Introgressive hybridization in the west Pacific pen shells (genus Atrina): Restricted interspecies gene flow within the genome

MOLECULAR ECOLOGY

Introgressive hybridization in the west Pacific pen shells (genus *Atrina*): restricted interspecies gene flow within the genome

Journal:	*Molecular Ecology*
Manuscript ID	MEC-22-0741.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Sekino, Masashi; Japan Fisheries Research and Education Agency, Fisheries Resources Institute Hashimoto, Kazumasa; Japan Fisheries Research and Education Agency, Fisheries Technology Institute Nakamichi, Reiichiro; Japan Fisheries Research and Education Agency, Fisheries Resources Institute Yamamoto, Masayuki; Kagawa Prefectural Government, Fisheries Division Fujinami, Yuichiro; Japan Fisheries Research and Education Agency, Goto Field Station, Fisheries Technology Institute Sasaki, Takenori; Tokyo Daigaku, University Museum
Keywords:	Hybridization, Genomics/Proteomics, Molluscs, Population Genetics - Empirical, Invertebrates

Title: Introgressive hybridization in the west Pacific pen shells (genus *Atrina*): restricted interspecies gene flow within the genome

Running title: Introgression in the west Pacific pen shells

Authors: Masashi Sekino[1] | Kazumasa Hashimoto[2] | Reiichiro Nakamichi[1] | Masayuki Yamamoto[3] | Yuichiro Fujinami[4] | Takenori Sasaki[5]

[1]Fisheries Resources Institute, Japan Fisheries Research and Education Agency, 2-12-4 Fuku-ura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan

[2]Fisheries Technology Institute, Japan Fisheries Research and Education Agency. 1551-8 Taira, Nagasaki 851-2213, Japan

[3]Fisheries Division, Kagawa Prefectural Government, 4-1-10 Bancho, Takamatsu, Kagawa 760-8570, Japan

[4]Goto Field Station, Fisheries Technology Institute, Japan Fisheries Research and Education Agency, 122-7 Nunoura, Tamanoura, Goto, Nagasaki 853-0508, Japan

[5]The University Museum, the University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

#Correspondence

Masashi Sekino. Fisheries Resources Institute, Japan Fisheries Research and Education Agency. 2-12-4 Fuku-ura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan.

Email: sekino_masashi31@fra.go.jp

Abstract

A compelling interest in marine biology is to elucidate how species boundaries between sympatric free-spawning marine invertebrates such as bivalve mollusks are maintained in the face of potential hybridization. Hybrid zones provide the natural resources for us to study the underlying genetic mechanisms of reproductive isolation between hybridizing species. Against this backdrop, we examined the occurrence of introgressive hybridization (introgression) between two bivalves distributed in the western Pacific margin, *Atrina japonica* and *Atrina lischkeana*, based on single-nucleotide polymorphisms (SNPs) derived from restriction site-associated DNA sequencing. Using 1,066 ancestry-informative SNP sites, we also investigated the extent of introgression within the genome to search for SNP sites with reduced interspecies gene flow. A series of our individual-level clustering analyses including the principal component analysis, Bayesian model-based clustering, and triangle plotting based on ancestry–heterozygosity relationships for an admixed population sample from the Seto Inland Sea (Japan) consistently suggested the presence of specimens with varying degrees of genomic admixture, thereby implying that the two species are not completely isolated. The Bayesian genomic cline analysis identified 10 SNP sites with reduced introgression, each of which was located within a genic region or an intergenic region physically close to a functional gene. No or very few heterozygotes were observed at these sites in the hybrid zone, suggesting that selection acts against heterozygotes. Accordingly, we raised the possibility that the SNP sites are within genomic regions that are incompatible between the two species. Our finding of restricted interspecies gene flow at certain genomic regions gives new insight into the maintenance of species boundaries in hybridizing broadcast-spawning mollusks.

KEYWORDS

Atrina pectinata, hybrid incompatibility gene, population genomics, RAD, underdominance

1 | INTRODUCTION

Natural or anthropogenic hybridization between divergent lineages has been observed in many animal taxa (Barton & Hewitt, 1985; Payseur & Rieseberg, 2016; McFarlane & Pemberton, 2019). Hybridization provides a golden chance for us to delve into the genomic architecture of reproductive isolation, the data from which in turn enhances our understanding of speciation processes. Hence, hybrid zones, where divergent lineages come into reproductive contact producing multiple generations of hybrids through introgressive hybridization (introgression), have drawn much interest in evolutionary biology (Barton & Hewitt, 1985; McFarlane & Pemberton, 2019). The genomic architecture of reproductive isolation has been made accessible owing to the advent of massively parallel DNA sequencing technologies, which have aided genotyping of an array of single-nucleotide polymorphisms (SNPs) across the genomes of organisms (Gompert *et al*., 2017).

An exciting finding in the analyses of hybrid zones is that the level of introgression varies within the genome; that is, some allele of a given lineage can easily flow into the gene pool of the other, whereas other loci show restricted inter-lineage exchange of alleles (Payseur, 2010; Gompert *et al*., 2012, 2017; Harrison & Larson, 2014; Payseur & Rieseberg, 2016). Identifying genes with a low rate of introgression is especially interesting since such genes may play a critical role in maintaining reproductive isolation in the face of potential hybridization (Harrison & Larson, 2014). Moreover, genes causing reproductive separation could be a group of speciation drivers (Wu & Ting, 2004).

The occurrence of hybridization has also been reported in broadcast-spawning bivalve mollusks such as clams (Pfenninger *et al*., 2002; Hurtado *et al*., 2011), mussels (Bierne *et al*., 2003; Riginos & Cunningham, 2005; Kijewski *et al*., 2006), oysters (Huvet *et al*., 2004; Zhang *et al*., 2017), and pen shells (Yokogawa, 1996; Liu *et al*., 2011). However, it is not well understood how their species boundaries (sensu Harrison & Larson, 2014) are maintained despite their capabilities of interbreeding. Two hybridizing lineages of the west Pacific pen shells (genus *Atrina* Gray, 1842) offer the natural resources for us to look for the genes responsible for their reproductive isolation.

An *Atrina* species, so-referred *Atrina pectinata* (Linnaeus, 1767), points to an Indo-Pacific pen shell. In the past, there were implications that discernible *Atrina* morphotypes should be synonymized en masse with *A. pectinata* (*e.g*., Rosewater, 1961). The entity of the nominal *A. pectinata* is nonetheless unclear, as the *Atrina* morphotypes have variously been classified (Liu *et al*., 2011; Xue *et al*., 2012, 2021; Hashimoto *et al*., 2018, 2021). An isozyme-based survey in the Seto Inland Sea (Japan; Fig. 1) indicated a large genetic divergence between sympatric population samples of a “non-scaly” form of *A. pectinata*, which bears a smooth outer shell surface, and a “scaly” form with fine ribs and densely packed spines (or scaly protrusions) on its shell surface (Yokogawa, 1996; see Fig. 2). Thus, the author Yokogawa concluded that the taxonomic connections between the morphotypes is at the level of neither subspecies (Habe, 1977) nor local variety (Okutani, 1994), but deserves the status of separate species. Liu *et al*. (2011) identified six different mitochondrial lineages (L1 to L6) within the *A. pectinata* species complex in the western Pacific margin based on the mitochondrial cytochrome *c* oxidase subunit I gene

(*coxI*). This definition, with their nuclear DNA data (internal transcribed spacer I of the nuclear rRNA family) and morphological evidence, led the authors Liu *et al*. to postulate that the *A. pectinata* species complex comprises of at least five cryptic species, mostly in line with the mitochondrial lineages.

Based on the definition of the six mitochondrial lineages, the primary *A. pectinata* lineages in Japanese waters are L1 and L2 (we refer to the lineages as mt-L1 and mt-L2), which correspond to the non-scaly and scaly forms of *A. pectinata*, respectively (Hashimoto *et al*., 2018, 2021). We treat hereafter the two lineages as different species, presuming that the lineage mt-L1 represents *Atrina japonica* (Reeve, 1858), and the lineage mt-L2, *Atrina lischkeana* (Clessin, 1891), according to suggestions by many authors (Huber, 2010; Japanese Association of Benthology, 2012; Schultz & Huber, 2013; Kurozumi, 2017; Xue *et al*., 2021). Both *A. japonica* and *A. lischkeana* typically occur in inner bays (low tidal zones to the depth of 30 m; Japanese Association of Benthology, 2012), but *A. japonica* is found in higher latitudes than is *A. lischkeana* with some exceptions (Liu *et al*., 2011; Hashimoto *et al*., 2021; Xue *et al.,* 2021). Allopatric *A. japonica* and *A. lischkeana* can be distinguished by assessing if their shells are finely ribbed and spinous (Fig. 2). However, we have encountered individuals with intermediate shell traits in the Seto Inland Sea and the Ariake Sea (Fig. 1), where both *A. japonica*-type and *A. lischkeana*-type shells have been identified. Those shells have weak to moderate ribs and no or few spines, making them appear as hybrid between *A. japonica* and *A. lischkeana*. A previous genetic study based on eight isozyme markers with nearly complete allelic substitution between the two species also suggested the presence of their hybrids (Yokogawa, 1996).

