SPECIAL ISSUE

Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats

Mark Ravinet 1 | Manabu Kume 2 | Asano Ishikawa 3 | Jun Kitano 3

1 School of Life Sciences, University of Nottingham, Nottingham, UK
2 Kyoto University Field Science Education and Research Center, Kyoto, Japan
3 Ecological Genetics Laboratory, National Institute of Genetics, Mishima, Japan

Abstract
With only a few absolute geographic barriers in marine environments, the factors maintaining reproductive isolation among marine organisms remain elusive. However, spatial structuring in breeding habitat can contribute to reproductive isolation. This is particularly important for marine organisms that migrate to use fresh- or brackish water environments to breed. The Japanese Gasterosteus stickleback species, the Pacific Ocean three-spined stickleback (G. aculeatus) and the Japan Sea stickleback (G. nipponicus) overwinter in the sea, but migrate to rivers for spawning. Although they co-occur at several locations across the Japanese islands, they are reproductively isolated. Our previous studies in Bekanbeushi River showed that the Japan Sea stickleback spawns in the estuary, while the Pacific Ocean stickleback mainly spawns further upstream in freshwater. Overall genomic divergence was very high with many interspersed regions of introgression. Here, we investigated genomic divergence and introgression between the sympatric species in the much shorter Tokotan River, where they share spawning sites. The levels of genome-wide divergence were reduced and introgression was increased, suggesting that habitat isolation substantially contributes to a reduction in gene flow. We also found that genomic regions of introgression were largely shared between the two systems. Furthermore, some regions of introgression were located near loci with a heterozygote advantage for juvenile survival. Taken together, introgression may be partially driven by adaptation in this system. Although, the two species remain clearly genetically differentiated. Regions with low recombination rates showed especially low introgression. Speciation reversal is therefore likely prevented by barriers other than habitat isolation.

1 | INTRODUCTION

Differential spatial use of breeding habitats can contribute to reproductive isolation (Coyne & Orr, 2004; Mayr, 1963). Such habitat isolation can reduce the encounter rates of heterospecific males and females, thereby directly reducing hybridization. In speciation studies where the relative contribution of each isolating barrier is calculated, divergence in breeding sites often contributes the most to total reproductive isolation (Kitano et al., 2009; Lackey & Boughman, 2017; Matsubayashi & Katakura, 2009; Ramsey, Bradshaw, & Schemske, 2003). One of the reasons for the largest relative contribution is simply because across the life cycle of an
organism habitat isolation can act before any other isolating barriers, such as assortative mating and post-zygotic isolation (Coyne & Orr, 2004). Differential use of breeding sites can also drive divergent adaptation and secondarily reduce hybridization (Schluter, 2000). For example, once two species adapt to different breeding habitats, immigrants from dissimilar environments will have lower survival rates or reproductive success in foreign sites, thereby reducing the chances of heterospecific mating between native residents and immigrants (Nosil, Vines, & Funk, 2005).

Marine environments present a conundrum in this regard; there are seemingly high levels of dispersal and few geographical barriers to gene flow. The factors driving speciation and maintaining reproductive isolation in marine organisms remain elusive (Knowlton, 1993; Palumbi, 1994; Potkamp & Fransen, 2019; Puebla, 2009). Some marine organisms, such as anadromous fishes, use brackish or freshwater environments for breeding and closely related species may diverge in breeding habitat use (Hendry & Stearns, 2004; Kitano et al., 2009; Kume, Kitano, Mori, & Shibuya, 2010). For migratory species with divergent spawning preferences, variation in the spatial distribution and segregation of breeding sites within and among river systems may modulate the extent to which secondary contact occurs. What are the potential consequences of the breakdown of habitat isolation in this scenario? One possible outcome would be speciation reversal, where two species collapse into a single species by extensive hybridization (Elmer, 2019; Seehausen, Takimoto, Roy, & Jokela, 2008; Taylor et al., 2006). A second possibility is that a level of reproductive isolation is maintained in the face of ongoing gene flow. In this case, introgression is likely to be suppressed near loci involved in reproductive isolation (Elmer, 2019; Nosil, 2012; Ravinet et al., 2017), whereas alleles that are beneficial to both populations (Anderson et al., 2009; Pardo-Diaz et al., 2012; Whitney, Randell, & Rieseberg, 2006) or those with heterozygote advantage are likely to introgress elsewhere in the genome (Kim, Huber, & Lohmueller, 2018; Nadachowska-Brzyska, Zieliński, Radwan, & Babik, 2012; Wolfe et al., 2019). Finally, if reproductive isolation is strong enough, breakdown of habitat isolation would neither influence the hybridization rates nor erode overall genomic divergence between species. Such strong reproductive isolation may be achieved by other multiple isolating barriers and/or irreversible isolating barriers, such as complete hybrid sterility and inviability (Nosil, Harmon, & Seehausen, 2009).

