Molecular Phylogenetics and Evolution 127 (2018) 613–625

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Worldwide phylogeny of three-spined sticklebacks

Bohao Fang[a,*], Juha Merilä[a], Filipe Ribeiro[b], Carlos M. Alexandre[c], Paolo Momigliano[a]

[a] Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, FI-00014 Helsinki, Finland
[b] MARE – Marine and Environmental Sciences Centre, Faculty of Science, University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal
[c] MARE – Marine and Environmental Sciences Centre, University of Évora, Largo dos Colegiais 2, 7004-516 Évora, Portugal

ARTICLE INFO

Keywords:
Coalescent
Gasterosteus
Phylogenomics
RAD-seq
SNP
Threespine stickleback

ABSTRACT

Stickleback fishes in the family *Gasterosteidae* have become model organisms in ecology and evolutionary biology. However, even in the case of the most widely studied species in this family – the three-spined stickleback (*Gasterosteus aculeatus*) – the worldwide phylogenetic relationships and colonization history of the different populations and lineages remain poorly resolved. Using a large collection of samples covering most parts of the species distribution range, we subjected thousands of single nucleotide polymorphisms to coalescent analyses in order to reconstruct a robust worldwide phylogeny of extant *G. aculeatus* populations, as well as their ancestral geographic distributions using Statistical-Dispersal Vicariance and Bayesian Binary MCMC analyses. The results suggest that contemporary populations originated from the Pacific Ocean in the Late Pleistocene, and the Atlantic was colonized through the Arctic Ocean by a lineage that diverged from Pacific sticklebacks ca 44.6 Kya. This lineage contains two branches: one that is distributed in the Mediterranean area, from the Iberian Peninsula to the Black Sea ('Southern European Clade'), and another that is comprised of populations from northern Europe and the east coast of North America ('Trans-Atlantic Clade'). Hence, the results suggest that the North American East Coast was colonized by trans-Atlantic migration. Coalescence-based divergence time estimates suggest that divergence among major clades is much more recent than previously estimated.

1. Introduction

During the past two decades, the three-spined stickleback (*Gasterosteus aculeatus*), a small teleost easily reared in laboratory conditions, has become one of the most important model species in ecology and evolutionary biology (Bell and Foster, 1994; Gibson, 2005; Hendry et al., 2013; Östlund-Nilsson et al., 2006). With a nearly circumpolar marine distribution in the Northern Hemisphere, this species has colonized freshwater habitats multiple times since the last glacial maximum, evolving dramatic morphological, physiological and behavioral adaptations in remarkably short periods of time (Barrett et al., 2011). In recent years, the development of restriction site-associated DNA (RAD) and whole-genome sequencing (WGS) has resulted in a plethora of studies exploring the genomic basis of these adaptations (Baird et al., 2008; Hohenlohe et al., 2010; Jones et al., 2012, etc.). As such, the combination of a broad geographic distribution, high degree of morphological, physiological and behavioral diversity, and the availability of extensive genomic resources has made the three-spined stickleback an ideal candidate species to test hypotheses in the fields of biogeography (Mäkinen et al., 2006; Orti et al., 1994), physiology (Barrett et al., 2011; Bell, 2001; Kitano et al., 2010), developmental biology (Shapiro et al., 2004), population genetics (Cresko et al., 2004; Leinonen et al., 2006), comparative genomics (Guo et al., 2013), ethology (Huntingford and Ruiz-Gomez, 2009; Von Hippel, 2010), adaptive evolution and speciation (Gibson, 2005; Jones et al., 2012; McKinnon and Rundle, 2002).

A well-supported and highly resolved phylogeny is a prerequisite for understanding evolutionary processes and testing ecological and evolutionary hypotheses, and in particular, for untangling processes that have shaped genomic divergence, such as natural selection and demographic changes (Delsuc et al., 2005). Interestingly, despite its importance as a model species for ecological and evolutionary studies, the worldwide phylogenetic relationships and the colonization history of three-spined sticklebacks in different ocean basins has yet to be fully resolved. As a result, the dispersal routes and refugial origins of various lineages and populations have been disputed. There are a number of reasons for this.

The first reason is restricted geographic sampling. Previous phylogeographic studies have mostly been based on limited geographic sampling, with focus on specific geographic regions (Cano et al., 2008; Lucek et al., 2010; DeFaveri et al., 2012; Deagle et al., 2013; Ravinet et al., 2014; Sanz et al., 2015; Vila et al., 2017; Table 1). Although there are a few comprehensive studies that encompass a broader geographic

* Corresponding author.
E-mail address: bohao.fang@helsinki.fi (B. Fang).

https://doi.org/10.1016/j.ympev.2018.06.008
Received 12 December 2017; Received in revised form 16 April 2018; Accepted 4 June 2018
Available online 12 June 2018

Table 1
A review of previous representative large-scale phylogenetic studies for *G. aculeatus*. The table contains the type of genetic marker used, the number of sampling localities across the regions listed, and brief synopsis of the main results of each study. The most comprehensive studies conducted so far have focused on few mitochondrial gene fragments and a handful of microsatellite loci, or based on limited geographic sampling.

Studies	Years	Genetic marker	Location number	Sampling range					
				Pacific		Iberian Mediterranean			
				Western Pacific basin	Eastern Pacific basin	Iberian peninsula	Mediterranean	Adriatic sea	Black sea
Haglund et al.	1992	Allozyme	16	✓	✓		✓		
Orti et al.	1994	mtDNA	25	✓	✓				
Colosimo et al.	2005	SNPs	25	✓	✓				
Mäkinen et al.	2006	Microsatellite	73			✓	✓	✓	✓
Mäkinen and Merilä	2008	mtDNA	49			✓	✓	✓	✓
Cano et al.	2008	mtDNA & Microsatellite	18					✓	
DeFaveri et al.	2012	mtDNA & Microsatellite	10*					✓	
Ravinet et al.	2014	mtDNA & Microsatellite	46*			✓	✓	✓	✓
Sanz et al.	2015	mtDNA	7*			✓	✓	✓	✓
Liu et al.	2016	SNPs	5*		✓				
Vila et al.	2017	mtDNA & Microsatellite	17*			✓		✓	✓
This study	**2018**	**Genome-wide SNPs**	**70**	✓	✓	✓	✓	✓	✓

Studies	Sampling range								Main findings
	Europe								
	Western Atlantic	Atlantic	Norwegian Sea	North sea	Barents sea	White sea	Baltic sea	Mainland Europe	
Haglund et al.	✓			✓			✓		Interpopulation divergence between Pacific and Atlantic
Orti et al.	✓	✓					✓	✓	Two divergent mitochondrial clades: *Japanese* and *North American-Atlantic* clade; Colonization of Atlantic dates to 260–90 Kya.
Colosimo et al.	✓	✓	✓	✓					Two major lineages: *Atlantic* and *Pacific* lineage
Mäkinen et al.	✓	✓	✓	✓	✓	✓	✓	✓	Marine ancestors colonize freshwater during post-glaciation; Mediterranean region colonization dates back to the Pleistocene
Mäkinen and Merilä	✓	✓	✓	✓	✓	✓	✓	✓	Three major mtDNA lineages: *Trans-Atlantic*, *European* and *Black Sea* lineage; The divergence between major lineages occurred 170–130 Kya based on a molecular clock calibration
Cano et al.			✓	✓	✓	✓	✓	✓	Adriatic lineage may have a long independent evolutionary history.
DeFaveri et al.			✓	✓	✓		✓	✓	Two divergent lineages: *Adriatic* and *European* lineage; Adriatic lineage is most divergent and ancient
Ravinet et al.	✓	✓	✓	✓	✓	✓	✓	✓	Two major lineages: *European* and *Trans-Atlantic* lineage
Sanz et al.	✓	✓	✓	✓	✓	✓	✓	✓	Two independent lineages in the Mediterranean region are divergent from Atlantic (Portuguese) lineage; Divergence among Iberian Mediterranean populations precedes last glaciation
Liu et al.	✓			✓			✓		Four lineages: *Japanese*, *Euro-American*, *Transatlantic* and *Europe* lineage; Freshwater populations are founded by marine ancestors post-glacially
Vila et al.		✓	✓	✓	✓	✓	✓	✓	Iberian Peninsula has a history of multiple colonization
This study	✓	✓	✓	✓	✓	✓	✓	✓	See discussion

* Other published haplotype data from previous literature were also included in the analyses.

range, these have still failed to include either Pacific populations (Mäkinen et al., 2006; Mäkinen and Merilä, 2008; Ravinet et al., 2014; Sanz et al., 2015; Vila et al., 2017) or Iberian-Mediterranean populations (Orti et al., 1994), leading to an incomplete picture of the overall phylogenetic relationships between distant lineages. The second reason is that most phylogenetic studies conducted thus far have used either one or two mitochondrial gene fragments and/or a few microsatellite loci (Mäkinen et al., 2006; Mäkinen and Merilä, 2008; Orti et al., 1994; Ravinet et al., 2014; Sanz et al., 2015). While mitochondrial DNA (mtDNA) has remained an important marker for species phylogenies, it is nevertheless a single, non-recombining locus that is maternally inherited. Thus, any tree that is reconstructed with mtDNA will only reflect the phylogenetic history of female lineages. Moreover, the high and variable substitution rates in mtDNA can lead to longer branch lengths, which may generate aberrant attraction between lineages that is independent of the actual underlying phylogeny (Felsenstein, 1978). Finally, mtDNA-based phylogenies (as well as phylogenies based on single nuclear markers) are often affected by introgression and hence may not reflect the dominant phylogenetic history of a species or population (Leache et al., 2016). When it comes to nuclear markers, quickly evolving microsatellites lack the power to resolve deep and complex phylogenetic relationships (Goldstein and Pollock, 1997). However, with the exception of a few recent SNP-based studies (e.g., Colosimo et al. (2005); Deagle et al. (2013)), microsatellites have remained the more widely used marker among stickleback phylogenetic studies. Hence, despite the numerous phylogenetic studies of three-spined sticklebacks (see Table 1 for synopsis), our understanding of the evolutionary history of this widely distributed species remains fragmented and incomplete, sometimes with conflicting evidence of ancestry.

