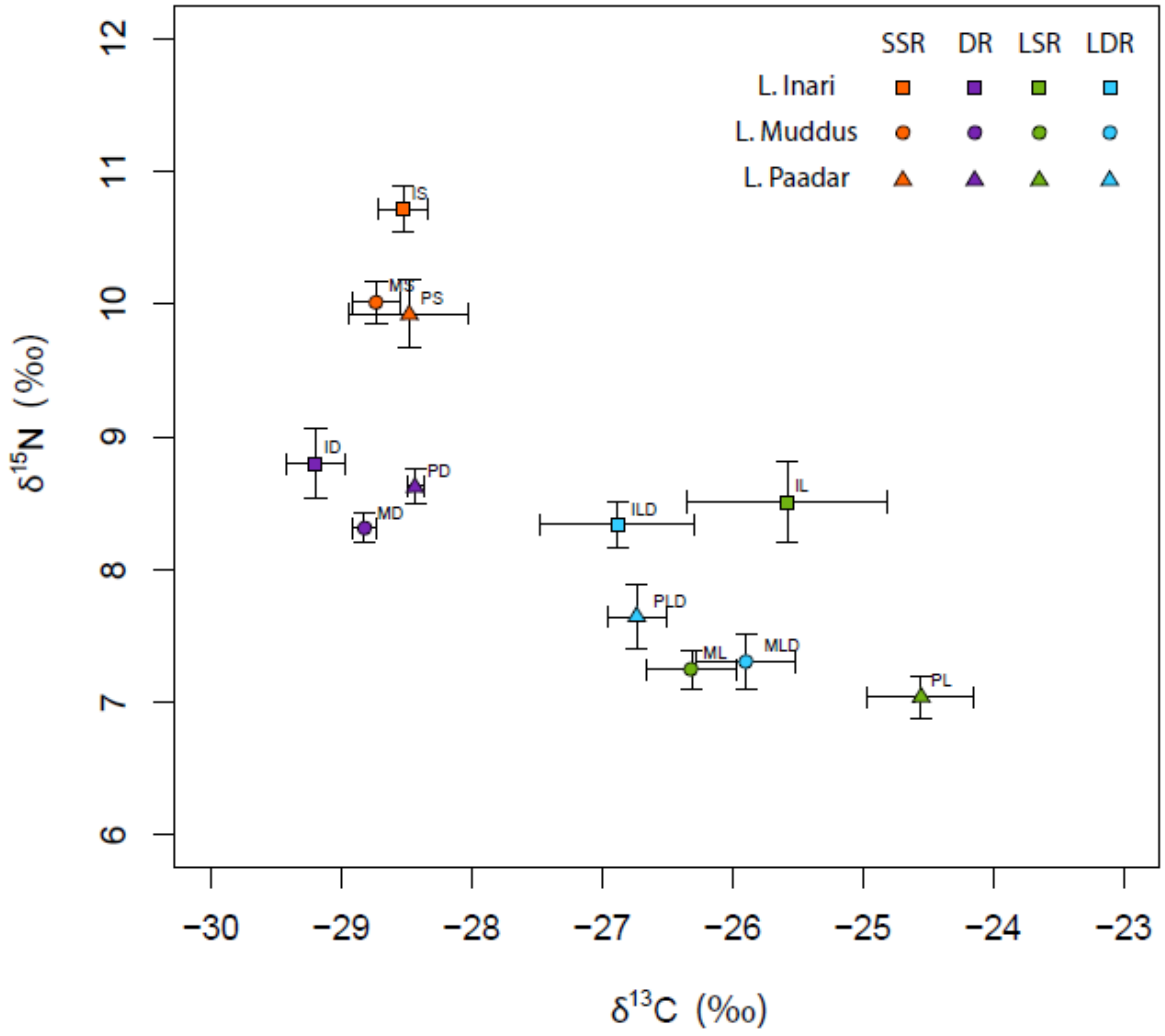


Figure 4 Isotopic signatures for each sympatric whitefish morph in three lakes from the Pasvik watercourse (mean ± 95% confidence interval). Population codes are explained in Table 2. The pair-wise mean differences of the stable isotope values have been tested using a Tukey HSD test and the p-values are given in the Table S2.

GENETIC ANALYSES

Hardy-Weinberg equilibrium and linkage disequilibrium

Homozygote excess were indicated by MICRO-CHECKER at BFRO018 (ML), BWF1 (IS, MD), C2_157 (ID, MD), ClaTet3 (ID), ClaTet6 (ILD, IL), ClaTet13 (ID, PD), ClaTet17 (ID, ILD, IL, ML, PD, PL), ClaTet18 (IS), Cocl-Lav10 (IS), and Cocl-Lav52 (PD) all indicated as caused by the presence of null alleles. However, the homozygote excess may also reflect genetic effects associated with the population history, e.g. inbreeding, expansion, unidirectional gene flow and general founder effects, and this study concerns populations that are expected to be influenced by these factors. No locus showed homozygote excess for more than two populations except ClaTet17 which exhibited this deviation for 6 out 12 populations.

GENEPOP data revealed departures from HWE after Bonferroni corrections only for the locus ClaTet17 ($p < 0.001$) in 4 out of 12 populations (ILD, IL, ML, PD). This result was concordant with the MICRO-CHECKER output, and the locus ClaTet17 was thus removed from further analyses. As the other loci exhibiting a homozygote excess were found to be in HWE, they were maintained in the subsequent analyses.

No locus-pair out of the 1836 tested was associated with LD after Bonferroni corrections ($p > 0.05$).

Genetic diversity and variation

In total, 242 alleles were observed across the 18 microsatellite loci for the 12 populations. The highest number of alleles ($N_A = 162$) and the lowest ($N_A = 98$) were found in L. Paadar for PD and PS, respectively. The number of alleles per locus across samples varied between 1 and 22 (Table S3). Among the 18 microsatellite loci, the locus Cocl-Lav27 displayed an interesting pattern with seven different alleles across populations, the allele 185 being fixed in four populations (MS, IL, IS, and PS). MLD and PLD exhibited only two alleles (179 and 185), which were found in MD, ML, PD, and PL as well. However, ILD was characterized by four alleles (117, 179, 181, and 185), two of them also being exclusively found in MD (Figure 5).

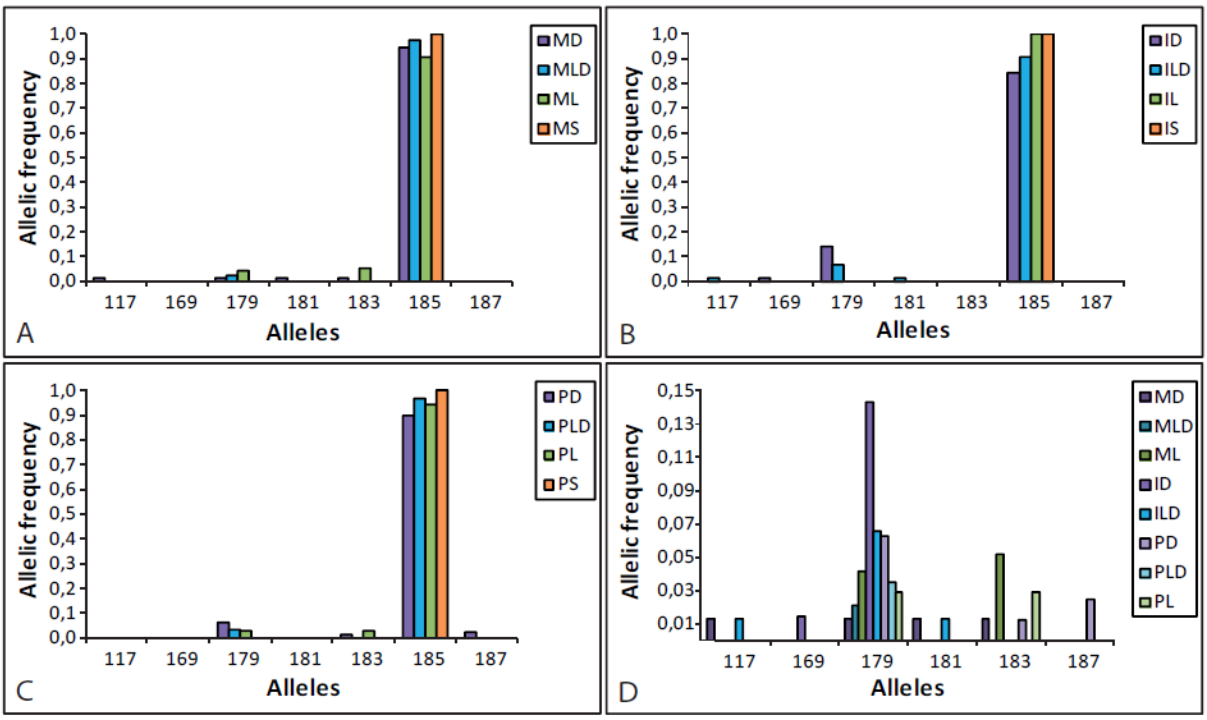


Figure 5 Frequency of the different alleles for the locus Cocl-Lav27 in each population in L. Muddus (A), L. Inari (B), and L. Paadar (C). As the allele 185 is fixed in some populations, the graph D represents only the frequencies of the other alleles in all the polymorphic populations. Note the different y-axis scale in D compared to A-C. The population codes are explained in Table 2.