The Japanese anadromous stickleback species pair composed of the Japan Sea stickleback (Gasterosteus nipponicus) and Pacific Ocean populations of the three-spined stickleback (G. aculeatus; hereafter referred to as Pacific Ocean stickleback) are an excellent system to address this question. Our previous demographic analyses using whole genome sequence data showed that the Japan Sea and the Pacific Ocean stickleback species diverged about 0.68–1 million years ago (Ravinet et al., 2018), most likely triggered by sea level fluctuation during the Pleistocene (Mix et al., 1995). Furthermore, gene flow between the two species has probably been maintained for most of their evolutionary history (Ravinet et al., 2018). These two anadromous stickleback species share spawning sites in rivers at several locations across the Japanese islands (Ishikawa et al., 2019; Kume et al., 2010; Ravinet, Takeuchi, Kume, Mori, & Kitano, 2014; Yamada, Higuchi, & Goto, 2007). In Bekanbeushi River on Hokkaido Island, where our previous work focused, the most downstream and upstream of stickleback spawning sites are over 5 km apart (Figure 1), contributing a substantial level of habitat isolation (Kitano et al., 2009; Kume, Kitamura, Takahashi, & Goto, 2005; Kume et al., 2010). Isolation of spawning sites contributes to more than 90% of total reproductive isolation between these two species in the Bekanbeushi River (Kitano et al., 2009; Kume et al., 2010). Both species overwinter in the sea but migrate to the river to reproduce. The Japan Sea stickleback spawns in the brackish estuary, while the Pacific Ocean stickleback spawns further upstream in a more freshwater environment (Kitano et al., 2009; Kume et al., 2010). In

![Map of collection sites.](image)

**FIGURE 1** Map of collection sites. (a) Map of the Japanese archipelago showing the location of Akkeshi Town. (b) Map of the Bekanbeushi River. Magenta indicates the breeding habitats of the Pacific Ocean stickleback, whereas green indicates those of the Japan Sea stickleback. Both sticklebacks are distributed in the midstream (Mid1 and Mid2). (c) Map of Tokotan River. Both sticklebacks are distributed from the river mouth to 200 m upstream in Tokotan River, Hokkaido, Japan. Above this point, no sticklebacks are thought to occur and we were unable to collect any.
the midstream, there is a hybrid zone, where behavioural isolation and hybrid male sterility likely play an important role in reproductive isolation (Kitano, Mori, & Peichel, 2007; Kitano et al., 2009). Additionally, because the Japan Sea and Pacific Ocean sticklebacks have morphological and physiological characteristics adapted for marine and freshwater habitats, respectively (Ishikawa et al., 2019; Ishikawa, Kusakabe, Kume, & Kitano, 2016; Kume et al., 2010; Ravinet et al., 2014), there is likely selection against immigrants, which can further contribute to reproductive isolation (Nosil et al., 2005).

Our previous genomic analysis of a sympatric pair from the Bekanbeushi River system revealed high levels of overall genomic divergence interspersed with multiple, small regions of introgression (Ravinet et al., 2018). Introgression rates were low in regions with low recombination (Ravinet et al., 2018), consistent with previous theoretical (Butlin, 2005; Nachman & Payseur, 2012; Noor & Bennett, 2009) and empirical studies on the association between rates of recombination and introgression (Brandvain, Kenney, Flagel, Coop, & Sweigart, 2014; Janoušek, Munclinger, Wang, Teeter, & Tucker, 2015; Juric, Aeschbacher, & Coop, 2016; Martin, Davey, Salazar, & Jiggins, 2019; Samuk et al., 2017; Schumer et al., 2018). In low recombination regions, negative selection against maladaptive alleles has a greater power to reduce the rate of effective migration at linked loci (Nachman & Payseur, 2012). Alternatively, reduced recombination may be a consequence of selection for maintaining a combination of locally adaptive alleles (Kumar, 1956; Kirkpatrick & Barton, 2006; Nei, 1967; Yeaman, 2013). Additionally, the Japan Sea stickleback has a neo-sex chromosome resulting from chromosomal fusion between the Y chromosome and an autosome (Chromosome 9), where recombination between the neo-X and neo-Y is suppressed near the fusion point (Kitano et al., 2009; Yoshida et al., 2014).

Introgression rates have been previously shown to be lower on the neo-X chromosomes than in autosomes (Ravinet et al., 2018).

In contrast to Bekanbeushi River, in the nearby Tokotan River (Figure 1) both stickleback species use only approximately 200 m of the river for spawning, indicating no apparent isolation of breeding habitats. Studying the Tokotan system alongside our previous analysis therefore makes it possible to compare instances of the Japanese stickleback species pairs with and without spatial isolation of breeding sites. A comparative approach will help develop a better understanding of how differences in spawning sites contribute to reproductive isolation and the maintenance of genomic divergence between anadromous fishes. Furthermore, comparative genomic analysis can also provide insights into the adaptive significance of introgression. Using replicate contact zones, we can test whether the similar regions and/or regions with similar gene functions show more introgression compared to other genomic regions. Additionally, by combining quantitative trait loci (QTL) analysis with population genomics, we can explicitly test whether regions that show introgression harbour QTL or fitness-related traits.

Here, our first aim was to show how the presence and absence of breeding habitat isolation influence the genome-wide patterns of divergence and introgression. To this end, we conducted whole genome resequencing of sticklebacks caught in Tokotan River and compared the data with our previously published whole genome sequences of sticklebacks collected from Bekanbeushi River (Ravinet et al., 2018; Yoshida et al., 2014). We specifically compared the levels of overall interspecies genomic divergence between the two river systems. Our second aim was to test whether the introgression identified may have any functional significance. To this end, we examined whether introgression sites are shared by two systems and overlap with any fitness-related QTLs. Our previous QTL mapping of juvenile survival using an interspecies cross derived from the Bekanbeushi River system found several loci showing heterozygote advantage (Ishikawa et al., 2019). We tested whether these QTL overlapped with genomic signatures of introgression between the two species. Finally, we asked whether regions with lower recombination rates show less introgression.

