Estimating uncertainty in divergence times among three-spined stickleback clades using the multispecies coalescent

Bohao Fang[a,*], Juha Merilä[a], Michael Matschiner[b,c], Paolo Momigliano[a]

[a] *Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, FI-00014 Helsinki, Finland*
[b] *Department of Palaeontology and Museum, University of Zurich, Karl-Schmid-Strasse 4, 8006 Zurich, Switzerland*
[c] *Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, 0316 Oslo, Norway*

ARTICLE INFO

ABSTRACT

Incomplete lineage sorting (ILS) can lead to biased divergence time estimates. To explore if and how ILS has influenced the results of a recent worldwide phylogeny of three-spined sticklebacks (*Gasterosteus aculeatus*), we estimated divergence times among major clades by applying both a concatenation approach and the multispecies coalescent (MSC) model to single-nucleotide polymorphisms. To further test the influence of different calibration strategies, we applied different calibrations to the root and to younger nodes in addition to the ones used in the previous study. Both the updated calibrations and the application of the MSC model influenced divergence time estimates, sometimes significantly. The new divergence time estimates were more ancient than in the previous study for older nodes, whereas the estimates of younger nodes were not strongly affected by the re-analyses. However, since the applied MSC method employs a simple substitution model and cannot account for changes in population size, we suggest that different analytical approaches and calibration strategies should be used in order to explore uncertainty in divergence time estimates. This study provides a valuable reference timeline for the ages of worldwide three-spined stickleback populations and emphasizes the need to embrace, rather than obscure, uncertainties around divergence time estimates.

1. Introduction

The development of genome complexity reduction methods (such as restriction site associated DNA sequencing, RAD-seq), dropping sequencing costs and new analytical approaches have brought phylogenomics within the reach of most researchers. However, constructing phylogenies from multiple loci scattered along the genome poses serious challenges: different areas of the genome may undergo different sorting processes, resulting in individual gene trees that are in conflict with each other and with the true species tree (Maddison and Knowles, 2006). When lineage sorting is incomplete (ILS), phylogenetic analyses based on the concatenation of multiple unlinked loci may inflate statistical support, and in some cases lead to biases in branch lengths and divergence time estimates (Mendes and Hahn, 2016; Ogilvie et al., 2017). The multispecies coalescent (MSC) model incorporates incomplete lineage sorting, and may therefore often provide a better alternative for inferring phylogenies using multiple loci, yet most methods based on the MSC are computationally prohibitive to run on more than a few hundred loci and few tens of individuals.

Stange et al. (2018) extended the MSC method implemented in SNAPP (Bryant et al., 2012) to allow Bayesian divergence time estimation with SNP (single nucleotide polymorphism) data and found that their approach can overcome biases resulting from ILS. However, to achieve feasible run times with SNAPP's computationally demanding MSC implementation, the model chosen by Stange et al. (2018) had to make a few simplistic assumptions, which may also introduce novel biases under certain conditions. For example, the model of Stange et al. (2018) assumes an identical population size (Θ) for all lineages, an assumption that is often violated (for example in repeated range expansions) and may lead to overestimating divergence time in smaller populations (Stange et al., 2018). Previous work has shown that short-term mutation rates decay exponentially to long-term substitution rates within a few hundred thousand to a few million years (Ho et al., 2011), and therefore the application of a strict molecular clock when lineages of very different ages are included in the analyses (for example representing a recent radiation) may overestimate divergence times of younger nodes. This time-dependency of rates is usually attributed to the fact that slightly deleterious mutations are removed by purifying

