

RESEARCH ARTICLE

Open Access



# Genome-wide signatures of local adaptation among seven stoneflies species along a nationwide latitudinal gradient in Japan

Maribet Gamboa \* and Kozo Watanabe

## Abstract

**Background:** Environmental heterogeneity continuously produces a selective pressure that results in genomic variation among organisms; understanding this relationship remains a challenge in evolutionary biology. Here, we evaluated the degree of genome–environmental association of seven stonefly species across a wide geographic area in Japan and additionally identified putative environmental drivers and their effect on co-existing multiple stonefly species. Double-digest restriction-associated DNA (ddRAD) libraries were independently sequenced for 219 individuals from 23 sites across four geographical regions along a nationwide latitudinal gradient in Japan.

**Results:** A total of 4251 candidate single nucleotide polymorphisms (SNPs) strongly associated with local adaptation were discovered using Latent mixed models; of these, 294 SNPs showed strong correlation with environmental variables, specifically precipitation and altitude, using distance-based redundancy analysis. Genome–genome comparison among the seven species revealed a high sequence similarity of candidate SNPs within a geographical region, suggesting the occurrence of a parallel evolution process.

**Conclusions:** Our results revealed genomic signatures of local adaptation and their influence on multiple, co-occurring species. These results can be potentially applied for future studies on river management and climatic stressor impacts.

**Keywords:** Adaptation, Environmental associations, Landscape genetics, ddRAD

## Background

Understanding the influence of natural selection on genomic variation in natural populations is one of the primary goals for addressing questions regarding adaptive evolution. Principally, divergence selection derived from natural selection may result in allele frequency shifts at loci under selective pressures, such as heterogeneous climatic factors, to maximize fitness in local environment [1]. Therefore, the capacities of organisms to respond to changing selective pressures (such as climate) may widely vary due to their potential pending genetic adaptation and the rate at which new genomic variations arise [2, 3].

Studies investigating the genomic basis of adaptation often compare genomic information of populations along heterogeneous environments [2]. First, the genome was screened for outlier loci by determining associations

between allele distributions and environmental variables that were presumed to be selection drivers [4–6]. Then, correlations among environmental variables combined with population genomic approaches (i.e., fixation index and inbreeding coefficient) may complement each other in terms of power and accuracy [7]. However, discerning the relative contribution of environmental conditions in shaping genetic diversity remains a challenge. Identifying factors such as natural selection, genetic drift, and dispersion that modify patterns of genomic variation remains difficult [8].

Genome–environmental association analysis along environmental gradients [9] provides an opportunity to examine the influence of environmental conditions in shaping genetic variation of natural populations. Environmental information, as a combination of landscape effects [10] and environmental clines [11], provides stronger evidence for examining adaptive divergence. This type of analysis considers the effect of population structure and geographical scale and expects that genome signatures of

\* Correspondence: [gamboa@cee.ehime-u.ac.jp](mailto:gamboa@cee.ehime-u.ac.jp)  
Department of Civil and Environmental Engineering, Ehime University,  
Matsuyama 790-0871, Japan



selection become more pronounced with increasing geographical scale (isolation by distance) owing to larger environmental gradient differences [12]. Genetic response to environmental gradients has been evidenced in a wide range of taxa, e.g. including plants [13, 14], fish [15], and invertebrates [16, 17]. Overall, reported studies have highlighted that the selective pressure resulting from environmental changes can be explained by the evolutionary response of population along these gradients. However, these studies investigating this association have been based on a single species [14, 15, 18, 19], thus neglecting the potential effect of this association on multiple species that co-occur in the same habitat.

The occurrence of multiple species in a community remains another challenging aspect of determining the influence of environment on genomic variation. Adaptive genomic variation is strongly influenced by the complex dynamics between environmental and community effects [20]. Due to evolutionary dynamics, the assessment of genomic variation in organisms that co-occur with other species may not match that predicted using single species approaches [21] because of the effect of competition, predation, and co-evolution. For example, using multiple-species observations, it is possible to observe genome–genome interactions (e.g., hybridization and mutualistic), genome–environment interactions (e.g., parallel adaptation, recombination rate, and demography), and genome–genome–environment interactions (e.g., interactions among species) [20] between organisms. Data on the impact of environmental change on evolutionary processes at the community and ecosystem scale is lacking [20]. In order to gain a better understanding of adaptive genomic variation, genetic consequences of co-existing species in a local community should be studied [22].

Aquatic insects are ideal organisms for observing genome–environment interactions. Stoneflies (Plecoptera) are aquatic insects that exhibit limited airborne dispersal within stream corridors and their dispersal is mainly dominated by in-stream drift [23]. Stoneflies are considerably more sensitive to environmental changes, such as low oxygen concentration and high water temperature, than other aquatic insects [24]. Until now, studies evaluating genome–environmental association in aquatic insects have often used mitochondrial DNA sequences with a particular molecular marker (Sanger sequencing) [25], microsatellite [26], or Amplified Fragment Length Polymorphism makers [27]; however, NGS data has been scarcely used except for in few studies such as those on damselfly [16] and chironomidae [17].

In this study, we examined the genome-wide signatures of adaptive divergences in seven species of stoneflies using a double-digest restriction-site-associated DNA protocol (ddRAD). Several techniques have been applied for the detection of adaptive genes, such as direct sequencing of

genomic regions of interest [28], genome re-sequencing [18], genotyping by sequencing (GBS; [14]), transcriptomics [19], and restriction site associated DNA sequencing (RAD-Seq; [15]). Among these techniques, RAD-Seq allows the analysis of multiple organisms at the same time [29] and the genomic analysis of organisms without a reference genome [30]. In addition, it allows for the identification of thousands of genetic markers randomly distributed across the genome [31], which further increases the chances of detecting loci under natural selection. Therefore, RAD-Seq has been employed for understanding the pattern of genetic structure of populations across environmental heterogeneity (e.g. [32]).

Here, we performed population and landscape genomic studies using ddRAD so as to screen imprints of selection on a national-wide scale in Japan archipelago. We primarily aimed to evaluate the relative strength of selection in influencing the distribution of genomic variation at different spatial scales, to identify putative environmental drivers, and to observe their effects on the community. We hypothesized that organisms living under similar meteorological and hydrological conditions would experience similar genomic evolution in their adaptive genes, and therefore, similar patterns of adaptive divergence appear among co-inhabiting species. For this, we collected samples along a nationwide latitudinal gradient in Japan with varying meteorological and hydrological conditions. We applied an individual-based Latent Factor Mixed Model (LFMM) and a multivariate redundancy analysis (RDA) to identify putative single nucleotide polymorphisms (SNPs) under selection and the main environmental population differentiation driver. The resulting species-specific association of SNPs with environmental variables were compared to determine the degree of nucleotide changes (i.e., nucleotide substitution rate) among the seven species so as to possibly explain their local adaptation. Although the use of SNPs has been the standard method to determine loci putatively under selection, integrating environmental gradients and multiple species can be reliably used to identify such loci and selective agents underlying the observed patterns either in a population or community.

## Methods

### Sample collection

We ad hoc collected a total of 47 species. Of these, eight species that commonly occurred throughout four regions in Japan were shortlisted to test the potential effect of climatic variation in these regions. The eight species of stream stoneflies selected (*Perlodini incertae*, *Haploperla japonica*, *Nemoura ovocercia*, *Rhabdiopteryx japonica*, *Obipteryx femoralis*, *Isoperla nipponica*, *Amphinemura longispina*, and *Stavsolus japonicus*) have different biological requirements, such as feeding

behavior, and habitat preferences. An average of 31 stonefly nymphs per species were collected in January–March 2014 along a latitude cline in Japan using D-flame nets (mesh size = 250 μm). Collected specimens were identified using the taxonomic key of Japanese aquatic insects [33].