2 | MATERIALS AND METHODS

2.1 | Fish sampling and genome sequencing

Breeding sticklebacks were distributed from the river mouth (42.9974 N, 144.8673 E) through the 200 m upstream in Tokotan River, Hokkaido, Japan (Figure 1). Above this point, there was a steep slope and no sticklebacks were observed while beach seineing at 6 sampling points within the upstream pond in 2007, although we note there is a report that beach seineing captured one stickleback in 2006 (Akkeshi Town, 2006). Using hand nets, 30 individuals were randomly collected in June 2006, at the peak of the breeding season (Kume et al., 2005). These fish turned out to be nine females and two males of the Japan Sea stickleback and eight females and 11 males of the Pacific Ocean stickleback, based on morphological characteristics (Higuchi, Sakai, & Goto, 2014; Kitano et al., 2007). Among them, eight females and two males of the Japan Sea stickleback and eight females and two males of the three-spined stickleback were used for whole genome resequencing. Genomic DNA was isolated from pectoral fins using DNeasy Blood & Tissue Kits (Qiagen). NEB Next Ultra II DNA Library Preparation Kit was used for making genome sequence libraries with each fish having a unique index barcode (NEB, Ipswich). Libraries were run on one lane of Illumina HiSeqX at the 150 bp × 2 paired-end mode (Illumina) at Macrogen Japan. Sequence reads are available from DDBJ (DRA009272).

Obtained reads were trimmed and mapped to the stickleback reference sequence BROAD51 (Jones et al., 2012), using CLC Genomics Workbench 10.1.1 (Qiagen) as described previously with the same parameters (Ishikawa et al., 2019). Average sequencing coverage (±SD) per fish was × 14.30 (±2.01). The exported bam files were used for subsequent analysis.

2.2 | Variant calling, filtering, liftover and phasing

Prior to variant calling, bam files were sorted, indexed and realigned around indels using GATK 3.7 (McKenna et al., 2010). Samples from
our previous study on the Bekanbeushi system were sequenced to 70–80× (Ravinet et al., 2018), compared to the much lower coverage used for samples from Tokotan (see above). In order to address this, we used samtools 0.1.19 (Li et al., 2009) to downsample all bams to adjust the mean depth to 12× to ensure similar levels of coverage among individuals and to prevent variant calling biases. Following downsampling, we called variants at all positions across the genome using the bcftools 1.9 multiallelic caller (Danecek & McCarthy, 2017).

Following variant calling, we filtered our data set using vcftools 0.1.17 (Danecek et al., 2011). We removed all sites with >20% missing data, a site Phred quality score <30, a minimum genotype depth of 5×, a maximum genotype depth of 25×, a minimum mean site depth of 5× and a maximum mean site depth of 25×. Further filters, such as minor allele frequency or minimum number of alleles were applied downstream when required for specific analyses and are detailed in the relevant sections.

Prior to analysis, we used the LiftOverVcf module of Picard 2.9 to lift our variant calls to the linkage map-corrected physical positions of the stickleback genome (Roesti, Moser, & Berner, 2013). We also phased a subset of our data containing only biallelic SNPs using Shapeit2 (O’Connell et al., 2014). Phasing was informed by the recombination map estimated by Roesti et al. (2013) and was conducted using default parameters.

We conducted our resequencing data from 20 individuals from Tokotan (PO-T) with previously published sequences of 20 individuals from Bekanbeushi (PO-B) (Yoshida et al., 2014), six individuals from the Atlantic (AT) (Feulner et al., 2015), a single individual from Little Campbell River, North America (PO-NA) (Ishikawa et al., 2017) and a nine-spined stickleback (Pungitius pungitius) as an outgroup (White, Kitano, & Peichel, 2015), resulting in 48 individuals in total. Using GATK 3.7 Haplotype Caller, we called 389,448,440 positions across the genome. After filtering, we retained 213,462,870 sites (i.e., variant and invariant). Applying further filters to keep only biallelic sites reduced the data set to 36,182,627 positions.

Only autosomes were used for most of the subsequent analyses. We only included sex chromosomes (neo-sex chromosome [Chromosome IX] and Ancestral-sex chromosome [Chromosome XIX]) for the analysis of introgression (fXY and TWISS, see below); when included, we used only females (i.e., X chromosomes) to exclude the confounding effects of intraspecific genetic differentiation between X and Y chromosomes.

2.3 Population structure and admixture

To identify the major axes of differentiation among our sequenced individuals, we used principal components analysis (PCA) implemented in plink 1.9 (Purcell et al., 2007). PCA is a parameter-free approach that has no a priori assumptions of population structure. Prior to population structure analysis, we only allowed sites with a <5% missing data. Since PCA also requires independence among sites, we also performed linkage pruning, removing all sites with an $r^2 > .1$. This resulted in 282,612 loci. We did not include P. pungitius for PCA.

In addition to PCA, we used ADMIXTURE, a model-based approach to assessing population structure in a Maximum Likelihood framework (Alexander, Novembre, & Lange, 2009). In short, ADMIXTURE attempts to fit a model of K population clusters to the data, fitting individuals in a way that minimizes deviation from Hardy–Weinberg equilibrium. Like PCA, ADMIXTURE also requires independence among sites, so we reused the linkage-pruned data set for this analysis. We ran ADMIXTURE for $K = 1–10$ and also performed cross-validation error estimation in order to assess the most suitable value of K (Alexander & Lange, 2011). We did not include P. pungitius in the ADMIXTURE analysis.

