ARTICLE

Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback

David A. Marques 1,2,3,7, Kay Lucek 4,7, Vitor C. Sousa 3,5, Laurent Excoffier 3,6 & Ole Seehausen 1,2

Ecological speciation can sometimes rapidly generate reproductively isolated populations coexisting in sympatry, but the origin of genetic variation permitting this is rarely known. We previously explored the genomics of very recent ecological speciation into lake and stream ecotypes in stickleback from Lake Constance. Here, we reconstruct the origin of alleles underlying ecological speciation by combining demographic modelling on genome-wide single nucleotide polymorphisms, phenotypic data and mitochondrial sequence data in the wider European biogeographical context. We find that parallel differentiation between lake and stream ecotypes across replicate lake-stream ecotones resulted from recent secondary contact and admixture between old East and West European lineages. Unexpectedly, West European alleles that introgressed across the hybrid zone at the western end of the lake, were recruited to genomic islands of differentiation between ecotypes at the eastern end of the lake. Our results highlight an overlooked outcome of secondary contact: ecological speciation facilitated by admixture variation.

https://doi.org/10.1038/s41467-019-12182-w

1 Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland. 2 Department of Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center for Ecology, Evolution and Biogeochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland. 3 Computational and Molecular Population Genetics, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland. 4 Department of Environmental Sciences, University of Basel, Schönbeinstrasse 6, CH-4056 Basel, Switzerland. 5 Centre for Ecology, Evolution and Environmental Changes, University of Lisbon, Campo Grande 016, 1749-016 Lisbon, Portugal. 6 Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. 7 These authors contributed equally: David A. Marques, Kay Lucek. Correspondence and requests for materials should be addressed to O.S. (email: ole.seehausen@eawag.ch)
Contemporary speciation studies have shown that speciation can sometimes be surprisingly fast even in the face of gene flow, allowing for sympatric divergence or persistence of incipient species. Many of these cases involve divergent natural selection or habitat-dependent sexual selection. However, the origin of genetic variants underlying rapid speciation and reproductive isolation has remained unknown in all but a few cases, e.g., whether loci that contribute to reducing gene flow between populations are derived from de novo mutation, from standing genetic variation or from admixture variation acquired through introgression between divergent lineages. Understanding the process of speciation and its constraints requires knowledge on the origin of alleles under ecological, sexual or incompatibility selection. In turn, understanding the origin of alleles requires reconstructing the history of populations undergoing speciation.

We recently documented a case of contemporary ecological speciation in threespine stickleback (Gasterosteus aculeatus complex) of Lake Constance, Central Europe. Lake- and stream-adapted ecotypes differ in predator defense and feeding morphology, ecology, nuptial coloration, migration behavior, and life history: lake stickleback grow larger than stream stickleback, are covered with larger or more lateral bony plates and possess longer spines as protection from predators. Migratory stream rakers and jaws adapted to feeding on zooplankton instead of benthiic invertebrates, migrate to lower reaches of streams to benthic invertebrates, migrate to lower reaches of streams to breed in contrast to resident stream stickleback and start breeding 1 year later and die older than stream stickleback.

In a South-Eastern tributary of Lake Constance, both ecotypes breed in sympathy and they maintain phenotypic and genomic differentiation despite ongoing gene flow. Sympatric breeding of stickleback ecotypes is very rare and occurs here at a surprisingly early stage of speciation, given that historical records document the presence of stickleback in the Lake Constance catchment for the past 150 years only, and their prior absence from the basin. The genomic architecture of this case is characterized by an undifferentiated genomic background interspersed by strong differentiation across multiple chromosomes, especially in low recombination regions and inversions enriched with quantitative trait loci (QTL) for divergent traits. The lack of reduced diversity within genomic islands and of genomic background differentiation led us to hypothesize that ecotypes likely diverged in situ in Lake Constance, either from selection on standing genetic variation or admixture variation. From Lake Constance data alone, however, we were unable to distinguish these alternative origins of alleles.

A second study, including tributaries North and West of Lake Constance, has documented a similar genomic architecture but stronger genomic background differentiation in some parapatric lake vs. stream ecotype comparisons. While standing genetic variation was also suggested as substrate for ecotype divergence, the study came to different conclusions regarding the mode and age of ecotype divergence. The authors estimated that ecotypes have been diverging for ~4500 generations, translating to ~9000 years or an early post-glacial divergence. Such a long time for ecological speciation corresponds to what has been reported in other well-studied cases of sympatric stickleback species, but is at odds with the historical ichthyological literature. A mode of ecotype divergence termed ‘ecological vicariance’ was proposed in which a stream-adapted stickleback lineage colonized the streams of the area first, becoming isolated in different Lake Constance tributary streams due to the lake acting as a barrier for stream-adapted fish, followed by reconnection of populations once a lake-adapted ecotype had evolved from standing genetic variation.

Here, we re-evaluate the population history of Lake Constance stickleback and investigate the origin of genetic variants underlying contemporary ecological speciation and early persistence in sympathy. We place all previously studied lake and stream populations from the Lake Constance catchment in a wider European phylogeographic context, using new and published genome-wide single nucleotide polymorphism (SNP), mitochondrial, microsatellite and plate morph data. We compare the fit of alternative demographic models to genomic data and estimate demographic parameters from the best-fitting model. The models we compare are: (1) primary divergence in situ, in which lake and stream ecotypes have recently diverged from standing genetic variation in a single lineage that colonized Lake Constance; (2) ecological vicariance (outlined above); (3) secondary contact, in which ecotypes correspond to West and East European lineages that have diverged in allopatry, recently met and sorted between lake and stream habitats with exchange of genes at the lake-stream boundaries; and (4) hybrid origin, in which one of the ecotypes has recently evolved through hybridization between divergent West and East European lineages or sorting of admixture variation following introgression from one divergent lineage into the other lineage.

Our analyses reveal admixture variation as genetic source of ecotype differentiation and thus contemporary ecological speciation, derived from hybridization between two divergent lineages from at least two European watersheds. We find that stream ecotypes in Lake Constance show a gradient of admixture between divergent European stickleback lineages: near the zone of secondary contact, stream ecotypes show a 50:50 hybrid origin, while far from the zone of contact, novel stream ecotypes evolved in situ from predominantly one genomic background, aided by sorting of introgressed alleles. Our analyses provide evidence for the hypothesis that admixture variation can be an important facilitator of rapid ecological speciation and adaptive radiation and uncover an unexpected outcome of secondary contact: repeated ecological speciation beyond the contact zone. Our results imply that caution should be taken when inferring modes and times of speciation from population genome sequence data in the absence of a thoroughly sampled phylogeographic context.

