Species introduction promotes hybridization and introgression in *Coregonus*: is there sign of selection against hybrids?

KIMMO K. KAHILAINEN,*† KJARTAN ØSTBYE,‡§ CHRIS HARROD,¶** TAKAHITO SHIKANO,††
TOMMI MALINEN† and JUHA MERILÄ††

*Kilpisjärvi Biological Station, University of Helsinki, Käsivarrentie 14622, FI-99490 Kilpisjärvi, Finland, †Department of Environmental Sciences, PO Box 65, FI-00014 University of Helsinki, Finland, ‡Department of Biology, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, PO Box 1066, Blindern, NO-0316 Oslo, Norway, §Department of Forestry and Wildlife Management, Hedmark University College, Campus Evenstad, NO-2418 Elverum, Norway, ¶Department of Ecological Genetics, Max Planck Institute for Limnology, D-24302 Plön, Germany, **School of Biological Sciences, Medical Biology Centre, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK, ††Department of Biosciences, Ecological Genetics Research Unit, PO Box 65, FI-00014 University of Helsinki, Finland*

Abstract

Species introductions are considered one of the major drivers of biodiversity loss via ecological interactions and genetic admixture with local fauna. We examined two well-recognized fish species, native whitefish (*Coregonus lavaretus*) and introduced vendace (*Coregonus albula*), as well as their morphological hybrids in a single lake to test for selection against hybrids and backcrosses in the wild. A representative random subsample of 693 individuals (27.8%) was taken from the total catch of coregonids. This subsample was examined with the aim to select *c.* 50 individuals of pure whitefish ($n = 52$), pure vendace ($n = 55$) and putative hybrid ($n = 19$) for genetic analyses. The subsequent microsatellites and mitochondrial (mt) DNA analyses provided compelling evidence of hybridization and introgression. Of the 126 fish examined, four were found to be F_1, 14 backcrosses to whitefish and seven backcrosses to vendace. The estimates of historical gene flow suggested higher rates from introduced vendace into native whitefish than vice versa, whereas estimates of contemporary gene flow were equal. Mitochondrial introgression was skewed, with 18 backcrosses having vendace mtDNA and only three with whitefish mtDNA. Hybrids and backcrosses had intermediate morphology and niche utilization compared with parental species. No evidence of selection against hybrids or backcrosses was apparent, as both hybrid and backcross growth rates and fecundities were high. Hybrids (F_1) were only detected in 2 year-classes, suggesting temporal variability in mating between vendace and whitefish. However, our data show that hybrids reached sexual maturity and reproduced actively, with backcrosses recorded from six consecutive year-classes, whereas no F_2 individuals were found. The results indicate widespread introgression, as 10.8% of coregonids were estimated to be backcrosses.

Keywords: fitness, gene flow, gill raker, morphology, natural selection, niche

Received 13 March 2010; revised received 30 May 2011; accepted 15 June 2011

Introduction

The traditional biological species concept states that a species consists of actually, or potentially, inter-breed-

Correspondence: Kimmo K. Kahilainen, Fax: +358 16 3202100;
E-mail: kimmo.kahilainen@helsinki.fi

ing populations, which are reproductively isolated from other such populations and that interspecific mating would lead to lowered reproductive success via hybrid sterility and inviability (Mayr 1942). This traditional viewpoint has been challenged by the frequent observations of natural hybridization, and in some cases, genetic evidence indicates recent or past events of

introgression (Mallet 2005). The lower fitness of hybrids is generally related to the loss of advantageous adaptations via mixing with other species, which is an especially common outcome during the introduction or invasion by a new species (Rhymer & Simberloff 1996). Introductions are currently considered one of the major drivers of biodiversity loss both via ecological interactions and genetic admixture with local fauna (Seehausen *et al.* 2008).

The frequency of human introduction and invasions by new species has accelerated in an increasingly connected world; this has resulted in frequent gene flow with native taxa (Rahel 2002). The secondary contact between closely related species via translocation is especially prone to promote hybridization and introgression in plants and animals (Wilson & Bernatchez 1998; Blair & Hufbauer 2010; Senn *et al.* 2010; Steeves *et al.* 2010). In plants, such translocation could quickly promote hybridization and introgression with native species that may lead to complete admixture or in some cases, lead to the formation of hybrid species (Arnold 2006; Zalapa *et al.* 2010). The outcome of hybridization is highly diverse and depends on both the invasive species and receiving environments. According to Allendorf *et al.* (2001), there are three general patterns in hybridization between introduced and native species: i.e. hybridization without introgression, widespread introgression and complete admixture. Here, the last two patterns provide pathways to genetic admixture with introduced species that may potentially reduce phenotypic adaptations and fitness in native species. Salmonids fisheries have a long history of introductions with multiple adverse ecological and evolutionary effects on native species on a global scale (Crawford & Muir 2008; Korsu *et al.* 2010). Hybridization and introgression have frequently been detected, e.g. in the salmonid genera *Oncorhynchus*, *Salvelinus* and *Coregonus* across the Northern Hemisphere (Rogers *et al.* 2001; Redenbach & Taylor 2002; Rubidge & Taylor 2004). However, detailed studies comparing the ecological performance of pure and admixture individuals in the wild are surprisingly scarce.

The Holarctic coregonid fishes are extremely diverse in terms of behaviour, phenotypic traits and niche occupation. These factors have led to repeated taxonomic re-evaluation and an ongoing debate about the phylogenetic relationship between and within three recognized genera (i.e. *Coregonus*, *Prosophium* and *Stenodus*; Bernatchez 2004). Diversity is especially evident within European whitefish (*Coregonus lavaretus* (L.); Svärdson 1979; Østbye *et al.* 2005; Hudson *et al.* 2007). Coregonids were among the first fishes to colonize the Northern Hemisphere after the retreat of the last glacial ice sheets and faced ecological opportunities promoting their adaptive radiation in newly formed lakes (Østbye *et al.* 2006; Bernatchez *et al.* 2010; Siwertsson *et al.* 2010). The adaptive radiation of *Coregonus* was rapid in northern Fennoscandia where a single mtDNA ancestral lineage probably radiated into sympatric morphs within a time span of only *c.* 10 000 years (Østbye *et al.* 2005, 2006). Furthermore, whitefish have repeatedly diverged into the two major niches in the pelagic–littoral axis of these lakes, and in a few cases, also into a third major niche, the profundal zone (Amundsen *et al.* 2004; Harrod *et al.* 2010). In marked contrast, the highly specialized zooplanktivore vendace [*Coregonus albula* (L.)] has not diverged along these resource axes.

Whitefish and vendace are commonly found sympatrically in lakes in northern Europe. Although their reproduction may overlap both in time and location, size-assortative mating, among other prezygotic barriers, is thought to generally prevent interspecific hybridization (Svärdson 1976, 1979). However, hybridization between these taxa has been reported both in the wild and in the experiments (Garside & Christie 1962; Pethon 1974; Todd & Stedman 1989), indicating the lack of complete intrinsic isolation barriers. In experimental crossings between whitefish and vendace, there was no indication of lowered fecundity in F_1 hybrids, whereas some reduction of fecundity was observed in F_2 hybrids (Svärdson 1965). However, some experimental crossings of lake whitefish morphs (*Coregonus clupeaformis*, Mitchill) have reported reduced intrinsic F_1 hybrid viability and markedly reduced backcross viability compared with pure morphs (Rogers & Bernatchez 2006). Thus, the combined evidence suggests that both intrinsic and extrinsic isolation mechanisms are important to maintain the integrity of pure species. Here, we concentrate on extrinsic isolation barriers to hybridization, defined as the ecological factors following the successful hatching of eggs that could be related to niche utilization reflecting life history traits, e.g. growth and fecundity. Such detailed studies of putative selection against hybrids and backcrosses in relation to the most closely related resource competitors, namely their parental pure species, are largely lacking in the literature.

Here, we conducted a field survey in a subarctic lake, which harbours a natural whitefish population and an introduced vendace population. We aimed to examine, with the aid of morphology, mtDNA and microsatellite data, whether these closely related taxa hybridize and introgress in sympatry. To evaluate whether the taxonomic/genetically defined groups had diverged in terms of niche utilization, we compared both short-term (habitat use and diet) and long-term (stable isotopes and habitat-specific endoparasites) measurements. Finally, we searched for apparent signs of selection against hybrids and introgressed fish. We assumed that

only two optimal and divergent niches exist in the lake (viz. the pelagic and the benthic niche) and that 'genetically pure' vendace and whitefish will perform best in their respective niches. If hybrids had an intermediate morphology between the parental species, they would be less adapted to either the benthic or pelagic niche than the 'genetically pure' parental taxa (Hatfield & Schluter 1999; Gow et al. 2007). This maladaptation should be detectable in life history traits commonly used as fitness proxies: hybrids should have slower growth rates, rarely reach sexual maturity and have lower fecundity compared with 'genetically pure' species. The reduced fitness of hybrids should also be detected as an overall lower abundance of hybrids and the lack of backcrosses. Finally, we evaluate our results in comparison to the alternative hypothesis that coregonids in this lake represent a hybrid swarm where genetic mixing is not selected against.

Materials and methods

Study area and field sampling

Lake Rahajärvi (68°45'N, 27°17'E; L. Raha) is located in the subarctic Paatsjoki watercourse in Finnish Lapland at 132 m a.s.l. (Fig. S1, Supporting information). It is a relatively large (23 km^2) and deep (max 46 m, mean depth 14.1 m) oligotrophic lake (totP 4 μg/L, totN 100 μg/L) with clear water (colour 8 Pmg/L, Secchi depth 6.5 m, 1% of surface light left at depth of 9 m) surrounded by rocky terrain, mainly scattered with erratic boulders and pine (Pinus sylvestris) forest. The shoreline is predominantly rocky with a few sandy bays. The natural fish fauna is dominated by salmonids including whitefish, arctic charr [Salvelinus alpinus (L.)], brown trout [Salmo trutta (L.)] and grayling [Thymallus thymallus (L.)]. Other species include perch (Perca fluviatilis L.), pike (Esox lucius L.), burbot [Lota lota (L.)], minnow [Phoxinus phoxinus (L.)] and nine-spined stickleback [Pungitius pungitius (L.)]. Whitefish naturally occur as a monomorphic population.