We additionally investigated individual variation in admixture and ancestry by estimating hybrid index using the R package introgress (Gompert & Alex Buerkle, 2010). For this, we used Japan Sea and Pacific Ocean individuals from the Bekanbeushi River as ‘pure’ parental populations. Our rationale for this was that these individuals were previously assigned as pure ancestry based on microsatellite markers, prior to genome resequencing (Kitano et al., 2009). We first identified autosomal sites with an absolute allele frequency difference of greater than 0.9 between the two populations and then performed linkage pruning (as above) to reduce our data set to 5,044 ancestry informative loci. We then used these sites to calculate hybrid index and interspecific heterozygosity among Tokotan individuals. We additionally used adegenet (Jombart & Ahmed, 2011) to simulate $F_1$, $F_2$ and backcross hybrids in order to assess whether any Tokotan individuals fell close to these classes on the joint distribution of interspecific heterozygosity and hybrid index.

2.4 Genetic differentiation, divergence and introgression

We estimated genetic differentiation ($F_{ST}$) and divergence ($d_{XY}$) using the popgenWindows.py script (https://github.com/simonhmartin//genomics_general), $F_{ST}$ calculated with this script is also known as $\gamma_{ST}$ (Hahn, 2019). Both statistics were estimated in 10 and 100 kb sliding windows with a 2.5 and 25 kb step respectively. We note here that sliding window estimates were generated for visualization purposes only; we removed overlapping windows to prevent pseudoreplication in downstream statistical analyses. Mean-pairwise genome-wide estimates of these statistics were thus derived from the 10 kb nonoverlapping window data using the dplyr package in R (Wickham, François, Henry, & Müller, 2018). We did not include Little Campbell River fish and P. pungitius for analysis of $F_{ST}$ and $d_{XY}$.

We further examined admixture and introgression using Patterson’s D statistic implemented in ADMIXTOOLS (Patterson et al., 2012) and the R package admixr (Petr, Vernot, & Kelso, 2019). To calculate D statistics and an excess of sharing of discordant site patterns between taxa, we included the P. pungitius as an outgroup in the test topology, (([P1, P2], P3), O) and only focused on autosomal sites. We first tested for evidence of introgression between
any combination of Pacific Ocean or Japan Sea population (P1 = AT, P2 = PO-T, PO-B or PO-NA, P3 = JS-T or JS-B). We next focused the test topology on the Pacific Ocean populations from Japan (P1 = PO-B, P2 = PO-T) and asked which of the two populations shares an excess of sites with either Japan Sea population (P3 = JS-T or JS-B). We then repeated this test but with P3 set to either Pacific Ocean from North America or G. aculeatus from the Atlantic Ocean. For the last test, we set P1 and P2 to JS-B and JS-T respectively in order to investigate site sharing with all possible G. aculeatus populations separately set as P3. By using multiple configurations of the D statistic with different populations with known divergence times (Ravinet et al., 2018), our rationale was that we would be able to broadly infer when introgression likely occurred and also gain some insight into the direction of gene flow. The significance of all D statistics was calculated using block jack-knifing to determine a Z-score distribution, where an absolute value of Z > 3 is equivalent to an approximate p-value of .001.

2.5 | Characterization of introgressed regions

To investigate the physical location of introgression in the genome, we used $f_d$, a derivative of ABBA-BABA statistics that is reliable and robust to sliding window estimates (Martin, Davey, & Jiggins, 2014). $f_d$ was calculated in 10 and 100 kb sliding windows with a 2.5 and 25 kb step respectively using ABBABABAwindows.py (https://github.com/simonhmartin/genomics_general). As for estimates of genetic differentiation and divergence, sliding windows were generated for visualization purposes only; we removed overlapping windows to prevent pseudoreplication in downstream statistical analyses. In order to calculate $f_d$, the test topology ($[[P1, P2], P3], O$) must be used, where P1 and P2 are sister taxa and O is an outgroup—thus, $f_d$ can be interpreted as the proportion of sites in a window shared between P2 and P3 as a result of introgression. We estimated $f_d$ for two different topologies where P1 was Atlantic G. aculeatus and nine-spined stickleback (P. pungitius) was the outgroup. Our first test was for introgression between G. aculeatus (PO-B) and G. nipponicus (JS-B) in Bekanbeushi (i.e., P2 = PO-B, P3 = JS-B) and our second test was for interspecific introgression between G. aculeatus (PO-T) and G. nipponicus (JS-T) in Tokotan (i.e., P2 = PO-T, P3 = JS-T). To define peaks of $f_d$ in the genome, we took the 99th percentile and identified all windows with a value of $f_d$ greater than or equal to this. We then clustered neighbouring windows above this threshold together, stopping when there were no neighbouring windows within 30 kb. This enabled us to count the number of unique peaks and to measure the average size of $f_d$ peaks observed.

One caveat to both $f_d$ and ABBA-BABA statistics is that they assume monophyletic relationships within the groups at the tips of the test phylogeny. An alternative method that attempts to account for this is topology weighting—that is, the weighting of different topologies in a sliding window across the genome. Using our phased, biallelic site data set, we first used phym (Guindon et al., 2010) to estimate phylogenies in 50 SNP windows across the genome for G. aculeatus and G. nipponicus of Bekanbeushi and Tokotan (i.e., JS-B, PO-B, JS-T and PO-T). With four groups, there are three possible unrooted phylogenies—a species topology where Japan Sea and Pacific Ocean populations cluster together with the same species, a geographical topology where species clustered based on their geographical proximity (i.e., Japan Sea and Pacific Ocean in Tokotan are more closely related) and an interspecific topology where Japan Sea and Pacific Ocean from different river catchments group more closely with one another (Figure 4a). We then used TWISSST to weight the contribution of each of these topologies to all the possible subtrees at each window across the genome (Martin & Van Belleghem, 2017).

In order to examine how introgression varied with recombination rate, we used a previously published three-spined stickleback linkage map (Roesti et al., 2013). Following Roesti et al. (2013), we fit a loess model with a span of 0.25 in order to interpolate recombination rate for 10 kb windows across the genome. We then compared mean recombination rate for $f_d$ peaks versus the genome background using a one-tailed permutation test, resampling 10,000 times.