Sampling of stonefly nymphs was carried out at 23 sites across four regions with different climatic conditions, including: Matsuyama (6 sites), Gifu (4 sites), Sendai (6 sites), and Sapporo (7 sites) (Fig. 1, Additional file 1: Tables S1 and S2). Sampling sites in Matsuyama located in Ehime Prefecture in Shikoku Island, South of Japan (geographical distance to other regions ranged 404–1295 km) has a mean temperature of 17 °C and an altitude range of 154–277 m above sea level (masl). The Gifu region, situated in Gifu prefecture at the central area of Japan (geographical distance to other regions range 374–959 km) has a mean temperature of 15 °C and an altitude range of 251–720 masl. The Sendai region, located in Miyagi prefecture in the northeast area of Japan (geographical distance to other regions range 374–895 km) has a mean temperature of 12 °C and an altitude range of 261–394 masl. The sampling sites in the Sapporo region, in Hokkaido prefecture positioned at the Northern most part of Japan (geographical distance to other regions range 418–1295 km) has a mean temperature of 7 °C and an altitude range of 95–298 masl.

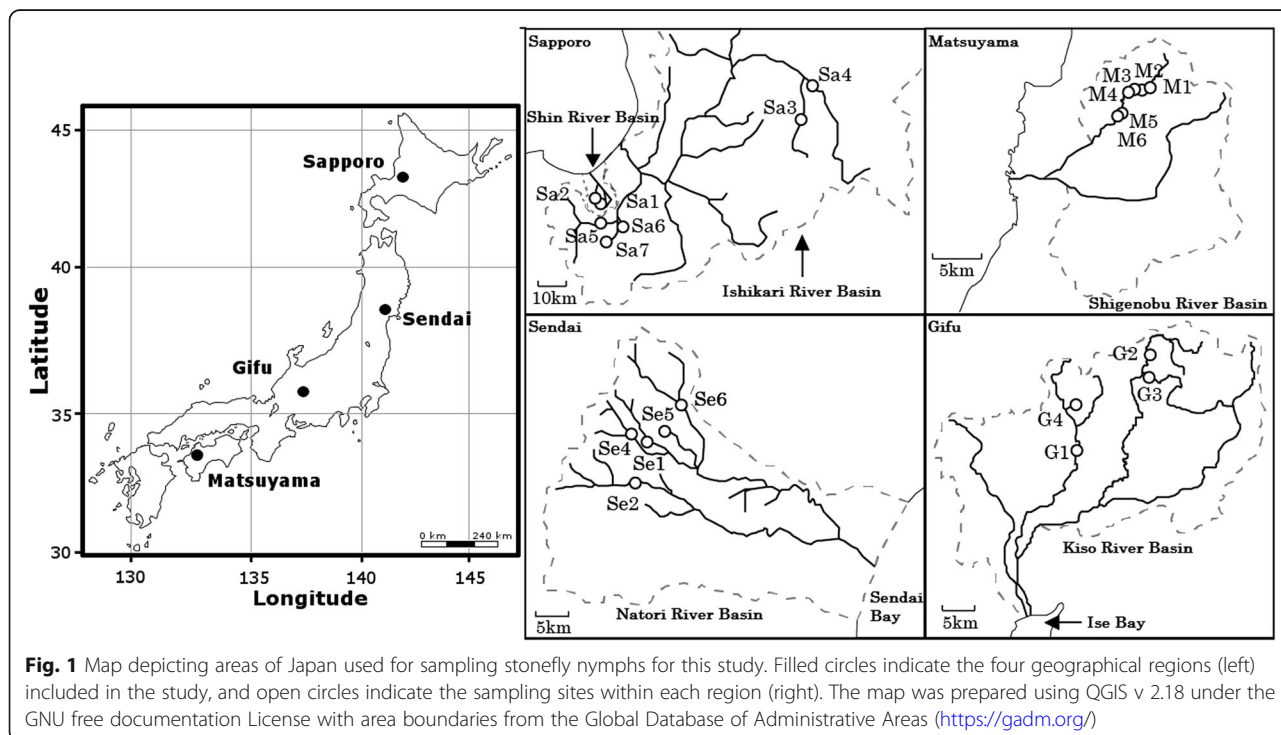
Data on precipitation (annual mean value of the amount of rainfall that occurs over a unit area; unit mm), river water level (annual mean value of the level of the water of

a river above a horizontal surface used as zero point of measurement of water level; unit m), discharge (annual mean volume of water flowing through a river channel; unit m<sup>3</sup>/s), water temperature (annual mean temperature of the water; unit °C), air temperature (annual mean temperature of atmospheric air; unit: °C), and depth of snow cover (annual mean value of the total amount of snow that accumulates on the ground; unit cm) for the sampling sites (Additional file 1: Table S1) were obtained from the Ministry of Land, Infrastructure, Transport and Tourism of Japan (<http://www1.river.go.jp/>) and the NASA Giovanni website (<https://giovanni.sci.gsfc.nasa.gov/giovanni/>) within standard NASA estimate algorithms. Environmental data for each sampling site was collected from a different meteorological station. The locations of the meteorological stations were situated within 2.5–19.8 km in Matsuyama, 7–52 km in Gifu, 1.3–16 km in Sendai, and 17–103 km in Sapporo regions.

Geographic distances between sampling sites were calculated from the transformation of GPS coordinates [34] using great-circle distances calculated for the Vincenty Ellipsoid in geosphere package of R v.3.3 [35] and used for principal components analysis (PCA) using the stats package in R v.3.3 (R Core Team, <https://www.r-project.org/>). The results obtained from PCA were used to statistically separate geographical regions.

**ddRAD library preparation and sequencing**

A total of 270 stoneflies belonging to eight species (average = 31 individuals per species; range = 20–52



individuals per species) were used for RAD-Seq. All individuals were washed thrice with ethanol to eliminate any contamination. Genomic DNA was extracted from each individual using DNeasy tissue kit (Qiagen GmbH, Hilden, Germany) following manufacturer's instructions. The whole body of the specimen was fully submerged in lysis buffer (Qiagen GmbH, Hilden, Germany) and incubated at 56 °C overnight. This DNA extraction method involved the use of the whole body without any destruction of the specimen, thus reducing the probability of contamination (gut content contamination). ddRAD libraries were constructed as previously described [29] with slight modifications. Briefly, 100 ng genomic DNA per sample was digested with restriction endonucleases, *SbfI* (NEB) and *HaeIII* (Takara) at 37 °C for 3 h. Restriction fragments of each sample were ligated to a unique 6-mer adaptor. Ninety samples with equimolar DNA concentrations were combined and electrophoresed on agarose gels. Subsequently, 300–500 bp fragments were excised and purified using MinElute Gel Extraction Kit (Qiagen).