As with most bivalves, *A. japonica* and *A. lischkeana* are broadcast spawners with high fecundity. A rearing experiment using individuals from the Seto Inland Sea revealed that the annual reproductive cycles of the two species are synchronized (peak spawning duration from late June to early July; Matsumoto *et al*., 2019). These observations imply that the two species in sympatry have many chances of interbreeding. Despite their prospect of hybridization, they generally maintain their phenotypic difference. Therefore, we presumed that there are some reproductive barriers between them. Yokogawa (1996) claimed that the two species easily hybridize in the Seto Inland Sea, but introgression seldom occurs there due to the sterility of F_1 hybrids, which enables a successful division between their gene pools. Due to the difficulty in rationalizing Yokogawa's contention suggesting little occurrence of introgression (see the discussion section), we hypothesized that other mechanisms of reproductive isolation restrict interspecies gene flow. The presence of genes that block introgression is a possible cause of their reproductive isolation.

In the present study, we aimed at documenting the occurrence of introgression between *A. japonica* and *A. lischkeana* based on nuclear SNP genotypes. We also searched for SNP sites with restricted interspecies gene flow. To those ends, we constructed a draft genome sequence of *A. japonica*, which served as a reference sequence in our SNP identification and prediction of open reading frames (ORFs). Then, we obtained SNP profiles for specimens derived from five areas along the coast of Japan using the restriction site-associated DNA (RAD) sequencing method (Baird *et al*., 2008). Several lines of evidence substantiated that introgression occurs between the two species. A notably low rate of introgression was observed at 10 sites, which were

156 commonly characterized by no or very few heterozygotes in a hybrid zone.
157 Accordingly, we argue about the possibility that the sites are within genomic
158 regions responsible for the maintenance of reproductive isolation between the
159 two species. Our finding of restricted interspecies gene flow at certain genomic
160 regions casts new light on the underlying genetic mechanisms of reproductive
161 isolation in hybridizing free-spawning mollusks.
162

163 ## 2 | MATERIALS AND METHODS

164

165 ### 2.1 | Genomic DNA for genome assembly

166 For *de novo* genome assembly, we used a specimen with an *A. japonica*-
167 type morph (shell length: 275 mm) collected from off Hakodate in Hokkaido
168 Prefecture in 2021 (the sampling site corresponds to "HKDT" in Fig. 1). We
169 presumed the specimen as pure *A. japonica* (see below). To obtain tissue
170 sample for DNA extraction, we sliced off the surface of the posterior adductor
171 muscle with a disposable scalpel blade. Subsequently, we excised a piece of
172 muscle tissue from the newly exposed surface with another clean blade to
173 minimize the risk of contamination by other organisms. The tissue was kept in
174 TNES buffer containing 8 M urea (Asahida *et al*., 1996) and 5 μl of proteinase K
175 solution bundled in QuickGene DNA Tissue Kit S (Fujifilm Wako Chemicals) at
176 37°C for six days. Genomic DNA extraction followed the phenol-chloroform
177 method with RNase A treatment (Sambrook *et al*., 1989). Both valves of the
178 specimen were deposited in the University Museum, the University of Tokyo

(registration ID: RM33908).

2.2 | Geographical samples

In RAD sequencing, we used specimens with their *coxI*-based mitochondrial lineage of either mt-L1 or mt-L2 (Hashimoto *et al*., 2021 and the present study; see below). These specimens were sampled from five areas along the coast of Japan (*n* = 355; Table 1 and Fig. 1): off Hakodate in Hokkaido Prefecture (abbreviated area/sample name: HKDT), off Himakajima Island (Aichi Prefecture) in Mikawa Bay (abbr. HMKJ), off Kagawa Prefecture in the Seto Inland Sea (abbr. KGWA), off Saga Prefecture in the Ariake Sea (abbr. SAGA), and off Fukuejima Island (Nagasaki Prefecture) of the Goto Islands (abbr. GOTO). In all areas but GOTO, the specimens were caught by diving, which involved experienced divers or fishermen, at the depth of < 25 m. The GOTO specimens were hand-picked from a tidal flat at low tide (by YF). All the HKDT and HMKJ specimens had typical *A. japonica*-type morphs, whereas the GOTO specimens were morphologically *A. lischkeana*. The KGWA and SAGA samples contained specimens with undetermined shells (see Supporting information S1 for the results of our visual inspection). Genomic DNA was extracted in the same manner as described above.

The geographical distributions of *A. japonica*- and *A. lischkeana*-type shells are locality-dependent. In the HKDT and GOTO areas, *A. japonica*- and *A. lischkeana*-type shells have been exclusively found, respectively. Additionally,

Hashimoto *et al*. (2021) demonstrated that *coxI* haplotypes of all specimens from the HKDT and GOTO areas fell into a clade of mt-L1 and mt-L2, respectively. Therefore, we assumed that our population samples from the two areas were made up of pure species (HKDT, *A. japonica*; and GOTO, *A. lischkeana*). In the HMKJ area, *A. japonica*-type shells are dominant. Following an interview with local fisherfolk, however, *A. lischkeana*-type shells have been fished, though at a very low frequency. Hashimoto *et al*. (2021) found that one of their 67 specimens from the HMKJ area was morphologically *A. lischkeana* with an mt-L2 *coxI*-haplotype. Although *A. lischkeana*-type shells dominate the pen shell resource in the SAGA area (Aramaki, 2013), it appears that *A. japonica*-type shells were abundant there in the past (Ito, 2004; Koga, 1992). Moreover, recent *coxI*-based surveys (Aramaki 2013; Hashimoto *et al*., 2018, 2021) observed mt-L1 haplotypes in the Ariake Sea at a far lower frequency (3.5–16.1%) than that of mt-L2 haplotypes. Among our study areas, only the KGWA area is where both types of shells and both mitochondrial lineages are frequently found (Hashimoto *et al*., 2021). Overall, there is the possibility of hybridization between *A. japonica* and *A. lischkeana* in the HMKJ, KGWA, and SAGA areas to a lesser or greater extent.

2.3 | Long- and short-read sequencing for genome assembly

We collected long-read (LR) and short-read (SR) sequencing data to build a genome assembly using the hybrid genome assembly method (Bashir *et al*.,

2012). A DNA library for LR sequencing (CLR mode in PacBio Sequell II System; Pacific Biosciences) and paired-end SR sequencing (NovaSeq 6000 System; Illumina) was prepared using SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences) and TruSeq DNA PCR-Free Library Prep Kit (Illumina), respectively. The library construction and the subsequent sequencing were conducted by DNA Link, Inc. (https://www.dnalink.com; last accessed October 11, 2021).

2.4 | Sanger sequencing for the mitochondrial *coxI* gene

Partial *coxI* sequences for all but the SAGA specimens and the source specimen of genome assembly were previously determined (Hashimoto *et al*., 2021; Supporting information S2). We conducted *coxI* amplification in a polymerase chain reaction (PCR) for specimens without data on their mitochondrial lineage (the PCR protocol is detailed in Supporting information S3). Using BigDye Terminator v3.1 Cycle Sequencing Kit combined with a 3730xl DNA Analyzer (Applied Biosystems in Thermo Fisher Scientific), the PCR amplicons were sequenced from both directions.

2.5 | RAD library construction and sequencing

We constructed RAD libraries, each of which included DNA fragments from 20 or fewer specimens (Sekino *et al*., 2016). Briefly, we used the restriction enzyme SbfI (SbfI-HF; New England Biolabs), which recognizes a stretch of

245 eight nucleotides (5'-CCTGCAGG-3'; referred to as SbfI-seq henceforth), to
246 digest genomic DNA. A modified Solexa P1 adapter was ligated to the cleaved
247 DNA with T4 DNA Ligase (New England Biolabs). The P1-ligated fragments
248 from each specimen were pooled and sheared by sonication with a focal size of
249 300 bp (S220 Focused-ultrasonicator; Covaris). Fragments between 250 bp and
250 600 bp were retrieved from 1.8% agarose gel, and their termini were repaired
251 with Mighty Cloning Reagent Set Blunt End (Takara). We used Exo-minus
252 Klenow DNA Polymerase (Epicentre) to produce 3'-adenine-protruding ends of
253 the repaired fragments. A Y-shaped P2 adapter (Coyne *et al*., 2004; Baird *et al*.,
254 2008) was ligated to the fragments. The P1- and P2-ligated fragments were put
255 through 12 cycles of PCR (Phusion High-Fidelity DNA Polymerase; New
256 England Biolabs) using the P5 and P7 PCR primers (Illumina) in eight separate
257 tubes. The post-PCR mixtures were combined and purified (AMPure XP beads;
258 Beckman Coulter), and the resulting DNA pool was referred to as a RAD library.
259 Using a NextSeq 500 sequencer and NextSeq 500 High Output Kit (75 cycles;
260 Illumina), we conducted single-end sequencing for a mixture of DNA containing
261 the library and PhiX Control v3 (Illumina) at a molar concentration ratio of three
262 to one.
263

264 **2.6 | Construction of genome assembly and ORF prediction**

265 We used Trimmomatic version 0.39 (Bolger *et al*., 2014) for adapter clipping
266 and quality filtering for SRs (minimum average Phred-scale base-quality score

of 25 in a window size of four; minimum LEADING and TRAILING base-quality score of 20; and minimum read length of 50). With Filtlong version 0.2.1 (https://github.com/rrwick/Filtlong; last accessed October 6, 2021), we removed reads with the length of < 500 bases from LR data. We mapped both filtered SRs and LRs onto a reported complete *A. pectinata* mitogenome sequence (lineage mt-L1; accession number, KC153059) using the MEM algorithm in Burrows-Wheeler Aligner (BWA) version 0.7.12 (Li & Durbin, 2010). Unmapped reads extracted with the subprogram *view* of SAMtools version 0.1.19 (Li *et al*., 2009) were used for genome assembly in HASLR version 0.8a1 (Haghshenas *et al*., 2020). HASLR expects a predicted genome size as an input parameter. Because little was known about the genome size of *A. japonica*, we applied the *k*-mer method to our SR data (KmerGenie version 1.7016; Chikhi & Medvedev, 2014) to obtain a tentative estimate of the genome size (*ca.* 900 Mb). Among the other modifiable parameters, we set the minimum coverage depth of 50 for LRs required for genome assembly. With POLCA (Zimin & Salzberg, 2020) available in MaSuRCA version 3.4.2 (Zimin *et al*., 2017), we enhanced the precision of the resulting genome assembly using SR data. The polished genome assembly was named a reference sequence of the *A. japonica* nuclear genome (ref-genome). Using gVolante version 2.0.0 (Nishimura *et al*., 2017), we evaluated the completeness of the ref-genome based on the BUSCO analysis (Simão *et al*., 2015) against a molluscan ortholog gene set (5,295 query genes).