---

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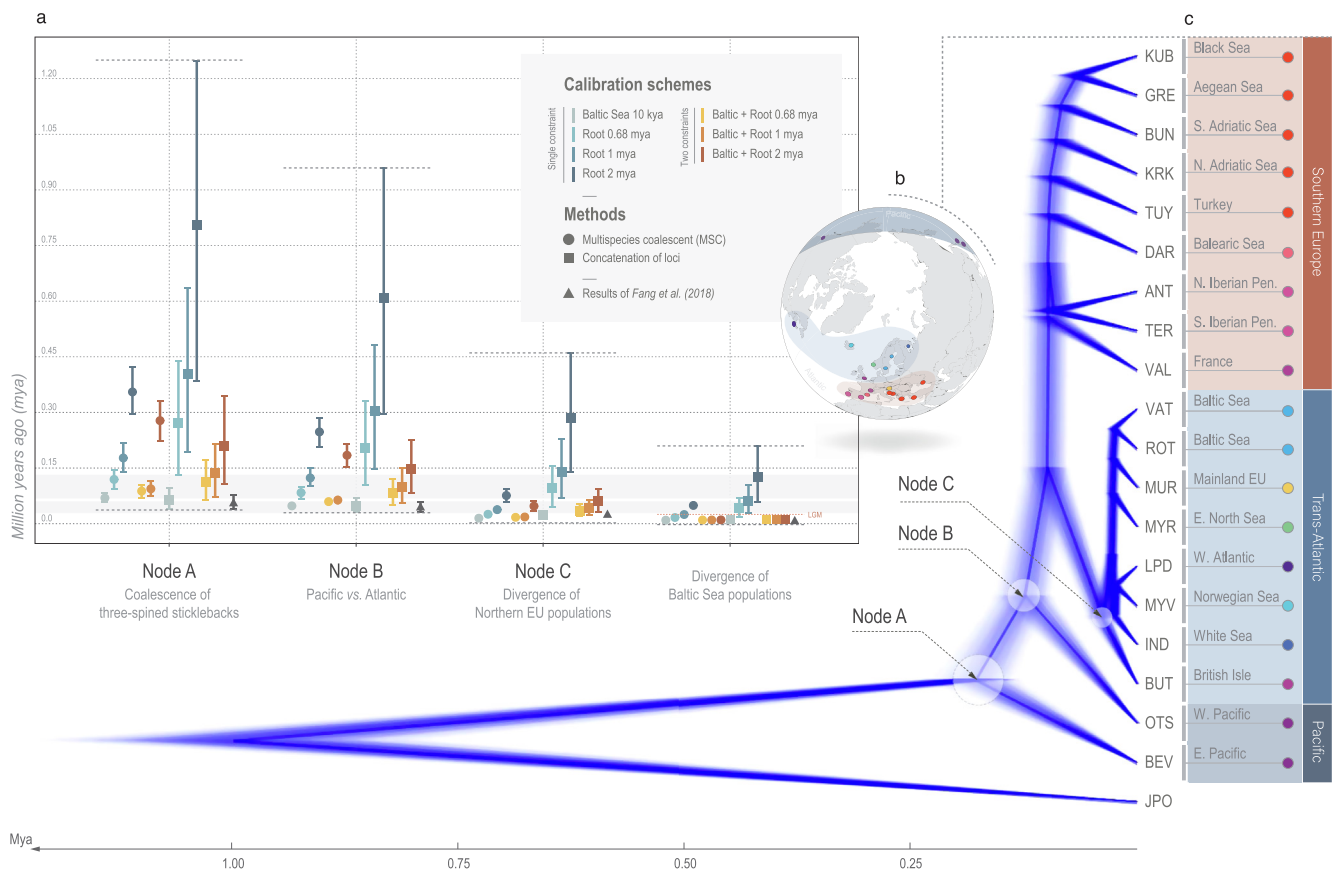*E-mail address:* bohao.fang@helsinki.fi (B. Fang).