Purified DNA samples were PCR amplified using two sequential PCRs. The first PCR was carried out in a 20  $\mu$ l reaction, consisting of 12  $\mu$ l of Phusion High-Fidelity PCR Master Mix with HF buffer (NEB), 2  $\mu$ l DNA template, 2  $\mu$ l of each forward and reverse primers specific to Illumina adaptors, 1  $\mu$ l of Betaine (Sigma), 1  $\mu$ l of DMSO (NEB), and 2  $\mu$ l of water, using the following conditions: initial denaturation at 98 °C for 1 min, followed by 18 cycles of 98 °C for 10 s, 62 °C for 30 s and 72 °C for 30 s, and a final extension at 72 °C for 10 min. The second PCR was carried out to eliminate non-specific sequences using 4  $\mu$ l of Phusion HF buffer (NEB), 2  $\mu$ l of dNTPs (Promega), and 2  $\mu$ l of each forward and reverse Illumina sequencing primers, using the following conditions: 98 °C for 3 min, 60 °C for 2 min, and 72 °C for 12 min.

Five PCR replicates were pooled to form a single library. Each library was quantified using KAPA library quantification kit (Roche). Three DNA libraries each comprising 90 individuals were sequenced per lane using HiSeq 2500 Illumina sequencer (paired end, 2  $\times$  100 bp) at the Beijing Genomics Institute, China. One species (*Perlodini incertae*, 51 individuals) was removed from the analysis because of its low read recovery (< 8,000,000 total reads). Therefore, we used seven species (219 individuals) in the analysis.

#### Data processing

Forward and reverse raw data (789,587,765 reads) were screened for tags and adapters using CUTADAPT [36] (Additional file 1: Figure S1). A phred-type quality score of Q20 was used to trim sections of the reads using FASTX toolkit v.0.0.13 ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Reads shorter than 50 bp were excluded.

The number of reads was reduced to 675,870,356 (86% of the total reads).

RAD sequences were processed using a pipeline in STACKS v 2.0 that incorporates a maximum likelihood statistical model to identify sequence polymorphisms [37]. Sequences were demultiplexed and quality filtered using the process\_RADtags program. Reads shorter than 50 bp with a quality read score of < 20 or with ambiguous barcodes (i.e. erroneous barcodes) were removed. The number of reads was reduced to 675,100,758 (85% of the total reads).

De novo assembly of short paired-end reads was performed independently using three programs in the STACKS software package. The ustacks program uses a set of reads as input and aligns them into exactly matching (i.e. 100% nucleotide sequence identity) putative alleles. Different parameter combinations were evaluated (Additional file 1: Appendix S1, Tables S3 and S4). The analysis revealed similar genomic comparisons [i.e. observed and expected homozygotes and heterozygotes, fixation index ( $F_{st}$ , Additional file 1: Table S5), and inbreeding coefficient ( $F_{is}$ )] but different number of loci (Additional file 1: Table S4). For the following analysis, parameter combinations with a high number of polymorphic loci with a depth read coverage ranging from 30X to 60X were used. We applied the following conditions as previously recommended by [38, 39]: minimum read depth to create a putative allele group = 3, number of mismatches allowed between putative alleles within individuals = 2, and number of mismatches allowed between putative alleles within a putative allele group = 2. Reads that failed to meet these criteria were removed. This process reduced the number of reads to 499,574,561 (63% of the total reads).

The output file from ustacks was a catalog of putative alleles per individual. This catalog was used as an input file for the cstacks program implemented in STACKS. This program merges putative alleles together and creates a set of consensus putative loci. The default parameters, namely permitting mismatches between putative alleles of two nucleotides and removing putative alleles considered unique alleles were applied. This process reduced the number of reads to 310,722,904 (41% of the total reads).

The output list of consensus putative loci was used as an input file for the sstacks program implemented in STACKS. This program matches each individual putative locus list against the consensus putative loci for each species. The default parameters were applied, following the base matching on the alignment position, which allowed the detection of sequences errors. Consensus putative loci for each species that were found among all individuals were retained. The tsv2 and gstacks programs were implemented in STACKS to assemble and merge paired-end reads. These processes reduced the

number of reads to 100,980,315 (16% of the total reads) among 50,514 loci (Additional file 1: Figure S1).

Following de novo assembly, we performed an additional data-filtering step per species using the populations program implemented in STACKS. Consensus putative loci were excluded from the study using the following criteria: 1) present in <80% of individuals; 2) minor allele frequency <0.05; 3) <30X – >60X read coverage; and 4) >one SNP per locus. We examined whether a high read depth affected our data in the filtering as well as the downstream analysis (Additional file 1: Appendix S1) as previously reported [40]. A cut-off filtering process was employed using vfilter by examining whether the high depth coverage reads affected our data (<https://github.com/bcbio/bcbio-nextgen>) following a previous study's suggestion [40] (the output removed 0.001% of the total reads). The number of reads was reduced to 64,502,837 (8% of the total reads, Additional file 1: Table S6) among 24,318 loci (Additional file 1: Table S7 and Figure S1).

The observed ( $H_o$ ) and expected heterozygosity ( $H_e$ ),  $F_{is}$ , and  $F_{st}$  were estimated per region and per species using popGenome package in R v3.3 [41]. The significance of pairwise  $F_{st}$  values was tested using 10,000 permutations.

Population genetic structure was explored using fastSTRUCTURE [42] without a priori information of the geographical origin of each sample. Analyses were performed using the admixture model with correlated allele frequencies and a burn-in period of 200,000 MCMC interactions, followed by 300,000 interactions per run using the logistic parameter. We selected a number of K (putative populations) ranging from 1 to 12 with all loci. Five replicate analyses were performed for each putative population (K). The number of cluster was inferred using the complexity model of fastSTRUCTURE. The results of 5 replicates of the selected K values were pooled into a single result using CLUMPAK web version (<http://clumpak.tau.ac.il/index.html>) and their consistency was examined following [43] recommendations using CLUMPP V. 1.1.2 [44] by the Greedy algorithm.

### Power analysis

To determine the power of RAD datasets to resolve population structure, we performed a power analysis using POWSIM v. 4.1 [45]. This program estimates the statistical power of genetic homogeneity in individual species. To reflect our sampling design, we set the number of subpopulations to four, with 5–10 samples per subpopulation for SNPs (based on the distribution of the individuals per sampling site; Additional file 1: Table S2). We set the effective population size of the subpopulations to 1000, 2000, and 3000, and we adjusted the generation time (t) to assess power at multiple  $F_{st}$  values (10 and 20 generations) as previously described [46].  $F_{ST}$

in Powsim assumes the independence of the subpopulations [45]. Power was expressed as the proportion of significant outcomes for 1000 replicates and a statistically significant test ( $p < 0.05$ ).

Additionally, the power of our SNP data to the environmental association was determined using Quanto v.1.2.4 [47]. We adjusted the parameters based on our sample size (Additional file 1: Table S5) and implemented independent individuals for the gene-environment interaction, a marginal effect, a population mean of 20 per region, and a marginal R squared model.