The development of genome complexity reduction protocols such as RAD-seq, which allows sampling a random selection of markers across the genome, together with a rich repertoire of analytical tools has opened the door for large phylogenomic analyses using thousands of genome-wide polymorphisms (Andrews et al., 2016). Phylogenetic studies that employed supermatrices of concatenated RAD loci played a pivotal role in resolving difficult phylogenetic challenges, such as defining species boundaries in recent and rapid adaptive radiations (Wagner et al., 2013). Constructing phylogenies from multiple loci does pose challenges, since different genomic regions likely undergo different sorting processes and can therefore depict different phylogenies that do not accurately represent the species tree (Maddison, 1997; Takahata, 1989). Rapid adaptive radiations may result in incomplete lineage sorting (ILS), causing incongruence between phylogenetic trees for different genomic segments and the overall population tree (Pamilo and Nei, 1988; Rogers and Gibbs, 2014). However, under a wide range of scenarios, methods based on the concatenation of multiple loci into supermatrices accurately recover the underlying species tree (DeGiorgio and Degnan, 2009; Lambert et al., 2015; Liu et al., 2014; Rivers et al., 2016; Tonini et al., 2015), even in cases where incomplete lineage sorting causes incongruences between individual gene trees and the species tree (Mirarab et al., 2014). Lambert et al. (2015) show that relaxed clock Bayesian analyses of concatenated data such as the analyses performed in BEAST2 (Bouckaert et al., 2014) can yield results which are nearly identical to species tree methods which are still computationally prohibitive to run on large datasets.

The aims of this study were to (1) construct a comprehensive and robust phylogenetic hypothesis of historical relationships among three-spined stickleback populations collected throughout their worldwide distribution range, (2) estimate divergence times among the major lineages, and (3) reconstruct their colonization routes. To these ends, we utilized a large amount of RAD-seq data from samples collected from 70 locations covering most of the species range, and reconstructed their worldwide phylogeny and colonization routes using coalescent and biogeographic modeling methods, respectively.

2. Materials and methods

2.1. Samples

A total of 126 three-spined sticklebacks from 70 locations spanning the entire species' distribution range were used in this study (Fig. 1). Sampling included populations in the Eastern and Western Pacific Ocean, the east coast of North America (Western Atlantic), and multiple locations across most of Europe, covering both northern and southern parts of the continent (Fig. 1 and Supplementary Table S1). Samples from Northern Europe and the Pacific area included fish from both marine (N = 27) and freshwater (N = 99; ponds, lakes and rivers)

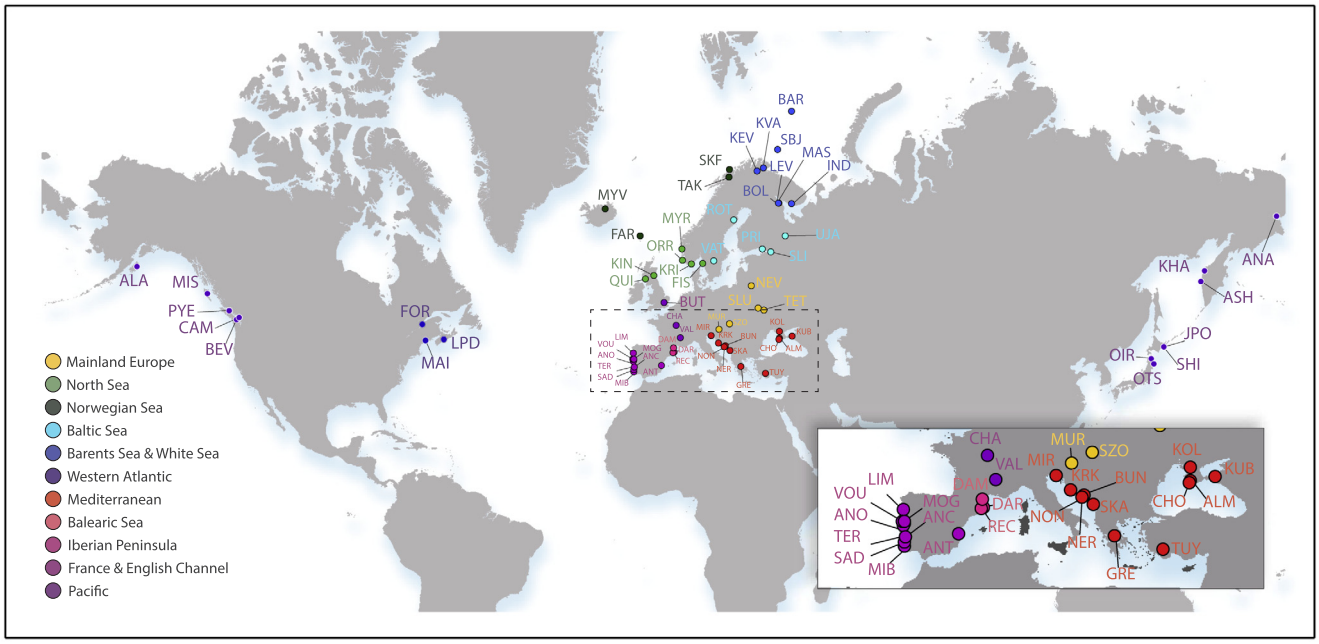


Fig. 1. Map showing the 70 worldwide sampling localities of *G. aculeatus* spanning most parts of the species distribution range. The three-letter abbreviations refer to sample codes used in the phylogenetic tree. Different colors refer to different regions of the sampling. See Supplementary Table S1 for sampling location details.

habitats; only freshwater populations were sampled from Southern Europe (Iberian Peninsula and Mediterranean; see Supplementary Table S1 for details on habitat and sampling coordinates). Most samples had been previously used in earlier microsatellite and mtDNA studies (DeFaveri et al., 2013; DeFaveri et al., 2012; Mäkinen et al., 2006; Mäkinen and Merilä, 2008), however several new locations were sampled specifically for this study. Fish were collected with seine nets, minnow traps or by electrofishing, and preserved in ethanol after an overdose of MS-222. Two individuals of *Gasterosteus nipponicus* (Higuchi et al., 2014), a sister species of *G. aculeatus*, were included as an outgroup taxon to root the tree.

2.2. DNA extraction, restriction site associated DNA library construction and sequencing

High molecular weight DNA was extracted from ethanol-preserved pectoral fins or muscle tissues using a standard phenol-chloroform (Sambrook and Russell, 2006) or modified salting out (Sunnucks and Hales, 1996) protocol. A total of 128 individuals were used to build the genomic libraries. Restriction site associated DNA (RAD) library preparation and sequencing were carried out by Beijing Genome Institute, (BGI Tech Solutions Co., Ltd, Hong Kong), following a protocol modified (see details below) from Baird et al. (2008). Briefly, genomic DNA was digested using the restriction enzyme PstI (5′-CTGCAG-3′ recognition site) and a P1 adapter containing a unique barcode, forward amplification and sequencing primer were ligated. Ligation products were then pooled in equimolar concentrations, randomly sheared and size selected using a 300–500 bp window through agarose. Y-adaptors (P2) were ligated and PCR amplification selectively enriched fragments with P1 and P2 adaptors. The barcoded RAD libraries were paired-end sequenced (read length of 100 bp) on four lanes of an Illumina HiSeq2000 platform. The sequence quality was checked via Fastx-toolkit (Gordon and Hannon, 2010). The latest reference genome of the three-spined stickleback (release-88) was retrieved from Ensembl database (Hubbard et al., 2005).

2.3. SNP calling

SNP calling was performed using the *ipyrad* pipeline (v.0.7.11; http://ipyrad.readthedocs.io/; Eaton (2014)) with the reference assembly method. First, adaptors were trimmed using *cutadapt* and all bases with a Phred quality score below 20 were converted in Ns. Low-quality bases at the 3′ end of the reads were trimmed, and reads that contained five or more Ns after trimming (or were shorter than 50 bp) were discarded. Second, filtered reads were mapped to the reference genome in *BWA* using the function *BWA mem* (Li, 2013) with default parameters in the *ipyrad* pipeline. The rest of the pipeline was run with default *ipyrad* parameters, and in the final step of the pipeline the data were filtered to retain only loci with a minimum read depth of six, a maximum read depth of 1000, no more than 20 SNPs per locus, $H_E < 0.5$, and no more than 25% missing data as suggested by Takahashi et al. (2014). Given that our populations share a relatively recent ancestry, and that we used the reference assembly method rather than a *de novo* assembly of RAD loci, extensive testing of *ipyrad* parameters that are known to be affected by the extent of divergence among lineages (such as the clustering threshold, see Eaton, 2014) were not deemed necessary.

Considering potential sources of genomic heterogeneity in sex chromosomes (e.g., Hedrick (2007); Schaffner (2004)), we used VCFtools v0.1.15 (Danecek et al., 2011) to exclude markers on linkage groups IX and XIX (neo-sex chromosome system in *G. aculeatus*; Kitano et al. (2009); Natri et al. (2013)). VCFtools was used to thin the dataset for computational efficiency; a subset of unlinked markers was generated by selecting sites that were greater than 10 kb from each other. We further filtered the final dataset by retaining only bi-allelic SNPs with a Minor Allele Count (MAC) greater than or equal to two, and removed SNPs with read depth exceeding 200 to exclude highly repetitive regions. The final dataset consisted of a supermatrix (de Queiroz and Gatesy, 2007), including 8079 concatenated SNPs which were converted to a FASTA alignment with PGDspider v. 2.1.1.0 (Lischer and Excoffier, 2011).

