



OPEN

## Population transcriptomics reveals the effect of gene flow on the evolution of range limits

Katsunori Tamagawa<sup>1,3✉</sup>, Kotone Yoshida<sup>2</sup>, Shiori Ohru<sup>2</sup> & Yuma Takahashi<sup>1✉</sup>

One of the most important questions in evolutionary biology is how the spatial distribution of species is limited. Asymmetric gene flow from core populations is suggested to increase the number of poorly adapted immigrants in the populations at the range edge. Genetic load due to migration, i.e., migration load, should prevent adaptation to the local habitat, leading to decreases in distribution range via local extinction or the limiting range expansion. However, few experimental studies have examined the effects of immigration on fitness and natural selection within recipient populations. To investigate the influence of migration load on the evolution of distribution range, we performed field and laboratory observations as well as population transcriptomics for the common river snail, *Semisulcospira reiniana*. This species meets the conditions that migration from source populations can prevent local adaptation in a sink population because they inhabit the broader range of environments, including middle/upper reaches of a river and estuaries within a single river and they may be more vulnerable to being swept away by water currents due to lowered spontaneous (upward) locomotion activity. We found that river steepness was related to the lower distribution limit of *S. reiniana*, with a narrower distribution range in the steeper river. Population transcriptomic analysis showed that gene flow was heavily asymmetric from the upstream populations to downstream ones in the steep river, suggesting a greater migration load in the steep river. The number of genes putatively involved in adaptation to the local habitat was lower in the steep river than in the gentle river. Gene expression profiles suggested that individuals achieve better local adaptation in the gentle river. Laboratory experiments suggested that evolutionary differences in salinity tolerance among local populations were only found in the gentle river. Our results consistent with the hypothesis that migration load owing to asymmetric gene flow disturbs local adaptation and restricts the distribution range of river snails.

Traits within a species are known to change gradually across the geographical space. The existence of spatial variation in fitness-related traits indicates that populations can adapt to environmental continuums within the species' range or within a habitat and that they can adapt to novel environments with inhospitable conditions<sup>1</sup>. In practice, abundant evidence of ongoing adaptation to environmental continuums has been reported<sup>2,3</sup>. Thus, given evolutionary periods of time, any species can adapt to novel environments and extend its range permanently across the globe. However, the distribution of many species ends abruptly, without the presence of obvious ecological or physical barriers, suggesting that adaptation is prevented at the edge of the range<sup>4</sup>. Elucidation of the processes preventing adaptive evolution at range margins is therefore important to our understanding of the evolutionary mechanisms that determine species ranges and the pattern of species diversity on earth.

There are two theoretical explanations for the failure of adaptation at range margins. (i) If migration moves mainly from the central, well-adapted regions of the range into the margins, asymmetric dispersal increases the proportion of poorly adapted immigrants in the margins of a population. The range margins may have lower carrying capacity (i.e., migration load) and consequently, the immigrants fail to adapt to the novel environment. This hypothesis is called the migration load hypothesis<sup>5–8</sup>. (ii) If the range edge is highly fragmented, stochastic events and the low rate of mutational input into marginal populations may limit the availability of locally beneficial alleles, preventing adaptation and therefore range expansion (diversity loss hypothesis)<sup>9</sup>. In general, failure of adaptation along an environmental gradient due to diversity loss is recognized as the predominant process at

<sup>1</sup>Graduate School of Science, Chiba University, 1-33, Yayoi, Inage, Chiba 263-8522, Japan. <sup>2</sup>Graduate School of Science and Engineering, Chiba University, 1-33, Yayoi, Inage, Chiba 263-8522, Japan. <sup>3</sup>Present address: Graduate School of Life Sciences, Tohoku University, 6-3, Aoba, Aramaki, Sendai, Miyagi 980-8578, Japan. ✉email: tamagawa.katsunori@gmail.com; takahashi.yum@gmail.com

larger geographic scales because the effect of migration input and migration load is relatively small<sup>10</sup>. In practice, many empirical studies have detected indications that the loss of genetic diversity is the predominant mechanism preventing adaptation. Genetic diversity in a population gradually decreases, and genetic differentiation among the population gradually increases, along the core to edge transects at large geographical scales<sup>11–13</sup>. In contrast, if an environmental gradient is steep at a small spatial scale, the population often receives poorly adapted immigrants from adjacent populations<sup>14</sup>. Thus, the failure of adaptation along the environmental gradient due to migration load is suggested to be the predominant process at a small spatial scale and in species with higher dispersal ability<sup>8,15</sup>. Despite theoretical progress, the actual effect of migration load on the adaptation of populations at range margins remains unclear due to the lack of estimation of gene flow and the genetic and phenotypic signature of maladaptation, although some transplant studies and meta-analyses have provided circumstantial evidence for migration load<sup>14,16,17</sup>.

The lotic environment is a suitable system to examine the effects of asymmetric gene flow on local adaptation and range expansion since lotic organisms are exposed to unidirectional water flow and occasional floods. Water currents along a river encourage organisms to move downstream; hence, gene flow occurs most frequently downstream<sup>18–20</sup>. Despite relatively high mobility, the dispersal of fishes is also influenced by stream currents<sup>21–23</sup>. A previous study assessed the phenotypic traits of a stickleback, *Gasterosteus aculeatus*, population inhabiting a pond and its inlet and outlet rivers and showed that the population in the outlet (downstream) of the pond exhibits a phenotype intermediate between that of the river- and pond-adapted fish, whereas the upstream population has a phenotype optimized for the river environment<sup>23</sup>. This inconsistency between the optimal and realized phenotype resulted from frequent immigration from the pond population, following the water flow. Organisms with a limited ability to disperse, such as freshwater snails and other invertebrates, are assumed to be more susceptible to the effects of asymmetric gene flow because passive movements will frequently occur<sup>24–27</sup>. The extent of asymmetric gene flow is also affected by the geographic and physical features of each river since passive movement is caused by environmental factors, such as frequency of floods, nature of the substratum, and the velocity and steepness of the river<sup>18,28</sup>.

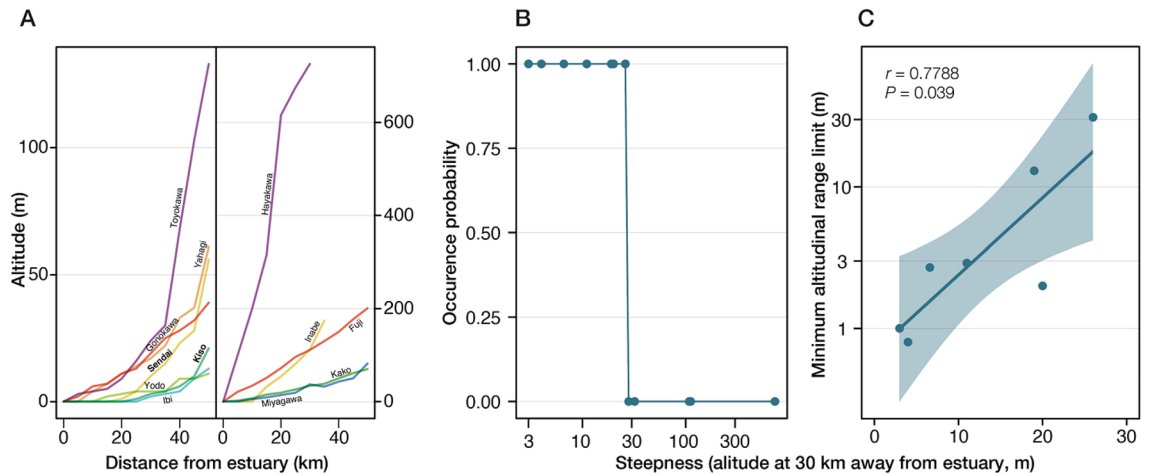
The freshwater snail, *Semisulcospira reiniana*, is a common gastropod species in Japan, inhabiting mainly sandy or muddy bottoms of flowing rivers and lakes<sup>29</sup>. This species is broadly distributed within the riverine system including estuaries. Field observation indicated that individuals transplanted into faster currents are prone to migrate downstream<sup>24</sup>. Therefore, comparative research into rivers inhabited by *S. reiniana*, which have different geographic features, is expected to clarify the relationship between gene flow, local adaptation, and the distribution range of species.