Results

The origin of Lake Constance stickleback. We identified five different mitochondrial DNA (mtDNA) haplotypes in the Lake Constance catchment (Fig. 1b), all of them matching haplotypes that today can be found in four main European river catchments, draining respectively into the Baltic Sea, the North Sea, the Mediterranean Sea, and the Black Sea. All five mtDNA haplotypes are part of central European lineages, highly divergent from Mediterranean and Black Sea lineages (Fig. 1a, b). Phylogenomic analyses based on concatenated genome-wide SNPs derived from SbfI-restriction-site-associated DNA (RAD) sequencing confirm the position of Lake Constance stickleback among the central European lineages (Fig. 1c, Supplementary Fig. 1). Among these are at least two old sublineages from East and West Europe that represent different nominal species: the West European Gasterosteus gymnurus lacking lateral bony plates except for the structural plates and the fully plated East and North European Gasterosteus aculeatus that also resembles the marine form distributed along the North European coasts.
tributaries to Lake Constance are nearly fixed for an mtDNA haplotype found in East Europe in the Vistula and Upper Danube catchments (Fig. 1b). The nuclear genome of these populations also resembles East European stickleback and phylogenetic analyses with genome-wide SNPs cluster them as sister lineages (Fig. 1c). Most individuals are fully plated, while partially or low-plated individuals occur at low frequency (Fig. 1e), consistent with an East European freshwater or a marine origin of stickleback from Lake Constance and its South-Eastern tributaries.

In contrast, some stream populations from Northern and Western tributaries of the lake appear to be of hybrid origin between West and East European stickleback lineages, meeting in a secondary contact zone where the former tend to inhabit streams and the latter the lake. West European haplotypes that are known otherwise only from the Rhine, the upper Rhone and the southern North Sea occur at high frequency in populations GRA and NID (Fig. 1a, b, see Supplementary Table 1 for population abbreviations). A phylogenetic analysis based on concatenated genome-wide autosomal SNPs derived from...
Fig. 2 Evidence for admixture between West and East European lineages in Lake Constance. a The D-statistic shows significant excess allele sharing between two West European lineages (P3: upper Rhine = FR54, Rhine = AGS1) and stream stickleback in three streams (S: S2, S1, NID) relative to lake stickleback (L: L1). The lake population L1 shows excess allele sharing with the East European sister lineage (P3: Vistula = PL51) only when compared with the stream population NID, suggesting that L1, S2, and S1 contain similarly large proportions of East European ancestry. Error bars indicate ±3 standard deviations around D-estimates. Japan Sea stickleback Gasterosteus nipponicus was used as outgroup (‘Out’), see Supplementary Fig. 3 for near-identical results with other Constance lake populations or outgroups. b Admixture proportions of six different Lake Constance stickleback populations estimated from SNPs that are fixed between East and West European stickleback populations (PLS1 vs. FR54 + AG51, n = 299 SNPs each spaced at least 100 kb apart). A hybrid index of 0 implies that all 299 loci are fixed for the East European allele, 1 implies fixation for the West European alleles. Black box plots delineate the 1st and 3rd quartile, with error bars extend these by max. 1.5 times the interquartile range. Source data are provided as a Source Data file.

SbfI-RAD sequencing clusters the stream population NID with West European lineages rather than with East European lineages, with very high bootstrap support (Fig. 1c). However, when all Lake Constance lake and stream populations are included into the phylogeny, the stream population NID clusters with the other Lake Constance samples as sister of the East European lineage, with slightly reduced bootstrap support (Supplementary Fig. 1), as expected for a population of hybrid origin23. Phenotypically, the two stream populations GR and NID are dominated by low and partially plated stickleback, respectively, in contrast to the fully plated stream populations from South-Eastern tributaries and the lake population (Fig. 1d).

We assessed the presence and extent of admixture between West and East European lineages in the Lake Constance catchment with D-statistics, hybrid index and clustering analyses, using a SbfI-RAD sequencing SNP dataset (see Methods). When Constance and East European populations are treated as sister lineages in a phylogenetic tree, all Lake Constance populations show a significant excess of derived allele sharing with West European populations from the Rhine and the upper Rhine (Supplementary Fig. 2), confirming admixture between West and East European lineages in Lake Constance. The stream population NID shows the strongest signal for the D-statistic, the other stream populations (S1, S2) also show D-statistics significantly different from zero and two lake populations have marginally significant D-statistics (L1, L2, but not ROM, Supplementary Fig. 2), indicating varying extents of admixture with a West European lineage. Overall, populations of the stream ecotype contain more West European alleles than the lake ecotype populations: with lake and stream populations as sisters, all stream populations showed excess allele sharing with West European lineages, with a particularly strong signal in NID (Fig. 2a, Supplementary Fig. 3). Only when contrasted with NID as the stream population, lake populations show excess allele sharing with East European lineages (Fig. 2a, Supplementary Fig. 3), in agreement with a hybrid origin of the NID population.

We estimated admixture proportions by computing the hybrid index from 299 divergently fixed SNPs between the East and West European populations (PLS1 vs. FR54, AGS1), Fig. 2b. Admixture estimates for individuals from population NID ranged from 36–53% West European origin with a mean of 44%, supporting a hybrid origin of this population. Lake and stream populations from South-Eastern tributaries showed lower individual admixture proportions of 21–36% (Fig. 2b), with the stream ecotype (29–30%) showing a slightly elevated hybrid index compared with the lake ecotype (25–26%). In a clustering analysis of SNP and microsatellite data (Supplementary Figs. 4, 5), the Lake Constance stream population GRA even showed a predominantly West European origin while the remaining Lake Constance lake and stream populations showed minor admixture from West European origin.