To investigate possible adaptive significance of introgression, we tested whether the regions showing signs of introgression were located near previously identified QTLs for juvenile survival rates. In our previous QTL mapping experiment, we made an $F_2$ intercross using $F_1$ hybrids between a Pacific Ocean female and a Japan Sea male from the Bekanbeushi system and monitored the juvenile survival rates (Ishikawa et al., 2019). The majority of QTL showed heterozygote advantage (overdominance) (Table S2). Following a previous study (Renaut, Nolte, & Bernatchez, 2009), we judged $d/a > 1.5$ as overdominance, where $d$ = mean trait value of heterozygote – (mean trait value of Pacific Ocean homozygote + mean trait value of Japan Sea homozygote)/2; $a$ = absolute value of difference in mean trait value between two homozygotes. For calculation of $d/a$, life duration was used as a trait value.

We also investigated what types of genes are enriched in regions showing signatures of introgression. We first calculated the midpoint of the $f_d$ peaks in each comparison and then identified all genes falling with 100 kb of this. Gene names were then used in a Gene Ontology (GO) analysis. Enriched GO terms were searched using g:Profiler (Raudvere et al., 2019). FDR-corrected p-values < .05 with a hypergeometric test were judged as significant.

3 | RESULTS

3.1 | Population structure and admixture

PCA revealed the primary axis of divergence, which explained 33.3% of variance, was between Japan Sea stickleback and all G. aculeatus (Pacific Ocean and Atlantic; Figure 2a and Figure S1). PC2 explained 6.16% of variance and reflected differentiation within G. aculeatus among the marine populations—the Japanese Pacific Ocean, North American Pacific Ocean and the Atlantic lineages (Figure 2a and Figure S1). Notably, Pacific Ocean individuals from Tokotan were closer in PC space to Japan Sea individuals from the same river.
system, especially in comparison to the same species pair from Bekanbeushi River; suggesting a possibility of a higher level of admixture between the two species in the Tokotan system.

To further investigate this, we used ADMIXTURE to characterize population structure. A value of $K = 2$ separated all $G. \text{aculeatus}$ from $G. \text{nipponicus}$ (see Figure S2). However, a value of $K = 3$ was best supported (Figure S3) and this further split $G. \text{aculeatus}$ marine fish into Atlantic and Pacific Ocean lineages. In both $K = 2$ and $K = 3$, ADMIXTURE showed that Pacific Ocean and Japan Sea individuals from Tokotan were more admixed than those sampled from the Bekanbeushi River system (Figure 2b). Most notably, two Tokotan Pacific Ocean females with the highest PC1 scores also had higher proportions of Japan Sea ancestry from the ADMIXTURE analysis (Figure 2a,b; individuals No. 38 and No. 39). Further analysis using ancestry informative markers showed that these two individuals have a higher level of interspecific heterozygosity and a lower hybrid index, indicating that they might be recent backcrosses (Figure S4).

### 3.2 Differentiation, divergence and introgression

Pairwise mean $F_{ST}$ was highest between Japan Sea fish from Bekanbeushi and Atlantic Ocean stickleback (0.618, see Figure 2c). Focusing on pairwise comparisons from our two Japanese river systems, mean $F_{ST}$ was substantially higher between the Pacific Ocean and Japan Sea fish in Bekanbeushi (0.617) than in Tokotan (0.476, one-tailed permutation test: $p < 1 \times 10^{-4}$). Similarly, genomic divergence measured as $d_{XY}$ was lower in Tokotan (0.0182) compared to Bekanbeushi (0.0199; $p < 1 \times 10^{-4}$; Figure 2d). Genomewide $F_{ST}$ in 100 kb sliding windows also clearly demonstrated that interspecies genomic differentiation is lower in Tokotan compared to Bekanbeushi (Figure 3). This pattern was less apparent with $d_{XY}$ except on chromosome XIX, the Ancestral-sex chromosome (see Figure S5).

To more formally test for the occurrence of introgression, we calculated Patterson’s $D$ statistics for the autosome. We first tested for sharing of alleles between any combination of Pacific Ocean or

---

**FIGURE 2** Genetic structure and differentiation among *Gasterosteus* populations. (a) Principal component analysis (PCA) plot of *G. nipponicus* (Japan Sea) and *G. aculeatus* populations (i.e., Pacific and Atlantic Ocean). Numbers denote individuals with mixed ancestry in (b) ADMIXTURE plot—Table S1 provides key to identify individuals. (c) Pairwise $F_{ST}$ between populations. (d) Pairwise $d_{XY}$ between populations. AT, Atlantic Ocean; JS-B, Japan Sea Bekanbeushi; JS-T, Japan Sea Tokotan; PO-B, Pacific Ocean Bekanbeushi; PO-T, Pacific Ocean Tokotan.
Japan Sea population, all of which resulted in a significantly negative $D$ value (i.e. an excess of ABBA site patterns) (test 1 – Table S2), suggesting introgression between Pacific Ocean and Japan Sea populations. We then altered the test topology to set P1 as PO-B and P2 as PO-T in order to ask which of the two populations shares an excess of sites with either Japan Sea population; this too produces significantly significant $D$ values (test 2 – Table S2), suggesting that there is greater sharing of alleles between Tokotan Pacific Ocean and G. nipponicus. This is further supported by positive $D$ values when P3 is set Atlantic or Pacific Ocean G. aculeatus from North America (test 3 – Table S2). This shows lower sharing between either of these populations and Pacific Ocean Tokotan, suggesting that at least some introgression is occurring from G. nipponicus into the Tokotan PO fish. Finally, we set P1 and P2 to JS-B and JS-T respectively and tested all combinations of G. aculeatus populations as P3 (test 4 – Table S2) in order to ask which of the two Japan Sea populations shares an excess of sites with any of the G. aculeatus populations. This again results highly significant negative $D$ values and an excess of ABBA sites, including when P3 is set to the Atlantic Ocean lineage. This is consistent with the hypothesis that there is more introgression from G. aculeatus into the Japan Sea in Tokotan compared to Bekanbeushi.