2.4. Phylogenetic inference

The most appropriate nucleotide substitution model for the final FASTA alignment containing 8079 genome-wide SNPs was determined using jModelTest 2 (Darriba et al., 2012). The results suggested the general time reversible substitution model with a gamma distribution of rates (GTR + G). We used the Bayesian Evolutionary Analysis by Sampling Trees (BEAST2) software package v.2.4.4 to construct a coalescent phylogeny (Bouckaert et al., 2014) using the Bayesian supermatrix approach (Ogilvie et al., 2016). The Bayesian phylogenetic inference was implemented through Markov Chain Monte Carlo (MCMC) simulations based on the relaxed log-normal molecular clock (Drummond et al., 2006), an approach that has been shown to be less affected by ILS than ML methods (Lambert et al., 2015). The input data for BEAST2 was the concatenated alignment of 8079 SNPs, in FASTA format. We ran 15 independent chains for 200 million generations, sampling the chain every 1000 trees after discarding the first 40,000 trees (20%) as burn-in. Stationarity and convergence of MCMC chains were visually checked in TRACER v1.6 (http://beast.bio.ed.ac.uk/tracer/). TREEANNOTATOR (Drummond et al., 2012) was used to summarize posterior parameters from tree samples and compose a consensus tree which was visualized with FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

2.5. Divergence time estimation

Divergence time was estimated from analyses of the concatenated dataset in BEAST2; for recent examples of other studies that have used concatenated genome-wide SNPs to estimate divergence time see Liu et al. (2015), Longo and Bernardi (2015), Tariel et al. (2016) and Zhou et al. (2018) . Time to the most recent common ancestor (TMRCA) for lineages was estimated through BEAST2 with the joint inference of phylogeny. To date the tree, three calibration points were used, including two geological events and the divergence time to the outgroup taxon.

As to the first geological calibration point, we note that the Mediterranean Sea became connected to the Black Sea following the end of the last glaciation. While the exact timing of this event remains a contentious issue, multiple studies provide a rough time estimate for the birth of the Black Sea-Mediterranean connection. The Black Sea deluge hypothesis (Ryan and Pitman, 2000) suggests a massive flooding event occurred about 7550 ya through the Bosphorus, a strait connecting the Black Sea and the Mediterranean. A radiocarbon dating study of mollusk shells in the Black Sea continental shelf indicated that the flooding of the Black Sea took place around 7000 ya (ca. 7460–6820 yr, Ballard et al. (2000)). Also, a study on the origin of the Bosphorus suggests the Black Sea was in its present condition about 7000 ya (Gökaşan et al., 1997). Hence, we used 7000 ya for Black Sea lineages as the first calibration point. A tentative tree inferred by BEAST2 without any calibration prior supported this choice: it revealed comparable branch lengths between Black Sea and post-glacial northern European lineages, suggesting a recent rather than a pre-Pleistocene origin of sampled Black Sea populations.

As to the second geological calibration point, the Baltic Ice Lake – a historical stage of the Baltic Sea – was a large water body formed ca. 12 kya when the modern Gulf of Bothnia was still covered by the ice cap. The lake was connected to the Atlantic Ocean ca. 10.3 kya (Nilsson et al., 2001) via a waterway through the Närke Strait, which might have facilitated three-spined stickleback colonization from the Atlantic to the Yoldia Sea (the following stage of the Baltic Sea after Baltic Ice Lake)

through Lake Vänern. Under this scenario, the TMRCA of the Baltic Sea stickleback populations was assumed to be around ca. 10 kya.

The third calibration point was the divergence between *G. aculeatus* and the outgroup taxon *G. nipponicus*, which according to Higuchi et al. (2014) started two million years ago due to reproductive isolation caused by geological changes and sea level fall in the ancient Sea of Japan.

From the above, we assigned three calibration points respectively for the Black Sea, Baltic Sea and the outgroup divergence. Specifically, the monophyletic group including populations in the Black Sea (KUB, CHO, ALM, KOL) and Aegean Sea (GAS) was assigned a TMRCA of 7 ky, with a normally distributed prior having a mean of 0.007 (mya) and sigma 0.00035 in BEAST2. The TMRCA of the monophyletic group consisting of the Baltic Sea populations (Gulf of Bothnia (SWE); Gulf of Finland (ROU, PRI, SLI, UJA)) was assumed to be 10 ky, with a normally distributed prior with a mean of 0.01 (mya) and sigma 0.00035. Finally, we set the divergence time from the outgroup to 2 mya, again with a normally distributed prior with a mean of 2 mya and sigma 0.18.

To verify the accuracy of time divergence estimation based exclusively on variable sites, we estimated divergence times using both the SNP data as well as a subset of 1000 concatenated RAD loci (i.e., including both SNPs and invariant bases) selected using the R package Radami (Hipp et al., 2014) from the "loci" output file from *ipyrad*. The latter dataset was used exclusively to calculate divergence times given the tree topology estimated using the SNP data, and solely for the purpose of confirming divergence time estimates generated with the SNP dataset. In BEAST2, priors were given according to the topology of the SNP tree. The rest of the settings, as well as the numbers of chains and running generations, were identical to those used in the analysis of the abovementioned SNP dataset. This verification analysis showed that the two sets of markers, (i.e., SNPs and concatenated loci) yielded similar divergence time estimates for all nodes, having overlapping and very similar HPD intervals (Supplementary Table S2).

2.6. Biogeographic analyses

The ancestral geographic ranges at each node were reconstructed by Statistical-Dispersal Vicariance Analysis (S-DIVA) and Bayesian Binary MCMC (BBM) analysis in the program Reconstruct Ancestral States in Phylogenies v4.0 (RASP; Yu et al. (2015)). In RASP, the BBM method inputs posterior distribution of Bayesian inference, in this study the consensus tree from BEAST, to reconstruct the possible ancestral distributions of given nodes via a hierarchical Bayesian approach (Ronquist and Huelsenbeck, 2003). Enhanced S-DIVA method reconstruct s the frequencies of ancestral distribution at each node by utilizing all posterior distribution to account for phylogenetic uncertainty and DIVA optimization (Yu et al., 2010).

Using the R package BioGeoBEARS we collapsed the phylogeny to a monophyletic population tree (Matzke, 2013), whose tips are monophyletic populations instead of samples. Outgroups were excluded as the input phylogeny should include only monophyletic groups (Matzke, 2013) without outgroup (Yu et al., 2015).

Distribution ranges in the analysis were divided into 10 areas, according to geographic proximity shown in the coalescent tree (Fig. 3B). These areas were: A (Norwegian, Barents & White Seas), B (France inland), C (North Sea), D (Mainland Europe), E (Baltic Sea), F (Western Atlantic), G (English Channel), H (Mediterranean & Black Sea), I (Iberian Peninsula), J (Pacific Ocean). For both methods, the number of maximum areas for each node was kept at two. In the Bayesian analysis, fixed JC + G (Jukes-Cantor + Gamma) model was used with 5,000,000 cycles, 10 chains, a temperature parameter of 0.1 and sampling every 100 generations.

2.7. Computational process

All computations were conducted in the High-Performance Computing system provided by the Finnish IT Center for Science (CSC). For runs with BEAST software, a BEAGLE library (Ayres et al., 2011) was used to enhance the computational efficiency by using multi-core and parallel graphics processors.

3. Results

3.1. Genome-wide sequencing data

The *ipyrad* pipeline identified a total of 15,678 loci with an average read depth of 25.09 (average read depth ranged from 7.1 to 71.8 across samples; see Supplementary Table S3). After filtering and thinning in VCFtools, the final dataset for phylogenetic analysis contained 8079 genome-wide SNPs.

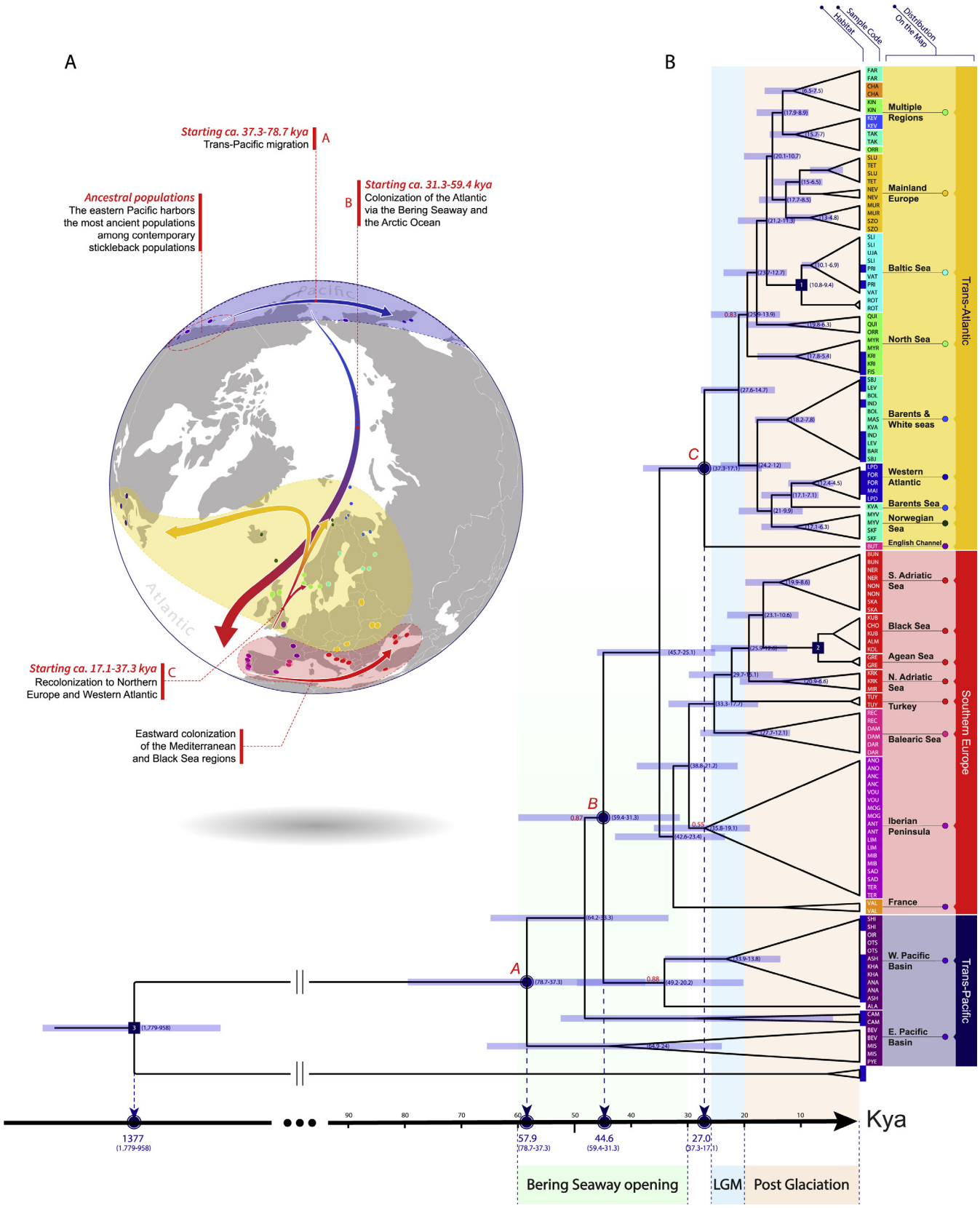
3.2. Worldwide phylogenetic relationships

