Supplementary Information to

Speciation in stickleback facilitated by admixture – where is the evidence?

By Daniel Berner

Contents:

Supplementary Methods

Phylogeny of European freshwater threespine stickleback

The phylogenetic analysis involves stickleback samples from 39 localities in and around central Europe, most of which are represented by two individuals (69 individuals in total; Supplementary Table 1). Because of strong adaptive genetic divergence between marine and freshwater stickleback populations, all 38 European localities concern freshwater habitat; only a single locality (CLU; Cluxewe Estuary, Vancouver Island, Canada) included as outgroup represents saltwater habitat (anadromous marine stickleback). All populations are at least potentially natural. The only exceptions are the populations CHE and SAS, which originate from human introduction to the Lake Geneva basin around 1900 (Fatio 1882; Bertin 1925). The sequence data underlying this analysis are Sbf1 or Pst1 enzyme restriction site-associated DNA (RAD) sequences generated specifically for this study, or retrieved from published investigations (Roesti et al. 2012b, 2014, 2015; Ferchaud & Hansen 2016; Marques et al. 2016; Fang et al. 2018). In the latter case, individuals with high read coverage were given priority. The full data set is described in detail in Supplementary Table 1.

To obtain single-nucleotide polymorphisms (SNPs) for phylogenetic analysis, all raw fastq files were initially filtered for reads starting with the exact Sbf1 restriction residual (TGCAGG; Sbf1 restriction sites are covered by the Pst1 enzyme too), and the reads were trimmed to 70 base pairs (bp). The fastq data thus obtained were aligned to the Glazer et al. 2015 threespine stickleback reference genome assembly with Novoalign v3.00 (http://www.novocraft.com/products/novoalign), using the alignment parameters from Roesti et al. 2012a (key settings: -t180 -g40 -x15). The alignments were then converted to BAM format and accessed with the R package *Rsamtools* (Morgan et al. 2017). At each RAD locus in each individual, haploid genotyping was performed by retrieving the leading haplotype, defined as the single sequence exhibiting the highest read count among all unique sequences present at the RAD locus. RAD loci at which this leading haplotype did not occur in at least two copies, or exhibiting an excessive read depth beyond 4.5 times the expected read depth across all genome-wide RAD loci (estimated by the total number of reads divided by the total number of RAD loci), were excluded from analysis. This haploid genotyping approach was chosen because it avoids potential bias in the identification of heterozygous positions and is

therefore highly reliable. RAD loci successfully genotyped in every single individual were then used for SNP detection (hence, the final SNP data set contained no missing data). I accepted SNPs along a RAD locus only if they were at least 8 bp away from the previous polymorphic position, thus avoiding pseudo-SNPs arising from indels. For each individual, the nucleotides present at all SNPs were concatenated to a single string, and these strings combined across all 69 individuals in a single fasta file. Overall, the fasta file contained genotype information from 7,121 SNPs from 4,429 genome-wide RAD loci shared among all individuals. As a robustness check, the above SNP detection and genotyping protocol was repeated using more stringent criteria: for an individual, the leading haplotype at a RAD locus was accepted only if present in at least five (as opposed to two) copies, and the minimum spacing threshold for SNPs located on the same RAD locus was increased from 8 to 12 bp. This latter approach, yielding 797 SNPs from 563 RAD loci, produced a very similar phylogenetic tree topology leading to the same conclusions (details not presented). Both fasta files are provided as Supplementary Data 1 and 2.

Analyses based on the above fasta files were carried out using the R packages *ape* (Paradis & Schliep 2018) and *phangorn* (Schliep 2011). For the phylogeny, I first determined the most appropriate substitution model ('GTR+G+I'), estimated the maximum likelihood tree, and visualized this tree as phylogram. In addition, genetic similarity among the individuals was quantified by ordination of a genetic distance matrix using Principal Coordinates Analysis (PCoA). Individuals were plotted along the first two PCoA axes.