Population genetics using single nucleotide polymorphisms (SNPs) is a powerful method used to survey population structures and demography in freshwater environments<sup>26,30,31</sup>. Recent technological developments in next-generation sequencing allow us to analyze both genome-wide SNPs and the transcriptome of wild organisms<sup>14,32–34</sup>. The transcriptomic analysis provides evidence for adaptation and phenotypic change by identifying SNPs located within the mRNA and changes in gene expression<sup>35</sup>. By comparing rivers with steep and gentle gradients, we could infer the impact of the river gradient on local adaptation and the distribution range along the river. Recently, the genetic basis of response to salinity stress in *S. reiniana* was studied by transcriptomic analysis<sup>36</sup>; however, the influence of gene flow on local adaptation and range limit is not known yet in this species. Here we performed population transcriptomics for *S. reiniana* populations collected from rivers of different steepness in order to investigate the influence of gene flow on local adaptation and distribution range. If river steepness influences the direction of migration and local adaptation at range margins, populations in each river should show distinct properties of genetic structure and transcriptional regulation. We also conducted a field survey to examine the influence of river steepness on phenotypic traits. To provide support for the hypothesis that migration load influences the evolution of range limit, the relationship between the river steepness and the lower range limit was analyzed using 12 independent rivers. Our results consistently imply that asymmetric gene flow affects local adaptation and the evolution of range limit.

## Results

**Relationship between river steepness and distribution limits.** The lower distribution limit of *S. reiniana* was investigated in 12 rivers in Japan—Ibi, Kiso, Yodo, Sendai, Gonokawa, Yahagi, Toyokawa, Miya-gawa, Kako, Fuji, Inabe, and Hayakawa Rivers—between 2017 and 2020 (Fig. 1A). The elevation at 30 km from the estuary, a measure of the river steepness, was strongly related to the occurrence of *S. reiniana* in each river (Fig. 1B). In some of the quite steep rivers, no *S. reiniana* were observed up to 50 km from the estuary, while we found that the lower distribution limit was from 10 to 35 km from the estuary in other rivers. The river steepness was significantly correlated with the lower distribution limit of *S. reiniana* (Fig. 1C). The expansion of distribution range into the intertidal zone was observed only in the Kiso River.

**Evaluation of phenotypic adaptation in the peripheral population.** In the present study, we selected two rivers, the Kiso and Sendai Rivers, with different steepness and enough snails to generate reliable results. In the Sendai River with a relatively steeper slope, the distribution of this species is limited to freshwater habitat, while in the Kiso River with a gentle gradient, the distribution extends into the intertidal zone. The Kiso and Sendai Rivers are referred to as the gentle and steep rivers, respectively. The shell of the juvenile snails was large in the peripheral, downstream population in the gentle river, and a significant switching point for shell size was detected between the brackish water and freshwater populations (Fig. 2A,B). Juvenile snails born from females in each population were subjected to 0%, 1%, 2%, or 3% saline water. Although most of the juveniles exposed to 0% saline water survived, several juveniles died in the water with a higher salinity (altitude,  $\chi^2 = 23.32$ ,  $P < 0.001$ ; river,  $\chi^2 = 18.93$ ,  $P < 0.001$ ; concentration,  $\chi^2 = 574.05$ ,  $P < 0.001$ ; size,  $\chi^2 = 8.44$ ,  $P = 0.004$ ;



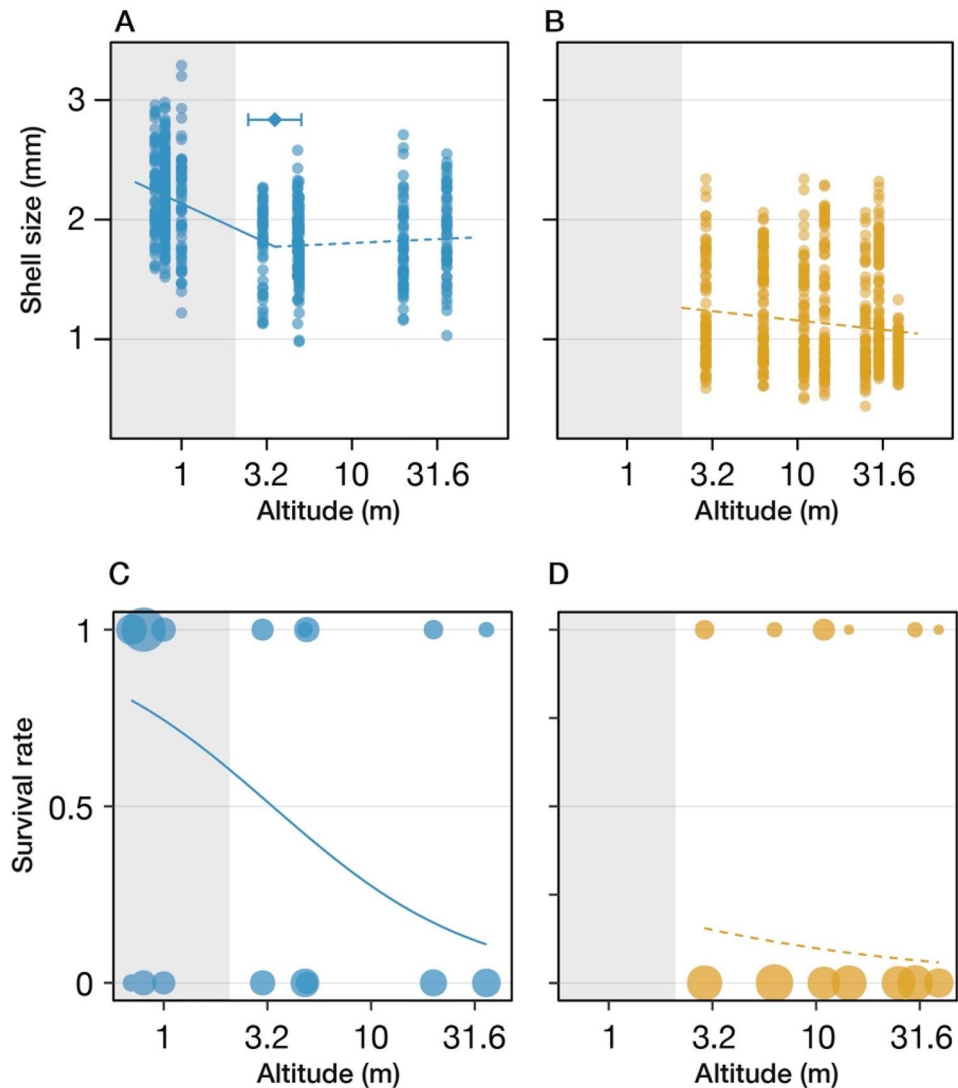
**Figure 1.** Altitudinal distribution of *Semisulcospira reiniana*. (A) Steepness of each river surveyed in this study. Relatively gentle and steep rivers are represented in the left and right panel, respectively. (B) Occupancy probability of *S. reiniana* by steepness in each river. (C) The relationship between the river steepness and the lower distribution limit of *S. reiniana*. Altitude at 30 km away from estuary is used as a measure of river steepness. Correlation coefficient and *P*-value are shown in the upper left of the panel.

altitude  $\times$  river,  $\chi^2 = 9.54$ ,  $P = 0.002$ ; altitude  $\times$  concentration,  $\chi^2 = 0.26$ ,  $P = 0.61$ ; river  $\times$  concentration,  $\chi^2 = 1.48$ ,  $P = 0.22$ , Fig. S1). The juveniles of downstream populations in the gentle river had higher survival rates than those of upstream populations in water with 3% salinity, whereas most of the juveniles in the steep river died, irrespective of their origins (Fig. 2C,D). A significant effect of body size and altitude of the location of the maternal individuals was detected only in the gentle river (gentle: altitude,  $\chi^2 = 17.863$ ,  $P < 0.001$ ; shell size,  $\chi^2 = 15.903$ ,  $P < 0.001$ , the steep river: altitude,  $\chi^2 = 1.66$ ,  $P = 0.197$ ; size,  $\chi^2 = 0.87649$ ,  $P = 0.35$ ), indicating genetic differences in salinity tolerance among populations in the gentle river.