Contrasting modes of ecotype divergence in Lake Constance. We compared different modes of stickleback ecotype divergence in Lake Constance using a coalescent demographic modeling framework, and taking advantage of the West and East European sister lineages as discussed above. We tested the fit of multiple neutral demographic models to the observed site-frequency spectra (SFS) computed from sites in high recombination rate regions in order to minimize effects of selection on the SFS32. We compared models on three hierarchical levels of complexity (Fig. 3). First, we established the relationships among the three allopatric European stickleback lineages contributing to Lake Constance stickleback using three population models and three-dimensional (3D) SFS (Fig. 3a). This allowed us to estimate split times unaffected by gene flow in Lake Constance at low model and data complexity. Then, we increased complexity by adding single Constance lake or stream populations to the model in order to quantify major (Fig. 2b) and minor (Fig. 2c) contributions of the European lineages in four population models optimized on joint two-dimensional (2D) SFS for all population pairs. Finally, we compared different modes of incipient speciation: primary divergence of ecotypes, ecological vicariance, persistence in secondary contact and hybrid origin, by including one lake/stream population pair into each five population models optimized on joint 2D-SFS (Fig. 2d). We optimized model parameters on observed SFS from two RAD sequencing sets, an SbfI-RAD dataset3,33 including European sister lineages and lake (L2) and stream (S2, NID) populations, and an NsiI-RAD dataset14.
We infer West and East European lineages to have split ~3665 (95% confidence interval (CI): 3636–4877) generations ago, while the upper Rhone and Rhine populations split 1710 (1710–2239) generations ago according to our best 3-population model (Fig. 3a, Supplementary Fig. 6, Table 1). Admixture/gene flow between allopatric European lineages did not significantly improve the model fit, supporting that the three populations we use (FRS4, AGS1, PLS1) are non-admixed (Supplementary Figs. 7, 8, Supplementary Table 2). Adding any single Lake Constance population (L2, S2, NID) to the trio of allopatric European lineages revealed that the respective best supported models all suggest that the majority of their genomes is of East European origin (Fig. 3b, Table 1). Likewise, models with additional contributions from both West European populations (Rhine, Rhone) were not better supported than models with contributions from only one of the two lineages (Fig. 3c, Table 1). Finally, we compared different modes of ecotype divergence: primary divergence in situ, ecological vicariance, secondary contact and hybrid origin. For all pairwise lake-stream comparisons, a hybrid origin model clearly outperformed the other modes of divergence (Supplementary Fig. 8, Table 1). Pink rectangles highlight the best supported models. Little letters in model names denote models with admixture from West European lineages Rhone (<sup>‘o’</sup>), Rhine (<sup>‘i’</sup>), or from both (<sup>‘x’</sup>). Source data are provided as a Source Data file.
Admixture variation fueled incipient speciation. Our analyses reveal that stickleback in Lake Constance and South-Eastern tributaries originate from a phenotypically ‘marine’-like, fully plated East European freshwater lineage (*Gasterosteus aculeatus*), while tributaries North and West of the lake are of hybrid origin with the ‘classical freshwater’-type, low-plated stickleback from West European streams (*Gasterosteus gymnuratus*). Even though stream ecotypes differ in admixture proportions and plate morph distributions (Fig. 1e), all stream ecotypes from around Lake Constance and South-Eastern tributaries differ in admixture proportions and plate morph distributions (Fig. 1e).

Table 1 Fit of demographic models to the observed data

| Dataset | SbfI(1-3) | SbfI(1-4) | NsiI(1-3,7) | SbfI(1-3,6) | SbfI(1-3,7) |
|---------|-----------|-----------|-------------|-------------|-------------|
| Model   | ΔLL       | ΔAIC      | ΔLL         | ΔAIC        | ΔLL         |
| PD      | 650       | −1063     | 1532        | −1304       | −313        |
| EVa     | 642       | −1028     | 1535        | −1321       | −318        |
| EVb     | 647       | −1053     | 1528        | −1289       | −310        |
| HOax*   | 437       | −89       | 1421        | −799        | −205        |
| HOai*   | 440       | −101      | 1420        | −790        | −160        |
| HOao    | 461       | −199      | 1428        | −827        | −278        |
| HObx*   | 424       | −29*      | 1337        | −411        | −7*         |
| HObi*   | 418       | 0*        | 1337        | −410        | −9*         |
| HOb*    | 453       | −160      | 1355        | −491        | −31*        |
| HOcx*   | 421       | −20*      | 1249        | −12*        | 0*          |
| HOct*   | 417       | 3*        | 1247        | 0*          | 191         |
| HSc*    | 454       | −170      | 1283        | −166        | −37*        |
| SCI     | 711       | −1346     | 1436        | −860        | −1811       |
| SCo     | 784       | −1862     | 1529        | −1288       | −434        |

Best-fitting models and models with very similar likelihood are marked with an asterisk (*). Shown are log₂ likelihood differences between observed and expected site-frequency spectra (ΔLL) and difference in Akaike information criterion (ΔAIC) between the best and all models for a given dataset. For NsiI-data, the three West European lineages were modeled as unsampled (‘ghost’) populations, with parameters for the latter fixed to best estimates of model 3a (Supplementary Fig. 6). Numbers in brackets indicate populations used: West (Rhine: 1F54, Rhone: 2A505) and East European (3: PL51), Lake Constance lake (411, 570) and stream populations (6:52, 7:19, 8:GRA, 9:BOH). Source data are provided as a Source Data file.

Note: upper Rhone) were best supported for the lake and both stream populations (Fig. 3c, Table 1). Strikingly, estimates suggest that the lake population received lower contributions from West Europe (Rhine: 6.7%, upper Rhone: 0.7%) than the two stream populations S2 (Rhine: 16.6%, upper Rhone: 0.1%) and NID (Rhine: 27.2%, upper Rhone: 0.2%), in line with D-statistic and hybrid index analyses.

The mode of ecotype divergence ‘hybrid origin’ fits the observed data considerably better than models of primary divergence in situ, ecological vicariance or of secondary contact (Fig. 3d, Table 1, Supplementary Fig. 8). This is true for all lake-stream comparisons tested with two independent and partially overlapping RAD sequencing datasets: S2 vs. L2 and NID vs. L2, stream comparisons tested with two independent and partially overlapping RAD sequencing datasets: S2 vs. L2 and NID vs. L2, again with the hybrid index results.

In summary, demographic modeling suggests that West and East European lineages diverged ~4000 generations ago, translating to ~8000 years assuming 2 years average age of reproduction as generation time. After several thousand generations of isolation, these West and East European lineages met in the Lake Constance system, perhaps at lake-stream boundaries North and West of the lake. Stream stickleback there are of hybrid origin, having received most of their genome from a West or East European lineage depending on the population. Introgression across the secondary contact zone contributed some West European alleles to the lake ecotype and also to stream ecotypes in South-Eastern tributaries.
We seek the answer to this question by identifying the origin of alleles in genomic islands of differentiation between lake and stream ecotypes breeding in sympathy. In a previous study, we identified 19 genomic islands with unexpectedly high differentiation between lake and stream ecotypes and parallel allele frequency changes in two South-Eastern tributaries to Lake Constance, one tributary with sympatric (S1) and one with parapatric reproduction (S2). These genomic islands are thus candidate loci for reproductive isolation or ecological adaptation to either the lake or the stream habitat due to their persistence in sympathy, parallel allele frequency shifts, habitat associations and their enrichment with QTL controlling adaptive traits divergent between ecotypes. Before identifying the origin of alleles in genomic islands, we aim to confirm that these genomic islands represent regions that resist gene flow between ecotypes due to either divergent selection or reproductive isolation, rather than representing regions that diverged due to background selection.