Based on *A. pectinata* transcriptome data retrieved from the Sequence Read Archive (DRR209159, DRR348924, DRR348925, DRR348926, and SRR2016653; as of March 10, 2022), we located ORFs in the ref-genome. We

292 used Prinseq++ version 1.2 (Cantu *et al*., 2019) to quality-filter the raw reads.
293 Reads with a length of < 50, a mean base-quality score of < 20, and/or missing
294 bases of > 10 were omitted. Poly A/T stretches were trimmed with the minimum
295 threshold A/T length of five. We treated the clean paired-end reads as single-
296 end reads, and mapped them onto the ref-genome (Hisat2 version 2.2.1; Kim *et*
297 *al*., 2019). Mapped reads with a mapping quality score of ≥ 20 and the "NH:i:1"
298 tag (no alternative hit) were selected with SAMtools *view* and the UNIX *grep*
299 command, and the extracted reads were assembled into transcripts (StringTie
300 version 2.1.7; Pertea *et al*., 2015). The resulting gtf-output was used to predict
301 ORFs (≥ 100 amino acids) with TransDecoder version 5.5.0
302 (https://github.com/TransDecoder; last accessed March 11, 2022).
303

304 **2.7 | Data processing of mitochondrial *coxI* sequences**

305 The forward and reverse *coxI* sequences were trimmed and aligned with
306 DNASIS Pro version 2.02 (Hitachi Software Engineering). We performed
307 multiple alignment among the sequences based on the Clustal W algorithm
308 (Thompson *et al*., 1994) in MEGA X version 10.0.4 (Kumar *et al*., 2018). The
309 aligned sequences were trimmed to a constant length of 606 bases.
310

311 **2.8 | Nuclear SNP genotyping**

312 RAD sequences (85 bases per read) containing six bases of the SbfI-seq
313 were de-multiplexed and truncated to 64 bases with the subprogram
314 *process_radtags* of Stacks version 1.35 or higher (Catchen *et al*., 2011). Reads

that had a base-quality score of < 20 at 5% or more of the bases were removed with FASTX-Toolkit version 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/download.html; last accessed July 10, 2020). The retained reads were aligned to the ref-genome (BWA-MEM), and reads with a possible alternative hit ("XA" tag) and/or chimeric reads ("SA" tag) were excluded. After converting the resulting bam-data to the mpileup format (SAMtools *mpileup*), we conducted variant calling using the subprogram *mpileup2snp* of VarScan 2 version 2.4.4 (Koboldt *et al*., 2012). We set the following parameters, leaving the other parameters unchanged: 30 as the minimum base-quality score, five as the minimum number of reads that supported a variant, and 0.05 as the threshold probability in calling a variant (Fisher's exact test).

Among the called SNP sites, we rejected non-biallelic sites and those with a coverage depth of < 30 for each specimen using VCFtools version 0.1.16 (Danecek *et al*., 2011). Sites with a maximum coverage depth of $D + 3\sqrt{D}$ were allowed, where *D* is the average depth over all SNP sites and specimens (Li, 2014; *D* was 227). We removed specimens with more than 10% of missing genotypes (VCFtools). Subsequently, we selected sites with minor allele frequency (MAF) of > 0.05 across the population samples (MAF filtering) and those available in ≥ 90% of the specimens across the population samples as well as in each population sample (shared-SNP filtering) using VCFtools. With our own R script (R Core Team, 2016), we thinned the retained sites so that

337 neighboring sites were at least 1 Kb apart in a contig of the ref-genome
338 (Supporting information S4). Moreover, we pruned the surviving sites according
339 to Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD). We
340 applied these filtering procedures only to HKDT and GOTO, each of which was
341 presumed to represent a random mating population of pure species. Thus, sites
342 with no variation in both samples were out of our consideration. The HWE
343 testing used an exact test (Wigginton *et al*., 2005) available in VCFtools (critical
344 *P* of 0.05 without correction of significance level for multiple comparisons). Sites
345 that failed to meet HWE in either sample or both were omitted. Our LD pruning
346 with BCFtools version 1.8 (Li, 2011) was based on the LD statistic of r^2
347 (threshold r^2 = 0.1) estimated for pairs of sites with MAF of > 0.05 in each
348 sample.

349

350 **2.9 | Statistical analyses based on mitochondrial *coxI* sequences**

351 We constructed a *coxI*-based haplotype network (a parsimony algorithm;
352 Templeton *et al*., 1992) using Popart version 1.7 (Leigh & Bryant, 2015). With
353 Arlequin version 3.5.2.2 (Excoffier & Lischer, 2010), we calculated pairwise F_{ST}
354 (Φ_{ST}) between population samples (mtF_{ST}) with the Kimura 2-parameter model
355 (Kimura, 1980). The significance of mtF_{ST} values was evaluated with 10^4
356 permutations of haplotypes between samples. We applied the false discovery
357 rate (FDR) correction (Benjamini & Hochberg, 1995) to the resulting probability
358 values to obtain FDR-corrected probability values (FDR-*q*; Pike, 2011).

359

360 **2.10 | Population-level statistical treatments based on nuclear SNPs**

361 The observed and expected heterozygosities at nuclear SNP sites (H_O and
362 H_E, respectively) were calculated in Arlequin. We also used Arlequin to estimate
363 pairwise nuclear F_{ST} (ncF_{ST}) with 10^4 permutations. The abovementioned HWE
364 testing was applied to samples other than HKDT and GOTO. The number of
365 sites with complete allelic substitution between samples was manually counted.

366

367 **2.11 | Individual-level clustering for hybrid detection**

368 We employed three individual-level clustering analyses to detect hybrids.
369 We note that assigning individuals into simple hybrid classes such as F_1, F_2,
370 and first-generation backcross (BC_1) can be problematic and misleading,
371 especially for specimens from hybrid zones with a long history of introgression
372 (Boecklen & Howard, 1997; Anderson & Thompson, 2002; Fitzpatrick, 2012).
373 Therefore, categorical assignment of our specimens into predefined hybrid
374 classes was out of our scope.

375 First, we carried out the principal component analysis (PCA) using
376 Adegenet version 2.1.3 (Jombert, 2008). Second, we performed a Bayesian
377 model-based clustering with Structure version 2.3.4 (Pritchard *et al.*, 2000)
378 under the admixture and the correlated allele frequency models (Falush *et al.*,
379 2003). We changed the number of clusters (*K*) from one to six (the number of
380 population samples; see below). We set an initial burn-in period of 5×10^4

381 followed by 25 × 10^4 replications in Markov chain Monte Carlo (MCMC)
382 simulations (10 independent MCMC runs at each *K*). We inferred the most likely
383 *K* according to the delta *K* statistic (Evanno *et al.,* 2005) calculated in Structure
384 Harvester (Earl & vonHoldt, 2012) and four other statistics (Puechmaille, 2016)
385 obtained with StructureSelector (Li & Liu, 2018). After determining the best *K*,
386 we conducted a single run of Structure with the *K*, and visualized the distribution
387 of membership coefficient (*q*) in each specimen (Distruct version 1.1;
388 Rosenberg, 2004). Third, we evaluated relationships between individual
389 heterozygosity (H_{ind}) and genome-wide mean ancestry (hybrid index; *HI*) for
390 each specimen from HMKJ, KGWA, and SAGA, which had the possibility of
391 containing hybrids. We estimated H_{ind} by manually counting heterozygous sites.
392 The maximum likelihood estimate of *HI* (Buerkle, 2005) was calculated in
393 Introgress version 1.2.3 (Gompert & Buerkle, 2010), in which HKDT and GOTO
394 were defined as parental populations of pure species (*HI* = 0.000 in *A.*
395 *lischkeana* and 1.000 in *A. japonica*). In this analysis, we selected SNP sites
396 with an allele frequency difference of > 0.300 between HKDT and GOTO, since
397 the accuracy of the maximum likelihood *HI* is affected by the magnitude of allele
398 frequency differences between parental species (Buerkle, 2005; the selected
399 sites are referred to as ancestry-informative sites). We estimated H_{ind} and *HI* for
400 simulated hybrids to evaluate the extent to which H_{ind}–*HI* relationships based on
401 the ancestry-informative sites deviated from those expected with diagnostic
402 markers (see Supporting information S5 for the expected H_{ind} and *HI* in early

generations of hybrids based on diagnostic markers). We produced simulated individuals for two classes of pure species (P_0J, *A. japonica*; and P_0L, *A. lischkeana*; *n* = 200 for each) using the observed genotypes in HKDT and GOTO (Hybridlab version 1.1; Nielsen *et al*., 2006). Based on the simulated genotypes of parental species, we created eight hybrid classes (*n* = 200 for each) using Hybridlab as follows: F_1, F_2, BC_1 (BC_1J: $F_1 \times P_0J$; and BC_1L: $F_1 \times P_0L$), second-generation backcross (BC_2J: $BC_1J \times P_0J$; and BC_2L: $BC_1L \times P_0L$), and cross between BC_1J or BC_1L and the other parental species (BC_1J_L: $BC_1J \times P_0L$; and BC_1L_J: $BC_1L \times P_0J$). For each simulated individual, H_{ind} and *HI* were calculated in the abovementioned manner. The H_{ind}–*HI* relationships in each of the real and simulated datasets were evaluated by drawing a triangle plot (ggplot2 version 3.3.6; Wickham, 2016).