### Genome-environmental associations

The latent factor mixed model (LFMM; [48]) was used to identify local adaptations among stonefly populations. LFMM detects correlations between environmental variables and genetic variation through the estimation of latent factors and regression coefficient. For LFMM analysis, data on SNPs, population structure, and environmental variables and spatial eigenvectors among sampling sites (8–22 (mean, 14.4) sites per species; Additional file 1: Table S2) were used with the LEA package in R v.3.3 [49]. To ensure that the selected candidate SNPs were not due to an erroneous false positive detection, we employed Benjamin Hochberg false discovery rate (FDR) algorithm to partial LFMM results of randomly selected three locations among the regions and three species (i.e. *Rhabdiopteryx japonica*, *Obiopteryx femoralis*, *Isoperla nipponica*). We compared three FDR q-values calculated from three different expected q-values, 0.01, 0.05, and 0.10 (i.e., 0.01 represents a critical threshold above which 1% of loci are false positives) using a G-test implemented in the q value package in R v.3.3 [50]. The calculated FDR q-values were not significantly different from the expected ones for the three species (*Rhabdiopteryx japonica*,  $G = 2.0$ , d.f. = 2,  $P = 0.4$ ; *Obiopteryx femoralis*,  $G = 1.98$ , d.f. = 2,  $P = 0.41$ ; and *Isoperla nipponica*,  $G = 1.8$ , d.f. = 2,  $P = 0.39$ ), supporting the fact that there is no influence of false positive selection of candidate SNPs. Following the manual of LEA package, we applied a FDR of 0.10 to all species [49]. Multidimensional scaling (MDS) of all candidate SNPs generated using LFMM and their correlation with environmental variables was examined to determine the proportion of genomic variation affected by environment. Candidate SNPs were extracted using VCFtools v.0.1.14 [51] and used to calculate the allele frequencies (identity by state; IBS) at each site using Hamming distance matrix with PLINK v.1.9 [52]. The IBS matrix was used for Principal Coordinate Analysis (PCoA). The resulting eigenvalues of PCoA were used for the subsequent distance-based redundancy analysis (db-RDA; [53]). Statistical significance of db-RDA models was tested using ANOVA. The db-RDAs were conducted using capscale and ordistep

functions and by plotting MDS1 and MDS2 in the vegan package in R v3.0.2 (R Core Team; [54]). MDS analysis was also used to enable the identification of a SNP associated with the respective environmental variable; therefore, we examined correlations between environmental variables and individual candidate SNPs using MDS and db-RDA.

In order to determine if the genome variation of candidate SNPs were affected by isolation by distance, we conducted mantel test using the vegan package in R v.3.2.2 [54]. We compared matrices of pairwise geographic distances (refer methods: sample collection) and pairwise genetic distance (IBS) among individuals. We conducted separate analyses for each species with 10,000 permutations to obtain corrected *p*-values.

#### Adaptive genetic variation among species

To determine the degree of adaptive genetic variation in stonefly genomes, sequences of loci of seven stonefly species harboring candidate SNPs were aligned using bowtie2 v.2.3 [55]. Default parameters were used in end-to-end alignment that allowed up to two mismatches per 90 base reads. Subsequently, the best match sequences for each locus were identified. Additionally, these sequences under putative selection were used to compute mean nucleotide substitution rate within species (i.e., individuals within species) and within or between geographical regions (i.e., species per region) by pooling the seven species [41]. Mean nucleotide substitution rate was calculated using pairwise nucleotide substitution rate, where the expected value ranged from 0 (similar) to 1 (dissimilar).

We identified the gene function of the sequences of the candidate loci using web version of tBLASTx against the nucleotide databases of GenBank, EMBL, DDBJ, PDB, RefSeq, PDB, SwissProt, PIR, and PR in NCBI (<http://blast.ncbi.nlm.nih.gov>) under default parameters. Sequences with an *e*-value  $\leq 1e-5$  and  $\geq 90\%$  sequence similarity were used in Gene Ontology analysis using web service Uniprot database (<https://www.uniprot.org>). We excluded a possible contamination and an erroneous gene

function in our sequences by processing our DNA sequences against the entire NCBI database using DeconSeq web interface (<http://deconseq.sourceforge.net/>).

## Results

### Data filtering

RAD-Seq generated a total of 789,587,765 raw reads with an average of 263,195,922 reads per library, 1,802,711 reads per individual, and 3,527,060 per species per region. The length of raw reads ranged from 90 to 100 bp with an average length of 98 bp. In total, 113,717,409 (14%) reads with low quality, ambiguous barcode and short length were discarded. On an average, 3,606,439 reads per individual (range: 2,600,058–4,016,315 reads) with an average length of 98 bp (range = 90–100 bp) were successfully aligned generating 53,835 consensus loci (Additional file 1: Figure S1). After loci identification and SNP calling, a dataset of 50,154 loci representing 16% of the total reads was generated, which was equivalent to 222–4359 loci per species (Additional file 1: Figure S1). After final data filtering (i.e., removing consensus putative loci that were present in less than 80% of individuals), 24,318 SNP loci were identified with 202–2167 SNP loci per species (mean = 868 SNP loci per species, Additional file 1: Table S7) and an average 45-fold coverage per species (range = 30–60-fold).

The statistical power of our SNP dataset to population structure (POWSIM program) and to gene-environmental association (Quanto program) was evaluated. Both programs provided consistently high power to detect genetic differences (90% mean statistical significance) and detection of association (57% average chance of detection) among the populations.

### Population structure

Population statistics are summarized in Table 1. Fixation index ( $F_{st}$ ) and inbreeding coefficient ( $F_{is}$ ) within species varied between 0.043–0.123 and 0–0.032, respectively. Expected heterogeneity ( $H_e$ ) was an average of 0.33 for all species without spatial differentiation between regions.

**Table 1** Summary of sample size and molecular metrics. Where: *n*, number of individuals;  $H_o$ , average observed heterozygosity;  $H_e$ , average expected heterozygosity;  $F_{st}$ , mean genetic divergence of average pair geographical locality;  $F_{is}$ , inbreeding coefficient

| Species                       | <i>n</i> | Total reads | Total loci | $H_o$ | $H_e$ | $F_{st}$ | $F_{is}$ |
|-------------------------------|----------|-------------|------------|-------|-------|----------|----------|
| <i>Nemoura ovocercia</i>      | 20       | 2,532,235   | 10,448     | 0.71  | 0.3   | 0.106    | 0.032    |
| <i>Haploperla japonica</i>    | 32       | 495,817     | 1624       | 0.75  | 0.20  | 0.078    | 0        |
| <i>Stavsolus japonicus</i>    | 52       | 798,464     | 7018       | 0.77  | 0.22  | 0.043    | 0.027    |
| <i>Rhabdiopteryx japonica</i> | 22       | 2,320,398   | 9248       | 0.75  | 0.27  | 0.093    | 0.014    |
| <i>Obipteryx femoralis</i>    | 31       | 3,318,488   | 12,411     | 0.80  | 0.20  | 0.069    | 0.011    |
| <i>Isoperla nipponica</i>     | 32       | 942,014     | 3086       | 0.70  | 0.29  | 0.081    | 0.003    |
| <i>Amphinemura longispina</i> | 30       | 572,899     | 6319       | 0.70  | 0.26  | 0.120    | 0.050    |

Based on fastSTRUCTURE analysis, bayesian clustering analysis revealed 3 to 6 populations (K) as the most parsimonious partitioning of individuals within species. Whereas  $K = 3$  in *H. japonica* [ $\ln(K) = -0.079$ ] and *S. japonicus* [ $\ln(K) = -0.040$ ];  $K = 4$  in *A. longispina* [ $\ln(K) = -0.044$ ], *O. femoralis* [ $\ln(K) = 0.038$ ], *R. japonica* [ $\ln(k) = 0.013$ ], and *N. ovocercia* [ $\ln(K) = -0.003$ ]; and  $K = 6$  in *I. nipponica* [ $\ln(K) = 0.008$ ].

**Genome-environmental association**

Of the 24,318 SNPs, 4251 (18%) candidate SNPs (148–1910 SNPs per species) identified by LFMM were potentially associated with regional environment (Fig. 2, Table 2). The highest number of SNPs associated with environmental variables was identified in *R. japonica* (1910 candidate SNPs), followed by *N. ovocercia* (922 candidate SNPs).