Independent runs of Monte-Carlo Markov chains converged to equal posterior likelihood and with high Effective Sample Size (ESS > 500). The phylogenetic analysis revealed a well-supported topology whereby populations generally clustered according to their geographic proximity (Fig. 2). All but three of the interior nodes had very high posterior probability (P > 0.99), and all had posterior probabilities higher than 0.8 (Fig. 2). Only a few peripheral nodes had lower support (see posterior probabilities highlighted in red in Supplementary Fig. S1). The phylogenetic tree identified three distinct clades in the Pacific Ocean, with populations from the Eastern Pacific being a sister lineage for the rest of the lineages (Fig. 2). All populations from the Western Atlantic, the Eastern Atlantic as well as southern and northern Europe form a monophyletic clade that diverged from the Pacific lineages within the last 60 ky (Fig. 2). Within this monophyletic group there were two major lineages: an older Southern European Clade, which includes all populations from the Iberian Peninsula, the Mediterranean and the Black Sea regions, and a Trans-Atlantic Clade composed of mostly young populations (< 20 ky old) from the North Sea, the Baltic Sea, the White Sea, the Barents Sea and the Western Atlantic (Fig. 2). As for the result of divergence time estimation, TMRCA and its 95% highest probability density (HPD) intervals for each clade were obtained from the MCMC stationary parametric distribution (Fig. 2). The analyses using the dataset of concatenated loci yielded similar divergence time estimates, with all nodes having overlapping and very similar HPD intervals to those obtained using variable sites only (Supplementary Table S2). The probabilities of ancestral distribution areas for each node from BBM and S-DIVA methods were basically consistent in terms of most likely states (MLS) in nodes (MRCA) for all major lineages, with some minor variations (see Supplementary Fig. S2 for comparison of BBM and S_DIVA results). In view of the high degree of concordance between BBM and S-DIVA results, only the BBM result are shown in Fig. 3.

3.2.1. Pacific Clades

Within the Pacific, populations were grouped into three distinct clades. The first clade included freshwater populations (PYE, MIS, BEV) in the Eastern Pacific Basin in Vancouver and Queen Charlotte Islands, British Columbia; the second clade consists of a marine population (CAM) in the Eastern Pacific Basin around Vancouver Island, British Columbia; and the third clade consists of populations in the Western Pacific Basin, including a freshwater population in the Alaskan Kodiak (ALA), marine populations in the Russian Anadyr Bay of the Bering Sea (ANA) and Russian Poluostrov Kamchatka bordering the Sea of Okhotsk (KHA, ASH), as well as populations from the East Coast of Japan (OTS, OIR, SHI; Fig. 2). According to the BEAST analyses, the Eastern Pacific populations are phylogenetically the most ancient ones (diverged between ca. 78.7–37.3 kya), and hence placed on a sister lineage position with respect to the rest of the worldwide populations (Fig. 2).

The biogeographic analyses suggest the Pacific was the ancestral



(caption on next page)

Fig. 2. Time-calibrated worldwide phylogenetic tree of *G. aculeatus*. (A) The schematic globe shows the sampling localities (colored dots) and hypothesized major colonization routes; shaded colors indicate the distribution of major clades and colors correspond to those in the phylogenetic tree: blue (Pacific Clades), red (Southern European Clade) and yellow (Trans-Atlantic Clade). (B) The time-calibrated worldwide phylogenetic tree of *G. aculeatus* including sampling codes, sampling regions and habitat. 'Distribution' indicates the sea basins the samples originate from. Marine habitats are labeled with blue rectangles on the left side of the sampling code. The horizontal axis of evolutionary time is labeled by geographic events: Bering Seaway opening period (ca. 60–30 kya) and the Last Glacial Maximum (LGM, ca. 26–20 kya). Almost all nodes had extremely high posterior probability and unlabeled nodes have posterior probability values ≥0.99. Purple horizontal bars depict the 95% highest posterior density (HPD) intervals around the mean ages and values are presented in the parentheses. Time calibration points are given in nodes with squares. Nodes with black dots refer to major colonization events at (A). Closely related samples are collapsed and marked with triangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

region for the worldwide populations since the ancestral range for nodes 137 and 134 was Pacific Ocean in both BBM and S-DIVA with almost 100% marginal probability (Fig. 3, marginal probability in parenthesis hereafter). Node 133 was the postulated ancestral range of Pacific Ocean (area J, 98.92%) in BBM (Fig. 3), and the Pacific Ocean plus Iberian Peninsula (area IJ, 100%) in S-DIVA (Supplementary Fig. S2). Hence, S-DIVA suggests the early colonization from the Pacific Ocean to the Iberian Peninsula.

3.2.2. Southern European Clade

All samples from southern Europe formed a well-supported monophyletic group composed of three clades: a French, an Iberian and a Mediterranean clade (Fig. 2). The French clade includes only one population (VAL) in central Le Rhône River, which started to diverge from the rest of the European populations between ca. 42.6–23.4 kya. The Iberian clade includes eight freshwater populations from the Iberian Peninsula in Portugal (TER, SAD, MIB, LIM, MOG, VOU, ANC, ANO) as well as a population in the Xúquer River discharging into the Mediterranean Sea in Southeast Spain (ANT; Fig. 2). The Mediterranean clade, which started to diverge from the Iberian clade between 38.8 and 21.1 kya, is comprised of freshwater populations bordering the Mediterranean Sea, including populations in rivers draining into the Balearic Sea (DAR, DAM, REC), a population in the Dalaman Stream in southwest Turkey (TUY), Adriatic freshwater populations distributed along the western margin of the Balkans (MIR, KRK, SKA, NON, NER, BUN), populations in the Black Sea drainage (KOL, ALM, KUB, CHO), as well as a population in the Eastern Spercheios River connecting to the Aegean Sea in southern Greece (GRE; Fig. 2). The Adriatic Sea populations were split into two separate groups, one of which (the southern Adriatic Sea populations) forms a monophyletic group with the Black Sea populations (Fig. 2).

In the biogeographic analyses at node 126, shared by all Atlantic lineages, S-DIVA suggested English Channel and Iberian Peninsula (area GI, 100%) as ancestral areas, while BBM mostly favored Pacific Ocean (area J, 79.26%). At node 125, representing southern European lineages, Iberian Peninsula plus inland France (area BI, 100%) were suggested as ancestral areas by S-DIVA while BBM postulated Inland France (area B, 65.87%) as most likely ancestral area. In spite of slightly different results from BBM and S-DIVA at node 125, the ancestral area for Southern European Clade was most likely in the Iberian Peninsula and inland France, which were not glaciated before and during LGM.

3.2.3. Trans-Atlantic Clade

The Trans-Atlantic Clade structure was characterized by two major subclades, with a freshwater population (BUT) in southeast Britain as sister lineage (Fig. 2). One subclade ('Trans-Atlantic subclade') – which diverged from the other subclades ('North Sea-Baltic Sea subclade') ca. 27.6–14.7 kya – includes freshwater populations in the Norwegian Sea area (SKF, MYV), marine and freshwater populations from the East Coast of America (LPD, MAI, FOR), as well as populations from the Barents and White Seas (SBJ, BOL, LEV, IND, MAS, KVA, BAR; Fig. 2). The North Sea-Baltic Sea subclade included marine and freshwater populations from the eastern part of the North Sea Basin (FIS, KRI, MYR, ORR), a freshwater population in Scotland (QUI), marine and freshwater populations from the Baltic Sea drainages (ROT, VAT, PRI, SLI, UJA), mainland European freshwater populations (SZO, MUR, NEV, SLU, TET), as well as populations from various northern European regions (Norwegian Sea: TAK, FAR; Barents and Whites Seas: KEV; North Sea: KIN, ORR; France: CHA; Fig. 2). The Trans-Atlantic Clade is younger than all other major phylogeographic lineages, as all of its populations, with the exception of the population from the English Channel (BUT), were estimated to have originated following the end of the last glaciation. In all cases, divergence time HPD intervals were residing well within the post-glacial period that commenced in the Northern Hemisphere approximately 20 kya (Clark et al., 2009).

The biogeographic analyses suggest that the ancestral area for node 101 was the English Channel (BBM: area G, 64.68%) or areas AG (50%) and CG (50%; S-DIVA). The possible ancestral range for the node 100 was area AC (100%; S-DIVA) or area A (47.85%; BBM). Putting together nodes 126, 101 and 100, BBM suggested a colonization route from the Pacific to the English Channel then through the Atlantic to Northern Europe, while S-DIVA indicated colonization from the Pacific plus Iberian Peninsula to English Channel, and through the Atlantic to Northern Europe. Hence, both results indicated a colonization scenario from south to north in the Atlantic region after dispersal from the Pacific, and that the ancestral area for the trans-Atlantic populations was around the English Channel.

At node 99, which represents the most recent common ancestor for populations from the Western Atlantic (area F) and the majority of Norwegian Sea, Barents Sea and White Sea areas (area A), area A was postulated as the ancestral region by both BBM (97.19%) and S-DIVA (100%). Hence, the agreement between both analyses suggests that the Western Atlantic populations were colonized from northern Europe. The possible ancestral area for the node 85 was North Sea (BBM: 98.42% S-DIVA; 100%), suggesting colonization from the North Sea to the Baltic Sea and Inland Europe.

