MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2727–2742 doi: 10.1111/j.1365-294X.2012.05561.x

Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.)

MAGNUS W. JACOBSEN,*†, MICHAEL M. HANSEN,* LUDOVIC ORLANDO,† DORTE BEKKEVOLD,‡ LOUIS BERNATCHEZ,§ ESKE WILLERSLEV† and M. THOMAS P. GILBERT†

Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark, †Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark, ‡National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, DK-8600 Silkeborg, Denmark, §Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon Charles-Eugène-Marchand, 1030, Avenue de la Médecine, Québec City, Québec, Canada G1V 0A6

Abstract

The advent of second-generation sequencing has made it possible to quickly and economically generate whole mitochondrial genome (mitogenome) sequences. To date, mitogenome studies of nonmodel organisms have demonstrated increased power for resolving interspecies relationships. We explored an alternate use of such data to recover relationships and population history of closely related lineages with a shallow evolutionary history. Using a GS-FLX platform, we sequenced 106 mitogenomes from the *Coregonus lavaretus* (Europe) and *Coregonus clupeaformis* (North America) species complexes to investigate the evolutionary history of the endangered Danish North Sea houting (NSH) and other closely related Danish and Baltic European lake whitefish (ELW). Two well-supported clades were found within both ELW and NSH, probably reflecting historical introgression via Baltic migrants. Although ELW and NSH are not reciprocally monophyletic, they share no haplotypes, suggesting recent, but strong, reproductive isolation. The divergence time between NSH and the geographically closest ELW population was estimated using IMa, assuming isolation with migration and a new mutation rate estimate chosen to avoid time-dependency effects. The estimate of *c.* 2700 BP was remarkably similar to results obtained using microsatellite markers. Within North American *C. clupeaformis*, the divergence time between the two lineages (Atlantic and Acadian) was estimated as between 20 000 and 60 000 BP. Under the assumption that NSH and ELW colonized Denmark following the last glacial maximum, Bayesian Serial SimCoal analysis showed consistency with a scenario of long-term stability, resulting from a rapid initial sixfold population expansion. The findings illustrate the utility of mitogenome data for resolving recent intraspecific divergence events and provide evidence for recent reproductive isolation of the phenotypically divergent NSH.

Keywords: *Coregonus*, demographic history, divergence, mitochondrial genomes, mutation rate, second-generation sequencing

Received 17 October 2011; revision received 21 January 2012; accepted 27 January 2012

Correspondence: Magnus W. Jacobsen, Fax: +45 87150201;
E-mail mwj@biology.au.dk

Introduction

The advent of second-generation sequencing has made it possible to quickly and economically generate whole mitochondrial genome (mitogenome) sequences, creating new possibilities within the field of phylogenetics and population genetics (e.g. Morin *et al.* 2010). Previous mitogenome-based studies in nonmodel organisms have almost exclusively focussed on research questions at the interspecific level using phylogenetic-based analysis (e.g. Doiron *et al.* 2002; Azuma *et al.* 2008; Yasuike *et al.* 2010; Willerslev *et al.* 2009; Inoue *et al.* 2010; Vilstrup *et al.* 2011). Intraspecific mitogenome-based studies investigating population structure and demography are still in their infancy, and there are currently few studies published on extant (e.g. Carr & Marshall 2008; Morin *et al.* 2010) or extinct fauna (e.g. Gilbert *et al.* 2008; Stiller *et al.* 2010). Except in cases of incomplete lineage sorting, both theoretical and empirical analyses have demonstrated that the use of longer mtDNA sequences will generally resolve phylogenies better than those based on shorter sequences (Saitou & Nei 1986; DeFilippis & Moore 2000; Rokas & Carroll 2005), in particular for recently diverged species, where added data can substantially impact the resolution of topology and benefit estimation of the time of divergence. For example, a recent study of killer whales (*Orcinus orca*) showed that mitogenome data have high potential to resolve shallow evolutionary histories on a timescale of 150 000 to 700 000 BP (Morin *et al.* 2010).

Despite its increasing popularity as a molecular marker, one area where mitogenome analysis has yet to be exploited is with regard to the study of recent divergence events, for example, between taxa and lineages diverged in glaciated regions in the time since the last glacial maximum (*c.* <18 000 BP). In this regard, most authors have relied on microsatellite markers due to their fast mutation rates, but the additional variation that mitogenomes can offer could provide important additional insights.

Whitefishes (*Coregonus* spp.) represent a group of organisms with tremendous phenotypic variation and are important models for studying the genetics of speciation (e.g. Pigeon *et al.* 1997; Rogers & Bernatchez 2005, 2007; Østbye *et al.* 2006; Hudson *et al.* 2007; Whiteley *et al.* 2008). Several distinct morphotypes are known to co-occur both allopatrically and sympatrically, usually differentiated by the number of gill rakers, body size, habitat choice and food preferences (Himberg & Lehtonen 1995; Bernatchez *et al.* 1996; Kahilainen *et al.* 2004; Østbye *et al.* 2005a,b, 2006), which has led some authors to suggest that such morphs should be recognized as distinct taxa (e.g. Himberg & Lehtonen 1995; Stott & Todd 2005; Kottelat & Freyhof 2007).

The anadromous North Sea houting (*Coregonus oxyrhinchus*) (abbreviated NSH) is a phenotypically divergent form or species belonging to the European lake whitefish complex (*C. lavaretus*) (abbreviated ELW). It is the only known representative within the ELW complex able to tolerate seawater of oceanic salinities (>33‰) (Grøn 1987). NSH also deviates morphologically from closely related ELW notably with regard to snout length, number of gill rakers, and life-history traits such as mean age at reproduction (Christensen & Hvidt 1990; Hansen *et al.* 2008). Comparative data are, however, scarce, and some authors have proposed that the NSH is not morphologically distinct if comparisons are made with ELW over broader geographical scales (Otterstrøm 1922; Freyhof & Schöter 2005). Today, the NSH is considered highly endangered. Until recently, its range included the Wadden Sea area, a shallow coastal region characterized by huge tidal flats, and its adjoining rivers, from the Netherlands to the west coast of the Jutland Peninsula, Denmark (Jensen *et al.* 2003). Due to pollution, overexploitation and habitat destruction, NSH populations declined drastically during the 20th century, and in the late 1970s, the only self-sustaining NSH population persisted in Vidaa river in southern Denmark (Jensen *et al.* 2003). In 2005, a major re-establishment project was initiated, funded by the EU LIFE programme (http://ec.europa.eu/environment/life/) to restore habitat and re-introduce NSH into its former range in Denmark (http://www.snaebel.dk/English/).

The taxonomic status of the NSH is controversial. Freyhof & Schöter (2005) argued that the NSH was historically only distributed in the Rhine River region and is now extinct, a conclusion that has also been adopted by IUCN (International Union for the Conservation of Nature). These authors instead suggested that the current Vidaa river NSH and ELW of the region belong to the same species, *C. maraena*. However, Freyhof & Schöter's taxonomy was based on two morphological traits, that is, gill raker numbers and snout length, of which at least the former has repeatedly been shown to be highly homoplastic in whitefish (e.g. Bernatchez *et al.* 1996; Østbye *et al.* 2005a, 2006; Etheridge *et al.* 2012) and therefore unsuitable when adopting the phylogenetic species concept (Cracraft 1983). Hence, Freyhof & Schöter's (2005) conclusion must be considered disputable, but on the other side it cannot be ruled out that the now extinct NSH populations of the Wadden Sea region may in fact have been polyphyletic, representing parallel evolution as is often observed in whitefish (Østbye *et al.* 2005a, 2006).

Previous studies that employed both short mtDNA sequences and microsatellite markers suggest close genetic relationships between NSH and ELW in Denmark (Hansen *et al.* 1999, 2008; Østbye *et al.* 2005a).

The results further suggest that they descend from a single ancestral population that colonized Denmark after the last ice age around 13 000 BP (Hansen *et al.* 2008). Within the European whitefish complex, studies on mtDNA have shown the presence of two distinct lineages, presumably representing two glacial refugia: one west of the Ural Mountains and one north of the Alps (Bernatchez & Dodson 1994; Østbye *et al.* 2005a; Hudson *et al.* 2011). These two lineages have since introgressed as seen in many whitefish populations across central and northern Europe (Bernatchez & Dodson 1994; Østbye *et al.* 2005a; Hudson *et al.* 2011) including in Danish ELW and NSH (Østbye *et al.* 2005a). Analyses using microsatellite markers show statistically significant genetic differentiation between Danish NSH and ELW (Hansen *et al.* 1999, Hansen *et al.* 2008), but splitting time estimation using the IMa framework (Hey & Nielsen 2007) suggests very recent divergence between the NSH and the geographically most proximate ELW population in Ringkøbing fjord [2482 BP; 90% highest probability density (HPD) interval, 641–4344 BP] (Hansen *et al.* 2008).