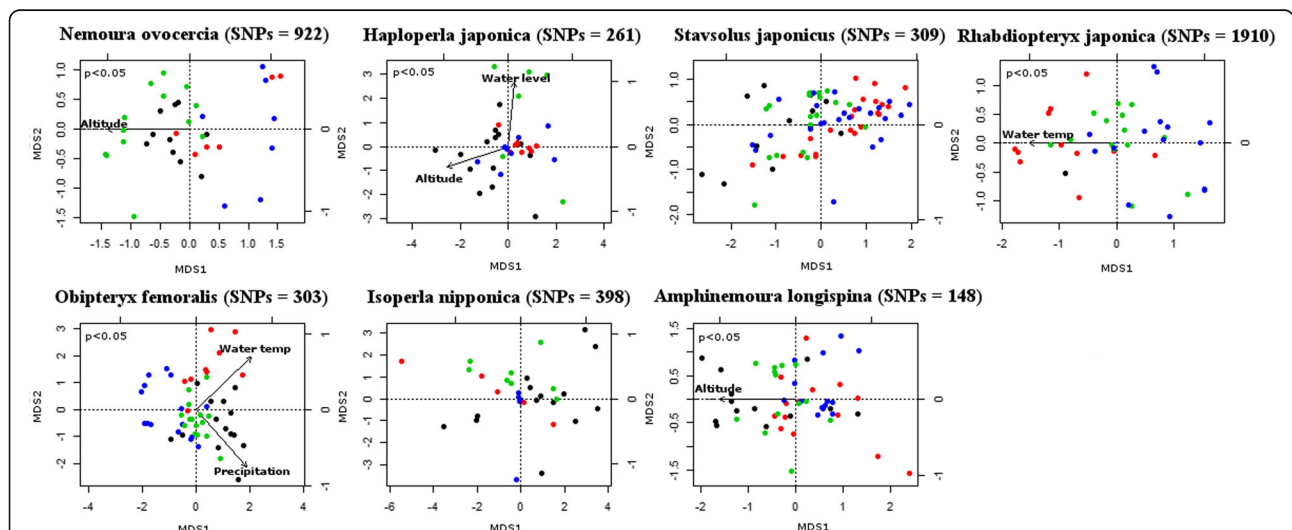
Multidimensional scaling (MDS) of all candidate SNPs showed genomic variation affected by environment (Fig. 2). The average amount of genotypic variation explained by the first two MDS axes was 76% for seven species. MDS2 (overall species average of 33% of the variation) discriminated between genomic variation in *H. japonica* and *O. femoralis*. A combination of MDS1 and MDS2 discriminated between the genomic variation in *N. ovocercia*, *S. japonicus*, *A. longispina*, *R. japonica*, and *I. nipponica*. Environmental variables, including altitude, water level, water temperature, and precipitation significantly affected the genomic variation in five of seven species (*N. ovocercia*, *H. japonica*, *R. japonica*, *O. femoralis*, and *A. longispina*) ( $p < 0.05$ ). No correlation was observed between candidate SNPs and environmental variables for the remaining two species, *S. japonicus*

and *I. nipponica*. Highest genomic variation was detected in Gifu region along a spatially structured environmental gradient, whereas the least was detected in Sendai region. We found no significant correlations between geographical and genetic distance for all species (Mantel test, *H. japonica*  $r = 0.2$ ,  $P = 0.98$ ; *N. ovocercia*  $r = 0.04$ ,  $P = 0.75$ ; *R. japonica*  $r = 0.068$ ,  $P = 0.67$ ; *O. femoralis*  $r = 0.075$ ,  $P = 0.68$ ; *I. nipponica*  $r = 0.2$ ,  $P = 0.98$ ; *A. longispina*  $r = 0.3$ ,  $P = 0.98$ ; and *S. japonicus*  $r = 0.15$ ,  $P = 0.99$ ), which indicated that the genome variation of the candidate SNPs was not affected by geographical distance.

db-RDA analysis of allelic variation revealed 294 SNPs (6–92 SNPs per species) significantly associated with six environmental variables ( $p < 0.05$ ; Table 2). Thirteen SNPs were correlated with more than one environmental variable. Overall, precipitation and altitude were the most frequently selected environmental variables by the db-RDA models (Table 2), and showed strong correlations with 65 and 47 SNPs, respectively, except in the case of *H. japonica* and *I. nipponica*. Among the seven species, *I. nipponica* displayed the highest number of correlations between environmental variables and SNPs.

**Comparative genome analysis**

Multiple sequence alignments of candidate SNPs sequences using bowtie revealed an average of 98% nucleotide sequence similarity among the seven species. Mean nucleotide substitution rate within species per geographical region (Table 3) ranged from 0 to 0.137 (mean = 0.091), where *S. japonicus* had the largest intra-specific genetic variation throughout the four studied regions (overall mean for four regions = 0.137). Six of seven species showed the highest mean nucleotide



**Fig. 2** Genome-wide multivariate genome-environmental associations. Panels show the multidimensional scaling analysis (MDS1 and MDS2) of 4251 candidate SNPs of seven stoneflies species across four regions of Japan. Colored circles represent different regions: red, Matsuyama; black, Gifu; green, Sendai; and blue, Sapporo

**Table 2** Summary of candidate SNPs analysis results and their correlation with environmental variables for each species. Where: H, number of haplotypes; k, number of populations; spatial and environmental correlation: a, altitude; p, precipitation (mm); wl, water level (m); d, discharge (m<sup>3</sup>/s); sc, snow cover (cm); wt, water temperature (°C); at, atmospheric temperature (°C)

| Species                       | SNPs | H      | k | Candidate SNPs | Unique SNPs | Correlated SNPs | Spatial and environmental correlation |    |    |    |    |    |    |
|-------------------------------|------|--------|---|----------------|-------------|-----------------|---------------------------------------|----|----|----|----|----|----|
|                               |      |        |   |                |             |                 | a                                     | p  | wl | d  | sc | wt | at |
| <i>Nemoura ovocercia</i>      | 5073 | 51,230 | 4 | 922            | 30          | 66              | 24                                    | 5  | 5  | 11 | 5  | 5  | 11 |
| <i>Haploperla japonica</i>    | 943  | 5987   | 3 | 261            | 8           | 6               | 3                                     |    | 3  |    |    |    |    |
| <i>Stavsolus japonicus</i>    | 3283 | 58,180 | 3 | 309            | 16          | 10              | 2                                     | 6  |    | 2  |    |    |    |
| <i>Rhabdiopteryx japonica</i> | 4572 | 71,510 | 4 | 1910           | 113         | 52              | 4                                     | 12 | 4  | 4  | 12 | 12 | 4  |
| <i>Obipteryx femoralis</i>    | 6445 | 18,320 | 4 | 303            | 8           | 24              | 4                                     | 5  | 3  |    | 3  | 5  | 4  |
| <i>Isoperla nipponica</i>     | 1434 | 24,405 | 6 | 398            | 8           | 92              |                                       | 31 | 31 |    |    |    | 30 |
| <i>Amphinemura longispina</i> | 2568 | 17,512 | 4 | 148            | 22          | 44              | 10                                    | 6  |    |    | 10 | 8  | 10 |

substitution rate in Gifu region (Table 3). Mean nucleotide substitution rate within geographical regions (range = 0.046–0.160, mean = 0.089) for seven species was significantly lower than that between geographical regions (range = 0.094–0.300, mean = 0.19) ( $p < 0.05$ ). The Gifu region showed the highest mean nucleotide substitution rate between species (0.16) compared with other regions (range = 0.046–0.075).