4. Discussion

4.1. Worldwide phylogeny and colonization scenario

In this study we reconstructed the phylogeographic history of three-spined sticklebacks using genomic data from populations representing the entire species distribution range. By employing the most geographically comprehensive sampling design to date and the analyses of thousands of independently assorting nuclear loci we unveiled the relationships and estimated divergence time for all the major lineages, overcoming the limitations of previous studies due to sparse sampling and the inherent biases associated with the use of mtDNA data. The results suggest that all sampled populations of three-spined sticklebacks share a very recent ancestry, and are the result of multiple waves of colonization that started from the Eastern Pacific in the Late Pleistocene (< 80 kya), much more recently than previous estimates based on mtDNA data (Mäkinen and Merilä, 2008; Orti et al., 1994). Northern European as well as eastern American populations were likely recolonized from marine populations in western Europe advancing along the margin of the retreating ice sheet, rather than from a northern range expansion of Balkan populations as previously suggested (DeFaveri et al., 2012; Hewitt, 2000). Furthermore, the discovery that the Black Sea lineage forms a monophyletic group with populations from the Aegean Sea and is very closely related to the Southern Adriatic lineage disproves the hypothesis of a pre-Pleistocene colonization of the Black

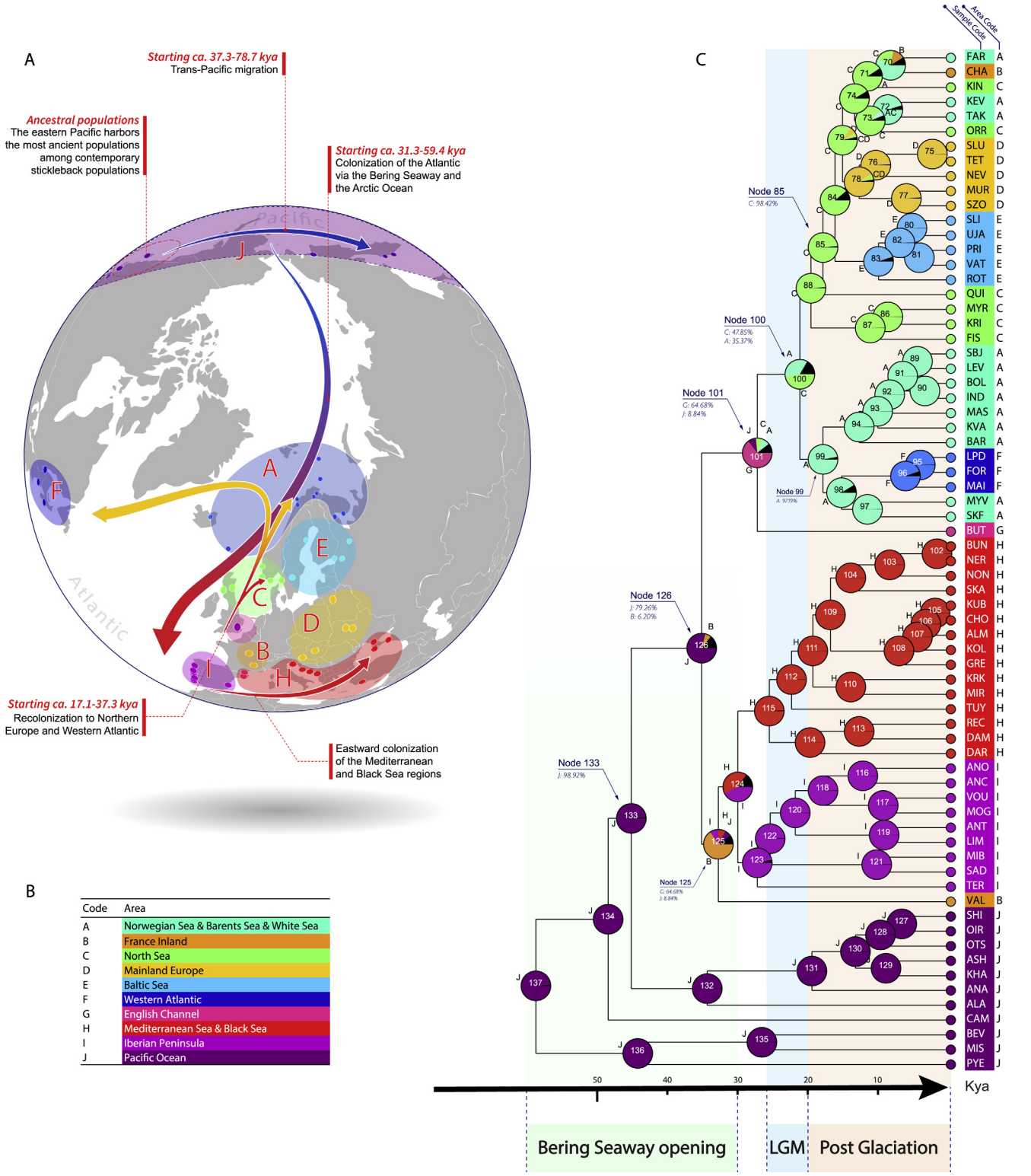


Fig. 3. Reconstruction of ancestral distribution using RASP. (A) The schematic globe shows the sampling localities (colored dots) and hypothesized major colonization routes inferred in this study. Shaded colors and letters indicate ten distribution ranges considered in the analysis. (B) The key indicating the 10 distribution ranges, corresponding coded letters (A to J) and colors in this figure. (C) The results of the Bayesian Binary MCMC (BBM) analysis. Pie charts illustrate the probability of distribution areas of the most recent common ancestor (MRCA). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Sea from the Atlantic through the Baltic Sea and northeastern Europe (Crivelli and Britton, 1987; Münzing, 1963; Sanz et al., 2015) and supports the view that the Black Sea was colonized very recently from the Mediterranean through the Bosphorous as suggested by Mäkinen et al. (2006) and Mäkinen and Merilä (2008), settling a decade-long controversy. This result is unlikely to be driven by a constraint imposed by the calibration point, as preliminary analyses with no time calibration revealed that the Black Sea populations formed a monophyletic lineage nested within the Mediterranean clade. Hence the results presented in this study change our understanding of the evolutionary history of three-spined stickleback populations, particularly in Europe and the Eastern Atlantic. The results suggest that molecular dating based on long-time phylogenetic rates of mtDNA in past studies led to greatly overestimated divergence times among major *G. aculeatus* lineages, a hypothesis first discussed by Mäkinen and Merilä (2008) but never tested.

4.1.1. Pacific region

Our results confirm the Eastern Pacific lineage is the sister lineage of other Pacific lineages, which coincides with the mtDNA Trans-North-Pacific (TNP) lineage *sensu* Lescak et al. (2015) first identified by Orti et al. (1994), but suggest that geographic expansion from ancestral populations in the Eastern Pacific occurred in the Late rather than Middle Pleistocene. The ancestors of present-day three-spined sticklebacks likely originate from a refuge in the Eastern Pacific Basin, from which dispersal to the west occurred between ca. 78.7–37.3 kya, in the Late Pleistocene but before the LGM. During the following westward colonization course, the three-spined sticklebacks reached the Bering Sea region and likely migrated to the Atlantic Ocean following a Trans-Arctic connection (ca. 59.4–31.3 kya) through the Bering Seaway (Bering Strait) during its opening period (ca. 60–30 kya) preceding the LGM.

This interpretation postulates a far more recent colonization scenario than previously hypothesized. The Bering Seaway opened around 3 mya creating the possibility for marine fauna to colonize the Atlantic from the Pacific Ocean via a trans-Arctic route (Herman and Hopkins, 1980; Laakkonen et al., 2013; Orti et al., 1994; Väinölä, 2003). According to fossil evidence, the earliest migration of three-spined sticklebacks from the Pacific to the Atlantic could be traced back to the late Tertiary or Early Pleistocene (Bell and Foster, 1994), approximately 2 mya. Orti et al. (1994) dated the colonization from the Pacific to the Atlantic ca. 90–260 kya using mtDNA data, and suggested a more recent replacement of the original Atlantic mitochondrial lineage as a result of local extinctions or severe bottleneck followed by a more recent invasion from the Pacific.

Orti et al. (1994) used cytochrome b sequence data and a molecular clock calibration based on the divergence between three- and nine-spined sticklebacks, which initiated ca. 10 mya (Bell and Foster, 1994). As it is now widely recognized, mutation rates can be strongly time-dependent, a phenomenon known as heterotachy: rapid short-term mutation rates decay exponentially to long-term phylogenetic rates over periods of hundreds of thousands to millions of years (Burridge et al., 2008; Ho and Larson, 2006). This phenomenon, in part due to the disappearance of transient polymorphism (see review in Ho et al. (2011)), has important implications when reconstructing recent evolutionary events, as the use of distant calibrations based on fossil records might lead to greatly overestimate divergence times. Mäkinen and Merilä (2008), who also used the divergence of three- and nine-spined sticklebacks to calibrate the molecular clock in their study, discussed these limitations.

Within the Pacific, the Alaskan population (ALA) on Kodiak Island diverged prior to the rest of the Western Pacific lineages, suggesting that the three-spined sticklebacks firstly reached the Northern Pacific from the east. Russian populations in the Bering Sea and Sea of Okhotsk (ANA, KHA, ASH), and the Japanese populations (SHI, OIR, OTS) are younger and share a recent ancestry with the Kodiak Island population, suggesting both phylogenetic groups are the result of a westward colonization from ancestral populations in the Eastern Pacific. The Atlantic basin lineages are phylogenetically closer to the Western Pacific lineages than the Eastern Pacific ones. This pattern is in contrast with the previous findings of a Japanese mtDNA clade (Orti et al., 1994), where sticklebacks in Japan (major Japanese clade in that study) were more divergent from the Atlantic basin sticklebacks. In support of our suggestion, Colosimo et al. (2005) identified two divergent lineages across the trans-Pacific region using 193 SNPs from 25 random nuclear loci, and indicated the Western Pacific lineage is closely related to the Atlantic sticklebacks.