Genetic diversity

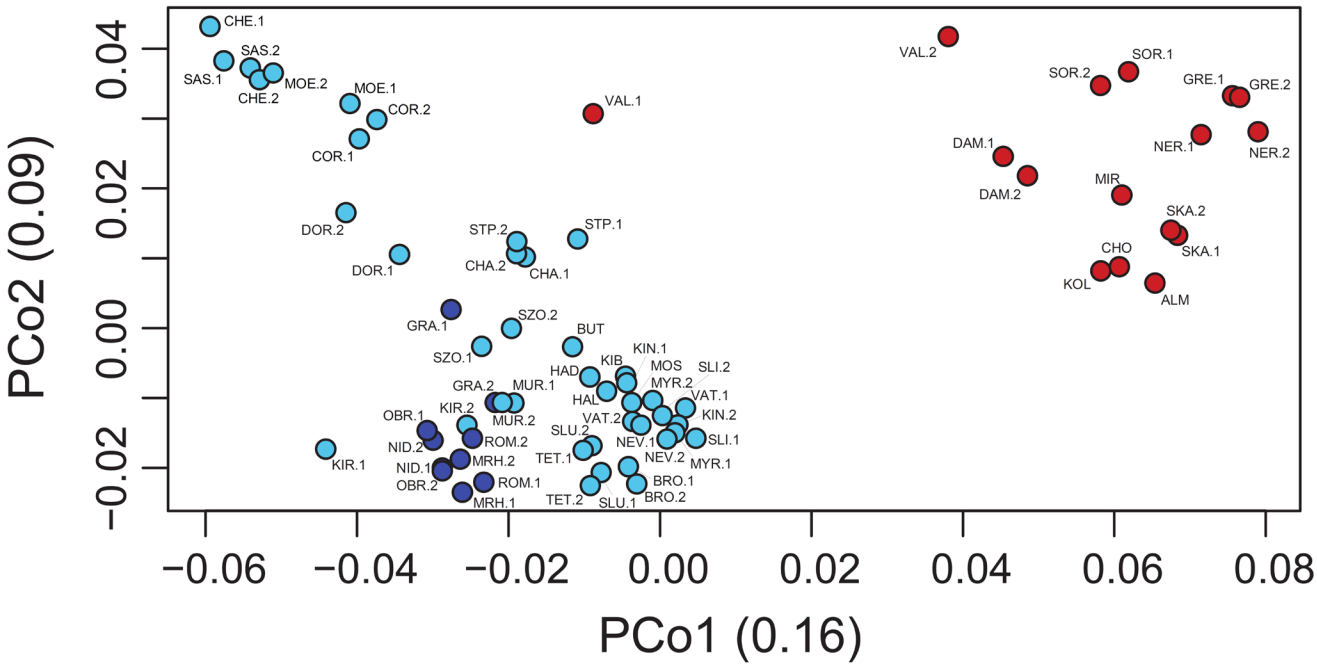
Genetic diversity within populations was quantified based on the haploid individual-level SNP genotype data generated as described for the phylogenetic analysis (minimum leading haplotype coverage = 2; minimum SNP spacing = 8 bp). Diversity was expressed by the proportion of SNPs at which two (haploid) individuals from a given population exhibited a distinct nucleotide. This metric, representing a close analog of within-population heterozygosity, required genotype data from two individuals from each population, thus excluding all study populations for which sequence data from only a single individual was available (Supplementary Table 1). Because genetic diversity is typically greater in marine stickleback populations – considered large and well-connected – than in derived freshwater

populations (Mäkinen et al. 2006; Hohenlohe et al. 2010; Catchen et al. 2013; Haenel et al. 2019), I included samples from two marine localities (NOR, Norsminde Fjord, North Sea, Ferchaud & Hansen 2016; PRI, Primorsk, Baltic Sea, Fang et al. 2018) as a robustness check. Overall, this analysis used genotype data from 7,247 SNPs located on 4,474 RAD loci, and considered 32 localities. For visualization, genetic diversity was standardized by using the sample with the greatest diversity (CLU; Pacific marine stickleback) as reference. The analysis of genetic diversity was repeated using data generated with more stringent thresholds for genotyping (minimum leading haplotype coverage = 5; minimum SNP spacing = 12 bp), which produced very similar results supporting the same conclusions.