DNA sequences of 4251 candidate SNPs were subjected to tBLASTx search using the NCBI database in order to identify their gene functions. Of these, 3868 sequences (91% sequences of the candidate SNPs, 4% of the total loci) were found to have matched with functional gene regions of other organisms (such as *Drosophila* and *Lepidoptera*, > 90% similarity) based on Gene Ontology analysis using the UNIPROT database.

We identified an average of 205 SNPs unique to each geographical region (range = 1–98 SNPs per species per region) (Table 4). Across all species, the highest number of unique SNPs occurred in Gifu (133 SNPs) and the lowest number in Sendai (10 SNPs). Among the seven species, *R. japonica* in Gifu showed the highest number of unique SNPs (98 SNPs) (Table 4). Unique SNPs showed high mean pairwise nucleotide substitution rate within species ranging from 0.223 in *H. japonica* to 1.012 in *S. japonicus*.

**Table 3** Mean nucleotide substitution rate within seven species distributed across four regions of Japan based on 4251 candidate SNPs. Numbers in bold indicate the highest values

| Species                       | Matsuyama    | Gifu         | Sendai | Sapporo |
|-------------------------------|--------------|--------------|--------|---------|
| <i>Nemoura ovocercia</i>      | 0.015        | <b>0.058</b> | 0.018  | 0.069   |
| <i>Haploperla japonica</i>    | 0            | <b>0.317</b> | 0      | 0       |
| <i>Stavsolus japonicus</i>    | <b>0.315</b> | 0.092        | 0.058  | 0.197   |
| <i>Rhabdiopteryx japonica</i> | 0.028        | <b>0.247</b> | 0.009  | 0.008   |
| <i>Obipteryx femoralis</i>    | 0.018        | <b>0.115</b> | 0.026  | 0.004   |
| <i>Isoperla nipponica</i>     | 0.116        | <b>0.301</b> | 0      | 0.259   |
| <i>Amphinemura longispina</i> | 0.035        | <b>0.213</b> | 0.037  | 0.013   |

## Discussion

Genome-wide RAD-Seq analysis of seven stream stonefly species conducted in this study successfully identified candidate SNPs associated with environmental conditions along a nationwide latitudinal gradient in Japan. A total of 4251 candidate SNPs in the stonefly genome were potentially associated with local environmental variations, of which 294 SNPs were significantly associated with six environmental variables and 205 were unique for a particular geographical region.

A substantial proportion (> 70%) of genomic variation at the candidate SNPs was shaped by environment conditions for five of the seven species of stoneflies. Precipitation and altitude were most commonly correlated with allelic variation at candidate SNPs. Precipitation in the Japanese archipelago is abundant throughout the islands all year round [56]. Regions in southern Japan, such as Matsuyama, experience the highest amount of rainfall in summer, whereas those in northern Japan, such as Sendai and Sapporo, experience homogenous rainfall throughout the year (Japan Meteorological Agency, [www.jma.go.jp](http://www.jma.go.jp)). Precipitation has been acknowledged as one of the principal environmental variables that shape aquatic insect distribution and diversity [57]. Precipitation is also known to influence adaptive genetic divergence among populations in a variety of plants (e.g. *Sorghum bicolor*, [14]) based on SNPs analysis. Our results showed that precipitation explained a substantial proportion of SNP variation, which was likely associated with the local adaptations of stonefly populations.

In addition to precipitation, altitude of the geographical regions was also associated with the genomic differentiation of stonefly species. Altitude is a long-term studied environmental variable implicated in the distribution [24] and genetic differentiation [58] of stoneflies species. In this study, sampling sites within a geographical region were located at different altitudes (Additional file 1: Table S1). It is, therefore, possible that population differentiation and local adaptation of



**Table 4** Numbers of unique SNPs to each geographical region per species and mean nucleotide substitution rate of the unique SNPs within species

| Species                       | Matsuyama | Gifu | Sendai | Sapporo | Mean nucleotide substitution rate within species |
|-------------------------------|-----------|------|--------|---------|--|
| <i>Nemoura ovocercia</i>      | 6         | 6    | 4      | 14      | 0.407  |
| <i>Haploperla japonica</i>    | 0         | 8    | 0      | 0       | 0.223  |
| <i>Stavsolus japonicus</i>    | 8         | 6    | 0      | 2       | 1.012  |
| <i>Rhabdiopteryx japonica</i> | 6         | 98   | 2      | 7       | 0.331  |
| <i>Obiptyx femoralis</i>      | 0         | 5    | 0      | 3       | 0.363  |
| <i>Isoperla nipponica</i>     | 1         | 6    | 0      | 1       | 0.903  |
| <i>Amphinemura longispina</i> | 4         | 9    | 4      | 5       | 0.511  |

stonefly species were affected by altitude variation within a region. Similar results have been commonly observed in aquatic insects based on mitochondrial DNA sequence variation [59], and in other taxa, including anemonefish [15] and sorghum (*Sorghum bicolor*; [14]) using SNP genotyping. Together these data suggest that environmental requirements of stoneflies species play an essential role in their local adaptation and consequently, in adaptive genetic variation.

Among the four sampling regions of Japan, Gifu demonstrated the highest intraspecific genomic variation of candidate SNPs for six of seven species (range, 0.09–0.3). Much higher nucleotide substitution rates has been reported for the genomes of human [60], bacteria [60], and insects [61], ranging from 0.8 to 0.9, as a consequence of coding region sequences. Additionally, Gifu harbored 67% of total unique SNPs observed. This suggests that Gifu harbors a high level of genetic diversity. Geographically, Gifu is located in the middle of Japan surrounded by the Japanese Alps. This region has high insect diversity because of the geological formation history of the Japanese islands [56, 57]. Ancestral Japanese landmasses were located along the borders of two major tectonic plates, namely the Eurasian (south Japanese landmass) and North American (north Japanese landmass) plate [62]. Both landmasses had independent geological histories until their union approximately 20 million years ago [63] and the consequent formation of the Japanese Alps. The formation of the Japanese Alps has greatly contributed to the evolutionary process (species diversification) of Japanese aquatic insects [56] and the proteome variations of stoneflies species [64]. However, no prior studies explored the geographic effect on intra-specific evolutionary patterns. Our findings provide an invaluable resource for the identification of genes associated with diverse traits of environment importance. Natural SNP variation influenced by environmental conditions has helped to detect genes associated with domestication [65] and climate change [66]. Thus, multiple-species studies at the molecular level have the potential to detect hotspots of genetic diversity, thereby having implications for future research on river

management and climate change. Moreover, multiple-species approaches may be useful in understanding how environmental pressure acts to shape genomic variation of a community in a region rather than that of an isolated species.

Genome–genome comparison of the candidate loci among the seven species showed high genomic similarity (98% nucleotide sequence similarity) and low mean nucleotide substitution rate (0.089) within their geographical regions. An evolutionary process that could lead to similar genomic profiles as a consequence of environmental conditions is parallel evolution [67]. Parallel evolution occurs when phenotypes evolve in the same way in different species descending from the same ancestor because of their adaptation to similar habitats, and is a strong evidence of natural selection in adaptive evolution [68–70]. Parallel divergence could reflect similar selective pressures acting on regions in different localities, geographically non-associated genetic traits, and dependent genetic drift [68, 70]. Previous studies have demonstrated that similar environmental pressure can lead to parallel evolution of the genes under selection. For example, similar pattern of local adaptation of two stick insect populations under similar conditions of nutritional balance (*Timema cristinae*; [68]), similar expression patterns of latitude-driven genes (as cytochrome b) in two *Drosophila* species (*D. melanogaster* and *D. hydei*; [69]), and genes associated with morphological traits (as body shape, etc.) in two populations of three-spined stickleback (*Gasterosteus aculeatus*; [70]) were reported. It is considered that mutations and evolutionary forces, as stabilizing selections that are invoked to maintain the genetic polymorphism among populations, will only play a role in maintaining genetic stability in species in a stable environmental over long periods of time [71].