Fig. 1. Estimates of divergence times among major three-spined stickleback clades. (a) Summary of divergence time estimates derived from the multispecies coalescent-based and concatenation-based methods. Seven sets of calibration schemes were applied in the two methods in parallel. Calibration points included three root calibrations (0.68 mya, 1 mya and 2 mya; the divergence time between *G. aculeatus* and the outgroup taxon *G. nipponicus*) and one geographic calibration (10 kya, TMRCA of Baltic Sea populations). Different calibration settings are marked with different colors. The gray-shaded regions denote periods of Bering Seaway existence, providing a geographical reference relevant for the estimates of node B. Divergence time estimates are represented by mean ages (dots) and 95% highest posterior density (HPD) intervals (vertical lines). See statistical summary in Appendix Table 1. (b) Schematic globe showing the sampling locations. (c) SNAPP tree, as an example, inferred solely with the 1 mya root calibration point. The geographic regions of populations are given on the right, distinguished by colors corresponding to colored dots in (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

selection over time (Ho et al., 2005; Burridge et al., 2008; Membrebe et al., 2019). It is therefore not clear how important this phenomenon is for RADseq data: while the sampled SNPs are unlikely to be under selection, it is likely that at least a portion of the sampled polymorphisms are affected by linked background selection (Kern and Hahn, 2018).

In a recent study, Fang et al. (2018) reconstructed the worldwide phylogenetic relationships and colonization history of the three-spined stickleback *Gasterosteus aculeatus,* a key model organism in ecology and evolutionary biology, using relaxed molecular clock Bayesian analyses. A time-calibrated phylogeny was built using a supermatrix of 8079 concatenated SNPs derived from RAD sequencing. The study identified three divergent major clades corresponding to the old Pacific lineage, a younger lineage including southern European populations and a third postglacial lineage including Northern European and Western Atlantic populations. Divergence times among these major clades were found to be far more recent than indicated by previous studies (Fang et al., 2018). Since the applied method using standard Bayesian inference in BEAST 2 did not account for ILS, it is possible, in light of Stange et al.'s (2018) findings, that estimates of divergence times between major lineages were biased. Stange et al. (2018) reported that when the age of the root is constrained, ILS may lead to overestimation of divergence times of younger nodes when concatenation is used. In Fang et al. (2018), the use of very recent calibration points could potentially have led to a bias in the opposite direction in the presence of ILS, leading to an underestimation of divergence times for intermediate nodes.

Three-spined sticklebacks have undergone rapid and repeated

"adaptive radiations" towards various phenotypes while colonizing freshwater habitats since the Last Glacial Maximum (LGM, ca. 26.5–20 kya, thousands of years ago; Bell and Foster, 1994; McKinnon and Rundle, 2002; Clark et al., 2009). Hence, these recently diverged populations and lineages, such as Northern European populations, are likely to exhibit a large degree of incomplete lineage sorting (Maddison and Knowles, 2006; Fang et al., 2018).

Here, we reanalyze a subset of the data from Fang et al. (2018) using Stange et al.'s (2018) method in order to test whether ILS may have led to biased estimates of divergence times. It should be noted that the assumption of a constant population size (Θ) across lineages is almost certainly not met in this system, since three-spined stickleback radiations have involved multiple bottlenecks (Taylor and McPhail, 1999, 2000), particularly when freshwater habitats were colonized. Depending on the strength of selection on genome-wide SNPs, we might expect heterogeneous mutation rates across lineages, reflective of the rapid exponential decay of mutation rate observed within the first few hundred thousands to millions of years (Ho et al., 2005). These biases could lead to underestimation or overestimation of the divergence times with respect to Fang et al.'s (2018) results when Stange et al.'s (2018) method is applied, depending on whether old or young nodes are used as calibration points. Furthermore, since the publication of Fang et al. (2018), the divergence time estimate between the outgroup taxon (*G. nipponicus*) and *G. aculeatus* has been refined from 2 mya to 0.68–1.0 mya (Ravinet et al., 2018), and this could affect divergence times within *G. aculeatus*. Therefore, the results of the extensive

reanalysis conducted in this study, together with those from Fang et al. (2018), should present a range of minimum and maximum ages for worldwide stickleback populations, giving rise to a useful reference timeline for the research on the three-spined stickleback.