2.12 | Genomic cline analysis

Using the ancestry-informative sites, we employed the Bayesian genomic cline analysis (Gompert & Buerkle, 2011) to find sites with a low rate of introgression. This method estimates two genomic cline parameters (α and β) at each locus in an admixed population. As a function of *HI*, the parameter α (cline center) represents an increase or decrease in the probability of ancestry from one parental population, whereas the parameter β (cline rate) reflects the rate of change in ancestry probability from one parental population to the other (Gompert & Buerkle, 2011). Positive and negative values of α indicate an excess of ancestry from one parental population and from the other parental

population, respectively (in our computational setting, a positive *α* supported an excess of *A. japonica* ancestry, and negative *α*, excess of *A. lischkeana* ancestry). A low rate of introgression results in a positive *β* (steeper cline), whereas a negative *β* (wider cline) indicates a high rate of introgression. Therefore, the parameters *α* and *β* can represent the direction of introgression and the amount of introgression, respectively (Gompert & Buerkle, 2011; Janoušek *et al*., 2015; McFarlane *et al*., 2021).

We estimated the cline parameters for an admixed sample (KGWA; see below) with bgc version 1.03 (Gompert & Buerkle, 2012), in which HKDT and GOTO were set as a reference sample of pure *A. japonica* and *A. lischkeana*, respectively. We used default parameters for computation in bgc, except for the MCMC parameters (2×10^6 iterations including the first half of burn-in). We conducted MCMC simulation with a 50-thinning interval in three independent chains. The resulting three sets of MCMC samples (2×10^4 samples per set) were combined with ClineHelpR (Martin *et al*., 2021). We investigated if there were remarkably high (positive) or low (negative) values of the cline parameters relative to the average over all the used sites (outlier). Using ClineHelpR, we determined outliers when the obtained cline parameters met both of the following criteria (Gompert & Buerkle, 2012; Martin *et al*., 2021): 1) the 95% credibility interval for the posterior probability distribution of the cline parameters did not include the neutral expectation of zero; and 2) median posterior values of the cline parameters fell outside the interval bounded by the $\frac{N}{2}$ and $\frac{1-N}{2}$ quantiles of the posterior probability distribution concerning the random locus-effect prior for the cline parameters ($qN_{0.975}$ threshold, N = 97.5%; and $qN_{0.990}$, N = 99.0%).

3 | RESULTS

3.1 | Genome assembly

The filtered LRs comprised approximately 119 Gb, and the paired-end SRs, 106 Gb. A plot for GC-content (%) against read coverage (SRs) exhibited a normal distribution, suggesting that there was little concern about contaminants (Supporting information S6). The final genome assembly (ref-genome) consisted of 835,332,209 bases in 3,391 contigs. Thus, the nuclear genome size of *A. japonica* would be around 835 Mb. The maximum, mean, and N50 length was 6,330,338, 246,338, and 911,473 (260 of 3,391 contigs), respectively. The completeness of the ref-genome (BUSCO analysis) was 91.2% (93.7% when partially matched genes were included). Nevertheless, the ref-genome is still rough, given the haploid chromosome number of 17 in *A. pectinata* (Liqing *et al*., 2020). The ref-genome contained 5,040 SbfI-recognition sites spread across 1,156 contigs. The number of SbfI-recognition sites per contig and contig length were positively correlated (Supporting information S7). We found 51,399 possible ORFs in the ref-genome.

3.2 | Selection of specimens and sample subdivision

Of the 355 specimens, we omitted 19 from KGWA after processing RAD data, due to the paucity of reads that were effectively mapped onto the ref-genome (< 250,000 reads; n = 4), low proportion of mapped reads (2.1%; n = 1), or failure to qualify the 10%-threshold for missing genotypes (n = 14). The remaining 336 specimens were used in our downstream analyses.

The separation between the mitochondrial lineages mt-L1 and mt-L2 was evident according to the aligned *coxI* sequences without insertion/deletion polymorphism (Fig. 3). All the HKDT and HMKJ specimens, as well as the source specimen of the ref-genome, were from the lineage mt-L1, whereas all the GOTO and SAGA specimens had mt-L2 haplotypes. Only KGWA contained both mt-L1 and mt-L2 haplotypes (mt-L1, n = 112: and mt-L2, n = 71). We divided the KGWA specimens into two population subsamples based on their mitochondrial lineage (KGL1, lineage mt-L1; and KGL2, mt-L2), resulting in the redefinition of six population samples (mt-L1: HKDT, HMKJ, and KGL1; and mt-L2: GOTO, KGL2, and SAGA). All the following statistical analyses, except for the genomic cline analysis, were based on this sample definition. The MAF filtering and shared-SNP filtering were re-applied to SNP data after selecting specimens and subdividing KGWA.

3.3 | Nuclear SNPs used for statistical analyses

After a series of filtering, 1,474 SNP sites were retained; however, five of them were found in contigs with no SbfI-seq in the ref-genome, most likely owing to the mutation(s) within the SbfI-seq in the corresponding genomic regions of the ref-genome's source specimen. We excluded these five sites and kept the remaining 1,469 sites. We presumed that the SNP set was not contaminated by mitochondrial SNPs, as no SbfI-recognition site was found in the abovementioned *A. pectinata* mitogenome sequence. The proportion of missing genotypes per specimen (missingness) ranged from 0.0 to 9.3% (94% of the specimens had a missingness of < 1.0%).

501 **3.4 | Within-population genetic variability based on nuclear SNPs**

502 A high proportion of polymorphic sites in KGL1 (Table 2; 99.9%) and KGL2
503 (98.7%) contrasted with those of the other samples (32.1–54.4%). The number
504 of non-HWE sites was 43 out of 507 polymorphic sites in HMKJ and 75 of 799
505 in SAGA. A HWE departure could occur by chance at approximately 25 and 40
506 sites in HMKJ and SAGA, respectively, at an uncorrected threshold probability
507 of 0.05. Thus, non-HWE sites in the two samples could largely be described by
508 the type I errors (false positive), but large numbers of non-HWE sites in KGL1
509 (922 of 1,467 polymorphic sites) and KGL2 (479 of 1,450) were not the case.
510 Consequently, KGL1 and KGL2 exhibited an obvious deficiency in heterozygote
511 (H_O/H_E in Table 2).

512

513 **3.5 | Genetic population divergence**

514 We estimated pairwise mtF_{ST} for three pairs within each mitochondrial
515 lineage (Supporting information S8 provides an F_{ST} matrix). All pairs of the mt-
516 L1 samples yielded a near-zero mtF_{ST} (from -0.004 to 0.001; FDR-q > 0.500),
517 whereas mtF_{ST} values estimated for all pairs of the mt-L2 samples significantly
518 deviated from zero (0.040–0.138; FDR-q < 0.020). In the same manner with the
519 estimation of mtF_{ST}, pairwise ncF_{ST} was calculated for sample pairs within each
520 mitochondrial lineage. Consistent with the extent of mitochondrial divergence,
521 pairwise ncF_{ST} values between the mt-L1 samples (0.006–0.053) were
522 generally smaller than those estimated for pairs of the mt-L2 samples (0.052–
523 0.093), although all the six ncF_{ST} values were significantly high (FDR-q <
524 0.002). When comparisons were extended to sample combinations between
525 mitochondrial lineages (Fig. 4), four pairs that did not involve KGL1 and KGL2

526 (GOTO *vs*. HKDT and HMKJ; and SAGA *vs*. HKDT and HMKJ) generated
527 similar ncF_{ST} values (0.847–0.866). More reduced ncF_{ST} values were observed
528 in pairs containing KGL1 (*vs*. GOTO and SAGA; 0.689 and 0.680, respectively)
529 and KGL2 (*vs*. HKDT and HMKJ; 0.726 and 0.719). The sympatric pair (KGL1
530 and KG2) gave the smallest ncF_{ST} (0.589).

531 Complete allelic substitutions were observed in four pairs between
532 mitochondrial lineages: HKDT *vs*. GOTO (517 sites), HKDT *vs*. SAGA (293),
533 HMKJ *vs*. GOTO (577), and HMKJ *vs*. SAGA (327). No allelic substitution was
534 found between samples within each mitochondrial lineage. There was no site
535 with allelic substitution between KGL1 and KGL2. Moreover, each of the two
536 samples shared alleles at all the sites with all the samples even from the
537 different mitochondrial lineage.