To support local seine fisheries, vendace were introduced into the Paatsjoki watercourse in the 1950s and subsequently colonized the large Lake Inari (Fig. S1, Supporting information) during the following decades (Amundsen et al. 1999). In the late 1980s, vendace stocks became extremely abundant in L. Inari and subsequently invaded the lakes in the lower Paatsjoki watercourse (Bøhn & Amundsen 2001). At the same time, it was also intentionally transferred as a potential prey resource for piscivorous salmonids from L. Inari to L. Raha on three occasions during late autumn and winter. In 1987, 1989 and 1990, approximately 7000, 116 000 and 21 500 vendace were introduced to L. Raha

(M. Ahonen, personal communication). Vendace immediately began to reproduce, resulting in a rapid increase in population size. However, vendace experienced a demographic collapse at the end of the 1990s, resulting in a markedly low population size in the 2000s (M. Seppänen, personal communication).

The fish samples for this study were collected in September 2005, 18 years after the first introduction of vendace into L. Raha. Sampling was performed using a combination of gill net series, pelagic trawling and vertical echosounding (Kahilainen et al. 2004). We used a 240-m-long gill net series randomly combined of eight individual nets with mesh sizes (knot to knot) 12, 15, 20, 25, 30, 35, 45 and 60 mm. The trawl was 4 m high, 8 m wide, and the mesh size of the cod end was 3 mm (knot to knot). Trawl and gill net catches were divided to species and all coregonids (n = 2490) were measured (total length ±1 mm and mass ±0.1 g). Random subsamples from gill net and trawl catches (total of 693 coregonids) were frozen at −20 °C prior to morphological classification and subsequent analyses. This random sample was used also as a sampling population for the genetic and morphological analyses.

Genetic analyses

Initially, we partitioned all individuals (n = 693) into three a priori groups based on their external morphology: putative 'whitefish' having a subterminal mouth and the lowest number of short gill rakers; putative 'vendace' having a superior mouth and the highest number of long gill rakers; and putative 'interspecific hybrids' possessing intermediate characteristics. The initial aim was to sample c. 50 specimens from each of the pure species and their putative hybrids. As a result, all putative hybrids (n = 20), 53 putative whitefish and 56 putative vendace, were selected for genetic analyses. Three fish were removed because of failures in amplification (one putative vendace) and simultaneous amplification of three alleles per locus at some microsatellite loci (one putative whitefish and one putative hybrid). The amplification of three alleles was confirmed in three different PCR with three independent DNA extractions per individual. Thus, genetic assignments were conducted for 52 whitefish, 55 vendace and 19 putative hybrids. These same individuals were used for all subsequent morphological, ecological and life historical analyses.

Total DNA was extracted from muscle with a silica-fine-based purification method (Elphinstone et al. 2003) following proteinase K digestion. A fragment of the cytochrome b gene was amplified and sequenced using a forward (5'-GTTATTGCAGCTGCAACAGTCAT-3') and a reverse (5'-GGCTCATTCGAGGGCTTTAT-3') primer (Østbye et al. 2005). The forward primer was designed

in a conserved region of published cytochrome *b* sequences of salmonid species available at GenBank. PCR amplification and sequencing reaction were performed according to Shikano *et al.* (2010). The annealing temperature was set to 53 °C. Cycle-sequencing products were purified by ethanol precipitation and analysed on an ABI 3730*xl* DNA Analyzer (Applied Biosystems). The sequences were edited in MEGA 4 (Tamura *et al.* 2007). The best nucleotide substitution model was found in MODELTEST 3.7 (Posada & Crandall 1998) after an initial analysis in PAUP 4.0b10 (Swofford 2001). Here, the TrN + G model (Tamura & Nei 1993) fitted the data best using the hierarchical likelihood ratio test (hLRT). A total of 26 characters were phylogenetically informative. We ran a maximum-likelihood (ML) analysis in PAUP 4.0b10 with 10 random stepwise addition replicates and tree-bisection–reconnection swapping (TBR). Here, 100 bootstrap replicates were used to assess the statistical support. *Stenodus leucichthys* (Genbank Accession no.: DQ185400) was used as the genetic outgroup in the phylogenetic analysis.

Nuclear genetic variation was assessed at seven microsatellite loci: Bwf2 (Patton *et al.* 1997), Cisco-90, Cisco-126 (Turgeon *et al.* 1999), Cocl23, Cocl-Lav4, Cocl-Lav27 and Cocl-Lav61 (Rogers *et al.* 2004). Each forward primer was labelled with a fluorescent dye (FAM, HEX or TET) and the 5′-end of reverse primer was modified with a GTTT-tail (Brownstein *et al.* 1996). PCRs were performed using the Qiagen Multiplex PCR Kit (Qiagen) arranging loci in two multiplex PCR panels with nonoverlapping size ranges in each dye. The reactions were carried out following Shikano *et al.* (2010). The annealing temperature was set to 55 °C. PCR products were visualized on MegaBACE 1000 automated sequencers (Amersham Biosciences). The alleles were scored using Fragment Profiler 1.2 (Amersham Biosciences) with visual inspection and manual corrections of alleles by Takahito Shikano.

Microsatellites were run in Micro-Checker 2.2.3 (Van Oosterhout *et al.* 2004) on a priori groups (whitefish, vendace, hybrids) by locus basis, where no discrepancy was found regarding nonpositive integers, out-of-range values or zero values, or values with inconsistent modulus based on repeat motifs. Moreover, no evidence was found for scoring error because of stuttering, large allele dropout or for the presence of null alleles.

The genetic relationship among three morphologically determined groups was analysed using STRUCTURE 2.2.3 (Pritchard *et al.* 2000). The most likely number of genetic clusters was analysed using an admixture model with 2 500 000 burn-in steps and 7 500 000 MCMC replicates, with 33 replicates for each of the runs for $K = 1$–3, (i.e. putatively; '*whitefish*', '*vendace*' and '*interspecific hybrids*') based on methods in Evanno

et al. (2005) and Pritchard *et al.* (2000). We used a posterior probability of >0.85 for each fish over runs to denote a 'significant' assignment to either whitefish or vendace. The choice of this value was based on the concordance with the gill raker trait association as a highly heritable trait (Svärdson 1970; Rogers & Bernatchez 2007; Bittner *et al.* 2010), which together can be viewed as an evaluation of the biological reality of such theoretical genetic assignments. Ideally, one should aim at the highest value attainable in a given study or as a priori value such as 0.95 as in statistics for *P* values. However, an extensive diversity may exist in hybrid zones comprising older than two generation hybrids and thus we argue that our choice of a cut-off value on 0.85 is justified. In addition, we performed similar analysis in BAPS 3.2 (Corander & Marttinen 2006). However, as results were concordant between the two methods, only STRUCTURE data are presented.

To obtain more details with regard to the pattern of gene flow between taxa, we used NewHybrids 1.0 (Anderson & Thompson 2002) to estimate the genotype posterior probability that individual belonged to one of the six a posteriori categories; '*pure whitefish (hereafter referred to as WF)*', '*pure vendace (V)*', 'F_1', 'F_2', '*backcross to whitefish (BCWF)*' and '*backcross to vendace (BCV)*'. This program assumes that the sample consists of pure individuals and recent hybrids of species (or populations). This may cause uncertain assignments of later generation hybrids and backcrosses as whitefish and vendace have been able to interbreed during 18 spawning seasons before our field sampling. However, in lack of other available software that could cope with more than two generations of hybridization and introgression, we used NewHybrids 1.0, despite shortcomings and potential violation of assumptions. Here, a set of five analyses was performed using 100 000, 115 000, 175 000, 390 000 and 1 000 000 burn-in steps followed by an equal number of MCMC iterations, where a posterior probability >0.85 for each fish was decided to denote a 'significant' assignment to one of six categories. Moreover, we tested the ability of NewHybrids 1.0 to identify six categories by generating these events from simulated mating between 30 WF and 30 V as well as 30 of their offspring in each of the remaining four categories using Hybridlab 1.0. (Nielsen *et al.* 2006). Here, we used 10 replicates of 20 000 burn-in steps/MCMC iterations in NewHybrids. For ease of graphical interpretation of the real data with regard to the genetic admixture levels, we ran these simulated six categories (WF, V, F_1, F_2, BCWF, BCV) in STRUCTURE and plotted them (Fig. 1c). To evaluate these genetic assignment probabilities and group adherence in terms of biological realism, we compared six categories from NewHybrids (assignment as whitefish) and gill raker
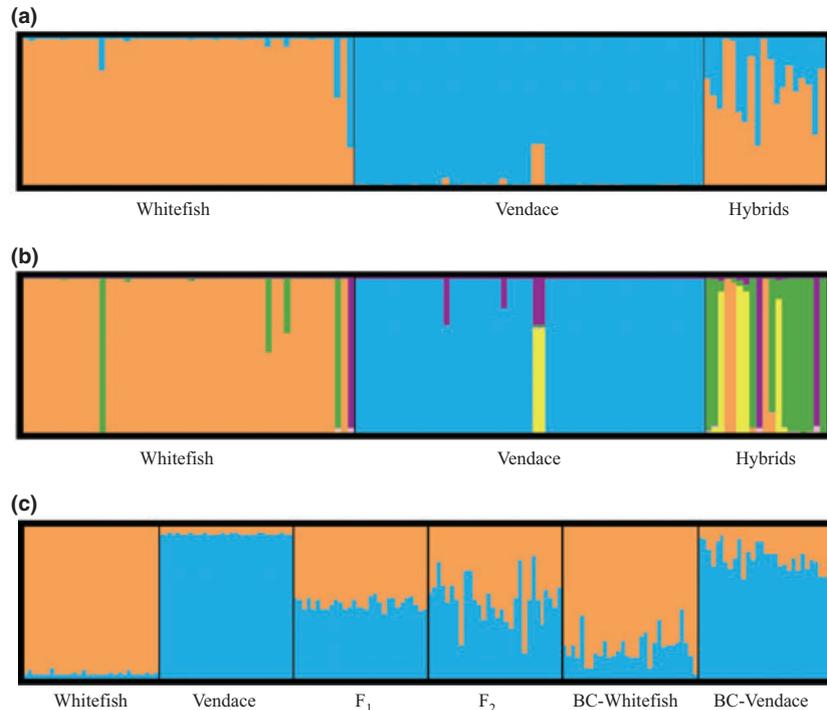
Fig. 1 Genetic classification of coregonid individuals. Genetic assignment based on STRUCTURE results (a). Probability of belonging to six a posteriori genetic groups based on NewHybrids analyses (b). Genetic assignment in STRUCTURE for the simulated six a posteriori groups using real parents as generated in Hybridlab (c). Colour codes are: orange = 'whitefish', blue = 'vendace', yellow = 'F$_1$', pink = 'F$_2$', green = 'backcross to whitefish' and purple = 'backcross to vendace'. In addition the a priori phenotypic/genetic groups are displayed in text below each figure.

counts (which is a highly heritable trait in coregonids; Svärdson 1970; Rogers & Bernatchez 2007) using Pearson correlation.