**Reference transcriptome construction and evaluation.** Total RNA was extracted from 87 individuals collected from the Kiso and Sendai Rivers. We obtained a total of around 2.5 billion pairs of high-quality reads using Illumina sequencing, and de novo assembly produced 547,731 contigs. After removal of redundancy and screening for protein-coding sequences, 75,670 contigs were retained. We also obtained 184,501 high-quality, full-length Iso-Seq sequences from five individuals, of which 46,891 sequences were retained after removing redundancy. Based on these full-length transcripts, reference-guided assembly was performed to equalize the mapping result among populations, resulting in 18,878 sequences. Reference-guided and de novo assembly were merged to generate 89,850 reftigs with a mean length of 1,425 bp. The reftigs had 96.1% comprehensiveness for the metazoan dataset in Benchmarking Universal Single-Copy Orthologs (BUSCO) v3, and 51,009 sequences had high homology with *C. gigas* protein sequences (E value  $< 1.0 \times 10^{-4}$ ). The basic information about the reference construction is summarized in Table S1.

**Population genetic analysis in each river.** Variant calling and filtering resulted in 7,820 SNPs from the gentle river and 10,721 SNPs from the steep river. Most SNPs were in the UTR, and there were fewer non-synonymous SNPs than synonymous SNPs (Table S2). Using these SNPs, we conducted population genetic analyses (see Table S3 for basic population genetic parameters). To examine the effect of the steepness on the population genetic structure in the two rivers, we analyzed the relationship between geographic distance and pairwise  $F_{ST}$  among populations. Unlike our expectation, we could not detect the isolation-by-distance (IBD) patterns in gentle river ( $r = -0.408$ ,  $P = 0.808$ ) while IBD was detected in steep river ( $r = 0.965$ ,  $P = 0.042$ , Fig. 3A). However, IBD pattern was also detected in the gentle river when we excluded brackish water populations of the river from the analysis ( $r = 0.795$ ,  $P = 0.042$ ) since all pairwise  $F_{ST}$  values between a brackish population and freshwater populations were low in the gentle river (Fig. 3A).  $F_{ST}$  values per geographic distance along a river were significantly higher in the gentle river than the steep river ( $P = 0.002$ ), though such pattern tended to be ambiguous in a case where brackish water population was included ( $P = 0.056$ ). Migration probability estimated using BaysAss3-SNPs indicated that moderate to high gene flow occurs between populations within a river (Fig. S2). Downward migration was significantly stronger than upward migration only in the steep river (Fig. 3C). Asymmetric gene flow, which was assessed by the log-transformed ratio of the downward to upward migration probability, was detected only in the steep river (Fig. 3D). The estimation of genetic population structure was performed using the sparse nonnegative matrix factorization (*snmf*) function implemented in Landscape and Ecological Association (LEA) package of the *R* statistical language. This analysis did not identify different genetic population structures within rivers, and the best number of genetic clusters, *K*, was one (Fig. S3).

The number of outlier SNPs which were strongly associated with elevation (Table S4, S5), was relatively greater in the gentle river (1,528 SNPs) than in the steep river (1,045 SNPs). In the gentle river, 232 SNPs were non-synonymous and involved in a metabolic process, while in the steep river, 170 SNPs were non-synonymous and enriched in Gene Ontology (GO) terms related to response to stimulus and cell development (Table S6,

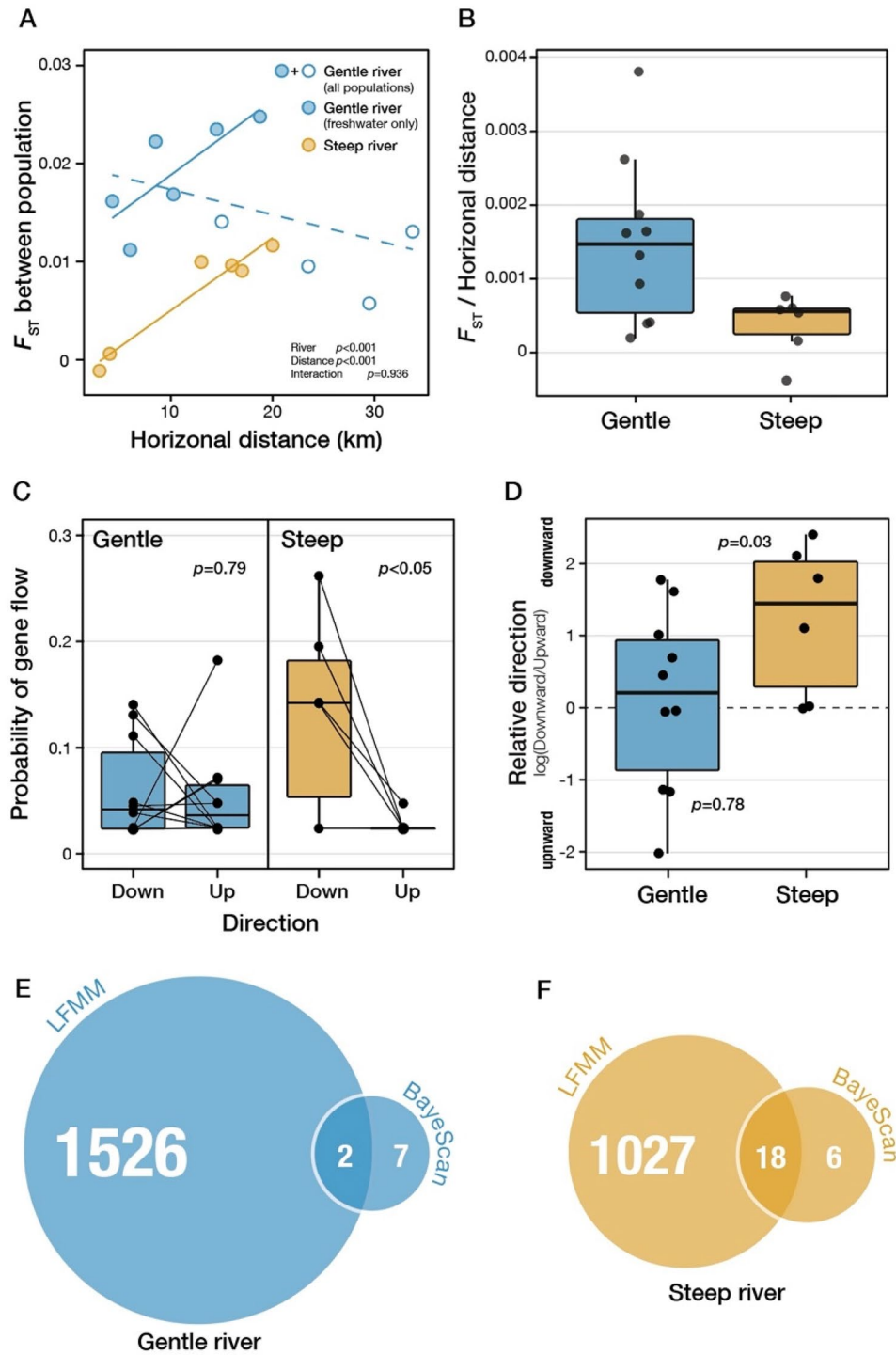


**Figure 2.** Shell size of juveniles in each population in the gentle (A) and steep rivers (B). Each line was estimated using a segmented regression method, and the rhombus and whisker represent the significant segmented point and the confidence interval, respectively. The shaded area in each panel indicates the intertidal region in the gentle river (Kiso River). The survival rate of juveniles exposed to 3% saline water in the gentle (C) and steep rivers (D). The size of each point shows the number of juvenile dead (bottom of each panel) or alive (top of each panel). The curves in each panel were estimated using a generalized linear model, and the line type represents significance.