Although other studies on coregonid fishes have demonstrated superior resolution of microsatellite markers (as opposed to mtDNA) for resolving questions involving genetic differentiation (Patton *et al.* 1997; Lu *et al.* 2001), such mtDNA-based studies relied on small fragments of mtDNA (in general ≤1000 bp). We hypothesized that because the complete mitochondrial genome contains more variation, it could potentially be more informative. In this study, 106 mitogenomes were sequenced using second-generation sequencing techniques to refine our understanding of the recent evolutionary history of morphologically and ecologically distinct European whitefish. The individuals represent Danish and Baltic Sea ELW and Danish NSH populations, as well as individuals from both the Acadian and Atlantic evolutionary lineages (Lu *et al.* 2001) of the closely related North American lake whitefish (ALW) (*C. clupeaformis*). In addition to providing a general data set with which to investigate the usefulness of mitogenome data for resolving phylogeographic questions on a recent timescale, the data allowed us to target four specific questions related to the NSH.

Our first research objective was to assess the genetic distinctiveness of NSH, in particular, whether (i) the increased resolution provided by mitogenomes supports the pronounced sharing of haplotypes between ELW and NSH, previously observed at smaller mtDNA segments (Østbye *et al.* 2005a) and (ii) whether mitogenome data suggest a recent splitting time between ELW and NSH, as supported by microsatellite data. Although palaeontological evidence suggests that ELW colonized the Jutland peninsula rapidly, immediately following the last glaciation [13 000 BP, (Aaris-Sørensen 1995)], previous analyses of smaller mtDNA targets show no evidence for a population expansion (Østbye *et al.* 2005a). We therefore (iii) assessed whether signals of an expansion were evident when analysing the demographic histories of populations based on mitogenomes. Finally, previous attempts to date the divergence time between the main clades within ELW have yielded conflicting results, including 360 000 BP, 108 000–929 000 BP and >1 Ma (Bernatchez & Dodson 1994; Winkler *et al.* 2010; Hudson *et al.* 2011). These discrepancies are primarily a result of the conflicting mutation rates used, as no estimates are available for coregonids and few for salmonids due to the lack of good calibration points in the fossil record. Moreover, the dating of recent events (e.g. at the intraspecific level) using mutation rates that are calibrated using much more ancient time points (e.g. at the speciation level) is contentious due to time-dependency effects (Ho *et al.* 2005; Burridge *et al.* 2008). We therefore (iv) attempted to date the divergence events within this complex using previously published as well as newly estimated substitution rates.

Materials and methods

Samples and populations

A total of 106 mitogenomes were sequenced from fish representing 11 populations across Denmark and the Baltic Sea area, along with six individuals of the closely related ALW (*C. clupeaformis*) from Canada, representing the Acadian and Atlantic lineages (Bernatchez *et al.* 1991). One population consisted of 21 NSH, from the Danish Vidaa River, while the remainder consisted of ELW and ALW (Fig. 1, Table 1). All NSH were caught using electrofishing, while ELW and ALW were caught using nets. The European individuals were sampled between 1995 and 1996, and the ALW were sampled between 2007 and 2010. The Danish samples have been used in previously published studies on the population genetics and phylogenetic relationships of North European whitefish, based on microsatellite markers and short mtDNA sequences (Hansen *et al.* 1999; Østbye *et al.* 2005a; Hansen *et al.* 2008). An additional previously published ELW mitogenome sequence representing an individual sampled in the Czech Republic was added to the data set (GenBank accession no. NC_002646.1) (Miya & Nishida 2000).

Mitogenome sequencing

DNA was extracted between 1997 and 2010 from tissue samples using a phenol–chloroform method and kept frozen at −20 °C (Hansen *et al.* 2008). Mitogenomes

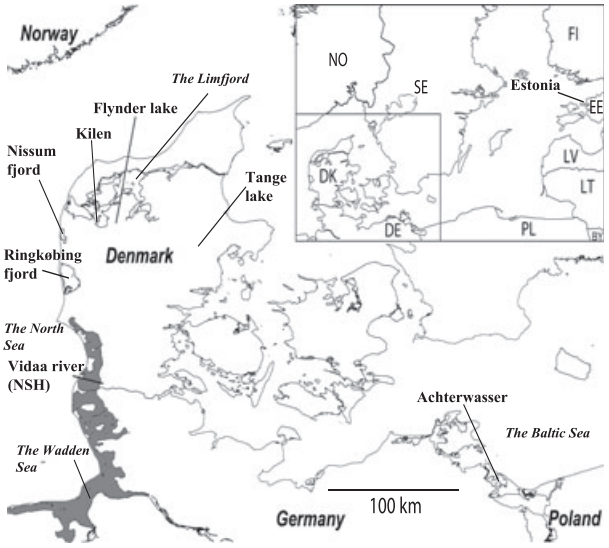

Fig. 1 Map showing the European sample localities. The North Sea houting (NSH) population inhabits the Danish Vidaa river, while the European lake whitefish (ELW) are found in all of the other indicated locations. The Wadden Sea, where the North Sea houting was previously distributed, is indicated by grey.

were subsequently sequenced through a combined long-range PCR amplification and high-throughput DNA sequencing approach on a Roche GS-FLX sequencing platform (Roche, Basel, CH), following Morin *et al.* (2010) and Vilstrup *et al.* (2011). To design species-specific primers, mitogenome sequences from GenBank records of Atlantic salmon (*Salmo salar*) (accession no. U12143.1) (Hurst *et al.* 1999), rainbow trout (*Oncorhynchus mykiss*) (accession no. L29771.1) (Zardoya *et al.* 1995), and a European whitefish (*Coregonus lavaretus*) (Accession no. NC_002646.1) were aligned, and primers were designed to target conserved sites using Primer3 (Rozen & Skaletsky 2000) as implemented in Geneious Pro 4.7.6 (Drummond *et al.* 2009). Two different primer sets, P1 and P2, were designed, each covering just over half of the mitogenome with small overlaps. A third nested primer set, P2Nested, was also designed to increase the specificity of some of the PCR products amplified from some of the samples (see Table S1 and Fig. S1, Supporting information).

Each P1/P2 25 μL reaction mixture contained 1× High Fidelity PCR Buffer, 2 mM $MgSO_4$, 200 nM mixed dNTPs, 400 pM of each primer, 1 μL 1:10 diluted DNA extract and 0.2 μL High Fidelity Platinum Taq (Invitrogen, Carlsbad, CA, USA). PCR amplifications were performed using a Peltier Thermal cycler (DNA Engine, DYAD) with a 2-min activation step at 94 °C, followed by 36 cycles of 94 °C for 2 min, 60 °C for 30 s, 68 °C for 9 min, finished by a final extension step of 72 °C for 7 min. When P2Nested was used, conditions were identical except for the adjustment of $MgSO_4$ concentration to 1.6 mM and annealing temperature to 64 °C.

The concentration of the purified amplicons was measured using a NanoDrop spectrophotometer (NanoDrop Products, Wilmington, DE, USA), and subsequently, the two amplicons from each DNA extract were pooled at equimolar concentration. Pooled samples were converted into MID-tagged GS-FLX libraries following the manufacturer's guidelines with minor modifications. Subsequently, libraries were sequenced in pools of 12 per 1 of 8 of a PicoTitre plate using either LR70 or Titanium-sequencing chemistry (sample-dependent).

Mitogenome assembly

Post sequencing, individual reads were sorted by MID tag, and the tag was removed from both ends (if present) using a custom tag-removal Perl script (Morten

Table 1 Information about the sampled populations, location, year of sampling, species, life history and sample size

Population (location)	Year of sampling	Taxonomy	Geographical region	Life history	Sample size
Vidaa river, Denmark*	1995	*Coregonus oxyrhinchus*	Wadden (North) Sea	Anadromous (high salinity)	21
Ringkøbing fjord, Denmark	1995	*Coregonus lavaretus*	North Sea	Anadromous (brackish)	20
Nissum fjord, Denmark	1995	*C. lavaretus*	North Sea	Anadromous (brackish)	19
Kilen, Denmark	1995	*C. lavaretus*	The Limfjord	Lake	8
Flyder lake, Denmark	1995	*C. lavaretus*	The Limfjord	Lake	8
Tange lake, Denmark	1996	*C. lavaretus*	Kattegat Sea	Lake	8
Achterwasser, Germany	1996	*C. lavaretus*	Baltic Sea	Anadromous (brackish)	8
Baltic sea, Estonia coast	1996	*C. lavaretus*	Baltic Sea	Anadromous (brackish)	8
Czech Republic†	2000	*C. lavaretus*	Czech Republic*	Lake	1
Témiscouata Lake, Canada	2007–2010	*Coregonus clupeaformis*	Canada (Acadian)	Lake (dwarf form)	3
Aylmer lake, Canada	2007–2010	*C. clupeaformis*	Canada (Atlantic)	Lake (normal form)	2
Ross Lake, Canada	2007–2010	*C. clupeaformis*	Canada (Atlantic)	Lake (normal form)	1

*North Sea houting.
†GenBank accession no. NC_002646.1 (Miya & Nishida 2000).

