MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 4061–4076 doi: 10.1111/j.1365-294X.2010.04801.x

Phylogeography of ninespine sticklebacks (*Pungitius pungitius*) in North America: glacial refugia and the origins of adaptive traits

JACLYN T. ALDENHOVEN, MATTHEW A. MILLER, PATRICE SHOWERS CORNELI and MICHAEL D. SHAPIRO

Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT, 84112, USA

Abstract

The current geographical distribution of the ninespine stickleback (*Pungitius pungitius*) was shaped in large part by the glaciation events of the Pleistocene epoch (2.6 Mya–10 Kya). Previous efforts to elucidate the phylogeographical history of the ninespine stickleback in North America have focused on a limited set of morphological traits, some of which are likely subject to widespread convergent evolution, thereby potentially obscuring relationships among populations. In this study, we used genetic information from both mitochondrial DNA (mtDNA) sequences and nuclear microsatellite markers to determine the phylogenetic relationships among ninespine stickleback populations. We found that ninespine sticklebacks in North America probably dispersed from at least three glacial refugia—the Mississippi, Bering, and Atlantic refugia—not two as previously thought. However, by applying a molecular clock to our mtDNA data, we found that these three groups diverged long before the most recent glacial period. Our new phylogeny serves as a critical framework for examining the evolution of derived traits in this species, including adaptive phenotypes that evolved multiple times in different lineages. In particular, we inferred that loss of the pelvic (hind fin) skeleton probably evolved independently in populations descended from each of the three putative North American refugia.

Keywords: glacial refugia, phylogeography, population genetics—empirical, *Pungitius*, stickleback

Received 23 March 2010; revision received 9 July 2010; accepted 19 July 2010

Introduction

The glaciation events of the Pleistocene epoch (2.6 Mya–10 kya) played a major role in the dispersion and evolutionary history of many northern temperate fishes. Expanding ice sheets excluded many species from large parts of their ranges, and others were forced into refugia at the glacial margins (Hewitt 2000). In North America, as the continental glaciers melted at the beginning of the current interglacial ~8000–15 000 years ago (Bernatchez & Wilson 1998), large, interconnected proglacial lakes formed and allowed the dispersal of fish over vast geographical regions. New trophic, locomotor, and physiological opportunities led to rapid adaptive radiations in many fish lineages, including the sticklebacks (family Gasterosteidae) (e.g. Bell & Foster 1994; Orti *et al*. 1994; Ward & McLennan 2009).

The ninespine stickleback (*Pungitius pungitius*) is a small (typically 4–6 cm standard length), euryhaline fish with a circumpolar distribution in the northern hemisphere, including freshwater habitats and northern coasts of North America and northern Eurasia. This species exhibits striking morphological variation throughout its range, including differences in the numbers of dorsal spines, lateral plates, and gill rakers; body size and shape; the presence or absence of the pelvic (hind) fin complex; and several behavioural and physiological traits (McPhail 1963a; Nelson 1971; Gross 1979; Ayvazian & Krueger 1992; Blouw & Boyd 1992;

Correspondence: Michael D. Shapiro, Fax: +18015814668; E-mail: shapiro@biology.utah.edu

Ziuganov & Zotin 1995; Gonda *et al.* 2009; Herczeg *et al.* 2009b; Kendal *et al.* 2009; Waser *et al.* 2010). This enormous diversity has captured the attention of biologists for well over a century (Day 1878; Morris 1958; McPhail 1963a; Nelson 1971; Wootton 1976, 1984) and has led to the emergence of the ninespine stickleback as a model for comparative genetics and convergent evolution (Shapiro *et al.* 2006, 2009; Herczeg *et al.* 2009a).

A long-standing hypothesis for the postglacial colonization of North America by ninespine sticklebacks was proposed by McPhail (1963a) who, building on earlier work by Walters (1955), suggested that modern populations dispersed from distinct Bering and Mississippi glacial refugia. He hypothesized that fish from the Bering refugium (located near modern day western Alaska) dispersed around the coastal margins of North America, including Alaska and the Aleutian Islands, and eastward through the Arctic Ocean to the Atlantic coast. In contrast, sticklebacks trapped in the freshwater Mississippi refugium (located south of the modern Great Lakes, in the upper Mississippi valley) lost their ability to migrate through saltwater, and therefore dispersed exclusively into lakes and streams throughout the continental interior, but not into coastal marine habitats. This hypothesis was inferred from morphological analysis, especially numbers of dorsal spines and gill rakers. However, the use of morphological traits is often inconsistent with information obtained from genetic markers (Hansen *et al.* 1999), as different lineages may exhibit similar morphological features as a result of adaptation to similar habitats, not shared ancestry (Lindsey 1981; Orti *et al.* 1994). While McPhail noted that the differences between these two forms could have a genetic basis, the appropriate genetic resources were not available to critically test this hypothesis.

In this study, we present a phylogeographical and population genetic analysis of *Pungitius* across the northern hemisphere, with a focus on North American populations. We sampled ninespine sticklebacks from several populations across North America, northern Europe, and the Kuril Islands (East Asia) and then assessed genetic variation and phylogenetic relationships among populations, using mitochondrial DNA (mtDNA) control region (CR) sequences and nuclear microsatellite genotypes. This combined approach of using mtDNA sequence and multiple, neutral, genetic markers allowed us to resolve genetic signatures of both long-term isolation and recent postglacial expansion of species into previously glaciated regions.

Materials and methods

Sample collection

Ninespine sticklebacks (*Pungitius pungitius*) were sampled from 23 locations in North America. Samples from nine locations in Ireland, Sweden, and the Kuril Islands (East Asia) were also included to infer relationships with coastal Eurasian populations (Fig. 1, Table 1). Fish were preserved in ethanol, and DNA was extracted from either the pectoral fins or liver using a standard phenol chloroform method.

Mitochondrial DNA amplification

We used PCR to amplify and sequence a 1100-bp fragment of mtDNA CR from 169 ninespine sticklebacks using previously published primers (L-Thr and H-12S (Takahashi & Goto 2001)). PCR was performed in a PTC-200 DNA Engine thermocycler (BioRad) in 50 μl reactions, using 10 μl buffer (HF, Finnzymes), 0.1 mM dNTPs, 0.2 μM each of forward and reverse primer, 1 U Phusion Taq polymerase (Finnzymes), and 20 ng of genomic DNA. Thermal cycling consisted of 98 °C for 2 min; 30 cycles of 98 °C for 10 s, 56 °C for 30 s, 72 °C for 30 s, and 72 °C for 7 min. PCR products were excised from a 1% agarose gel and purified (QIAEX II Gel Extraction Kit; Qiagen) for direct sequencing. We designed additional primers internal to the amplified fragment to sequence through repeat regions (INTF 5′-TTGTCAGCTAAACCCGTGTG, INTR 5′-TGGTCG-GTTCTCGGAGTAGT). Sequences were analysed using

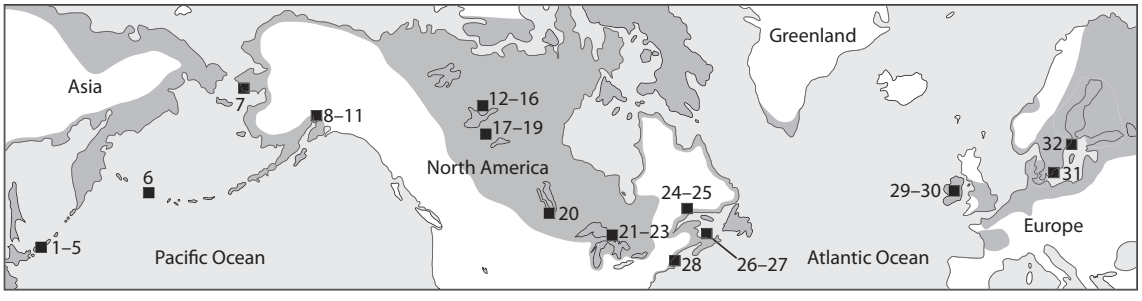


Fig. 1 Location of the 32 ninespine stickleback sample sites. Species range (shaded) is based on Wootton (1976) and McPhail (1963a). Details for each population are listed in Table 1.

