MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 5187–5202 doi: 10.1111/mec.13827

Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences

SHENGLIN LIU,[1] MICHAEL M. HANSEN[1] and MAGNUS W. JACOBSEN
Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark

Abstract

We analysed 81 whole genome sequences of threespine sticklebacks from Pacific North America, Greenland and Northern Europe, representing 16 populations. Principal component analysis of nuclear SNPs grouped populations according to geographical location, with Pacific populations being more divergent from each other relative to European and Greenlandic populations. Analysis of mitogenome sequences showed Northern European populations to represent a single phylogeographical lineage, whereas Greenlandic and particularly Pacific populations showed admixture between lineages. We estimated demographic history using a genomewide coalescence with recombination approach. The Pacific populations showed gradual population expansion starting >100 Kya, possibly reflecting persistence in cryptic refuges near the present distributional range, although we do not rule out possible influence of ancient admixture. Sharp population declines ca. 14–15 Kya were suggested to reflect founding of freshwater populations by marine ancestors. In Greenland and Northern Europe, demographic expansion started ca. 20–25 Kya coinciding with the end of the Last Glacial Maximum. In both regions, marine and freshwater populations started to show different demographic trajectories ca. 8–9 Kya, suggesting that this was the time of recolonization. In Northern Europe, this estimate was surprisingly late, but found support in subfossil evidence for presence of several freshwater fish species but not sticklebacks 12 Kya. The results demonstrate distinctly different demographic histories across geographical regions with potential consequences for adaptive processes. They also provide empirical support for previous assumptions about freshwater populations being founded independently from large, coherent marine populations, a key element in the Transporter Hypothesis invoked to explain the widespread occurrence of parallel evolution across freshwater stickleback populations.

Keywords: demographic history, pairwise sequentially Markovian coalescent analysis, phylogeographical lineage, postglacial recolonization, threespine stickleback, whole genome sequencing

Received 20 May 2016; revision received 18 August 2016; accepted 22 August 2016

Introduction

Current populations of living organisms are the results of demographic histories that reach back in time, through ancestors within populations, founders of the populations derived from other populations and ultimately encompassing the whole species. Reconstruction of such demographic histories allows for assessing the effects of past biotic and abiotic factors on population sizes, such as glaciations and postglacial expansions and recolonizations, and may ultimately contribute to understanding the forces shaping current genetic composition and standing variation that ongoing and future evolution depends on (Avise 2000; Knowles 2009).

Correspondence: Michael M. Hansen, Fax: +45 87150201;
E-mail: mmh@bios.au.dk
[1]Joint first authors.

Developments in next-generation sequencing have made it possible to analyse genomewide variation (Davey *et al.* 2011; Ellegren 2014), and this has yielded unprecedented possibilities for reconstructing the demographic history of populations. In particular, new methods have been developed based on models of genomewide coalescence with recombination (McVean & Cardin 2005) that allow for tracking effective population size across different time intervals (Li & Durbin 2011; Sheehan *et al.* 2013; Schiffels & Durbin 2014). Among these, PSMC (Pairwise Sequentially Markovian Coalescent) (Li & Durbin 2011) is at present the most applied method. It estimates effective population size over time based on the genome sequence of a single individual and has been used to reconstruct demographic histories in organisms covering a broad phylogenetic range from vertebrates to arthropods (Li & Durbin 2011; Green *et al.* 2014; Moura *et al.* 2014; Prufer *et al.* 2014; Wallberg *et al.* 2014; Nadachowska-Brzyska *et al.* 2015, 2016; Palkopoulou *et al.* 2015; Xue *et al.* 2015).

Here we focus on the threespine stickleback (*Gasterosteus aculeatus*), a fish species that is widely distributed across the Northern Hemisphere and has emerged as an important model for adaptive divergence and speciation in evolutionary biology (Bell & Foster 1994; McKinnon & Rundle 2002; Colosimo *et al.* 2005; Jones *et al.* 2012; Roesti *et al.* 2014). One of its prominent features consists in the propensity for forming local morphologically and genetically divergent populations (Hendry *et al.* 2002; McKinnon & Rundle 2002; Jones *et al.* 2012; Deagle *et al.* 2013). Particularly, in the case of marine and freshwater populations, important morphological differences exist that are replicated across different populations, with freshwater populations mostly exhibiting reduced armour as compared to marine sticklebacks. This parallel evolution at the phenotypic level is also reflected at the genomic level (Hohenlohe *et al.* 2010; Jones *et al.* 2012; Deagle *et al.* 2013; Terekhanova *et al.* 2014), although exceptions to these patterns exist (DeFaveri *et al.* 2011; Raeymaekers *et al.* 2014; Ferchaud & Hansen 2016).

The colonization dynamics of freshwater populations from marine ancestral populations have been explained by the Transporter Hypothesis (Schluter & Conte 2009). It assumes that freshwater populations have been continuously and repeatedly founded by marine populations with high levels of standing genetic variation, including rare alleles subject to positive selection in freshwater environments. In turn, gene flow from established freshwater populations resupplies marine populations with freshwater-adapted alleles that are subsequently available for founding of new freshwater populations. In previously glaciated regions, freshwater stickleback populations are assumed to have been founded as the ice receded (Schluter & Conte 2009). However, analysis of newly founded stickleback populations in freshwater habitats that have only recently become available due to, for example, earth quakes or extinction of the original populations shows that colonization and adaptation to freshwater conditions is an ongoing process not restricted to initial postglacial recolonization (Bell *et al.* 2004; Lescak *et al.* 2015a).

Tremendous progress has been made towards understanding interactions between variation at the genomic and phenotypic levels and adaptation to environmental conditions (Colosimo *et al.* 2005; Chan *et al.* 2010; Jones *et al.* 2012; Arnegard *et al.* 2014), but comparatively less is known about the phylogeography and particularly the demographic history of stickleback populations. This is an important gap in our knowledge, as selection and adaptation to a significant extent depends on standing genetic variation that is again the result of a long demographic history. Hence, it has for a long time been known that glaciations and postglacial expansions have been associated with repeated bottlenecks, depending on the specific dynamics of the species in question (Hewitt 1996, 2000; Bernatchez & Wilson 1998). Whereas this on one side would be expected to reduce standing variation, it could on the other side also increase quantitative genetic variation due to changed epistatic interactions (Hewitt 2000). In any case, however, the consequences would be assumed to be different across geographical regions and phylogeographical lineages depending on the specific demographic and phylogeographic histories; for example, persistence in refugia, recolonization from nearby refugia, or recolonization over long distances mediated by a small number of founders. In this context, threespine sticklebacks are widely geographically distributed across the Northern Hemisphere, but it is not known to which extent populations from, for example, Europe and Pacific North America show comparable demographic histories. The Transporter Hypothesis is also an important aspect of stickleback evolutionary biology and our general understanding of the dynamics of parallel evolution and gene reuse (Schluter & Conte 2009; Conte *et al.* 2012). However, it is not known if its assumptions regarding demographic history are fulfilled, particularly the assumption of large, continuous marine populations from which freshwater populations have been founded.

Previous phylogeographical studies of threespine sticklebacks, based on mitochondrial DNA analysis, revealed the presence of two distinct lineages, a 'Japanese' and a 'Euro-American' lineage, with both lineages found in populations in the North American Pacific region from Alaska to British Columbia (Deagle *et al.* 1996; Lescak *et al.* 2015b), but with only the latter

lineage found in Europe (Orti *et al.* 1994). Analysis of nuclear single nucleotide polymorphisms (SNPs) have further supported this large-scale pattern (Colosimo *et al.* 2005). In addition, the divergence between Pacific and Atlantic sticklebacks belonging to the same major lineage has been dated to ca. 90–260 000 years bp (Orti *et al.* 1994). Mäkinen & Merilä (2008) extended the resolution of lineages in the Atlantic range and identified a 'Trans-Atlantic' lineage distributed along the North American East Coast and parts of Europe, a widespread 'European' lineage and an isolated 'Black Sea' lineage, all of which showed evidence for postglacial demographic expansion.

In this study, we generated 15 whole genome sequences and retrieved an additional 66 previously published genome sequences from the European Nucleotide Archive (ENA) (Feulner *et al.* 2013, 2015). The samples represented different geographical regions: the temperate to Subarctic Pacific North American region, the Greenland West Coast region, which represents an Arctic environment, and Denmark, Northern Germany and Norway, a temperate region in Northern Europe (see Fig. 1 and Table 1). A total of three marine and 13 freshwater populations were included.