Rasmussen, University of Copenhagen, unpublished). Subsequently, mtDNA genomes were assembled against the GenBank reference sequence (accession no. NC_002646.1) using gsRefMapper (Roche). Assembled sequences were visually inspected using GENIOUS PRO 4.7.6 (Drummond *et al.* 2009) to assess overall mitogenome coverage and quality of SNPs identified. Differences from the reference sequence were called manually, in general, adhering to a majority rule (>50% of reads for any single SNP/insertion/deletion), although with the following modifications. First, obviously duplicated sequences (reads with identical start and end positions) were not counted as independent reads. Second, in cases where conflict occurred at a particular position, where one state was present at the 3′ end of some reads vs. in the middle of other reads, the state in the latter was prioritized to account for the poor quality of 3′ FLX sequences. In total, an average coverage of 30.5 reads per base (range, 0–1396; data not shown) was achieved for the 106 mitogenomes (GenBank accession no. JQ661382–JQ661487). Gaps in the consensus sequence after the initial sequencing were filled either through additional FLX sequencing or using conventional ABI sequencing, using the commercial service provided by Macrogen (Seoul, South Korea). None of the gaps showed any discrepancies in the reference genome. One region (np 16 205–16 216 with reference to *Coregonus laveretus* accession no. NC_002646.1) constituted of a 10–12 base thymine homopolymer stretch, which was difficult to sequence accurately using the GS-FLX. In all samples, this region was disregarded from subsequent analyses.

Data reliability

To test the reliability of the data, a haplotype data set consisting of the congregated 13 coding genes from the 106 sequenced mitogenomes was made in GENIOUS PRO 4.7.6 (Drummond *et al.* 2009). We subsequently used this data set as an independent means to test the quality of the data, by comparing whether mutations observed in single individual fish ('singletons') exhibited a different distribution of synonymous and nonsynonymous substitutions as compared to mutations present in multiple fish ('shared mutations'). The validity of this test is supported by the fact that while sequencing errors should occur in a random manner [thus *c.* 2 of 3 of these changes should lead to nonsynonymous substitutions (e.g. Nei & Gojobori 1986)], the mitochondrial genes in fish are under strong purifying selection (Sun *et al.* 2011). To undertake this test, the observed mutations were therefore divided as nonsynonymous and synonymous, taking into account whether each mutation was observed as a singleton (thus possibly a sequencing error) or in multiple individuals (thus unlikely to be a sequencing error). To investigate whether the substitution pattern in the 'singleton' data set was different from that in the total data set (thus possibly deriving from sequencing error) chi-square tests were used to compare the 'singleton' data with the data of all 'shared' mutations. Furthermore, as the ratio of nonsynonymous substitutions might change through time due to changes in purifying selection (Ho *et al.* 2005), we also investigated this by testing the 'singleton' mutations against the restricted data set of 'shared' mutations in terminal branches, thereby constituting the most recently emerged mutations.

Phylogenetic and population structure analyses

Analysis of pairwise differentiation (Φ_{ST}) was performed between all ELW and NSH populations using Arlequin (version 3.5.2.1, Excoffier & Lischer 2010). The number of haplotypes was calculated for the total data set using DnaSP version 5.1 (Librado & Rozas 2009), and the distribution of haplotypes among populations was recorded. Bayesian phylogenetic analysis was performed on all the mitogenomes ($N = 107$), as well as the haplotypes ($N = 53$), using beast (version 1.5.4, Drummond & Rambaut 2007). Prior to the analysis, jModeltest version 0.1.1 (Posada 2008) was used to estimate the substitution model that best fitted the data using the AIC (Akaike information criterion) and eight gamma categories: TrN + G. Preliminary analyses using an uncorrelated log-normal clock model showed no significant evidence for rate heterogeneity among branches (Drummond *et al.* 2007) and subsequent analyses were made using a strict clock approach. The constant size population model was chosen over exponential growth. Both models gave similar estimates of divergence, but calculation of Bayes Factor using TRACER v1.5 (Drummond & Rambaut 2007) showed slight support for the constant model (BF <2). The times of the most recent common ancestor (MRCA) for major phylogenetic clades were estimated as substitutions/site and later divided by two different mutation rate estimates to produce unscaled time values (as detailed later).

The final MCMC sample was based on a run for 40 000 000 generations, and genealogies were sampled every 4000 generations with 10% discarded as burn-in. Examination of convergence and effective sample size (ESS) values was conducted using TRACER, and all parameters had ESS values >1000. Two additional runs gave similar results. The maximum clade credibility tree with mean heights for branches was estimated in the program TREEANNOTATOR (Drummond *et al.* 2007) with 10% burn-in and visualized and edited in the program FigTree v1.3.1 (Andre Rambaut, University of

Edinburgh, http://tree.bio.ed.ac.uk/software/figtree/). To assess the robustness of the tree topology, a maximum-likelihood analysis was conducted using MEGA 5, using the 53 unique haplotypes (Tamura *et al.* 2011). The analysis was based on the same data and substitution model as implemented in the BEAST analysis, with 1000 replicates to calculate branch support (bootstrap values).

Analyses of demographic history and mutation rate estimation

The Bayesian skyline plot (BSP) was applied to investigate fluctuations in effective population size through time, using BEAST (Drummond *et al.* 2005). Sequences representing the north-eastern phylogeographic lineage [thought to represent historical immigration from the Baltic Sea basin (Østbye *et al.* 2005a) were omitted from the analysis of Danish samples. jModeltest version 0.1.1 (Posada 2008) was used to estimate the substitution model that best fitted the data using the Akaike information criterion (AIC). Examination of convergence and priors on clock model and rate were identical to the phylogenetic analysis. Each MCMC sample was based on a run of 20 000 000 generations for the population data sets and 40 000 000 for the combined data sets of different haplotype-clades. Genealogies were sampled every 1000 generations for the population data set and 2000 for the combined data set, with 10% discarded as burn-in. The scaled values of N_e ($N_e \times \mu$) and time (measured as substitutions per site) were subsequently unscaled by the two different mutation rate estimates used (as detailed later).

Bayes factors were used to test the alternative hypotheses of stable vs. growing population size using TRACER. The model comparison provided no evidence to support a growing population over one of constant sizes (Jeffreys 1961) for any of the populations (all BF <10, see Table S2, Supporting information). This observation is, however, in contrast with palaeontological data that indicate that ELW were among the first fish to colonize Denmark, *c.* 13 000 BP (Aaris-Sørensen 1995) and underwent a rapid expansion either prior to or immediately after this event. We therefore hypothesized that this lack of signal may relate to limitations in the power of such analyses to recover signals of extremely rapid, short-lived growth. Bayesian Serial SimCoal (BSSC) (Excoffier *et al.* 2000) was therefore used to investigate whether the Danish whitefish data (excluding the Baltic migrants) were consistent with a model of rapid expansion at 13 000 BP and to estimate parameters associated with such an event. In brief, the models tested were (i) a constant population of current size and (ii) a population that underwent immediate expansion of between 1 and 1000 times (sampled under a uniform distribution) up to the current population size, commencing post colonization, 13 000 BP (N_e, sampled under a uniform distribution from 500 to 10 000 individuals; see Fig. S6, Supporting information). Different prior ranges were used in preliminary analyses to better define a credible range of values. Priors were wide to encompass all realistic values of the given parameter [N_e: 500–20 000 individuals (here female effective population size); $\mu = 0.1–10 \times 10^{-8}$ substitutions per site per year (sub/site/year); expansion rate = 1–1000X]. Hansen *et al.* (2008) estimated the historical N_e for the Danish NSH and the ELW from Ringkøbing fjord to be 352 (349–3143) and 1816 (1344–6720). The other four Danish populations in this study are assumed to be of equal sizes, thereby yielding a total effective population size from around a few to many thousands.

Mitochondrial substitution rates in salmonid fishes have been estimated to be $0.375–0.55 \times 10^{-8}$ sub/site/-year (Koskinen *et al.* 2002) for the control region, 0.97×10^{-8} sub/site/year (Wilson & Turner 2009) for the ND2 gene and 0.5×10^{-8} sub/site/year (Smith 1992) for the whole mitochondrial genome. However, rates may vary across species (Pulquério & Nichols 2006) and across timescales (Ho *et al.* 2005; Burridge *et al.* 2008), thus a wide prior was chosen.

Final analyses were performed using a normal distribution for the estimated mutation rate (mean = 0.0009, sub per locus per generation; standard deviation = 0.0056, equivalent to 1.537×10^{-8} sub/site/year) and uniform distributions for N_e (500–10 000) and the expansion rate (1–1000X; see Fig. S5, Supporting information). The simulations were performed using 2 000 000 simulations in BSSC, assuming a sequence length of 16 725 sites and a generation time of 3.5 years [to account for the difference between NSH and ELW (Hansen *et al.* 2008)], and a K2 mutation model with a kappa ratio of 4.57. Summary statistics (haplotypic diversity, Hd; Tajima's *D*, *tD*) were calculated using DNAsp 5.0 (Librado & Rozas 2009) and further used for estimating posterior distributions of model parameters (current population effective size, N_e; expansion factor) using Approximate Bayesian Computation. The best 2000 simulations were selected from a local linear weighted regression procedure, and for both models, the posterior probability was estimated using categorical regression following Beaumont (2008).