Table 1 Sampling information and microsatellite and mtDNA statistics for each of the 32 populations

				Microsatellite			mtDNA				
Population	Symbol	Population no.	Location	N	H_E	N_A	N	No. Haps	π	hd	Fu's Fs
Kuril islands											
Paramushir Island	PM	1	50°01′ N, 155°40′ E	5	0.340 ± 0.288	2.000 ± 1.134	2	1	0	0	–
Polonskogo Island	PS	2	43°63′ N, 146°30′ E	5	0.573 ± 0.229	3.214 ± 1.206	3	1	0	0	–
Shikotan Island	SK	3	43°82′ N, 146°71′ E	10	0.573 ± 0.202	4.500 ± 1.880	6	4	0.0030	0.800 ± 0.172	0.275
Tanfilyeva Island	TF	4	43°44′ N, 145°94′ E	5	0.619 ± 0.287	3.643 ± 1.757	3	3	0.0103	1.000 ± 0.272	1.207
Urup Island	UR	5	45°80′ N, 149°90′ E	5	0.298 ± 0.320	1.786 ± 0.860	2	2	0.0048	1.000 ± 0.500	1.609
Alaska—Aleutian Islands											
Attu Island	AI	6	52°49′ N, 173°10′ E	20	0.186 ± 0.212	1.857 ± 0.990	9	2	0.0002	0.222 ± 0.166	−0.263
Alaska—Mainland											
Nome	NO	7	64°33′ N, 165°25′ W	16	0.627 ± 0.253	5.500 ± 2.872	9	5	0.0040	0.806 ± 0.119	0.751
Matanuska-Susitna Valley											
Point MacKenzie	PT	8	61°18′ N, 150°01′ W	40	0.356 ± 0.247	3.786 ± 2.273	9	4	0.0011	0.694 ± 0.147	−0.722
Mud Lake	ML	9	61°33′ N, 148°56′ W	20	0.332 ± 0.235	3.286 ± 1.868	9	6	0.0015	0.917 ± 0.072	−2.872*
Glenn Highway Slough	GH	10	61°28′ N, 149°18′ W	24	0.458 ± 0.290	4.929 ± 3.494	6	5	0.0015	0.933 ± 0.122	−2.711*
Church Road, Wasilla	CH	11	61°35′ N, 149°30′ W	20	0.105 ± 0.162	1.429 ± 0.623	10	1	0	0	–
Northwest Territories											
Snap Lake	SN	12	63°36′ N, 110°52′ W	2	0.393 ± 0.461	1.214 ± 1.221	2	2	0.0106	1.000 ± 0.500	2.397
Great Slave Lake	SL	13	62°24′ N, 114°20′ W	20	0.439 ± 0.319	4.500 ± 2.946	9	8	0.0034	0.973 ± 0.064	−3.760*
Alexie Lake	AE	14	62°40′ N, 114°9′ W	3	0.428 ± 0.392	2.286 ± 0.939	2	2	0.0048	1.000 ± 0.500	1.609
Drygeese Lake	DL	15	62°44′ N, 114°10′ W	11	0.539 ± 0.221	7.786 ± 2.425	5	5	0.0021	1.000 ± 0.127	−2.862*
Chitty Lake	CL	16	62°42′ N, 114°7′ W	10	0.574 ± 0.260	4.571 ± 2.412	4	4	0.0034	1.000 ± 0.177	−1.012
Fox Holes Lake	FH	17	60°03′ N, 112°27′ W	20	0.507 ± 0.269	4.286 ± 2.462	7	4	0.0013	0.714 ± 0.181	−0.914
Salt River	SR	18	59°57′ N, 112°17′ W	19	0.624 ± 0.159	5.929 ± 2.738	10	5	0.0012	0.756 ± 0.129	−1.758*
Pine Lake	PL	19	59°33′ N, 112°15′ W	18	0.490 ± 0.246	4.000 ± 2.171	7	4	0.0013	0.714 ± 0.181	−0.914
Manitoba											
Delta Marsh	DM	20	50°12′ N, 98°09′ W	2	0.619 ± 0.359	2.357 ± 0.972	2	2	0.0125	1.000 ± 0.500	2.564
Great Lakes											
Lake Huron, Port Dolomite	LH	21	45°59′ N, 84°16′ W	4	0.668 ± 0.275	3.571 ± 1.545	4	4	0.0020	1.000 ± 0.177	−1.741*
Lake Superior	LS	22	46°58′ N, 84°10′ W	2	0.559 ± 0.376	2.143 ± 0.990	1	1	0	1.000 ± 0.000	–
St Marys River, DeTour Village	MR	23	46°01′ W 83°55′ W	2	0.571 ± 0.343	2.286 ± 1.030	2	2	0.0039	1.000 ± 0.500	1.386
St Lawrence River Area											
Rivière St Marguerite	RM	24	48°26′ N, 69°94′ W	1	0.071 ± 0.256	0.929 ± 0.457	1	1	0	0	–
St Lawrence River	ST	25	48°01′ N, 69°35′ W	5	0.691 ± 0.143	3.786 ± 1.423	5	3	0.0010	0.800 ± 0.164	−0.475
Prince Edward Island											
Glenfinnan River	GF	26	46°17′ N, 62°59′ W	26	0.554 ± 0.349	9.929 ± 8.908	10	5	0.0009	0.667 ± 0.163	−2.259*
O'Keefe's Lake	OK	27	46°15′ N, 62°49′ W	22	0.548 ± 0.325	7.500 ± 5.539	7	6	0.0023	0.952 ± 0.096	−2.809*

Table 1 (Continued)

Population	Symbol	Population no.	Location	Microsatellite			mtDNA				
				N	H_E	N_A	N	No. Haps	π	hd	Fu's Fs
Massachusetts											
Little River—Newbury	NB	28	42°45′ N, 70°52′ W	20	0.715 ± 0.264	9.214 ± 3.913	7	6	0.0071	0.952 ± 0.096	−0.424
Ireland											
Lough Ennell	LE	29	53°30′ N, 7°32′ W	19	0.484 ± 0.306	5.643 ± 3.791	4	4	0.0048	1.000 ± 0.177	−0.524
River Brosna	RB	30	53°22′ N, 7°22′ W	4	0.413 ± 0.331	2.357 ± 1.288	4	1	0	0	–
Sweden											
Barsebäckshamn, Öresund	OS	31	55°46′ N, 12°54′ E	10	0.613 ± 0.232	4.786 ± 2.596	5	5	0.0025	1.000 ± 0.126	−2.517*
Askö, Northern Baltic Proper	AS	32	58°49′ N, 17°38′ E	10	0.536 ± 0.204	4.071 ± 1.869	3	3	0.0077	1.000 ± 0.272	0.901

N, number of fish; H_E, expected heterozygosity; N_A, number of microsatellite alleles; π, nucleotide diversity; hd, haplotype diversity.
*$P < 0.05$.

Sequencher v4.9 (Gene Codes Corp., Ann Arbor, MI, USA) and aligned with ClustalW (Larkin *et al.* 2007) in MEGA (Kumar *et al.* 2008).

Mitochondrial DNA analysis

Phylogenetic relationships among CR haplotypes were estimated with a coalescent (skyline) prior, using the Bayesian MCMC algorithm implemented in BEAST v1.5.2 (Drummond & Rambaut 2007). The model was selected first using PAUP 4.0 (Swofford 2002) and methods described by Posada & Crandall (1998). The generalized time reversible plus invariant sites plus gamma (GTR+I+G) and Hasegawa-Kishino-Yano plus invariant sites plus gamma (HKY+I+G) models were not significantly different from each other ($P = 0.20$); therefore, the most comprehensive model (GTR+I+G) was chosen and used with a strict molecular clock. Two runs of 20 million steps were combined after removal of 2 million burn-in steps. No long-term temporal trends were evident in the likelihood and other parameter estimates that were examined using TRACER v1.4 (Rambaut & Drummond 2007). A consensus tree was created with TREEANNOTATOR v1.5.2 (Drummond & Rambaut 2007) and visualized using FIGTREE v1.2.3 (Rambaut 2006). We also inferred a maximum likelihood (ML) tree using PHYML v3.0 (Guindon & Gascuel 2003). Unlike the Bayesian analysis, the GTR+I+G analysis failed to resolve the tree, therefore the HKY+I+G model was used. Support for the ML tree was evaluated from 1000 bootstrap replicates of the data.