### 3.3 Characterizing the introgression sites

We first investigated introgression regions using topology weighting with TWISST (Figure 4). Across the genome, the species topology had the highest mean weight (0.926) as expected for species with high levels of genomic divergence. However, the next highest weight was the geographical topology (0.0402) which indicates introgression between the species in one or both river systems (Figure 4b).

We next investigated finer-scale patterns of introgression with the aim to identify where in the genome gene flow might be occurring. We did this by calculating $f_d$, a statistic derived from ABBA-BABA site frequencies that estimates the proportion of a genome window where sites are shared between two focal taxa. By setting P1 to G. aculeatus from the Atlantic Ocean and the outgroup as P. pungitius, higher values of $f_d$ suggest an excess of allele shared between the Japanese species pairs. In accordance with population structure and $F_{ST}$ analyses, mean $f_d$ (±SD) was highest between Japan Sea and Pacific Ocean in Tokotan (0.078 ± 0.052), compared to Bekanbeushi (0.015 ± 0.048; $p < 1 \times 10^{-4}$). Sliding window analyses identified multiple regions across the genome where $f_d$ was high in both river systems (Figure 4). Using the 99th percentile as a threshold, we identified 469 $f_d$ peaks in the Bekanbeushi system and 543 $f_d$ peaks.
in the Tokotan system, with significantly more peaks in the latter (proportional test: $X^2 = 5.28, df = 1, p = .021$). Of the peaks identified, 348 occurred in both systems, and the overlap was significantly higher than expected by random sampling ($p = 0$; hypergeometric test). The mean size of peaks did not differ significantly between the two comparisons ($p = .051$), but was slightly larger in Bekanbeushi (33.7 kb) compared to Tokotan (29.0 kb). In general, 10 kb windowed estimates of $f_d$ from between-species comparisons in both systems were highly correlated ($r = .72, p < 2.2 \times 10^{-16}$; Figure S6), suggesting either a shared history of introgression or multiple independent introgression events in two systems; nonetheless, a substantial number of windows showed high $f_d$ in Tokotan but not Bekanbeushi.

Theory and empirical observations suggest introgression more likely to be observed when recombination rate is high as linked neutral or adaptive alleles are able to recombine away from maladaptive alleles. Following this, mean recombination rate for $f_d$ peaks versus the genome background was almost two times higher in Bekanbeushi (mean cM/Mb ± SD; peaks = 8.32 ± 19.5; background = 4.57 ± 9.36; permutation test, $p < 1 \times 10^{-4}$; Figure S7a) and 1.5 times higher in Tokotan (peaks = 6.92 ± 14.0; background = 4.58 ± 9.49; $p < 1 \times 10^{-4}$; Figure S7b). We additionally examined the distribution of $f_d$ peaks in the genome and found similar numbers of peaks per Mb on autosomes in both Bekanbeushi and Tokotan (Figures 5 and 6). However, in Tokotan, the highest proportion of $f_d$ peaks for the entire genome

![Figure 4](image-url)

(Figure 4) (a) Three possible tree topologies among four populations. (b) Genome-wide sliding window analysis of weighting of each topology calculated with the TWISST. Blue, a species topology where the same species from different rivers cluster together; red, a geographical topology where different species from the same river cluster together; and green, an interspecific topology where different species from different rivers cluster together.
was found on Chromosome XIX (2.27 peaks per Mb), the Ancestral-sex chromosome (Figure 6) and this was substantially higher than the proportion of peaks on the same chromosome in Bekanbueshi (0.62 peaks per Mb). This is also reflected by a higher overall mean \( f_d \) for Chromosome XIX compared to the rest of the genome in Tokotan (general linear model on 100 resampled windows per chromosome, \( F_{20.1971} = 2.43, p = .004 \); Figure 6c). In contrast, the nonrecombining region of the neo-X (<7.5 Mb of Chromosome IX), where recombination between the neo-X and the neo-Y is suppressed, had no apparent \( f_d \) peaks in either river system (Figure 5).

None of the peaks of LOD scores for juvenile survival rates overlapped with the identified introgression sites in either Bekanbueshi or Tokotan systems. However, five QTLs on Chromosomes IV, VIII, XII and XVI were located within 0.2-1 Mb of introgression sites (Table S3), and two were seen in both river systems. For Tokotan, four QTLs were located close to \( f_d \) peaks, as opposed to three in Bekanbueshi. Among the five total QTLs, three showed heterozygote advantage. We additionally identified a total of 4,667 genes occurring within 100 kb of \( f_d \) peaks in both comparisons with 3,411 seen in Bekanbueshi and 3,866 in Tokotan. GO analysis showed that genes in the close vicinity of introgression sites from both sites were enriched for GO terms related to olfactory receptor activity and detection of chemical stimuli (Table S4).

**FIGURE 5** Genome-wide sliding window analysis of \( f_d \). Red and blue colours indicate the species pairs in Bekanbueshi and Tokotan systems, respectively. For analysis of sex chromosomes (Chromosome IX and XIX), we used only females.