Observed heterozygosity (H$_o$), expected heterozygosity (H$_E$), allele counts, *F*-statistics and genotypic linkage disequilibrium for the assigned microsatellite groups detected in the NewHybrid analysis were estimated with Genepop 4.0.10 (Raymond & Rousset 1995). Here, deviations from Hardy–Weinberg equilibrium were tested by the exact (probability) test for locus and groups (Guo & Thompson 1992). Genetic differentiation was estimated using the log-likelihood-based exact test on the alleles. Tests were made across the groups and combined across loci using Fisher's combined probabilities. Rarefaction of alleles (where N genes were set to the lowest sample in the compared data set = 8 genes in the F$_1$ population, i.e. four fish) were performed using HP-Rare (Kalinowski 2005).

To estimate gene flow between whitefish and vendace, we combined backcrosses within their respective morphological and microsatellite-assigned parental pure species and removed all the F$_1$ individuals. The rationale for removing the four recognized F$_1$ individuals is that they are first generation hybrids that belong to both the 'whitefish genetic cluster' and the 'vendace genetic

cluster'. In the analyses, one then would need to randomly assign the F$_1$ individuals in the two groups. Instead of performing this arbitrary allocation, we choose to exclude the F$_1$ hybrids and obtain a more conservative estimate of gene flow. We further compared groups using Migrate-n version 3.0.3 (Beerli & Felsenstein 1999, 2001), estimating historical gene flow with the Brownian model on microsatellite repeat size, default values and a ML strategy, where the mutation rate was treated as constant for all the seven loci. Also, we estimated contemporary gene flow using BAYESASS 1.3 (Wilson & Rannala 2003), which calculates migration rates over the past two generations based on inferred proportion of immigrants. Here, three independent runs with 6×10^7 MCMC steps, where the first 10^6 steps were burn-in, seemed sufficient to obtain sufficient convergence (data not shown).

Morphological analyses

Body morphology was measured on the left side of each fish using a dial calliper (±0.01 mm). The following traits were measured: caudal peduncle depth, eye diameter, pelvic fin length, pectoral fin length, head length, snout length, upper jaw length and lower jaw

length (see Kahilainen & Østbye 2006 for details). In addition, the first left gill arch was dissected out from each fish to count the numbers of gill rakers and to measure gill arch length, length of the middle raker and width of middle raker under a microscope. A crude estimate of the gill raker space was also calculated dividing the gill arch length by the number of gill rakers. The morphological traits measured from head and gill rakers are strongly associated with foraging efficiency, and these traits reflect the differences between whitefish and vendace (Kahilainen *et al.* 2011). Gill raker number is known to be a heritable trait in *Coregonus* (Svärdson 1970; Rogers & Bernatchez 2007; Bittner *et al.* 2010) and significant heritability of other foraging-related head traits has been documented, for example in cichlids (Parsons & Albertson 2009). Vendace have a superior mouth and the highest number of gill rakers, whereas whitefish have a subterminal mouth and the lowest number of gill rakers. The morphological data set was length-calibrated using the allometric method (Thorpe 1975), which is an efficient way to remove size effects in morphological data (Reist 1985, 1986) and facilitates comparisons of morphological/genetic groups. This length-adjusted data set was analysed using a stepwise discriminant function analysis (DFA), where morphological traits were included in analysis until the discriminant power was not further improved. The selection threshold for traits to be included in analysis was $P < 0.05$.

Niche occupation

The niche occupation of species represents the sum of solutions in a given lake environment to avoid predators, and at the same time, to obtain food. As such, it includes both habitat use and diet preferences, which may also change during the ontogeny for a given fish. The niche occupation of a fish, or a population, is thus inherently difficult to delimit, as sample collection represents only a temporal snapshot. As an approximation, one can compare short-time habitat use from capture location, with measures of both short- and long-term consumption patterns to obtain more representative representation of individual niche occupation.

To obtain estimates of habitat use of a posteriori genetic groups, we examined the presence–absence of group-specific individuals in the different principal habitat types based on total catch data. Here, the principal habitat types (littoral, pelagic and profundal) were classified according to the 1% light level at 9 m (measured using LI-1400, Li-Cor, Inc., USA, with a wavelength range 400–700 nm), which defines the compensation depth i.e. at depth 0–9 m, the photosynthetic rate is higher than the respiration rate (Lampert & Sommer

2007). The littoral and pelagic zone (light level >1%) are productive habitats providing photosynthetically derived energy by algae and plants for higher trophic organisms, whereas the dark and deep profundal zone (light level <1%) relies on the settling subsidies (dead organisms, particular organic matter) from overlaying productive habitats. We used the measured compensation depth (9 m) and respective 9-m depth contour from bathymetric map to calculate the proportions of principal habitat types. In L. Raha, the littoral and pelagic/profundal areas contribute to 42.6% and 57.4% of the total surface area, respectively. The proportion of group-specific individuals present in each habitat type was calculated from catch data. The similarity of habitat use in a posteriori genetic groups was then compared using Schoener's (1970) similarity index.

To illuminate the short-time dietary niche, we removed the stomach and estimated fullness using the points method (Hynes 1950) on a scale from 0 (empty) to 10 (extended full stomach). We identified prey items to family or higher levels and estimated their relative contribution visually. The prey items were further pooled into four categories: cladocerans (*Bosmina*, *Daphnia*, *Holopedium*), copepods (Cyclopoida, Calanoida), terrestrial insects (adults and pupae) and benthic prey (*Pisidium*, *Eurycercus*, Trichoptera, chironomid larvae). We compared overlap in the short-term diet of the a posteriori genetic groups using the Schoener's (1970) similarity index, in which values ≥0.60 were considered biologically significant (Wallace 1981). Prey size was further measured from 10 randomly sampled zooplankton specimens from each of the undigested stomach content to compare how efficient foragers are in pure species and admixture individuals. For copepods, body length was estimated from the rostrum to the furca, and for cladocerans, from the head to the root of the tail spine (details in Kahilainen *et al.* 2005).

The long-term niche utilization of a posteriori groups was estimated through stable isotope analysis (SIA) of carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$). Here, $\delta^{13}C$ indicates the utilization of resources along the pelagic–littoral resource axis and $\delta^{15}N$ refers to the trophic level of the consumer (Post 2002). In subarctic lakes of this region, the lower $\delta^{13}C$ values (*c.* −31‰ to −27‰) indicate pelagic habitat increasing towards littoral (*c.* −25‰ to −20‰) (Harrod *et al.* 2010). $\delta^{15}N$ values (*c.* 3–4‰ per trophic level) indicate trophic level of organisms, where primary producers are *c.* 1–2‰, primary consumers, zooplankton and benthic macroinvertebrates *c.* 4–8‰ and secondary consumers e.g. planktivorous and benthivorous fish *c.* 6–11‰ (Eloranta *et al.* 2010; Harrod *et al.* 2010). Both $\delta^{13}C$ and $\delta^{15}N$ were estimated from dorsal muscle. Samples were dried for 24 h at 60 °C, ground, weighed, encapsulated in tin cups and analysed with a stable

isotope ratio mass spectrometer (see Harrod *et al.* 2005 for details). Differences in C:N ratios among groups (ANOVA: $F_{4,115} = 10.78$, $P < 0.001$) indicated variation in lipid concentrations and thus data were arithmetically lipid-normalised (Kiljunen *et al.* 2006).

Selective endoparasite screening was used to detect the long-term utilization of pelagic or benthic niches. *Icthyocotylurus erraticus* flatworms transmitted through the water column in littoral areas (Harrod & Griffiths 2005) and *Diphyllobothrium* sp. tapeworms transmitted via pelagic copepods (Knudsen *et al.* 2003) were used as habitat-specific indicators of benthic and pelagic niche use, respectively. Coregonids encyst *I. erraticus* on the heart surface and *Diphyllobothrium* sp. on the stomach wall. Infection intensity (abundance per host) of cysts of these parasite taxa was calculated for each fish and prevalence (present or not present in the specimen) was estimated for each of a posteriori groups (whitefish, vendace, F_1 hybrid, BCWF, BCV).

Life history traits

Age of each fish was determined using sagittal otoliths and ventral scales collected from behind the pelvic fins (details in Kahilainen *et al.* 2003). We used these data to describe how many year classes of a posteriori groups were detected in L. Raha. The growth pattern of each a posteriori group was estimated using the Von Bertalanffy (1938) nonlinear growth equation. Sex and the maturation stage was recorded according to the gonad stage (0 = immature, 1 = mature). Because of the relatively low sample sizes, sexes were pooled in the maturity analyses, estimating size and age, where 50% of each of the a posteriori groups reached sexual maturity. Pooling of the sexes in this case is justified as there is little variation in maturation trajectories between the sexes in both whitefish and vendace (Bøhn & Amundsen 2004; Bøhn *et al.* 2004). To obtain fecundity estimates, mature female gonads were weighed and then *c.* 10 subsamples from different parts of the gonads were weighed and eggs were counted (Bagenal 1978). In addition, egg diameter was measured from 10 random eggs. Because of differences in body size among populations, fecundity was presented as eggs per gram of wet mass.

Population densities

Coregonid densities were estimated from areas deeper than 5 m using a SIMRAD EY-500 echosounder equipped with a down-facing split-beam transducer (120 kHz). Hydroacoustic data were sampled along 10 transects situated 0.5 km apart during darkness (Fig. S1, Supporting information), which is a suitable time frame for coregonid population estimation (Kahilainen *et al.* 2004; Jensen *et al.* 2008). Coregonids were the main species captured in simultaneous trawl catches: very small-bodied nine-spined sticklebacks were also captured, but were excluded from hydroacoustic analyses using a target strength threshold of −53 dB. Hydroacoustic data were analysed using EP500 software (SIMRAD). Targets between 2 m from the surface to 0.5 m above the bottom were used in the analyses. The overall coregonid density (all *Coregonus*) was calculated using the observed target strength distributions (see Kahilainen *et al.* 2004 for details). Then the coregonid total density was first divided into each a priori group (*whitefish, hybrids, vendace*) according to the proportions in the random subsample ($n = 693$). Then, we used higher resolution from genetic data to calculate the proportions for a posteriori group (WF, BCWF, F_1, BCV, V) and finally densities for these groups.