We estimated nucleotide substitution rates and divergence times in a subsequent analysis using BEAST v1.5.2 (Drummond & Rambaut 2007). We combined the distinct haplotypes found in this study with 41 Asian haplotypes, including two other *Pungitius* species, *P. tymensis* and *P. sinensis* (Takahashi & Goto 2001) and two outgroups, the brook stickleback (*Culaea inconstans*, Genbank accession number AB445125) and threespine stickleback (*Gasterosteus aculeatus*, AB054361). When using a large number of in-group samples and divergent out-groups, estimated model parameters cannot be used in all parts of the tree because of the variance in the heterogeneous dataset (Guiher & Burbrink 2008). To overcome this issue, we used a combined approach incorporating Yule speciation priors to determine rate and Bayesian skyline priors to determine topology, as a majority of the samples was from a single species (Drummond *et al.* 2005). To date nodes among species, a Yule process prior was used with a reduced data set containing representatives of major lineages (*G. aculeatus, C. inconstans, P. tymensis, P. sinensis*, and North American and Asian *P. pungitius*) (Guiher & Burbrink 2008). The first calibration point was 13.3 Mya, representing the minimum age of diver-

gence between ninespine and threespine sticklebacks based on the fossil record (Bell *et al.* 2009). The second calibration point was the split between ninespine and brook sticklebacks, based on the oldest known fossil of a ninespine stickleback dated to ~7 Mya (Rawlinson & Bell 1982). We used a lognormal distribution for TMRCA, which assumes that the actual divergence date is earlier than the appearance of fossil data (Ho 2007). The full dataset was then used to determine the relationship among all haplotypes. Priors were set as for the first haplotype tree, and the clock rate and height was set as in Yule prior analysis. Likelihoods were checked, and trees were combined and drawn as in the first haplotype tree. Haplotype and nucleotide diversity was determined in Arlequin 3.1 (Schneider *et al.* 2000), and Fu's *Fs* test (Fu 1997) was used to detect population demographic expansions.

Microsatellite amplification

Fourteen microsatellite markers were selected from different linkage groups (LG) across the ninespine stickleback genome (Table 2) (Shapiro *et al.* 2009) to generate multi-locus genotypes for 400 individuals from 32 populations. Forward primers were either labelled with a fluorescent dye or an M13 tag sequence was added to the 5′ end to anneal with a universal fluorescent tag primer (Schuelke 2000; Protas *et al.* 2006). PCR and genotyping were performed as described by Peichel *et al.* (2001) for directly labelled primers or Protas *et al.* (2006) for the M13 primer scheme. Allele sizes were determined manually using GeneMapper v3.7 (Applied Biosystems).

Table 2 Linkage group (LG) location, expected heterozygosity (H_E), number of alleles (N_A), and F_{ST} of nuclear microsatellite markers

Locus	LG	H_E	N_A	F_{ST}
Pun44	14A	0.832	62	0.500
Pun117	19	0.792	39	0.441
Pun171	13	0.878	28	0.309
Pun68	8	0.826	24	0.467
Pun203	14B	0.899	26	0.329
Pun255	12	0.855	19	0.535
Pun19	5A	0.827	25	0.562
Pun157	3	0.785	22	0.521
Pun212	17	0.912	32	0.439
Pun261	16	0.808	26	0.415
Pun78	7B	0.922	31	0.285
Pun134	1B	0.860	31	0.576
Pun238	9A	0.668	13	0.452
Stn433	4	0.885	16	0.586
Mean		0.839	28.14	0.458
SD		0.063	11.43	0.098

Microsatellite data analysis

Estimates of allelic diversity and expected and observed heterozygosity and tests for linkage disequilibrium and Hardy–Weinberg equilibrium were performed for each population using Arlequin v3.1 (Schneider *et al.* 2000). The level of genetic differentiation, F_{ST} (Reynolds *et al.* 1983; Slatkin 1995), based on allele frequency distributions, between all pairs of populations was estimated and tested for significance. Negative F_{ST} values were constrained to zero. Genetic structuring was assessed using an analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) with populations partitioned by geography, habitat (marine or freshwater), and lineage. Mantel tests (Smouse *et al.* 1986) were also performed to determine the relationship between F_{ST} values and geographical distance between geographical groups/lineages.

Genetic population structure was inferred using the clustering method in Structure (Pritchard *et al.* 2000). The number of putative populations (K) from 1 to 20 was evaluated with admixture allowed, and no *a priori* information for individuals. A 50 000-iteration burn-in followed by 500 000 iterations was used, with each K repeated three times to ensure stability.

Phylogenetic relationships among populations were estimated by neighbour-joining with chord distances D_{CE} (Cavalli-Sforza & Edwards 1967; Takezaki & Nei 1996). Analyses were performed using the subprograms in PHYLIP 3.68 (Felsenstein 2008). The chord distance matrix was estimated using GENDIST, and relationships between populations were inferred by NEIGHBOR. Support was determined by bootstrapping over the 14 loci (SeqBoot). The resulting tree was drawn using FIGTREE v1.2.3 (Rambaut 2006).

Topology comparison between mtDNA and microsatellite trees

Prior studies using both mtDNA and microsatellites have shown topology differences in the resulting phylogenetic trees (e.g. Hansen *et al.* 1999; Lu *et al.* 2001). To determine whether the microsatellite and mtDNA trees were significantly different, we manually constructed alternative mtDNA haplotype trees in MacClade (Maddison & Maddison 2001) to match topological differences observed in the microsatellite tree. Next, PAUP (Swofford 2002) was used to fit ML models to each of the trees. Finally, we employed the Shimodaira–Hasegawa test (Shimodaira & Hasegawa 1999) to determine whether the altered mtDNA haplotype trees (based on the microsatellite tree) were significantly different from the original mtDNA haplotype tree.

Results

Analysis of mitochondrial DNA control region

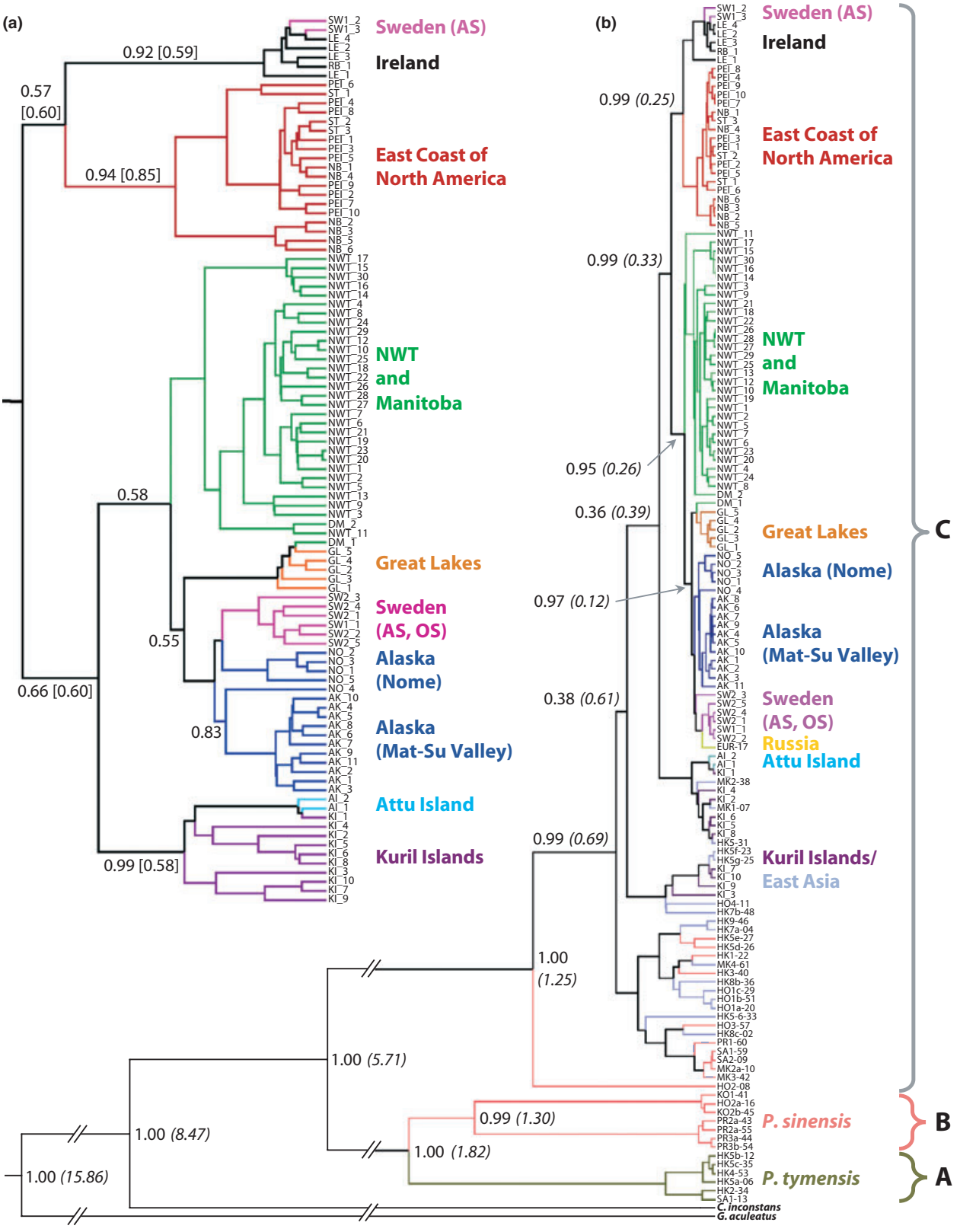
Genetic and haplotype diversity. We detected 97 haplotypes in 169 mtDNA CR region sequences belonging to 32 populations of ninespine sticklebacks; each haplotype was specific to a particular geographical region (haplotype diversity, hd = 0.989 ± 0.002; Tables 1 and S1 (Supporting information); GenBank accession numbers HM483400–HM483508). The overall nucleotide diversity was low, but comparable to those seen in other studies of sticklebacks (Orti *et al.* 1994; Mäkinen & Merilä 2008; Shikano *et al.* 2010) and northern hemisphere fishes that were displaced by glaciation events (Hansen *et al.* 1999; Lu *et al.* 2001; van Houdt *et al.* 2005). An estimate of Fu's Fs was significantly negative over all the populations (−24.19, $P = 0.001$) and in several individual populations (Table 1). These results suggest that ninespine sticklebacks underwent rapid and recent population expansions, as would be expected following postglacial dispersal.