Supplementary Table 1. Description of the RAD sequence data underlying the population genetic analyses. Accession codes refer to the NCBI Sequence Read Archive.

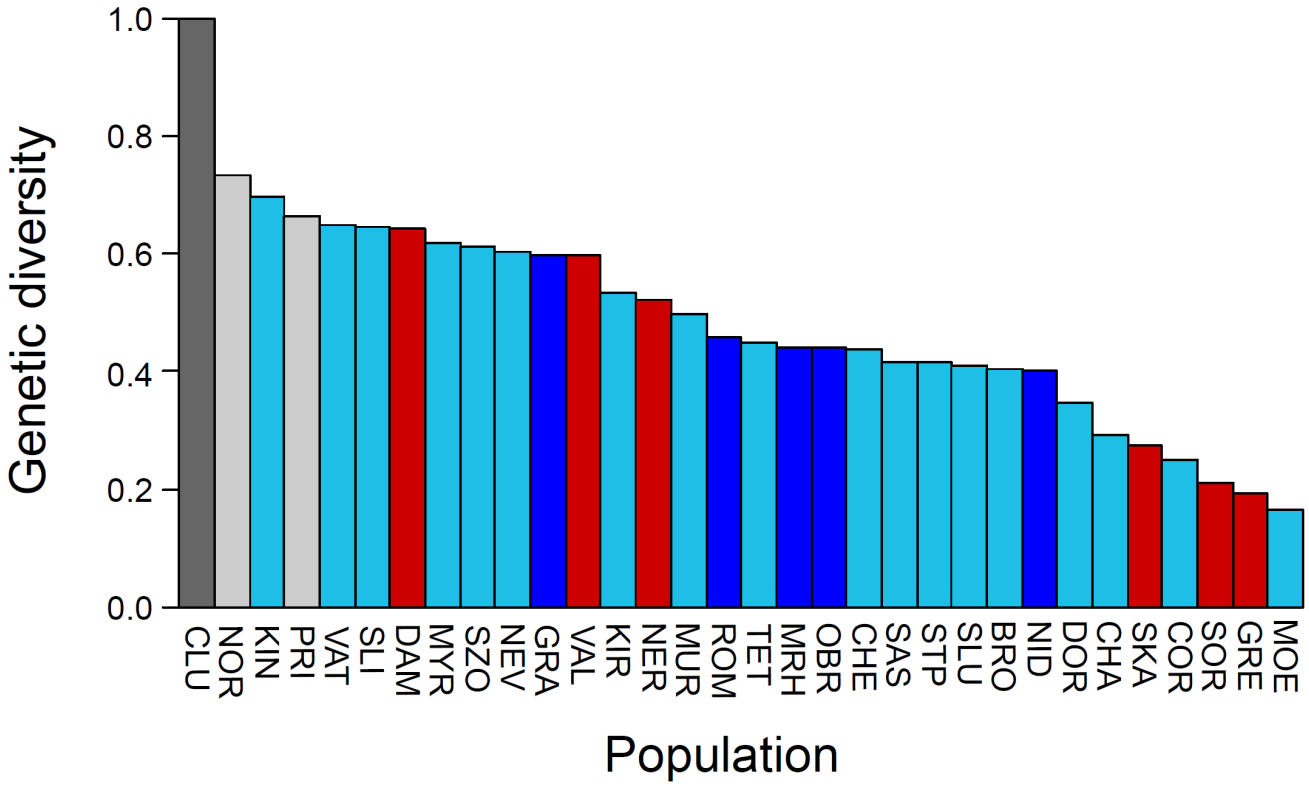
Individual	Locality	Country	Restriction enzyme	Accession	Reference
ALM	Alma	Ukraine	Pst1	SRX3998066	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
BRO.1	Broszkowice	Poland	Sbf1	SRX6084984	Marques et al. 2019, Nat. Commun. 10: 4240
BRO.2	Broszkowice	Poland	Sbf1	SRX6084934	Marques et al. 2019, Nat. Commun. 10: 4240
BUT	Butler pond	UK	Pst1	SRX3997979	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
CHA.1	Butler pond	France	Pst1	SRX3997970	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
CHA.2	Chamoux	France	Pst1	SRX3997983	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
CHE.1	Chessel	Switzerland	Sbf1	SRX1038796	Roesti et al. 2015, Nat. Commun. 6: 8767 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
CHE.2	Chessel	Switzerland	Sbf1	SRX1038804	Roesti et al. 2015, Nat. Commun. 6: 8767 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
CHO	Chornaya	Ukraine	Pst1	SRX3998069	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
CLU.1	Cluxewe Estuary, Vancouver Island	Canada	Sbf1	SRX456700	Roesti et al. 2014, Mol. Ecol. 23: 3944-3956
CLU.2	Cluxewe Estuary, Vancouver Island	Canada	Sbf1	SRX456701	Roesti et al. 2014, Mol. Ecol. 23: 3944-3956
COR.1	Cormoz	France	Sbf1	SRX6084965	Marques et al. 2019, Nat. Commun. 10: 4240
COR.2	Cormoz	France	Sbf1	SRX6084963	Marques et al. 2019, Nat. Commun. 10: 4240
DAM.1	Treilles	France	Pst1	SRX3998041	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
DAM.2	Treilles	France	Pst1	SRX3998036	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
DOR.1	Basel	Switzerland	Sbf1	SRX6864080	this study (locality described in Moser et al. 2012, PLoS ONE 7: e50620)
DOR.2	Basel	Switzerland	Sbf1	SRX6864081	this study (locality described in Moser et al. 2012, PLoS ONE 7: e50620)
GRA.1	Grasbeuren	Germany	Sbf1	SRX6864092	this study (locality described in Moser et al. 2012, PLoS ONE 7: e50620)
GRA.2	Grasbeuren	Germany	Sbf1	SRX6864103	this study (locality described in Moser et al. 2012, PLoS ONE 7: e50620)
GRE.1	Spercheios	Greece	Pst1	SRX3998070	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
GRE.2	Spercheios	Greece	Pst1	SRX3998071	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