A rich literature has shown that heterogeneous differentiation ($F_{ST}$) across the genome will arise as by-product of background selection in the absence of gene flow. Background selection removes proportionately more linked variation in low recombination regions, leading to negative correlations of differentiation ($F_{ST}$) with recombination rate, absolute divergence ($d_{XY}$) and diversity levels ($\pi$), and correlated genome-wide differentiation across populations and taxa. Allopatric stickleback populations from different European watersheds lack gene flow, but we find no correlation of $F_{ST}$ with recombination rate or absolute divergence, even though differentiation landscapes are weakly correlated and $F_{ST}$ and diversity show negative correlations in some populations (Supplementary Fig. 9). The lack of association with recombination rate suggest that background selection might not be a major driver of genome-wide differentiation ($F_{ST}$) between allopatric European stickleback populations, even in the absence of gene flow.

In contrast to those, Lake Constance ecotypes breed either in sympathy or parapatry with ample opportunity for gene flow, or in alloparity between different streams with limited gene flow between them. We find negative correlations between $F_{ST}$ and recombination rate among symp- and parapatric lake and stream ecotypes, but not between allopatric stream populations around Lake Constance (Supplementary Fig. 9). However, neither differentiation landscapes between lake-stream ecotypes or between allopatric streams are correlated with diversity or absolute divergence, nor are they with allopatric differentiation landscapes from outside Lake Constance, both genome-wide (Supplementary Fig. 9) and in genomic islands. These combined patterns are best explained by a scenario of divergent selection where lake and stream ecotype differences persist against gene flow aided by low recombination, rather than by background selection.

If admixture variation facilitated ecotype divergence in South-Eastern tributaries of Lake Constance, we expect differential sorting of West- and East-derived alleles in genomic islands of differentiation between lake and stream ecotypes. We thus compared genome-wide differentiation and differentiation in genomic islands between Constance lake/stream stickleback and West European populations (Fig. 5b, c, Supplementary Figs. 10, 11). Genome-wide, both ecotypes from South-Eastern Lake Constance differ strongly from West European populations (weighted mean $F_{ST}$ (L1 vs. AGS) = 0.50, $F_{ST}$ (S1 vs. AGS) = 0.46, $F_{ST}$ (S2 vs. AGS) = 0.48, Fig. 5b) and are more similar to East European populations ($F_{ST}$ (L1 vs. PLS) = 0.31, $F_{ST}$ (L1 vs. pLS) = 0.30, $F_{ST}$ (S2 vs. PLS) = 0.33), consistent with a largely East European genomic background of both ecotypes. In genomic islands of ecotype differentiation, however, stream stickleback are
Here we have demonstrated that secondary contact between divergent lineages and the re-arrangement of introgressed alleles into ecotypes underlie recent ecological speciation across lake-stream habitat boundaries in Lake Constance. Our analysis reveals an unexpected outcome of secondary contact between old allopatric lineages: rapid in situ ecological speciation outside the secondary contact zone fueled by introgression of admixture variation beyond the hybrid zone. It also explains and reconciles contrasting conclusions of previous studies regarding the origin, age and mode of ecotype divergence in Lake Constance. Finally, our results raise interesting questions about the evolutionary potential arising from recombining old alleles into new combinations and how distinctive the new combinations might be from parental combinations.

Analyzing the incipient radiation of Lake Constance stickleback in a European-wide biogeographic and phylogeographic context revealed contributions of at least two old central European lineages of freshwater stickleback that had evolved in isolation for several thousand generations before coming into secondary contact. Secondary contact has not been invoked previously in this system and explains the old, early Holocene (post-glacial) divergence times that had been estimated between Lake Constance ecotypes by applying a primary divergence model representing an ecological vicariance scenario. Our analyses of mitochondrial, microsatellite, and genome-wide SNP data in a larger phylogeographic context using an array of different methods such as phylogenetic reconstruction, demographic modeling, D-statistics, hybrid index, and cluster analysis clearly reject an origin from a single lineage and identify more than one source of origin. The colonization of the Lake Constance catchment by multiple lineages is consistent with historical accounts reporting introductions of stickleback into streams of North-Western Lake Constance between 1920 and 1940, in which the
Fig. 6 New combinations of old West and East European alleles made lake and stream ecotypes. Stream stickleback from South-Eastern tributaries of Lake Constance represent new combinations of West- and East-derived alleles and not a simple reassembly of a West European stickleback genotype. (a, b) Genomic differentiation between lake and stream stickleback from South-Eastern tributaries of Lake Constance, including stream 1 where ecotypes breed in sympathy (S1). Genomic islands of elevated differentiation are highlighted in light gray backgrounds, with SNPs shown in black, blue or red. Blue SNPs indicate lake-stream differentiated SNPs where the lake ecotype carries the East European allele and the stream ecotype the West European allele. Red SNPs indicate the opposite pattern—where the stream ecotype carries an East European allele and the lake ecotype a West European allele. Ancestry assignment of alleles is based on strong allele frequency differences between West (AGS1 + FRS4) and East (PLS1) European populations (panels c, d). (a) Genome-wide differentiation. (b) Magnification of chromosome VII. (c, d) Most lake-stream differentiated SNPs show the blue pattern, but red pattern SNPs make a significant proportion. Source data are provided as a Source Data file.

introduction of both fully plated and low-plated individuals was documented17. This is additional to historical accounts of introductions from unknown sources into streams South of Lake Constance in the late 19th century16. Consistent with an earlier mitochondrial phylogeography,22 and the biogeographic context of a highly plated freshwater phenotype30,44, our analyses based on genome-wide SNPs identified East Europe as the major source of origin. The same lineage is also found in the nearby upper Danube drainage12 and it is one of the central European lineages26–28 as opposed to the highly divergent Black Sea or Mediterranean lineages. Such a central European rather than Black Sea origin is also in agreement with the historical ichthyologic record, which documents the absence of stickleback from the upper Danube until the early 20th century15,18 and thus rather a more recent colonization of the upper Danube and the Lake Constance basin from Northern catchments aided by human introductions instead of a natural colonization from the Black Sea.