Phylogenetic analysis. Bayesian coalescent and ML phylogenetic trees were built with the 97 new CR haplotypes (Fig. 2a). In general, clades contained populations from adjacent geographical regions. Notably, the separation of Atlantic populations (North American East Coast and Ireland) from other North American populations [Alaska, Great Lakes, and Northwest Territories (NWT)] and the Aleutian and Kuril Islands populations was highly supported (posterior probability, *pp* = 0.94, ML bootstrap value = 0.85). This separation divides the Alaskan and Atlantic members of the putative 'Bering' form of ninespine sticklebacks into multiple clades; thus, the monophyly of this morphotype is not supported. Three results in this analysis were not consistent with geography. First, the two fish in our sample from Manitoba (DM) were split between the NWT and Great Lakes clades. Second, our sample from Rivière St Marguerite (RM), which is geographically proximate to the St Lawrence River population (ST) in the East Coast clade, also groups with the Great Lakes populations. Third, samples from two populations in Sweden were separated: most fish from Askö (AS) clustered with fish from Ireland, whereas all fish from Öresund (OS) and one from AS formed their own clade allied with fish from Alaska and the Great Lakes. While intriguing, this latter branch of the tree was poorly supported (*pp* = 0.24), and this association between geographical groups was not observed in the nuclear marker analysis. We expect that further sampling of populations across northern Eurasia will help resolve this grouping and determine whether this relationship is historically correct.

An expanded Bayesian tree was built combining 97 CR haplotypes from this study, 41 haplotypes from a previous study of East Asian ninespine sticklebacks, including the additional species *Pungitius sinensis* and *P. tymensis* (Takahashi & Goto 2001), and threespine and brook stickleback mtDNA sequences as outgroups (Fig. 2B). Three major lineages of ninespine sticklebacks were named following the convention of Takahashi & Goto (2001) and are indicated in Fig. 2b. Lineage A comprises *P. tymensis*, Lineage B contains *P. sinensis*, and Lineage C contains both *P. pungitius* and *P. sinensis*. The paraphyly of Lineage C was discussed previously by Takahashi & Goto (2001). We found high support (*pp* = 0.99) for the divergence of Asian populations from North American and European populations in Lineage C. In this expanded tree, the haplotypes from the Kuril Islands populations were intermixed with those from Lineage C in Asia, which is expected as the result of the geographical proximity of the Kuril Islands to Japan. The North American/European clade within Lineage C shows a similar topology to that of the 97 haplotypes tree (Fig. 2a); however, the inclusion of the Asian samples shows that the Kuril and Aleutian Island samples diverged earlier in the phylogeny.

Molecular clock and divergence dating. We estimated a nucleotide substitution rate of 0.028 substitutions/site/Myr, based on mtDNA sequences from ninespine, threespine, and brook sticklebacks. Based on this rate, divergence between ninespine and threespine sticklebacks likely occurred 15.86 Mya (95% CI: 11.43–20.66 Mya; Fig. 2B). This age is consistent with Bell *et al.* (2009), who suggested that the divergence between these species predates a 13.3 Myr-old Miocene threespine stickleback fossil. The more recent divergence between ninespine and brook sticklebacks was estimated at 8.47 Mya (95% CI: 7.50–9.44 Mya), which is older than the oldest known *Pungitius* fossil at

Fig. 2 Bayesian phylogenetic trees based on mitochondrial DNA control region sequence. (a) Phylogenetic tree based on 97 new CR haplotypes. Posterior probabilities from the Bayesian analysis are listed on the branches. Bootstrap values from the maximum likelihood analysis are in square brackets. (b) Extended tree containing the 97 haplotypes from this study, 41 from East Asian populations (Takahashi & Goto 2001), and threespine and brook sticklebacks as outgroups. Dashed lines denote haplotypes present in both *Pungitius pungitius* and *Pungitius sinensis*. Posterior probabilities and divergence dates in Mya (in *italics* and parentheses) are listed on the branches. Major lineages in the ninespine stickleback phylogeny ('A', 'B', and 'C') are denoted as in Takahashi & Goto (2001).



~7 Mya (Rawlinson & Bell 1982). Divergence between *P. tymensis* and *P. sinensis* was estimated at 1.82 Mya (95% CI: 1.32–2.31), which is more recent than a previous estimate of 2.34 Mya (Takahashi & Goto 2001). The major split between the Asian + Aleutian Island clade and the North American + European group occurred ~390 000 years ago (95% CI: 297 000–508 000), while Atlantic populations from North America and Europe diverged about 250 000 years ago (95% CI: 162 000–340 000). Interestingly, even the most recent divergence estimates greatly predate the Wisconsin glaciations, suggesting ancient origins for the major modern clades of ninespine sticklebacks in North America.

Analysis of nuclear microsatellite loci

Measures of genetic diversity are listed in Table 1. Pairwise comparisons among microsatellite marker loci revealed no linkage disequilibrium, and populations generally appeared to be in Hardy–Weinberg equilibrium. Just three loci in eight populations showed individually significant deviations from H–W equilibrium (1.7% of all locus population combinations, which is lower than expected by chance).

F_{ST} values between populations were generally high (overall $F_{ST} = 0.458$), indicating low levels of recent migration (Table S2, Supporting information). Significant pairwise F_{ST} values varied between 0.04 and 0.835. High F_{ST} values were found for many comparisons, especially those involving isolated freshwater lakes with little or no migration. In contrast, a low F_{ST} value (0.078, $P < 0.0001$) was detected between the anadromous Glenfinnan River (GF) and freshwater O'Keefe's Lake (OK) populations from Prince Edward Island (PEI). These two populations also shared mtDNA haplotypes, suggesting an ancestral population similar to GF recently colonized OK. We did not detect significant divergence among the populations from the Great Lakes region ($F_{ST} = 0$–0.09, ns); however, sample sizes from these populations were small. Larger samples will be required to determine the levels of genetic differentiation more accurately and to infer whether migration has occurred recently among these populations.

AMOVA revealed that a majority of the variation existed within populations (51.3–88.3%) and the degree of genetic differentiation varied among geographical regions and phylogenetic lineages (Table S3, Supporting information). A significant amount of the variation was explained by geography (28.3%) and lineage (32.6%), whereas habitat (marine vs. freshwater) explained only 3.02%. The latter result is consistent with a study of European ninespine sticklebacks, in which habitat accounted for 3.7% of the variation (Shikano *et al.* 2010). Isolation by distance analyses did not detect any significant relationships between populations and physical distance within any of the geographical groups (Table S4, Supporting information). Together, these results show that geographical region is a major determinant of genetic similarity, but that the location and habitat of a population within a region plays little role in genetic differentiation.