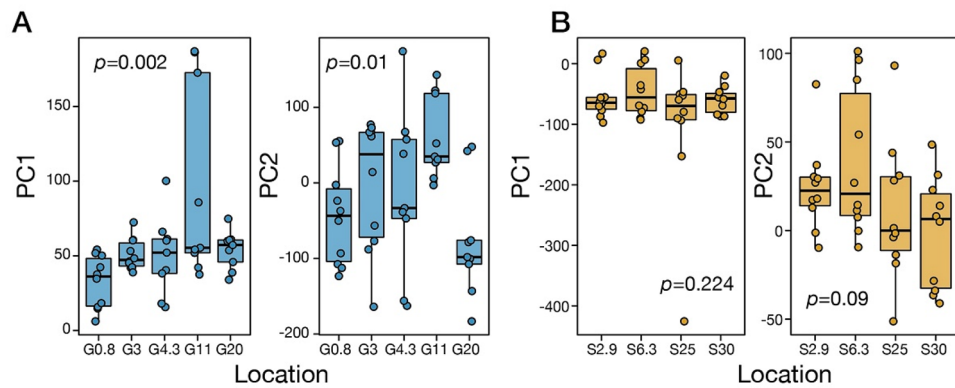
S7). Second, we used *BayeScan*, an  $F_{ST}$ -based method, which detected nine and 24 outlier SNPs in the gentle and steep river populations (Table S8, S9). In the gentle river, two non-synonymous SNPs located in the genes *Bis(5'-nucleosyl)-tetrphosphatase* and *Caskin-1* were detected by both methods; in the steep river, five out of 18 SNPs detected by both methods were non-synonymous (Fig. 3E,F).

**Spatial variation in gene expression in each river.** Principal component analysis (PCA) based on gene expression profiles identified separation between the rivers, but not among populations within a river (Fig. S4). To clarify the differences in expression profiles among the populations in each river, we extracted and examined the variance of the principal component (PC) values. PC1 and PC2 values differed significantly among populations in the gentle river, but not in the steep river (Fig. 4).

The number of transcripts correlated with an elevation in each river was greater in the gentle river than in the steep river (Fig. S5, Table S10). In the gentle river, 21 transcripts were significantly correlated with elevation (positive: 1, negative: 20), while only five transcripts were correlated with elevation in the steep river (positive: 5, negative: 0).



**Figure 3.** Results of the population genetic analyses. **(A)** The relationship between horizontal distance between populations and the pairwise  $F_{ST}$  as genetic differences. Filled circles are values between freshwater populations, and open circles are values between fresh and brackish water populations.  $P$ -values based on generalized linear models are shown in the panel, and the blue dashed line represents the regression using all five populations of the gentle river. **(B)** Ratio of  $F_{ST}$  to horizontal distance in each river. **(C)** Probability of the downward and upward gene flow between populations in each river.  $P$ -values based on paired  $t$ -tests are shown in each panel. **(D)** The ratio of downward to upward gene flow. Differences between the upward and downward gene flow in each population were tested by one-sample  $t$ -tests, and  $P$ -values are shown beside each box. **(E), (F)** The number of outlier SNPs detected by LFMM and BayeScan in each river.



**Figure 4.** Principal component analysis using expression profiles of both rivers. Boxplots show the variance of PC1 and PC2 values of each population in the gentle (A) and steep rivers (B). In each panel,  $p$ -values based on analysis of variance are shown.

**Genetic basis of high tolerance to salinity downstream in the gentle river.** To detect the genetic basis of tolerance to the highly stressful estuary environment, which includes high salinity and exposure to air during low tide, we conducted an in-depth analysis of the gentle river. The salinity of river water measured by the refractometer was 5, 0, 1, 1, and 1‰ at G0.8, G3, G4.3, G11, and G20, respectively, which show the population at brackish water is exposed to extremely high salinity stress (Table S3). We detected 812 SNPs that were significantly correlated with the salinity in each location, and 147 out of these were non-synonymous (Tables S11, S12). We also identified five genes that were differentially expressed in at least three comparisons between brackish and freshwater populations within the gentle river, and four of the five genes were downregulated in the population living in brackish water (Fig. S6, Table S13). Note that, salinity at tidal (brackish water) area might be underestimated because it was measured at low tide to access the habitat of the species.

## Discussion

In the present study, we demonstrated a relationship between river steepness and the lower distribution limit of a river snail using field data and population transcriptomic data, focusing on SNPs in transcripts and gene expression levels. Although no clear separation of population structures was observed within the rivers, the estimated direction of gene flow and differentiation among the populations supported the assumption that strong downward migration homogenizes the population in the steep river. In the saline tolerance experiment, gene expression analysis and field data suggested that local evolutionary adaptation could be attributed to the expansion of the distribution range to a novel environment, in this case, the intertidal environment, in the gentle river. In contrast, a recent review suggested that most of the studies regarding the effect of gene flow on local adaptation were consistent with the diversity loss hypothesis, and few studies support the migration load hypothesis<sup>37</sup>. Some studies failed to detect any swamping effect on local adaptation under asymmetric gene flow<sup>38</sup>. One possibility is that the swamping effect of asymmetric gene flow is weaker than expected from theoretical studies, owing to complex environmental and ecological conditions, and is hence difficult to demonstrate except in quite simplified systems<sup>17,38</sup>. These results, which appear to contradict ours, might arise from the different consequences of positive and negative gene flow. While the negative effect leads to population collapse at the edge of the range<sup>4,15</sup>, the positive effect of gene flow leads to the formation of stable populations and subsequent range expansion<sup>37,39</sup>. Therefore, populations recognized by us, the abundant and fine population, have probably been subjected to the positive effects of contemporary gene flow, so it is hard to detect the negative effects of gene flow using the existing population. In the present study, we detected the negative effects of gene flow. We overcame this challenge by focusing on whether river snails expand into novel environments and by comparing rivers with different slopes. Our results supported the hypothesis that migration load owing to asymmetric gene flow, along with the flow of the stream, disturbs local adaptation and restricts the distribution range of a river snail. The lack of replication for genetic analysis (we only studied one steep and one gentle river) is a weakness of our study, but the analysis on the relationship between the steepness of river and lower range limit by using multiple rivers can strengthen our hypothesis.

**Genetic differentiation is more apparent in the gentle river.** Our results revealed that the population structure within each river is ambiguous, and thus genetic differentiation among the subpopulations is predicted to be small (Fig. S2), probably due to the small geographic distance (less than ~35 km) along a river. Previous studies have also suggested that population structure tends to be ambiguous in river snails because they tend to migrate to other locations, mainly downstream, due to water flow, including floods<sup>24–27</sup>. The results of the present study supported the presence of passive asymmetric downward migration and low genetic differentiation among populations in the steep river (Fig. 3C,D). Passive downward migration may predominate over the persistent, aggressive upstream movement in a steep river<sup>40</sup>. In a gentle river, however, aggressive upstream movement contributed to the balance of downstream and upstream migration. Thus, our results suggested that river steepness leads to heavily asymmetric gene flow and introduces poorly adapted individuals or alleles into

the downstream population. Field surveys using multiple rivers have identified a significant correlation between river steepness and the range of distribution toward the estuary (Fig. 1B,C). Taken together, in *S. reiniana*, range limitation and local adaptation to novel and harsh habitats probably depends upon the extent of influx from a different environment.