Our demographic models were not able to fully resolve the discrepancy between historical absence of stickleback in Lake Constance until 1860–7016,17 and ecotype divergence time estimates hundreds of generations ago. Our parameter estimates are still associated with high uncertainty, as reflected by their wide confidence intervals in each model and variation for the same parameters between different datasets (Fig. 4). The estimated divergence times, times of introgression (which are notably more recent for two streams) and amounts of introgression however may lead to multiple plausible estimates in parameter space. Selection, in particular on old variation as we document here, may have further biased these estimates toward older times, despite our attempts to exclude sites under selection by considering only high recombination region SNPs for the SFS computation. The rather sparse SbfI-RAD sequencing data on one hand and the lack of sister lineage data for denser Nsil-RAD sequencing data on the other hand may have limited our power to better estimate these parameters. Further model optimization incorporating additional variation in population sizes, e.g., bottlenecks and expansions, in combination with higher resolution SNP data may be able to generate better estimates for the time of secondary contact and ecotype divergence.

Importantly, secondary contact brought two lineages together that not only have evolved for many thousand generations in geographical isolation, but that are also phenotypically very different. They differ for example in their lateral plate phenotype, a nearly Mendelian trait encoded predominantly by the Eda gene on chromosome IV often under divergent selection between freshwater and marine environments43. In this case, one of the colonizing lineages is a freshwater lineage that has retained the otherwise marine, fully plated phenotype across its entire distribution range30. Fully plated East European freshwater stickleback may have been preadapted to environments with vertebrate predation regimes44 such as large lakes like Lake Constance harboring a rich community of piscivorous fish45. Fully plated stickleback have an increased probability, relative to low-plated individuals, to escape and survive after capture by toothed predators such as piscivorous fish46. Plate morph is also known to be correlated with schooling behavior in stickleback, such that carriers of the fully plated allele show a higher tendency for schooling47, which is common among marine stickleback and may be advantageous in the open waters of large lakes. Likewise, low-plated West European stickleback may have been relatively better pre-adapted to stream environments given the wide distribution of this lineage and its key phenotypes in rivers and streams of West Europe.

Introgression of West European alleles into streams in South-Eastern tributaries of Lake Constance may hence have been adaptive, as suggested by introgression into genomic regions that are enriched for QTL of lake-stream divergent traits such as jaw and dorsal spine length, lateral plate size and male coloration3. Whether lateral plate number and respective variation at the Eda locus is under selection in stream environments is unclear and
con founded by the origin of different stream populations: variation in genomic contributions from West East European lineages to stream ecotypes are correlated with lateral plate phenotypes (Fig. 1e), resulting in stream stickleback North and West of the lake being low plated and stream stickleback South-East of the lake being fully plated (Fig. 1e). Even though low-plated stickleback occur also in the South-East of Lake Constance, lake and stream ecotypes there are not divergent in this trait and stream ecotypes reduced lateral plate cover by reducing plate size instead of plate number3. A better understanding of ancestral phenotypes, of divergent selection between the lake and the stream habitat, the genomic architecture of traits and a higher genomic resolution will allow to identify the exact loci contributing to adaptation and reproductive isolation and to trace back their West, East European or recombined ancestry.

Secondary contact between old lineages can result in a complex mosaic of evolutionary outcomes. In classical work on contact zones between the West and East European species of firebelly toads (Bombina bombina, B. variegata), outcomes varied from persistence with leaky reproductive isolation in steep tension zones48, through classical `mosaic hybrid zones' patterns where each species occurs in habitat patches rich in the habitats they are best adapted to39, to situations resembling a hybrid swarm in which allelic combinations are sorted between parental species habitats on very small spatial scale, reminiscent of ecological speciation from a hybrid population50. In Lake Constance stickleback, we document a similar continuum of outcomes where secondary contact and environmental adaptation interact in diverse ways. Outcomes range from partial collapse at the zone of contact of the old lineages at habitat boundaries (as between the GRA, NID, BOH stream population and the lake ecotype) to the contact of the old lineages at habitat boundaries (as between the secondary contact and environmental adaptation interact in

Admixture has also facilitated the repeated emergence of blue and red Pandamilla cichlid species in Lake Victoria, East Africa45,55. Introg ression between divergent lineages and the resulting admixture variation may thus more generally be an important source of heritable variation for the rapid evolution of reproductively isolated species and for adaptive radiation23,25,36,37.

Methods

Sample collection. We used predominantly previously collected threepine stickleback populations listed in Supplementary Table 1 and collected three additional populations for this study (VAL, CHA, FRSl1). Hand nets or electrofishing were used to capture stickleback at these sites, in accordance with scientific fisheries permits issued to members of the departmental federations for fisheries who executed the collection (department 84, Vaucluse; 89, Yonne, 03, Allier). Fish were euthanized in the field with an overdose of clove oil or MS-222 in accordance with the respective fisheries regulations.

Mitochondrial DNA. We obtained partial mitochondrial control region sequences for two populations. We extracted DNA from fin tissue using a Qiagen blood and tissue extraction kit, amplified the partial control region fragment using previously published primers (forward: 5′-CCCTATGCTCTAAATCGAGT-3′, reverse: 5′-CCGTAAGCCTAGAAAAAGA-3′39 and sequenced the fragment on an ABI 3130XL DNA Analyzer (Applied Biosystems)22. Accessions are given in Supplementary Table 1. We combined these sequences with previously published mitochondrial sequences from the Lake Constance region12,21,22 and from populations across Europe22,36,39, resulting in a combined dataset of 254 individuals from the Lake Constance catchment, each sequenced for 253 overlapping base pairs of the mitochondrial control region (Supplementary Table 1). Sequences were aligned manually in Bioedit v7.0.5.348, collapsed into identical haplotypes in MEGA X 10.0.560 and matched to European reference haplotypes Eu27, Eu36, Eu9, At1, Bs1, Sor1, Ner1, Sk131, CH0132, and a new haplotype So17 identified in population DKM3. For these reference haplotypes, we concatenated partial mitochondrial control region and cytomeg b alignments of a total 1402 base pairs length, identified HKY + G + I as the best-fitting DNA substitution model based on BIC in MEGA and reconstructed the maximum-likelihood phylogeny under this model in MEGA, with support assessed from 1000 bootstrap replicates. We used the R-package ape 5.131 to visualize the phylogeny, using the following color code: yellow = So17, dark blue = Eu9, light blue = Eu36, green = CH01, orange = Eu27, black = At1, 2 = Sor1, 3 = Ner1, 4 = Sk1, 5 = Bs1 (Fig. 1).