Statistical analyses for the ecological data

We used ANOVA to test differences between compared groups, and Tukey's HSD tests to reveal pairwise differences. Relevant transformations were undertaken if the variables were not normally distributed. Analysis of covariance (ANCOVA) using age as a covariate was employed for the analysis of parasite infection intensity and egg size. Regression analyses were used to calculate the age and size at maturity for a posteriori population. A DFA using a stepwise selection procedure was used for the multivariate comparisons of morphological data. An alpha value of 0.05 was used throughout to represent statistical significance.

Results

Genetic analysis

In the following paragraphs of this section, we compare the results between putative groupings based on morphological analyses determined in the laboratory (52 whitefish, 55 vendace and 19 putative hybrids) and genetic assignments described here. All genetic analyses suggested that the three groupings reflecting laboratory-based morphological assignments were an oversimplification. Here, the combined genetic results from mtDNA and microsatellite data provided evidence for the presence of five different groups: WF, V, F_1 hybrids, BCWF and BCV. In more detail, the morphologically determined groups were partitioned into the following categories according to the combined evidence from the morphological and genetic data sets: 52 morphologically assigned whitefish (47 WF, 4 BCWF, 1 BCV), 55 morphologically assigned vendace (51 V, 4 BCV) and 19

morphologically assigned hybrids (4 F_1, 10 BCWF, 2 BCV, 3 WF). The detailed genetic results for this a posteriori grouping are presented later.

Ten haplotypes were observed from a total of 562 bp scored cyt *b* sequences (Table 1). Of the 10 observed haplotypes, five (h1–5) were found in the morphologically assigned whitefish, and five (h3, h5–h8) in the morphologically assigned vendace, while seven haplotypes were found in the morphologically assigned hybrids (h1–h3, h5, h6, h9, h10; Table 1). Two haplotypes (h9–10) were only observed in the morphologically assigned hybrids. Two haplotypes were shared among three groups (h3, h5). Four haplotypes were shared between whitefish and hybrids (h1–3, h5), while three haplotypes (h3, h5, h6) were shared between vendace and hybrids. Based on the bootstrap values of mtDNA haplotypes (Fig. S2, Supporting information; 98% for the whitefish cluster and 100% for the vendace cluster from outgroup *Stenodus leucichthys*), 49 of the 52 putative morphologically WF had whitefish mtDNA, providing a 94% match between morphological and mtDNA assignments. The remaining three fish had vendace mtDNA. For the putative 55 morphologically V, all had vendace mtDNA, giving a 100% match of morphological and mtDNA assignment. Of 19 a priori hybrids, only four had whitefish mtDNA, whereas 15 had vendace mtDNA.

The results from STRUCTURE (see Fig. S3 and Table S1, Supporting information) suggested that two genetic clusters ($K = 2$; *Mean LnP(K) = −2168, Ln'(K) = 865*)

comprising whitefish and vendace were more likely than either one (*Mean LnP(K) = −3033, Ln'(K) = NA*) or three genetic clusters (*Mean LnP(K) = −2158, Ln'(K) = 10*) (Fig. 1a). The NewHybrids analyses revealed that 47 of 52 a priori whitefish had genetic assignment values >0.99 back to WF, giving an 84% match between morphological and genetic assignments (Fig. 1b, for data see Table S2, Supporting information). Of the remaining five fish, two had assignment >0.97 to BCWF, while two fish had assignment values 0.52–0.64 to WF and 0.36–0.48 to BCWF, suggesting they belonged to a hybrid/backcross category not detected by NewHybrids. We grouped these two fish into a new category termed '*uncertain backcross to whitefish*'. The remaining single fish had assignment 0.96 to BCV.

In 55 a priori vendace, 51 fish had assignment values >0.99 to V, giving a 90.4% match of morphological and genetic assignments (Fig. 1b). Of the remaining four fish, two had an assignment 0.70–0.81 to V and 0.19–0.30 to BCV, and two had an assignment of 0.67 to 'F_1' and 0.30 to BCV, suggesting they belonged to a hybrid/backcross category not detected by NewHybrids. We therefore grouped these four fish into a new category termed '*uncertain backcross to vendace*'.

Of 19 a priori hybrids, three fish had assignment >0.97 to WF, four an assignment >0.87 to 'F_1', 10 an assignment >0.86 to BCWF, while two had an assignment >0.95 to BCV giving a 84.2% match between morphological and genetic assignments (Fig. 1b). Thus, of the 126 fish that were genetically classified, we found:

Table 1 Genetic and morphological assignments of coregonid individuals

| Haplotype | N | Morphological assignment | | | Microsatellite assignment | | | | |
		WF	H	V	WF	BCWF	F_1	BCV	V
Cluster 1									
h1	15	14	1	—	13	2	—	—	—
h2	35	34	1	—	34	1	—	—	—
h4	1	1	—	—	1	—	—	—	—
h9	1	—	1	—	1	—	—	—	—
h10	1	—	1	—	1	—	—	—	—
Cluster 2									
h3	20	1	3	16	—	3	1	—	16
h5	27	2	7	18	—	6	2	1	18
h6	24	—	5	19	—	3	1	5	15
h7	1	—	—	1	—	—	—	—	1
h8	1	—	—	1	—	—	—	—	1
Sum	126	52	19	55	50	15	4	6	51

Abbreviations in a priori morphological assignment are whitefish (WF), hybrid (H) and vendace (V). The first two columns indicate the number of cyt *b* haplotypes observed in the sample of 126 coregonids from L. Raha. Cluster 1 denotes a bootstrap cluster, which includes mostly whitefish based on either a priori morphological assignment or a posteriori Bayesian microsatellite assignment (BCWF, backcross to whitefish, F_1, hybrid, BCV, backcross to vendace) (Fig. S2, Supporting information). Similarly, Cluster 2 denotes a bootstrap cluster including mostly vendace, but also hybrids (Fig. S2, Supporting information).

50 WF, 51 V, 4 F_1, 12 BCWF, 2 'uncertain backcross to whitefish', 3 BCV and 4 'uncertain backcross to vendace'. 'Uncertain backcrosses' were grouped into their respective backcross groups according to the strong and supporting morphological patterns (i.e. gill raker counts and morphology).

The ability of NewHybrids 1.0 to identify six a posteriori categories in our simulated data set (Fig. 1c) including real parents, resulted in the following posterior probabilities (assignment range/mean assignment): WF (range 0.95–0.99/mean 0.98), V (0.97–0.99/0.99), F_1 (0.81–0.97/0.95), F_2 (0.01–0.99/0.69), and BCWF (0.01–0.96/0.78) and BCV (0.02–0.95/0.75) (see Fig. S4 and S5, Table S3, Supporting information). We further evaluated the power of NewHybrids to assign individuals back to the six a posteriori categories, selecting only fish with a lower assignment to its own group than 0.85 (see Fig. S6, Table S4, Supporting information). This was performed as the same value was used in the assignment of all real individuals. For the two categories WF and V, none of the fish had <0.85 assignment to its own group.

In the F_1 group, only one fish had a mean assignment value below this threshold (0.82), where the next highest assignment values were to BCV (0.09) and F_2 (0.07). This suggests a slight overestimation of backcross and hybrid frequency, a slight underestimation of F_1 values, but would have little impact on the overall frequency of hybrids in the real data set (but, a slight inaccuracy of estimates for each group). For F_2, in 16 fish, the mean assignment was 0.45 to F_2, 0.23 to BCWF and 0.21 to BCV. The same argument holds as for F_1, but the result suggests that this effect was stronger for F_2 than for F_1. For BCWF, the mean assignments for seven fish were 0.44 to W, 0.33 to BCWF and 0.13 to F_2, whereas for BCV, the mean assignments for 12 fish were 0.50 to BCV, 0.42 to V and 0.08 to F_2. These two backcross categories seem to have similar values regarding assignment to their own group as well as their closest backcross group and F_2. The simulated backcrosses were mostly misassigned to their closest parental group, which further implies that the extent of hybridization and introgression is actually underestimated in the real data set (as well as a slight overestimation of the F_2 assignment values).

Evidently, in retrospect, our capacity to satisfactorily assign wild individuals into the six a posteriori genetic categories in NewHybrids would have been much improved by increasing the number of microsatellite markers. Another issue relates to the violation of the NewHybrid detection limitation that comprises only the last two generations, where our real data comprise more than these two generations. Here, we would expect that NewHybrids have a low power to detect the genetic groups extending these two generations. However, the rather close association of our inferred NewHybrid groups with the gill raker counts strongly suggest that our classifications are meaningful as they have a biological realism, as gill raker counts are reflected in our expectations. Thus, in conclusion, despite the inaccuracy regarding hybrid categorization frequencies, our simulations imply an underestimation of the extent of hybridization and introgression in the wild. As such, our data provides a conservative estimate of the degree of hybridization and introgression in whitefish and vendace in L. Raha. A following up study of this system using more genetic markers will reveal a much finer resolution.

In the following section, we review the categories from NewHybrids, comparing data with mtDNA haplotypes. In WF (50 fish with assignment >0.97) and V (51 fish with assignment >0.99), all had mtDNA from their own species. In F_1 (four fish with assignment >0.87), all had mtDNA from V suggesting female $_V$ × male $_{WF}$ mating. F_2 hybrids were not found. In BCWF (14 fish with assignment >0.36), two scenarios emerged. In three cases, mtDNA originated from WF, implying: (i) the first F_1 was produced by female $_{WF}$ × male $_V$, and then female F_1 subsequently mated with male $_{WF}$, or (ii) that the first F_1 was produced by female $_{WF}$ × male $_V$, or (iii) by female $_V$ × male $_{WF}$, and then a male from either of these F_1 generations mated with female $_{WF}$. In the eleven remaining cases, mtDNA came from V, suggesting that: (iv) the first F_1 was produced by female $_V$ × male $_{WF}$, and then female F_1 subsequently mated with male $_{WF}$. In BCV (seven fish with assignment >0.19), all had mtDNA from V suggesting: (v) the first female F_1 individuals were produced by female $_V$ × male $_{WF}$, which subsequently mated with male $_V$; or (vi) that the first male F_1 individuals were produced by female $_V$ × male $_{WF}$, or (vii) by female $_{WF}$ × male $_V$, and then a male from either of these F_1 generations further mated with female $_V$.

No evidence ($P > 0.05$) for genotypic linkage disequilibrium was found in any of the five groups using microsatellite loci: WF, V, BCWF, BCV and F_1. Deviations from Hardy–Weinberg equilibrium after Bonferroni adjustment was evident only for the vendace locus Cisco90 (heterozygote deficit; see Table S5, Supporting information). Observed (H_o) and expected heterozygosities (H_E), as well as allele counts, are shown in Table S5 (Supporting information). Genetic differentiation was significant in all pairwise comparisons of the five groups, except in the case of BCWF × F_1 (G-test: $P = 0.514$) and BCV × F_1 ($P = 0.699$). Statistically significant F_{ST}-values ranged as follows: WF vs. V (0.495), WF vs. BCWF (0.086), WF vs. BCV (0.385), WF vs. F_1 (0.207), V vs. BCWF (0.286), V vs. BCV (0.050), V vs. F_1 (0.155) and BCWF vs. BCV (0.132).