**Detection of natural selection and inference of functional associations within each river.** Evidence of local adaptation could be detected in the form of outlier SNPs with different allele frequencies in each population if the population had adapted to the local environment at each elevation, including temperature, sediment, and salinity. In the steep river, functions such as morphogenesis, developmental process, and response to external stimulus were significantly enriched (Table S6), and in the gentle river, GO terms associated with RNA and other metabolic processes were significantly enriched (Table S7). In the present study, however, we found no shared non-synonymous SNPs associated with the elevation in both rivers, and only two GO terms (cellular process and regulation of biological process) detected by LFMM were shared in both rivers. While these findings suggest that local adaptation to the elevation has been accomplished by different genetic pathways in the different rivers, we should be careful to discuss the association between enrichment analysis and the mechanism of local adaptation. To verify the effect of each SNP on local adaptation associated with elevation, we require more evidence of related ecological features in this species and genetic studies for genes where SNPs are located.

In the steep river, 18 SNPs detected by BayeScan were included in 1,045 SNPs detected using LFMM (Tables S4, S8). Many of these SNPs were in the myosin heavy chain (MHC) in the striated muscle and exhibited a clear separation of allele frequency between the upstream and downstream populations. One contig (TRINITY\_DN279260\_c4\_g3\_i9) encoded one putative homolog of MHC and contained two non-synonymous alleles in its coding sequence<sup>41</sup>. Typical striated muscle has not been found in “probranchs”<sup>42,43</sup>; therefore, MHC could be involved in embryogenesis or early stage muscle development. Since the upstream region has a particularly steep flow and rocky substratum in the steep river, non-synonymous alleles located in the MHC, which are possibly associated with muscle development, were considered to be possible targets of selection. Our pilot analysis revealed that the speed of movement of newly-born individuals whose female parent came from the upstream population was greater than those whose female parent came from the downstream population in the steep river. This observation suggested that only the upstream population, which is a source rather than a sink population, may achieve local adaptation in the steep river (Fig. S7). This pattern is not consistent with the previous studies that have reported a higher dispersal ability at the range edge of populations<sup>44</sup>.

**Differences in transcriptional regulation were greater in the gentle river.** In the present study, we considered the expression level of each transcript to be a phenotypic trait and evaluated the state of local adaptation of each river. Since the global gene expression pattern differed significantly with elevation only in the gentle river, individuals inhabiting the gentle river presumably produced the appropriate transcriptional regulation for each local environment (Fig. 4). All individuals were subjected to a uniform environment before RNA fixation, and therefore, observed patterns were assumed to reflect genetic factors rather than phenotypic plasticity.

We observed that more transcripts exhibited expression levels correlated with an elevation in the gentle river than in the steep river (Table S10). These results suggest that local adaptation and habitat expansion rely on transcriptional regulation in *S. reiniana*. Although differences in expression are mainly governed by a mutational change in regulatory regions, in the present study, we focused only on SNPs within the mRNA, and thus we could not demonstrate the existence of SNPs in the regulatory region or investigate these effects on local adaptation. To further investigate the local adaptation in *S. reiniana*, we need to investigate the relationships between these mutational changes and differences in gene expression.

**High salinity tolerance downstream in the gentle river and the genetic basis of local adaptation.** In the gentle river, *S. reiniana* inhabits both fresh and brackish water around estuaries, despite the latter being quite a harsh environment, with widely fluctuating tidal levels and daily changes in salinity. The high tolerance to salinity observed in juveniles born from individuals inhabiting brackish water in the gentle river indicated that this population had genetically adapted to the local environment (Fig. 2A,B). While modification of the extracellular matrix, muscle structure, free amino acids, swelling of the cell, de-condensation of perinuclear chromatin, and the induction of stress responses are important functions for osmoregulation and the adaptation of marine mollusks to a hyposaline environment<sup>36,45–54</sup>, adaptation and the responses of freshwater mollusks to intertidal environments has been poorly studied.

Although no clear separation of population genetic structure was found between freshwater and brackish water populations (Fig. S2), 147 non-synonymous SNPs were detected as loci subjected to natural selection associated with salinity in each population (Table S11). These non-synonymous SNPs were enriched in genes associated with the regulation of signaling and response to stimuli (Table S12). They might have arisen during the process of adaptation to large differences in environmental stimuli, such as temperature, hypoxia, sediment, salinity, and tidal rhythm. One non-synonymous SNP was in Adenosine deaminase acting on RNA (*Adar*), which encodes an enzyme that edits adenosine to inosine in pre-mRNA. In *Drosophila melanogaster*, *Adar* is specifically expressed in neurons, and its substrate is primarily mRNAs encoding an array of voltage- and ligand-gated ion channels<sup>55</sup>. RNA editing by *Adar* is hypothesized to fine-tune the neurophysiological properties of animals and mutants and appears to be involved in locomotory behavior, regulation of circadian rhythm, and response to heat and oxidative stress<sup>55,56</sup>. Although the hypothesis that behavioral change based on the non-synonymous mutation contributes to adaptation to the severe environment is reasonable and intriguing, to further discussion, we need investigations including genetic experiments using model organisms.

We detected five genes (GTP-binding protein RAD, Sodium-dependent phosphate transport protein 2A, Heat shock protein 68, Heat shock protein 68 and SAP30-binding protein), which were differentially expressed between brackish and freshwater populations (Table S13). These differences might be the result of adaptation or responses to the intertidal environment, but their function is ambiguous in freshwater invertebrates, including *S. reiniana*. For example, heat shock protein 68 (*hsp68*) maintains homeostasis and prevents cellular degeneration in response to various stressors, and in Pacific oyster, is associated with stress responses to changes in salinity and heat shock<sup>47,50</sup>. The gene expression of *hsp68* increased in response to hyperosmotic, hypoosmotic and heat stress in the Pacific oyster, but our analysis showed lower expression of *hsp68* in populations from brackish environments (Fig. S6). Since all individuals used for the experiment were acclimated in the same conditions before RNA fixation, this result suggests that the difference in expression of *hsp68* between brackish and freshwater populations depends on genetic changes, probably local adaptation, rather than being a response to environmental stress.

In summary, our multifaceted results are consistent with the hypothesis that migration load owing to asymmetric gene flow along the water stream restricts the range expansion of river snails. High tolerance for salinity in downstream population in the Kiso River implies that evolutionary adaptation for local environment might be more easily achieved in the gentle river than in a steep river. To confirm the generality of such a process, we need to conduct similar population transcriptome analyses in several rivers in the future. To reveal the genetic basis of the observed changes in gene expression patterns among populations, we may need to conduct further in-depth analyses of organisms from the rivers.

## Materials and methods

**Relationship between river steepness and distribution limits.** The altitudinal distribution range was investigated in 12 rivers in Japan. Field surveys were conducted in nine rivers—the Kiso, Yodo, Sendai, Yahagi, Toyokawa, Miyagawa, Fuji, Inabe, and Hayakawa Rivers—between 2017 and 2020. Sampling sites in each river were selected every 5 km from the estuary to 50 km from the estuary, and sampling of adult snails was carried out for about 30 min at each site. Dense sampling, at sites less than 5 km apart from each other, were conducted around the lower altitudinal limit of the snails' distributions. For the rivers in which we found at least one individual of the species, the lowest site at which the adults were captured was defined as the lower range limit of the river. For three rivers—Gonokawa River, Ibi River, and Kako River—previously reported occurrence data were used to assess the lower range limits.

Since the downstream areas of large rivers flow through plains in Japan, the difference in steepness between rivers is small at a distance of 0–20 km from the estuary. In addition, the length of a couple of rivers that we studied is less than 40 km from the estuary. Thus, in the present study, elevations at 30 km horizontally away from the estuary were taken to estimate the “steepness” of each river (Table S14). All statistical analyses in the present study were conducted using R 4.0.0 (<https://www.r-project.org/>). The relationship between steepness and lower distribution limit in each river was analyzed using Spearman's correlation calculated using *cor* function.