Bayesian population structure analysis. Population structure was determined by Bayesian clustering analysis of multi-locus microsatellite genotypes using Structure (Pritchard *et al.* 2000) (Fig. 3). We tested models for $K = 2$ to $K = 20$. The log-likelihood reached a maxi-

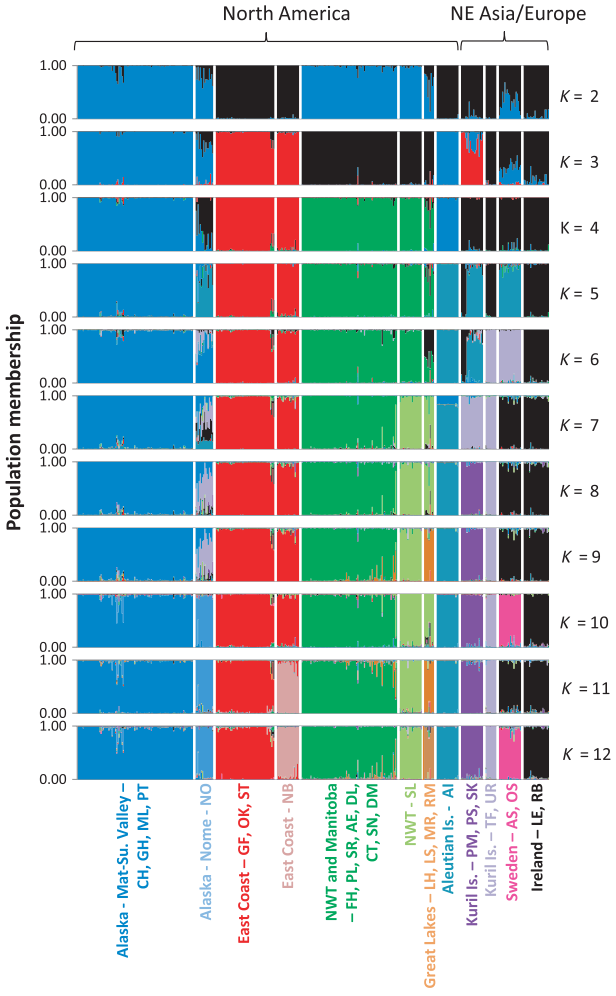


Fig. 3 Individual genetic population membership coefficients estimated by structure from $K = 2$–12 populations, with $K = 12$ being the most likely population structure. Each individual ($n = 400$) is represented by a single vertical line, with different colours representing membership in different genetic population groups. Geographical populations identified at the bottom of the figure.

mum at $K = 12$ and variance increased at $K \geq 13$; therefore, $K = 12$ was determined to be the most likely population number (Evanno *et al.* 2005; Pritchard *et al.* 2009). This analysis reveals a high level of genetic structure among the ninespine stickleback populations sampled. We also note that the populations from Alaska and the Atlantic coast are distinct from one another at all $K > 1$, indicating that these populations do not share the same genetic history (and by extension, refugial origin). This result is also consistent with our mtDNA results, which showed an ancient split between these two groups.

Phylogenetic relationships among populations—nuclear microsatellites. A phylogenetic tree of populations was inferred from D_{CE} distances of microsatellite loci for the 32 populations of ninespine sticklebacks (Fig. 4). Greatest support was for a distinct Alaskan clade (bootstrap value = 84.7). We found a low level of support (bootstrap value = 59.7) for the separation of the Great Lakes clade from the clade containing the NWT and Manitoba populations, but high support for separation of populations within each of these clades. The clade that includes the Irish, Swedish, Kuril Islands, and Alaskan populations was weakly supported, possibly indicating a weakening of the phylogenetic signal by allele homoplasy (Mäkinen *et al.* 2006). In short, the geographical groupings in the tree are consistent with the results of the Structure analysis (Fig. 3).

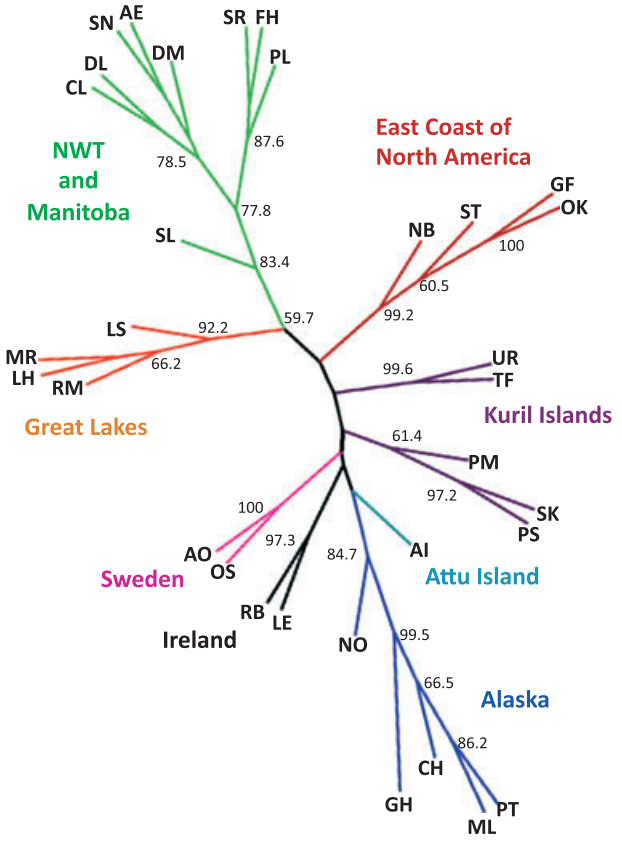


Fig. 4 Phylogenetic relationships among populations of ninespine sticklebacks based on 14 independent nuclear microsatellite markers. Neighbour-joining tree based on D_{CE} among ninespine sticklebacks grouped into their collection sites (populations). Per cent bootstrap support (1000 iterations) is listed on the branches.

Topology comparison between mtDNA and microsatellite trees. The mtDNA and microsatellite tree topologies did not differ significantly (-2lnL difference = 18.50, $P = 0.241$). This suggests that both provide fairly accurate and complementary accounts of ninespine stickleback population history.

Discussion

A principal objective of this study was to determine the genetic relationships among ninespine stickleback populations across the northern hemisphere, with a particular focus on North America. To this end, we have assessed the genetic variation and phylogenetic relationships among populations of ninespine sticklebacks using both mtDNA CR sequence and nuclear microsatellite markers. Analysis of these two independent data sets illuminates the phylogenetic origins of modern populations of this species and suggests that the postglacial history of ninespine sticklebacks is more complex than previously hypothesized. The presence of multiple major genetic lineages suggests that ninespine sticklebacks in North America survived the Wisconsin glaciations in at least three glacial refugia, rather than only two as inferred by morphology.

Refugial origins of modern ninespine stickleback populations

The precise dates and the waterways involved in the dispersal of fish species thousands of years ago necessarily involves some conjecture (McPhail 1963a; Underhill 1986). The case of the ninespine stickleback is especially challenging because of the lack of Pleistocene fossils from the regions we examined in this study. However, based on the previously proposed dispersal route of ninespine sticklebacks from the Bering refugium, we expected the fish from Alaska (GH, ML, CH, PT, NO), the Aleutian Islands and the East Coast (GF, OK, NB, ST) to have a close genetic relationship (McPhail 1963a). However, analysis of both mtDNA and microsatellites (phylogeny and population structure) suggests that this is not the case; in both trees, for example, we see a distinct North American East Coast group that is not allied with Alaskan populations.

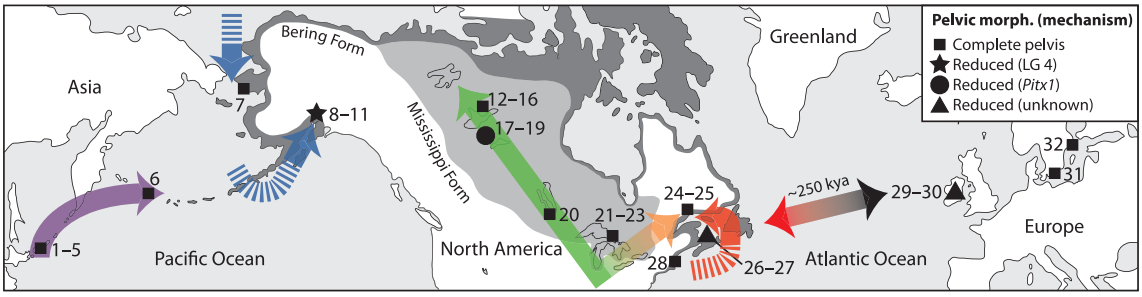


Fig. 5 Hypothesized postglacial dispersal patterns of ninespine sticklebacks in North America compared with McPhail's (1963a) two major morphotypes of ninespine stickleback, the Bering and Mississippi forms. Arrows in and around North America represent postglacial dispersal routes from Bering (blue), Mississippi (green and orange) and Atlantic (red) regions after the Wisconsin glaciations. The red/black arrow represents the earlier divergence of trans-Atlantic populations ~250 000 years ago, and the purple arrow represents colonization of the Aleutian Islands from East Asia. Dashed arrows indicate uncertainty about dispersal origins and routes, as discussed in the text. Regions containing populations with pelvic reduction and the associated genetic mechanisms (if known) are labelled.

Mainland Alaskan populations in our study likely dispersed from the Bering refugium (Fig. 5). These populations formed two clades (Nome and Mat-Su Valley) in the mtDNA and microsatellite analyses, with an estimated divergence time of 110 000 years ago (95% CI: 66 400–173 700). This extended separation time suggests that the ancestors of the Mat-Su Valley population might have survived Wisconsin times in an unglaciated area near Cook Inlet (McPhail 1963a; Lindsey & McPhail 1986), whereas the Nome population originated from a more northern Bering population. Subsequently, ninespine sticklebacks probably began invading new freshwater habitats in the Mat-Su Valley about 9000 years ago (Karlstrom 1964), and this invasion would have been contemporaneous with the incursion of threespine sticklebacks (Bell & Orti 1994). An inland migration at or near Cook Inlet was the only access point for these fish as high mountains surround the valley. The region around Nome attained its current topography beginning 11 000 years ago (Elias *et al.* 1996), as the Bering land bridge became inundated as the sea level rose. Bering ninespine sticklebacks also dispersed westward along the Arctic coast beginning 8000 years ago (McPhail 1963a), but apparently did not reach the East Coast, at least not as far south as the St Lawrence River or PEI.