2. Material and methods

The software SNAPP (Bryant et al., 2012) would require prohibitively long run-times for the analysis of the full dataset of Fang et al. (2018; see: Chifman and Kubatko, 2014). Therefore, a subset of 39 *G. aculeatus* individuals, originating from 19 different populations, as well as two individuals representative for the outgroup taxon *G. nipponicus* were extracted from the published dataset of Fang et al. (2018). These representative individuals were chosen so that the sampled populations covered the entire extant species distribution range (sampling locations given in Fig. 1b). Only populations from which at least two individuals had been sequenced by Fang et al. (2018) were included as population size estimates could otherwise be unreliable. Within the sampled populations, we chose the two individuals sequenced from Fang et al. (2018) that had the highest sequencing depth to minimize missing data as SNAPP requires a SNP matrix in which each site has data in at least one individual per population. Divergence times were then estimated with the MSC model in SNAPP following Stange et al. (2018). We further estimated divergence times using concatenation as in Fang et al. (2018), applied to the same subset of individuals as used in MSC analyses, to allow a direct comparison of the estimates from the two approaches. In these analyses, we concatenated the same number of loci and used the same calibration strategies (see below) as in the MSC analyses.

2.1. Specification of calibration points

As for the root calibration point, we re-evaluated the divergence time between *G. aculeatus* and the outgroup taxon *G. nipponicus*. Fang et al. (2018) accepted 2 mya as the divergence time between the two species based on genetic divergence and geological information (Higuchi and Goto, 1996; Higuchi et al., 2014). However, a subsequent study redefined this estimate to 0.68–1.0 mya on the basis of multiple demographic analyses (Ravinet et al., 2018). Given these widely varying estimates, the root calibration point of 2 mya used by Fang et al. (2018) might have led to an overestimation of divergence times of the major lineages. Therefore, we re-evaluated the divergence times among different stickleback clades by considering three possible root calibration points (*viz.* 2 mya, 1 mya and 0.68 mya).

In terms of geographic calibrations, Fang et al. (2018) initially used two calibration points which were the Black Sea (7 kya; Gökaşan et al., 1997; Ballard et al., 2000; Ryan and Pitman, 2000) and the Baltic Sea openings (10 kya; Leppäkoski et al., 2002; Nilsson et al., 2001; Voipio, 1981). However, as we ran some initial analyses using the MSC approach we noticed that all calibrations schemes (except the Black Sea calibration) resulted in more ancient times of most recent common ancestor (TMRCA) of the Black Sea and Aegean Sea populations than that of the northern European populations. While Fang et al. (2018) refuted a pre-Pleistocene origin of contemporary Black Sea sticklebacks on the basis of calibration-free phylogenetic analyses, the current Aegean Sea samples might not be the direct sister group of the Black Sea samples. Therefore, the divergence of current samples from the Aegean Sea and Black Sea could predate the origin of the Black Sea population. Consequently, the Black Sea calibration would not be applicable to the samples in this study, and therefore we omitted the Black Sea calibration point from the re-analyses.

Hence, on the basis of the latest research and a re-evaluation of the previously applied calibration points, we used the root divergence times 2 mya, 1 mya and 0.68 mya, and the TMRCA of Baltic Sea 10 kya, as calibration points. The four age constraints were implemented with normally distributed densities centered at 2 mya (SD: 0.08 mya), 1 mya (SD: 0.08 mya), 0.68 mya (SD: 0.05 mya), and 10 kya (SD: 0.00035 mya), respectively. The standard deviations of each calibration were chosen so that 99% of the prior distribution for each of the three root calibration points would not overlap. We first performed independent analyses with each one of the above four calibration points, and then combined each of the three root constraints separately with the young Baltic Sea age constraint in subsequent analyses. Consequently, seven sets of calibration schemes were applied for each of the two approaches (MSC and concatenation).