We analysed nuclear SNPs to assess the genetic relationships among populations, analysed mitogenome sequences to identify the phylogeographical lineages represented by the populations, and we used PSMC analysis of whole genome sequences (Li & Durbin 2011) for reconstructing demographic histories of populations. We asked the following questions: (i) Does PSMC analysis suggest similar or different demographic histories across different geographical regions and climates? Specifically, the Pacific Northwest has a complex glaciation history and several cryptic refugia may have existed close to the analysed populations (Shafer *et al.* 2010), potentially resulting in relatively stable population sizes over time. In contrast, Greenlandic and Northern European populations are supposed to have persisted in Southern refugia including the Mediterranean region (Mäkinen & Merilä 2008), although there are also indications of cryptic refugia closer to the ice margin (Ravinet *et al.* 2014). We therefore expected Greenlandic and Northern European populations to show distinct signals of postglacial recolonization involving rapid population expansion. (ii) Do marine and freshwater populations show different demographic histories? Marine and freshwater populations from the same regions should initially show the same patterns of postglacial expansion, but we would expect population declines of freshwater populations after founding and subsequent reproductive isolation from marine sticklebacks. Conversely, we would expect consistently high effective population size in marine sticklebacks if they form large, coherent populations. We additionally assessed how the choice of mutation rate may affect the results, and we discussed the possible impact of population subdivision on the PSMC results (Mazet *et al.* 2016).

Materials and methods

Populations and whole genome sequences

A total of 81 threespine stickleback genomes were included in the study, representing 16 populations (Table 1, Fig. 1). Among these, 15 individuals were sequenced as part of this study, representing five populations, each with a sample size of three. Three of the populations were from the Nuuk Fjord in Greenland, representing one marine (Kob_M) and two freshwater lake populations (Kob_L and Qar_L). The other two populations were sampled in Denmark, Europe and included one marine population (Ran_M) and one freshwater lake population (Hal_L).

Fig. 1 Map showing the location of the population analysed in the study. See Table 1 for a list of sample location abbreviations.

Table 1 Location of studied populations along with their abbreviations, environment types, sample size, nucleotide diversity and the source of the whole genome sequences

Population abbreviation	Location	Environment	Geographical coordinates	Sample size	Nucleotide diversity	Source of data
Ca_R	Misty Lake, inlet, Vancouver Island, Canada	River	50°36′10.40″N, 127°15′8.30″W	6	0.00179	Feulner *et al.* (2015)
Ca_L	Misty Lake, Vancouver Island, Canada	Lake	50°36′16.40″N, 127°15′42.30″W	6	0.00248	Feulner *et al.* (2015)
Us_R	Little Meadow Creek, Alaska, USA	River	61°34′8.76″N, 149°45′36.00″W	6	0.00273	Feulner *et al.* (2015)
Us_L	Long Lake, Alaska, USA	Lake	61°34′33.96″N, 149°46′25.50″W	6	0.00274	Feulner *et al.* (2015)
Kob_M	Kobbefjord, Nuuk Fjord, Greenland	Marine	64° 8′16.77″N, 51°23′27.52″W	3	0.00190	This study
Kob_L	Kobbefjord Lake, Nuuk Fjord, Greenland	Lake	64° 7′51.77″N, 51°22′18.18″W	3	0.00177	This study
Qar_L	Qarajat, Nuuk Fjord, Greenland	Lake	63°59′27.60″N, 51°26′45.60″W	3	0.00174	This study
No_R	Orraelva, Fusa, Norway	River	60°15′19.33″N, 5°55′35.68″E	6	0.00112	Feulner *et al.* (2015)
No_L	Skogseidvatnet, Fusa, Norway	Lake	60°14′41.57″N, 5°54′55.39″E	6	0.00183	Feulner *et al.* (2015)
Ran_M	Udbyhøj, Randers Fjord, Denmark	Marine	56°36′24.07″N, 10°18′7.27″E	3	0.00199	This study
Lm_M	Lemvig, Limfjord, Denmark	Marine	56°36′9.19″N, 8°18′1.94″E	6	0.00192	Feulner *et al.* (2013)
Hal_L	Lake Hald, Denmark	Lake	56°22′25.59″N, 9°20′47.10″E	3	0.00164	This study
G1_R	Malenter Au, Schleswig-Holstein, Germany	River	54°12′15.08″N, 10°33′41.90″E	6	0.00147	Feulner *et al.* (2015)
G1_L	Grosser Plönsee, Schleswig-Holstein, Germany	Lake	54°09′21.61″N, 10°25′48.52″E	6	0.00165	Feulner *et al.* (2015)
G2_R	Eider, Schleswig-Holstein, Germany	River	54°09′58.07″N, 10°04′31.05″E	6	0.00158	Feulner *et al.* (2015)
G2_L	Westensee, Schleswig-Holstein, Germany	Lake	54°16′39.70″N, 9°55′41.04″E	6	0.00192	Feulner *et al.* (2015)

Sequencing was outsourced to Beijing Genomics Institute, BGI (Hong Kong, China; all samples from Denmark) and AROS Applied Biotechnology (Aarhus, Denmark; all samples from Greenland). Paired-end Illumina sequencing with read lengths of 100 bp and average insert size of 400–500 bp was conducted using the Illumina HiSeq 2000 (BGI) and HiSeq 2500 (AROS) platforms, and we aimed for sequencing depths of ca. 20×.

The genome sequences of the remaining 66 individuals were derived from previous publications (Feulner *et al.* 2013, 2015) and retrieved from the European Nucleotide Archive (ENA) (Accession no. ERP004574). These samples covered a wide geographical range from the North American Pacific coast (Vancouver Island and Alaska) to Northern Europe (Norway, Denmark, and Germany) (Table 1, Fig. 1). A total of six geographical localities were represented, with one marine population (Lm_M) and five freshwater locations, each of which included two sympatric populations of lake and river sticklebacks, respectively (see Table 1 for details).

Mapping of sequence reads, summary statistics and principal component analysis

The sequence reads were mapped to the threespine stickleback reference genome (Jones *et al.* 2012). However, data from different sources were treated slightly differently. For the reads from AROS, after quality assessment with FASTQC v0.11.3 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/), the last five base pairs of each read were trimmed off with FASTX v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) due to high 'N' content. The reads were subsequently mapped to the reference genome using the 'aln' and 'sampe' subfunctions of BWA-0.7.12 (Li & Durbin 2009), allowing a maximum of two mismatches. The SAM format alignment file was filtered for a minimum mapping

quality of 20, and was sorted and converted into a BAM file. As the reads from BGI were of high sequencing quality (phred scores >20 over the entire length of the reads), they were directly mapped to the reference genome, and subsequent procedures were similar as for the reads from AROS.

For the reads retrieved from ENA, each individual had been sequenced in multiple rounds, some with paired-end sequencing and some also with mate-pair sequencing. For each subset of reads for an individual, the reads were mapped directly to the reference genome as described for the data from BGI. Finally, BAM files belonging to the same individual were merged with the 'merge' subfunction of SAMTOOLS-1.2 (Li *et al.* 2009).

To examine the genetic relationships among populations, a principal component analysis (PCA) was conducted based on SNPs within individuals. First, a VCF file containing all the variant sites were called from the BAM files created above using SAMTOOLS-1.2 and BCFTOOLS-1.2 (Li *et al.* 2009). Only the 20 autosomal chromosomes were used; that is, chromosome XIX and smaller scaffolds and contigs were excluded. Then, filtration using VCFutils.pl (Li *et al.* 2009) and VCFTOOLS v0.1.12b (Danecek *et al.* 2011) removed all indels and only biallelic SNPs were retained. SNPs with extreme depths (lower than 400 or higher than 5600 for all individuals in total, determined based on the distribution of depths across SNP), low mapping quality (<20) or missing values were excluded as well. A total of 8 570 690 SNPs were obtained, of which a random subset of 85 698 SNPs were used for PCA. This analysis was conducted with the R package 'ade4' (Dray & Dufour 2007). Finally, nucleotide diversity was estimated using VCFTOOLS (Danecek *et al.* 2011).