MLD and PLD were associated with a lower allelic richness (N_{AR} = 6.11 and N_{AR} = 6.18 respectively) and private allelic richness (N_{PAR} = 0.51 and N_{PAR} = 0.54 respectively) than the DR and LSR populations in the same lake (MD: N_{AR} = 6.74 and N_{PAR} = 0.79, ML: N_{AR} = 6.96 and N_{PAR} = 1.03, PD: N_{AR} = 7.48 and N_{PAR} = 1.28, PL: N_{AR} = 6.86 and N_{PAR} = 0.92), whereas ILD exhibited a higher allelic richness (N_{AR} = 7.45) and private allelic richness (N_{PAR} = 0.72) (Figure 6A and 6C). Despite their exact same value (H_e = 0.63), the expected heterozygosity of the LDR populations in L. Inari and L. Paadar were intermediate to those of the DR and LSR populations in the respective lakes, whereas in L. Muddus, the expected heterozygosity of MLD was the highest for the lake (Figure 6B). When compared between each other, MLD and PLD exhibited a lower private allelic richness (N_{PAR} = 0.52 and N_{PAR} = 0.53 respectively) than ILD (N_{PAR} = 1.78) (Figure 6D). These results suggest that the LDR population of L. Inari is more isolated than the two others.

The lowest allelic richness, private allelic richness, and expected heterozygosity were associated with the three SSR populations (MS: N_{AR} = 5.39, N_{PAR} = 0.54, H_e = 0.54; IS: N_{AR} = 5.66, N_{PAR} = 0.51, H_e = 0.58; PS: N_{AR} = 5.02, N_{PAR} = 0.34, H_e = 0.56) (Figure 6). In L.Muddus, the highest allelic and private allelic richness were found in ML (N_{AR} = 6.96, N_{PAR} = 1.03)

whereas in L. Paadar, the highest values were associated with PD (N_{AR} = 7.48, N_{PAR} = 1.28) (Figure 6).

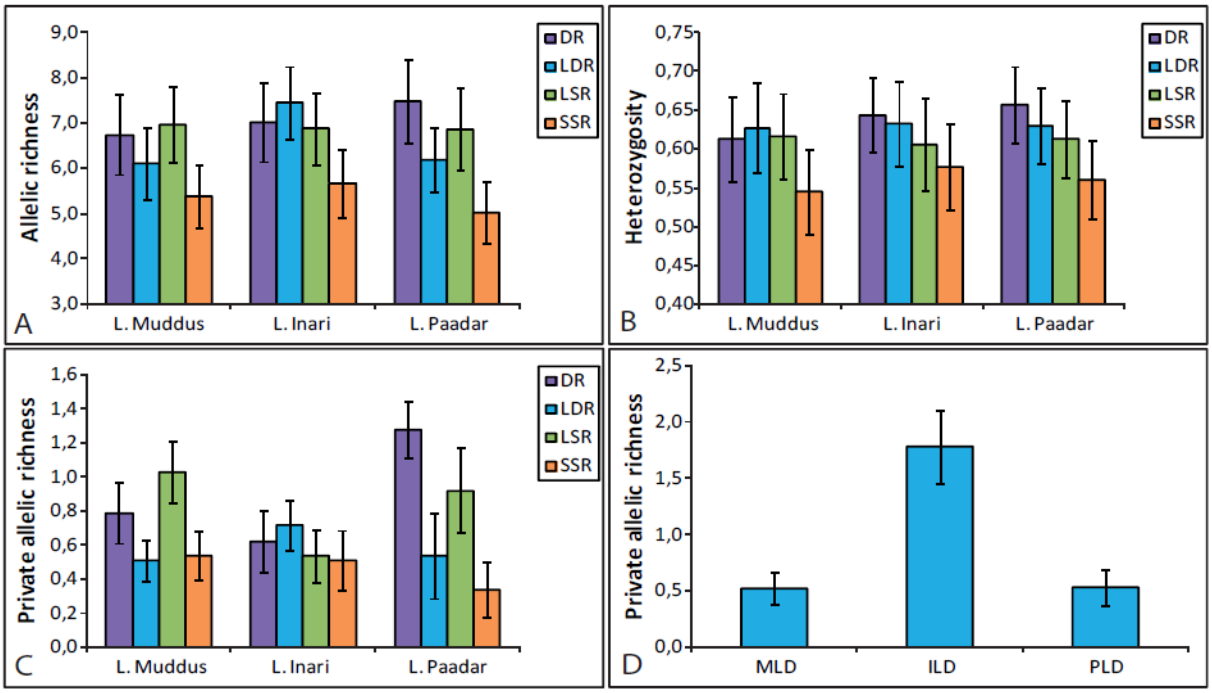


Figure 6 Mean allelic richness (A) and expected heterozygosity (B) of 12 populations from three lakes in northern Fennoscandia (L. Muddus, L. Inari, and L. Paadar). Private allelic richness was calculated by comparing morph within lakes (C) or by comparing the LDR morph between lakes (D). Each value is given with its standard error of mean (SEM).

Population structure

The LDR whitefish morphs were found to be significantly ($p < 0.001$) genetically different from all other morphs (Table 4). In each lake, the LDR population was less genetically differentiated from the LSR (MLD/ML: F_{ST} = 0.031; ILD/IL: F_{ST} = 0.017; PLD/PL: F_{ST} = 0.022) than from the DR population (MLD/MD: F_{ST} = 0.044; ILD/ID: F_{ST} = 0.033; PLD/PD: F_{ST} = 0.460). In L. Muddus, the pair MD/ML had a lower F_{ST} value (F_{ST} = 0.044) than the pair MLD/MD (Table 4). The less differentiated populations were PD/MD (F_{ST} = 0.009; p = 0.002), PLD/MLD (F_{ST} = 0.005; p = 0.020), and PS/IS (F_{ST} = 0.006; p = 0.045). The SSR populations were the most differentiated from the other morphs with F_{ST} values ranging from 0.063 to 0.107 ($p < 0.001$). The population-pair PL/ML could not be genetically differentiated with the only not significant (p = 0.153) F_{ST} value among all pair-wise comparisons (Table 4).

Table 4 Pair-wise F_{ST} values for four sympatric whitefish morphs in three Finnish lakes of the Pasvik watercourse below the diagonal with their corresponding significance p-values after Bonferroni corrections above the diagonal. Most F_{ST} values reflect highly significant differences ($p < 0.001$), whereas the underlined value is non-significant and bold values are not highly significant. Population codes are explained in Table 2.

	MD	**MLD**	**ML**	**MS**	**ID**	**ILD**	**IL**	**IS**	**PD**	**PLD**	**PL**	**PS**
MD	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.002**	0.000	0.000	0.000
MLD	0.044	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.020**	0.000	0.000
ML	0.044	0.031	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.153	0.000
MS	0.089	0.087	0.073	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ID	0.026	0.046	0.052	0.101	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ILD	0.038	0.024	0.024	0.071	0.033	-	0.000	0.000	0.000	0.000	0.000	0.000
IL	0.076	0.045	0.038	0.110	0.062	0.017	-	0.000	0.000	0.000	0.000	0.000
IS	0.071	0.073	0.077	0.023	0.075	0.063	0.107	-	0.000	0.000	0.000	**0.045**
PD	**0.009**	0.041	0.053	0.088	0.018	0.032	0.071	0.064	-	0.000	0.000	0.000
PLD	0.043	**0.005**	0.029	0.081	0.051	0.027	0.047	0.068	0.046	-	0.000	0.000
PL	0.047	0.026	0.002	0.073	0.050	0.025	0.038	0.072	0.055	0.022	-	0.000
PS	0.086	0.078	0.080	0.022	0.096	0.070	0.105	**0.006**	0.085	0.071	0.076	-

The STRUCTURE analysis revealed six clusters (K = 6, mean lnP(K) = -23397.42 ± 28.13, Figure S1). However, the mean probability for K = 5, despite having a more negative value (mean lnP(K) = -23438.25 ± 20.68, Figure S1), had an overlapping standard deviation with the one for K = 6. In both cases, the clustering was identical except for the DR populations, which were identified either as one (K = 5) or two (K = 6) clusters. PLD and MLD were grouped together in the same cluster whereas ILD was grouped with IL. One cluster regrouped all three SSR populations and another consisted of ML and PL (Figure 7).