538

539 **3.6 | Individual-level clustering**

540 On a PCA planar surface (Fig. 5), most of the specimens were explicitly
541 divided by the first component (PC1), consistent with their mitochondrial
542 lineages. The second component (PC2) separated the GOTO specimens from
543 those of KGL2 and SAGA within the lineage mt-L2. Therefore, we identified
544 three distinct clusters: one occupied by all the HKDT and HMKJ specimens, and
545 a large majority of the KGL1 specimens (cluster HHK), another by all the SAGA
546 specimens and a large majority of the KGL2 specimens (cluster SK), and the
547 third by the GOTO specimens exclusively (cluster GOTO). There were
548 specimens of KGL1 and KGL2 that were positioned between the clusters HHK

and SK along the PC1 axis. Three KGL1 specimens were settled in the cluster SK (mt-L2), despite them having mt-L1 haplotypes.

In the Bayesian clustering, the delta *K* statistic implied an optimal *K* of two, which was supported by almost all the four other statistics (Supporting information S9; the MaxMedK statistic proposed the most likely *K* of three at a threshold *q* of 0.5). According to the *q* plot with *K* = 2 (Fig. 6), the specimens from HKDT (putatively pure *A. japonica*) and GOTO (*A. lischkeana*) were assigned to groups A and B, respectively, with high *q* values in the respective groups (*q* in group A in HKDT: 0.976–1.000; *q* in group B in GOTO: 0.988–1.000). Thus, we defined group A as an *A. japonica* group and group B as an *A. lischkeana* group. This definition, together with the distribution of *q* within each specimen, suggested that KGL1 and KGL2 contained specimens with differing degrees of genomic admixture between the two species. The assignment of three KGL1 specimens (mt-L1) to the *A. lischkeana* group (*q* values of 0.991–1.000 in this group) agreed to the findings that the three belonged to the cluster SK (mt-L2) in the PCA plot (Fig. 5). We ran Structure with *K* = 3 as suggested by the MaxMedK estimator, expecting that the GOTO specimens could form another group as with the PCA results. However, our attempted clustering failed, as just five GOTO specimens had a very low *q* value in the third group (group C; *q* of 0.001–0.005; Supporting information S10). Instead, putatively pure *A. japonica* (HKDT specimens) had a relatively high *q* value in group C (*q* of 0.200–0.388). Accordingly, it was challenging to interpret the clustering

571 results with K = 3.

572 We visualized H_{ind}–HI relationships based on 1,066 ancestry-informative
573 sites (Fig. 7). In simulated hybrid classes (Fig.7a), overall, H_{ind} values slightly
574 deviated from those expected with diagnostic markers (*e.g.*, F_1 should have an
575 H_{ind} of 1.000), since the ancestry-informative sites included those with shared
576 alleles between the parental species (HKDT and GOTO). Nevertheless, the
577 SNP set was expected to have a high resolving power in classifying at least the
578 early generation hybrids (Fig. 7a). In the triangle plot for real data (Fig. 7b), the
579 configurations of specimens resembled those observed in the PCA plotting, in
580 that KGL1 and KGL2 contained specimens that did not form a particular cluster.

581

582 **3.7 | Genomic cline analysis**

583 In KGWA, an admixed sample according to the results of the clustering
584 analyses, the genomic cline analysis identified 104 and 124 outliers for the cline
585 parameters α and β, respectively, at the $qN_{0.975}$ threshold (Fig. 8a). A vast
586 majority of the β outliers (114) were negative, demonstrating a high rate of
587 introgression. Positive β outliers were found at mere 10 SNP sites (Fig. 8b).
588 With the highly stringent $qN_{0.990}$ threshold, no positive β outlier was detected. As
589 a supplementary analysis, we examined the goodness of fit of five cline models
590 (simple sigmoid cline, left-tailed cline, right-tailed cline, cline with mirrored tails,
591 and cline with independent tails; Derryberry *et al*. 2014) at the sites with a
592 significantly reduced introgression ($qN_{0.975}$) by estimating the maximum

593 likelihood clines. All the 10 sites commonly exhibited a sigmoid cline
594 (Supporting information S11).

595 No allele was shared between HKDT and GOTO at almost all the 10 SNP
596 sites (Table 3; at the sites S0033 and S0361, a heterozygote was found in
597 HKDT and GOTO, respectively). We refer to the dominant allele at each site in
598 HKDT and GOTO as *japonica*-type and *lischkeana*-type alleles, respectively. In
599 KGWA, very few heterozygotes were found at these sites, and even a complete
600 absence of heterozygote was observed at the site S0361 (Table 3; H_O ranged
601 from 0.000 to 0.016). These low estimates of H_O were determined as statistical
602 outliers (Supporting information S12). As with GOTO, *lischkeana*-type alleles
603 were dominant in SAGA. Despite the presence of one or more homozygotes of
604 the *japonica*-type allele at four of the 10 sites, there was no heterozygote at
605 these sites (Table 3).

606 Of the 10 SNP sites, six were located within genic regions, but none was in
607 the protein-coding region (Table 4). The remaining four sites were in intergenic
608 regions, however, three of which were particularly very close to genic regions
609 (distances of hundreds to thousands of bases; Table 4). According to our
610 BLASTP search against the clustered non-redundant database for the genic
611 regions within which the intragenic SNP sites were positioned and for the
612 nearest genic regions to the intergenic SNP sites, the amino acid sequences of
613 all the genic regions had a high similarity with those of predicted genes in other
614 bivalves (Table 4). As we explain later, all but two of the 10 genes had known

615 functions.

616

617 ## 4 | Discussion

618

619 ### 4.1 | Evidence of introgression

620 The present study provided ample evidence of hybridization between *A.*
621 *japonica* and *A. lischkeana*, notably in the Seto Inland Sea. The extremely large
622 numbers of polymorphic sites in KGL1 and KGL2 were consistent with our
623 presumption that the two samples contained hybrids. Our hybridization
624 hypothesis also accounts for the absence of SNP site with allelic substitution
625 and the reduced estimates of ncF_{ST} between the two samples and the other
626 samples from the different mitochondrial lineage. The clustering analyses
627 confirmed the presence of a genomic admixture in the Seto Inland Sea. The
628 thread-like clustering pattern of specimens from KGL1 and KGL2 in the PCA
629 and triangle plots, and the varying distributions of *A. japonica* and *A. lischkeana*
630 ancestries within the specimens of the two samples as represented by *q* values
631 (Structure analysis), illustrated how the genome from the two species admixed
632 varied depending on the specimens. The different degrees of genomic
633 admixture among the specimens of both mitochondrial lineages can be
634 explained by the presence of various generations of hybrids resulting from
635 reciprocal interspecies gene flow. Furthermore, the three instances of palpable
636 cyto-nuclear discordance (*i.e.*, mt-L1 specimens were apparently *A. lischkeana*

637 according to the nuclear SNPs) stood for the long-standing occurrence of
638 introgression.

639 Our results supporting the occurrence of introgression are at odds with
640 Yokogawa's (1996) claim that the reproductive success of F_1 hybrids between
641 *A. japonica* and *A. lischkeana* in the Seto Inland Sea, if any, would be
642 insignificant due to F_1 sterility. Yokogawa's logic behind his hypothesis of scarce
643 reproductive success of F_1 was that there were a few specimens with H_{ind} of
644 around 0.500 (expected H_{ind} in F_2/BC_1 based on diagnostic markers; Supporting
645 information S5) in his population samples containing hybrids, according to eight
646 isozyme markers with nearly complete allelic substitution between the two
647 species. With such a small number of markers, however, a considerable
648 variance in H_{ind} in F_2/BC_1 is predicted (Boecklen & Howard, 1997; Fitzpatrick,
649 2012; McFarlane & Pemberton, 2019). Hence, we consider that the Yokogawa's
650 contention was based on his unwarranted interpretation of data (see Supporting
651 information S13 for more detailed argument).

652 There were few strong signals of ongoing hybridization in samples other
653 than those from the Seto Inland Sea. However, the number of polymorphic sites
654 in SAGA from the Ariake Sea (799 sites) was obviously greater than those in
655 GOTO, HKDT, and HMKJ (471–581). In the Bayesian clustering, 32.5% of the
656 SAGA specimens had a *q* value of 1.000 in the *A. lischkeana* group, and this
657 low value contrasted with a high value observed in GOTO (84.8%).
658 Analogously, the proportion of specimens with *q* of 1.000 in the *A. japonica*

659 group was high in HKDT (86.7%) and HMKJ (97.1%). These findings may
660 indicate that hybridization occurred in the past in the Ariake Sea, and a weak
661 signature of hybridization (*i.e.*, the presence of low-frequency remnant *A.*
662 *japonica* alleles) is still maintained in the extant *A. lischkeana* population. This
663 hypothesis explains the occurrence of specimens with ambiguous shell
664 characteristics in SAGA (Supporting information S1). Despite the current
665 dominance of *A. lischkeana*-type shells in the Ariake Sea, *A. japonica*-type
666 shells were frequently observed there until the late 1980s (Ito, 2004). For
667 instance, Koga (1992) reported that *A. japonica*-type shells accounted for
668 approximately 20% of his specimens between 1986 and 1989 (as much as 75%
669 depending on his sampling sites). Consistent with the rapid decrease of the *A.*
670 *japonica* resource, the mitochondrial lineage mt-L1 has been observed at a low
671 frequency in the Ariake Sea (Aramaki, 2013; Hashimoto *et al.,* 2018, 2021),
672 although we failed to identify it in SAGA. Thus, *A. japonica* and *A. lischkeana*
673 may have had hybridization opportunities in the Ariake Sea in the past (we
674 return to this issue later).