Pairwise sequentially Markovian coalescent analysis

For each individual, a 'psmcfa' file was created from the BAM file following the instructions in the GitHub page of PSMC (https://github.com/lh3/psmc). Again, only the 20 autosomal chromosomes were used. The bin size was set to 100 bp, and the parameters for running PSMC were the same as the default in the GitHub page (*psmc -N25 -t15 -r5 -p '4 + 25*2 + 4+6'*). Confidence intervals were estimated using 100 bootstraps, and we show results both with and without bootstrapping.

Effective population size (N_e) and time estimates in PSMC have to be calibrated by generation time and mutation rate (μ). Generation length in Fennoscandian populations range from 2 to 4 years (DeFaveri & Merila 2013), and in Southern Greenland, it is ca. 2 years (Bergersen 1996). We settled for a generation length of 2 years, but note that it could differ across populations and over long timescales.

Mutation rate is a critical issue, as the specific choice of rate significantly affects the interpretation of results. We initially considered two approaches. First, we assumed $\mu = 6.6 \times 10^{-8}$, based on mutations detected in a pedigree of Midas cichlid fishes (*Amphilopus* spp.) (Recknagel *et al.* 2013). Second, we assumed $\mu = 1.42 \times 10^{-8}$ per generation, based on transcriptome sequencing of ninespine stickleback (*Pungitius pungitius*), synonymous substitutions between nine- and threespine stickleback and assuming a divergence time of 13 Mya between species (Guo *et al.* 2013). Purifying selection may act over longer timescales to remove slightly deleterious mutations, leading to time dependency (Ho *et al.* 2005). In that case, the first estimate of μ could represent an overestimate, and the second could represent an underestimate relatively to the timescale expected to be targeted by PSMC (a few thousands to several hundreds of thousands years back in time). We therefore undertook a third approach and calibrated μ from our data from Greenland. Analysis of sediment layers from Johs Iversen Lake in the Nuuk Fjord region has documented first occurrence of threespine sticklebacks ca. 9 Kya (Bennike 1997). We therefore made the assumption that the point in time when parapatric marine and freshwater sticklebacks show divergent demographic trajectories in PSMC plots marks the time of founding of the freshwater populations, and we roughly estimated mutation rate by assuming that this split occurred 9 Kya.

Mapping and analysis of mitogenomes

To assess the phylogenetic relationships among individuals and compare them with previous phylogeographic studies (Orti *et al.* 1994; Mäkinen & Merilä 2008), we constructed phylogenetic trees based on their mitogenomes. The mitogenomes can be obtained from the mapping results above. However, sequences obtained in this way often contain large segments of 'N' due to the existence of nuclear mitochondrial DNA sequence (Li *et al.* 2012). To obtain the complete sequence of the mitogenomes, we therefore chose to conduct de novo assembly.

For each of the two 'fastq' files per individual (due to paired-end sequencing), we used only the first 500 000 reads, which is equivalent to decreasing the sequencing coverage. This will fragmentize the assembly of nuclear DNA, whereas the assembly for the mitogenome will not be influenced due to much higher number of copies per cell relative to nuclear DNA. De novo assembly was conducted using SPADES v3.6.2 (Nurk *et al.* 2013), which is efficient in assembling relatively short sequences such as mitogenomes. From the assembly,

the mitogenome sequence was captured using a probe sequence ('ACTTACACATGCAAGTATCCGC'). The probe belongs to a conserved the region in the mitogenome which was identified by aligning mitogenome sequences of six stickleback species in GenBank (*Apeltes quadracus* AB445126, *Culaea inconstans* AB445125, *Gasterosteus aculeatus* AP002944, *Gasterosteus wheatlandi* AB445129, *Pungitius pungitius* AB445130, *Pungitius sinensis* EU332748).

The generated mitogenomes were aligned and manually proofed with MEGA6 (Tamura *et al.* 2013). The control region was removed due to high occurrence of indels. The mitogenome of *G. wheatlandi* (GenBank Accession no. AB445129) was added as an outgroup. We chose to use the Maximum Parsimony algorithm in MEGA6 to build the mitogenome phylogeny due to high sequence similarity among most individuals. The results were further corroborated by 1) constructing a Maximum Likelihood tree using PHYML 3.0 (Guindon *et al.* 2010), applying the GTR model for substitution, allowing variable proportions of invariable sites and mutation rates across sites (GTR + I + gamma) and using 100 bootstraps; and by 2) constructing a Bayesian tree using MRBAYES 3.2.6 (Ronquist *et al.* 2012) and the same GTR + I + gamma model, running 500 000 to 1 000 000 generations until the standard deviation of split frequencies was below 0.05.

Results

Mapping of sequence reads, summary statistics and principal component analysis

The 15 new whole genome sequences generated in the course of this study showed sequencing depths from 17.84× to 25.13×, and the 66 genome sequences included from ENA showed depths ranging from 10.28× to 33.99× (see Table S1, Supporting information for mapping summary statistics). Nucleotide diversity ranged from 0.00112 to 0.00274 across populations, with Pacific North American populations showing generally more variation (range 0.00179–0.00274) than Atlantic populations (range 0.00112–0.00199).

The principal component analysis (PCA) based on SNP variation within individuals showed that all populations could be separated along the first 20 axes (Fig. 2a; the eigenvalues of axes are shown in Fig. S1, Supporting information). Specifically, axis 1 separated Pacific and Atlantic sticklebacks, whereas axis 2 separated populations from Vancouver Island (CA) and Alaska (US) from each other, and also provided separation between individuals from the river and lake populations from Vancouver Island (CA_R and CA_L) but not from Alaska (US_L and US_R) (Fig. 2a,b). Axes 4 and 6 separated most, but not all Atlantic populations (Fig. 2a,c).

Demographic histories inferred by PSMC analysis

If we assumed that the time of founding of freshwater populations in the Nuuk Fjord region, Greenland, took place 9 Kya (Bennike 1997) and that this corresponds to the time when marine and freshwater populations diverged in terms of following different demographic trajectories in PSMC plots, then this would correspond to a mutation rate per generation, μ of ca. 3.7×10^{-8}. This rate is intermediate relative to $\mu = 6.6 \times 10^{-8}$ based on mutations detected in a pedigree (Recknagel *et al.* 2013) and $\mu = 1.42 \times 10^{-8}$ based on a phylogenetic approach (Guo *et al.* 2013). We proceeded with the analyses assuming this mutation rate, but also considered if $\mu = 6.6 \times 10^{-8}$, $\mu = 1.42 \times 10^{-8}$ or an even lower mutation rate could be compatible with the inferred demographic histories.

Demographic histories inferred by PSMC encompassing all individuals from each population are shown in Fig. S2 (Supporting information). Data for 16 individuals downloaded from ENA showed highly deviating patterns in PSMC plots compared to the other individuals in the same populations. Some of these showed a conspicuously shallow demographic history, and on closer inspection they were characterized by low sequence coverage (see Table S1, Supporting information). Others showed ancient asymptotically increasing N_e trajectories (see Us_L, No_R, G1-R, G1_L, G2_R, G2_L in Fig. S2, Supporting information). We suspect that these represent admixed individuals, as an asymptotic increase of N_e is exactly expected in individuals exhibiting different sets of chromosomes from two different populations (Li & Durbin 2011). In fact, PSMC analysis of F1 hybrids, either naturally occurring or synthetic hybrids, obtained by merging sets of phased chromosomes from individuals from two different populations can be used to estimate divergence time, which corresponds to the time of onset of the increase of N_e (Li & Durbin 2009). In this study, however, we could not with certainty identify the populations involved in admixture and ascertain that individuals were F1 hybrids, and hence, we did not consider this information for estimating divergence time. We chose to remove the 16 'outlier' individuals and not consider them for the interpretation of results. PSMC plots of the remaining 65 individuals in each population are shown in Fig. 3, and all demographic histories are superimposed in a single figure in Fig. S3 (Supporting information).