The historical gene flow as estimated using Migrate 3.0.3 was significantly higher from vendace into whitefish (maximum likelihood estimation (MLE); 5–95 percentiles = 1.18; 1.00–1.36) than from whitefish into vendace (0.44; 0.37–0.51). However, the results (based on three independent runs) from BAYESASS 1.3 suggested that the direction of contemporary gene flow did not differ significantly between the two species (i.e. from vendace into whitefish; 0.0174; −0.002 to 0.0362 and from whitefish into vendace; 0.0085; −0.0056 to 0.0236). The reported values here are from one of the three runs performed.

Morphology

The significant negative correlation between genetic assignment as whitefish and gill raker count (Fig. 2a, $r_{1,124} = -0.94$, $P < 0.01$) suggested a continuum from whitefish (the lowest number of gill rakers) towards vendace (the highest number of gill rakers). The genetically WF and V groups had the most distinct gill raker counts, whereas hybrids and backcrosses were intermediate between the pure parental species (Fig. 2a, b). The mean gill raker counts of WF, BCWF, F$_1$, BCV and V were 30, 34, 38, 40 and 43, respectively. Gill raker count differed between groups (ANOVA, $F_{4,121} = 206.21$, $P < 0.001$), where parental groups differed from each other, F$_1$ and backcrosses (Tukey's HSD tests, $P < 0.001$). BCWF and BCV also differed (Tukey's HSD tests, $P < 0.001$); however F$_1$ and backcrosses were not statistically different.

All morphological measurements were used in the DFA, which classified different groups using two significant functions, function 1 (eigenvalue = 15.35) and function 2 (eigenvalue = 0.92; Fig. 2c). Altogether, eight morphological traits minimizing Wilk's Lambda were included in the stepwise analysis. In order of decreasing significance, these were as follows: gill raker count, snout length, lower jaw length, gill raker space, gill arch length, head length, gill raker length and gill raker width. The stepwise DFA classified 94.4% of the individuals correctly according to their genetic assignment. All V and F$_1$ specimens were correctly classified. One WF was misclassified as BCWF, and two BCWF were classified as WF. BCV were misclassified to V (three individuals) or BCWF (one individual). In the scatterplot of the DFA scores (Fig. 2c), WF, F$_1$ and V represent the endpoints, whereas backcrosses were morphologically intermediate or overlapping with parental species. Some of the backcrosses had morphologies identical to that of the pure species. There were significant differences in centroid values in function 1 (ANOVA, $F_{4,120} = 460.75$, $P < 0.001$), where all population types differed (Tukey's HSD tests, $P < 0.05$), except BCWF and F$_1$. There were differences also in function 2 (ANOVA, $F_{4,120} = 27.67$, $P < 0.001$). Here, the centroids of WF, BCV and V did not differ ($P > 0.05$), but all other population types differed (Tukey's HSD tests, $P < 0.05$).

Niche dimensions

WF and BCWF used all major habitat types (Fig. 3a). The pelagic zone was the major habitat for all the groups, except WF. The profundal zone was also utilized by all the groups, whereas the littoral zone was the least used habitat and was mainly restricted to individuals with the lowest gill raker counts observed in WF and BCWF. Estimates of niche overlap revealed that
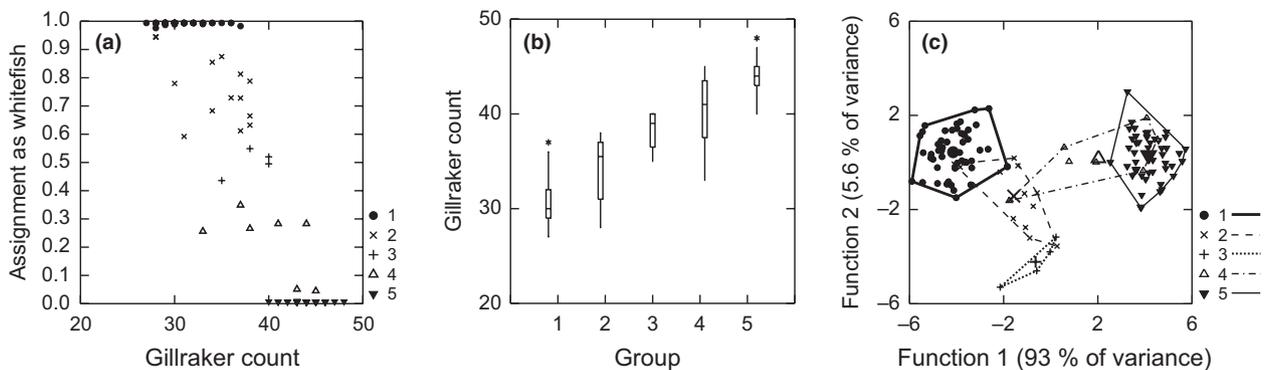
Fig. 2 The morphological characterization of genetically pure and admixed coregonid groups in L. Raha. Correlation between genetic assignment and gill raker count (a), mean gill raker count (b) and combined morphology (c) of whitefish (1), backcross to whitefish (2), F$_1$ hybrids (3), backcross to vendace (4) and vendace (5). In box-plot (b), the box indicates the range of central values (50%) between the first and third quartile. The horizontal line in the boxes indicates the median values and whiskers indicate the minimum and maximum values (outlier and extreme values discounted). The scatterplot (c) indicates morphological scores of each individual using two functions explaining 98.6% of total variance. The convex hulls differentiate the populations whilst the larger markers located in the middle of convex hulls show the population centroids.
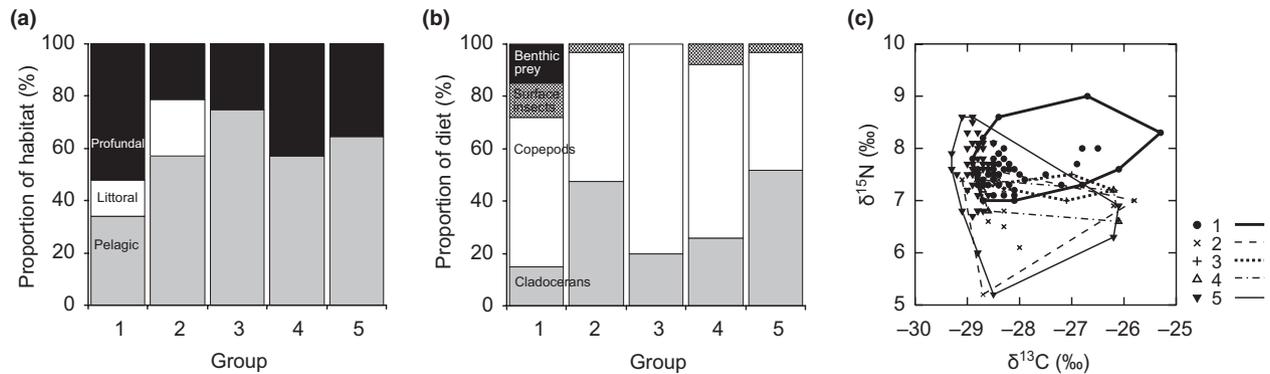
Fig. 3 The niche utilization of coregonids in L. Raha. Niche dimensions in habitat use (a), diet (b) and isotopic values (c) of white-fish (1), backcross to whitefish (2), F_1 hybrids (3), backcross to vendace (4) and vendace (5). Principal habitat types and major groups of prey are presented in bars (a, b). Individual isotopic values are presented with symbols and convex hulls combine the population values.

WF and F_1 used generally similar habitats ($D = 0.59$), whilst other groups used very similar habitats ($D = 0.69$–0.92). The highest overlap (0.92) was observed between BCV and V.

Pelagic prey dominated the diet of all groups, and only WF had consumed benthic prey (Fig. 3b). Feeding activity according to stomach fullness was roughly similar in all populations (Table 2, ANOVA, $F_{4,121} = 1.69$, $P = 0.15$). In addition, diet was biologically similar ($D = 0.63$–0.95) among all groups. Zooplankton was the main prey consumed by all groups; however, the size of consumed prey differed (ANOVA, $F_{4,303} = 57.22$, $P < 0.001$). The mean prey size of WF, BCWF, F_1, BCV and V was 1.77, 0.80, 2.18, 1.71 and 0.74 mm, respectively. V and BCWF used smaller prey than other groups (Tukey's HSD tests, $P < 0.05$).

Stable isotope data indicated that all groups utilized pelagic habitats over the long-term. Hybrids and back-crosses utilised the isotope niche space between that of their pure parental species (Fig. 3c). The differences among the average $\delta^{13}C$ values for the different groups (Table 2, ANOVA, $F_{4,120} = 7.90$, $P < 0.01$) indicated that V were more pelagic (i.e. $\delta^{13}C$-depleted) than WF, F_1 and BCV (Tukey's HSD tests, $P < 0.05$), but no other comparisons were nonsignificant. There were also differences in the average $\delta^{15}N$ values (Table 2, ANOVA, $F_{4,120} = 6.12$, $P < 0.01$). In the pairwise comparisons, WF and V had higher $\delta^{15}N$ values than BCWF (Tukey's HSD tests, $P < 0.05$), whereas no other differences were detected.