### DISCUSSION

Our results clearly indicate that genomic divergence is lower and introgression rates are higher in the Tokotan system compared to the nearby Bekanbueshi system. In Bekanbueshi, the most extreme breeding habitats of the two species are separated by nearly 5 km, whereas in Tokotan, both species spawn in a 200 m stretch (Figure 1). Therefore, our results are consistent with the hypothesis that isolation of breeding habitats contributes to and maintains reproductive isolation between anadromous fishes. Although we observed that these two species are evenly distributed in Tokotan River, further detailed analysis of microhabitat use in Tokotan River is necessary to quantitatively measure the strength of habitat isolation in this river system. Additionally, it is possible that local adaptation to more divergent breeding habitats in Bekanbueshi might further contribute to reduced effective introgression in this river system relative to the Tokotan River system. Analysis of other nearby rivers containing the species pairs and differing in length would also help to test the role of habitat isolation in reproductive isolation of anadromous species.

Although introgression can be adaptive (Arnold, Sapir, & Martin, 2008; Hedrick, 2013; Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015), it is not always straightforward to conclusively show that this is the case. The integration of QTL analysis of...
fitness-related traits alongside population genomics is a good means of testing whether introgression has indeed played a role in adaptation. In the present study, we found that three QTLs with heterozygote advantage on juvenile survival are located near to introgression sites. This is consistent with the idea that loci with heterozygote advantage can introgress at higher rates than other genomic regions (Kim et al., 2018; Nadachowska-Brzyska et al., 2012; Wolfe et al., 2019), although we caution that the QTL peaks do not overlap with introgressed regions. We further found that the introgression sites we identified were enriched for genes related to olfactory reception and the detection of chemical stimuli. These genes were located on Chromosome I (257,080–372,834 bp) and Chromosome XVI (4,224,687–4,301,585 bp and 13,791,276–15,682,213 bp).

Interestingly, genes linked to olfactory receptor activity show signatures of overdominance in humans (Alonso, López, Izagirre, & de la Rúa, 2008). A plausible alternative explanation is that olfactory receptors are involved in conspecific mate recognition mechanism and are potentially acting as a one-allele barrier (Felsenstein, 1981; Garner, Goulet, Farnitano, Molina-Henao, & Hopkins, 2018; Smadja & Butlin, 2009). Positive selection for mate recognition where breeding grounds overlap would promote adaptive introgression in this case. Nonetheless, further work is now necessary to distinguish between these two alternative explanations. Furthermore, we found that many introgression sites were shared between both Bekanbeushi and Tokotan although the latter system has a larger number of unique regions. This suggests that introgression between the two species has likely occurred throughout their evolutionary history (Ravinet et al., 2018) but that it is also ongoing. In addition, shared regions of introgression and the presence of introgression close to genes and QTLs of functional significance are consistent with the hypothesis that some introgression in this system may be adaptive.

We additionally found that the strongest signatures of introgression occurred more often in regions with a higher recombination rate. Association between recombination rates and introgression has been reported in both theoretical (Butlin, 2005; Nachman & Payseur, 2012; Noor & Bennett, 2009) and empirical studies (Brandvain et al., 2014; Janoušek et al., 2015; Juric et al., 2016; Martin et al., 2019; Samuk et al., 2017; Schumer et al., 2018). In low recombination regions, selection purging deleterious foreign alleles is more likely to remove nearby physically linked loci, including both neutral and slightly advantageous alleles (Barton & Bengtsson, 1986; Ravinet et al., 2017). In high recombination regions, the probability that linked loci will recombine away from deleterious introgressed alleles is higher and thus signatures of introgression are more likely (Schumer et al., 2018). It should be noted, however, that our present analysis assumed that recombination rates were conserved among stickleback populations and species. The linkage map we used to determine the recombination rate was derived from a cross between sticklebacks from the Atlantic Ocean lineage (Roesti et al., 2013). Thus, it is necessary to test whether this assumption is valid or not using a recombination rate map estimated using Japanese crosses.
in the future. Additionally, we also cannot exclude the possibility that regions with signatures of introgression result from introgression that occurred before the split between the Tokotan and Bekenbeushi populations or from ancestral polymorphism (Nelson & Cresko, 2018). The apparent overlap of 'geography' and 'interspecific' topologies in the TWISST analysis (Figure 4) may reflect ancestral polymorphism or ancestral introgression before the geographical split between the Bekenbeushi and Tokotan systems. In our D statistics test, when we used (((JS-B, JS-T), P3), O) and varied P3 for all G. aculeatus populations (test 4 in Table S2), D is significantly negative in all cases, indicating admixture between the Japan Sea in Tokotan and G. aculeatus groups. Because the Japan Sea species are not genetically structured (i.e., genetically similar among populations) (Ishikawa et al., 2019), the divergence between JS-Tokotan and JS-Bekenbeushi is likely to be more recent than the split among PO, NA, and AT of G. aculeatus. Therefore, these D statistics results likely indicate that gene flow has occurred from G. aculeatus to Tokotan after the split between the Tokotan and Bekenbeushi Japan Sea sticklebacks. Next, when we set P1 to PO-B and P2 to PO-T, all D values became positive regardless of whether P3 is AT or PO-NA, which may be due to introgression from G. nipponicus to PO-T. These data indicate that introgression may occur in both directions. A further comparative analysis of introgression patterns is necessary to distinguish this more clearly.