Assuming a generation length of 2 years and mutation rate of 3.7×10^{-8}, the demographic histories

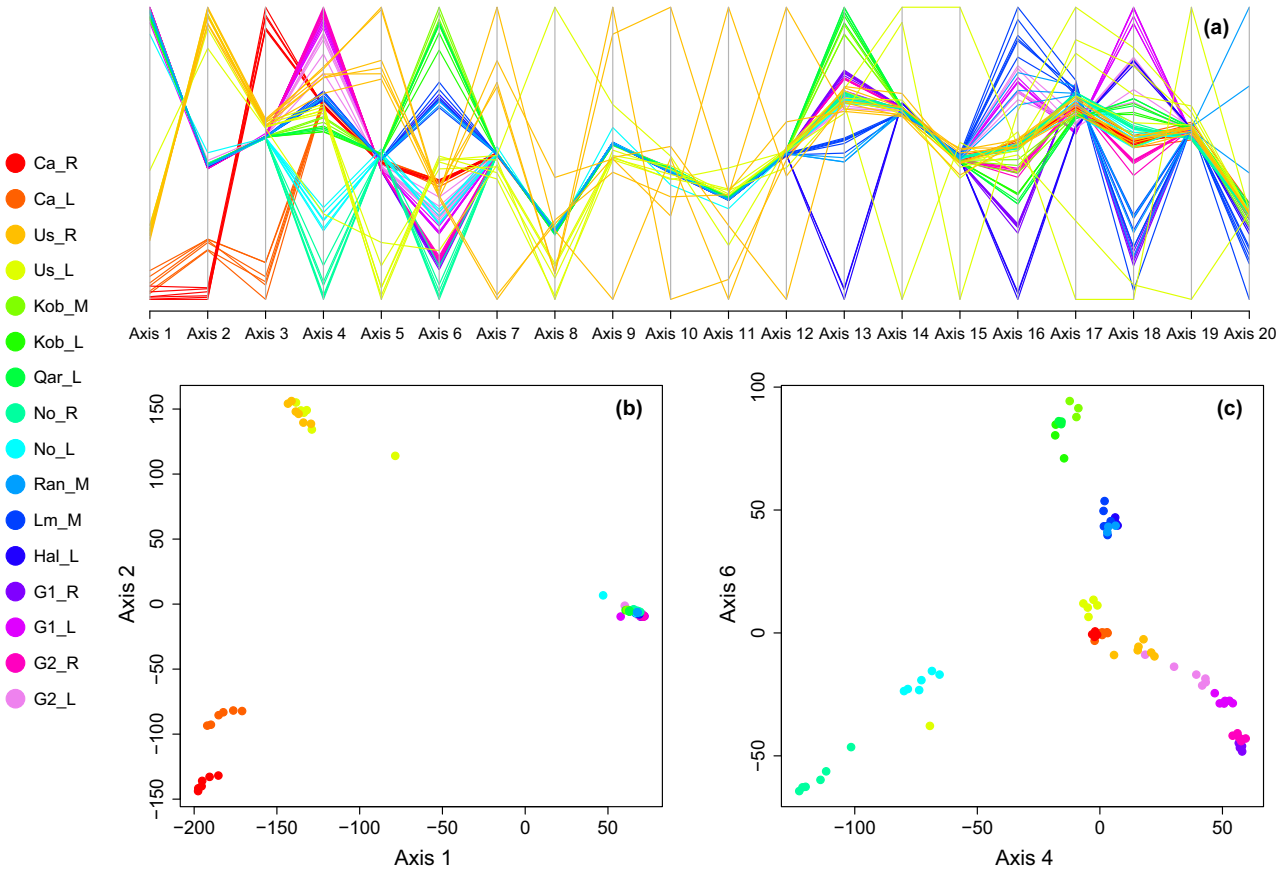

Fig. 2 Principal component analysis of all individuals based on 85698 SNPs. (a) Parallel coordinates plot of the first 20 axes (principal components). (b) Two-dimensional plot of axis 1 and axis 2. (c) Two-dimensional plot of axis 4 and axis 6.

encompassed a time period from ca. 400–500 Kya to ca. 1–2 Kya. All North American sticklebacks showed similar demographic histories (Fig. 3; Fig. S3, Supporting information). Effective population size (N_e) was initially ca. 10 000, but 150–100 Kya a gradual expansion started. This continued and accelerated until ca. 14–15 Kya, when N_e peaked at ca. 25 000–30 000 and subsequently declined. The decline ended ca. 3–5 Kya and N_e stabilized at values of ca. 1000–5000. The latter N_e values varied among populations, with US_R and US_L showing the highest and CA_L and CA_R showing the lowest values.

There was high similarity of demographic histories of sticklebacks from Northern Europe and Greenland (Fig. 3; Fig. S3, Supporting information). In all cases, N_e was initially around 10 000, but ca. 20–25 Kya a pronounced population expansion occurred, which accelerated and culminated ca. 8–9 Kya at N_e values of ca. 20 000–25 000. Following this, marine populations were stable (Lm_M and Kob_M) or expanded further (Ran_M), whereas the freshwater populations declined over a period of ca. 5000 years, after which N_e remained at a few thousands. Similar to Pacific sticklebacks the most recent N_e values showed differences among populations (e.g. 4000–5000 in Hal_L and ca. 1000 in G1_R; see Fig. 3). There were no major differences in demographic histories between parapatric lake and river populations, neither in the Pacific nor Northern European regions.

In summary, we interpreted the PSMC plots to show the following phases: (1) a stable phase from ca. 400–500 Kya; (2) expansion from ca. 150–100 Kya to ca. 14–15 Kya in Pacific populations and from ca. 20–25 Kya to 8–9 Kya in Greenlandic and Northern European sticklebacks; (3) continued stability or expansion of Atlantic marine populations, but decline of freshwater populations from ca. 14 Kya to 3–5 Kya in the Pacific region and from ca. 8–9 Kya to 4 Kya in freshwater populations from Greenland and Northern Europe; (4) stability of freshwater populations towards the present. We assume that phases (1) and (2) occurred in ancestral marine populations and that the declines in freshwater populations in phase (3) marked the founding of the populations from marine ancestors. The three main types of demographic histories exhibited by the populations; Pacific freshwater, Atlantic marine and Atlantic

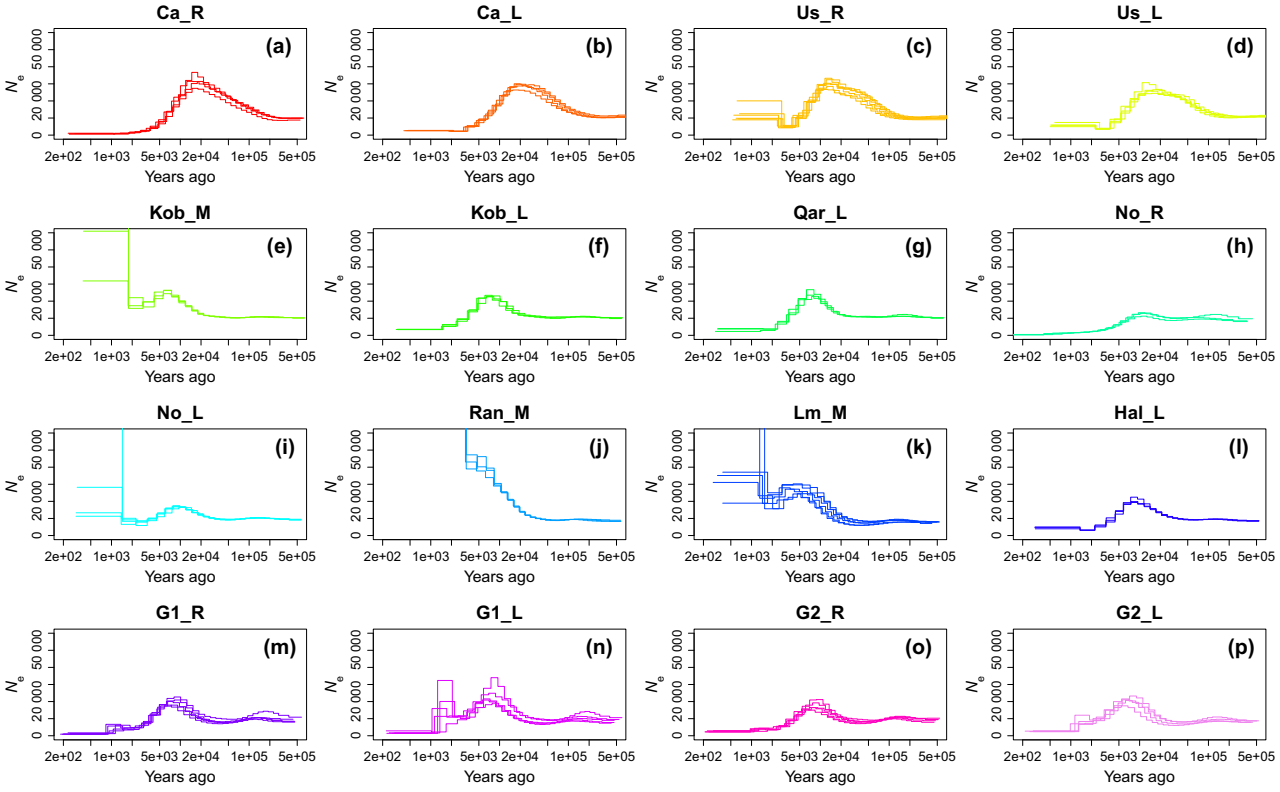

Fig. 3 Results of PSMC analyses of all individuals in all populations, shown for each population separately. Demographic histories are shown as effective population size (N_e) plotted against time.

freshwater are shown in Fig. 4, each represented by PSMC analysis of a single individual and with bootstrapped confidence intervals.