The Bayesian- and coalescence-based program IMa (Hey & Nielsen 2007) was used to estimate historical (θ_1 and θ_2) and ancestral (θ_A) effective population sizes, migration rates (m_1 and m_2) and splitting times (t) between populations. The method assumes stable population sizes and a model of isolation with gene flow, where a single panmictic population had split into two

at some time in the past, but subsequently remained connected by a degree of gene flow. As all parameters are scaled by mutation rate, two different mutation rate estimates were used to produce unscaled estimation (as detailed in following section), and generation time was set to 3 years for the ELW and 4 years for the NSH (Hansen *et al.* 2008). An approach using a linear heating scheme ($g = 0.01$) with 10 different chains to explore the parameter space was found to give solid estimates of convergence and ESS >100 for the parameters. Priors were set to $\theta_A = \theta_1 = \theta_2 = 40$ and $m_1 = m_2 = 10$ to explore the full range of the parameters. The prior on t was set to 5, corresponding to *c.* 20 000 BP using the BSSC mutation rate estimate of 1.537×10^{-8} sub/site/-year, to encompass the proposed time frame for the whitefish colonization of Demark after last ice age *c.* 13 000 BP (Hansen *et al.* 2006, 2008). Moreover, analysis assuming a higher value of $t = 10$ was also conducted to assess the robustness of the results with respect to this prior. Analyses were run for 40 000 000 iterations, with the first 10% of simulations discarded as burn-in. All runs were based on the HKY substitution model. To estimate the parameters for the whole diploid population, an inheritance scalar of 0.25 was implemented. Three runs were made for each individual data set with a different number of seeds.

Two IMa analyses were conducted: one between the NSH and the closest situated Danish ELW population from Ringkøbing fjord as in the study of Hansen *et al.* (2008) and a second between the NSH and the combined ELW samples from Denmark.

Unscaling divergence times and population sizes

To produce unscaled estimates of divergence times and population sizes, two different mutation rates were used. The first rate of 0.5×10^{-8} sub/site/year was originally estimated using mitochondrial RFLP divergences between different pacific salmonid species (*Onchorhyncus* spp.) calibrated against fossil data (Smith 1992). The second rate was generated in this study using BSSC (Excoffier *et al.* 2000; Anderson *et al.* 2005) and relied on the assumption of a rapid increase in effective population size of the combined Danish whitefish population 13 000 BP as evident from the paleontological record. This approach produced a mutation rate estimate of 0.0009 sub/locus/Ma corresponding to 1.537×10^{-8} sub/site/years.

Results

Data reliability

The coding gene data set contained 313 mutations, of which 42 were categorized as 'singleton' and 271 as 'shared' mutations. A total of 10.3% of the 'shared' mutations were nonsynonymous (Table 2), a significantly lower level than observed among 'singleton' mutations (Table 2). However, restricting the test to include only 'shared' mutations in terminal branches (subject to only short time of purifying selection) yielded nonsignificant differences (Table 2), suggesting that sequencing error has not affected the data.

Table 2 Distribution of substitutions within the 13 coding genes

	Total mutations in the data set	Shared* mutations-total†	Shared mutations-internal branches‡	Shared mutations-terminal branches§	Singleton mutations¶
Nonsynonymous	47	28	20	8	19
Synonymous	266	243	229	14	23
Total mutations	313	271	249	22	42
% Nonsynonymous	15.0%	10.3%	8.0%	36.4%	45.2%
dN/dS**	0.061	0.040	0.030	0.199	0.287
χ^2-test: column vs. unshared mutations††	—	χ^2-value = 34.72 $P < 0.001$	NA	χ^2-value = 0.47 $P > 0.30$	—

*Refers to those mutations observed in multiple individuals, thus unlikely to arise due to sequencing error.
†The total number of mutations that are shared between at least two individuals.
‡The number of mutations that are present in multiple individuals and multiple haplotypes.
§The number of mutations found in multiple individuals but single haplotypes.
¶Refers to mutations only observed in single individuals, thus potentially reflecting sequencing error.
**dN/dS ratio calculated using the mean number of synonymous and nonsynonymous positions in the data (estimated using DnaSP v 5.1).
††Results of chi-square tests comparing the distribution of synonymous and non-synonymous substitutions against the distribution observed for 'singleton' mutations.

Phylogeny and population structure

Fifty-three different haplotypes were observed among the 107 mitogenome sequences. Shared haplotypes were only observed between the Danish ELW populations in Kilen, Tange lake, Ringkøbing fjord and Nissum fjord, while the ELW population in Flynder lake only shared haplotypes with the closely situated population in Kilen. The NSH sequences from the Vidaa river were not shared with any of the other populations (Fig. 2). Furthermore, all haplotypes observed in the two Baltic ELW and the three North American ALW populations were private and not shared among populations (data not shown). Significant Φ_{ST} was found between the Baltic ELW populations when compared to Danish ELW and NSH populations as well as between NSH and all Danish ELW populations except Flynder lake. None of the pairwise comparisons between the Danish ELW were significantly different (Table 3), although statistical power is expected to be compromised by low sample sizes.

Similar phylogenetic topologies were recovered using both maximum-likelihood and Bayesian approaches, with the European samples distant from the samples of the ALW (*C. clupeaformis*) (Figs 3 and S2, Supporting information). Within Europe, the samples segregated into two clades, one principally containing ELW specimens from the Baltic Sea [designated the north-eastern clade after Østbye *et al.* (2005a)] and the other containing only Danish ELW and NSH samples as well as the previously published Czech Republic ELW sample [designated the south-western clade after Østbye *et al.* (2005a)] (Fig. 3). The NSH was not a monophyletic group as both European clades were represented (Fig. 3). Within the south-western clade, the Danish samples were both distinct from the Czech sample and subdivided into two subclades. Both subclades were represented in all populations with the exception of Nissum fjord. All three major clades (the ALW, the south-western clade and the north-eastern clade) showed similar intraclade genetic diversity, illustrated by similar TMRCA (Fig. 3). Using BEAST and the BSSC estimated mutation rate of 1.537×10^{-8} sub/site/years from this study, the mean TMRCA of the three clades was estimated to be *c.* 39 000–64 000 BP and 120 000–197 000 BP using the rate from Smith (1992) of 0.5×10^{-8} sub/site/years (Fig. 3, group 3, 5, 7). Within the ALW, the Acadian lineage samples were monophyletic from the Atlantic lineage, with an estimated TMRCA of *c.* 20 000 BP using the mutation rate of 1.537×10^{-8} sub/site/years and *c.* 61 000 BP using the rate from Smith (1992) (Fig. 3, group 8). Furthermore, using these mutation rates, the TMRCA of all ELW was estimated to be *c.* 159 000 and 490 000 BP, and the TMRCA of ELW and ALW to *c.* 431 000 and 1 326 000 BP (Fig. 3, group 1, 2). Finally, the phylogeny also shows that as with other Danish ELW populations, the NSH population has undergone historical introgression with Baltic migrants, as some haplotypes belong to the north-eastern clade (Fig. 3). The event was estimated to have occurred as recently as between 5400 and 16 600 BP, according to the mutation rate used (Fig. 3, group 6).

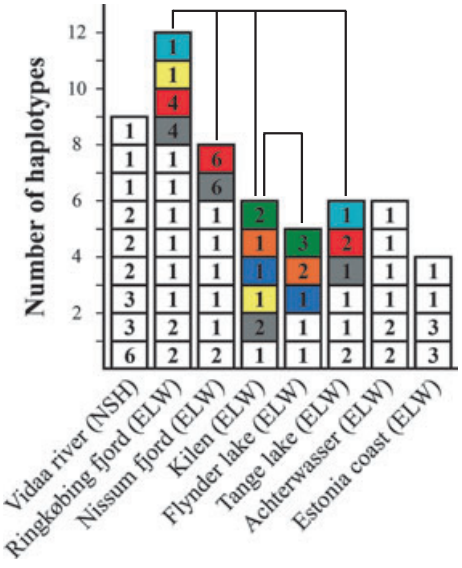

Fig. 2 Histogram showing the haplotype sharing between the sampled European populations. The number of haplotypes within each population is shown on the *y*-axis, with the number of identical haplotypes denoted within the bars. Haplotypes illustrated in white are unique, those in colours are shared between 2 or more populations (specific colours representing identical haplotypes). Lines interconnect populations that share haplotypes.

Demographic history

Bayesian skyline plots did not show any noticeable departures from the constant size model when all sequences from the south-western clade were analysed (Fig. 4A). The inclusion of all sequences from the north-eastern clade showed a small increase in effective population size, with a decrease near the present (Fig. 4B). However, like in all other analyses, Bayes factor tests showed little support in favour of the BSP models (Table S2 and Figs S3 and S4, Supporting information).