Of the candidate loci sequences, 27% showed significant similarity to coding sequences of known proteins. This low proportion may be due to incomplete alignment matching with the genome of model organisms (fruit flies and butterflies) and poor representation of insects proteins in public databases. Functions of most of

these genes include metabolism (i.e., cellular amino acid metabolism, catabolism, small molecule metabolism, cellular nitrogen metabolism, organic substance metabolism, biosynthetic metabolism, cellular catabolism, and primary metabolism) and magnesium and citrate transport in the cuticle. Interestingly, the candidate SNPs sequences correlated with water temperature and discharge showed functional annotation to magnesium and citrate transport in the cuticle as previously observed [64]. The remaining annotated sequences showed similar metabolism functionality among species and geographical regions. High degree of candidate loci sharing same or similar functions is commonly observed on RAD-seq studies, such as in snails [72], and sunflowers [73]. The identical descriptions of many annotated candidate loci sequences may be associated to biological functions with related biological pathways. Moreover, RAD-seq method focusses on less expressed genome regions than transcriptome methods. Overall, confirming the expression of these functions in the insect genome requires additional examination along with other approaches, such as gene expression analysis and proteomics.

## Conclusions

Our study demonstrated that local environmental conditions and ecological requirements of organisms play key roles in their evolutionary adaptation. Given the advantage of RAD-Seq, as finding similar genome regions [38, 39] and the repeatability of the experiments on multiple-species, we could affirm that our approach could detect signals of genome variation throughout local adaptation and could be used in future studies. To further understand the influence of environmental conditions in the genetic variation of locally adapted species, studies incorporating a larger number of samples and /or environmental variables combined with gene expression analysis should be conducted. Additionally, we recommend that future studies also include further tests based on a regional scale sampling design with a local environmental condition, and include further observations on the potential effect of variability of the environmental variables on the selection of candidate loci that may provide clearer insights into the genetic responses of stoneflies to local environmental conditions.

## Additional file

**Additional file 1: Table S1.** Sampling sites and associated meteorological data. **Table S2.** Number of individuals per sampling site. Geographical Regions: M, Matsuyama; G, Gifu; S, Sendai; and Sa, Sapporo. Sampling site codes are shown as described in Additional file 1. **Appendix S1.** de novo assembly filtering and post-assembly filtering. **Table S3.** Number of loci by different parameter combinations evaluated for the de novo assembly. **Table S4.** Number of loci after final data filtering using the population program of the STACKS program. **Table S5.**  $F_{st}$  values after final data

filtering using the population program of the STACKS program. **Table S6.** Numbers reads per species per region using the parameter combination m3 M2 n2 for the de novo assembly after the population program by STACKS. **Table S7.** Numbers of loci per species per region using the parameter combination m3 M2 n2 for de novo assembly after population program by STACKS. **Figure S1.** Flow chart of data filtering steps. (DOCX 4012 kb)

## Abbreviations

db-RDA: Distance-based multivariate redundancy analysis; GO: Gene ontology; LFMM: Latent factor mixed model; PCoA: Principal coordinate analysis; SNPs: Single nucleotide polymorphism

## Acknowledgements

For logistic support in the fieldwork, we kindly thank Sueyoshi Masanao, Terutaka Mori, Ryota Kawanishi, and Kei Nukazawa. We also acknowledge Hiroki Hata for the assistant with RAD-seq analysis. Finally, we are grateful to Joeselle Serrana and anonymous reviewers for constructive comments on an earlier version of the manuscript.

## Funding

This study was supported by the Japan Society for the Promotion of Science (JSPS) (grants: 13F03366 and 17F17908). MG was supported by JSPS International Research Fellowship (ID: P13366, PU17908). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

The RAD-seq data were deposited in Sequence Read Archive (SRA) database under the Bioproject accession number (PRJNA496336) (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA496336/>).

## Authors' contributions

MG and KW designed the study. Sampling, laboratory analyses, and bioinformatics analyses were conducted by MG. Both the authors contributed to the writing of the manuscript. Both authors read and approved the final manuscript.

## Ethics approval and consent to participate

The field sampling was conducted without the necessity of any specific permit.

## Consent for publication

No applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 12 September 2018 Accepted: 14 January 2019

Published online: 24 January 2019

## References

- Hoffmann AA, Sgro CM. Climatic change and evolutionary adaptation. *Nature*. 2011;470:479–85.
- Holt RD, Gaines MS. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. *Evol Ecol*. 1992;6:433–47.
- Hendrick PW. Genetic polymorphism in heterogeneous environments: the age of genomics. *Ann Rev Ecol Evol Syst*. 2006;37:67–93.
- Holderegger R, Kamm R, Gugerli F. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecol*. 2006;21:797–807.
- Jones MR, Forester BR, Teufel AL, Adams RV, Anstett DN, Goodrich BA, et al. Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*. 2013;67:3455–68.
- Manuel S, Holderegger R. Ten years of landscape genetics. *Trends Ecol Evol*. 2013;28:614–21.