If we instead assume $\mu = 1.42 \times 10^{-8}$, then this would correspond to expansion starting ca. 500 Kya in Pacific and ca. 40 Kya in Atlantic sticklebacks and founding of freshwater populations to have occurred ca. 40 Kya in the Pacific and ca. 20 Kya in the Atlantic basin (see Fig. S4, Supporting information). Assuming $\mu = 6.6 \times 10^{-8}$ expansion should have started ca. 80 Kya in Pacific and 10 Kya in Atlantic sticklebacks, and freshwater populations should have been founded ca. 9–10 and 4–5 Kya in the Pacific and Atlantic basins, respectively (see Fig. S5, Supporting information).

Phylogeographical lineages identified from mitogenome sequences

The maximum parsimony tree generated from mitogenome sequences showed the presence of 4 distinct lineages (Fig. 5), and the Maximum Likelihood and Bayesian trees based on the same data set corroborated the results (see Figs S6 and S7, Supporting information). Two distinct lineages (A and B) were found in North American Pacific sticklebacks. By analysing only the part of the Cytochrome b gene that has previously been studied by Orti *et al.* (1994), it turned out that the lineage found in most individuals (B) corresponded to the 'Euro-American' lineage (a composite lineage which also encompassed lineages from the Atlantic Basin that Mäkinen & Merilä (2008) later showed to be distinct), whereas the five individuals in lineage A found in the Pacific North American populations represented the 'Japanese' lineage (Fig. S8, Supporting information). Hence, these results are in accordance with previous studies suggesting ancient admixture (Orti *et al.* 1994; Lescak *et al.* 2015b). All Northern European and all freshwater sticklebacks from Greenland belonged to a third lineage (D), whereas a fourth lineage (C) was represented by 2 of 3 marine sticklebacks from Greenland (Fig. 5). By comparison with another set of published Cytochrome b sequences (Mäkinen & Merilä 2008) the most numerous lineage (D) corresponded to the 'European' lineage, whereas the two sticklebacks from Greenland (C) belonged to the 'Trans-Atlantic' lineage (see Fig. S9, Supporting information). Hence, there was evidence of either admixture or incomplete lineage sorting in Greenland.

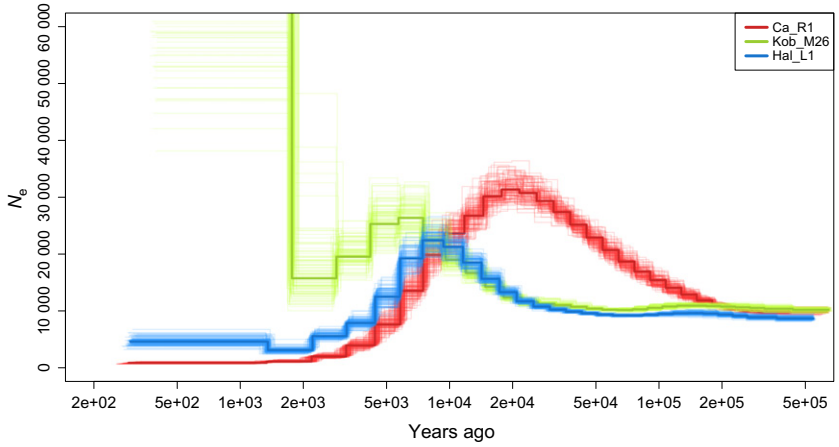

Fig. 4 PSMC plots illustrating the three main categories of demographic histories observed in the study: Pacific freshwater (represented by an individual from CA_R), Atlantic marine (represented by and individual from Kob_M) and Atlantic freshwater (represented by an individual from Hal_L). Thick lines represent point estimates, whereas thin lines represent 100 bootstraps.

Discussion

The results of our study provide novel insights into the demographic history of threespine stickleback populations. First, all populations showed evidence for postglacial expansions, but with important differences among geographical regions. Second, marine and freshwater populations from the same regions showed similar demographic histories during the last glaciation and the initial postglacial expansion, but highly different trajectories of effective population size towards the later postglacial period. We discuss these issues in more detail in the following, but first assess the reliability of the results, particularly considering the evidence for ancient admixture in some of the regions, along with general aspects of interpreting demographic history based on PSMC.

Reliability of results and interpretation of demographic histories

Despite the overall usefulness of PSMC and related methods, they nevertheless have some potential shortcomings that need to be addressed for our discussion of results. This particularly involves the choice of mutation rate and how this affects interpretation of results, and the possibility of complex population histories, such as population subdivision and gene flow not accounted for in PSMC (Mazet *et al.* 2016).

The applied estimate of $\mu = 3.7 \times 10^{-8}$ is moderately higher than genomewide mutation rate estimates in humans ($1.25–2.5 \times 10^{-8}$) (Nachman & Crowell 2000; Altshuler *et al.* 2010) and a mammal like panda (*Ailuropoda melanoleuca*) (1.29×10^{-8}) (Zhao *et al.* 2013), but much higher than estimates from, for example, crocodiles (7.9×10^{-9}) (Green *et al.* 2014) and birds (1.4×10^{-9}) (Ellegren 2007). These results may certainly represent genuine differences among taxa, but can we rule out that our assumed μ represents a severe overestimate, leading to entirely erroneous inference of the timing of demographic history events?

In the case of stickleback, a comparison of demographic histories of parapatric marine and freshwater populations could resolve whether a mutation rate $>10^{-8}$ or closer to 10^{-9} is the most realistic. There is strong evidence suggesting that freshwater populations have been founded repeatedly and independently from marine populations and that there is consequently not a dichotomy between marine and freshwater sticklebacks (Colosimo *et al.* 2005). The inferred demographic histories of marine and freshwater sticklebacks from both Greenland and Northern Europe followed similar trajectories until ca. 8–9 Kya (assuming $\mu = 3.7 \times 10^{-8}$), when freshwater populations declined, whereas marine populations either were stable or expanded. If μ was instead ca. 10^{-9}, then this would shift the major part of the timescale back from the Weichselian/Wisconsinan Glaciation (ca. 115–12 Kya) to the Saalian Glaciation (ca. 300–130 Kya), and the timing of population expansion would rather coincide with the Eemian Interglacial (ca. 130–115 Kya) than the recent postglacial time period (<12 Kya). In this case, however, marine and freshwater sticklebacks should exhibit different demographic histories more than 100 Kya back in time in regions that were covered by ice during the Weichselian/Wisconsinan Glaciation and have subsequently been recolonized afterwards. This would again imply that a major dichotomy should exist between marine and freshwater sticklebacks, which is clearly not the case (Colosimo *et al.* 2005). In total, we therefore find that the applied