2.2. Divergence time estimation

We subsampled 37 individuals representing 19 populations, as well as two individuals of the outgroup taxon *G. nipponicus*, using VCFtools v.0.1.15 (Danecek et al., 2011). From a full dataset of 8079 concatenated SNPs for 128 samples (Fang et al., 2018), a set of 1183 biallelic SNPs without any missing data was retained in variant call format (VCF). The genome-wide distribution of 1183 SNPs is shown in Appendix Fig. 1. The VCF file was converted to a PHYLIP format SNP-matrix using the Python script "vcf2phylip" (https://github.com/edgardomortiz/vcf2phylip), and the XML input file for SNAPP analyses was prepared using the Ruby script "snapp_prep.rb" (https://github.com/mmatschiner/snapp_prep).

The input for the Ruby script included the above SNP matrix, a starting tree, population information, divergence-time constraints and the number of Markov-Chain Monte Carlo (MCMC) iterations. For the starting tree, we trimmed the full phylogenetic tree of 128 individuals (Fang et al., 2018) to 39 individuals, and collapsed the trimmed tree into a tree of 19 monophyletic populations, using the R packages APE and BioGeoBEARS (Paradis et al., 2004; Matzke, 2013). To reduce computational demand, the tree topology was set to be fixed in the SNAPP analyses (as in Stange et al., 2018). All input files, pipelines and produced XML files for SNAPP analyses are available from the DRYAD repository (doi: 10.5061/dryad.nvx0k6dn5).

The generated XML files were used for analyses with the SNAPP v.1.3.0 (Bryant et al., 2012) package in BEAST2 v.2.4.7 (Bouckaert et al., 2014). For each calibration strategy, we ran two independent SNAPP analyses with 100 million MCMC iterations. Stationarity and convergence of chains were visually checked in TRACER v1.6 (ESS > 1000; http://beast.bio.ed.ac.uk/tracer/). The program DensiTree v.2.2.6 (Bouckaert, 2010) was used to visualize the SNAPP trees after discarding the first 10% of each MCMC chain as burn-in (Appendix Fig. 1). A maximum-clade-credibility summary tree was generated with TreeAnnotator (Drummond et al., 2012) and visualized in FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) (DRYAD doi: 10.5061/dryad.nvx0k6dn5).

To compare divergence time estimates obtained from MSC and the concatenation approach, a subset of 1183 randomly selected RAD loci (the same number as the unlinked SNPs used in SNAPP) was concatenated from the iPyRAD "loci" output file from Fang et al. (2018) using the R package Radami (Hipp et al., 2014). Employing the seven calibration schemes, the 1183 concatenated loci of the same set of samples (39 individuals) were used for concatenation-based phylogenetic inference in BEAST2 v.2.4.7 as in Fang et al. (2018).

In addition, to assess the possible effect of subsampling individuals, we performed the analyses described above with the full dataset (128 individuals) of Fang et al. (2018; i.e. phylogenetic inference based on 1183 concatenated loci of 128 individuals under seven calibration schemes). Comparison of the divergence times inferred from the full and the subsampled datasets suggested that the two sets of samples yielded similar divergence time estimates for all nodes, showing highly overlapping estimates (bounds of the 95% highest probability density, HPD) under all seven calibrations schemes (Appendix Fig. 2). Hence, the subsample of 39 individuals used in the study was considered to yield reliable and comparable divergence time estimates as the full dataset used by Fang et al. (2018).

3. Results and discussion

The inferred divergence times were influenced both by the use of different calibration schemes and by the choice of analytical approach (concatenation *vs.* MSC; Fig. 1a). Combining the estimates from the two approaches, the estimated TMRCA for the extant Northern European lineages (node C) varied from 11.3–18.3 kya (single Baltic Sea calibration under MSC) to 137.7–460.5 kya (single 2 mya root calibration under concatenation). The divergence time between the Atlantic and Pacific lineages (node B), on the other hand, was estimated ranging from 29.5–71.9 kya (single Baltic Sea calibration under concatenation) to 295.3–959.3 kya (single 2 mya root calibration under concatenation). Moreover, the estimates for the TMRCA of all three-spined sticklebacks (node A) varied between 36.9–96.7 kya (single Baltic Sea calibration under concatenation) and 383.1–1250.1 kya (single 2 mya root calibration under concatenation; Fig. 1a).