Anadromous ninespine sticklebacks are widespread on the East Coast, so postglacial dispersal probably did not originate from an exclusively freshwater refugium. Anadromous sticklebacks could have dispersed northward from an Atlantic refugium (*sensu* Bernatchez 1997) when the ice began to melt ~18 000 years ago (Curry 2007) (Fig. 5). Alternatively, or perhaps in addition, they also might have survived in the Acadian refugium located on the Grand Banks and dispersed in a route similar to that of the anadromous rainbow smelt (*Osmerus mordax*) (Bernatchez 1997). We cannot currently distinguish among these possibilities, but additional sampling of East Coast freshwater populations could help resolve this uncertainty. The lone East Coast freshwater population in our study (OK) is from PEI. This island was probably colonized by anadromous fish ~11 000–8000 years ago following deglaciation, with lake fish becoming isolated ~6000 years ago when the land mass rose and limited access to the sea (Curry 2007). The mtDNA relationship of the East Coast populations with those from Europe shows a common history between these regions until a split occurred ~250 000 years ago (95% CI: 162 000–340 000); however, we cannot currently determine in which direction movement between these two groups occurred. This trans-Atlantic association was also found in the threespine stickleback (Orti *et al.* 1994; Mäkinen & Merilä 2008), a close relative of the ninespine stickleback.

Ninespine sticklebacks also survived the Wisconsin glacial period in the upper Mississippi Valley and dispersed to the east to colonize the Great Lakes region and to the northwest to colonize the continental interior. The phylogenetic affinities of presumptive Mississippi populations from the Great Lakes, Manitoba, and NWT were confirmed by analysis of mtDNA (Fig. 2) and microsatellites (Figs 3 and 4). Notably, these populations are more closely related to fish from Alaska than to Atlantic populations, with a vicariance time of up to 260 000 years ago (95% CI: 180 000–342 500). These relationships suggest that preglacial populations dispersed from the Pacific, where this genus has been present for at least 7 Myr (Rawlinson & Bell 1982). However, the topology of these relationships in the mtDNA tree is also puzzling. We expected a close relationship among only the NWT, Manitoba, and Great Lakes populations because of the geographical proximity and presumed shared ancestry of these regions;

indeed, this prediction is supported by the microsatellite analyses (Figs 3 and 4). In contrast, the mtDNA tree shows a close relationship between the geographically distant Alaskan and Great Lakes groups. Based on known routes of dispersal between the Bering and Mississippi refugia, as well as the current distribution of ninespine sticklebacks, we cannot surmise a feasible route for interchange between these two regions that bypasses the NWT. Briefly, northwestward dispersal of ninespine sticklebacks into the vast interior of North America probably originated from glacial Lake Agassiz. From there, they could disperse northward to the Mackenzie River via the Agassiz-Clearwater corridor from 9.9–9.5 Kya (Smith & Fisher 1993; Rempel & Smith 1998), or possibly earlier (McPhail 1963b). This would provide a clear route to glacial Lake McConnell, which covered most of our NWT collection sites at that time (Craig 1965). Ninespine sticklebacks also dispersed eastward to the Great Lakes region, possibly through Lake Kelvin (Leverington & Teller 2003), and as far east as the salinity border at the Champlain Sea ~10 Kya (Underhill 1986). We suspect that the large distances between our sampling sites and the low number of samples from the Great Lakes region might have a misleading effect on our mtDNA tree topology. We expect that additional sampling in the interior and along the northern coasts of North America will further elucidate the mtDNA relationships among the populations in question.

The inclusion of the Manitoba haplotypes in both the Great Lakes and NWT clades is also intriguing. It is possible that at least two major mitochondrial lineages of ninespine sticklebacks occurred in Lake Agassiz, and both are still present in Manitoba. Alternatively, the Great Lakes and NWT populations could have radiated from two different refugia, the Missouri and Mississippi, and their dispersal routes overlapped in Manitoba. However, the modern distribution and inferred postglacial dispersal routes of this species cast doubts on their survival in the Missouri Valley (van Houdt *et al.* 2005).

The close relationship between fish from Attu Island, a western island in the Aleutian Islands chain, to the Kuril Islands populations indicates that (i) these two groups split ~47 000 years ago (95% CI: 14 800–86 900) (Fig. 2) and (ii) the Attu population originated from East Asia, not from North America (Fig. 5). Nevertheless, this does not rule out the Bering refugium as the source of colonization for the other islands in the Aleutian chain. Additional sampling will help determine the pre- and postglacial history of fish from the Aleutian Island, including identification of the eastern extent of colonization from the Bering refugium and the western extent of colonization from Asia. However, our data already suggest that the Aleutians represent the boundary between the eastward radiation of the Asian ninespine sticklebacks and the westward movement of the North American group.

Mitochondrial DNA and nuclear microsatellite phylogenies are complementary

We generated phylogenetic hypotheses for ninespine sticklebacks using both mtDNA and nuclear microsatellite data. While we found that the trees constructed from each type of data were statistically indistinguishable, we also note that these separate analyses are complementary. Our mtDNA tree utilizes relatively slow changing sequence data and shows strong support for more ancient divergence events (for example, between the trans-Atlantic group and other North American populations), but weaker support for more recent splits (for example, among NWT haplotypes) (Fig. 2). In contrast, size polymorphisms in nuclear microsatellites evolve rapidly and provide much more data to confidently resolve recent divergences (Fig. 3). Nevertheless, such rapid changes can also result in homoplasy, rendering microsatellites less informative for deeper divergences. As a result, our microsatellite tree shows strong support for the relationships among recently diverged populations within a geographical region, but not necessarily among populations in different geographical regions. In summary, our complementary phylogenetic analyses provide strong statistical support for both ancient and recent divergences among ninespine stickleback lineages.

A notable difference between the mtDNA and microsatellite trees concerns populations from the Baltic Sea coast of Sweden (AS and OS). In the mtDNA tree, haplotypes from both populations group with haplotypes from central and northwestern North America and Russia, but two haplotypes from the AS population group with Irish haplotypes in the trans-Atlantic clade. However, the microsatellite tree shows strong support (100% of trees) for the close alliance of these two Swedish populations. Why do the two methods produce such seemingly different results? One possibility is that two major lineages of ninespine sticklebacks are admixed in the Baltic Sea. Our mtDNA tree provides evidence that one Swedish population (AS) includes fish from the trans-Atlantic clade and another eastern lineage. This latter lineage also includes a previously reported haplotype from Russia (Takahashi & Goto 2001) (Fig. 2B). Nuclear microsatellite markers, on the other hand, do not show wide separation between the two Swedish populations because of relatively recent gene flow between these two putative mitochondrial lineages. This interpretation is consistent with the admixture between

eastern and western European lineages inferred by a recent phylogeographical study of northern European ninespine sticklebacks (Shikano *et al.* 2010). Unfortunately, our study and the northern European study utilized different microsatellites and sequences from different parts of the mtDNA, making a comprehensive analysis impossible at this time. Additional sequencing and genotyping with common markers will help elucidate the global phylogeography of ninespine sticklebacks, including the possible admixture of trans-Atlantic and Eurasian lineages in the Baltic Sea.

Molecular clock rate and divergence dates

Mitochondrial DNA CR clock rates vary greatly among major taxonomic groups of bony fishes (Liu *et al.* 2006). The clock rate in this study was determined to be 5.6% (difference/Myr), which is faster than that determined in a study of East Asian ninespine sticklebacks (~2.71%) (Takahashi & Goto 2001). However, our rate is similar to another northern hemisphere fish affected by glaciation, Arctic charr (*Salvelinus alpinus*), at 5–10% (Brunner *et al.* 2001). In our study of the ninespine stickleback, calibration dates based on fossil data were much more ancient than all of our intraspecific divergence times. If we had a more recent calibration point, the substitution rate might have appeared faster, resulting in more recent divergence estimates (Ho & Larson 2006); therefore, divergence times reported here may be overestimated. However, if our estimate is reasonably accurate, then most major lineages of *P. pungitius* diverged from a common ancestor during the Pleistocene, with major clades in place well before the most recent Wisconsin glaciations. The high degree of divergence, combined with the absence of shared haplotypes from geographically distant fish, suggests that little or no intermixing of these lineages occurred between major glaciation events.