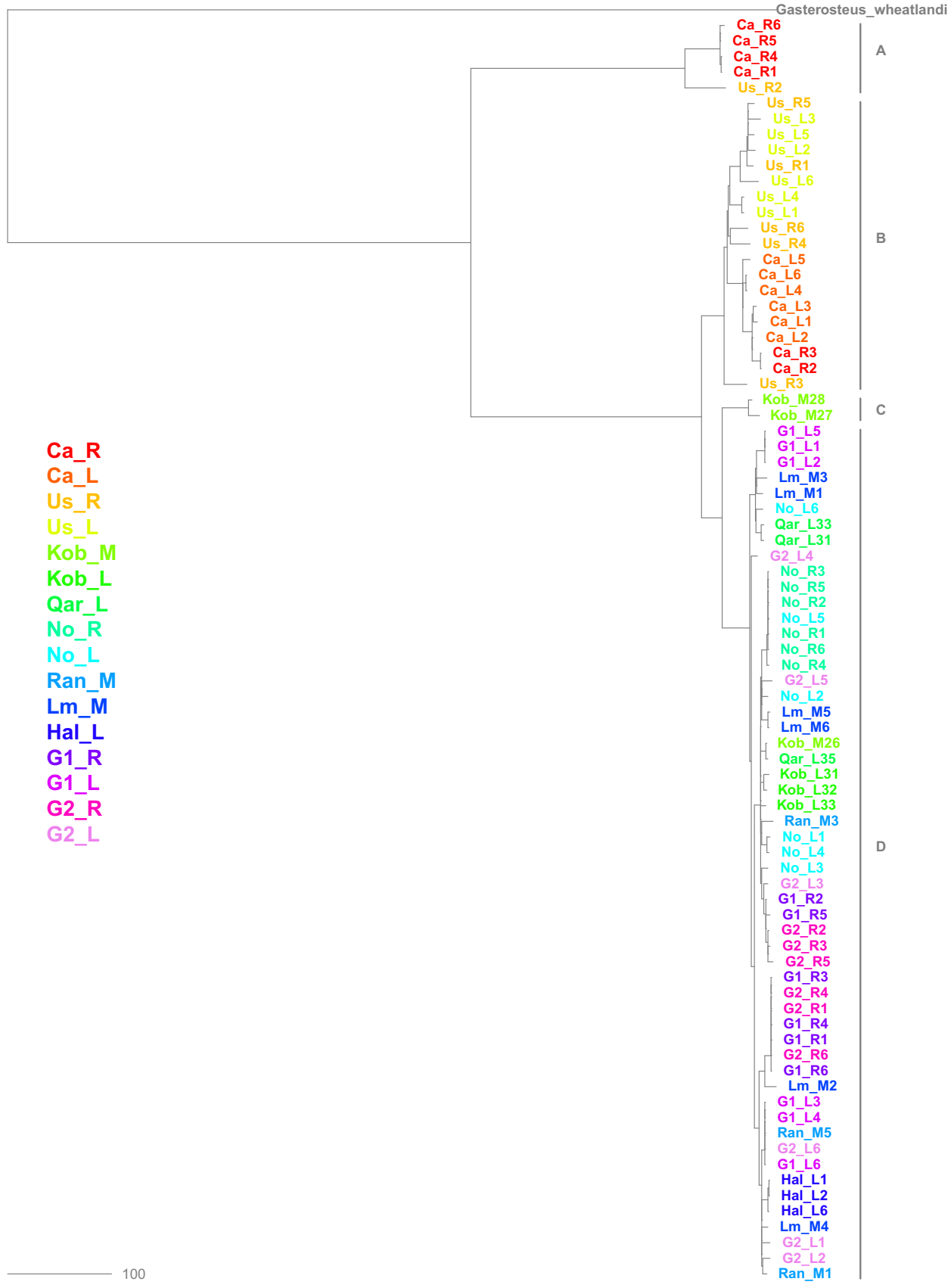

Fig. 5 Maximum Parsimony tree based on mitogenome sequences, excluding the control region, from all individuals. A, B, C and D represent the four major lineages identified. By comparison with data from other studies (see Figs S3 and S4, Supporting information), lineage A corresponds to the 'Japanese Lineage', lineage B to the Pacific part of a major 'Euro-American' lineage, lineage C to the 'Transatlantic Lineage' and lineage D to the 'European Lineage' (Orti *et al.* 1994; Mäkinen & Merilä 2008). Hence, the total 'Euro-American' lineage originally defined by Orti *et al.* (1994) would encompass lineages B, C, and D of this study.

mutation rate of 3.7×10^{-8} is realistic and falls between two different estimates from fishes of 6.6×10^{-8} (Recknagel *et al.* 2013) and 1.42×10^{-8} (Guo *et al.* 2013), respectively. These may represent over- and underestimates due to time dependency effects (Ho *et al.* 2005), and the lowest estimate would suggest early founding of freshwater populations that is difficult to reconcile with known glaciation history, whereas the highest estimate would suggest very recent founding of the Atlantic freshwater populations (Figs S4 and S5, Supporting information). The ideal option would have been to estimate genomewide mutation rate for threespine stickleback, using the split between Pacific and Atlantic populations as a calibration point. Unfortunately, however, such an approach would be questionable considering the mitochondrial DNA evidence for admixture between the 'Euro-American' and 'Japanese' phylogeographic lineages that appears to have occurred after colonization of the Atlantic basin (Orti *et al.* 1994; Lescak *et al.* 2015b).

PSMC implicitly assumes a panmictic population structure, and it has recently been demonstrated that deviations from this, that is population subdivision, can affect results and their interpretation (Mazet *et al.* 2016). Specifically, PSMC measures 'inverse instantaneous coalescence rate' (IICR) that in a panmictic system equals effective population size, but has a more complex definition in systems with population subdivision. Simulations show that populations may exhibit stable effective population size, but changes in migration rate can affect IICR and lead to signals in PSMC analyses that are erroneously interpreted as population size change (Mazet *et al.* 2016). In the case of threespine stickleback, some gene flow between marine and freshwater populations is expected, as is indeed assumed in the Transporter Hypothesis (Schluter & Conte 2009) invoked to explain the distribution of standing variation across marine and freshwater populations. The time of founding of freshwater populations would coincide with both the establishment of population subdivision and with population declines in freshwater populations, thus making it difficult to distinguish between the two. However, it makes sense that population size should decline in restricted freshwater habitats as compared to large, continuous marine habitats. We also note that PSMC plots for marine and freshwater populations, respectively, are remarkably similar across Northern Europe and Greenland. If population subdivision was an issue, then we would expect to see more differences across regions, resulting in different PSMC plots. Instead, the high similarity of inferred demographic histories suggest that a common factor underlies the results, that is glaciation history and subsequent founding of populations affecting population sizes.

In the Greenlandic and Pacific North American populations, there was evidence at the mitogenome level for admixture of different phylogeographical lineages. In Greenland, different lineages were observed among marine sticklebacks, but it is possible that larger sample sizes would have revealed the same in the freshwater populations. The inferred demographic histories of the Greenland populations were nevertheless quite similar to those of the nonadmixed Northern European populations, suggesting that effects of admixture on the results from PSMC analysis were modest if at all present.

In the Pacific North American populations, there was a signal of gradual population expansion from >100 Kya to ca. 14–15 Kya (Fig. 3). At the same time, admixture is well documented at the mitochondrial DNA level in previous studies (Orti *et al.* 1994; Lescak *et al.* 2015b) and in the present results (Fig. 5), but the time at which admixture took place is not known. It is therefore difficult to rule out that admixture could have affected the PSMC results for these populations, although we note that the results make sense in the light of the Glaciation history and presence of refugia in the region (see below).

As a whole, given uncertainties about the true mutation rate the timing of events should be treated with caution. Furthermore, deviations from a panmictic population model should be considered when interpreting demographic histories, although in the case of sticklebacks we still find population size changes to be the most likely factors underlying the results.

Demographic histories across regions

There was a striking difference of demographic histories between Pacific North American populations on one hand and Atlantic (Greenlandic and Northern European) populations on the other. Pacific North American populations showed much more gradual and prolonged expansion, commencing already during the Weichselian/Wisconsinan Glaciation >100 Kya. This would correspond well with suggestions of cryptic marine refugia in the North American Pacific (Shafer *et al.*

2010) close to the extant populations. Hence, northern Vancouver Island from where the samples Ca_L and Ca_R are derived may have been an offshore refugium, and there is further evidence of refugial status of the Haida Gwaii archipelago ca. 250 km further North-West (Shafer *et al.* 2010). In the case of sticklebacks, this archipelago harbours very high levels of phenotypic variation among populations, which could in itself argue for long persistence of sticklebacks in this region, although extensive population genomics analysis neither provided strong evidence for nor against this hypothesis (Deagle *et al.* 2013).