However, several of these estimates set the origin of the Baltic Sea populations to predate the LGM (Fig. 1a), which is highly unlikely given the fact that the Baltic Sea did not exist during the LGM (Bell and Foster, 1994; McKinnon and Rundle, 2002; Clark et al., 2009; Stroeven et al., 2016; Fang et al., 2018). For instance, in the analyses using concatenated data, all the three strategies with single root calibration (in this case with age constraints at 0.68, 1 and 2 mya) produced divergence time estimates for the Baltic Sea population that predated the formation of the Baltic Sea. Stange et al. (2018) found that root calibration in combination with concatenated data can lead to extreme overestimation of young divergence times, and in light these known biases, we conclude that the divergence time estimates of young nodes based on analyses of concatenated data with root calibration are likely to be strongly biased. In the analyses performed under the MSC, on the other hand, only the application of the single most ancient root age constraint (2 mya) produced divergence times in conflict with known geological history (Fig. 1a). Whenever both calibration points (root calibration and Baltic Sea) were used, both methods yielded similar (i.e. with overlapping HPD) divergence time estimates, though concatenation yielded always wider HPDs.

Using analytical approaches that are likely to under- or overestimate divergence times due to different potential biases, and using seven sets of calibrations reflecting uncertainty about the true divergence time between *G. aculeatus* and *G. nipponicus,* we can propose the range of divergence times among major three-spined stickleback clades. We frame the divergence time ranges for nodes A, B and C, by casting the maximum and minimum divergence time among those estimates (HPDs). If we exclude the analyses of concatenated data using only root calibration, which have been shown by Stange et al. (2018) to overestimate young divergence times and produced in this study estimates that clearly conflict with the known geological history of the Baltic Sea, we conclude that the divergence among major clades (i.e. nodes A, B and C; Fig. 1) of the three-spined sticklebacks occurred as follows: The TMRCA of all *G. aculeatus* populations is 36.9–346.5 kya, the divergence between Pacific and Atlantic lineages occurred 29.5–226.6 kya, and the TMRCA for the extant Northern European lineages is set at 11.3–95.2 kya.

The new TMRCAs for nodes A, B and C are older than the original estimates in Fang et al. (2018; Fig. 1a). Likewise, the upper bounds of the new estimates are higher than those of the original estimates, with greater increments for older nodes. For nodes A, B and C, the upper bounds for the new estimates are 4.4 times (346.5 *vs.* 78.7 kya), 3.8 times (226.6 *vs.* 59.4 kya) and 2.5 times (95.2 *vs.* 38.0 kya) higher than the Fang et al (2018) estimates, respectively, while the lower bounds of the estimates are similar (node A: 36.9 *vs.* 37.3 kya; node B: 29.5 *vs.* 31.3 kya; node C: 11.3 *vs.* 17.0 kya). As re-analyses of the dataset from Fang et al. (2018) using the same concatenation approach, but excluding the Black Sea age constraint, also produced older estimates (Appendix Fig. 2), we conclude that not only the use of concatenation, but also the incorrect choice of one of the calibration points, may have

biased previous estimates. However, it should also be noted that TMRCA estimates under the MSC method in this study for all three-spined sticklebacks (node A), as well as for the Pacific and Atlantic lineages (node B), could be biased because the implemented model assumes a constant and equal population size across all lineages. This assumption is almost certainly violated as some of the stickleback lineages have undergone recent range expansions likely involving population size bottlenecks (McKinnon and Rundle, 2002). Nevertheless, our analyses confirm that the range expansion from which all sampled populations of *G. aculeatus* originate took place no earlier than ca. 1.25 mya. Under the assumption that the divergence between *G. aculeatus* and *G. nipponicus* took place 0.68 mya (Ravinet et al. 2018), the range expansion of *G. aculeatus* likely happened around 62.7–439.7 kya. This confirms the hypothesis put forth by Orti et al. (1994), and later corroborated by Fang et al. (2018), that extant Atlantic populations originate from a recent re-colonization event from the Pacific, and the Late Pliocene/Early Pleistocene Atlantic populations of *G. aculeatus* – for which there is fossil evidence (Bell and Foster, 1994; Foster, 1995) – had gone extinct before the Atlantic Ocean became re-colonized.