Phylogenetic origins of adaptive traits

A previous hypothesis for the postglacial dispersal of North American ninespine sticklebacks focused on shared—and presumably derived—morphological traits, especially the numbers of dorsal spines and gill rakers (McPhail 1963a), which vary with latitude as well as between marine and freshwater populations. These observations led to the reasonable conclusion that populations with high numbers of gill rakers and dorsal spines probably shared a common ancestry in one glacial refugium, while modern populations with fewer spines and rakers likely had a different origin (Fig. 5). However, our phylogenies contradict this interpretation, instead suggesting that similar morphological changes occurred independently in different lineages. For example, Alaskan and East Coast populations from McPhail's presumptive Bering group actually belong to different mitochondrial and nuclear genetic lineages and probably originated from separate Bering and Atlantic ancestors (Fig. 5). These results highlight the widespread convergent evolution of quantitative morphological traits in ninespine sticklebacks that might be linked to ecological, rather than phylogenetic factors (Ostbye *et al.* 2006).

Our phylogeny also suggests widespread convergent evolution of pelvic reduction, another important adaptive phenotype. The pelvic skeleton shows robust development in marine and most freshwater populations, but has been lost in several freshwater populations of both threespine and ninespine sticklebacks as an adaptation to local predators or water chemistry (Nelson 1971; Reimchen 1980; Giles 1983; Blouw & Boyd 1992; Bell *et al.* 1993; Ziuganov & Zotin 1995; Shapiro *et al.* 2004; Hunt *et al.* 2008; Marchinko 2009). Under appropriate selection conditions, this phenotype can be rapidly driven to high frequency in a population (Ziuganov & Zotin 1995). To date, two genomic regions have been implicated in the evolution of pelvic reduction in ninespine sticklebacks. The probable involvement of *Pitx1* in a NWT population was inferred through complementation crosses and expression studies, while a locus on LG 4 was detected in a genetic mapping study of an Alaskan population (Shapiro *et al.* 2006, 2009). Based on our new phylogenetic information, we can now see that these two populations, each with a potentially unique genetic change leading to this adaptive phenotype, also have different phylogenetic histories and may have diverged over 300 years ago.

Pelvic reduction has also evolved in the trans-Atlantic lineage, including two populations in our analysis (OK and Lough Ennell (LE)) (Nelson 1971; Blouw & Boyd 1992). Is pelvic reduction in these populations controlled by genetic mechanisms that are similar or different to the NWT and Alaskan populations? While we have not yet genetically mapped the origins of reduction in the trans-Atlantic lineage, we can use our phylogenetic and molecular clock results to make predictions about shared versus *de novo* mutations leading to this derived phenotype. One possibility is that one or both of the OK and LE populations share a common pelvic reduction haplotype with either the Alaskan or NWT pelvic-reduced populations. For example, an ancestral (presumably marine) population might have maintained pelvic reduction alleles at a low frequency, and multiple populations subsequently selected on this standing genetic variation. This scenario has striking precedent in threespine sticklebacks, in which the derived haplotypes of *Eda* and *Kitlg* that underlie variation in bony armour and pigmentation, respectively, probably spread

from marine fish to multiple freshwater populations where they were swept to high frequency (Colosimo *et al.* 2005; Miller *et al.* 2007). However, a key aspect of these traits in threespine sticklebacks is that their geographical distribution is essentially continuous in freshwater populations along the west coast of North America (*Kitlg*, pigmentation) or the northern hemisphere (*Eda*, armour). Interbreeding of anadromous fish with resident freshwater populations provides an annual mechanism for flow of derived alleles back into marine populations and maintenance of allelic variation (Colosimo *et al.* 2005). In contrast, pelvic reduction in ninespine sticklebacks occurs in relatively few populations with a geographically sporadic distribution and with limited opportunities for gene flow.

A second possibility is that different mutations, either in the same or different genes, underlie pelvic reduction in different ninespine stickleback lineages. This scenario parallels the repeated evolution of pelvic reduction in threespine sticklebacks by *de novo* mutations in a *cis*-regulatory element of *Pitx1* (Chan *et al.* 2010). Similarly, other studies of natural populations of vertebrates implicate different mutations in the same gene, or mutations in different genes, in the convergent evolution of adaptive traits (e.g., Protas *et al.* 2006; Steiner *et al.* 2007; Gross *et al.* 2009). Collectively, these examples demonstrate that different lineages, even those in geographical proximity, repeatedly converge on similar phenotypes by different mutations in the same or different genes.

Based on available phylogenetic and molecular clock evidence, as well as the sporadic geographical distribution of pelvic reduction in ninespine sticklebacks, we predict that this derived trait probably arose by different mutations in the Alaskan, NWT, and trans-Atlantic lineages. (Indeed, based on the aforementioned examples, we should not be surprised to find that different mutations or genes control pelvic reduction in different populations *within* each of these clades.) These lineages are widely separated phylogenetically, geographically, and temporally. Moreover, high pairwise F_{ST} values argue against widespread gene flow among the pelvic-reduced populations from these three regions. It is possible that allele(s) underlying pelvic reduction arose before the major North American lineages split over 300 BP and remained at a low frequency in each of the clades and that some cases of pelvic loss resulted from selection on standing genetic variation; however, the genetic origins of this trait already appear to be different in the Alaskan and NWT populations studied thus far. Our prediction of independent mutations remains to be confirmed by genetic linkage mapping studies and identification of mutations in East Coast and European populations, and these experiments are under way.

In summary, we used mitochondrial and nuclear genetic data to examine the phylogeography of ninespine sticklebacks. Our analyses suggest that modern populations in North America originated from at least three glacial refugia, and these results challenge a long-standing hypothesis of two refugia based on morphological traits. Based on our phylogenetic analysis and molecular clock estimates, we also predict that pelvic reduction likely evolved independently in multiple clades. This study provides a robust phylogeographical framework for future evolutionary genetic studies of this emerging model organism.

Acknowledgements

We thank Katherine Maslenikov, Mary Burridge, Richard Winterbottom, Frank von Hippel, Justine Crowe, Hillary Matchens, Bertil Borg, Anjanette Bowen, Mike Bell, the Shannon Regional Fisheries Board, and Wood Buffalo National Park for providing samples and collecting assistance. We are grateful to Louis Bernatchez, Gábor Herczeg, Jon Seger, Sydney Stringham, and an anonymous reviewer for insightful comments on earlier drafts of the manuscript. This project was supported by grants from NSF (IOS-0744974) and the Burroughs Wellcome Fund (MDS); and grants from the University of Utah Biology Undergraduate Research Program and the Biology Environmental Science Trust (MAM).

References

Ayvazian SG, Krueger WH (1992) Lateral plate ontogeny in the North American ninespine stickleback, *Pungitius occidentalis*. *Copeia*, **1992**, 209–214.

Bell MA, Foster SA (1994) *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press, Oxford.

Bell MA, Orti G (1994) Pelvic reduction in threespine sticklebacks from Cook Inlet lakes: geographical distribution and intrapopulation variation. *Copeia*, **1994**, 314–325.

Bell MA, Orti G, Walker JA, Koenings JP (1993) Evolution of pelvic reduction in threespine stickleback fish: a test of competing hypotheses. *Evolution*, **47**, 906–914.

Bell MA, Stewart JD, Park PJ (2009) The world's oldest fossil threespine stickleback fish. *Copeia*, **2009**, 256–265.

Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St Lawrence River estuary (Quebec, Canada). *Molecular Ecology*, **6**, 73–83.

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

Blouw DM, Boyd GJ (1992) Inheritance of reduction, loss, and asymmetry of the pelvis in *Pungitius pungitius* (ninespine stickleback). *Heredity*, **68**, 33–42.

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001) Holarctic phylogeography of Arctic Charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **55**, 573–586.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chan YF, Marks ME, Jones FC *et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx 1* 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.

Craig BG (1965) Glacial Lake McConnell, and the superficial geology of parts of Slave River and Redstone River map areas, District of Mackenzie. *Bulletin Geological Survey of Canada*, **122**, 33.

Curry RA (2007) Late glacial impacts on dispersal and colonization of Atlantic Canada and Maine by freshwater fishes. *Quaternary Research*, **67**, 225–233.

Day F (1878) On some Irish Gasterostei. *Journal of the Linnean Society of London (Zoology)*, **13**, 110–114.

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

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

Elias SA, Short SK, Nelson CH, Birks HH (1996) Life and times of the Bering land bridge. *Nature*, **382**, 60–63.