**Sample collection for laboratory experiments.** In the present study, we selected two rivers, the Kiso and Sendai Rivers, which have different steepness and enough individuals to conduct experiments. Though Kiso and Sendai Rivers are relatively gentle in 9 rivers that we studied, Sendai River is one of the steepest rivers in which individuals of *S. reiniana* inhabit at a wide range of elevations. Thus, these rivers are referred to as the gentle river and the steep river, respectively. Adult individuals of *S. reiniana* were collected at the river edges along each river between June and August 2017. For both rivers, since the steepness is similar within 20 km of the estuary, salinity intrusion is caused up to around 20 km from the estuary. The sampling sites were located at a distance of 0–50 km from the estuary, and all sites were around 5 km apart from each other. Individuals were captured by hand. Adults used to examine the tolerance to salinity were collected from 10 and eight locations in the gentle and steep rivers, respectively. Samples used for transcriptomic population genetic analysis came from nine locations, which were 0.8, 3, 4.3, 11 and 20 m above sea level (asl) for the gentle river and 2.9, 6.3, 25, and 30 m asl for the steep river. Sampling locations are named as a combination of the steepness of the rivers (gentle: G, steep: S) and the altitude. For example, the location at 25 m asl in the gentle river and that at 30 m asl in the steep river were referred to as G25 and S30, respectively.

**Saline tolerance experiment and size measurement.** To remove environmental effects, saline tolerance was quantified by using newborn snail. Adults captured in the natural populations were maintained in a freshwater aquarium (15 × 20 × 15 cm) in the laboratory at room temperature (~25 °C), and juvenile snails (< 5 days old) were collected. The juvenile snails were put into 1.5-ml microtubes filled with 1 ml of 0%, 1%, 2%, or 3% saline water, and stably stored in the dark in the incubator for 5 h. After the 5 h exposure, juvenile snails were transferred to 1.5 ml microtubes filled with 1 ml freshwater to enhance recovering from salinity shock. Since in general a living individual can recover from salinity shock and come out of their shells within an hour after exposure to freshwater, we recorded the viability (dead or alive) and shell size (width) of each individual three hours later after exposure to freshwater. The relationship between the elevation of sampling locations and the survival rates were tested using generalized linear models (GLMs). Shell width was used as the covariate since body size are predicted to influence the survival rate. To identify non-linear changes of shell width, the relationship between the elevation and the shell size was analyzed using segmented regression using the R package “segmented” (<https://cran.r-project.org/web/packages/segmented/index.html>).

**RNA extraction.** All individuals from each river were transported to the laboratory, rinsed with freshwater, and kept at constant temperature in an incubator until their dissection around 24 h after rinsing. The columellar



muscle tissue was excised and immersed in RNAlater™ Stabilization Solution (preserved at  $-80^{\circ}\text{C}$  until RNA extraction). Total RNA was extracted using Maxwell® 16 LEV Plant RNA purification kits in accordance with the manufacturer's instructions. The concentration of each RNA extraction was confirmed using Qubit Fluorometer with Qubit RNA BR Assay kits (Invitrogen, Carlsbad, CA). In the gentle river, 10, 10, 9, 9, and 9 individuals at G0.8, G3, G4.3, G11, and G20 respectively, and in the steep river, 10 individuals at each location were subjected to RNA extraction and RNA-sequencing. Also, individuals for Iso-Seq sequencing using PacBio Sequel system were taken from 0.8 and 1.0 m asl in the gentle river and 2.9, 6.3, and 25 m asl in the steep river to take regionality of transcript repertoire into consideration.

#### 4.5. Reference dataset construction.

For RNA-seq, library preparation and 100 bp paired-end sequencing using the NovaSeq 6000 System (Illumina, San Diego, CA) were outsourced to Filgen (Nagoya, Japan). Sequence read quality was confirmed using *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and low-quality reads and bases ( $Q < 20$ ) were trimmed and discarded using *Trimmomatic*<sup>57</sup>. Filtered short reads were subjected to de novo assembly using *Trinity* v2.6.5 with default parameter<sup>58</sup>. To remove redundancy from de novo assembly contigs, the *CD-hit* program (*cd-hit-est*) was conducted with parameter of sequence identity threshold  $-c$  0.99<sup>59</sup>. Subsequently, *TransDecoder* was conducted to screen protein-coding sequencing, with default parameters<sup>58</sup>.

Iso-Seq was outsourced to the National Institute of Genetics to obtain full-length transcripts. The sequences of each individual were merged using *cd-hit-est* with  $-c$  0.99 and assigned to be reference sequences for reference-guided assembly. Filtered short reads from the RNA-seq were mapped to these reference sequences using *Tophat2* with default parameter<sup>60</sup>, and the reference-guided assembly was constructed from mapped reads using *cuffmerge* and *gffread* implemented in *Cufflinks*<sup>61</sup>. In this study, the contigs that were merged de novo and the reference-guides assembly using *cd-hit-est* with  $-c$  0.99 was defined as "reftigs" for the following analyses. The reftigs were annotated by searching homology with the protein dataset of *Crassostrea gigas* (oyster\_v9) using *BLASTX* program. We used only reftig sequences with homology to *C. gigas* proteins for the subsequent analyses ( $E$ -value  $< 10^{-4}$ ). Orthology comprehensiveness was assessed using *BUSCO* v3 (<https://busco.ezlab.org/>).

We identified SNPs in samples from each river in accordance with the best practice for calling variants in RNA-seq of Genome Analysis Tool Kit (*GATK*, <https://software.broadinstitute.org/gatk/documentation/article.php?id=3891>). For each individual, all reads after quality control were mapped to reftigs using *STAR*<sup>62</sup>, and duplicated reads were identified using *Picard* (<http://broadinstitute.github.io/picard>). The variant sites were detected using HaplotypeCaller with minimum confidence threshold of 20 and merged for each river using GenotypeGVCFs<sup>63</sup>. Then, variants were filtered using VariantFilteration with threshold of FisherStrand  $> 30$  and QualByDepth  $< 2$ . Loci that had alleles observed in at least 5% of individuals existed in all individuals within each river, or were bi-allelic, were retained using *VCFtools*<sup>64</sup>. To estimate the impact of each SNP on phenotypes, the SNPs were classified according to their positions in the mRNA: synonymous, non-synonymous, or UTR, using *snpEff*<sup>65</sup>. Classification was conducted according to the GFF file output from *TransDecoder*.

**Population genetic analysis.** To avoid the effect of artifacts or sequencing error, deviation from the Hardy–Weinberg equilibrium were calculated using the  $-hwe$  option implemented in *VCFtools* and were excluded from the dataset ( $q < 0.05$ ). Linkage disequilibrium was calculated using *Plink* v1.9<sup>66</sup> and one member of each tightly linked SNP pair was excluded from subsequent analyses ( $r^2 > 0.8$ ). The pairwise  $F_{ST}$  between populations within each river, mean of the allele number, and the observed and expected heterozygosity were calculated using *Arlequin*<sup>67</sup>. The existence of isolation-by-distance (IBD) was tested using the pairwise  $F_{ST}$  by using *mantel* function in *vegan* package (<https://cran.r-project.org/web/packages/vegan/index.html>). The extent of IBD within a river would reflect the migration rate of *S. reimiana* and the connectivity between populations, and thus we were able to infer differences in gene flow between rivers. Also, the degree of IBD between rivers was compared using GLM. In this analysis, we used only freshwater populations to compare IBD between the rivers.  $F_{ST}$  values per horizontal distance along a river were compared between rivers using a Wilcoxon rank sum test in *exactRankTests* package (<https://cran.r-project.org/web/packages/exactRankTests/index.html>). Gene flow within each river was estimated using a Bayesian method implemented in *BA3-SNPs*<sup>68,69</sup>. Using this method, we were able to estimate the extent and direction of gene flow in the past few generations using a large SNP dataset. We estimated gene flow using the Bayesian estimation and sampled every 1,000th iteration from 10,000,000 iterations after a burn-in of 1,000,000 steps. Asymmetric migration was tested using one-sample t-tests for the ratio of downward to upward estimated migrations between locations within each river. For the assignment test, we used the *sNMF* function implemented in *LEA* package<sup>70</sup> in R, which is fast and efficient compared with conventional methods<sup>71</sup>. The population structure of each river was estimated using *sNMF* with 50 repetitions, using neutral loci, after removing loci that deviated from HWE and linkage disequilibrium. In the gentle river, there were 4,544 SNPs, and in the steep river, there were 6,142 SNPs).