Subsequently, BSSC simulations (Anderson *et al.* 2005) were used to further investigate the past demographic history of the combined Danish ELW/NSH population under the assumption that all individuals belonged to a single deme. More specifically, a model of a population

Table 3 ϕ_{ST} between Danish and Baltic European lake whitefish and North Sea houting (NSH) populations

	Vidaa r.	Ringkøbing f.	Nissum f.	Kilen	Flynder l.	Tange l.	Achterwasser	Estonia c.
Vidaa river (NSH)	—	**0.176**	**0.240**	**0.181**	0.067	**0.111**	**0.649**	**0.625**
Ringkøbing fjord	—	—	−0.021	−0.048	0.013	−0.005	**0.830**	**0.806**
Nissum fjord	—	—	—	−0.032	0.097	0.041	**0.904**	**0.862**
Kilen	—	—	—	—	0.003	0.031	**0.904**	**0.862**
Flynder lake	—	—	—	—	—	−0.014	**0.863**	**0.821**
Tange lake	—	—	—	—	—	—	**0.870**	**0.829**
Achterwasser	—	—	—	—	—	—	—	**0.261**
Estonia coast	—	—	—	—	—	—	—	—

Bold values are significant ($P < 0.05$).

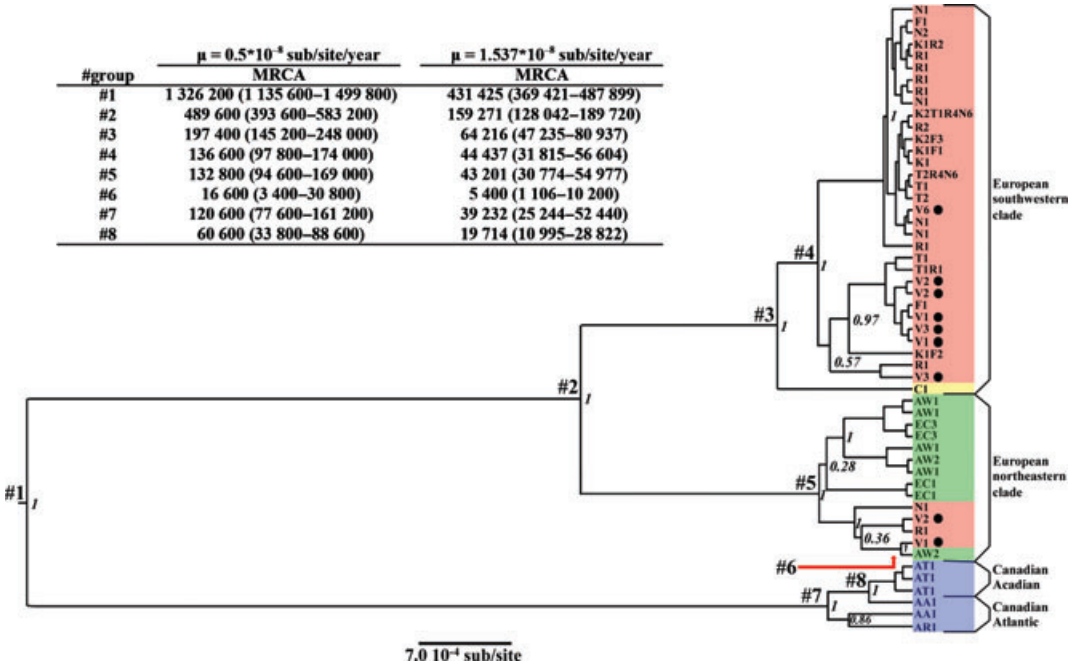

Fig. 3 Bayesian haplotype phylogeny with MCRA estimates for the major clades using all 107 genomes and estimates using the two alternate mutation rates discussed (including error estimates). Clades are numbered on nodes. Posterior support is shown for the main clades (above the branching points). NSH are represented in all European clades and are denoted by black dots. Colour codes represent the region from which the samples have been collected. Red = Denmark, green = the Baltic Sea, yellow = the Czech Republic and blue = Canada. Geographic localities are indicated as R = Ringkøbing fjord, N = Nissum fjord, V = Vidaa river, K = Kilen, F = Flynder lake, T = Tange lake, AW = Achterwasser, EC = Estonia coast, AT = Tésmiscouata lake, AA = Aylmer lake, AR = Ross lake. Values after the letters denote the number of identical haplotypes.

expansion from 1- to 1000-fold at 13 000 BP was compared to a model of constant population size, and key model parameters (N_e and expansion rate) were derived from posterior distributions in an Approximate Bayesian Framework (see Materials and Methods). Categorical regression was used to estimate the posterior probability of both models following Beaumont (2008). The expansion model received highest support (posterior probability = 0.623) compared to the null model (posterior probability = 0.377), suggesting that the genetic diversity present in the current Danish population is compatible with a rapid expansion occurring at the onset of the Holocene (see Figs S5 and S6, Supporting information). According to the maximal posterior probability values, this expansion resulted in a *c.* 6-fold increase in the female effective population size resulting in the current estimate of 8500.

Estimates of historical population size differences and migration were estimated for the data using IMa (Hey & Nielsen 2007), although with large credible intervals

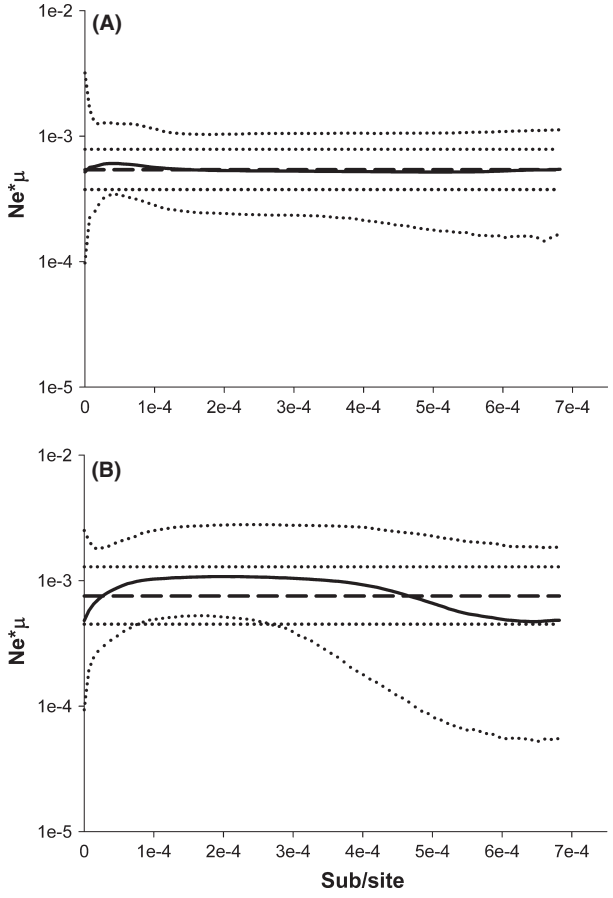

Fig. 4 Bayesian skyline plots (BSPs) of mitogenome data vs. a model of constant size. Axes are scaled by mutation rate. Solid lines represent median inferred BSP values while thick dashed lines represent median values estimated using a model of constant size. The *y*-axis is on logarithmic scale, and dashed lines represent 95% highest probability density (HPD) intervals. Plots show changes in effective population size through time (for unscaled plots, see Fig. S3, Supporting information). (A) BSP made from the combined Danish sequences belonging to the south-western clade and (B) BSP made from the combined sequences belonging to the north-eastern clade.

(Table 4). The t-parameter (splitting time × μ) was more difficult to assess, and 90% HPD intervals could not be identified. However, additional runs yielded highly identical estimates of all parameters also when changing the prior on divergence time (from *c.* 20 000 to 40 000 BP using the BSSC estimated mutation rate), indicating convergence (not shown) (Hey & Nielsen 2007). The analysis of the splitting time between the NSH and the ELW from Ringkøbing fjord showed two distinct peaks in the probability space (not shown). The most recent peak showed the highest probability and was used to estimate divergence. Using the rate of 1.537×10^{-8} sub/site/years, the divergence was estimated to be 2676 BP while the divergence between the NSH and the combined Danish ELW samples was estimated to be 9920 BP (Table 4). Using the alternate lower rate from Smith (1992), deeper divergence times and higher effective population sizes were estimated (Table 4).

The effective population size of the ELW in Ringkøbing fjord was higher than that of the Vidaa river NSH (Table 3). Estimates of the effective number of migrants between the NSH in Vidaa river and Ringkøbing fjord and the combined Danish ELW data set provided values of 0.26–2.4 migrants per generation (Table 4).