675 The speciation between *A. japonica* and *A. lischkeana* dates back to the
676 geological epoch from the late Miocene to the Pliocene (Yokogawa, 1996; Liu *et*
677 *al.,* 2011). What drove their speciation remains a mystery, but geographical
678 isolation in sync with the paleo-topographical evolution around the western
679 Pacific margin is a probable cause (Yokogawa, 1996). The distributions of the
680 mitochondrial lineages mt-L1 and mt-L2 are geographically separated even at

present, if not completely, along the coasts of Japan and the East Asian Continent (Liu *et al*., 2011; Hashimoto *et al*., 2021), supporting the hypothesis of their allopatric divergence. Thus, the frequent occurrence of both *A. japonica* and *A. lischkeana* in the Seto Inland Sea and the Ariake Sea (but only in the past) is unusual. Both the Seto Inland and Ariake Seas are semi-enclosed with a shallow geological history. The last glacial maximum coincided with the global sea-level decline (Lambeck *et al*., 2002), and the sea level around the East China Sea was up to 155 m lower than the current level, approximately 15 kya (Emery *et al*., 1971; Wang & Wang, 1980; see Fig. 1 for the contemporary water depth). The last glacial epoch followed a period of ice sheet melting, during which the sea level increased dramatically worldwide (Lambeck *et al*., 2002). The present-day Seto Inland and Ariake Seas were developed associated with circumlittoral retreats caused by the Jomon transgression (Shimoyama, 2000; Shioya *et al*., 2007), which peaked at around 5 kya (Maeda, 2007). The far younger ages of the sea areas relative to the divergence time between *A. japonica* and *A. lischkeana* suggest that the two initially allopatric species colonized there long after their speciation. Thus, their hybridization could have begun through secondary contact, possibly prompted by the confined nature of the sea areas.

4.2 | SNP sites with restricted interspecies gene flow

A common observation in our clustering analyses was that the majority of

703 the KGWA specimens from the Seto Inland Sea formed distinct clusters that
704 conformed to *A. japonica* and *A. lischkeana.* This finding hints that the two
705 species cannot hybridize unrestrainedly, even if they are in sympatry. There
706 should be some underlying reproductive barriers between them despite
707 potential hybridization. Our genomic cline analysis revealed site-specific
708 patterns of introgression, and 10 SNP sites exhibited a significantly low rate of
709 introgression. These results raise the possibility that genomic regions containing
710 the 10 sites are involved in their reproductive isolation.

711 Very few heterozygotes were observed at the 10 SNP sites in KGWA.
712 Widespread null alleles (allele dropout) due to polymorphisms within the SbfI-
713 seqs (Gautier *et al*., 2013) are unlikely to account for these observations, as no
714 missing genotype (homozygote of null alleles) was found in KGWA at all but
715 one site (only the site S0112 had a missing genotype; Table 3). A more
716 plausible explanation for the deficits of heterozygotes is that selection acts
717 against heterozygotes (underdominance or heterozygote disadvantage). Six of
718 the 10 sites were found in functional genes but not protein-coding regions. The
719 four other sites were in intergenic regions, albeit proximal to functional genes.
720 Therefore, the polymorphisms at the sites would not have a functional
721 significance. Instead, we hypothesize that an LD block within which each of the
722 sites is positioned is subjected to selection. Hence, the heterozygote
723 deficiencies at the "proxy" sites can be described by presuming that the
724 corresponding genomic regions of *A. japonica* and *A. lischkeana* are
725 incompatible; that is, hybrids that receive a copy of the genomic regions from
726 each species are at a disadvantage to their survival.

727 The predicted genes related to eight of the 10 SNP sites (Table 4) have

known functions, including the regulations of transcriptional processes (*ETS2*, Wasylyk *et al*., 2002; *ZBTB49*, Jeon *et al*., 2014; *HDGFL2*, Gao *et al*., 2015; *SP3*, Ihn & Trojanowska, 1997) and translational processes (*EIF3B*; Lee *et al*., 2015). A subunit of tRNA-methyltransferase encoded by the *TRMT6* gene controls gene expression through post-transcriptional base modification (Ozanick *et al*., 2005). The nucleotide-sugar transporter gene *UTR7* is essential for maintaining the endoplasmic reticulum in a nematode, and deletion mutants of the gene resulted in a defective larval growth or larval death with intestinal malformation (Dejima *et al*., 2009). The tumor necrosis factor receptor-associated factors including protein encoded by the *TRAF3* gene have signaling functions in the immune system, thereby regulating immune and inflammatory responses (Shi & Sun, 2018). Altogether, these genes are believed to play an important role in maintaining vital functions for the viability of individuals. Despite the functional importance of the predicted genes, however, it is still uncertain if they act as hybrid incompatibility genes (Johnson, 2010), because other genes that are in LD with the SNP sites, but not the predicted genes per se, can be incompatible between the two species (Pavlidis *et al*., 2012).

In SAGA from the Ariake Sea, in which *lischkeana*-type alleles were dominant, homozygote(s) of *japonica*-type allele occurred without a heterozygote at four of the 10 SNP sites. Eight specimens at the most prominent site, S0361, were homozygotes of the *japonica*-type allele (allele *C*; Table 3). If within-species polymorphisms gave rise to the allele *C* in a random mating population of pure *A. lischkeana*, this genotype frequency is very unlikely. Although there were three specimens with missing genotype at this site in SAGA (*i.e*., homozygotes of possible null alleles), the complete absence of

753 heterozygote would not solely be ascribable to the presence of null alleles in a
754 panmictic *A. lischkeana* population (Supporting information S14). Because the
755 chance of ongoing hybridization in the Ariake Sea is limited, we propose that
756 *japonica*-type alleles at the genomic regions related to the four sites were
757 introduced into the *A. lischkeana* population through past hybridization with *A.*
758 *japonica* and that most of the *A. japonica*-derived alleles have been maintained
759 mainly by being homozygous because of underdominance. The survival of an
760 allele at an underdominance locus depends on the frequency of the allele in a
761 population (Altrock *et al*., 2011; Reed *et al*., 2013; Chamber *et al*., 2020). Thus,
762 the low-frequency copies of the corresponding genomic regions from *A.*
763 *japonica* will finally be purged from the *A. lischkeana* population, unless
764 hybridization following the resurgence of the *A. japonica* resource takes place.

765 Some caution is warranted in interpreting our results of genomic cline
766 analysis. Genetic drift can cause a differential introgression across the genome,
767 thereby leading to a false discovery of adaptive introgression (α outlier) and
768 reproductive isolation (β outlier) (Gompert *et al*., 2017 and references therein),
769 although the parameter α is more susceptible to drift than our focal parameter of
770 β (Gompert & Buerkle, 2011). This problem is acute especially in the analysis of
771 hybrid zones with recent and rare hybridization (McFarlane *et al*., 2021).
772 Stochastic drift, however, would not explain at least the absence or deficiency of
773 heterozygotes commonly observed at four of the 10 SNP sites in KGWA and
774 SAGA. Conversely, it remains the possibility that we overlooked genomic
775 regions associated with reproductive isolation. We obtained moderate estimates
776 of β at the 10 SNP sites ($\beta \approx 2$), and the significances of the β values
777 disappeared at the conservative threshold ($qN_{0.990}$). These observations may

indicate that many genomic regions (traits) are subjected to multifarious selection, resulting in a moderate estimate of the cline parameter at the affected sites (Gompert & Buerkle, 2011). Because of the relativistic nature of outlier detection in the genomic cline analysis, we cannot reject the possibility that *β* values at such sites were not qualified as a statistical outlier (Fitzpatrick, 2013). In addition, Gompert and Buerkle (2011) pointed out the difficulty in identifying epistatic loci involved in reproductive isolation (loci with Bateson–Dobzhansky–Muller incompatibility; Johnson, 2010) based on the Bayesian genomic cline analysis.

4.3 | Concluding remarks

In the present study, we revealed that introgression occurs between *A. japonica* and *A. lischkeana*. We also raised the possibility that their interspecies gene flow is incompletely blocked by certain genomic regions (genes) with hybrid incompatibility. Our findings provide an important cue in furthering our understanding of the genetic mechanisms of reproductive isolation between the two species. In this regard, it would be interesting to assess if differential introgression is a common phenomenon in hybridizing free-spawning mollusks. Nevertheless, the present study fell short of validating our hypothesis of reduced fitness in heterozygotes at the genomic regions containing the 10 SNP sites with reduced introgression. This issue should be addressed by evaluating functional constraints around the genomic regions via interspecies crossing experiments.

ACKNOWLEDGEMENTS

We are indebted to Atsushi Ito, Daisuke Ojima, Katsuhisa Eguchi, Masaei Kanematsu, Masanori Tsukuda, Masataka Tomimori, and Yohei Shimizu, who helped collect pen shell specimens analyzed in the present study. Thanks are due to Katsumasa Yamada and Mayumi Kobayashi for their assistance to our visual inspection of shell characteristics. We also express our gratitude to K. Yamada and Haruka Ito for their involvement in measuring shells. Maika Ozawa aided laboratory experiments including DNA extraction and RAD library construction. Many criticisms and recommendations provided by three anonymous reviewers and the editor Cynthia Riginos significantly improved the manuscript. The corresponding author was financially supported by the Japan Society for the Promotion of Science to conduct this study (JSPS KAKENHI 16K07859).