The two analytical approaches (concatenation versus MSC) independently produced comparable but non-identical TMRCA estimates, adding uncertainty to the results. As both methods have their specific shortcomings and may bias age estimates in opposite directions when the model assumptions are violated, we argue that the application of both methods provides an opportunity to embrace the underlying uncertainties. Therefore, we suggest that in the absence of more realistic models for divergence time inference based on SNP data, analysis of directly comparable datasets with different models, each making different assumptions about the data, is the most reasonable approach to quantify the uncertainty in divergence time estimates.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106646.

References

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.

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.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372–1373.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.

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.

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.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324.

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., et al., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

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.

Fang, B., Merila, J., Ribeiro, F., Alexandre, C.M., Momigliano, P., 2018. Worldwide phylogeny of three-spined sticklebacks. Mol. Phylogenet. Evol. 127, 613–625.

Foster, S.A., 1995. Understanding the evolution of behavior in threespine stickleback: the value of geographic variation. Behaviour 132, 1107–1129.

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.

Higuchi, M., Goto, A., 1996. Genetic evidence supporting the existence of two distinct species in the genus Gasterosteus around Japan. Environ. Biol. Fish. 47, 1–16.

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., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol. Biol. Evol. 22, 1561–1568.

Kern, A.D., Hahn, M.W., 2018. The neutral theory in light of natural selection. Mol. Biol. Evol. 35, 1366–1371.

Leppäkoski, E., Gollasch, S., Gruszka, P., Ojaveer, H., Olenin, S., Panov, V., 2002. The Baltic – a sea of invaders. Can. J. Fish Aquat. Sci. 59, 1175–1188.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30.

Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2, p. 2013.

McKinnon, J.S., Rundle, H.D., 2002. Speciation in nature: the threespine stickleback model systems. Trends Ecol. Evol. 17, 480–488.

Mendes, F.K., Hahn, M.W., 2016. Gene tree discordance causes apparent substitution rate variation. Syst. Biol. 65, 711–721.

Membrebe, J.V., Suchard, M.A., Rambaut, A., Baele, G., Lemey, P., 2019. Bayesian inference of evolutionary histories under time-dependent substitution rates. Mol. Biol.

Evol. msz094.

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.

Ogilvie, H.A., Bouckaert, R.R., Drummond, A.J., 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34, 2101–2114.

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.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., Kitano, J., 2018. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. PLoS Genet. 14, e1007358.

Ryan, W., Pitman, W., 2000. Noah's Flood: The New Scientific Discoveries About the Event that Changed History. Simon and Schuster.

Stange, M., Sanchez-Villagra, M.R., Salzburger, W., Matschiner, M., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (*Ariidae*) supports Miocene closure of the Panamanian Isthmus. Syst. Biol. 67, 681–699.

Stroeven, A.P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B.W., Harbor, J.M., Jansen, J.D., Olsen, L., Caffee, M.W., Fink, D., Lundqvist, J., Rosqvist, G.C., Strömberg, B., Jansson, K.N., 2016. Deglaciation of Fennoscandia. Quart. Sci. Rev. 147, 91–121.

Taylor, E.B., McPhail, J.D., 1999. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. Biol. J. Linn. Soc. 66, 271–291.

Taylor, E.B., McPhail, J.D., 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. Proc. Biol. Sci. B 267, 2375–2384.

Voipio, A., 1981. The Baltic Sea. Elsevier.