Discussion

Phylogeny and population structure

A salient finding of this study concerns the distinct differences in mitogenome composition between NSH and ELW populations. NSH showed in general significant differences in Φ_{ST} from the ELW populations (Table 3), which suggests reproductive isolation of the NSH and supports earlier studies based on microsatellite data (Hansen *et al.* 2008). The only population not showing any significant Φ_{ST} was Flynder lake. This is probably

Table 4 Results of the IMa analyses

Population1	Population2	μ (sub/site/year)	θ_1 (×10^3)	θ_2 (×10^3)	θ_3 (×10^3)	$2N_1m_1$	$2N_2m_2$	*t* (years)
Vidaa (NSH)	Ringkøbing fjord	0.50×10^{-8}	13.1 (2.1–32.5)	38.5 (13.1–88.7)	98.0 (47.7–191.6)	1.67	0.26	8221 (2348–?)
Vidaa (NSH)	Ringkøbing fjord	1.537×10^{-8}	4.3 (0.7–10.6)	12.5 (4.3–28.9)	31.9 (15.5–62.3)	1.67	0.26	2676 (1148–?)
Vidaa (NSH)	Rest DK	0.50×10^{-8}	16.0 (5.5–37.1)	51.3 (31.8–78.1)	100.7 (49.7–23.2)	2.04	2.40	30 493 (4454–?)
Vidaa (NSH)	Rest DK	1.537×10^{-8}	5.2 (1.8–12.1)	16.7 (10.4–25.4)	32.8 (16.2–75.5)	2.04	2.40	9920 (2135–?)

Values are high point density values (HPD) with 90% credible intervals in brackets. For the *t*-parameter 95% credible intervals. Population1 and Population2, the respective populations used in the analysis; μ (sub/site/year), the mutation rate used to unscale results: 0.5×10^{-8} is from Smith (1992), 1.537×10^{-8} is estimated in this study; θ_1, θ_2 and θ_A, the effective population size for the population1, population2 and the ancestral population; $2N_1m_1$ and $2N_2m_2$, the effective number of migrant gene copies per generation into either population1 or population2; *t*, time in years at which the ancestral population split into the two sampled populations. '?' denotes that upper credible intervals could not be defined.

an artefact due to low sample size, as Hansen *et al.* (2008) found it to be genetically distinct from all ELW and NSH populations, except Kilen, using microsatellite markers. Moreover, although NSH was not monophyletic (Fig. 3), it did not share any mitogenome haplotypes with other ELW populations (Fig. 2); however, NSH haplotypes clearly were closely related to those found among ELW (Fig. 3). In this regard, it is noteworthy that previous studies of the same populations that analysed shorter mtDNA segments did not find haplotypes unique to NSH (Hansen *et al.* 1999, Østbye *et al.* 2005a), thus highlighting the additional power that the complete mitogenome sequences confer to the analysis. Although increased sample sizes might recover shared haplotypes, the results provide strong evidence for recent divergence followed by a high degree of reproductive isolation, something further supported by the IMa analyses (Hey & Nielsen 2007) (Table 4), as discussed below. The results are also consistent with a previous study that used STRUCTURE (Pritchard *et al.* 2000) analysis of 12 microsatellite markers (Hansen *et al.* 2008) in the same populations and identified the Vidaa river NSH as a distinct nonadmixed genetic cluster. Moreover, the mitogenome-based observation of haplotype sharing between Ringkøbing fjord, Nissum fjord, Tange lake and Kilen, and between Kilen and Flynder lake (Fig. 2) is also supported by the study of Hansen *et al.* (2008), which showed clustering or high degree of admixture within the exact same populations. Thus, the close agreement between the results obtained using the two different genetic markers illustrates the strength of mitogenome sequencing for analysing population structure within a shallow temporal scale.

The increased power of using complete mitogenomes is also clear from the results of the phylogenetic analyses that support monophyly of the European whitefish clade. While this itself has been noted in previous studies of smaller amounts of sequence data (Bernatchez & Dodson 1994; Østbye *et al.* 2005a; Hudson *et al.* 2011), the resolution recovered here is much higher, and there is maximum support for the monophyly of the main clades (Fig. 3). Furthermore, as also noted previously, European whitefish segregate into two clades: one mainly found in the Baltic Sea area (north-eastern clade) and one found throughout Denmark and including one sample from the Czech Republic (south-western clade) (Fig. 3). In principle, such results could occur due to incomplete random lineage sorting, a possibility that cannot be ruled out when considering only the sampling range of the present study. However, inspection of the distribution of lineages throughout Europe presented by Østbye *et al.* (2005a) clearly suggests a south-western distribution of the one clade, a north-eastern distribution of the second clade and a wide zone of admixture encompassing the Jutland peninsula. Hence, this suggests that the recent European whitefish descend from two main lineages that diverged from each other during the past ice age as seen in other studies with broader geographical sampling (Bernatchez & Dodson 1994; Østbye *et al.* 2005a; Hudson *et al.* 2011).

ELW/NSH mutation rates

The estimated mitogenome mutation rate of 1.537×10^{-8} sub/site/year is substantially higher than that of the 0.5×10^{-8} sub/site/year estimated by Smith (1992), as well as other published rates of salmonid control region and ND2 sequences of $0.375–0.55 \times 10^{-8}$ sub/site/year and 0.97×10^{-8} sub/site/year, respectively (Koskinen *et al.* 2002; Wilson & Turner 2009). Several previous studies have shown evidence for a time-dependency effect on mutation rates, in which rate estimates calibrated on recent divergence events show higher rates than those calibrated on deeper splits (Ho *et al.* 2005; Burridge *et al.* 2008). One explanation underlying this is change in the efficiency of purifying selection (Ho *et al.* 2005), which might need time to remove slightly deleterious mutations. As a result, the observed mutation rate will be higher in studies of recent divergence compared to deeper divergence. However, sequencing errors have also been proposed as a potential explanation for changes in rates (Ho *et al.* 2005), as data of recent divergence will be more affected by these than deeper splits. The data from this study exhibit an increase in dN/dS ratio for mutations present on terminal branches (Table 2) in a manner that cannot be explained through sequencing error. This observation therefore supports an increase in the efficiency of purifying selection over time and indicates that the mutation rate estimated in this study may reflect relaxed purifying selection in the short term, resulting in a higher mutation rate in comparison with those calculated using fossil calibrations.

The estimate of divergence time between the Vidaa river NSH and Ringkøbing ELW of 2676 BP (Table 4), assuming our estimated mutation rate, is remarkably similar to the 2482 BP estimated by Hansen *et al.* (2008) using 12 microsatellite markers. This further validates the new mutation rate estimate from this study over the previous from Smith (1992), at least with regard to analyses of events that happened within the recent past (Ho *et al.* 2005; Pulquério & Nichols 2006). The applicability of this rate to questions relating to deeper timescales is not known, and therefore, does not exclude the possibility that the rate from Smith (1992) might be a more accurate approximation for deeper coalescence events

such as the MRCA of the ALW and the ELW (1 159 260–1 526 140 BP), as well for the MRCA of all European samples (408 380–603 240). The same is possible for the estimates of MRCA of the north-eastern and south-western ELW lineages and the ALW. A further unknown at this point, is whether the close similarity in divergence estimates observed is a result of the methodology applied (including sample size) or an organism-specific feature. The results seem, however, robust even when other priors were used, and we do not see other systematic features that could have affected the analysis.

On a more recent timescale, we estimated that the minimum time of introgression of Baltic migrants into the Danish populations (Fig. 3, group 6) is 5400 (1106–10 200) BP, if the new mutation rate is used. If the uncertainty of using only two sequences to estimate divergence is taken into account, this split might reflect the recent introgression proposed by Østbye *et al.* (2005a) around 8000–12 000 BP. It should be mentioned that although only the three most sampled populations show signs of introgression in this study, Østbye *et al.* (2005a) observed further introgression in the Danish ELW populations of Tange and Flynder lakes that did not show introgression in this study. This supports the notion that the introgression must have happened prior to the split of NSH and ELW and that small samples sizes are responsible for our lack of demonstration of introgression in these same populations. Hudson *et al.* (2011) argued for an even earlier introgression among the two European lineages, as north-eastern haplotypes seem to have been present prior to the recolonization of the Alps *c.* 12 000–20 000 BP. Most of the introgressed Danish samples coalesce prior to this point (Fig. 3), which might reflect such an earlier migration event. Hence, our data do not conclusively support a scenario of early (Hudson *et al.* 2011) or more recent (Østbye *et al.* 2005a) introgression, and in fact several introgression events may have occurred. In the latter case, some of the enormous morphological variation seen between European whitefish populations might be a by-product of differences in introgression history, which would alter the genetic composition and thereby the phenotype, a mechanism proposed more than 40 years ago by Svärdson (1970).