AUTHOR CONTRIBUTIONS

MS conceived this study. KH, MS, MY, and YF devoted their efforts to the collection of specimens. KH, MS, and TS dealt with visual inspection for shell characteristics under the supervision by TS. KH and MS performed mitochondrial DNA sequencing, and MS took on statistical analyses based on the resulting sequence data. MS was involved in the construction of draft genome sequence, prediction of ORFs, and the related bioinformatics. MS also conducted nuclear SNP analyses, including RAD experiments, data processing, and statistical analyses. RN developed an R script for SNP pruning based on physical distance, and prepared input data for cline-model fitting. MS wrote the draft manuscript, and all authors contributed to the completion of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest to disclose

DATA AVAILABILITY STATEMENT

All sequence data obtained in the present study were deposited in the DDBJ/EMBL/GenBank DNA database with accession numbers as follows: mitochondrial *coxI* sequences, LC715263–LC715303 (Supporting information S2); draft genome sequence, BROG01000001–BROG01003391 (BioProject, PRJDB13725; BioSample, SAMD00495586; Sequence Read Archive, DRR380487 and DRR380488); and RAD-derived sequences after demultiplexing (a constant length of 64 bases in the Sequence Read Archive), DRR384710–DRR385045 (BioProject, PRJDB13745; BioSample, SAMD00506414–SAMD00506749). SNP genotype data (vcf format) were submitted to the DRYAD database (doi:10.5061/dryad.6m905qg2m).

BENEFIT-SHARING STATEMENT

Benefits Generated: A research collaboration was developed with researchers and fisherfolk providing specimens, and collaborators are included as co-authors, or are described in the ACKNOWLEDGEMENTS. Benefits from this research accrue from the sharing of our data and results on public databases.

ORCID

Masashi Sekino, https://orcid.org/0000-0003-3603-8310

Kazumasa Hashimoto, https://orcid.org/0000-0002-3319-8899

Reiichiro Nakamichi, https://orcid.org/0000-0001-5789-7689
Masayuki Yamamoto, https://orcid.org/0000-0002-9302-5623
Takenori Sasaki, https://orcid.org/0000-0002-4050-8728

REFERENCES

Altrock PM, Traulsen A, Reed FA (2011) Stability properties of underdominance in finite subdivided populations. *PLoS Computational Biology*, **7**, e1002260. doi: 10.1371/journal.pcbi.1002260

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229. doi:10.1093/genetics/160.3.1217.

Aramaki H (2013) The relationship between an effect of DNA analysis and proportion of pen-shells, *Atrina* spp. in the innermost area of Ariake Bay. *Bulletin of Saga Prefectural Ariake Fisheries Research and Development Center*, **26**, 89–91 (in Japanese).

Asahida T, Kobayashi T, Saitoh K, Nakayama I (1996) Tissue preservation and total DNA extraction form fish stored at ambient temperature using buffers containing high concentration of urea. *Fisheries Science*, **62**, 727–730. doi: 10.2331/fishsci.62.727

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376. doi: 10.1371/journal.pone.0003376

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

878 doi: 10.1146/annurev.es.16.110185.000553

879 Bashir A, Klammer A, Robins WP, Chin CS, Webster D, Paxinos E, Hsu D,
880 Ashby M, Wang S, Peluso P, Sebra R, Sorenson J, Bullard J, Yen J,
881 Valdovino M, Mollova E, Luong K, Lin S, LaMay B, Joshi A, Rowe L, Frace M,
882 Tarr CL, Turnsek M, Davis BM, Kasarskis A, Mekalanos JJ, Waldor MK,
883 Schadt EE (2012) A hybrid approach for the automated finishing of bacterial
884 genomes. *Nature Biotechnology*, **30**, 701–707.
885 doi: 10.1038/nbt.2288

886 Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical
887 and powerful approach to multiple testing. *Journal of the Royal Statistical*
888 *Society Series B*, **57**, 289–300.
889 doi: j.2517-6161.1995.tb02031.x

890 Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F, David P (2003)
891 Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and
892 *M. galloprovincialis. Molecular Ecology*, **12**, 447–461.
893 doi: 10.1046/j.1365-294x.2003.01730.x.

894 Boecklen WJ, Howard D (1997) Genetic analysis of hybrid zones: numbers of
895 markers and power of resolution. *Ecology*, **78**, 2611–2616.
896 doi: 10.1890/0012-9658(1997)078[2611:GAOHZN]2.0.CO;2

897 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for
898 Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
899 doi: 10.1093/bioinformatics/btu170

900 Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on
901 molecular markers. *Molecular Ecology Notes*, **5**, 684–687.
902 doi: 10.1111/j.1471-8286.2005.01011.x

Cantu VA, Sadural J, Edwards RA (2019) PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets. *PeerJ Preprints*, **7**, e27553v1.
doi: 10.7287/peerj.preprints.27553v1

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: building and genotyping loci *de novo* from short-read sequences. *G3 Genes|Genomes|Genetics*, **1**, 171–182.
doi: 10.1534/g3.111.000240

Champer J, Zhao J, Champer SE, Liu J, Messer PW (2020) Population dynamics of underdominance gene drive systems in continuous space. *ACS Synthetic Biology*, **9**, 779–792.
doi:10.1021/acssynbio.9b00452

Chikhi R, Medvedev P (2014) Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics*, **30**, 31–37.
doi: 10.1093/bioinformatics/btt310

Coyne KJ, Burkholder JM, Feldman RA, Hutchins DA, Cary SC (2004) Modified serial analysis of gene expression method for construction of gene expression profiles of microbial eukaryotic species. *Applied and Environmental Microbiology*, **70**, 5298–5304.
doi: 10.1128/AEM.70.9.5298-5304.2004

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
doi: 10.1093/bioinformatics/btr330

Dejima K, Murata D, Mizuguchi S, Nomura KH, Gengyo-Ando K, Mitani S, Kamiyama S, Nishihara S, Nomura K (2009) The ortholog of human solute carrier family 35 member B1 (UDP-galactose transporter-related protein 1) is involved in maintenance of ER homeostasis and essential for larval development in *Caenorhabditis elegans*. *The FASEB Journal*, **23**, 2215–2225.
doi: 10.1096/fj.08-123737

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14,** 652–663.
doi: 10.1111/1755-0998.12209

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
doi: 10.1007/s12686-011-9548-7

Emery KO, Niino H, Sullivan B (1971) Post-Pleistocene levels of East China Sea. In: *Late Cenozoic Glacial Ages* (ed Trekian KK), pp. 381–390. Yale University Press, New Haven.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

953 doi: 10.1111/j.1755-0998.2010.02847.x
954 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure
955 using multilocus genotype data: linked loci and correlated allele frequencies.
956 *Genetics*, 164, 1567–1587.
957 doi: 10.1093/genetics/164.4.1567
958 Fitzpatrick BM (2012) Estimating ancestry and heterozygosity of hybrids using
959 molecular markers. *BMC Evolutionary Biology*, **12**, 131.
960 doi: 10.1186/1471-2148-12-131
961 Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and*
962 *Evolution*, **3**, 1951–1966.
963 doi: 10.1002/ece3.609
964 Gao K, Xu C, Jin X, Wumaier R, Ma J, Peng J, Wang Y, Tang Y, Yu L, Zhang P
965 (2015) HDGF-related protein-2 (HRP-2) acts as an oncogene to promote cell
966 growth in hepatocellular carcinoma. *Biochemical and Biophysical Research*
967 *Communications*, **458**, 849–855.
968 doi: 10.1016/j.bbrc.2015.02.042
969 Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM,
970 Estoup A (2013) The effect of RAD allele dropout on the estimation of genetic
971 variation within and between populations. *Molecular Ecology*, **22**, 3165–3178.
972 doi: 10.1111/mec.12089
973 Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping
974 components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–
975 384.
976 doi: 10.1111/j.1755-0998.2009.02733.x
977 Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines.

Molecular Ecology, **20**, 2111–2127.
doi: 10.1111/j.1365-294X.2011.05074.x

Gompert Z, Buerkle CA (2012) bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.
doi: 10.1111/1755-0998.12009.x

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. *Philosophical Transaction of the Royal Society B*, **367**, 439–450.
doi: 10.1098/rstb.2011.0196

Gompert Z, Mandeville EG, Buerkle CA (2017) Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 207–229.
doi: 10.1146/annurev-ecolsys-110316-022652

Habe T (1977) *Systematics of Mollusca in Japan—Bivalvia and Scaphopoda*. Hokuryukan, Tokyo, Japan (in Japanese).

Haghshenas E, Asghari H, Stoye J, Chauve C, Hach F (2020) HASLR: fast hybrid assembly of long reads. *iScience*, **23**, 101389.
doi: 10.1016/j.isci.2020.101389

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.
doi: 10.1093/jhered/esu033

Hashimoto K, Yamada K, Nagae A, Matsuyama Y (2018) Lineage specific detection of the scaly form of the pen shell *Atrina* spp. by a loop‐mediated isothermal amplification method. *Fisheries Science*, **84**, 837–848.
doi: 10.1007/s12562-018-1231-4

Hashimoto K, Yamada K, Sekino M, Kobayashi M, Sasaki T, Fujinami Y,

Yamamoto M, Choi K-S, Henmi Y (2021) Population genetic structure of the pen shell *Atrina pectinata* sensu lato (Bivalvia: Pinnidae) throughout East Asia. *Regional Studies in Marine Science*, **48**, 102024. doi: 10.1016/j.rsma.2021.102024

Huber M (2010) *Compendium of Bivalves*. ConchBooks, Hackenheim, Germany.