HAD	Hadsten	Denmark	Sbf1	SAMN03076274 (Had3)	Ferchaud & Hansen 2016, Mol. Ecol. 25: 238-259
HAL	Hald	Denmark	Sbf1	SAMN03076274 (Hal3)	Ferchaud & Hansen 2016, Mol. Ecol. 25: 238-259
KIB	Kibaek Molledam	Denmark	Sbf1	SAMN03076274 (Kib7)	Ferchaud & Hansen 2016, Mol. Ecol. 25: 238-259
KIN.1	Kinness Burn	UK	Pst1	SRX3998013	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
KIN.2	Kinness Burn	UK	Pst1	SRX3998012	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
KIR.1	Kirchbierlingen	Germany	Sbf1	SRX6864114	this study (locality described in Moser et al. 2012, PLoS ONE 7: e50620)
KIR.2	Kirchbierlingen	Germany	Sbf1	SRX6864125	this study (locality described in Moser et al. 2012, PLoS ONE 7: e50620)
KOL	Kolanraes	Ukraine	Pst1	SRX3998068	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
MIR	Mirna	Croatia	Pst1	SRX3998033	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
MOE.1	Moehlin	Switzerland	Sbf1	SRX6084954	Marques et al. 2019, Nat. Commun. 10: 4240
MOE.2	Moehlin	Switzerland	Sbf1	SRX6084993	Marques et al. 2019, Nat. Commun. 10: 4240
MOS	Mosso	Denmark	Sbf1	SAMN03076274 (Mos4)	Ferchaud & Hansen 2016, Mol. Ecol. 25: 238-259
MRH	Marina Rheinhof	Switzerland	Sbf1	SRS1271407	Marques et al. 2016, PLoS Genet. 12: e1005887
MRH	Marina Rheinhof	Switzerland	Sbf1	SRS1271408	Marques et al. 2016, PLoS Genet. 12: e1005887
MUR.1	Mura	Slowenia	Pst1	SRX3997988	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
MUR.2	Mura	Slowenia	Pst1	SRX3997989	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
MYR	Myrdalsvatnet	Norway	Pst1	SRX3998020	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
MYR	Myrdalsvatnet	Norway	Pst1	SRX3997992	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
NER.1	Neretva	Croatia	Sbf1	SRX6864129	this study (locality: Norin stream, 43.053475 N, 17.595780 E)
NER.2	Neretva	Croatia	Sbf1	SRX6864130	this study (locality: Norin stream, 43.053475 N, 17.595780 E)
NEV.1	Nevezis	Lithuania	Pst1	SRX3997981	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
NEV.2	Nevezis	Lithuania	Pst1	SRX3997990	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
NID.1	Nideraach	Switzerland	Sbf1	SRS257619	Roesti et al. 2012, BMC Evol. Biol. 12: 94 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
NID.2	Nideraach	Switzerland	Sbf1	SRS257621	Roesti et al. 2012, BMC Evol. Biol. 12: 94 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