Loci subjected to natural selection were detected using two different methods using loci detected in at least 90% of individuals within each river (gentle: 25,134 SNPs, steep: 24,665 SNPs). First, LFMM function implemented in *LEA* package were used to screen SNPs whose frequencies correlated with elevation. The number of latent factors was determined using the genetic structure identified by the *sNMF* function in *LEA* package. We ran LFMM for 10,000 iterations with 5,000 burn-in iterations and regarded SNPs that had a false discovery rate (FDR) of  $< 0.05$  as calculated by the *p.adjust* function in R as outliers associated with elevation. Second,  $F_{ST}$  outlier tests was conducted using a Bayesian method with *BayeScan* 2.1<sup>72</sup>. *BayeScan* with a prior odds value of 10 was run for 5,000 iterations with 500,000 burn-in and identified SNPs that had an FDR of  $< 0.05$ . We detected genes related to habitat expansion into brackish water in the Kiso River using LFMM and the salinity measured by refractometer at each elevation. GO terms were linked to each reftig using homology with the top-hit to a

*Drosophila melanogaster* protein dataset (FlyBase r6.36, BLASTP E-value < 10<sup>-4</sup>), and GO enrichment analysis was performed using Panther<sup>73</sup>.

**Expression analysis.** To examine the degree of local adaptation produced by transcriptional regulation, we quantified the expression levels of the transcripts in each individual and investigated the differences within and between the river under constant conditions. The transcript abundance in reftigs was quantified using RSEM<sup>74</sup> with bowtie2<sup>75</sup> as the alignment program. First, PCA based on fragments per kilobase of exon per million mapped reads (FPKM) values of all reftigs that has homology with *C. gigas* protein dataset was conducted using the *prcomp* function in R, and the pattern of gene expression (PC values) within rivers was analyzed by ANOVA using car package in R. Genes responsible for the habitat expansion into brackish water were explored by comparing expression levels between brackish water and freshwater populations in the gentle river. We tested the differences of expression levels after trimmed mean M-value (TMM) normalization between brackish water and freshwater using edgeR and TCC package with FDR of < 0.05<sup>76,77</sup>. To investigate the relationships of the local environment in the population and adaptation with transcriptional regulation, we estimated Pearson's correlation between the elevation and TMM normalized expression values of each transcript within each river.

Received: 26 July 2021; Accepted: 10 January 2022

Published online: 25 January 2022

## References

- Hoffmann, A. A. & Willi, Y. Detecting genetic responses to environmental change. *Nat. Rev. Genet.* **9**, 421–432 (2008).
- Endler, J. A. *Geographic Variation, Speciation and Clines* (Princeton, 1977).
- Huey, R. B. Rapid evolution of a geographic cline in size in an introduced fly. *Science*. **287**, 308–309 (2000).
- Bridle, J. R. & Vines, T. H. Limits to evolution at range margins: when and why does adaptation fail?. *Trends Ecol. Evol.* **22**, 140–147 (2007).
- Holt, R. D. & Gomulkiewicz, R. How does immigration influence local adaptation? A reexamination of a familiar paradigm. *Am. Nat.* **149**, 563–572 (1997).
- Ronce, O. & Kirkpatrick, M. When sources become sinks: Migrational meltdown in heterogeneous habitats. *Evolution* **55**, 1520–1531 (2001).
- Bridle, J. R., Gavaz, S. & Kennington, W. J. Testing limits to adaptation along altitudinal gradients in rainforest *Drosophila*. *Proc. R. Soc. B Biol. Sci.* **276**, 1507–1515 (2009).
- Bridle, J. R., Polechová, J., Kawata, M. & Butlin, R. K. Why is adaptation prevented at ecological margins? New insights from individual-based simulations. *Ecol. Lett.* **13**, 485–494 (2010).
- Holt, R. D. & Keitt, T. H. Alternative causes for range limits: A metapopulation perspective. *Ecol. Lett.* **3**, 41–47 (2000).
- Takahashi, Y. *et al.* Lack of genetic variation prevents adaptation at the geographic range margin in a damselfly. *Mol. Ecol.* **25**, 4450–4460 (2016).
- Arnaud-Haond, S. *et al.* Genetic structure at range edge: Low diversity and high inbreeding in Southeast Asian mangrove (*Avicennia marina*) populations. *Mol. Ecol.* **15**, 3515–3525 (2006).
- Pujol, B. & Pannell, J. R. Reduced responses to selection after species range expansion. *Science* **321**, 96 (2008).
- Cahill, A. E. & Levinton, J. S. Genetic differentiation and reduced genetic diversity at the northern range edge of two species with different dispersal modes. *Mol. Ecol.* **25**, 515–526 (2016).
- Bachmann, J. C., van Rensburg, A. J., Cortazar-Chinarro, M., Laurila, A. & Van Buskirk, J. Gene flow limits adaptation along steep environmental gradients. *Am. Nat.* **195**, E67–E86 (2020).
- Polechová, J. & Barton, N. H. Limits to adaptation along environmental gradients. *Proc. Natl Acad. Sci. U. S. A.* **112**, 6401–6406 (2015).
- Sexton, J. P., Strauss, S. Y. & Rice, K. J. Gene flow increases fitness at the warm edge of a species' range. *Proc. Natl Acad. Sci. U. S. A.* **108**, 11704–11709 (2011).
- Angert, A. L., Bontrager, M. G. & Aringgren, J. What do we really know about adaptation at range edges?. *Annu. Rev. Ecol. Evol. Syst.* **51**, 341–361 (2020).
- Ciborowski, J. J. H. Downstream and lateral transport of nymphs of two mayfly species (Ephemeroptera). *Can. J. Fish. Aquat. Sci.* **40**, 2025–2029 (1983).
- Bilton, D. T., Freeland, J. R. & Okamura, B. Dispersal in freshwater invertebrates. *Annu. Rev. Ecol. Syst.* **32**, 159–181 (2001).
- Markwith, S. H. & Scanlon, M. J. Multiscale analysis of *Hymenocallis coronaria* (Amaryllidaceae) genetic diversity, genetic structure, and gene movement under the influence of unidirectional stream flow. *Am. J. Bot.* **94**, 151–160 (2007).
- Congdon, B. C. Unidirectional gene flow and maintenance of genetic diversity in mosquitofish *Gambusia holbrooki* (Teleostei: Poeciliidae). *Copeia* **1995**, 162 (1995).
- Schaefer, J. Riffles as barriers to interpool movement by three cyprinids (*Notropis boops*, *Camptostoma anomalum* and *Cyprinella venusta*). *Freshw. Biol.* **46**, 379–388 (2001).
- Moore, J. S., Gow, J. L., Taylor, E. B. & Hendry, A. P. Quantifying the constraining influence of gene flow on adaptive divergence in the lake-stream threespine stickleback system. *Evolution* **61**, 2015–2026 (2007).
- Urabe, M. Diel change of activity and movement on natural river beds in *Semisulcospira reiniana*. *VENUS* **57**, 17–27 (1998).
- Hastie, L. C., Boon, P. J., Young, M. R. & Way, S. The effects of a major flood on an endangered freshwater mussel population. *Biol. Conserv.* **98**, 107–115 (2001).
- Alp, M., Keller, I., Westram, A. M. & Robinson, C. T. How river structure and biological traits influence gene flow: A population genetic study of two stream invertebrates with differing dispersal abilities. *Freshw. Biol.* **57**, 969–981 (2012).
- Terui, A. *et al.* Asymmetric dispersal structures a riverine metapopulation of the freshwater pearl mussel *Margaritifera laevis*. *Ecol. Evol.* **4**, 3004–3014 (2014).
- Holomuzki, J. R. & Biggs, B. J. F. Distributional responses to flow disturbance by a stream-dwelling snail. *Oikos* **87**, 36 (1999).
- Urabe, M. Phenotypic modulation by the substratum of shell sculpture in *Semisulcospira reiniana* (Prosobranchia: Pleuroceridae). *J. Molluscan Stud.* **66**, 53–60 (2000).
- Gu, Q. H., Husemann, M., Ding, B., Luo, Z. & Xiong, B. X. Population genetic structure of *Bellamya aeruginosa* (Mollusca: Gastropoda: Viviparidae) in China: Weak divergence across large geographic distances. *Ecol. Evol.* **5**, 4906–4919 (2015).
- Davis, C. D., Epps, C. W., Flitcroft, R. L. & Banks, M. A. Refining and defining riverscape genetics: How rivers influence population genetic structure. *Wiley Interdiscip. Rev. Water* **5**(2), e1269 (2018).