The evidence for the utilization of the pelagic habitat was indirectly apparent from other long-term biological

Table 2 Niche utilization, life history and density indicators for different coregonid groups [whitefish (WF), backcross to whitefish (BCWF), hybrids (F_1), backcross to vendace (BCV) and vendace (V)] in L. Raha

Data	WF	BCWF	F_1	BCV	V
Stomach fullness	4.4	4.5	2.5	4.8	5.1
$\delta^{13}C \pm 95\%$ CI	−28.1 ± 0.2	−28.2 ± 0.6	−27.2 ± 1.6	−27.6 ± 1.3	−28.7 ± 0.2
$\delta^{15}N \pm 95\%$ CI	7.6 ± 0.1	6.9 ± 0.4	7.3 ± 0.3	7.0 ± 0.3	7.5 ± 0.2
Diphyllobothrium sp.	31.3 (92)	3.8 (71)	23 (100)	2 (42)	0.7 (31)
Ichthyocotylurus sp.	15.8 (62)	0.1 (14)	10 (50)	0 (0)	0.3 (17)
Growth rate (K) ± 95% CI	0.18 ± 0.01	0.12 ± 0.03	—	0.32 ± 0.18	0.34 ± 0.05
Asymptotic length (L_∞) ± 95% CI	40.3 ± 1.1	42.1 ± 6.0	—	21.6 ± 3.3	20.0 ± 0.6
Model age at zero size (t_0)	0.01	−1.03	—	−0.82	−0.76
Sexual maturity size (cm)	26.8	19.8	—	13–14	12.8
Sexual maturity age (years)	6+	4+	—	2+ to 3+	2+
Year class diversity (n)	10	5	2	6	9
Population density (ind./ha)	245	27	2	12	76
Proportion in lake (%)	67.7	7.5	0.5	3.3	21

Stomach fullness indicates short-term feeding activity (min 0, max 10), whereas mean stable isotope values ($\delta^{13}C$ and $\delta^{15}N$) and parasite values (mean abundance and prevalence in the parenthesis) represent long-term niche use. Life history traits come from the von Bertalanffy growth model estimates and probit regressions. Year class diversity indicates the number of observed year classes based on age determinations. The density of coregonid groups was assessed through echosounding and trawling.

indicators i.e. benthic and pelagic transmitted parasites. The prevalence and infection intensity of *Diphyllobothrium* sp. tapeworms (pelagic) was higher than *Icthyocotylurus* sp. flatworms (benthic) in all population types (Table 2). WF had the highest and V the lowest amount of both parasite types. F_1 had an intermediate amount of parasites (Table 2). Backcrosses had relatively low parasite abundance and prevalence compared with F_1. Population type had an effect on *Diphyllobothrium* sp. abundance after age correction (ANCOVA, $F_{4,120} = 9.82$, $P < 0.001$). V had lower amount of *Diphyllobothrium* sp. than WF (Tukey's HSD test, $P < 0.05$), but other comparisons were not significant. *Icthyocotylurus* sp. abundance had a large overall variance and differences were not significant (ANCOVA, $F_{4,120} = 2.05$, $P = 0.091$).

Life history

According to growth curves, WF and BCWF attained the largest size and fastest growth, observed F_1 individuals attained an almost similar size, whereas BCV and V were the smallest (Fig. 4a, Table 2). There were differences in the observed number of year classes between six a posteriori groups (Table 2). The highest number of year classes was observed for WF (years 1995–1999, 2001–2005) and V (1997–2005), whereas for F_1, only individuals from 1994 to 1997 were found. The observed number of year classes for BCWF was 1998, 2002–2005 and for BCV 2000–2005. The size at sexual maturity at the population level (50% mature) closely followed the order of growth variation, where WF, BCWF and V reached sexual maturity at 26.8, 19.8 and 12.8 cm (Probit regression, length term, $P < 0.10$). For BCV, our raw data indicated that at the population level, individuals reached sexual maturity at lengths 13–14 cm, but statistical analyses did not support this ($P > 0.10$). For F_1, all individuals collected (range 31.3–35.5 cm) were sexually mature females, hence estimation of age or length at sexual maturity was not possible. The age at sexual maturity for WF, BCWF and V were 6+, 4+ and 2+ (Probit regression, age term, $P < 0.05$). Probit regression suggests that BCV mature at age 2+ to 3+, but there was no statistical support ($P > 0.05$).

Mean egg diameters differed between the population types (ANOVA, $F_{3,40} = 68.62$, $P < 0.001$) apart from BCV and V (Tukey's HSD tests, $P > 0.05$). WF had the largest mean egg diameter (1.76, $n = 10$), F_1 were intermediate (1.37, $n = 4$), whereas BCV (1.07, $n = 3$) and V (1.06, $n = 27$) had the smallest eggs (Fig. 4b). Estimates of relative fecundity differed between population types (ANOVA, $F_{3,40} = 60.48$, $P < 0.001$), except BCV and V (Tukey's HSD tests, $P > 0.05$; Fig. 4c). Here, relative fecundity was the lowest in whitefish (345 eggs/g), intermediate in F_1 (694), and markedly high in both BCV (1851) and V (1858). We were unable to provide estimates for BCWF because of a lack of mature females.

Coregonid density

According to the combined hydroacoustic and genetic analyses, density was the highest for WF (245 ind./ha) and V (76 ind./ha; Table 2). F_1 had the lowest density (2 ind./ha) followed by BCV (12 ind./ha) and BCWF (27 ind./ha). The estimated proportions of pure species, F_1 hybrids and backcrosses, in L. Raha at the time of sampling were 88.7%, 0.5% and 10.8%, respectively. The genetic assignment increased the accuracy of estimated proportions: according to the morphological analysis only, the proportion of hybrids (including backcrosses) was only 3.5%. The difference between results reflects the fact that the abundant putative whitefish and vendace groups actually included 9.6% and 7.2% of backcrosses. The proportion of BCWF was
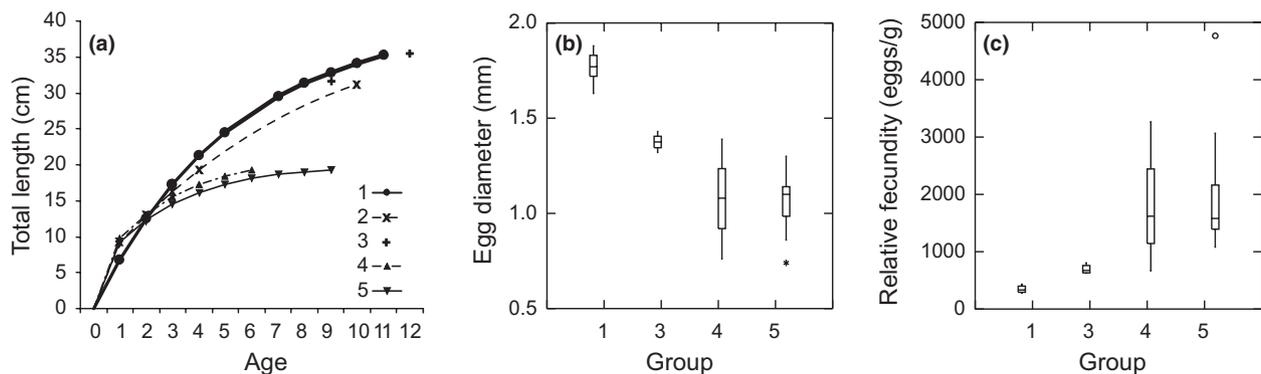
Fig. 4 Life history data used as fitness proxies for different coregonid groups [whitefish (1), backcross to whitefish (2), F_1 hybrids (3), backcross to vendace (4) and vendace (5)]. Growth curves (a) were calculated with the von Bertalanffy growth model, except for F_1 hybrids where observed lengths at age are shown. Egg diameter (b) and relative fecundity (c) are presented with box-whisker plots.

higher than BCV, closely following the abundances of pure species in the lake at the sampling year.

Discussion

The genetic data confirmed the existence of hybridization events between putative 'good biological species', as well as introgression back to both parental species. The occurrence of F_1 hybrids (*c.* 0.5%) and backcrosses (10.8%) in the multiple year classes of coregonids indicated widespread introgression and thus an apparent lack of efficient reproductive isolation barriers between these two well-recognised species. However, the biased direction of gene flow could suggest a lower fitness of hybrid males because of genomic maladaptation or alternatively because of inferiority during mate choice. Despite the fact that the ecomorphology of hybrids and backcrosses was intermediate between their parental species, there was no detectable evidence for reduced fitness in these individuals.

Species introduction and hybridization

The results of the genetic assignment analyses conformed remarkably well to the a priori morphological classifications of vendace and whitefish. However, some of the morphologically classified whitefish and vendace included backcrosses that could not be detected without the use of genetic tools. These individuals had morphologies typical of whitefish or vendace, but genetic results indicated the presence of gene flow from the other species. The morphological hybrid category mainly contained backcrosses, as only four F_1 individuals were confirmed with the genetic analyses. The combined use of mtDNA and microsatellites has proven to be a powerful method not only to confirm or identify hybrids but also to detect introgression in the wild (e.g. Salzburger *et al.* 2002; Senn *et al.* 2010). The use of mtDNA facilitated the identification of parental origin, and we found that F_1 (four individuals) originated from mating between female vendace and male whitefish. Genetic data from backcrosses indicated bidirectional gene flow, but in this case, the main direction of historical gene flow was from vendace to whitefish. Such unidirectional gene flow from introduced to native species or from abundant to rare species is typical (Rhymer & Simberloff 1996; Fitzpatrick *et al.* 2010; Cabria *et al.* 2011). Observed hybrids came from 2 year classes (1994 and 1997), when the population density of vendace was very high, and apparently enhanced hybridization, as detected from gene flow calculations. However, potential hybridization between whitefish and vendace has been possible since autumn 1987, and we can not exclude the possibility that the pure species already interbred during the year of introduction. Some support was obtained for the presence of higher historical gene flow from the introduced vendace to the native whitefish. The first vendace invaders in a system can attain considerably large sizes (>25 cm) via the consumption of superabundant, underutilized pelagic resources (Amundsen *et al.* 1999), that may further increase the likelihood of hybridization with whitefish via size-assortative mating. Apart from one individual that hatched in 1998, backcrosses hatched from the year 2000 onwards. If the initial hybridization between whitefish and vendace occurred already in autumn 1987, the first backcross generation might have already hatched in the early 1990s. In autumn 2005, backcrosses to whitefish were more abundant than backcrosses to vendace, which may reflect the current numerical dominance of whitefish over vendace (WF:V > 3:1). The probability of hybridization and introgression might also be elevated because of habitat modification, as L. Raha was dammed for hydroelectric power production in 1957, and exhibits large seasonal fluctuations in water level. Water level is elevated during the whitefish spawning period in late autumn, but falls through the winter: a practice likely to cause exposure and subsequent freezing of eggs laid on spawning sites located in the shallow littoral zone. Furthermore, water level fluctuations cause reduction of littoral-benthic macroinvertebrates, which may have subsequently forced whitefish to switch to pelagic habitats and associated prey items. The introduction of vendace, a superior pelagic competitor (Svärdson 1976; Bøhn & Amundsen 2001) to the already heterogeneous environmental conditions of L. Raha evidently promoted gene flow with native whitefish. Hybridization and introgression have led to both genetic and morphological shifts in the coregonid populations.