OBR.1	Oberriet	Switzerland	Sbf1	SRS1271421	Marques et al. 2016, PLoS Genet. 12: e1005887
OBR.2	Oberriet	Switzerland	Sbf1	SRS1271424	Marques et al. 2016, PLoS Genet. 12: e1005887
ROM.1	Romanshorn	Switzerland	Sbf1	SRS257608	Roesti et al. 2012, BMC Evol. Biol. 12: 94 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
ROM.2	Romanshorn	Switzerland	Sbf1	SRS257609	Roesti et al. 2012, BMC Evol. Biol. 12: 94 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
SAS.1	Saint-Sulpice	Switzerland	Sbf1	SRX1038830	Roesti et al. 2015, Nat. Commun. 6: 8767 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
SAS.2	Saint-Sulpice	Switzerland	Sbf1	SRX1038827	Roesti et al. 2015, Nat. Commun. 6: 8767 (locality described in Berner et al. 2010, Mol. Ecol. 19: 4963-4978)
SKA.1	Skadar	Albania	Pst1	SRX3997967	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SKA.2	Skadar	Albania	Pst1	SRX3997966	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SLI.1	Nevan	Russia	Pst1	SRX3998001	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SLI.2	Nevan	Russia	Pst1	SRX3997998	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SLU.1	Sluch	Ukraine	Pst1	SRX3997982	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SLU.2	Sluch	Ukraine	Pst1	SRX3998063	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SOR.1	Sorgue	France	Sbf1	SRX6084942	Marques et al. 2019, Nat. Commun. 10: 4240
SOR.2	Sorgue	France	Sbf1	SRX6084941	Marques et al. 2019, Nat. Commun. 10: 4240
STP.1	Saint-Pourcain-sur-Sioule	France	Sbf1	SRX6101711	Marques et al. 2019, Nat. Commun. 10: 4240
STP.2	Saint-Pourcain-sur-Sioule	France	Sbf1	SRX6101716	Marques et al. 2019, Nat. Commun. 10: 4240
SZO.1	Szödliget	Hungary	Pst1	SRX3997987	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
SZO.2	Szödliget	Hungary	Pst1	SRX3997986	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
TET.1	Teterev	Ukraine	Pst1	SRX3998065	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
TET.2	Teterev	Ukraine	Pst1	SRX3998062	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
VAL.1	Valence	France	Pst1	SRX3997984	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
VAL.2	Valence	France	Pst1	SRX3997985	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
VAT.1	Vättern	Sweden	Pst1	SRX3998050	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625
VAT.2	Vättern	Sweden	Pst1	SRX3998051	Fang et al. 2018, Mol. Phylogenet. Evol. 127: 613-625