4.1.2. Europe and the Atlantic region

The phylogeny of European and Atlantic three-spined sticklebacks suggests that the latest colonization of both Northern Europe and the Western Atlantic followed the end of the last glaciation. The biogeographic analyses indicate that the three-spined stickleback colonization of Northern Europe and the Western Atlantic occurred most likely from southern European refugia. The divergence of Southern European, Northern European and Eastern Atlantic lineages is also more recent than previous studies suggested. Mäkinen and Merilä (2008) estimated that major lineages within the trans-Atlantic clade diverged roughly 130–170 kya, much earlier than we estimated in our study (45–25 kya). Mäkinen and Merilä (2008) admitted that with a more recent calibration point (between 1 and 2 Mya, as we have used in this study), the substitution rate could have been faster and the divergence time of the main lineages more recent. They noted that if the short-term substitution rate for cytochrome b was approximately 0.1/Mya (Mäkinen and Merilä, 2008) – a more likely short-term substitution rate for this gene (Burridge et al., 2008) – then different lineages within the Atlantic clade would have diverged between 32 and 25 kya, congruent with our findings.

The nuclear phylogeny uncovered in this study differs in topology from previous studies from the Atlantic and Mediterranean regions. In southern Europe, two independent Adriatic lineages as well as a distinct Black Sea lineage were identified, consistent with the two independent Adriatic lineages in mtDNA phylogenies found by DeFaveri et al. (2012) and Mäkinen and Merilä (2008). However, there are several incongruences between our phylogenetic reconstruction and previously published mtDNA phylogenies. For example, in our tree both Adriatic lineages are grouped in the Southern European Clade while in the previous mtDNA tree the northern Adriatic sublineage belonged to the present Trans-Atlantic Clade (the European lineage *sensu* Mäkinen and Merilä (2008)). Our phylogeny also supports the hypothesis that the Black Sea was recently colonized from the Mediterranean (Mäkinen et al., 2006; Mäkinen and Merilä, 2008), and refutes the hypothesis of a pre-Pleistocene colonization of the Black Sea from the Atlantic through North-Eastern Europe, as suggested by Sanz et al. (2015) on the basis of mtDNA data.

The scenario of colonization of Northern European populations we present in this study is also different from previous hypotheses based on mtDNA data. For example, noticing that based on mtDNA data northern Adriatic populations were similar to Northern European lineages, DeFaveri et al., (2012) hypothesized that Northern Europe might have been re-colonized following the LGM via a northern expansion of Balkan ancestral populations, as it has been suggested for terrestrial fauna and flora (Buj et al., 2008; Cooper et al., 1995; Demesure et al., 1996; Konnert and Bergmann, 1995; Ursenbacher et al., 2008). Our results, however, suggest that colonization of Northern Europe was achieved by an expansion of marine populations (see below).

Such discrepancy between phylogenetic reconstruction from nuclear data and mtDNA haplotypes are common (Toews and Brelsford, 2012), and often reflect the fact that mtDNA is inherited as single linkage group providing only one independent estimate of the species tree. This issue is particularly important when there is incomplete lineage sorting and/or gene flow and is exacerbated by the higher

genetic drift due to lower effective size of the mitochondrial genome and sex-biased dispersal (Hoelzer, 1997).

Of particular interest is the finding that the Western Atlantic populations share a recent ancestry with Northern European populations, suggesting a very recent colonization from Europe by way of surface current circulation in the North Atlantic during the Holocene (12.4–4.5 kya). The Trans-Atlantic Clade includes many closely related populations from a wide range of open sea regions, which covers the northwestern Atlantic, the Norwegian Sea, the Barents Sea and the White Sea. The ability of long-distance migration of *G. aculeatus*, even across the oceans, has been confirmed by phylogenetic analysis and genetic evidence that identical haplotypes appear on opposite sides of the Pacific ocean (Johnson and Taylor, 2004; Mäkinen and Merilä, 2008; Orti et al., 1994; Vila et al., 2017). The trans-Atlantic subclade diverged ca. 27–14 kya, at the end of the LGM period at a time when the surface glacial North Atlantic Currents in the central North Atlantic were flowing counterclockwise (Seidov et al., 1996). When they reached the western margin of Europe and eastern margin of North America, there was a northward flow towards the Norwegian Sea area in the west of Ireland (Seidov et al., 1996). The existence of a monophyletic Trans-Atlantic Clade including Northern European as well as Eastern American populations, along with the proven long-distance dispersal ability of the species and the fact that past current patterns provided a plausible mean of dispersal and colonization strongly suggest that the Norwegian, Barents and White Sea as well as North American populations were colonized from the western margin of Europe following the end of the last glaciation periods via dispersal along oceanic currents.

Eastern North Sea sticklebacks could have been the first to colonize the modern North Sea and Baltic Sea region, since they were identified as the earliest diverging lineage in the North Sea-Baltic Sea subclade (diverged ca. 25.9–13.9 kya). This interpretation is concordant with the physical geography of the North Sea region during later Weichselian glaciation and Bølling-Allerød interglaciation (before ca. 13 kya), when there was just a Norwegian Trench (nowadays Eastern North Sea part) surrounding the southern border of Norway and above the Doggerland landmass (Coles, 2000). It is plausible that three-spined sticklebacks initially reached the North Sea area through either the Norwegian Sea or continental river catchments flowing to the North Sea basin.

The biogeographic analyses suggest that the ancestral area for the Trans-Atlantic Clade was around the English Channel, and that for the Southern European Clade in the Iberian Peninsula and inland France. It is noteworthy that the British Isles and France went through major geological changes during the Pleistocene. Before 17 kya there was a landmass, known as 'Doggerland', connecting the British Isle to continental Europe across the North Sea (Gaffney et al., 2007). The location around the English Channel harboring a population (BUT; 51°43′43″N, 0°9′38″E) was not covered by the ice cap during the LGM according to the reconstructions of the maximum extent of the Fennoscandian ice-sheet (Boulton and Hagdorn, 2006). The southern British Isle was once connected to northern France, and all the indigenous freshwater fishes are supposed to have invaded England during the existence of the Doggerland landmass through a connected river system from the southern part of continental Europe, or survived in refugia (Ravinet et al., 2014; Wheeler, 1977).

Refugia for *G. aculeatus* and other taxa have been identified across southern Europe (Hewitt, 2000). In our phylogeny, the most ancient population (VAL) in the Southern European Clade diverged from the rest of South European lineages ca. 42.6–23.4 kya, and resides in the Le Rhône River drainage in southern France, a known refuge during the last glaciation (DeFaveri et al., 2012; Mäkinen and Merilä, 2008; Sanz et al., 2015). Moreover, plenty of studies have suggested multiple colonization events in southern Europe, which is also corroborated by the latest mtDNA phylogeny studies (Lescak et al., 2017; Vila et al., 2017). In particular, Vila et al. (2017) detected a haplotype (H19) which appeared in both northwestern Spain and the British Isles, but did not determine an explicit dispersal direction for this haplotype. On the whole, we suggest that ancestral three-spined sticklebacks survived in the area of English Channel and France through the LGM and dispersed thereafter across the Atlantic and towards northern Europe.

Interestingly, the Trans-Atlantic Clade also includes a lineage comprising of freshwater populations from mainland Europe, and a group of populations from multiple scattered regions. This complicated phylogenetic pattern in the post-glacial clade might be accounted by repeated recolonization events, as well as by numerous glacial refugia sources or ancestral population origins for present populations. The phenomenon of complex and shallow phylogenetic relationships in post-glacial lineages has also been noted in previous studies, as for instance in a limited geographic region like Switzerland (Lucek et al., 2010) and across vast regions of Europe (Sanz et al., 2015). Accordingly, it is not surprising that a freshwater population in France (CHA), which would be expected to belong to the Southern Europe Clade, clusters in this mixed group. This might be explained by a post-glacial migration event from the south.

4.2. Analytical considerations

Our analyses suggest substantially more recent divergence among major three-spined lineages than indicated by previous studies. However, divergence time estimates are only as good as the underlying data, and various unaccounted factors can influence them.

First, there may be heterogeneity in nucleotide substitution rates among different lineages (Martin and Palumbi, 1993; Weir and Schluter, 2008), owing, for instance, to differences in their physiology and life history. Generation length in southern and northern three-spined sticklebacks are known to differ (Baker, 1994; DeFaveri and Merilä, 2013), but unfortunately, the generation time for most of our study populations is unknown. Nevertheless, to reduce the possible effects of variation in generation time to the time calibration, we adopted the relaxed-clock phylogenetic model in BEAST2 to estimate lineage-specific substitution rates. Hence, variations in evolutionary rates among lineages were accounted for in phylogeny construction.

Second, also complex population histories involving gene flow and admixture could affect genetic distances and divergence time estimation (McKinnon and Rundle 2002). In fact, cases of introgression have been reported in three-spined stickleback (Lescak et al., 2015; Liu et al., 2016; Yamada et al., 2001). However, divergence in the presence of gene flow is a general challenge for constructing calibrated phylogenies (Nosil, 2008). Incorporating genetic data from multiple nuclear loci can help to reduce the bias due to this effect (Hare, 2001), at least compared to single locus trees. While relaxed clock Bayesian analyses of concatenated data are robust to issues such as incomplete lineage sorting (Drummond et al., 2006; Lambert et al., 2015), the possibility that introgression is affecting our results cannot be completely discarded. However, gene flow between the major phylogenetic lineages (Pacific *vs.* Europe/Atlantic) seems very unlikely. Previous studies have demonstrated complete mitochondrial lineage sorting (Orti et al., 1994) and very high F_{ST} estimated from microsatellite data (Mäkinen et al., 2006) among the major lineages, suggesting that large-scale gene flow between the Pacific and the Atlantic, or between Southern and Northern European populations is unlikely. Hence, our divergence time estimates of major lineages are unlikely to be affected by recent or contemporary gene flow. Within the southern European lineage, only freshwater populations exist today. Hence gene flow among these locations and recent events of introgression can be discarded. The main regions where introgression could affect phylogenetic reconstruction are Northern Europe and the north Pacific. In both regions we have some of the least supported nodes in our phylogeny (Fig. S1). Lambert et al. (2015) noted that relationships strongly affected by incomplete lineage sorting usually appear as weakly supported nodes in concatenated trees, particularly in short branches. Previous studies have uncovered extensive patterns of admixture between the Eastern and Western Pacific lineages, particularly in the northern populations of

British Columbia and Alaska (Lescak et al., 2015). In this light, it seems at least possible that the phylogenetic position of two populations in our tree (CAM, British Columbia and ALA, Alaska, Fig. 2), which form two groups separated from both the Eastern and Western Pacific lineages and somehow intermediate between them, are in part a reflection of introgression between the Eastern and Western Pacific lineages.