Divergence time and population size estimates

The combined insights gained from the BSSC and BSP analyses indicate that following a rapid initial expansion, the long-term effective population sizes within the Danish NSH and ELW populations have been stable, which renders the data suitable for population size analysis using IMa. In general, we note that the IMa results (Table 4) reflect the prior knowledge on current census population sizes. The population size of Ringkøbing fjord ELW is higher than that of NSH and probably constitutes the largest population of whitefish in Denmark, with an estimated spawning population of up to 40 000 individuals in 1987 (Berg 1987)—considerably higher than the 4000 NSH estimated to spawn in the Vidaa river in 2000 (Jensen *et al.* 2003). Furthermore, although the unscaled estimates of historical effective population size were higher than IMa estimates based on microsatellite markers [Hansen *et al.* 2008; NSH 352 (349–3143), Ringkøbing ELW 1816 (1344–6720)], the 90% credible intervals overlap.

As noted previously, divergence time calculated in this study (2676 BP; Table 4) also corresponds well with divergence time estimated using microsatellite markers [2482 (641–4344 BP); Hansen *et al.* (2008)]. As our divergence estimate is based on an independent genetic marker and uses a different approach to estimate mutation rate, this provides further support for divergence of NSH after the last glaciation, possibly within the last few thousand years. The Wadden Sea area, the environment inhabited by the NSH, is characterized by tidal flats and oceanic salinities, and was formed after the last glaciation (Larsen 2006). This area is extremely dynamic due to sediment inflow from the North Sea, which has changed the physical environment of this region drastically. The geographically most proximate ELW population forages in the Ringkøbing fjord, a shallow brackish lagoon, and spawns in the rivers flowing into the fjord, primarily the Skjern river. The outlet of the Ringkøbing fjord is situated just 45 km from the northernmost range of the Wadden Sea and ~120 km from the outlet of the Vidaa river, but the environmental conditions experienced by the ELW and NSH populations differ considerably. It is therefore likely that past changes in this area, specifically the formation of the Wadden Sea, have lead to the isolation of the NSH from the ELW as seen in this study.

The estimate of divergence time between NSH and the combined Danish ELW populations was higher: 9920 BP (Table 4). Around 7500–9000 BP, the North American Wisconsinan ice-shield melted, leading to drastic changes of sea level and flooding the land west of modern day Denmark (Larsen 2006), possibly isolating some Danish whitefish populations from each other (and the NSH). This event may underlie the estimated divergence time, although we note that pooling all ELW populations represents a simplification of the actual history of populations, as evidenced by the more recent estimate of divergence between the geographically close Ringkøbing fjord ELW and Vidaa NSH populations.

A previous study using mitochondrial RFLP analysis estimated the divergence between ALW from the Acadian and Atlantic mitochondrial lineage to *c.* 150 000 BP and argued that it was caused by isolation during the last glaciation (Bernatchez & Dodson 1990). Our data supports a more recent estimate of divergence between the monophyletic Acadian clade and the closest related Atlantic whitefish sample from Aylmer lake—60 000 and 20 000 BP, depending on the mutation rate used (Fig. 3). Taking the credible intervals into account, these estimates coincide with a warming period around 50 000 BP (followed by a colder period) and the end of last glaciation (Larsen 2006), both events that could potentially account for migration followed by isolation and divergence. The sample sizes are, however, low for the ALW, and in addition, introgression history might bias this finding (Lu *et al.* 2001).

Demographic history

The Bayesian skyline analysis of the combined north-eastern samples showed increase in effective population size around the time of the MRCA (Fig. 4.B). The analysis did not, however, describe the data noticeably better than a model of constant size (Table S2, Supporting information). In this regard, it is interesting to note that using mismatch distribution analysis of 528 bp of combined CYTB and ND3 gene, Østbye *et al.* (2005a) estimated that an expansion had occurred in the north-eastern lineage around 5849–11 698 BP. That analysis relied on many more populations across Europe, with most samples from Finland, Norway and Sweden, potentially rendering a more recent expansion signal evident.

Bayesian skyline analysis of the Danish south-western samples showed no noticeable evidence of change in effective population size (Fig. 4A, Table S2, Supporting information), suggesting that the population has been stable in the long term. However, unless Denmark was initially colonized by a population of similar size to that found today, this cannot be a completely accurate picture of the ELW/NSH history in the region. Given the possibility that a rapid recent expansion may not have produced a signal detectable using BSPs, we used BSSC simulations to investigate whether the genetic diversity present in the current Danish population is compatible with a short-lived rapid expansion shortly following the first opportunity to colonize Denmark following the retreat of the ice sheets *c.* 13 000 BP. In this regard, we find evidence that a rapid expansion of *c.* 6-fold occurring at the onset of the Holocene would be consistent with the data. Thus, our data highlights the power of BSSC as a tool for detecting potential growth events.

Conclusion

This study provides substantial support for the utility of mitogenome data in resolving relationships and divergence times between recently isolated populations, using the European whitefish complex as an example. Overall, the tree topology was much more resolved than in previous studies that relied on smaller fragments of mitochondrial DNA, highlighting the increased power generated by including a larger length of sequence in the analysis. Using a new estimate of the mitogenome mutation rate that was based on a relatively recent geological event, the divergence time between the endangered North Sea houting and the geographically closest Danish whitefish population was estimated to be 2676 BP, almost the exact same time as that estimated in a previous microsatellite-based study. As such, a recent time of this divergence seems extremely well supported. We caution that the applicability of this mutation rate might be limited to recent events due to time-dependency probably caused by relaxed purifying selection in the terminal branches. However, the rate presented here may be useful in other studies of recent divergence within the salmonid fishes where reliable calibration points are scarce.

With regard to the endangered NSH, this study supports previous findings that provide evidence for a recent divergence between the NSH and the Danish ELW but at the same time strong evidence for reproductive isolation, illustrated by the finding that none of the haplotypes found in Vidaa NSH were found in ELW. The species status of NSH remains controversial and depends strongly on the species concept applied (see Hansen *et al.* 2008 for detailed discussion). Regardless of species status, however, NSH has a unique biology among whitefishes, is genetically and ecologically distinct from Danish ELW and should be considered an independent unit for conservation (Hansen *et al.* 2008).

Acknowledgements

The authors acknowledge the Danish Nature Agency (Naturstyrelsen), the EU-LIFE Programme (LIFE05 NAT/DK/000153—Actions for Houting) and the Danish National Research Foundation 'GeoGenetics' award for funding. We also thank Dr Andrew Foote and Dr Rute da Fonseca for helpful discussion, Dr Morten Rasmussen for bioinformatics assistance, BSc. Christoffer Plum for assistance in drawing the sampling map and three anonymous reviewers and the subject editor, Prof. Diethard Tautz for valuable comments on the manuscript.

References

Aaris-Sørensen K (1995) Paleoecology of late Weichselian vertebrate fauna from Nørre Lyngby, Denmark. *Boreas*, **24**, 355–365.

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SIMCOAL: a population genetics model for data from multiple populations and points in time. *Bioinformatics*, **21**, 1733–1734.

Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M (2008) Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evolutionary Biology*, **8**, 215.

Beaumont MA (2008) Joint determination of topology, divergence time and immigration in population tree. In: *Simulation, Genetics and Human Prehistory* (eds Matsumura S, Forster P, Rebfrew C), pp. 135–154. McDonald Institute for Archaeological Research, University of Cambridge, Cambridge.

Berg S (1987) *Heltbestanden i Ringkøbing og Stadil fjorde. Opgangen nov.-dec. 1987*. Report from Ringkøbing Amtskommune, Ringkøbing, Denmark.

Bernatchez L, Colombani F, Dodson JJ (1991) Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. *Journal of Fish Biology*, **39**, 283–290.

Bernatchez L, Dodson JJ (1990) Allopatric origin of sympatric populations of Lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. *Evolution*, **44**, 1263–1271.

Bernatchez L, Dodson JJ (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 JA, Bodaly RA, Dodson JJ (1996) Genetic evidence for reproductive isolation and multiple origins of sympatric trophic ecotypes of whitefish (Coregonus). *Evolution*, **50**, 624–635.

Burridge CP, Craw D, Fletcher D, Waters JM (2008) Geological and molecular rates: fish DNA sheds light on time dependency. *Molecular Biology and Evolution*, **25**, 624–633.

Carr SM, Marshall HD (2008) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gardus morhua*): origins of the ''Codmother,'' transatlantic vicariance and mid-glacial population expansion. *Genetics*, **180**, 381–389.

Christensen IG, Hvidt CB (1990) *Taksonomisk placering af nordsøsnæbel (*Coregonus oxyrhinchus *L.) fra Vidå-systemet i relation til helt (*Coregonus lavaretus *L.) fra Skjern å*. M.Sc. Thesis, Institute of Biology, Department of Ecology and Genetics, University of Aarhus.

Cracraft J (1983) Species concepts and speciation analysis. In: *Current Ornithology* (ed. Johnson RF), pp. 159–187. Plenum Press, New York.

DeFilippis VR, Moore WS (2000) Resolution of phylogenetic relationships among recently evolved species as a function of amount of DNA sequence: an empirical study based on woodpeckers (Aves: Picidae). *Molecular Phylegenetics and Evolution*, **16**, 143–160.