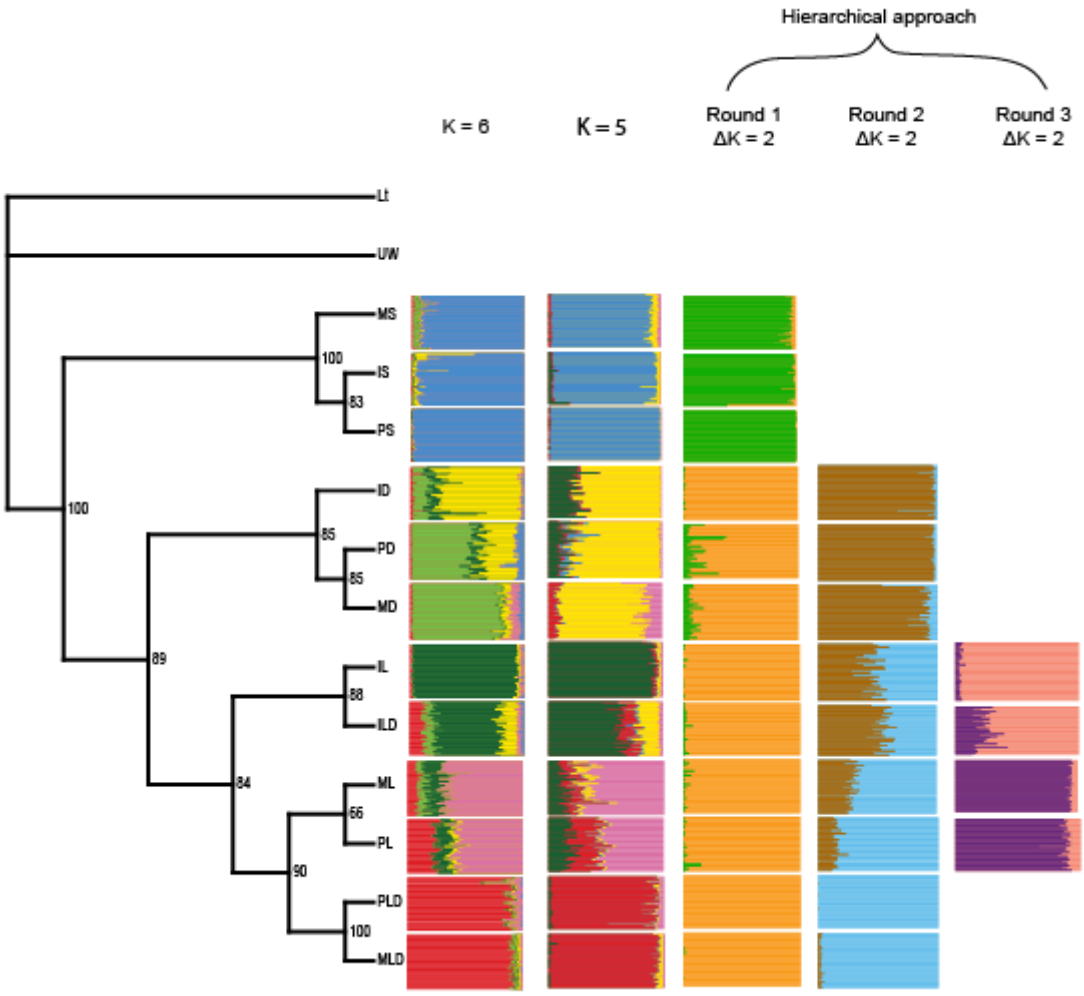


Figure 7 Phylogeny from the Cavalli-Sforza chord measure (with bootstrap values as % obtained with 1,000 permutations) and structure plots of four sympatric whitefish morphs of three lakes from Pasvik watercourse. Lt and UW are two LSR whitefish populations from Scotland used as out-groups. Population codes are explained in Table 2.

The hierarchical approach revealed two clusters after the first round with the SSR populations being separated from the others. The nine populations left were run in another round and two clusters appeared. One consisted of the PLD and MLD whitefish populations, whereas the other contained the DR whitefish morphs. The remaining individuals were run in a third run where PL and ML were grouped together in one cluster, and IL was part of another cluster with ILD (Figure 7). The total number of clusters obtained by the hierarchical approach match the five clusters of the phylogeny and K = 5 of the conventional approach.

Phylogeny

As expected, the phylogenetic trees with the best bootstrap support were given by Reynolds *et al.* genetic distances and Cavalli-Sforza chord measure. These two trees exhibited the same structure and only differed in regard to the bootstrap support. The tree obtained with Cavalli-Sforza chord measure is displayed in Figure 7 and the two others are given in appendix (Figure S2). The populations were grouped in five different clusters, which supported the overall morph and population structure revealed by STRUCTURE. As for the conventional approach (K) and the hierarchical approach (ΔK), PLD and MLD were grouped together and ILD was grouped with IL. The first group to be differentiated was formed by the three SSR populations, as in both of the STRUCTURE plots. Then the three DR populations diverged. ML and PL were always grouped together (Figure 7).

Principal component analysis

The PCA analysis revealed three main groups (Figure 8). The LDR populations could not be differentiated from the LSR populations but they were nonetheless grouped separately. The first principal component was responsible for separating the SSR populations from the others (37.7%, $p < 0.001$) whereas the second component divided the DR populations from LSR and LDR (20.8%, $p < 0.001$).

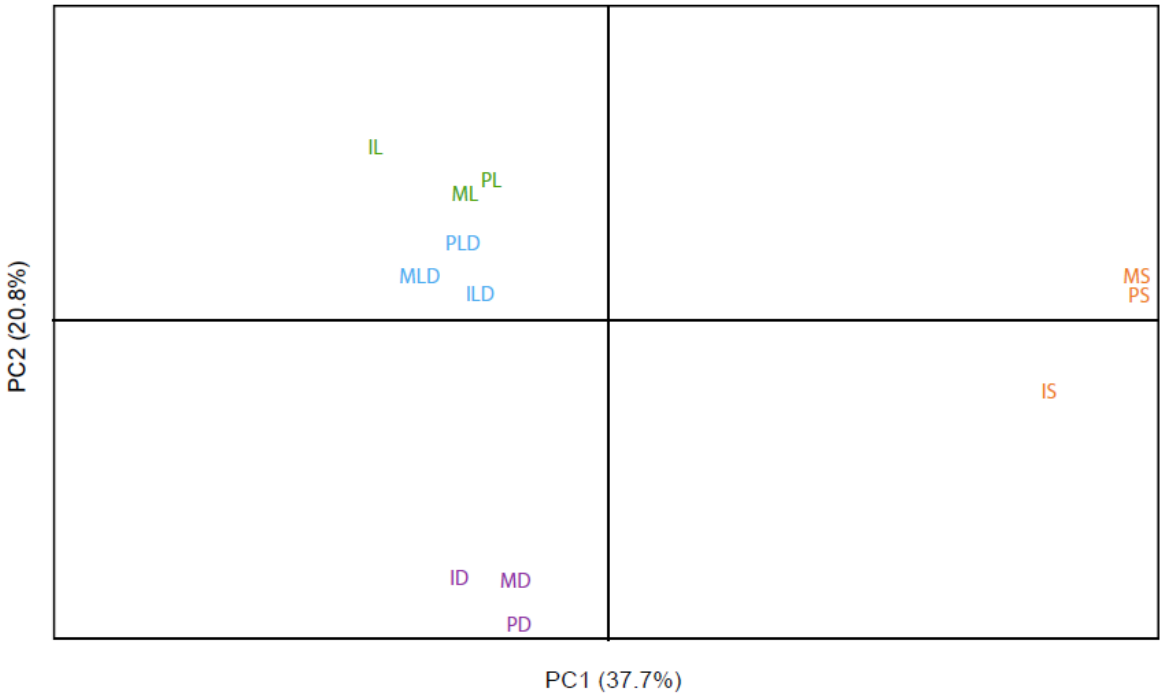


Figure 8 Plot of the first two canonical variables generated from principal component analysis of all populations. Codes are explained in Table 2.

DISCUSSION

The present study reveals that all four whitefish morphs identified were genetically differentiated within each lake. These results support earlier studies where three of them, the small sparsely rakered (SSR), the large sparsely rakered (LSR), and the densely rakered (DR) morphs have recently been described (Kahilainen & Østbye 2006; Harrod *et al.* 2010) and shown to be genetically different in northern Fennoscandia (Præbel *et al.* in press). Here, the emphasis is directed towards the recently identified large densely rakered (LDR) morph. In each of the three studied lakes, the newly found LDR populations could be genetically differentiated from the other sympatric whitefish morphs. All pair-wise F_{ST} values within each lake were highly significant ranging from 0.017 in L. Inari (ILD/IL) to 0.087 in L. Muddus (MLD/MS). These values correspond to those found in previous studies between whitefish morphs in the Pasvik watercourse (Østbye *et al.* 2006; Præbel *et al.* submitted). The LDR populations were found to be more genetically different from the SSR morph and closer to the LSR morph in each lake. The high differentiation between SSR and LDR was expected since they do not seem to share morphological or physiological characteristics and are assumedly occupying different niches. However, a closer genetic relation could have been expected between DR and LDR than between LSR and LDR since they share equivalent gill-raker counts, and gill-rakers are known to be a highly heritable trait (Svärdson 1970; Hermida *et al.* 2002). This aspect is highly interesting regarding the origin of this new morph and will be discussed in more details below. The genetic differentiation of the LDR morph is also supported by the STRUCTURE analysis, the phylogeny and the presence of private alleles. These private alleles support a scenario of limited gene flow between populations which has allowed fixation of different alleles by genetic drift, or has prevented the transfer between populations of new alleles appeared through mutation. Combined, these results strongly suggest that reproductive isolation has evolved between the LDR morph and the other whitefish morphs within each of the three lakes where they were originally identified.

This study also confirmed that the LDR morphs exhibit a gill-raker count similar to the DR whitefish. A similar number of gill-rakers was revealed between MD, MLD and PLD, whereas a slight shift towards less gill-rakers was observed for the LDR population in L. Inari. Moreover, the LDR gill-raker counts were always significantly different from the LSR distributions. The gill-raker numbers obtained for the three previously described morphs corresponded to other studies in northern Fennoscandia (Amundsen *et al.* 2004; Kahilainen & Østbye 2006; Harrod *et al.* 2010; Siwertsson *et al.* 2010). As gill-raker numbers have been

shown to be associated with diet and habitat choice (Kahilainen *et al.* 2011), it was expected that the LDR morphs would exhibit a zooplanktivorous diet and occupy the same niche as the DR populations. However, the stable isotope analysis (SIA) in L. Paadar and L. Inari revealed that the LDR morphs occupy their own isotopic niche, with more carbon enriched values than the PD and ID morphs. In L. Muddus, the SIA signal for the LDR morph was not significantly different from the LSR morph, meaning that the LDR individuals share an isotopic niche with the LSR individuals. Moreover, the large variability in $\delta^{13}C$ enrichment for the LSR and LDR populations (especially in L. Inari) suggests that they use a broader foraging niche and feed on more different prey types than the other two morphs (DR and SSR). Typically in northern Scandinavian lakes, the DR morphs have a narrow diet niche and include exclusively zooplankton in their diet whereas the SSR morphs are profundal soft-sediment benthos specialists (Amundsen *et al.* 2004; Kahilainen & Østbye 2006; Harrod *et al.* 2010; Siwertsson 2012). This low diversity in prey types in their diet explains the narrow SIA signals. The ecological data support the genetic data, revealing that the LDR populations form a new morph which occupies a distinct niche in L. Inari and L. Paadar. Despite their gill-raker distributions, which are closer or similar to the DR whitefish, the LDR morph seems to have an ecological niche closer to the LSR morph, suggesting a possible overlap in time and place of spawning which could lead to gene flow between these two populations. The intermediate $\delta^{13}C$ values for the LDR populations may reflect a mixed diet, feeding on both zooplankton and e.g. insect pupae or benthos.