Microsatellites. We combined data for four microsatellite markers in common between two previous studies22,23, featuring lake and stream populations from Lake Constance, populations from upper Danube, the upper Rhone catchment (Lake Geneva) and the River Rhine (Supplementary Table 1). We included 321 individuals without missing genotypes and used an admixture model in STRUCTURE 2.3.421 to infer population structuring with 50,000 burn-in steps followed by 300,000 steps in the MCMC chain. We ran STRUCTURE assuming 1–6 genetic clusters (K) with 10 replicates for each K. We identified the best number of genetic clusters supported by the data using the Evanno method53.

Morphological data. We combined lateral plate morph data from lake and stream populations around Lake Constance, from the upper Rhone drainage (Lake Geneva), the Rhine and the upper Danube32,21,22,26, in which individuals were scored as low, partially or fully plated morph, based on the presence of 0–3 lateral plates posterior the pelvic girdle (low plated), more than three lateral plates but with a gap of at least two plates to the caudal peduncle (partially plated) or the presence of a continuous series of lateral plates up to the caudal peduncle (fully plated)54. We embedded this lateral plate morph data in the context of historical phenotype distributions22,39,65.

Restriction-site-associated DNA. We generated standard1,8 restriction-site-associated (RAD) DNA sequence data for East (Vistula) and West (upper Rhone, Rhine) European stickleback populations (Supplementary Table 1) using the SbfI restriction enzyme, by single-end sequencing 100 bp up- and downstream of each restriction site on an Illumina HiSeq 2000. Accessions are given in Supplementary Table 1. We combined this data with previously published SbfI data from Lake Constance for a Western tributary population and the adjacent lake site33 and two South-Eastern tributary populations and their adjacent lake sites5. We additionally included overlapping PsiI-derived RAD sequencing data for further European populations and North American outgroups27. We also reanalyzed a PsiI-derived RAD sequencing dataset14, which, however, could not be merged with the SbfI (+ PsiI) data due to non-overlapping restriction sites. The NsiI dataset includes three tributary populations from North and West of Lake Constance and adjacent lake sites, including two sites overlapping with the SbfI dataset (NID, ROM, see Supplementary Table 1). Both SbfI and NsiI datasets were aligned to an improved version of the threespine stickleback reference genome6,79,86 using bowtie 2 v2.0.0 with default parameters69. With Phix reads available for SbfI libraries only, we used
We jointly called variants and genotypes in each SbfI and NslI datasets using GATK's unified genotyper. We identified a minor/major genotyping SNP and inserted number read imbalance strongly deviating from 1:1 (Supplementary Fig. 12). We used RAXML 8.0.2673 to reconstruct the phylogenetic relationship containing Lake Constance populations L2, ROM, NID, S1, and all combined, 8173, and 8436 SNPs across 25, 24, 25, and 47 individuals for the SbfI and NslI datasets, respectively.

For phylogenetic analysis, we used subsets of the SbfI dataset with four to five individuals each for Lake Constance populations (ROM: 4, others: 5), three individuals each for European lineages and two each for North American outgroups (Fig. 1c), by selecting individuals with the lowest proportion of missing genotypes. In these phylogeny subsets, we removed monomorphic sites, sites with >25% missing data and sites with less than one homozygote for each allele and genotypes. In these phylogeny subsets, we removed monomorphic sites, sites with frequency >5%, <5% missing genotypes and no linkage disequilibrium exceeding the physical map for each chromosome with a spline parameter of 0.7 and calculating first derivatives (= recombination rates) for positions of interest. Next, we randomly subsampled a fixed number of genotypes from each population (L2: 6, S1: 3, ROM: S1, 4, FR4: 6, ROM: 7, BOH: 5: the same loci as shown in the Hardy–Weinberg equilibrium using vcftools and discarded sites with evidence of heterozygote excess plus minus a buffer of 100 bp left and right of each site from the dataset in order to eliminate putative duplicated regions in the genome and remove the sum of all sites. Then, we computed the FST value and recombined the results of the five groups to avoid biases due to linked selection. We estimated local recombination rates on the FTC cross recombination map68 by cubic spline smoothing of the genetic information on the X chromosome with a spline parameter of 0.7 and calculating first derivatives (= recombination rates) for positions of interest. Next, we randomly subsampled a fixed number of genotypes from each population (L2: 6, S1: 3, ROM: 4, AGS1: 5, FR4: 6, ROM: 7, BOH: 5: the same loci as shown in the Hardy–Weinberg equilibrium using vcftools and discarded sites with evidence of heterozygote excess plus minus a buffer of 100 bp left and right of each site from the dataset in order to eliminate putative duplicated regions in the genome and remove the sum of all sites. Then, we computed the FST value and recombined the results of the five groups to avoid biases due to linked selection. We estimated local recombination rates on the FTC cross recombination map68 by cubic spline smoothing of the genetic. In these phylogeny subsets, we removed monomorphic sites, sites with frequency >5%, <5% missing genotypes and no linkage disequilibrium exceeding the physical map for each chromosome with a spline parameter of 0.7 and calculating first derivatives (= recombination rates) for positions of interest. Next, we randomly subsampled a fixed number of genotypes from each population (L2: 6, S1: 3, ROM: 4, AGS1: 5, FR4: 6, ROM: 7, BOH: 5: the same loci as shown in the Hardy–Weinberg equilibrium using vcftools and discarded sites with evidence of heterozygote excess plus minus a buffer of 100 bp left and right of each site from the dataset in order to eliminate putative duplicated regions in the genome and remove the sum of all sites. Then, we computed the FST value and recombined the results of the five groups to avoid biases due to linked selection. We estimated local recombination rates on the FTC cross recombination map68 by cubic spline smoothing of the genetic.