32. De Wit, P. & Palumbi, S. R. Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Mol. Ecol.* **22**, 2884–2897 (2013).
33. Sun, Y.-B. *et al.* Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. *Proc. Natl. Acad. Sci.* **115**, 201813593 (2018).
34. Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T. & Christie, M. R. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Mol. Ecol.* **27**, 4041–4051 (2018).
35. De Wit, P. *et al.* The simple fool's guide to population genomics via RNA-Seq: An introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* **12**, 1058–1067 (2012).
36. Yokomizo, T. & Takahashi, Y. Changes in transcriptomic response to salinity stress induce the brackish water adaptation in a freshwater snail. *Sci. Rep.* **10**, 1–9 (2020).
37. Kottler, E. J., Dickman, E. E., Sexton, J. P., Emery, N. C., & Franks, S. J. Draining the swamping hypothesis: Little evidence that gene flow reduces fitness at range edges. *Trends Ecol. Evol.* 1–12 (2021).
38. Moore, J. S. & Hendry, A. P. Can gene flow have negative demographic consequences? Mixed evidence from stream threespine stickleback. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 1533–1542 (2009).
39. Ingvarsson, P. K. Restoration of genetic variation lost - The genetic rescue hypothesis. *Trends Ecol. Evol.* **16**, 62–63 (2001).
40. Shimada, K. & Urabe, M. Drift and upstream movement of *Semisulcospira libertina* (Caenogastropoda: Pleuroceridae) in a natural stream. *Virus* **63**, 49–59 (2004).
41. Nyitrai, L., Goodwin, E. B. & Szent-Gyorgyi, A. G. Complete primary structure of a scallop striated muscle myosin heavy chain: Sequence comparison with other heavy chains reveals regions that might be critical for regulation. *J. Biol. Chem.* **266**, 18469–18476 (1991).
42. Ponder, W. F., Lindberg, D. R. & Ponder, J. M. *Shell, Body, and Muscles* (CRC Press, Taylor and Francis Group, Boca Raton, 2019).
43. Lesoway, M. P., Abouheif, E. & Collin, R. Comparative transcriptomics of alternative developmental phenotypes in a marine gastropod. *J. Exp. Zool. Part B Mol. Dev. Evol.* **326**, 151–167 (2016).
44. Sexton, J. P., McIntyre, P. J., Angert, A. L. & Rice, K. J. Evolution and ecology of species range limits. *Ann. Rev. Ecol. Evol. Syst.* **40**, 415–436 (2009).
45. Berger, V. J. & Kharazova, A. D. Mechanisms of salinity adaptations in marine molluscs. *Hydrobiologia* **355**, 115–126 (1997).
46. Rivera-Ingraham, G. A. & Lignot, J. H. Osmoregulation, bioenergetics and oxidative stress in coastal marine invertebrates: Raising the questions for future research. *J. Exp. Biol.* **220**, 1749–1760 (2017).
47. Jo, P. G., Choi, Y. K., An, K. W. & Choi, C. Y. Osmoregulation and mRNA expression of a heat shock protein 68 and glucose-regulated protein 78 in the Pacific oyster *Crassostrea gigas* in response to salinity changes. *J. Aquac.* **20**, 205–211 (2007).
48. Eierman, L. E. & Hare, M. P. Transcriptomic analysis of candidate osmoregulatory genes in the eastern oyster *Crassostrea virginica*. *BMC Genomics* **15**, 1–15 (2014).
49. X. Zhao, H. Yu, L. Kong, Q. Li, Transcriptomic responses to salinity stress in the Pacific oyster *Crassostrea gigas*. *PLoS ONE* **7** (2012).
50. Zhang, Y. *et al.* Proteomic basis of stress responses in the gills of the Pacific oyster *Crassostrea gigas*. *J. Proteome Res.* **14**, 304–317 (2015).
51. Veiga, M. P. T., Gutierrez, S. M. M., Castellano, G. C. & Freire, C. A. Tolerance of high and low salinity in the intertidal gastropod *Stramonita brasiliensis* (Muricidae): Behaviour and maintenance of tissue water content. *J. Molluscan Stud.* **82**, 154–160 (2016).
52. Muraeva, O. A., Maltseva, A. L., Mikhailova, N. A. & Granovitch, A. I. Mechanisms of adaptation to salinity stress in marine gastropods *Littorina saxatilis*: a proteomic analysis. *Cell Tissue Biol.* **10**, 160–169 (2016).
53. Muraeva, O., Maltseva, A., Varfolomeeva, M., Mikhailova, N. & Granovitch, A. Mild osmotic stress in intertidal gastropods *Littorina saxatilis* and *Littorina obtusata* (Mollusca: Caenogastropoda): A proteomic analysis. *Biol. Commun.* **62**, 202–213 (2017).
54. Maynard, A., Bible, J. M., Pespeni, M. H., Sanford, E. & Evans, T. G. Transcriptomic responses to extreme low salinity among locally adapted populations of Olympia oyster (*Ostrea lurida*). *Mol. Ecol.* **27**, 4225–4240 (2018).
55. Ma, E., Gu, X. Q., Wu, X., Xu, T. & Haddad, G. G. Mutation in pre-mRNA adenosine deaminase markedly attenuates neuronal tolerance to O<sub>2</sub> deprivation in *Drosophila melanogaster*. *J. Clin. Invest.* **107**, 685–693 (2001).
56. Jepson, J. E. C. *et al.* Engineered alterations in RNA editing modulate complex behavior in *Drosophila*: Regulatory diversity of adenosine deaminase acting on RNA (ADAR) targets. *J. Biol. Chem.* **286**, 8325–8337 (2011).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
59. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
60. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**(4), 1–13 (2013).
61. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
62. Dobin, A. *et al.* STAR: ULTRAFast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).
64. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
65. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
66. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16 (2015).
67. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
68. Wilson, G. A. & Rannala, B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177–1191 (2003).
69. Mussmann, S. M., Douglas, M. R., Chafin, T. K. & Douglas, M. E. BA3-SNPs: contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods Ecol. Evol.* **10**, 1808–1813 (2019).
70. Frichot, E. & François, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
71. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).
72. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993 (2008).
73. Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).
74. Parrish, N., Hormozdiari, F., & Eskin, E. Assembly of non-unique insertion content using next-generation sequencing. *BMC Bioinformatics* **12**, S3 (2011).
75. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

76. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
77. Sun, J., Nishiyama, T., Shimizu, K. & Kadota, K. TCC: An R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* **14**(1), 1–14 (2013).

### Acknowledgements

This work was supported by JSPS KAKENHI (Grant No. 17H03729 and 20K06822) with grants from Fujiwara Natural History Foundation and Asahi Glass Foundation. Iso-Seq was supported by JSPS KAKENHI Grant No. 16H06279 (PAGS). Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

### Author contributions

K.T. and Y.T. designed research; K.T., K.Y., S.O. and Y.T. performed research; K.T., K.Y., S.O. and Y.T. analyzed data; and K.T. and Y.T. wrote the paper.

### Funding

This work was supported by JSPS KAKENHI (Grant No. 17H03729 and 20K06822) with grants from Fujiwara Natural History Foundation and Asahi Glass Foundation.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-05248-1>.

**Correspondence** and requests for materials should be addressed to K.T. or Y.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022