Several possible scenarios can be provided for the origin of the LDR whitefish morph. First, because extensive stocking of whitefish has been performed in L. Inari (Mutenia & Ahonen 1990; Mutenia & Salonen 1994; Salojärvi & Mutenia 1994), the LDR morph could originate from one of these stocked populations. However, this hypothesis seems unlikely since stocking only occurred in L. Inari, and started 40 years ago. In that case, the LDR morph would have been created in L. Inari and some individuals could have migrated upstream to form new LDR populations in L. Muddus and L. Paadar. However, stocked populations are expected to exhibit a lower genetic diversity since only a fraction of the wild parents were brought into captivity for reproduction (Ryman & Laikre 1991). No decrease in allelic richness or heterozygosity has been observed for ILD, suggesting that this hypothesis is unlikely. Moreover, the stocked whitefish come from an external gene pool (the individuals brought in captivity for breeding do not come from L. Inari), meaning that the genetic differentiation expected between the stocked population and the natural population should be higher than the genetic differences between two populations of L. Inari. Nevertheless, the F_{ST} values revealed that it is not the case, suggesting that the hypothesis of a stocked origin is unlikely.

Another hypothesis is secondary contact after colonization of the lake. In a speciation context, the secondary contact is often referred to as the sympatric phase reinforcing the premating reproductive isolation between two populations that have started to accumulate genetic and phenotypic differences in allopatry (Schluter 2001). However, in this study, the term secondary contact would relate to the occurrence in sympatry of two different lineages after expansion of their geographical range (e.g. Pigeon *et al.* 1997). Nonetheless, the immigration of allopatric populations in the lake is unlikely since all northern Fennoscandian whitefish morphs belonged to the same monophyletic clade according to a recent study by Østbye *et al.* (2005a), indicating a common ancestor. This assumption is supported by an ongoing study using new mitochondrial genes which showed no differences between the four morphs in L. Muddus (K. Præbel pers. comm.).

Thirdly, the LDR morph could have arisen by transgressive hybridization. In general, hybrids have a lower fitness than both their parents in their respective niche because they exhibit intermediate characteristics (Schluter 2000). However, recent studies have shown that hybrids from two divergent species may be more suited to exploit a third niche, either because their intermediate phenotype is more adapted to this particular environment or because the recombination of the genes in the hybrid genome created a new phenotype (Salzburger *et al.* 2002; Smith *et al.* 2003; Mavarez *et al.* 2006). In this study, hybridization between LSR and DR could have resulted in individuals exhibiting a new phenotype that would have colonized a new available niche, thus creating a new population on the early steps of speciation. This scenario is supported by the morphology of the LDR morph that seems to show characteristics of both the LSR (body shape) and the DR (gill-rakers) populations, and by the stable isotope data suggesting that the LDR morph utilizes a different trophic niche than both the LSR and the DR morph. The genetic proximity to the LSR population could be explained by backcross of the first hybrid generation with only LSR and not DR. Moreover, transgressive hybridization seems to be a common case in adaptive radiations and is thought to facilitate the divergence (Seehausen 2004). However, in such a case, the hybrid populations (LDR) should display a higher allelic richness and heterozygosity than both their parents. This is not the case in L. Muddus and L. Paadar, as the LDR morphs exhibited a lower allelic richness than both the LSR and the DR morphs. This pattern could nonetheless be explained by founder effects if a small number of hybrid individuals had the ability to colonize a new niche and form the new population. However, these conditions are fulfilled for the LDR morph in L. Inari since this morph exhibits higher allelic richness and heterozygosity than both the LSR and the DR morphs within the same lake. On the other hand, such hybridization could have occurred between the LSR morph and a stocked population or between the LSR morph and some vendace in L. Inari. However, the vendace and the stocked populations are new in this lake (Mutenia & Salonen 1992;

Mutenia & Salonen 1994), and not present in L. Muddus and L. Paadar, which would suggest a different origin for MLD and PLD. In any case, the possibility of a transgressive hybridization can neither be confirmed nor rejected by the results of this study and further analyses (particularly phylogenetic studies) are needed to assess the likelihood of this theory.

Finally, the last scenario is a sympatric divergence of the LDR morph from the LSR or from the DR morphs. Adaptive radiation has been proposed in previous studies to explain the occurrence of sympatric whitefish morphs in northern Fennoscandia (Østbye *et al.* 2006; Præbel *et al.* submitted), suggesting that the LSR whitefish might have been the ancestral form. As gill-rakers have been proved to be a highly heritable trait (Svärdson 1970; Hermida *et al.* 2002; Rogers & Bernatchez 2007), a divergence from the DR whitefish morph was expected. In this case, reproductive isolation would build up due to a strong disruptive selection between the LDR and the DR morphs while the LDR whitefish started to utilize a new niche closer to the littoral zone already occupied by the LSR morph. The similarity of this new available niche with the LSR niche revealed by the stable isotopes would explain the similarity in body shape between LDR and LSR whitefish, and could have led to similarities in reproduction (time and place of spawning), suggesting a possible gene flow between these two populations. This would explain the genetic proximity between these two morphs revealed by all the statistical approaches. However, in the divergence of the other European whitefish morphs, gill-rakers seem to be the key trait on which divergent natural selection acts (Østbye *et al.* 2005b; Præbel *et al.* submitted). This would imply that a difference in gill-raker numbers is needed for a disruptive selection to occur. This hypothesis as well as the genetic proximity between the LSR and the LDR morphs would suggest that the LDR morph diverged from the LSR morph. The results of this study support a scenario of sympatric divergence as the most plausible explanation for the existence of the LDR morph. However, more detailed studies are needed to evaluate the ancestral form from which the LDR morph derived.

In a case of sympatric divergence, two scenarios remain possible. Either the divergence occurred within the lakes directly with one population dividing into two, or the same morph colonized the same lake repeatedly (i.e. double invasion). The second population to arrive had to find another niche to exploit since its preferred niche was already occupied, leading to a divergence in phenotype followed by reproductive isolation (Bernatchez *et al.* 1996). Both cases are difficult to differentiate regarding their genetic signature, the only notable difference being the time at which they occur (a double invasion would imply an older divergence). The results of this study do not enable us to support one or the other and more research is required to evaluate their respective likelihood.

Regardless of the mechanism responsible for the foundation of the LDR whitefish morph, the question whether this phenomenon appeared repeatedly in the three lakes or if some LDR individuals migrated from lake to lake remains to be answered. Parallel evolution has been suggested in previous studies as a mechanism responsible for the repeated appearance of the same whitefish morph in different lakes in northern Fennoscandia (Østbye *et al.* 2006; Præbel *et al.* submitted). In such a case, the same divergence would have occurred repeatedly in several lakes, and the different morphs within the same lake should be genetically more similar than when comparing the same morphs across lakes. The other explanation for the occurrence of similar morphs could be a single divergence event with dispersal of individuals from the newly formed population into several lakes. In this case, the similar morphs would be genetically closer across lakes than when comparing the different morphs within the same lake. In this study, the three LDR morphs were found to be genetically closer to each other across lakes than to the other morphs within the same lake. The same pattern was also observed for the other three morphs. These results suggest that the three LDR populations in the three studied lakes have a common origin. This conclusion is supported by the history of the Inarijärvi basin. After the last ice retreat (around 10,000 years ago), the three lakes concerned by this study were part of the same ice-dammed lake. As the ice melted and sedimentation happened, the area of L. Inari decreased and several smaller lakes were created around it (including L. Muddus and L. Paadar) (Kujansuu *et al.* 1998). In the light of the results of this study, it is thus reasonable to suggest that the LDR morph observed in these three lakes were formed during the deglaciation period in the Inarijärvi basin, and segregated into the newly formed lakes after the ice-dam broke. However, the LDR population of L. Inari is closer to the LSR morph of the same lake than the LDR populations of L. Muddus and L. Paadar. This pattern is most likely related to the invasion and stocking histories of this lake and will be discussed later.

When examining the STRUCTURE analysis and the phylogeny, a special pattern was observed for the LSR and the LDR populations of L. Inari. In fact, these populations are closer to each other than the other LSR/LDR pairs in the two other lakes. Moreover, the STRUCTURE analysis grouped IL and ILD together in the same cluster, as opposed to the other LSR and LDR populations in L. Muddus and L. Paadar. This particular pattern may reflect the extensive stocking that happened in L. Inari (Mutenia & Salonen 1994; Salojärvi & Mutenia 1994). The LSR morph in L. Inari did not exhibit any departure from HWE, suggesting no substructuring of the population, thus suggesting that no stocked individuals were sampled. However, the genetic structure of the IL population could have been modified if hybridization occurred between the natural and the stocked populations. Furthermore, the genetic proximity between ILD and IL suggests a possible ongoing gene flow between these

two morphs, which would explain the difference between ILD and the other two LDR populations. Moreover, ILD exhibited a higher allelic richness than MLD and PLD, and even higher than the other morphs in L. Inari. The same pattern was observed for the intralacustrine private allelic richness. The private allelic richness only between LDR morphs revealed that ILD displayed a higher number of private alleles compared to MLD and PLD, indicating a higher genetic differentiation. This pattern may be the result of the possible hybridization between IL and stocked individuals.

When looking more into details at the locus Cocl-Lav27, an interesting pattern can be discerned for the LDR morph in L. Inari. Two alleles in particular (117 and 181) are found only for ILD and MD, indicating a possible relation between these two populations. An explanation could be that these alleles were present in other populations but since their frequency was low, they were lost by genetic drift. On the other hand, this could be an indication that LDR was formed by sympatric divergence from the DR population rather than from the LSR whitefish. After building up their reproductive isolation with the DR morph, the LDR whitefish could have starting reproducing with the LSR morph because their ecological niches are similar. Another interesting point is that the allele 117 is also frequently found in the vendace populations (Præbel *et al.* in press). This could indicate some kind of gene flow between ILD and a vendace population as suggested in a previous paragraph. However, the presence of this allele in the MD morph could not be explained by the same mechanism since vendace are absent of L. Muddus. However, the frequencies of the alleles 117 and 181 found in Cocl-Lav27 are low and the fact that they are found only in ILD and MD might be due to sampling effect.