7. Weinig C, Ewers BE, Welch SM. Ecological genomics and process modeling of local adaptation to climate. *Curr Opin Plant Biol.* 2014;18:66–72.
8. Lenormand T. Gene flow and the limits to natural selection. *Trends Ecol Evol.* 2002;17:183–9.
9. Thomas D. Gene–environment-wide association studies: emerging approaches. *Nat Rev Genet.* 2010;11:259–72.
10. Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, et al. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol Ecol.* 2010;19:3760–72.
11. Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, et al. Ecological genetics in the north Atlantic: environmental gradients and adaptation at specific loci. *Ecology.* 2008;89:591–5107.
12. Leempoel K, Parisod C, Geiser C, Joost S. Multiscale landscape genomic models to detect signatures of selection in the alpine plant *Biscutella laevigata*. *Ecol Evol.* 2018;8:1794–806.
13. De Kort H, Vandepitte K, Mergeay J, Mijnsbrugge KV, Honnay O. The population genomic signature of environmental selection in the widespread insect-pollinated tree species *Frangula alnus* at different geographical scales. *Heredity.* 2015;115:415–25.
14. Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash T, Bonnette J, et al. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv.* 2015;1:e1400218.
15. Saenz-Agudelo P, Dibattista JD, Piatek MJ, Gaitner MR, Harrison HB, Nanninga GB, et al. Seascape genetics along environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of anemonefishes. *Mol Ecol.* 2015;24:6241–55.
16. Swaegers J, Mergeay J, Van Geystelen A, Therry L, Larmuseau MHD, Stoks R. Neutral and adaptive genomic signatures of rapid poleward range expansion. *Mol Ecol.* 2015;24:6163–76.
17. Waldvogel AM, Wieser A, Schell T, Patel S, Schmidt H, Hankeln T, et al. The genomic footprint of climate adaptation in *Chironomus riparius*. *Mol Ecol.* 2018;27:1439–56.
18. Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK, et al. Population genomic footprint of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol Ecol.* 2013;22:5594–607.
19. Limborg MT, Helvar SJ, De Bruyn M, Taylor MI, Nielsen EE, Ogden R, et al. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol Ecol.* 2012;21:3686–703.
20. Hand BK, Lowe WH, Kovach RP, Muhlfeld CC, Luikart G. Landscape community genomics: understanding eco-evolutionary process in complex environments. *Trends Ecol Evol.* 2015;30:161–8.
21. Chamberlain SA, Bronstein JL, Rudgers JA. How context dependent are species interactions? *Ecol Lett.* 2014;14:881–90.
22. Gugerli F, Brandl R, Castagneryol B, Franc A, Jactel H, Koelewijn HP, et al. Community genetics in the time of next-generation molecular technologies. *Mol Ecol.* 2013;22:3198–207.
23. Stewart KW, Stark BP. Plecoptera. In: Merritt RW, Cummins KW, Berg MB, editors. An introduction to the aquatic insects of North America. Iowa: Kendall/Hunt Publishing Co; 2008. p. 311–84.
24. Prenda J, Gallardo-Mayenco A. Distribution patterns, species assemblages and habitat selection of the stoneflies (Plecoptera) from two Mediterranean river basins in South Spain. *Int Rev Hydrobiol.* 1999;84:595–608.
25. Heino J. Biodiversity of aquatic insects: spatial gradients and environmental correlates of assemblage-level measures at large scales. *Fr Rev.* 2009;2:1–29.
26. Phillipsen IC, Lytle DA. Aquatic insects in a sea of desert: population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. *Ecography.* 2012;36:731–43.
27. Watanabe K, Kazama S, Omura T, Monaghan MT. Adaptive genetic divergence along narrow environmental gradients in four stream insects. *PLoS One.* 2014. <https://doi.org/10.1371/journal.pone.0093055>.
28. Keller SR, Levsen N, Olson MS, Tiffin P. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Mol Biol Evol.* 2012;29:3143–52.
29. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoeksstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One.* 2012. <https://doi.org/10.1371/journal.pone.0037135>.
30. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology on non-model organisms. *Heredity.* 2011;107:1–15.
31. Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, et al. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One.* 2011. <https://doi.org/10.1371/journal.pone.0019315>.
32. Tiffin P, Ross-Ibarra J. Advances and limits of using population genetics to understand local adaptation. *Trends Ecol Evol.* 2017;32:801–2.
33. Kawai T, Tanida K. Aquatic insects in Japan: manual with keys and illustration. Japan: Tokai University press; 2005. p. 236–90.
34. Borcard D, Legendre P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model.* 2002;153:51–68.
35. Karney CFF. Algorithms for geodesics. *J Geod.* 2013;87:43–55.
36. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2.
37. Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 2013;22:3124–40.
38. Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Pinero D, Emerson BC. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol Ecol Resour.* 2015;15:28–41.
39. Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for STACKS. *Methods Ecol Evol.* 2017. <https://doi.org/10.1111/2041-210x-12775>.
40. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics.* 2014;30:2843–51.
41. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lecher MJ. PopGenome: an efficient swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 2014;31:1929–36.
42. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics.* 2014;197:573–89.
43. Gillbert KJ, Andrew RL, Bock DG, Franklin MT, Kane NC, Moore JS, et al. Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Mol Ecol.* 2012;20:4925–30.
44. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007;23:1801–6.
45. Ryman N, Palm S. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Mol Ecol Notes.* 2006;6:600–2.
46. Vendrami DLJ, Telesca L, Weigand H, Weiss M, Fawcett K, Lehman K, et al. RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *R Soc Open Sci.* 2017;4:160548.
47. Gauderman WJ, Morrison JM. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006. <http://hydra.usc.edu/gxe>. Accessed 01 June 2018.
48. Frichot E, Schoville SD, Bouchard G, Francois O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol.* 2013;30:1687–99.
49. Frichot E, Francois O. LEA: an R package for landscape and ecological association studies. *Methods Ecol Evol.* 2015;6:925–9.
50. Bass JD, Swcf AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control. R package. 2015. <http://github.com/jdstorey/qvalue>. Accessed 10 Apr 2018.
51. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
52. Chang CC. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4:1–7.
53. Legendre P, Anderson MJ. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr.* 1999;69:1–24.
54. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, et al. Vegan. Community Ecology. R Package. 2012. <https://cran.r-project.org/web/packages/vegan/index.html>. Accessed 10 Feb 2018.
55. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
56. Tojo K, Sekine K, Takenaka M, Isaka Y, Komai S, Suzuki T, et al. Species diversity of insects in Japan: their origins and diversification processes. *Entomol Sci.* 2017;20:357–81.
57. Mori T, Murakami M, Saitoh T. Latitudinal gradients in stream invertebrate assemblages at a regional scale on Hokkaido Island, Japan. *Freshw Biol.* 2010;55:1520–32.
58. Boumans L, Hogner S, Brittain J, Johnsen A. Ecological speciation by temporal isolation in a population of the stonefly *Leuctra hippopus* (Plecoptera, Leuctridae). *Ecol Evol.* 2017;7:1635–49.

59. Dijkstra KDB, Monaghan MT, Pauls SU. Freshwater biodiversity and insect diversification. *Annu Rev Entomol.* 2014;59:143–63.
60. Chen J-Q, Wu Y, Yang H, Bergelson J, Kreitman M, Tian D. Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. *Mol Biol Evol.* 2009;26:1523–31.
61. Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Mol Biol Evol.* 2016;34:119–30.
62. Taira A. Tectonic evolution of the Japanese Island arc system. *Annu Rev Earth Planet Sci.* 2001;29:109–34.
63. Otofujii Y, Matsuda T, Nohda S. Opening mode of the Japan Sea inferred from paleomagnetism of the Japan arc. *Nature.* 1985;317:603–4.
64. Gamboa M, Tsuchiya MC, Matsumoto S, Iwata H, Watanabe K. Differences in protein expression among five species of stream stonefly (Plecoptera) along a latitudinal gradient in Japan. *Arch Insect Biochem Physiol.* 2017;96:e21422.
65. Bajaj D, Das S, Badoni S, Kumar V, Singh M, Bansal KC, et al. Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci Rep.* 2015;5:12468.
66. Rodriguez-Quilon I, Santos-del-Blanco L, Serra-Varela MJ, Koskela J, Gonzalez-Martinez SC, Alia R. Capturing neutral and adaptive genetic diversity for conservation in a highly structured tree species. *Ecol Appl.* 2016;26:2254–66.
67. Schluter D, Clifford EA, Nemethy M, McKinnon JS. Parallel evolution and inheritance of quantitative traits. *Am Nat.* 2004;163:809–22.
68. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, et al. Stick insect genomes reveal natural selection role in parallel speciation. *Science.* 2014;344:738–42.
69. Zhao L, Begun DJ. Genomics of parallel adaptation at two timescales in *Drosophila*. *PLoS Genet.* 2017. <https://doi.org/10.1371/journal.pgen.1007016>.
70. Stuart YE, Veen T, Weber JN, Hanson D, Ravinet M, Lohman BK, et al. Contrasting effect of environment and genetics generate a continuum of parallel evolution. *Nat Ecol Evol.* 2017;1:0158.
71. Lieberman BS, Dudgeon S. An evaluation of stabilizing selection as a mechanism for stasis. *Paleogeogr Palaeoclimatol Palaeoecol.* 1996;127:229–38.
72. Ravinet M, Westram A, Johannesson K, Butlin R, Andre C, Panova M. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol Ecol.* 2016;25:287–305.
73. Renault S, Owens GL, Rieseberg LH. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Mol Ecol.* 2014;23:311–24.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