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

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Felsenstein J (2008) *PHYLIP (Phylogeny Inference Package) Version 3.68*. Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Giles N (1983) The possible role of environmental calcium levels during the evolution of phenotypic diversity on Outer Hebridean populations of the three-spined stickleback, *Gasterosteus aculeatus*. *Journal of Zoology*, **4**, 535–544.

Gonda A, Herczeg G, Merila J (2009) Adaptive brain size divergence in nine-spined sticklebacks (*Pungitius pungitius*)? *Journal of Evolutionary Biology*, **22**, 1721–1726.

Gross HP (1979) Geographic variation in European ninespine sticklebacks, *Pungitius pungitius*. *Copeia*, **1979**, 405–412.

Gross JB, Borowsky R, Tabin CJ (2009) A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genetics*, **5**, e1000326.

Guiher TJ, Burbrink FT (2008) Demographic and phylogeographic histories of two venomous North American snakes of the genus *Agkistrodon*. *Molecular Phylogenetics and Evolution*, **48**, 543–553.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

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

Herczeg G, Gonda A, Merila J (2009a) Evolution of gigantism in nine-spined sticklebacks. *Evolution*, **63**, 3190–3200.

Herczeg G, Gonda A, Merila J (2009b) Predation mediated population divergence in complex behaviour of nine-spined stickleback (*Pungitius pungitius*). *Journal of Evolutionary Biology*, **22**, 544–552.

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

Ho SYW (2007) Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology*, **38**, 409–414.

Ho SYW, Larson G (2006) Molecular clocks: when times are a changin'. *TRENDS in Genetics*, **22**, 79–83.

van Houdt JKJ, de Cleyn L, Perretti A, Volckaert FAM (2005) A mitogenic view on the evolutionary history of the Holarctic freshwater gadoid, burbot (*Lota lota*). *Molecular Ecology*, **14**, 2445–2457.

Hunt G, Bell MA, Travis MP (2008) Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. *Evolution*, **62**, 700–710.

Karlstrom TNV (1964) Quaternary geology of the Kenai lowland and glacial history of the Cook Inlet region, Alaska. *USGS Professional Paper*, **443**, 1–69.

Kendal JR, Rendell L, Pike TW, Laland KN (2009) Nine-spined sticklebacks deploy a hill climbing social learning strategy. *Behavioral Ecology*, **20**, 238–244.

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*, **9**, 299–306.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Leverington DW, Teller JT (2003) Paleotopographic reconstructions of eastern outlets of Lake Agassiz. *Canadian Journal of Earth Sciences*, **40**, 1259–1278.

Lindsey CC (1981) Stocks are chameleons: plasticity in gill-rakers of coregonid fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1497–1506.

Lindsey CC, McPhail JD (1986) Zoogeography of fishes of the Yukon and Mackenzie basins. In: *The Zoogeography of North American Freshwater Fishes* (eds Hocutt CH, Wiley EO), pp. 639–674. John Wiley & Sons, New York.

Liu J-X, Gao T-X, Zhuang Z-M, Jin X-S, Yokogawa K, Zhang Y-P (2006) Late Pleistoncene divergence and subsequent population expansion of two closely related fish species, Japanese anchovy (*Engraulis japonicus*) and Australian anchovy (*Engraulis australis*). *Molecular Phylogenetics and Evolution*, **40**, 712–723.

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

Maddison DR, Maddison WP (2001) *MacClade 4: Analysis of Phylogeny and Character Evolution, Version 4.06*. Sinauer, Sunderland, MA, USA.

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.

Marchinko KB (2009) Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution*, **63**, 127–138.

McPhail JD (1963a) Geographic variation in North American ninespine sticklebacks, *Pungitius pungitius*. *Journal of the Fisheries Research Board of Canada*, **20**, 27–44.

McPhail JD (1963b) *The Postglacial Dispersal of Freshwater Fishes in Northern North America*, PhD thesis, McGill University, Montreal, Quebec, Canada.

Miller CT, Beleza S, Pollen AA *et al.* (2007) cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell*, **131**, 1179–1189.

Morris D (1958) The reproductive behaviour of the ten-spined stickleback (*Pygosteus pungitius* L.). *Behaviour*, **Supplement**, 1–154.

Nelson JS (1971) Absence of the pelvic complex in ninespine sticklebacks, *Pungitius pungitius*, collected in Ireland and Wood Buffalo National Park Region, Canada, with notes on meristic variation. *Copeia*, **1971**, 707–717.

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.

Ostbye K, Amundsen PA, Bernatchez L *et al.* (2006) Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Molecular Ecology*, **15**, 3983–4001.

Peichel CL, Nereng KS, Ohgi KA *et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*, **414**, 901–905.

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

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

Pritchard JK, Wen W, Falush D (2009) Documentation for *structure* software: version 2. Available from: http://pritch.bsd.uchicago.edu/structure.html.

Protas ME, Hersey C, Kochanek D *et al.* (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, **38**, 107–111.

Rambaut A (2006) *FigTree Version 1.0*. Available from http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummond AJ (2007) Tracer v 1.4. Available from http://beast.bio.ed.ac.uk/Tracer.

Rawlinson SE, Bell MA (1982) A stickleback fish (*Pungitius*) from the Neogene Sterling Formation, Kenai Penisula, Alaska. *Journal of Paleontology*, **56**, 583–588.

Reimchen TE (1980) Spine deficiency and polymorphism in a population of *Gasterosteus aculeatus*: an adaptation to predators? *Canadian Journal of Zoology*, **58**, 1232–1244.

Rempel LL, Smith DG (1998) Postglacial fish dispersal from the Mississippi refuge to the Mackenzie River basin. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 893–899.

Reynolds M, Weir BS, Cockerham CC (1983) Estimation for the coancestry coefficient: basis for a short term genetic distance. *Genetics*, **105**, 767–779.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN Version 2.0: A Software for Population Genetics Data Analysis*. University of Geneva, Geneva.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Shapiro MD, Marks ME, Peichel CL *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.

Shapiro MD, Bell MA, Kingsley DM (2006) Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Science*, **103**, 13753–13758.

Shapiro MD, Summers BR, Balabhadra S *et al.* (2009) The genetic architecture of skeletal convergence and sex determination in ninespine sticklebacks. *Current Biology*, **19**, 1140–1145.

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

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Smith DG, Fisher TG (1993) Glacial Lake Agassiz: the northwestern outlet and paleoflood. *Geology*, **21**, 9–12.

Smouse P, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel Test of matrix correspondence. *Systematic Zoology*, **33**, 627–632.

Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, **5**, e219.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), Version 4.0b10*. Sinauer, Sunderland, MA, USA.

Takahashi H, Goto A (2001) Evolution of East Asian ninespine sticklebacks as shown by mitochondrial DNA control region sequences. *Molecular Phylogenetics and Evolution*, **21**, 135–155.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Underhill JC (1986) The fish fauna of the Laurentian Great Lakes, the St Lawrence Lowlands, Newfoundland and Labrador. In: *The Zoogeography of North American Freshwater Fishes* (eds Hocutt CH, Wiley EO), pp. 105–136. John Wiley & Sons, New York.

Walters V (1955) Fishes of western arctic America and eastern arctic Siberia. *Bulletin of the American Museum of Natural History*, **106**, 255–368.

Ward JL, McLennan DA (2009) Historical and ecological correlates of body shape in the brook stickleback, *Culaea inconstans*. *Biological Journal of the Linnean Society*, **96**, 769–783.

Waser W, Sahoo TP, Herczeg G, Merilä J, Nikinmaa M (2010) Physiological differentiation among nine-spined stickleback populations: effects of copper exposure. *Aquatic Toxicology*, **98**, 188–195.

Wootton RJ (1976) *The Biology of the Sticklebacks*. Academic Press, London.

Wootton RJ (1984) *Functional Biology of Sticklebacks*. University of California Press, Berkeley.

Ziuganov VV, Zotin AA (1995) Pelvic girdle polymorphism and reproductive barriers in the ninespine stickleback *Pungitius pungitius* (L.) from northwest Russia. *Behaviour*, **132**, 1095–1105.

J.A. investigates the origin and genetic mechanisms responsible for quantitative traits. M.M. currently attends medical school at the University of Utah. P.S.C.'s primary interest is application of molecular phylogenetic techniques for detecting rapid evolutionary divergence events in deep phylogenies, including the evolutionary histories of mammals, cone snails, sticklebacks, and hymenopterans. M.S. studies the genetic, genomic, and developmental basis of vertebrate diversity.

Supporting information

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

Table S1 The distribution of the 97 haplotypes observed among the 169 ninespine sticklebacks CR sequences

Table S2 Pairwise population FST values

Table S3 Hierarchical analysis (AMOVA) of genetic divergence in ninespine stickleback geographical regions/lineages and habitats

Table S4 Isolation by distance analysis (Mantel test) between geographical regions and F_{ST}

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