CONCLUSIONS

This study revealed the occurrence of a new whitefish morph in three Finnish lakes of the Pasvik watercourse, which is genetically differentiated from all the other morphs. This newly described large densely rakered (LDR) whitefish exhibit gill-raker numbers similar to those of the densely rakered (DR) morph, but seem to feed in a different trophic niche, closer to the niche utilized by the large sparsely rakered (LSR) whitefish. The intermediate carbon values (SIA) of the LDR individuals indicate that they might feed on both zooplankton and benthic prey. From the several possible origins of this morph, the most likely seem to be a sympatric divergence from either the LSR morph or the DR morph; although the possibility of a transgressive hybridization between the LSR and DR populations could not be ruled out. The results of this study suggested a common origin for the three LDR populations. The most likely explanation is a sympatric divergence occurring during the last deglaciation, when the lakes Inarijärvi, Muddusjärvi and Paadarjärvi were part of the same big ice-dammed lake. Then, some sub-populations of each morph could have been trapped into the different lakes after the ice-dam broke.

The appearance of a new whitefish morph in northern Fennoscandia provides a new evidence for an ongoing adaptive radiation in this area. However, many questions remain regarding the precise origin of this morph. Further studies are needed to describe the large densely rakered whitefish in more details (i.e. morphology, diet, habitat, and genetics) and the appearance of this morph in other nearby lakes must also be investigated.

REFERENCES

Amundsen, P.-A., Bøhn, T. & Vaga, G.H. (2004) Gill raker morphology and feeding ecology of two sympatric morphs of European whitefish (*Coregonus lavaretus*). *Annales Zoologici Fennici,* **41,** 291-300.

Bernatchez, L. & Dodson, J.J. (1994) Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences,* **51,** 240-251.

Bernatchez, L., Vuorinen, J.A., Bodaly, R.A. & Dodson, J.J. (1996) Genetic evidence for reproductive isolation and multiple origins of sympatric trophic ecotypes of whitefish (*Coregonus*). *Evolution,* **50,** 624-635.

Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *Evolution,* **21,** 550-570.

Edelstein, L. (1944) Aristotle and the Concept of Evolution. *The Classical Weekly,* **37,** 148-150.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology,* **14,** 2611-2620.

Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* **10,** 564-567.

Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics,* **5,** 163-166.

Felsenstein, J. (2009) PHYLIP - Phylogeny Inference Package (version 3.69). Washington: Washington university.

Gillespie, R. (2009) Adaptive radiation. In: *The Princeton guide to ecology* (eds S.A. Levin & S.R. Carpenter), pp. 143-152. Princeton University Press, Princeton.

Glossary (2001) Speciation Glossary. *Trends in Ecology & Evolution,* **16,** 412-413.

Goudet, J. (2004) PCAGEN: computer program. Lausanne: Lausanne university.

Grant, P.R. & Grant, B.R. (1994) Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution,* **48,** 297-316.

Guo, S.W. & Thompson, E.A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics,* **48,** 361-372.

Harrod, C., Mallela, J. & Kahilainen, K.K. (2010) Phenotype-environment correlations in a putative whitefish adaptive radiation. *Journal of Animal Ecology,* **79,** 1057-1068.

Hartl, D.L. & Clark, A.G. (2007) *Principles of population genetics,* 4th edn. Sinauer Associates, Sunderland, Mass.

Hendry, A.P., Bolnick, D.I., Berner, D. & Peichel, C.L. (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology,* **75,** 2000-2036.

Hermida, M., Fernandez, C., Amaro, R. & San Miguel, E. (2002) Heritability and "evolvability" of meristic characters in a natural population of *Gasterosteus aculeatus*. *Canadian Journal of Zoology,* **80,** 532-541.

Himberg, K.J.M. (1970) A systematic and zoogeographic study of some North European Coregonids. In: *Biology of Coregonid fishes* (eds C.C. Lindsey & C.S. Woods), pp. 219-250. University of Manitoba Press, Winnipeg.

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,** 1322-1332.

Hudson, A.G., Vonlanthen, P., Mueller, R. & Seehausen, O. (2007) Review: The geography of speciation and adaptive radiation in coregonines. In: *Biology and Management of Coregonid Fishes* (eds M. Jankun, P. Brzuzan, P. Hliwa & M. Luczynski), pp. 111-146. E Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

Jensen, H., Kahilainen, K.K., Amundsen, P.-A., Gjelland, K.O., Tuomaala, A., Malinen, T. & Bøhn, T. (2008) Predation by brown trout (*Salmo trutta*) along a diversifying prey community gradient. *Canadian Journal of Fisheries and Aquatic Sciences,* **65,** 1831-1841.

Kahilainen, K.K. & Lehtonen, H. (2003) Piscivory and prey selection of four predator species in a whitefish dominated subarctic lake. *Journal of Fish Biology,* **63,** 659-672.

Kahilainen, K.K. & Østbye, K. (2006) Morphological differentiation and resource polymorphism in three sympatric whitefish *Coregonus lavaretus* (L.) forms in a subarctic lake. *Journal of Fish Biology,* **68,** 63-79.

Kahilainen, K.K., Malinen, T., Tuomaala, A., Alajaervi, E., Tolonen, A. & Lehtonen, H. (2007) Empirical evaluation of phenotype-environment correlation and trait utility with allopatric and sympatric whitefish, *Coregonus lavaretus* (L.), populations in subarctic lakes. *Biological Journal of the Linnean Society,* **92,** 561-572.

Kahilainen, K.K., Siwertsson, A., Gjelland, K.O., Knudsen, R., Bøhn, T. & Amundsen, P.-A. (2011) The role of gill raker number variability in adaptive radiation of coregonid fish. *Evolutionary Ecology,* **25,** 573-588.

Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes,* **5,** 187-189.

Klemetsen, A. (2010) The charr problem revisited: exceptional phenotypic plasticity promotes ecological speciation in postglacial lakes. *Freshwater Reviews,* **3,** 49-74.

Kujansuu, R., Eriksson, B. & Grönlund, T. (1998) Lake Inarijärvi, northern Finland: sedimentation and late Quaternary evolution. *Geological Survey of Finland, Report of Investigation 143*. Espoo.

Mavarez, J., Salazar, C.A., Bermingham, E., Salcedo, C., Jiggins, C.D. & Linares, M. (2006) Speciation by hybridization in *Heliconius* butterflies. *Nature,* **441,** 868-871.

Mayr, E. (1942) *Systematics and the origin of species*. Columbia University Press, New York.

McKinnon, J.S., Mori, S., Blackman, B.K., David, L., Kingsley, D.M., Jamieson, L., Chou, J. & Schluter, D. (2004) Evidence for ecology's role in speciation. *Nature,* **429,** 294-298.

Mutenia, A. & Ahonen, M. (1990) *Recent changes in the fishery on Lake Inari, Finland*. Pudoc, Wageningen.

Mutenia, A. & Salonen, E. (1992) The vendace (*Coregonus albula* L.), a new species in the fish community and fisheries of Lake Inari. *Polish Archives of Hydrobiology,* **39,** 797-805.

Mutenia, A. & Salonen, E. (1994) Rehabilitation of the fisheries of Lake Inari, northern Finland. In: *Rehabilitation of Freshwater Fisheries* (ed. I.G. Cowx), pp. 280-288. Fishing News Books Ltd, Farnham.

Nei, M. (1972) Genetic distance between populations. *American Naturalist,* **106,** 283-292.

Nosil, P. & Rundle, H.D. (2009) Ecological speciation: Natural selection and the formation of new species. In: *The Princeton guide to ecology* (eds S.A. Levin & S.R. Carpenter), pp. 134-142. Princeton University Press, Princeton.

Østbye, K., Bernatchez, L., Naesje, T.F., Himberg, K.J.M. & Hindar, K. (2005a) Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology,* **14,** 4371-4387.

Østbye, K., Naesje, T.F., Bernatchez, L., Sandlund, O.T. & Hindar, K. (2005b) Morphological divergence and origin of sympatric populations of European whitefish (*Coregonus lavaretus* L.) in Lake Femund, Norway. *Journal of Evolutionary Biology,* **18,** 683-702.

Østbye, K., Amundsen, P.-A., Bernatchez, L., Klemetsen, A., Knudsen, R., Kristoffersen, R., Naesje, T.F. & Hindar, K. (2006) Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Molecular Ecology,* **15,** 3983-4001.

Patton, J.C., Gallaway, B.J., Fechhelm, R.G. & Cronin, M.A. (1997) Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (*Coregonus nasus*) in the Colville and Sagavanirktok rivers in northern Alaska. *Canadian Journal of Fisheries and Aquatic Sciences,* **54,** 1548-1556.

Peakall, R. & Smouse, P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes,* **6,** 288-295.

Perga, M.E. & Gerdeaux, D. (2005) 'Are fish what they eat' all year round? *Oecologia,* **144,** 598-606.

Pigeon, D., Chouinard, A. & Bernatchez, L. (1997) Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution,* **51,** 196-205.

Post, D.M. (2002) Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology,* **83,** 703-718.

Præbel, K., Østbye, K., Siwertsson, A., Knudsen, R., Kahilainen, K.K., Fevolden, S.E. & Amundsen, P.-A. (2011) Parallel post-glacial re-colonisation of European whitefish morphs in northern Fennoscandia. *11th International Symposium on the Biology and Management of Coregonid Fishes*. Mondsee, Austria.

Præbel, K., Westgaard, J.I., Amundsen, P.-A., Siwertsson, A., Knudsen, R., Kahilainen, K.K. & Fevolden, S.E. (in press) A diagnostic tool for efficient analysis of population structure, hybridization and conservation status of European whitefish (Coregonus lavaretus (L.)) and vendace (C. albula (L.)). *Fundamental and Applied Limnology*.