In contrast to Ca_L and Ca_R, the US_R and US_L samples are from a region in Alaska that was covered by ice during the Last Glacial Maximum and further away from possible cryptic refugia. Yet they showed the same general demographic history as CA_L and Ca_R, and if we assume that the peak of the demographic expansion corresponds to the time of founding of the freshwater populations, then this would have occurred ca. 14–15 Kya in all the CA and US populations, and certainly earlier than in the Greenlandic and Northern European populations; this also finds support in the PCA plots, which show stronger divergence between Pacific than between Atlantic populations along the first two axes (Fig. 2). It corresponds well with suggestions that the region in Alaska represented by the US samples may have been deglaciated as early as 14–16 Kya (Mandryk *et al.* 2001). Also, this means that freshwater sticklebacks must have colonized lakes rapidly as the coastal areas became ice free. Indeed, sticklebacks are efficient dispersers, as can be seen from their current colonization of the Svalbard Archipelago in the Arctic Ocean, presumably as a result of climate change (Svenning *et al.* 2015).

The Greenlandic and Northern European populations showed similar demographic histories, but distinct from the Pacific populations. As two mitogenome lineages were present in Greenland and only one in Northern Europe, this suggests that the similarity of demographic patterns reflect exposure to similar climatic history and not necessarily expansion from the same refuges, although incomplete lineage sorting could also explain the finding of two lineages. The PSMC plots showed an expansion starting ca. 20–25 Kya and for the freshwater populations culminating ca. 8–9 Kya. The latter presumably reflects colonization of the freshwater bodies, and if we consider that PSMC tends to smooth changes in population size (Li & Durbin 2011), then the initiation of the expansion could correspond to the end of the Last Glacial Maximum ca. 20 Kya. In the Nuuk Fjord region, Greenland, the coastal areas were ice free ca. 10–9.5 Kya (Funder *et al.* 2011) and freshwater sticklebacks must have colonized freshwater lakes soon thereafter. The PSMC results suggested that they have persisted uninterrupted since then, despite unfavourable climatic events such as the more recent Little Ice Age from ca. 1300 to 1850 AD.

In Northern Europe and particularly the Jutland Peninsula, the ice had retreated from large areas already 14 Kya. As sticklebacks are found in Arctic regions and have previously been distributed as far North as Melville Bay, Greenland (75°N) (Bennike 1997), they would be expected to be among the first postglacial colonizers in Northern Europe. This does, however, not agree well with the PSMC results suggesting colonization ca. 8–9 Kya in all freshwater populations. Of course, uncertainties concerning the true mutation rate could account for this discrepancy. The results nevertheless find support in a study of subfossil remains from a lake in Northern Jutland, Denmark, dating back ca. 12 Kya (Aaris-Sørensen 1995), documenting the presence of freshwater fishes like Northern pike (*Esox lucius*), European lake whitefish (*Coregonus lavaretus*), burbot (*Lota lota*), smelt (*Osmerus operlanus*) and ruffe (*Acerina cernua*). These species are assumed to have dispersed from nearby glacial refuges, presumably freshwater bodies in the ice-free Doggerland region, most of which is now submerged in the North Sea. Threespine stickleback is, however, conspicuously absent, which suggests that it was not present in the freshwater refugia from where the other species were recruited, nor had it immigrated from the sea at this time. In fact, its presence is not documented until ca. 6–7 Kya, where it is found in high numbers in kitchen middens (Enghoff 1994). Both molecular and subfossil data therefore suggest a surprisingly late postglacial recolonization of sticklebacks. One possibility could be that multiple postglacial recolonizations occurred, but that most or all populations from the earliest recolonization events went extinct during cold periods at the transition between the Weichselian Glaciation and the Postglacial period, such as Younger Dryas ca. 13–12 Kya.

Demographic history of marine and freshwater populations

The consistently high N_e in marine populations support the assumptions of the Transporter Hypothesis (Schluter & Conte 2009) of large, connected marine populations maintaining high levels of standing genetic variation. The freshwater populations provide a contrast with population declines that we interpret to represent initial founder events, cessation of gene flow from marine sticklebacks and habitat sizes limiting population sizes.

It is curious that all freshwater populations show population declines extending over several thousands of years. If habitat size is the primary factor limiting N_e, then the effects should result in a more sudden rather than prolonged decline. It is possible, however, that ongoing gene flow from marine sticklebacks could cause a signal of a more prolonged decline. Even though it has been demonstrated that newly founded freshwater populations can evolve into phenotypes with reduced armour within a few generations (Bell *et al.* 2004; Le Rouzic *et al.* 2011; Lescak *et al.* 2015a), genomic footprints of divergence between marine and freshwater sticklebacks are consistent with some ongoing gene flow and diversifying selection acting on specific genomic regions (Roesti *et al.* 2014). In addition to these biological factors PSMC's tendency to smooth sudden demographic events (Li & Durbin 2011) could also have contributed to the results. The methods MSMC and diCal (Sheehan *et al.* 2013; Schiffels & Durbin 2014) might be used for increasing resolution of the demographic history towards the recent past by basing inferences on more than a single individual, but this requires phasing of genomes which again requires higher sample sizes.

Finally, it should be noted that the genome sequences derived from the study by Feulner *et al.* (2015) encompassed pairs of parapatric lake and river populations. We did not observe differences in demographic histories between populations within pairs. Presumably, they have diverged too recently compared to the timescale that PSMC can resolve or their demographic histories have tracked the same general developments in environmental conditions.

Conclusions

Our study showed distinctly different demographic histories of threespine stickleback populations from the Pacific and Atlantic basins. If we assume that the PSMC results from Pacific North American populations reflect population size changes and not admixture, then persistence in refugia near the present distributional range, the long-term population expansion and high N_e, and relatively early founding of freshwater populations should increase evolutionary potential and possibilities for adaptive divergence. This is in contrast to the Greenlandic and Northern European populations that showed classical patterns of rapid postglacial expansion possibly accompanied by loss of variation (Hewitt 1996, 2000). The high phenotypic diversity in some Pacific regions has often been highlighted (Orti *et al.* 1994; Deagle *et al.* 1996, 2013), and it would be interesting to test if diversity is in fact higher in this region as compared to the Atlantic Basin. As such, our results caution against making strong generalizations about adaptive processes for the species as a whole based on studies from just one major geographical region.

The populations from Greenland and Northern Europe showed very similar demographic histories, but we found it surprising that Northern European sticklebacks had apparently recolonized the region later than freshwater fishes like lake whitefish and Northern pike, as indicated both by our results and subfossil data. It would be interesting to investigate this in more detail by analysing populations from areas such as larger lake systems, where possible remnants of early stickleback recolonizations could have persisted and survived colder time periods. Finally, inference of demographic history in marine and freshwater populations illustrated a common demographic history until founding of the freshwater populations, after which freshwater populations declined whereas marine populations remained large.

In total, our results show that even in well-established model species like threespine stickleback, methods for analysing demographic histories based on genomic data, such as PSMC, can provide surprising new insights. A major part of adaptive divergence can undoubtedly be explained by current environmental and ecological parameters, but the role of long-term demographic history should not be dismissed, both for understanding adaptation dynamics within regions, such as assumptions underlying the Transporter Hypothesis, and for understanding possible differences across populations and regions.

Acknowledgements

We thank the authors who have published whole genome sequences that were included in the study, Annie Brandstrup for technical assistance, Rasmus Nygaard, Rasmus Hedeholm and Kim Mouritsen for sampling of Greenland sticklebacks, Michael Glad for keeping computers running and the Subject Editor and two anonymous reviewers for constructive comments and suggestions. Funding was obtained from the Danish Council for Independent Research, Natural Sciences (grant no. 1323-00158A to MMH), the Villum Foundation (grant no. VKR022523 to MMH) and EU Interreg (Øresund-Kattegat-Skagerrak) funds (MARGEN).

References

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

Altshuler D, Durbin RM, Abecasis GR *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.

Arnegard ME, McGee MD, Matthews B *et al.* (2014) Genetics of ecological divergence during speciation. *Nature*, **511**, 307–311.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge.