Ecomorphological shifts of hybrids and backcrosses

The intermediate morphological traits of hybrids as detected here between whitefish and vendace are common phenomena in fishes (Pethon 1974; Hatfield & Schluter 1999; Taylor *et al.* 2006). Our results further confirmed an apparently strong inheritance of gill raker traits in hybrids and backcrosses in *Coregonus* (Svärdson 1970; Rogers & Bernatchez 2007; Bittner *et al.* 2010). Gill raker traits are especially important in the evolution of polymorphic northern fish, where an increasing number of long gill rakers are correlated with the utilization of resources from pelagic habitats (Bernatchez *et al.* 1999; Amundsen *et al.* 2004; Bolnick & Lau 2008) and facilitate the consumption of smaller prey items (Kahilainen *et al.* 2007, 2011). However, the intermediate morphology of hybrids and backcrosses was not associated with

any evidence of the exploitation of new niches or any improved foraging ability relative to that of the pure parental species. Indeed, according to short-term (observed diet and habitat use), and long-term (stable isotope values and parasites) indicators of niche use, both hybrids and backcrosses exclusively used pelagic zooplankton prey, as did their parental species. The pelagic zone is the most profitable niche in dammed lakes, in which the strong seasonal water level fluctuations reduce significantly the production of littoral benthic prey (Haxton & Findlay 2008). The overlapping niches of native whitefish and introduced vendace may also have increased the overlap of spawning location and time, with subsequent hybridization. Despite a morphological continuum from whitefish to vendace, there was no evidence of a lowered feeding activity in neither hybrids nor backcrosses according to stomach fullness and growth.

Selection against hybrids and backcrosses according to life history traits

Heterosis is a general phenomenon observed in many freshwater fishes (Hubbs 1955), and we found that backcrosses grew longer and reached larger maximum sizes (von Bertalanffy asymptotic lengths), relative to their parental species. This has been observed in experimental hybrids between whitefish and vendace (Svärdson 1965), but here, we have extended the comparisons to both F_1 hybrids and backcrosses in the field, including the most evident resource competitors; their own parental species. Currently, a low total density of coregonids (362 ind./ha) as compared with other lakes harbouring whitefish and vendace (c. 2650 ind./ha) in this region (Jensen *et al.* 2008) suggests that the level of resource competition might be reduced in L. Raha. At high population densities, vendace is a superior competitor for shared pelagic resources compared with planktivorous whitefish, through more efficient foraging for small zooplankton (Bøhn & Amundsen 1998; Kahilainen *et al.* 2011).

We found high proportion of backcrosses in L. Raha that suggests frequent hybridization events and/or high survival rates of hybrids in 1990s. However, we are lacking data on embryonic mortality rates in hybrids and backcrosses, which may be elevated relative to parental species according to results from experimental studies (see Lu & Bernatchez 1998; Rogers & Bernatchez 2006; Woods *et al.* 2009). Experimental evidence from lake whitefish morphs indicated that genetic incompatibilities between the morphs increased posthatching mortality (Rogers & Bernatchez 2006). The same study also indicated that the severe reduction in survival may arise in backcrosses, because their hatching times differ from the pure morphs. Although we found gene flow

in neutral loci, and patterns of Mendelian heritability of many adaptive morphological traits, there might be underlying physiological traits such as respiration metabolism and growth that could affect negatively the survival of admixture individuals (Renaut *et al.* 2011). There might be also genetic barriers in nonneutral loci that are highly resistant to gene flow (Via & West 2008; White *et al.* 2010). Vendace and whitefish have been differentiated as species for a sufficient time period to accumulate strong genomic barriers in nonneutral loci. However, it is impossible to further evaluate the direction of the gene flow and the effects of selection on such genomic regions with the current data set. It is evident though that there are most likely both intrinsic and extrinsic barriers to gene flow, as year classes of F_1 were not observed annually.

Frequency of hybridization and introgression between species

The occurrence of F_1 hybrids in only 2 year classes suggests that either hybridization is an occasional phenomenon, or that hybrid mortality, because of genetic incompatibilities and selection pressures, is generally greater than that seen in pure species. An interesting contrast was the frequent introgression in both directions (altogether seen in seven different year classes) that also suggests the existence of F_2 hybrids. A lack of F_2 hybrids in our data set may further support the existence of elevated hybrid mortality rates (Svärdson 1965), or may simply reflect a very low natural density of F_2 hybrids in the lake. Alternatively, the number of microsatellite markers used here may limit the probability of detection of F_2 hybrids (Gow *et al.* 2006; Vähä & Primmer 2006; Sanz *et al.* 2009). As our simulation data shows, there were some inaccuracies in the six genetic class assignments in NewHybrids, which may have affected our estimates. This could partly explain the low amount of F_1 and F_2 individuals, as well as a slight overestimation of backcross frequencies. However, in general, the results seems to suggest an underestimation of the degree of hybridization and introgression, and therefore, the estimated frequency of these events should be viewed as conservative. In a follow-up study, next generation sequencing techniques would provide far increased resolution to hybrid detection, as well as our capacity to reveal the genomic level admixture in hybrids and backcrosses (e.g. Bernatchez *et al.* 2010; Wolf *et al.* 2010). The current conditions in L. Raha may be especially favourable for the survival of hybrids and backcrosses; however, this may not be maintained through time. Long-term studies of Darwin's finches have demonstrated that the fitness of hybrids fluctuates temporally (Grant & Grant 2008). Along the evolution-

ary history of coregonids, fusion and fission of gene pools may have been important for their highly successful adaptive radiations throughout Holarctic freshwater lakes (Bittner *et al.* 2010; Hudson *et al.* 2011).

In conclusion, we found strong evidence of hybridization and introgression between two well-recognized 'good biological species', whitefish and vendace, which dominate many lakes across northern Europe. Hybrids and backcrossed individuals had morphologies intermediate between parent species, but utilized the same niches, with no apparent signs of selection against them. Thus, under such labile environmental conditions as documented here, widespread introgression between native whitefish and introduced vendace was evident. Future studies should focus on the generality of hybridization and introgression between these species both in introduced and natural systems. In addition, more detailed approach to the genomic level admixture of hybrids and backcrosses would be needed to deeper understanding of mechanisms maintaining the separation between the pure species.

Acknowledgements

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Amundsen P-A, Staldvik FJ, Reshetnikov YS *et al.* (1999) Invasion of vendace *Coregonus albula* in a subarctic watercourse. *Biological Conservation*, **88**, 405–413.

Amundsen P-A, Knudsen R, Klemetsen A, Kristoffersen R (2004) Resource competition and interactive segregation between sympatric whitefish morphs. *Annales Zoologici Fennici*, **41**, 301–307.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Arnold ML (2006) *Evolution Through Genetic Exchange*. Oxford University Press, Oxford.

Bagenal T (1978) *Methods for Assessment of Fish Production in Fresh Waters. IBP Handbook No 3*. Blackwell Scientific Publications, Oxford.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Bernatchez L (2004) Ecological theory of adaptive radiation: empirical assessment from Coregonine fishes (Salmoniformes). In: *Evolution Illuminated: Salmon and their Relatives* (eds Hendry AP, Stearns SC), pp. 175–207. Oxford University Press, Oxford.

Bernatchez L, Chouinard A, Lu G (1999) Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus* sp., as a case study. *Biological Journal of the Linnean Society*, **68**, 173–194.

Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B*, **365**, 1783–1800.

Bittner D, Excoffier L, Largiadèr CR (2010) Patterns of morphological changes and hybridization between sympatric whitefish morphs (*Coregonus* spp.) in Swiss lake: a role for eutrophication? *Molecular Ecology*, **19**, 2152–2167.

Blair AC, Hufbauer RA (2010) Hybridization and invasion: one of North America's most devastating invasive plants shows evidence for a history of interspecific hybridization. *Evolutionary Applications*, **3**, 40–51.

Bøhn T, Amundsen P-A (1998) Effects of invading vendace (*Coregonus albula* L.) on species composition and body size in two zooplankton communities of the Pasvik River System, northern Norway. *Journal of Plankton Research*, **20**, 243–256.

Bøhn T, Amundsen P-A (2001) The competitive edge of an invading specialist. *Ecology*, **82**, 2150–2163.

Bøhn T, Amundsen P-A (2004) Invasion-mediated changes in the population biology of a dimorphic whitefish *Coregonus lavaretus* population. *Annales Zoologici Fennici*, **41**, 125–136.

Bøhn T, Sandlund OT, Amundsen P-A, Primicerio R (2004) Rapidly changing life history during invasion. *Oikos*, **106**, 138–150.

Bolnick DI, Lau OL (2008) Predictable patterns of disruptive selection in stickleback in postglacial lakes. *American Naturalist*, **172**, 1–11.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques*, **20**, 1004–1010.

Cabria MT, Michaux JR, Gómez-Moliner BJ *et al.* (2011) Bayesian analysis of hybridization and introgression between endangered European mink (*Mustela lutreola*) and polecat (*Mustela putorius*). *Molecular Ecology*, **20**, 1176–1190.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, **15**, 2833–2843.

Crawford SS, Muir AM (2008) Global introductions of salmon and trout in the genus *Oncorhynchus*: 1870-2007. *Reviews in Fish Biology and Fisheries*, **18**, 313–344.

Eloranta AP, Kahilainen KK, Jones RI (2010) Seasonal and ontogenetic shifts in the diet of Arctic charr *Salvelinus alpinus* in a subarctic lake. *Journal of Fish Biology*, **77**, 80–97.

Elphinstone MS, Hinten GN, Anderson MJ, Nock CJ (2003) An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes*, **3**, 317–320.

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.

Fitzpatrick BM, Johnson JR, Kump DK *et al.* (2010) Rapid spread of invasive genes into a threatened native species. *Proceeding of the National Academy of Sciences of the United States of America*, **107**, 3606–3610.

Garside ET, Christie WJ (1962) Experimental hybridization among three Coregonine fishes. *Transactions of American Fisheries Society*, **91**, 196–200.

Gow JL, Peichel CL, Taylor EB (2006) Contrasting hybridization rates between sympatric three-spined stickleback highlight the fragility of reproductive barriers between evolutionary young species. *Molecular Ecology*, **15**, 739–752.

Gow JL, Peichel CL, Taylor EB (2007) Ecological selection against hybrids in natural populations of sympatric threespined sticklebacks. *Journal of Evolutionary Biology*, **20**, 2173–2180.

Grant PR, Grant BR (2008) *How and Why Species Multiply: The Radiation of Darwin's Finches*. Princeton University Press, Princeton.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Harrod C, Griffiths D (2005) *Ichthyocotylurus erraticus* (Digenea: Strigeidae): factors affecting infection intensity and the effects of infection on pollan (*Coregonus autumnalis*), a glacial relict fish. *Parasitology*, **131**, 511–519.