We used fastsimcoal2 v2.681 and a hierarchical modeling approach to reconstruct the demographic history of Lake Constance back by comparing the fit of different demographic models to observed SFS and to estimate parameters for the best-fitting models. First, we optimized three population models (Fig. 3a, Supplementary Fig. 7) on the observed, SbfI derived, folded 3D-SFS with the putative sister lineages to Lake Constance stickleback from the Vistula (PLS1), Rhine (BOH), and upper Rhine (ROM), for each model, we computed the FST value and recombined the results of the five groups to avoid biases due to linked selection. We estimated local recombination rates on the FTC cross recombination map68 by cubic spline smoothing of the genetic.}

Demographic modeling. We used fastsimcoal2 v2.681 and a hierarchical modeling approach to reconstruct the demographic history of Lake Constance back by comparing the fit of different demographic models to observed SFS and to estimate parameters for the best-fitting models. First, we optimized three population models (Fig. 3a, Supplementary Fig. 7) on the observed, SbfI derived, folded 3D-SFS with the putative sister lineages to Lake Constance stickleback from the Vistula (PLS1), Rhine (BOH), and upper Rhine (ROM), for each model, we computed the FST value and recombined the results of the five groups to avoid biases due to linked selection. We estimated local recombination rates on the FTC cross recombination map68 by cubic spline smoothing of the genetic. In these phylogeny subsets, we removed monomorphic sites, sites with frequency >5%, <5% missing genotypes and no linkage disequilibrium exceeding the physical map for each chromosome with a spline parameter of 0.7 and calculating first derivatives (= recombination rates) for positions of interest. Next, we randomly subsampled a fixed number of genotypes from each population (L2: 6, S1: 3, ROM: 4, AGS1: 5, FR4: 6, ROM: 7, BOH: 5: the same loci as shown in the Hardy–Weinberg equilibrium using vcftools and discarded sites with evidence of heterozygote excess plus minus a buffer of 100 bp left and right of each site from the dataset in order to eliminate putative duplicated regions in the genome and remove the sum of all sites. Then, we computed the FST value and recombined the results of the five groups to avoid biases due to linked selection. We estimated local recombination rates on the FTC cross recombination map68 by cubic spline smoothing of the genetic.
population NID, we ignored singletons in the likelihood computation, the former to avoid possible remaining PCR artifacts in this population.

Population differentiation. We used a subset of the Sfs dataset to test for introgression of West European lineages into Lake Constance. This dataset contained one lake (L1) and two South-Eastern tributary populations (S2, S1), one East European (PLS1) and two West European populations (AGS1, FR54). We filtered the Sfs-dataset subset for each pairwise comparison to include only bi-allelic SNPs with at least three sequenced genotypes per population using scitoools. We estimated pairwise F-statistics for each SNP by performing locus-by-locus AMOVAs in ARLEQUIN V3.5.23. SNP-level F-statistics were averaged over non-overlapping windows containing at least 2500 sequenced bases without splitting RAD loci or over sliding windows of 1 Mbp size with 200 kbp step size. Non-overlapping windows were on average 344 kbp (73–1564 kbp) wide. We computed an average recombination rate for each window by computing the mean over 10 recombination rate estimates at equally spaced positions across each non-overlapping window. We computed weighted average F-statistics by averaging variance components and calculating the ratio of averages. We used permutation tests to assess whether differentiation in windows overlapping with genomic islands of differentiation persisting among sympatric populations identified in an earlier study differed from the genome-wide distribution of differentiation. For this, we permuted the position of genomic islands 10,000 times across the genome, generated a null distribution for the mean and computed empirical quantiles for the permuted data. For the second set of analyses, we resampled the Sfs- or Nsl-data subsets of population pairs to N genotypes per population corresponding to 75% of a population’s individuals in order to get a dataset without missing data, discarding sites with too few genotypes, with the custom script sampleGenotypesPerPop.py. We summarized the subsampled 1D-CF files for each non-overlapping window with the custom script vc2sfs.py and computed both \( F_{ST} \) and \( D_{XY} \) with custom scripts (wfs_dxy.py, v1.0; wsfs_pi.R, v1.1), respectively. Pearson’s correlation coefficient \( r \) (between statistics was computed in R for windows without missing values, for pairwise statistics as \( F_{ST} \) or \( D_{XY} \)) using statistics from non-overlapping population pairs to avoid autocorrelation. All data analysis was performed on the servers of the Genetic Diversity Centre (GDC), the Ubelex computer cluster, University of Bern and the Euler cluster, ETH Zurich, Switzerland.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All morphological and genetic data used in this study is available from previously published datasets or accessions detailed in Supplementary Table 1. These data are also available from the corresponding author on reasonable request. Sequence data has been deposited on GenBank under accessions MN082769–MN082781 and on Sequence Read Archive under accessions SRR3917386–SRR39317452, SRR3935375–SRR3935380 and SRA-BioProject accession PRJNAS49360. The data source underlying Figs. 1–6, Table 1, Supplementary Figs. 1–11 and Supplementary Table 2 are provided as a Source Data file.

Code availability
All custom scripts mentioned in the Methods section have been deposited on GitHub and are accessible under https://github.com/marqueda

Received: 31 August 2017 Accepted: 27 August 2019
Published online: 18 September 2019