5. Conclusions

In conclusion, our analyses based on genome-wide SNPs and worldwide sampling of populations resulted in a robust phylogenetic hypothesis of evolutionary relationships among sampled three-spined stickleback populations. The results confirm the hypothesis that the area of origin for this species resides in the Eastern Pacific, but suggest that geographic expansion was far more recent than previously thought. After dispersal to the Western Pacific, the species likely colonized Europe through the Arctic Ocean before the LGM, and subsequently (and post-glacially) recolonized Northern Europe and the North American East Coast. Within the three main clades identified, most of the populations were estimated to have diverged post-glacially, and even the divergence times between main clades appear to be more recent than suggested by previous studies. Furthermore, we resolved controversies over hypothesis of colonizations of both the Baltic Sea and the Black Sea, which we suggest to stem from inherent limitations of the mtDNA data that were used in previous studies.

6. Data availability

The short sequence reads (RAD sequences) used in this study have been deposited in the GenBank Short Read Archive (Bioproject PRJNA453151, Short Read Archive SRP142676, Accessions SRR7067148- SRR7067275). VCF file and the alignments used for phylogenomics analyses, *ipyrad* parameter file and phylogenetic trees from this study are accessible via the DRYAD digital repository (doi: 10.5061/dryad.2529hr1).

Acknowledgments

We thank Jacquelin DeFaveri and two anonymous referees for useful feedback on earlier versions of this manuscript, and Laura Hänninen, Sami Karja and Kirsi Kähkönen for help in laboratory work. We thank Abhilash Nair for program operation advice. We are also greatful to the following people who helped in obtaining the samples used in this study: Anders Adill, Windsor Aguirre, Arash Akbarzadeh, Theo Bakker, Alison Bell, Mike Bell, Janis Birzaks, Folmer Bokma, Bertil Borg, Jacquelin DeFaveri, Fredrik Franzén, Jörg Freyhof, Akira Goto, Andrew Hendry, Gabor Herczeg, Frank von Hippel, Aki Hirvonen, Jenni Hämäläinen, Deniz Innal, Markku Kaukoranta, Agnieszka Kijewska, David Kingsley, Yoshinobu Kosaka, Lotta Kvarnemo, Dmitry Lajus, Tuomas Leinonen, Arne Levsen, Scott McCairns, Antoine Millet, Jola Morozinska, Corey Munk, Hannu Mäkinen, Arne Nolte, Kjartan Østbye, Ioanna Papakosta, Wäinö Pekkola, Jouko Pokela, Mark Ravinet, Katja Räsänen, Radek Šanda, Nuria Sanz, Dolph Schluter, Mat Seymor, Takahito Shikano, Per Sjöstrand, Garrett Staines, Björn Stelbrink, Ilkka Syvänperä, Anti Vasemägi, Mike Webster, James Willacker, Helmut Winkler, Linda Zanella, and Linda Zaveik. Our research was supported by Chinese Scholarship Council (grant 201606270188 to BF), the Finnish Cultural Foundation (grant to PM) and the Academy of Finland (grants 129662, 134728, and 218343 to JM). This work was also partially supported by the Portuguese Science Foundation through the strategic plan of MARE – Marine and Environmental Science Centre (UID/MAR/04292/2013) and through a post-doctoral grant to CMA (SFRH/BPD/108582/2015).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.06.008.

References

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17, 81–92.
Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., 2011. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. 61, 170–173.
Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3, e3376.
Baker, J., 1994. Life history variation in female threespine stickleback. The evolutionary biology of the threespine stickleback. Oxford Univ. Press.
Ballard, R.D., Coleman, D.F., Rosenberg, G., 2000. Further evidence of abrupt Holocene drowning of the Black Sea shelf. Mar. Geol. 170, 253–261.
Barrett, R.D., Paccard, A., Healy, T.M., Bergek, S., Schulte, P.M., Schluter, D., Rogers, S.M., 2011. Rapid evolution of cold tolerance in stickleback. Proc. Biol. Sci. 278, 233–238.
Bell, A.M., 2001. Effects of an endocrine disrupter on courtship and aggressive behaviour of male three-spined stickleback, *Gasterosteus aculeatus*. Anim. Behav. 62, 775–780.
Bell, M.A., Foster, S.A., 1994. The Evolutionary Biology of the Threespine Stickleback. Oxford University Press.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 10, e1003537.
Boulton, G., Hagdorn, M., 2006. Glaciology of the British Isles Ice Sheet during the last glacial cycle: form, flow, streams and lobes. Quat. Sci. Rev. 25, 3359–3390.
Buj, I., Podnar, M., Mrakovcic, M., Choleva, L., Slechtova, V., Tvrtkovic, N., Caleta, M., Mustafic, P., Marcic, Z., Zanella, D., 2008. Genetic diversity and phylogenetic relationships of spined loaches (genus *Cobitis*) in Croatia based on mtDNA and allozyme analyses. Folia Zool. 57, 71.
Burridge, C.P., Craw, D., Fletcher, D., Waters, J.M., 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. Mol. Biol. Evol. 25, 624–633.
Cano, J.M., Mäkinen, H.S., Leinonen, T., Freyhof, J., Merilä, J., 2008. Extreme neutral genetic and morphological divergence supports classification of Adriatic three-spined stickleback (*Gasterosteus aculeatus*) populations as distinct conservation units. Biol. Conserv. 141, 1055–1066.
Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., McCabe, A.M., 2009. The last glacial maximum. Science 325, 710–714.
Coles, B.J., 2000. Doggerland: the cultural dynamics of a shifting coastline. Geol. Soc., London, Spec. Publ. 175, 393–401.
Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal Jr., G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., Kingsley, D.M., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307, 1928–1933.
Cooper, S., Ibrahim, K., Hewitt, G., 1995. Postglacial expansion and genome subdivision in the European grasshopper Chorthippus parallelus. Mol. Ecol. 4, 49–60.
Cresko, W.A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M.A., Kimmel, C.B., Postlethwait, J.H., 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. U. S. A. 101, 6050–6055.
Crivelli, A.J., Britton, R.H., 1987. Life history adaptations of *Gasterosteus aculeatus* in a Mediterranean wetland. Environ. Biol. Fishes 18, 109–125.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.
de Queiroz, A., Gatesy, J., 2007. The supermatrix approach to systematics. Trends Ecol. Evol. 22, 34–41.
Deagle, B.E., Jones, F.C., Absher, D.M., Kingsley, D.M., Reimchen, T.E., 2013. Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. Mol. Ecol. 22, 1917–1932.
DeFaveri, J., Jonsson, P.R., Merilä, J., 2013. Heterogeneous Genomic Differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. Evolution 67, 2530–2546.
DeFaveri, J., Merilä, J., 2013. Variation in age and size in Fennoscandian three-spined sticklebacks (*Gasterosteus aculeatus*). PLoS ONE 8, e80866.
DeFaveri, J., Zanella, L.N., Zanella, D., Mrakovcic, M., Merilä, J., 2012. Phylogeography of isolated freshwater three-spined stickleback *Gasterosteus aculeatus* populations in the Adriatic Sea basin. J. Fish Biol. 80, 61–85.
DeGiorgio, M., Degnan, J.H., 2009. Fast and consistent estimation of species trees using supermatrix rooted triples. Mol. Biol. Evol. 27, 552–569.
Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.
Demesure, B., Comps, B., Petit, R.J., 1996. Chloroplast DNA phylogeography of the