Præbel, K., Knudsen, R., Siwertsson, A., Kahilainen, K.K., Østbye, K., Peruzzi, S., Fevolden, S.E. & Amundsen, P.-A. (submitted) Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic and profundal lake habitats. *Molecular Ecology*.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155,** 945-959.

Quinn, G.P. & Keough, M.J. (2002) *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge, UK ; New York.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Reshetnikov, Y.S. (2004) Coregonid fishes in Arctic waters. *Annales Zoologici Fennici,* **41,** 3-11.

Reynolds, J., Weir, B.S. & Cockerham, C.C. (1983) Estimation of the co-ancestry coefficient: basis for a short-term genetic distance. *Genetics,* **105,** 767-779.

Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution,* **43,** 223-225.

Rogers, S.M., Marchand, M.H. & Bernatchez, L. (2004) Isolation, characterization and cross-salmonid amplification of 31 microsatellite loci in the lake whitefish (*Coregonus clupeaformis*, Mitchill). *Molecular Ecology Notes,* **4,** 89-92.

Rogers, S.M. & Bernatchez, L. (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus sp* Salmonidae) species pairs. *Molecular Biology and Evolution,* **24,** 1423-1438.

Rousset, F. (2007) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources,* **8,** 103-106.

Rundle, H.D., Nagel, L., Boughman, J.W. & Schluter, D. (2000) Natural selection and parallel speciation in sympatric sticklebacks. *Science,* **287,** 306-308.

Ryman, N. & Laikre, L. (1991) Effects of supportive breeding on the genetically effective population size. *Conservation Biology,* **5,** 325-329.

Salojärvi, K. & Mutenia, A. (1994) Effects of stocking fingerlings on recruitment in the Lake Inari whitefish (*Coregonus lavaretus* L. *s.l.*) fishery. In: *Rehabilitation of Freshwater Fisheries* (ed. I.G. Cowx), pp. 302-313. Fishing News Books Ltd, Farnham.

Salzburger, W., Baric, S. & Sturmbauer, C. (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology,* **11,** 619-625.

Sanderson, S.L., Cheer, A.Y., Goodrich, J.S., Graziano, J.D. & Callan, W.T. (2001) Crossflow filtration in suspension-feeding fishes. *Nature,* **412,** 439-441.

Schluter, D. (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

Schluter, D. (2001) Ecology and the origin of species. *Trends in Ecology & Evolution,* **16,** 372-380.

Schluter, D. (2009) Evidence for ecological speciation and its alternative. *Science,* **323,** 737-741.

Seehausen, O. (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution,* **19,** 198-207.

Siwertsson, A., Knudsen, R., Kahilainen, K.K., Præbel, K., Primicerio, R. & Amundsen, P.-A. (2010) Sympatric diversification as influenced by ecological opportunity and historical contingency in a young species lineage of whitefish. *Evolutionary Ecology Research,* **12,** 929-947.

Siwertsson, A. (2012) Diversity along a speciation continuum - ecology and morphology of northern European whitefish (*Coregonus lavaretus*). Philosophiae Doctor, University of Tromsø.

Skulason, S. & Smith, T.B. (1995) Resource polymorphisms in vertebrates. *Trends in Ecology & Evolution,* **10,** 366-370.

Slatkin, M. (1991) Inbreeding coefficients and coalescence times. *Genetical Research,* **58,** 167-175.

Smith, P.F., Konings, A. & Kornfield, I. (2003) Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity. *Molecular Ecology,* **12,** 2497-2504.

Sobel, J.M., Chen, G.F., Watt, L.R. & Schemske, D.W. (2010) The biology of speciation. *Evolution,* **64,** 295-315.

Susnik, S., Snoj, A. & Dovc, P. (1999) A new set of microsatellite markers for grayling: BFRO014, BFRO015, BFRO016, BFRO017 and BPRO018. *Animal Genetics,* **30,** 478-478.

Svärdson, G. (1957) The coregonid problem. VI. The Palearctic species and their intergrades. *Report of the Institute of Freshwater Research, Drottningholm,* **38,** 267-356.

Svärdson, G. (1970) Significance of introgression in coregonid evolution. In: *Biology of Coregonid Fishes* (eds C.C. Lindsey & C.S. Woods), pp. 33-59. University of Manitoba Press, Winnipeg, Canada.

Svärdson, G. (1979) Speciation of Scandinavian *Coregonus*. *Report of the Institute of Freshwater Research, Drottningholm,* **57,** 1-95.

Turgeon, J., Estoup, A. & Bernatchez, L. (1999) Species flock in the North American Great Lakes: Molecular ecology of Lake Nipigon Ciscoes (Teleostei : Coregonidae : *Coregonus*). *Evolution,* **53,** 1857-1871.

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,** 535-538.

Vander Zanden, M.J. & Rasmussen, J.B. (1999) Primary consumer $\delta^{13}C$ and $\delta^{15}N$ and the trophic position of aquatic consumers. *Ecology,* **80,** 1395-1404.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* **38,** 1358-1370.

Winkler, K.A. & Weiss, S. (2008) Eighteen new tetranucleotide microsatellite DNA markers for *Coregonus lavaretus* cloned from an alpine lake population. *Molecular Ecology Resources,* **8,** 1055-1058.

APPENDIX

Table S1 Pair-wise mean differences of gill-raker distribution between 12 whitefish populations in three lakes of the Pasvik watercourse calculated using a Tuskey's HSD test with the 95% family-wise confidence levels (FWCL) and the p-values adjusted for multiple comparisons. The codes for each population are listed in Table 2.

Population	*Mean difference*	*Lower 95% FWCL*	*Upper 95% FWCL*	*p adj*
IDL-ID	-3.8142857	-5.39696837	-2.23160306	0.0000000
IL-ID	-12.9714286	-14.56521790	-11.37763924	0.0000000
IS-ID	-17.1714286	-18.86189744	-15.48095970	0.0000000
MD-ID	-0.7879699	-2.34998618	0.77404633	0.8859766
MDL-ID	-1.4632219	-2.95180783	0.02536406	0.0590752
ML-ID	-12.0017857	-13.48373842	-10.51983300	0.0000000
MS-ID	-18.5047619	-20.03069902	-16.97882479	0.0000000
PD-ID	0.8607143	-0.68246560	2.40389417	0.7989200
PDL-ID	-2.0119601	-3.52981061	-0.49410966	0.0009992
PL-ID	-10.5714286	-12.16521790	-8.97763924	0.0000000
PS-ID	-16.5142857	-18.90496971	-14.12360172	0.0000000
IL-IDL	-9.1571429	-10.73982551	-7.57446021	0.0000000
IS-IDL	-13.3571429	-15.03714432	-11.67714139	0.0000000
MD-IDL	3.0263158	1.47563377	4.57699781	0.0000000
MDL-IDL	2.3510638	0.87437562	3.82775204	0.0000170
ML-IDL	-8.1875000	-9.65750128	-6.71749872	0.0000000
MS-IDL	-14.6904762	-16.20480905	-13.17614333	0.0000000
PD-IDL	4.6750000	3.14329373	6.20670627	0.0000000
PDL-IDL	1.8023256	0.29614166	3.30850950	0.0054913
PL-IDL	-6.7571429	-8.33982551	-5.17446021	0.0000000
PS-IDL	-12.7000000	-15.08329392	-10.31670608	0.0000000
IS-IL	-4.2000000	-5.89046887	-2.50953113	0.0000000
MD-IL	12.1834586	10.62144239	13.74547490	0.0000000
MDL-IL	11.5082067	10.01962074	12.99679263	0.0000000
ML-IL	0.9696429	-0.51230985	2.45159557	0.5867296
MS-IL	-5.5333333	-7.05927044	-4.00739622	0.0000000
PD-IL	13.8321429	12.28896297	15.37532274	0.0000000
PDL-IL	10.9594684	9.44161796	12.47731892	0.0000000
PL-IL	2.4000000	0.80621067	3.99378933	0.0000685
PS-IL	-3.5428571	-5.93354114	-1.15217314	0.0000995
MD-IS	16.3834586	14.72291195	18.04400535	0.0000000

Population	*Mean difference*	*Lower 95% FWCL*	*Upper 95% FWCL*	*p adj*
MDL-IS	15.7082067	14.11653817	17.29987521	0.0000000
ML-IS	5.1696429	3.58417624	6.75510947	0.0000000
MS-IS	-1.3333333	-2.95998776	0.29332109	0.2329244
PD-IS	18.0321429	16.38930242	19.67498330	0.0000000
PDL-IS	15.1594684	13.54039752	16.77853935	0.0000000
PL-IS	6.6000000	4.90953113	8.29046887	0.0000000
PS-IS	0.6571429	-1.79905146	3.11333718	0.9992801
MDL-MD	-0.6752520	-2.12976854	0.77926462	0.9328541
ML-MD	-11.2138158	-12.66154304	-9.76608854	0.0000000
MS-MD	-17.7167920	-19.20951245	-16.22407150	0.0000000
PD-MD	1.6486842	0.13834170	3.15902672	0.0190258
PDL-MD	-1.2239902	-2.70844311	0.26046269	0.2250123
PL-MD	-9.7834586	-11.34547490	-8.22144239	0.0000000
PS-MD	-15.7263158	-18.09593608	-13.35669550	0.0000000
ML-MDL	-10.5385638	-11.90674054	-9.17038711	0.0000000
MS-MDL	-17.0415400	-18.45724047	-15.62583957	0.0000000
PD-MDL	2.3239362	0.88966707	3.75820527	0.0000106
PDL-MDL	-0.5487382	-1.95571862	0.85824212	0.9810513
PL-MDL	-9.1082067	-10.59679263	-7.61962074	0.0000000
PS-MDL	-15.0510638	-17.37293671	-12.72919095	0.0000000
MS-ML	-6.5029762	-7.91170025	-5.09425213	0.0000000
PD-ML	12.8625000	11.43511653	14.28988347	0.0000000
PDL-ML	9.9898256	8.58986506	11.38978611	0.0000000
PL-ML	1.4303571	-0.05159557	2.91230985	0.0699864
PS-ML	-4.5125000	-6.83012581	-2.19487419	0.0000000
PD-MS	19.3654762	17.89247796	20.83847442	0.0000000
PDL-MS	16.4928018	15.04636142	17.93924213	0.0000000
PL-MS	7.9333333	6.40739622	9.45927044	0.0000000
PS-MS	1.9904762	-0.35551818	4.33647057	0.1887184
PDL-PD	-2.8726744	-4.33729375	-1.40805509	0.0000000
PL-PD	-11.4321429	-12.97532274	-9.88896297	0.0000000
PS-PD	-17.3750000	-19.73224621	-15.01775379	0.0000000
PL-PDL	-8.5594684	-10.07731892	-7.04161796	0.0000000
PS-PDL	-14.5023256	-16.84306812	-12.16158305	0.0000000
PS-PL	-5.9428571	-8.33354114	-3.55217314	0.0000000