Supplementary Fig. 1.



Supplementary Fig. 1. Genetic relationship among the 69 individuals included in the phylogeny (Fig. 1), as expressed by Principal Coordinates Analysis (PcoA) of individual SNP genotype data. The individuals are mapped along the first two ordination axes (Principal Coordinates 1 and 2). The relative importance of these axes (relative eigenvalues) is given in parentheses. To increase the visual resolution for the European individuals, the two CLU individuals from Canada were excluded from the graphic. The color coding follows Fig. 1. Further methodological detail is given in the Supplementary Methods.

Supplementary Fig. 2.



Supplementary Fig. 2. Genetic diversity within stickleback populations, quantified by an analog of heterozygosity (details given in the Supplementary Methods). The populations and color coding are as in Fig. 1, except that all localities from which only a single individual was available (Supplementary Table 1) were ignored. Gray bars show marine stickleback additionally included in this analysis (dark gray for Pacific fish, light for Atlantic fish). The populations are ranked by decreasing diversity, using the most variable population as benchmark. Consistent with Pacific stickleback being ancestral to all Atlantic stickleback Fang et al. 2018, the greatest genetic diversity is found in the Pacific marine outgroup sample (CLU), followed by Atlantic marine stickleback (NOR, PRI; the KIN sample is reported as freshwater in Fang et al. 2018 but originates from a river mouth in immediate proximity to the Ocean, hence this sample may represent anadromous marine fish as well). The expected high diversity observed in marine samples, particularly in the most ancestral one from the Pacific (Fang et al. 2018), confirms the soundness of the metric used for quantification. Note that the populations from the Lake Constance basin (dark blue) do not display elevated

genetic diversity relative to other European freshwater stickleback, contrary to expectations based on recent admixture between ancient lineages within this basin.

Supplementary References

Bertin, L. 1925. Recherches bionomiques, biométriques et systématiques sur les épinoches (Gastérostéidés). Blondel La Rougery, Paris.

Catchen, J. M., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q. & Cresko, W. A. 2013. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol. Ecol.* **22**: 2864-2883.

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

Fatio, V. 1882. Faune des Vertébrés de la Suisse - Histoire naturelle des Poissons. H. Georg, Genève.

Ferchaud, A.-L. & Hansen, M. M. 2016. The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: threespine sticklebacks in divergent environments. *Mol. Ecol.* 25: 238-259.

Haenel, Q., Roesti, M., Moser, D., MacColl, A. D. C. & Berner, D. 2019. Predictable genome-wide sorting of standing genetic variation during parallel adaptation to basic versus acidic environments in stickleback fish. *Evol. Lett.* **3**: 28-42.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. & Cresko, W. A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**: e1000862.

Mäkinen, H. S., Cano, J. M. & Merilä, J. 2006. Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. *Mol. Ecol.* **15**: 1519-1534.

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L. & Seehausen, O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet.* **12**: e1005887.

Morgan, M., Pages, H., Obenchain, V. & Hayden, N. 2017. Rsamtools: binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. *R package version 1.3.0 (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html)*.

Paradis, E. & Schliep, K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526-528.

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W. & Berner, D. 2014. The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23**: 3944-3956.

Roesti, M., Kueng, B., Moser, D. & Berner, D. 2015. The genomics of ecological vicariance in threespine stickleback fish. *Nat. Commun.* **6**: 8767.

Roesti, M., Hendry, A. P., Salzburger, W. & Berner, D. 2012a. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.* **21**: 2852-2862.

Roesti, M., Salzburger, W. & Berner, D. 2012b. Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evol. Biol.* **12**: 94.

Schliep, K. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592-593.