Although the rate of introgression is likely higher in Tokotan River, the two species remain clearly genetically differentiated (Figure 2) and have clearly not merged into a single species, in contrast with several well-known cases of speciation reversal (Seehausen et al., 2008; Taylor et al., 2006). This is likely because there are multiple isolating barriers present in the Japanese stickleback system, including hybrid male sterility, a post-zygotic incompatibility (Kitano et al., 2009). It is predicted that the presence of intrinsic incompatibilities and/or polygenic isolating barriers can lead to irreversible speciation (Barton & Bengtsson, 1986; Nosil et al., 2009; Seehausen, 2006) and our findings support this. Sex chromosome turnover in the Japanese stickleback has previously been linked to reproductive isolation in this system (Kitano et al., 2009; Yoshida et al., 2014). Here our analysis shows that the neo-X chromosome did not show signs of introgression except at the distal end away from the fusion site (Figure 5). This could potentially be explained by a lack of recombination in males carrying the neo-Y despite the presence of recombination in females, reducing the effective recombination rates in this chromosomal region. Additionally, the hemizygosity of the neo-X in males can expose recessively deleterious introgressed alleles to negative selection and efficiently purge them (Charlesworth, Coyne, & Barton, 1987; Johnson & Lachance, 2012). Furthermore, a previous QTL analysis showed that the neo-X chromosome contains QTL for the behavioural intensity of dorsal pricking and dorsal spine length, which is more sexually dimorphic in the Japan Sea stickleback and used for male courtship in this species (Kitano et al., 2009). Although these QTL are not located within the nonrecombining region of the neo-X, they are candidates for barrier loci that might act to reduce local introgression in the genome. In contrast to the neo-X, the ancestral-sex chromosome showed a larger number of introgression sites than the genomic background (Figure 6a) and also a distribution of f2 with a higher mean in Tokotan (Figure 6c). We are currently uncertain of what might cause this pattern, but higher introgression on sex chromosomes compared to autosomes has also been observed in other stickleback systems (Dixon, Kitano, & Kirkpatrick, 2019).

In conclusion, we found that the absence of habitat isolation substantially increases introgression and erodes genomic divergence between anadromous species. Additionally, we found that two species pairs with different levels of habitat isolation shared many introgression sites and some regions of introgression were located near QTL with heterozygote advantage on juvenile survival and genes that have previously been linked to heterozygote advantage in other species. Taken together, our findings suggest that at least some of the introgression observed may be adaptive. Despite the presence of introgression, these two species are still clearly differentiated. Other isolating mechanisms such as behavioural isolation and hybrid sterility likely contribute to reproductive isolation. Thus, the combination of multiple isolating barriers maintains reproductive isolation between the anadromous species, and the breakdown of one such barrier, here habitat isolation, is not enough to cause speciation reversal in this species pair. Overall, our data demonstrate that comparative genomic analysis of multiple species pairs with different levels of habitat isolation is informative for understanding how reproductive isolation and genomic divergence can be maintained by habitat isolation and other isolating barriers.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/jeb.13664.

ACKNOWLEDGMENTS
This research was supported by grant-in-aid for Scientific Research (A) (19H01003) to JK. We are grateful to the DNA Data Bank of Japan for providing high-performance computing access to perform analyses. We also thank Sean Stankowski, Thom Nelson and two anonymous reviewers for their comments on a previous draft of the manuscript.

ORCID
Mark Ravinet https://orcid.org/0000-0002-2841-1798
Jun Kitano https://orcid.org/0000-0001-8659-5698

REFERENCES
Akkeshi Town (2006). Report of Akkeshi Town Maritime Memorial Museum No.2. pp. 2. Akkeshi.
Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12, 246. https://doi.org/10.1186/1471-2105-12-246
Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664. https://doi.org/10.1101/gr.094052.109
Seehausen, O. L. E., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, 17, 30–44. https://doi.org/10.1111/j.1365-294X.2007.03529.x

Smadja, C., & Butlin, R. K. (2009). On the scent of speciation: The chemosensory system and its role in premating isolation. *Heredity*, 102, 77–97. https://doi.org/10.1038/hdy.2008.55

Taylor, E. B., Boughman, J. W., Groenenboom, M., Snitayski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15, 343–355. https://doi.org/10.1111/j.1365-294X.2005.02794.x

White, M. A., Kitano, J., & Peichel, C. L. (2015). Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. *Molecular Biology and Evolution*, 32, 1981–1995. https://doi.org/10.1093/molbev/msv078

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2006). Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist*, 167, 794–807.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A grammar of data manipulation*. R package version 0.7.6. Retrieved from https://CRAN.R-project.org/package=dplyr

Wolfe, M. D., Bauchet, G. J., Chan, A. W., Lozano, R., Ramu, P., Egesi, C., ... Jannink, J.-L. (2019). Historical introgressions from a wild relative of modern cassava improved important traits and may be under balancing selection. *Genetics*, 213, 1237–1253. https://doi.org/10.1534/genetics.119.302757

Yamada, M., Higuchi, M., & Goto, A. (2007). Long-term occurrence of hybrids between Japan Sea and Pacific Ocean forms of threespine stickleback, *Gasterosteus aculeatus*, in Hokkaido Island, Japan. *Environmental Biology of fishes*, 80, 435–443. https://doi.org/10.1007/s10641-006-9142-0

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, 110, E1743–E1751. https://doi.org/10.1073/pnas.1219381110

Yoshida, K., Makino, T., Yamaguchi, K., Shigenobu, S., Hasebe, M., Kawata, M., ... Kitano, J. (2014). Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genetics*, 10, e1004223. https://doi.org/10.1371/journal.pgen.1004223

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

---

**How to cite this article**: Ravinet M, Kume M, Ishikawa A, Kitano J. Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *J Evol Biol*. 2021;34:114–127. https://doi.org/10.1111/jeb.13664