Table S2 Pair-wise mean differences of δ13C and δ15N stable isotope values between 12 whitefish populations in three lakes of the Pasvik watercourse calculated using a Tuskey's HSD test with the p-values adjusted for multiple comparisons. The codes for each population are listed in Table 2.

Population	*Mean diff δ13C*	*p adj δ13C*	*Mean diff δ15N*	*p adj δ15N*
IDL-ID	2.31596526	0.0000000	-0.45920047	0.0597886
IL-ID	3.61941985	0.0000000	-0.28781859	0.6884523
IS-ID	0.67287180	0.4318136	1.91490554	0.0000000
MD-ID	0.37174792	0.9392466	-0.48653244	0.0136918
MDL-ID	3.30218081	0.0000000	-1.49168709	0.0000000
ML-ID	2.88257374	0.0000000	-1.55064209	0.0000000
MS-ID	0.46580100	0.7688111	1.21637252	0.0000000
PD-ID	0.76465762	0.0871299	-0.16974651	0.9808356
PDL-ID	2.46547837	0.0000000	-1.15288285	0.0000000
PL-ID	4.64106292	0.0000000	-1.75819126	0.0000000
PS-ID	0.71894267	0.8244242	1.12429640	0.0000168
IL-IDL	1.30345459	0.0000861	0.17138188	0.9885532
IS-IDL	-1.64309346	0.0000007	2.37410601	0.0000000
MD-IDL	-1.94421734	0.0000000	-0.02733197	1.0000000
MDL-IDL	0.98621555	0.0068620	-1.03248662	0.0000000
ML-IDL	0.56660848	0.4705019	-1.09144161	0.0000000
MS-IDL	-1.85016426	0.0000000	1.67557299	0.0000000
PD-IDL	-1.55130764	0.0000000	0.28945396	0.5449844
PDL-IDL	0.14951311	0.9999853	-0.69368238	0.0000252
PL-IDL	2.32509766	0.0000000	-1.29899078	0.0000000
PS-IDL	-1.59702259	0.0044829	1.58349687	0.0000000
IS-IL	-2.94654805	0.0000000	2.20272413	0.0000000
MD-IL	-3.24767193	0.0000000	-0.19871385	0.9393701
MDL-IL	-0.31723904	0.9858550	-1.20386850	0.0000000
ML-IL	-0.73684611	0.1177354	-1.26282350	0.0000000
MS-IL	-3.15361885	0.0000000	1.50419111	0.0000000
PD-IL	-2.85476223	0.0000000	0.11807208	0.9991697
PDL-IL	-1.15394148	0.0004384	-0.86506426	0.0000000
PL-IL	1.02164307	0.0024832	-1.47037267	0.0000000
PS-IL	-2.90047718	0.0000000	1.41211499	0.0000000
MD-IS	-0.30112388	0.9928455	-2.40143798	0.0000000
MDL-IS	2.62930901	0.0000000	-3.40659263	0.0000000
ML-IS	2.20970194	0.0000000	-3.46554763	0.0000000
MS-IS	-0.20707080	0.9997663	-0.69853302	0.0000715
PD-IS	0.09178582	0.9999999	-2.08465205	0.0000000

Population	Mean diff δ13C	p adj δ13C	Mean diff δ15N	p adj δ15N
PDL-IS	1.79260657	0.0000000	-3.06778839	0.0000000
PL-IS	3.96819112	0.0000000	-3.67309680	0.0000000
PS-IS	0.04607087	1.0000000	-0.79060914	0.0195198
MDL-MD	2.93043289	0.0000000	-1.00515465	0.0000000
ML-MD	2.51082582	0.0000000	-1.06410964	0.0000000
MS-MD	0.09405308	0.9999996	1.70290496	0.0000000
PD-MD	0.39290970	0.8431654	0.31678593	0.2603233
PDL-MD	2.09373045	0.0000000	-0.66635041	0.0000079
PL-MD	4.26931500	0.0000000	-1.27165882	0.0000000
PS-MD	0.34719475	0.9991673	1.61082884	0.0000000
ML-MDL	-0.41960707	0.8245640	-0.05895499	0.9999987
MS-MDL	-2.83637981	0.0000000	2.70805961	0.0000000
PD-MDL	-2.53752319	0.0000000	1.32194058	0.0000000
PDL-MDL	-0.83670244	0.0291182	0.33880424	0.2725086
PL-MDL	1.33888211	0.0000014	-0.26650416	0.6068013
PS-MDL	-2.58323814	0.0000000	2.61598350	0.0000000
MS-ML	-2.41677274	0.0000000	2.76701460	0.0000000
PD-ML	-2.11791612	0.0000000	1.38089558	0.0000000
PDL-ML	-0.41709537	0.8182687	0.39775923	0.0628350
PL-ML	1.75848918	0.0000000	-0.20754917	0.8536140
PS-ML	-2.16363107	0.0000028	2.67493849	0.0000000
PD-MS	0.29885662	0.9745939	-1.38611902	0.0000000
PDL-MS	1.99967737	0.0000000	-2.36925537	0.0000000
PL-MS	4.17526192	0.0000000	-2.97456377	0.0000000
PS-MS	0.25314167	0.9999619	-0.09207611	0.9999993
PDL-PD	1.70082075	0.0000000	-0.98313634	0.0000000
PL-PD	3.87640530	0.0000000	-1.58844475	0.0000000
PS-PD	-0.04571495	1.0000000	1.29404291	0.0000001
PL-PDL	2.17558455	0.0000000	-0.60530841	0.0000915
PS-PDL	-1.74653570	0.0006727	2.27717925	0.0000000
PS-PL	-3.92212025	0.0000000	2.88248766	0.0000000

Table S3 Summary statistics of the 18 microsatellite loci used in the analyses showing the number of diploid genotypes per population (N), the number of alleles per population (N_A), and the expected heterozygosity (H_e) per population. Population codes are listed in Table 2.

Pop		BWF1	BWF2	ClaTet3	ClaTet13	ClaTet18	Cocl-Lav4	Cocl-Lav6	Cocl-Lav10	Cocl-Lav27	BFRO018	Cocl-Lav18	Cocl-Lav49	Cocl-Lav52	ClaTet6	ClaTet9	ClaTet15	C2_157	ClaTet1
MD	**N**	36	38	38	38	36	38	38	38	38	38	38	37	36	33	34	35	36	37
	N_A	7	6	11	8	7	4	7	3	5	3	3	7	16	20	11	5	8	12
	H_e	0,802	0,723	0,586	0,652	0,438	0,319	0,738	0,554	0,102	0,148	0,506	0,703	0,812	0,885	0,790	0,691	0,807	0,766
MDL	**N**	47	47	47	47	47	47	47	47	47	47	47	47	47	46	46	46	45	45
	N_A	7	5	8	7	6	4	7	3	2	4	3	5	14	16	13	3	10	12
	H_e	0,746	0,713	0,797	0,708	0,502	0,280	0,776	0,510	0,042	0,195	0,493	0,758	0,725	0,882	0,837	0,618	0,842	0,856
ML	**N**	48	48	48	48	47	48	48	48	48	48	48	48	48	48	48	48	46	48
	N_A	8	5	11	10	10	6	7	4	3	4	2	6	15	23	13	5	10	13
	H_e	0,778	0,679	0,813	0,776	0,502	0,375	0,520	0,575	0,174	0,211	0,219	0,756	0,714	0,903	0,781	0,652	0,811	0,848
MS	**N**	40	42	41	42	41	42	42	42	42	41	42	42	41	40	41	42	40	40
	N_A	7	6	10	6	5	3	6	3	1	2	2	6	13	12	9	4	6	12
	H_e	0,761	0,579	0,642	0,657	0,544	0,384	0,369	0,513	0,000	0,176	0,375	0,671	0,811	0,770	0,680	0,343	0,676	0,855
ID	**N**	35	35	33	35	26	35	35	32	35	35	34	35	35	33	35	34	32	33
	N_A	6	6	10	9	6	4	8	4	3	4	2	6	16	17	14	6	9	13
	H_e	0,736	0,716	0,782	0,720	0,515	0,273	0,803	0,482	0,269	0,318	0,457	0,666	0,844	0,894	0,836	0,692	0,783	0,798
IDL	**N**	38	38	38	38	36	38	37	37	38	36	37	36	38	38	38	36	38	37
	N_A	7	6	10	10	12	6	8	3	4	5	2	8	16	17	12	8	9	13
	H_e	0,731	0,756	0,832	0,768	0,608	0,259	0,728	0,425	0,171	0,229	0,382	0,757	0,733	0,895	0,841	0,635	0,810	0,816
IL	**N**	35	35	33	34	32	35	35	34	35	35	35	34	35	34	34	33	35	35
	N_A	8	7	11	8	12	5	7	4	1	3	2	7	12	15	9	6	10	13
	H_e	0,353	0,717	0,833	0,770	0,643	0,407	0,718	0,419	0,000	0,299	0,320	0,773	0,764	0,912	0,814	0,502	0,816	0,849
IS	**N**	27	29	25	29	21	29	29	25	29	29	29	28	29	28	29	29	29	23
	N_A	6	4	10	6	5	3	6	3	1	2	2	8	14	12	8	3	7	9
	H_e	0,727	0,534	0,736	0,718	0,585	0,219	0,700	0,521	0,000	0,238	0,500	0,717	0,812	0,842	0,729	0,502	0,480	0,828