Doiron S, Bernatchez L, Blier P (2002) A comparative mitogenomic analysis of the potential adaptive value of arctic charr mtDNA introgression in brook charr populations. *Molecular Biology and Evolution*, **19**, 1902–1909.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Drummond AJ, Ho SYW, Rawlence N, Rambaut A (2007) *A Rough Guide to BEAST 1.4*. Available at http://beastmcmc.googlecode.com/files/BEAST14jyianual_6July2007.pdf.

Drummond AJ, Ashton B, Cheung M *et al.* (2009) *Geneious v4.8*. Available from http://www.geneious.com.

Etheridge EC, Adams CE, Bean CW *et al.* (2012) Are phenotypic traits useful for differentiating among a priori Coregonus taxa. *Journal of Fish Biology*, **80**, 387–407.

Excoffier L, Lischer HEL (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.

Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–509.

Freyhof J, Schöter C (2005) The houting *Coregonus oxyrinchus* (L.) (Salmoniformes: Coregonidae), a globally extinct species from the North Sea basin. *Journal of Fish Biology*, **67**, 713–729.

Gilbert MTP, Drautz DI, Lesk AM *et al.* (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Science, USA*, **105**, 8327–8332.

Grøn P (1987) Saving the North Sea houting. In: *Proceedings of the 5th International Wadden Sea Symposium* (eds Tougaard S, Asbirk S), Esbjerg, Denmark.

Hansen MM, Mensberg K-LD, Berg S (1999) Postglacial recolonization patterns and genetic relationships among whitefish (*Coregonus* sp.) populations in Denmark, inferred from mitochondrial DNA and microsatellite markers. *Molecular Ecology*, **8**, 239–252.

Hansen MM, Nielsen EE, Mensberg K-LD (2006) Underwater but not out of sight: genetic monitoring of effective population size in the endangered North Sea houting (*Coregonus oxyrhynchus*). *Canadian Journal of Fisheries and Aquatic Science*, **63**, 780–787.

Hansen MM, Fraser DJ, Als TD, Mensberg K-LD (2008) Reproductive isolation, evolutionary distinctiveness and setting conservation priorities: the case of European lake whitefish and the endangered North Sea houting (*Coregonus* sp.). *BMC Evolutionary Biology*, **8**, 137.

Hey S, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *PNAS*, **104**, 2785–2790.

Himberg MK-J, Lehtonen H (1995) Systematics and nomenclature of coregonid fishes, particularly in Northwest Europe. *Archiw für Hydrobiologie*, **46**, 39–47.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Hudson AG, Vonlanthen P, Mueller R, Seehausen O (2007) Review: the geography of speciation and adaptive radiation in coregonines. In: *Biology and Management of Coregonid Fishes*

– *2005*, vol. 60 (eds Jankun M, Brzuzan P, Hliawa P, Luczynski M), pp. 111–146. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, Germany.

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

Hurst CD, Barlett SE, Davidson WS, Bruce IJ (1999) The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. *Genetics*, **239**, 237–242.

Inoue JG, Miya M, Miller MJ *et al*. (2010) Deep-ocean origin of the freshwater eels. *Biology Letters*, **6**, 363–366.

Jeffreys H (1961) *Theory of Probability*, 3rd edn. Oxford University Press, Oxford.

Jensen AR, Nielsen HT, Ejbye-Ernst M (2003) *National Forvaltningsplan for Snæbel*. Report from Miljøministeriet, Skov- og Naturstyrelsen, Sønderlands Amt og Ribe Amt.

Kahilainen K, Malinen T, Toumaala 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**, 1–17.

Koskinen MT, Knizhin I, Primmer CR, Schlötterer C, Weiss S (2002) Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of fresh water, Lake Baikal. *Molecular Ecology*, **11**, 2599–2611.

Kottelat M, Freyhof J (2007) *Handbook of European Freshwater Fishes*. Kottelat, Cornol, Switzerland. Freyhof, Berlin, Germany.

Larsen G (2006) *Naturen i Danmark – Geologien*. Gyldendal, Copenhagen, Denmark.

Librado P, Rozas J (2009) DnaSP v5 software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, **10**, 965–985.

Miya M, Nishida M (2000) Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Molecular Phylogenetics and Evolution*, **17**, 437–455.

Morin PA, Archer FI, Foote AD *et al*. (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research*, **20**, 908–916.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

Østbye K, Bernatchez L, Næsje TF, Himberg JM, 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, Næsje TF, Bernatchez L, Sandlund OT, 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 *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.

Otterstrøm CV (1922) *Heltling (*Coregonus albula *L.) og helt (*Coregonus lavaretus *L.) i Danmark*. In: Undersøgelser af de ferske vandes fiskeriforhold II, 1–50.

Patton JC, Gallaway BJ, Fechhelm RG, Cronin MA (1997) Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (*Coregonus nasus*) in Colville and Sagavanirktok rivers in northern Canada. *Canadian Journal of Fisheries and Aquatic Science*, **54**, 1548–1556.

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

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

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

Pulquério MJF, Nichols RA (2006) Dates from the molecular clock: how wrong can we be? *Trends in Ecology and Evolution*, **22**, 180–184.

Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.

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.

Rokas A, Carroll S (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy *Molecular Biology and Evolution*, **22**, 1337–1344.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.

Saitou N, Nei M (1986) The number of nucleotides required to determine the branching order of three species with special reference to the human-chimpanzee-gorilla divergence. *Journal of Molecular Evolution*, **24**, 189–204.

Smith GR (1992) Introgression in fishes: significance for paleontology, cladistics and evolutionary rates. *Systematic Biology*, **41**, 41–57.

Stiller M, Baryshnikov G, Bocherens H *et al*. (2010) Whitening away – 25 000 years of genetic decline preceded cave bear extinction. *Molecular Biology and Evolution*, **27**, 975–978.

Stott W, Todd TN (2005) Genetic markers and the coregonid problem. *Advanced Limnology*, **60**, 3–23.

Sun YB, Shen YY, Irwin DM, Zhang YP (2011) Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution. *Molecular Biology and Evolution*, **28**, 39–44.

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, Manitoba.

Tamura K, Peterson N, Stecher G, Nei M, Kumar S (2011) *MEGA5*: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **24**, 1596–1599.

Vilstrup J, Ho SYW, Foote AD *et al.* (2011) Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evolutionary Biology*, **11**, 65.

Whiteley AR, Derome N, Rogers SM *et al.* (2008) The phenomics and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive divergence in lake whitefish species pairs (*Coregonus* sp.). *Genetics*, **180**, 147–164.

Willerslev E, Gilbert T, Binladen J *et al.* (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evolutionary Biology*, **9**, 95.

Wilson WD, Turner TF (2009) Phylogenetic analysis of the pacific cutthroat trout (*Oncorhynchus clarki* ssp.: Salmonidae) based on partial mtDNA ND4 sequences: A closer look at the highly fragmented inland species. *Molecular Phylogenetics and Evolution*, **52**, 406–415.

Winkler KA, Pamminger-Lahnsteiner B, Wanzenböck J, Weiss S (2010) Hybridization and restricted gene flow between native and introduced stocks of Alpine whitefish (*Coregonus* sp.) across multiple environments. *Molecular Ecology*, **20**, 456–472.

Yasuike M, Jantzen S, Cooper GA *et al.* (2010) Grayling (*Thymallinae*) phylogeny within salmonids: complete mitochondrial DNA sequences of *Thymallus arcticus* and *Thymallus thymallus*. *Journal of Fish Biology*, **76**, 395–400.

Zardoya R, Garrido-Pertierre A, Bautista JM (1995) The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, Oncorhynchus mykiss. *Journal of Molecular Evolution*, **41**, 942–951.

The research interests of M.W.J., M.M.H. and D.B. involve analyzing population structure and local adaptation using population genetics and genomics approaches, focusing particularly on freshwater and marine fishes. M.T.P.G., L.O. and E.W. have research interests related to the application of high throughput DNA sequencing to questions of evolutionary and anthropological interest. L.B. research focuses on understanding the patterns and processes of molecular and organismal evolution and their significance to conservation.

Data accessibility

Generated mitogenome sequences: GenBank accessions nos JQ661382–JQ661487. Sequence alignment uploaded as online supplemental material along with explanatory readme file.

Supporting information

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

Table S1 Primer information.

Table S2 Comparison of models in BEAST.

Fig. S1 The ELW reference sequence from GenBank (accession no. NC_002646.1) (Miya & Nishida 2000) showing the position of the ribosomal genes, protein coding genes, control region and the positions of the primers used for the long-range PCRs.

Fig. S2 Maximum likelihood phylogeny made in MEGA v5 with 1000 bootstrap indicated above branches.

Fig. S3 Bayesian skyline plots (BSPs) of mitogenome data vs. a model of constant size.

Fig. S4 Bayesian skyline plots (BSPs) of mitogenome data vs. a model of constant size.

Fig. S5 Overview of parameters and model settings for the final simulation.

Fig. S6 Distributions of model parameters (current population effective size, N_e; expansion factor) using Approximate Bayesian Computation for the final simulations.

Please note: Wiley-Blackwell is 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.