Harrod C, Grey J, McCarthy TK, Morrissey M (2005) Stable isotope analyses provide new insights into ecological plasticity in a mixohaline population of European eel. *Oecologia*, **144**, 673–683.

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

Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866–873.

Haxton TJ, Findlay CS (2008) Meta-analysis of the impacts of water management of aquatic communities. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 437–447.

Hubbs CL (1955) Hybridization between fish species in nature. *Systematic Zoology*, **4**, 1–20.

Hudson AG, Vonlanthen P, Müller R, Seehausen O (2007) Review: The geography of speciation and adaptive radiation of coregonines. *Archiv für Hydrobiologie Special Issues Advances in Limnology*, **60**, 111–146.

Hudson AG, Vonlanthen P, Seehausen O (2011) Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society London B*, **278**, 58–66.

Hynes HBN (1950) The food of freshwater sticklebacks (*Gasterosteus aculeatus* and *Pygosteus pungitus*), with a review of methods used in studies of the food of fishes. *Journal of Animal Ecology*, **19**, 36–58.

Jensen H, Kahilainen KK, Amundsen P-A *et al.* (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, Østbye K (2006) Morphological differentiation and resource polymorphism in three sympatric *Coregonus lavaretus* forms in a subarctic lake. *Journal of Fish Biology*, **68**, 63–79.

Kahilainen K, Lehtonen H, Könönen K (2003) Consequence of habitat segregation to growth rate of two sparsely rakered whitefish forms *Coregonus lavaretus* (L.) in a subarctic lake. *Ecology of Freshwater Fish*, **12**, 275–285.

Kahilainen K, Malinen T, Tuomaala A, Lehtonen H (2004) Diel and seasonal habitat and food segregation of three sympatric *Coregonus lavaretus* forms in a subarctic lake. *Journal of Fish Biology*, **64**, 418–434.

Kahilainen K, Alajärvi E, Lehtonen H (2005) Planktivory and diet-overlap of densely rakered *Coregonus* in a subarctic lake. *Ecology of Freshwater Fish*, **14**, 50–58.

Kahilainen KK, Malinen T, Tuomaala A *et al.* (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 KK, Siwertsson A, Gjelland KØ *et al.* (2011) The role of gill raker number variability in adaptive radiation of coregonid fish. *Evolutionary Ecology*, **25**, 573–588.

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

Kiljunen M, Grey J, Sinisalo T *et al.* (2006) A revised model for lipid-normalising $\delta^{13}C$ values from aquatic organisms, with implications for isotope mixing models. *Journal of Applied Ecology*, **43**, 1213–1222.

Knudsen R, Amundsen P-A, Klemetsen A (2003) Inter- and intra-morph patterns in helminth communities of sympatric whitefish morphs. *Journal Fish Biology*, **62**, 847–859.

Korsu K, Huusko A, Muotka T (2010) Impacts of invasive stream salmonids on native fish: using meta-analysis to summarize four decades of research. *Boreal Environmental Research*, **15**, 491–500.

Lampert W, Sommer U (2007) *Limnoecology*, 2nd edn. Oxford University Press, Oxford.

Lu G, Bernatchez L (1998) Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchill). *Proceeding of the Royal Society London B*, **265**, 1025–1030.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 129–237.

Mayr E (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University Press, New York.

Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, **6**, 971–973.

Østbye K, Bernatchez L, Næsje TF *et al.* (2005) The 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, Amundsen P-A, Bernatchez L *et al.* (2006) Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Molecular Ecology*, **15**, 3983–4001.

Parsons KJ, Albertson RC (2009) Roles for Bmp4 and CaM1 in shaping the jaw: evo-devo and beyond. *Annual Review of Genetics*, **43**, 369–388.

Patton JC, Gallaway BJ, Fechhelm RG, Cronin MA (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.

Pethon P (1974) Naturally occurring hybrids between whitefish (*Coregonus lavaretus* L.) and cisco (*Coregonus albula* L.) in Orrevann. *Norwegian Journal of Zoology*, **22**, 287–293.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rahel FJ (2002) Homogenization of freshwater faunas. *Annual Review of Ecology and Systematics*, **33**, 291–315.

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

Redenbach Z, Taylor EB (2002) Evidence for historical introgression along a contact zone between two species of char (Pisces: Salmonidae) in northwestern North America. *Evolution*, **56**, 1021–1035.

Reist JD (1985) An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. *Canadian Journal of Zoology*, **63**, 1429–1439.

Reist JD (1986) An empirical evaluation of coefficients used in residual and allometric adjustment of size covariation. *Canadian Journal of Zoology*, **64**, 1363–1368.

Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Molecular Ecology*, **20**, 545–559.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rogers SM, Bernatchez L (2006) The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology*, **19**, 1979–1994.

Rogers SM, 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.

Rogers SM, Cambell D, Baird SJE, Danzmann RG, Bernatchez L (2001) Combining the analyses of introgressive hybridization and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchill). *Genetica*, **111**, 25–41.

Rogers SM, Marchand MH, 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.

Rubidge EM, Taylor EB (2004) Hybrid zone structure and the potential role of selection in hybridizing populations of native westslope cutthroat trout (*Oncorhynchus clarki lewisi*) and introduced rainbow trout (*O. mykiss*). *Molecular Ecology*, **13**, 3735–3749.

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

Sanz N, Araguas RM, Fernandez R, Vera M, García-Marín JL (2009) Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. *Conservation Genetics*, **10**, 225–236.

Schoener TW (1970) Non-synchronous spacial overlap of lizards in patchy habitats. *Ecology*, **51**, 408–418.

Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Senn HV, Barton NH, Goodman SJ *et al.* (2010) Investigating temporal changes in hybridization and introgression in a predominantly bimodal hybridizing population of invasive sika (*Cervus nippon*) and native red deer (*C. elaphus*) on the Kintyre Peninsula, Scotland. *Molecular Ecology*, **19**, 910–924.

Shikano T, Shimada Y, Herczeg G, Merilä J (2010) History vs. habitat type: explaining the genetic structure of European nine-spined stickleback (*Pungitius pungitius*) populations. *Molecular Ecology*, **19**, 1147–1161.

Siwertsson A, Knudsen R, Kahilainen KK *et al.* (2010) Sympatric diversification as influenced by ecological opportunity and historical contingency in a young species lineage of whitefish. *Evolutionary Ecology Research*, **12**, 929–947.

Steeves TE, Maloney RF, Hale ML, Tylianakis JM, Gemmel NJ (2010) Genetic analyses reveal hybridization but no hybrid swarm in one of the world's rarest birds. *Molecular Ecology*, **19**, 5090–5100.

Svärdson G (1965) The Coregonid problem. VII. The isolating mechanisms in sympatric species. *Reports of the Institute of Freshwater Research Drottningholm*, **46**, 95–123.

Svärdson G (1970) Significance of introgression in coregonid evolution. In: *Biology of Coregonid Fishes* (eds Lindsey CC, Woods CS), pp. 33–59. University of Manitoba Press, Winnipeg.

Svärdson G (1976) Interspecific population dominance in fish communities of Scandinavian lakes. *Report of the Institute of Freshwater Research, Drottningholm*, **55**, 144–171.

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

Swofford DL (2001) *PAUP*version 4.0b10. Phylogenetic Analysis Using Parsimony (and other methods)*. Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Taylor EB, Boughman JW, Groenenboom M *et al.* (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, **15**, 343–355.

Thorpe RS (1975) Quantitative handling of characters useful in snake systematics with particular reference to intraspecific variation in the Ringed Snake, *Natrix natrix* (L.). *Biological Journal of the Linnean Society*, **7**, 27–43.

Todd TN, Stedman RM (1989) Hybridization of ciscoes (*Coregonus* spp.) in Lake Huron. *Canadian Journal of Zoology*, **67**, 1679–1685.

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.

Vähä J-P, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Von Bertalanffy L (1938) A quantitative theory of organic growth. *Human Biology*, **10**, 181–243.

Wallace RK (1981) An assessment of diet-overlap indexes. *Transactions of the American Fisheries Society*, **110**, 72–76.

White BJ, Cheng C, Simard F, Constantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, **19**, 925–939.

Wilson CC, Bernatchez L (1998) The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology*, **7**, 127–132.

Wilson GA, Rannala B (2003) Bayesian inference on recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wolf JB, Bayer T, Haubold B *et al.* (2010) Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, **19**, 162–175.

Woods PJ, Müller R, Seehausen O (2009) Intergenomic epistasis causes asynchronous hatch times in whitefish hybrids, but only when parental ecotypes differ. *Journal of Evolutionary Biology*, **22**, 2305–2319.

Zalapa JE, Brunet J, Guries RP (2010) The extent of hybridization and its impact on the genetic diversity and population structure of an invasive tree, *Ulmus pumila* (Ulmaceae). *Evolutionary Applications*, **3**, 157–168.

K.K.K. is a fish biologist particularly interested in adaptive evolution and ecological speciation. K.Ø.'s research interest is centred around polymorphic freshwater fish species, speciation, functional genomics and faunal history. C.H. is a fish and isotope ecologist interest in the evolution, function and ecology of aquatic consumers and ecosystems. T.S. is broadly interested in fish population genetics and genomics. T.M.'s research focuses on food-web interactions and hydroacoustic estimation of fish and invertebrate densities. J.M.'s research interest centre around genetic processes in wild vertebrate populations.

Data accessibility

A set of 10 haplotypes (h1–h10) detected in coregonids deposited at Genbank (Accession no.: JN003210–JN003219). Genetic, ecomorphology and life history data deposited at Dryad: doi:10.5061/dryad.83n0t.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 STRUCTURE summary statistics for evaluating the most likely *K* in the real data from L. Raha.

Table S2 The NewHybrid and STRUCTURE results using the real data from L. Raha.

Table S3 The NewHybrid assignment in the 10 runs of the simulated six genetic groups.

Table S4 The assignment distribution in NewHybrids for the miss-assigned individuals (<0.85) in the real data from L. Raha.

Table S5 Allelic variability at seven loci in the studied Coregonid group from L. Raha.

Fig. S1 Map of the study area.

Fig. S2 mtDNA 50% majority rule maximum likelihood consensus tree.

Fig. S3 The STRUCTURE summary results estimating the most likely *K*-value.

Fig. S4 The NewHybrids assignment for the simulated six genetic groups.

Fig. S5 The individual consistency in the assignment for the 10 simulations in the six genetic groups.

Fig. S6 The NewHybrids assignment for the miss-assigned individuals (<0.85) in the real data from L. Raha.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.