Pop		BWF1	BWF2	ClaTet3	ClaTet13	ClaTet18	Cocl-Lav4	Cocl-Lav6	Cocl-Lav10	Cocl-Lav27	BFRO018	Cocl-Lav18	Cocl-Lav49	Cocl-Lav52	ClaTet6	ClaTet9	ClaTet15	C2_157	ClaTet1
PDL	**N**	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43
	N_A	7	5	8	8	11	5	8	3	2	5	2	6	16	14	11	3	8	12
	H_e	0,745	0,686	0,792	0,757	0,553	0,289	0,733	0,568	0,067	0,416	0,483	0,762	0,813	0,747	0,785	0,563	0,742	0,844
PD	**N**	38	39	39	40	40	40	40	39	40	38	40	38	33	38	38	39	39	40
	N_A	8	7	13	9	8	6	8	3	4	3	3	8	18	20	12	6	10	16
	H_e	0,823	0,735	0,803	0,645	0,590	0,409	0,764	0,560	0,185	0,256	0,508	0,657	0,826	0,886	0,847	0,684	0,817	0,829
PL	**N**	33	35	34	35	35	35	35	35	34	35	34	35	34	30	34	34	35	35
	N_A	8	5	9	7	11	5	6	3	3	5	2	7	14	22	10	4	9	11
	H_e	0,759	0,654	0,767	0,697	0,499	0,464	0,479	0,527	0,112	0,302	0,344	0,775	0,768	0,881	0,722	0,617	0,833	0,831
PS	**N**	35	35	35	35	35	35	35	35	35	34	35	33	33	31	32	32	33	33
	N_A	6	4	10	5	3	3	6	2	1	2	2	5	11	11	9	3	6	9
	H_e	0,689	0,495	0,700	0,656	0,513	0,255	0,678	0,353	0,000	0,415	0,485	0,685	0,805	0,728	0,722	0,424	0,677	0,805

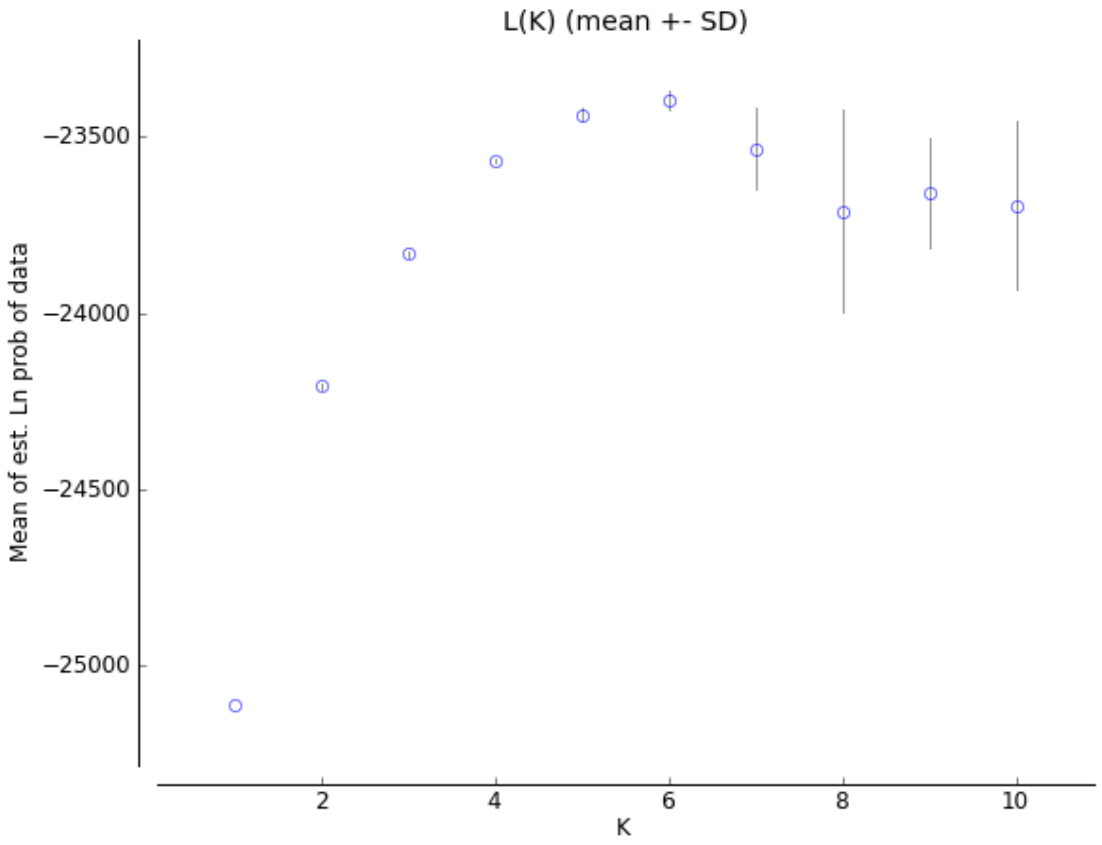


Figure S1 Probability of the number of clusters present in between the 12 whitefish populations. Most likely number of clusters (K) are in situations of highest lnP(K).

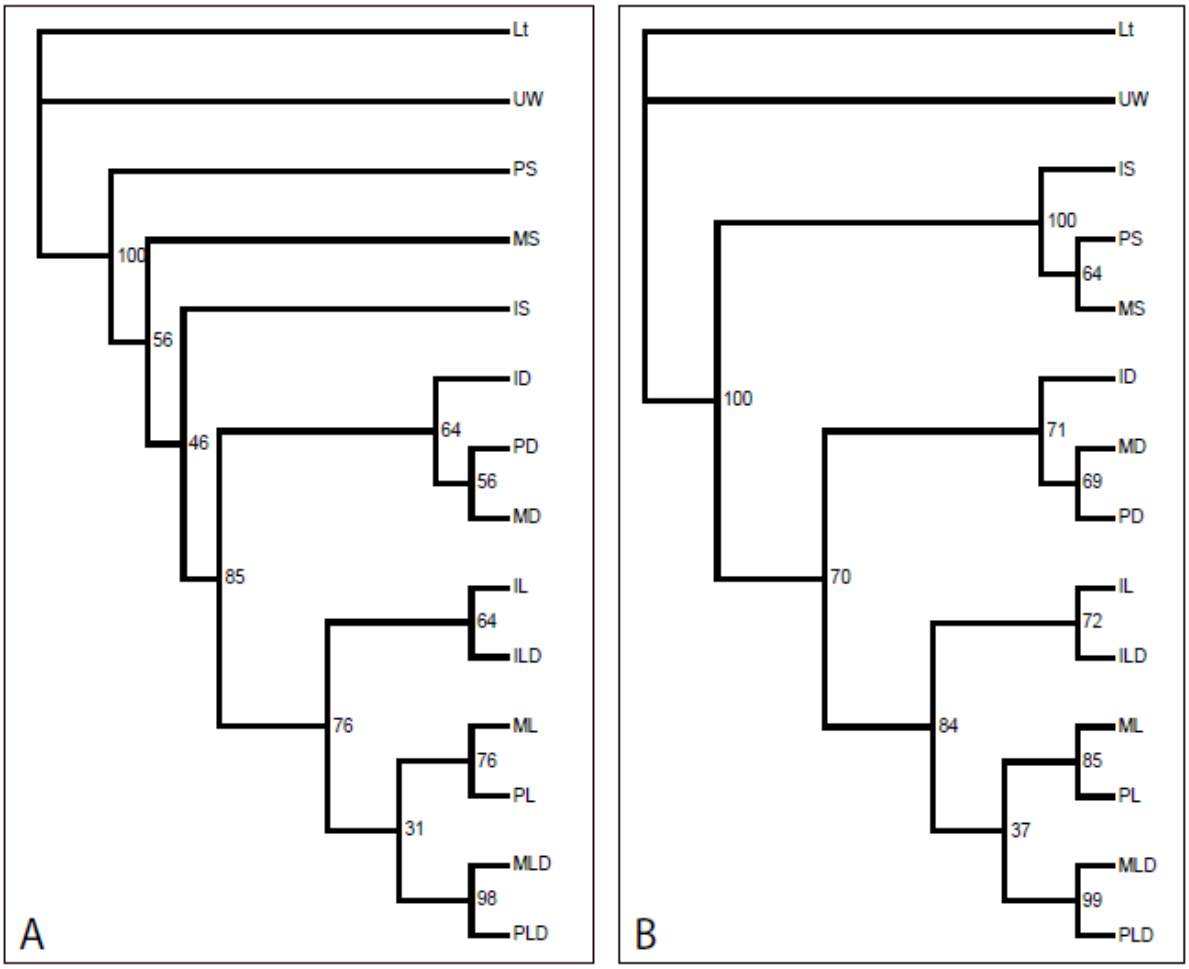


Figure S2 Phylogeny (with bootstrap values in % on 1,000 permutations) of 12 whitefish populations of three Finnish lakes in the Pasvik watercourse based on Nei's genetic distance (A) or Reynolds *et al.* genetic distance (B). Population codes are listed in Table 2. Lt and UW are Scottish LSR whitefish populations used as outgroups.