Bell MA, Foster SA (1994) Introduction to the evolutionary biology of the threespine stickleback. In: *The Evolutionary Biology of the Threespine Stickleback* (eds Bell MA, Foster SA), pp. 1–27. Oxford University Press, New York, New York.

Bell MA, Aguirre WE, Buck NJ (2004) Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution*, **58**, 814–824.

Bennike O (1997) Quaternary vertebrates from Greenland: a review. *Quaternary Science Reviews*, **16**, 899–909.

Bergersen R (1996) Sticklebacks from Greenland. *Journal of Fish Biology*, **48**, 799–801.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of nearctic and palearctic fishes. *Molecular Ecology*, **7**, 431–452.

Chan YF, Marks ME, Jones FC *et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science*, **327**, 302–305.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 5039–5047.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Deagle BE, Reimchen TE, Levin DB (1996) Origins of endemic stickleback from the Queen Charlotte Islands: mitochondrial and morphological evidence. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **74**, 1045–1056.

Deagle BE, Jones FC, Absher DM, Kingsley DM, Reimchen TE (2013) Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, **22**, 1917–1932.

DeFaveri J, Merila J (2013) Variation in age and size in Fennoscandian three-spined sticklebacks (*Gasterosteus aculeatus*). *PLoS One*, **8**, e80866.

DeFaveri J, Shikano T, Shimada Y, Goto A, Merila J (2011) Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution*, **65**, 1800–1807.

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Ellegren H (2007) Molecular evolutionary genomics of birds. *Cytogenetic and Genome Research*, **117**, 120–130.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.

Enghoff IB (1994) Fishing in Denmark during the Ertebølle Period. *International Journal of Osteoarchaeology*, **4**, 65–96.

Ferchaud AL, Hansen MM (2016) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: threespine sticklebacks in divergent environments. *Molecular Ecology*, **25**, 238–259.

Feulner PGD, Chain FJJ, Panchal M *et al.* (2013) Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology*, **22**, 635–649.

Feulner PG, Chain FJ, Panchal M *et al.* (2015) Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics*, **11**, e1004966.

Funder S, Kjeldsen KK, Kjær KH, Cofaigh CO (2011) The Greenland ice sheet during the past 300,000 years: a review. *Developments in Quaternary Science*, **15**, 699–713.

Green RE, Braun EL, Armstrong J *et al.* (2014) Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science*, **346**, 1335.

Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate Maximum Likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Guo BC, Chain FJJ, Bornberg-Bauer E, Leder EH, Merila J (2013) Genomic divergence between nine- and three-spined sticklebacks. *BMC Genomics*, **14**, 756.

Hendry AP, Taylor EB, McPhail JD (2002) Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. *Evolution*, **56**, 1199–1216.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

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.

Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics*, **6**, 23.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Knowles LL (2009) Statistical phylogeography. *Annual Review of Ecology Evolution and Systematics*, **40**, 593–612.

Le Rouzic A, Østbye K, Klepaker TO *et al.* (2011) Strong and consistent natural selection associated with armour reduction in sticklebacks. *Molecular Ecology*, **20**, 2483–2493.

Lescak EA, Bassham SL, Catchen J *et al.* (2015a) Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, E7204–E7212.

Lescak EA, Marcotte RW, Kenney LA *et al.* (2015b) Admixture of ancient mitochondrial lineages in three-spined stickleback populations from the North Pacific. *Journal of Biogeography*, **42**, 532–539.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature*, **475**, 493–U484.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li MK, Schroeder R, Ko A, Stoneking M (2012) Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Research*, **40**, e137.

Mäkinen HS, Merilä J (2008) Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe – Evidence for multiple glacial refugia. *Molecular Phylogenetics and Evolution*, **46**, 167–182.

Mandryk CAS, Josenhans H, Fedje DW, Mathewes RW (2001) Late Quaternary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. *Quaternary Science Reviews*, **20**, 301–314.

Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, **116**, 362–371.

McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution*, **17**, 480–488.

McVean GAT, Cardin NJ (2005) Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1387–1393.

Moura AE, Janse van Rensburg C, Pilot M *et al.* (2014) Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Molecular Biology and Evolution*, **31**, 1121–1131.

Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*, **156**, 297–304.

Nadachowska-Brzyska K, Li C, Smeds L, Zhang GJ, Ellegren H (2015) Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Current Biology*, **25**, 1375–1380.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H (2016) PSMC-analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology*, **25**, 1058–1072.

Nurk S, Bankevich A, Antipov D *et al.* (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of Computational Biology*, **20**, 714–737.

Orti G, Bell MA, Reimchen TE, Meyer A (1994) Global survey of mitochondrial DNA sequences in the threespine stickleback – evidence for recent migrations. *Evolution*, **48**, 608–622.

Palkopoulou E, Mallick S, Skoglund P *et al.* (2015) Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, **25**, 1395–1400.

Prufer K, Racimo F, Patterson N *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–49.

Raeymaekers JA, Konijnendijk N, Larmuseau MH *et al.* (2014) A gene with major phenotypic effects as a target for selection vs. homogenizing gene flow. *Molecular Ecology*, **23**, 162–181.

Ravinet M, Harrod C, Eizaguirre C, Prodohl PA (2014) Unique mitochondrial DNA lineages in Irish stickleback populations: cryptic refugium or rapid recolonization? *Ecology and Evolution*, **4**, 2488–2504.

Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (Amphilophus spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3-Genes Genomes Genetics*, **3**, 65–74.

Roesti M, Gavrilets S, Hendry AP, Salzburger W, Berner D (2014) The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, **23**, 3944–3956.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, **46**, 919–925.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9955–9962.

Shafer ABA, Cullingham CI, Cote SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.

Sheehan S, Harris K, Song YS (2013) Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. *Genetics*, **194**, 647–662.

Svenning MA, Aas M, Borgstrom R (2015) First records of three-spined stickleback *Gasterosteus aculeatus* in Svalbard freshwaters: an effect of climate change? *Polar Biology*, **38**, 1937–1940.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Terekhanova NV, Logacheva MD, Penin AA *et al.* (2014) Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genetics*, **10**, e1004696.

Wallberg A, Han F, Wellhagen G *et al.* (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.

Xue YL, Prado-Martinez J, Sudmant PH *et al.* (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*, **348**, 242–245.

Zhao SC, Zheng PP, Dong SS *et al.* (2013) Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, **45**, 67–U99.

M.M.H., S.L. and M.W.J. conceived and designed the study. S.L. conducted all bioinformatics and statistical analyses with inputs and discussion by M.M.H. and M.W.J. M.M.H., S.L. and M.W.J. wrote the manuscript. All authors read and approved the final manuscript.

Data accessibility

Whole genome sequences generated in this study are deposited in the EBI Sequence Read Archive (SRA) with Accession no. ERP016886. A VCF file containing the SNP data used for the PCA analyses, and files with the

alignments used for the analysis of mtDNA have been uploaded to DRYAD (http://dx.doi.org/10.5061/dryad.46fb1).

Supporting information

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

Table S1 Summary statistics for the individual whole genome sequences; Number of reads, percentage of reads that were mapped to the reference genome, percentage of paired reads mapped, sequencing depth/coverage, average size of inserts, individuals heterozygosity.

Fig. S1 Eigenvalues for PCA axes, based on analysis of 85 698 SNPs in 81 individuals.

Fig. S2 Results of PSMC analyses of all individuals in all populations, shown for each population separately.

Fig. S3 PSMC plots for all individuals, superimposed in a single figure.

Fig. S4 Results of PSMC analyses shown for each population separately and assuming a mutation rate, $\mu = 1.42 \times 10^{-8}$.

Fig. S5 Results of PSMC analyses shown for each population separately and assuming a mutation rate, $\mu = 6.6 \times 10^{-8}$.

Fig. S6 Maximum Likelihood tree based on mitogenome sequences, excluding the control region, from all individuals.

Fig. S7 Bayesian tree based on mitogenome sequences, excluding the control region, from all individuals.

Fig. S8 Maximum Parsimony Tree based on mitochondrial DNA Cytochrome b sequences, involving samples from the present study and previously published data (Orti *et al.* 1994).

Fig. S9 Maximum Parsimony Tree based on mitochondrial DNA Cytochrome b sequences, involving samples from the present study and previously published data (Mäkinen & Merilä 2008).