common beech (*Fagus sylvatica L.*) in Europe. Evolution 50, 2515–2520.
Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
Eaton, D.A., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849.
Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.
Gaffney, V.L., Thomson, K., Fitch, S., 2007. Mapping Doggerland: the Mesolithic Landscapes of the Southern North Sea. Archaeo Press.
Gibson, G., 2005. The synthesis and evolution of a supermodel. Science 307, 1890–1891.
Goldstein, D., Pollock, D., 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. 88, 335–342.
Gordon, A., Hannon, G., 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab.cshl. edu/fastx_toolkit 5.
Guo, B., Chain, F.J., Bornberg-Bauer, E., Leder, E.H., Merilä, J., 2013. Genomic divergence between nine-and three-spined sticklebacks. BMC Genom. 14, 756.
Gökaşan, E., Demirbag, E., Oktay, F.Y., Ecevitog, B., Şimşek, M., Yüce, H., 1997. On the origin of the Bosphorus. Mar. Geol. 140, 183–199.
Haglund, T.R., Buth, D.G., Lawson, R., 1992. Allozyme variation and phylogenetic relationships of Asian, North American, and European populations of the threespine stickleback, *Gasterosteus aculeatus*. Copeia 432–443.
Hare, M.P., 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. 16, 700–706.
Hedrick, P.W., 2007. Sex: differences in mutation, recombination, selection, gene flow, and genetic drift. Evolution 61, 2750–2771.
Hendry, A.P., Peichel, C.L., Matthews, B., Boughman, J.W., Nosil, P., 2013. Stickleback research: the now and the next. Evol. Ecol. Res. 15, 111–141.
Herman, Y., Hopkins, D.M., 1980. Arctic oceanic climate in late Cenozoic time. Science 209, 557–562.
Hewitt, G., 2000. The genetic legacy of the quaternary ice ages. Nature 405, 907.
Higuchi, M., Sakai, H., Goto, A., 2014. A new threespine stickleback, Gasterosteus nipponicus sp. nov. (Teleostei: *Gasterosteidae*), from the Japan Sea region. Ichthyol. Res. 61, 341–351.
Hipp, A.L., Eaton, D.A., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS ONE 9, e93975.
Ho, S.Y., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G., Cooper, A., 2011. Time-dependent rates of molecular evolution. Mol. Ecol. 20, 3087–3101.
Ho, S.Y., Larson, G., 2006. Molecular clocks: when times are a-changin'. Trends Genet. 22, 79–83.
Hoelzer, G.A., 1997. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. Evolution 51, 622–626.
Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6, e1000862.
Hubbard, T., Andrews, D., Cáccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F., 2005. Ensembl 2005. Nucleic Acids Res. 33, D447–D453.
Huntingford, F., Ruiz-Gomez, M., 2009. Three-spined sticklebacks *Gasterosteus aculeatus* as a model for exploring behavioural biology. J. Fish Biol. 75, 1943–1976.
Johnson, L.S., Taylor, E.B., 2004. The distribution of divergent mitochondrial DNA lineages of threespine stickleback (*Gasterosteus aculeatus*) in the northeastern Pacific Basin: post-glacial dispersal and lake accessibility. J. Biogeogr. 31, 1073–1083.
Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484, 55–61.
Kitano, J., Lema, S.C., Luckenbach, J.A., Mori, S., Kawagishi, Y., Kusakabe, M., Swanson, P., Peichel, C.L., 2010. Adaptive divergence in the thyroid hormone signaling pathway in the stickleback radiation. Curr. Biol. 20, 2124–2130.
Kitano, J., Ross, J.A., Mori, S., Kume, M., Jones, F.C., Chan, Y.F., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., 2009. A role for a neo-sex chromosome in stickleback speciation. Nature 461, 1079.
Konnert, M., Bergmann, F., 1995. The geographical distribution of genetic variation of silver fir (*Abies alba, Pinaceae*) in relation to its migration history. Plant Syst. Evol. 196, 19–30.
Laakkonen, H.M., Lajus, D.L., Strelkov, P., Väinölä, R., 2013. Phylogeography of amphiboreal fish: tracing the history of the Pacific herring Clupea pallasii in North-East European seas. BMC Evol. Biol. 13, 67.
Lambert, S.M., Reeder, T.W., Wiens, J.J., 2015. When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. Mol. Phylogenet. Evol. 82, 146–155.
Leache, A.D., Banbury, B.L., Linkem, C.W., de Oca, A.N., 2016. Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in north american spiny lizards (Genus *Sceloporus*)? BMC Evol. Biol. 16, 63.
Leinonen, T., Cano, J., Mäkinen, H., Merilä, J., 2006. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. J. Evol. Biol. 19, 1803–1812.
Lescak, E.A., Marcotte, R.W., Kenney, L.A., von Hippel, F.A., Cresko, W.A., Sherbick, M.L., Colgren, J.J., Andrés López, J., Riddle, B., 2015. Admixture of ancient mitochondrial lineages in three-spined stickleback populations from the North Pacific. J. Biogeogr. 42, 532–539.
Lescak, E.A., Wund, M.A., Bassham, S., Catchen, J., Prince, D.J., Lucas, R., Dominguez, G., von Hippel, F.A., Cresko, W.A., 2017. Ancient three-spined stickleback (*Gasterosteus aculeatus*) mtDNA lineages are not associated with phenotypic or nuclear genetic variation. Biol. J. Linn. Soc. 579–588.
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
Lischer, H.E., Excoffier, L., 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299.
Liu, L., Jin, X., Chen, N., Li, X., Li, P., Fu, C., 2015. Phylogeny of Morella rubra and Its Relatives (*Myricaceae*) and Genetic Resources of Chinese Bayberry Using RAD Sequencing. PLoS ONE 10, e0139840.
Liu, L., Xi, Z., Davis, C.C., 2014. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32, 791–805.
Liu, S., Hansen, M.M., Jacobsen, M.W., 2016. Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. Mol. Ecol. 25, 5187–5202.
Longo, G., Bernardi, G., 2015. The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. Mol. Phylogenet. Evol. 88, 55–63.
Lucek, K., Roy, D., Bezault, E., Sivasundar, A., Seehausen, O., 2010. Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in Switzerland. Mol. Ecol. 19, 3995–4011.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Martin, A.P., Palumbi, S.R., 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. 90, 4087–4091.
Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2 1, 2013.
McKinnon, J.S., Rundle, H.D., 2002. Speciation in nature: the threespine stickleback model systems. Trends Ecol. Evol. 17, 480–488.
Mirarab, S., Bayzid, M.S., Warnow, T., 2014. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380.
Münzing, J., 1963. The evolution of variation and distributional patterns in European populations of the three-spined stickleback, *Gasterosteus aculeatus*. Evolution 17, 320–332.
Mäkinen, H.S., Cano, J.M., Merilä, J., 2006. Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. Mol. Ecol. 15, 1519–1534.
Mäkinen, H.S., Merilä, J., 2008. Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe-evidence for multiple glacial refugia. Mol. Phylogenet. Evol. 46, 167–182.
Natri, H.M., Shikano, T., Merilä, J., 2013. Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes. Mol. Biol. Evol. 30, 1131–1144.
Nilsson, J., Gross, R., Asplund, T., Dove, O., Jansson, H., Kelloniemi, J., Kohlmann, K., Löytynoja, A., Nielsen, E.E., Paaver, T., 2001. Matrilinear phylogeography of Atlantic salmon (*Salmo salar L.*) in Europe and postglacial colonization of the Baltic Sea area. Mol. Ecol. 10, 89–102.
Nosil, P., 2008. Speciation with gene flow could be common. Mol. Ecol. 17, 2103–2106.
Ogilvie, H.A., Heled, J., Xie, D., Drummond, A.J., 2016. Computational performance and statistical accuracy of* BEAST and comparisons with other methods. Syst. Biol. 65, 381–396.
Orti, G., Bell, M.A., Reimchen, T.E., Meyer, A., 1994. Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. Evolution 48, 608–622.
Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.
Ravinet, M., Harrod, C., Eizaguirre, C., Prodohl, P.A., 2014. Unique mitochondrial DNA lineages in Irish stickleback populations: cryptic refugium or rapid recolonization? Ecol. Evol. 4, 2488–2504.
Rivers, D.M., Darwell, C.T., Althoff, D.M., 2016. Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. Cladistics 32, 672–681.
Rogers, J., Gibbs, R.A., 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat. Rev. Genet. 15, 347–359.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Ryan, W., Pitman, W., 2000. Noah's Flood: The new scientific discoveries about the event that changed history. Simon and Schuster.
Sambrook, J., Russell, D.W., 2006. Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harbor Laboratory Press.
Sanz, N., Araguas, R.M., Vidal, O., Viñas, J., 2015. Glacial refuges for three-spined stickleback in the Iberian Peninsula: mitochondrial DNA phylogeography. Freshwater Biol 60, 1794–1809.
Schaffner, S.F., 2004. The X chromosome in population genetics. Nat. Rev. Genet. 5, 43.
Seidov, D., Sarnthein, M., Stattegger, K., Prien, R., Weinelt, M., 1996. North Atlantic ocean circulation during the last glacial maximum and subsequent meltwater event: A numerical model. J. Geophys. Res. Oceans 101, 16305–16332.
Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428, 717–723.
Sunnucks, P., Hales, D.F., 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: *Aphididae*). Mol. Biol. Evol. 13, 510–524.
Takahashi, T., Nagata, N., Sota, T., 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80, 137–144.
Takahata, N., 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122, 957–966.
Tariel, J., Longo, G.C., Bernardi, G., 2016. Tempo and mode of speciation in Holacanthus

angelfishes based on RADseq markers. Mol. Phylogenet. Evol. 98, 84–88.
Toews, D.P., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21, 3907–3930.
Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Orti, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr. 7.
Ursenbacher, S., Schweiger, S., Tomović, L., Crnobrnja-Isailović, J., Fumagalli, L., Mayer, W., 2008. Molecular phylogeography of the nose-horned viper (*Vipera ammodytes*, Linnaeus (1758)): evidence for high genetic diversity and multiple refugia in the Balkan peninsula. Mol. Phylogenet. Evol. 46, 1116–1128.
Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22, 787–798.
Weir, J., Schluter, D., 2008. Calibrating the avian molecular clock. Mol. Ecol. 17, 2321–2328.
Wheeler, A., 1977. The origin and distribution of the freshwater fishes of the British Isles. J. Biogeogr. 1–24.
Vila, M., Hermida, M., Fernandez, C., Perea, S., Doadrio, I., Amaro, R., San Miguel, E., 2017. Phylogeography and Conservation Genetics of the Ibero-Balearic Three-Spined Stickleback (*Gasterosteus aculeatus*). PLoS ONE 12, e0170685.
Von Hippel, F., 2010. Tinbergen's legacy in behaviour: sixty years of landmark stickleback papers. Brill.
Väinölä, R., 2003. Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the Macoma balthica complex. Mar. Biol. 143, 935–946.
Yamada, M., Higuchi, M., Goto, A., 2001. Extensive introgression of mitochondrial DNA found between two genetically divergent forms of threespine stickleback, *Gasterosteus aculeatus*, around Japan. Environ. Biol. Fishes 61, 269–284.
Yu, Y., Harris, A., He, X., 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. Mol. Phylogenet. Evol. 56, 848–850.
Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87, 46–49.
Zhou, W., Ji, X., Obata, S., Pais, A., Dong, Y., Peet, R., Xiang, Q.-Y.J., 2018. Resolving relationships and phylogeographic history of the Nyssa sylvatica complex using data from RAD-seq and species distribution modeling. Mol. Phylogenet. Evol.
Östlund-Nilsson, S., Mayer, I., Huntingford, F.A., 2006. Biology of the Three-Spined Stickleback. CRC Press.