References
1. Powell, T. H. et al. Genetic divergence along the speciation continuum: the transition from host race to species in Rhagoletis (Diptera: Tephritidae). Evolution 67, 2561–2576 (2013).
2. Malinsky, M. et al. Genomic islands of speciation separate cichlid ecomorphs in East African lakes. Science 350, 1493–1498 (2015).
3. Marques, D. A. et al. Genomics of rapid incipient speciation in sympatric threespine stickleback. PLoS Genet. 12, e1005887 (2016).
4. Heier, J. I. et al. Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. Mol. Ecol. 26, 123–141 (2017).
5. Feder, J. L. et al. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. PLoS Natl. Acad. Sci. USA 100, 10314–10319 (2003).
6. Heier, J. I. et al. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8, 14853 (2017).
7. Richards, E. J. & Martin, C. H. Adaptive introgression from distant Caribbean lineages contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. PLoS Genet. 13, e1006919 (2017).
8. Wright, K. M., Lloyd, D., Lowry, D. B., Macnair, M. R. & Willis, J. H. Indirect evolution of hybrid lethality due to linkage with selected locus in Mimulus guttatus. PLoS Biol. 11, e1001497 (2013).
9. Lucek, K., Sivasundar, A. & Seehausen, O. Evidence of adaptive evolutionary divergence during biological invasion. PloS ONE 7, e49377 (2012).
10. Lucek, K., Sivasundar, A. & Seehausen, O. Disentangling the role of phenotypic plasticity and genetic divergence in contemporary ecotype formation during a biological invasion. Evolution 68, 2619–2632 (2014).
11. Lucek, K., Sivasundar, A., Roy, D. & Seehausen, O. Repeated and predictable patterns of ecotypic differentiation during a biological invasion: lake-stream divergence in parapatric Swiss stickleback. J. Evolut. Biol. 26, 2691–2709 (2013).
12. Moser, D., Roesti, M. & Berner, D. Repeated lake-stream divergence in stickleback life history within a Central European lake basin. PloS ONE 7, e50620 (2012).
13. Berner, D., Grandchamp, A. C. & Hendry, A. P. Variable progress toward ecological speciation in paraparik: stickleback across eight lake-stream transects. Evolution 63, 1740–1753 (2009).
14. Roesti, M., Kuenz, B., Moser, D. & Berner, D. The genetics of ecological divergence in threespine stickleback. Nat. Commun. 6, 8767 (2015).
15. Heckel, J. & Kner, R. Die Süsswasserfische der österreichischen Monarchie mit Rücksicht auf die angrenzenden Länder (Wilhelm Engelmann, 1858).
16. Heller, C. Die Fische Tirols und Vorarlbergs. Z. des. Ferdinandeums fär. Tirol. und Vorarl. 5, 295–369 (1870).
17. Mucke, R. Der dreistachlige Stichling (Gasterosteus aculeatus L.) im Bodensee. Schr. des. Ver. für. Gesch. des. Bodenspie und seiner Umge. 90, 249–257 (1972).
18. Ahnl, H. Zum Vorkommen des Dreistachligen Stichlings (Gasterosteus aculeatus, Pisces: Gasterosteidae) im österreichischen. Donauraum Ann. des. Nat. Mus. Wien. 88/89B, 309–314 (1986).
19. Peichl, C. L. & Marques, D. A. The genetic and molecular architecture of phenotypic diversity in sticklebaids. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. 372, 20150486 (2017).
20. Hendry, A. P., Bolnick, D. I., Berner, D. & Peichl, C. L. Along the speciation continuum in sticklebacks. J. Fish. Biol. 75, 2000–2036 (2009).
21. Berner, D., Roesti, M., Hendry, A. P. & Saltburger, W. Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. Mol. Ecol. 19, 4963–4978 (2010).
22. Lucek, K., Roy, D., Beuzart, E., Sivasundar, A. & Seehausen, O. Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in Switzerland. Mol. Ecol. 19, 3995–4011 (2010).
23. Seehausen, O. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207 (2004).
24. Abbott, R. et al. Hybridization and speciation. J. Evolut. Biol. 26, 229–246 (2013).
25. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. Trends Ecol. Evol. 34, 531–544 (2019).
26. Makinen, H. S. & Merila, J. Mitochondrial DNA phylogeny of the three-spined stickleback (Gasterosteus aculeatus) in Europe - evidence for multiple glacial refugia. Mol. Phylogenetics Evol. 46, 167–182 (2008).
27. Fang, B., Merila, J., Ribeiro, F., Alexandre, C. M. & Momigliano, P. Worldwide phylogeny of three-spined sticklebacks. Mol. Phylogenetics Evol. 127, 613–625 (2018).
28. Makinen, H. S., Cano, J. M. & Merila, J. Phylogeny of three-spined sticklebacks (Gasterosteus aculeatus) from western Mediterranean islands. Conserv Genet. 16, 1319–1333 (2015).
29. Meinzinger, J. Evolution of variation and distributional patterns in European populations of 3-spined stickleback, Gasterosteus aculeatus. Evolution 17, 320–332 (1963).
30. Kottelat, M. & Freyhof, J. Handbook of European Freshwater Fishes. (Kottelat, 2007).
31. Poirrier, J., Aeschbacher, S., Thiéry, A. & Excoffier, L. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. eLife 7, e36317 (2018).
33. Burri, R. et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Res.* 25, 1656–1665 (2015).

34. Vijay, N. et al. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Mol. Ecol.* 26, 4284–4295 (2017).

35. Wolf, J. B. & Ellegren, H. Making sense of genomic islands of differentiation in a mosaic hybrid zone between the re-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution* 40, 1114–1159 (1986).

36. Callum, C. J., Nürnberg, B., Barton, N. H. & Szymura, J. M. Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* 52, 227–239 (1998).

37. Vines, T. H. et al. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution* 57, 1876–1888 (2003).

38. Hedrick, P. W. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* 22, 4606–4618 (2013).

39. Lewontin, R. C., Birch, L. C. Hybridization as a source of variation for adaptation to new environments. *Evolution* 20, 315–336 (1966).

40. Bay, R. A. & Ruegg, K. Genomic islands of divergence or opportunities for introgression? *Proceedings Biol. Sci.* 284, 20162414 (2017).

41. Fraisse, C., Roux, C., Welch, J. J. & Bierne, N. Gene-flow in a mosaic hybrid zone: is local introgression adaptive? *Genetics* 197, 939–951 (2014).

42. Meier, J. L., Marques, D. A., Wagner, C. E., Excoffier, L. & Szymura, J. M. Genetic differentiation of freshwater three-spined sticklebacks. *Nature* 455, 461–464 (2008).

43. Wood, J. H. & Camp, J. T. Hybridizing or competing? *Mol. Ecol. Resour.* 12, 1659–1669 (2012).

44. Pakan, S. et al. Adaptive introgression in the re-bellied toad, *Bombina bombina*. *Mol. Ecol. Resour.* 13, 1547–1549 (2013).

45. Feulner, P. G. et al. Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet.* 11, e1004966 (2015).

46. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs for performing population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567 (2010).

47. Bhattacharya, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting Fst: The impact of rare variants. *Genome Res.* 23, 1514–1521 (2013).

Acknowledgements

We thank Henry Piers for facilitation of field work in France, Yann Monnier, Claude Chadaffaux, Michael Leèvire, Jean Louis Clerc for collecting stickleback in France and Rafael Bernas for collecting stickleback in Poland, Joanna Meier, Carmela Doenz and students of the Seehausen lab for assistance in the field. Salome Mwaiko for assistance in the lab, Aria Minder and Stefan Zoller from the Genetic Diversity Center, ETH Zurich/ Eawag for bioinformatics support and all members of the Fish Ecology and Evolution lab for discussion. This research was supported by the Swiss National Science Foundation (SNF) grants PMPFP3_134657 and 3103A3_163383 to O.S. and L.E. V.S. is funded by EU H2020 program (Marie Skłodowska-Curie grant 797729).

Author contributions

O.S., I.E., K.L. and D.AM. conceived the study. K.L., D.AM. and O.S. acquired samples. K.L. and D.A.M. generated and compiled morphological and genetic data. K.L. conducted morphological, microsatellite, and mitochondrial analyses. D.A.M. genomics analyses and demographic modeling, with input from V.S. L.E. and D.A.M. created code. D.A.M. and K.L. created figures. D.A.M. and K.L. wrote the paper, with input from O.S., L.E., and V.S. O.S, L.E., and V.S. acquired funding.