Hurtado NS, Pérez-García C, Morán P, Juan J.Pasantes JJ (2011) Genetic and cytological evidence of hybridization between native *Ruditapes decussatus* and introduced *Ruditapes philippinarum* (Mollusca, Bivalvia, Veneridae) in NW Spain. *Aquaculture*, **311**, 123–128. doi: 10.1016/j.aquaculture.2010.12.015

Huvet A, Fabioux C, McCombie H, Lapègue S, Boudry P (2004) Natural hybridization between genetically differentiated populations of *Crassostrea gigas* and *C. angulata* highlighted by sequence variation in flanking regions of a microsatellite locus. *Marine Ecology Progress Series*, **272**, 141–152. doi:10.3354/meps272141

Ihn H, Trojanowska M (1997) Sp3 is a transcriptional activator of the human *α*2(I) collagen gene. *Nucleic Acids Research*, **25**, 3712–3717. doi: 10.1093/nar/25.18.3712

Ito S (2004) The status and recovery of marine resources in the Ariake Sound. *Bulletin of Saga Prefectural Ariake Fisheries Research and Development Center*, **22**, 69–80 (in Japanese).

Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK (2015) Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, **32**, 1208–1220.

doi: 10.1093/molbev/msv011

Japanese Association of Benthology (2012) *Threatened Animals of Japanese Tidal Flats: Red Data Book of Seashore Benthos*. Tokai University Press, Kanagawa, Japan (in Japanese).

Jeon BN, Kim MK, Yoon JH, Kim MY, An H, Noh HJ, Choi WI, Koh DI, Hur MW (2014) Two ZNF509 (ZBTB49) isoforms induce cell-cycle arrest by activating transcription of *p21*/*CDKN1A* and *RB* upon exposure to genotoxic stress. *Nucleic Acids Research*, **42**, 11447–11461.
doi: 10.1093/nar/gku857

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
doi: 10.1093/bioinformatics/btn129

Johnson NA (2010) Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics*, **26**, 317–326.
doi: 10.1016/j.tig.2010.04.005

Kijewski TK, Zbawicka M, Väinölä R, Wenne R (2006) Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Marine Biology*, **149**, 1371–1385.
doi: 10.1007/s00227-006-0316-2

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915.
doi: 10.1038/s41587-019-0201-4

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of*

Molecular Evolution, **16**,111–120.
doi: 10.1007/BF01731581

Koboldt DC, Zhang Q, Larson DE,1 Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, **22**, 568–576.
doi: 10.1101/gr.129684.111

Koga H (1992) Studies on pen-shells in the Ariake Sea−VI. Differences of form distinguished by texture of shells, and its distribution. *Bulletin of Saga Prefectural Ariake Fisheries Research and Development Center*, **14**, 9−24 (in Japanese with English abstract).

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, **35**,1547–1549.
doi: 10.1093/molbev/msy096.

Kurozumi T (2017) Family Pinnidae. In: *Marine Mollusks in Japan*, 2nd edition (ed Okutani T), pp. 1185–1186, pl. 486. Tokai University Press, Kanagawa, Japan.

Lambeck K, Esat TM, Potter E-K (2002) Links between climate and sea levels for the past three million years. *Nature*, **419**, 199–206.
doi: 10.1038/nature01089

Lee ASY, Kranzusch PJ, Cate JHD (2015) eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*, **522**, 111–114.
doi: 10.1038/nature14267

Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.
doi: 10.1111/2041-210X.12410

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
doi: 10.1093/bioinformatics/btp352

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **26**, 589–595.
doi: 10.1093/bioinformatics/btp698

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.
doi: 10.1093/bioinformatics/btr509

Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, **30**, 2843–2851.
doi: 10.1093/bioinformatics/btu356

Li YL, Liu JX (2018) STRUCTURESELECTOR: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, **18**, 176–177.
doi: 10.1111/1755-0998.12719

Liqing Z, Xuemei W, Aiguo Y, Biao W, Xiujun X, Zhihong L, Siqing C, Dan Z (2020) Abnormal karyotype of pen shell (*Atrina pectinata*) during its early embryonic development in late breeding season. *Journal of Ocean University*

of China, **19**, 902–910.
doi: 10.1007/s11802-020-4280-0

Liu J, Li Q, Kong L, Zheng X (2011) Cryptic diversity in the pen shell *Atrina pectinata* (Bivalvia: Pinnidae): high divergence and hybridization revealed by molecular and morphological data. *Molecular Ecology*, **20**, 4332–4345.
doi: 10.1111/j.1365-294X.2011.05275.x

Maeda Y (2007) Sea level rise from 19 ka to 5 ka cal. BP (Jyomon transgression). *Transactions, Japanese Geomorphological Union*, **28**, 349–363 (in Japanese with English abstract).

Martin BT, Chafin TK, Douglas MR, Douglas ME (2021) ClineHelpR: an R package for genomic cline outlier detection and visualization. *BMC Bioinformatics*, **22**, 501.
doi: 10.1186/s12859-021-04423-x

Matsumoto S, Ojima D, Kanazawa K (2019) Temporal changes of maturation in broodstock parents of pen shell. In: *Guidebook for Seed Production and Aquaculture of Pen Shell* (eds Maeda Y and Kanematsu M), pp. 21–25. Fisheries Research and Education Agency, Yokohama, Japan (in Japanese).

McFarlane SE, Pemberton JM (2019) Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology and Evolution*, **34**, 315−326.
doi: 10.1016/j.tree.2018.12.013

McFarlane SE, Senn HV, Smith SL, Pemberton JM (2021) Locus-specific introgression in young hybrid swarms: Drift may dominate selection. *Molecular Ecology*, **30**, 2104−2115.
doi: 10.1111/mec.15862

Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, **6**, 971–973.
doi: 10.1111/j.1471-8286.2006.01433.x

Nishimura O, Hara Y, Kuraku S (2017) gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*, **33**, 3635–3637.
doi: 10.1093/bioinformatics/btx445

Okutani T (1994) Mollusks. In: *Fishery Invertebrates II—Description of Useful and Harmful Species* (ed Okutani T), pp. 1–192. Kouseisha kouseikaku, Tokyo, Japan (in Japanese).

Ozanick S, Krecic A, Andersland J, Anderson JT (2005) The bipartite structure of the tRNA m^1A58 methyltransferase from *S. cerevisiae* is conserved in humans. *RNA*, **11**, 1281–1290.
doi: 10.1261/rna.5040605

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genome scans. *Molecular Biology and Evolution*, **29**, 3237–3248.
doi: 10.1093/molbev/mss136

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.
doi: 10.1111/j.1755-0998.2010.02883.x

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

doi: 10.1111/mec.13557

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**, 290–295. doi: 10.1038/nbt.3122

Pfenninger M, Reinhardt F, Streit B (2002) Evidence for cryptic hybridization between different evolutionarylineages of the invasive clam genus Corbicula (Veneroida, Bivalvia). *Journal of Evolutionary Biology*, **15**, 818–829. doi: 10.1046/j.1420-9101.2002.00440.x

Pike N (2011) Using false discovery rates for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution*, **2**, 278–282. doi: 10.1111/j.2041-210X.2010.00061.x

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959. doi: 10.1093/genetics/155.2.945

Puechmaille SJ (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, **16**, 608–627. doi: 10.1111/1755-0998.12512

R Core Team (2016) R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reed FA, Traulsen A, Altrock PM (2013) Underdominance. In: *Brenner's Encyclopedia of Genetics*, volume 7, 2nd edition (eds Maloy S, Hughes K), pp. 2093–2094. Academic Press, Cambridge, MA.

Riginos C, Cunningham CW (2005) Local adaptation and species segregation

in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, **14**, 381–400. doi:10.1111/j.1365-294X.2004.02379.x

Rosenberg NA (2004). Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138. doi:10.1046/j.1471-8286.2003.00566.x

Rosewater J (1961) The family Pinnidae in the Indo-Pacific. *Indo-Pacific Mollusca*, **1**, 175–226.

Sambrook J, Fritsch EF, Manitaris T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Plainview, NY

Schultz PWW, Huber M (2013) Revision of the worldwide recent Pinnidae and some remarks on fossil European Pinnidae. *Acta Conchyliorum*, **13**, 1–164.

Sekino M, Nakamichi R, Iwasaki Y, Tanabe AS, Fujiwara A, Yasuike M, Shiraishi M, Saitoh K (2016) A new resource of single nucleotide polymorphisms in the Japanese eel *Anguilla japonica* derived from restriction site-associated DNA. *Ichthyological Research*, **63**, 496–504. doi: 10.1007/s10228-016-0518-7

Shi JH, Sun SC (2018) Tumor necrosis factor receptor-associated factor regulation of nuclear factor κB and mitogen-activated protein kinase pathways. *Frontiers in Immunology*, **9**, 1849. doi: 10.3389/fimmu.2018.01849

Shioya F, Mii T, Iwamoto N, Inouchi Y (2007) Late Pleistocene to Holocene variations in sea conditions within the Seto Inland Sea, offshore Matsuyama City, Japan. *Earth Science*, **61**, 103–115. doi: 10.15080/agcjchikyukagaku.61.2_103

Shimoyama S (2000) The geological history of Ariake Sea and occurrence of Ariake-endemic species. In: *Life in the Ariake Sea: Biodiversity in Tidal Flats and Estuaries* (ed Satou M), pp. 37–48. Kaiyusha, Tokyo, Japan (in Japanese).

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.
doi: 10.1093/bioinformatics/btv351

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. *Genetics*, **132**, 619–633.
doi: 10.1093/genetics/132.2.619

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.
doi: 10.1093/nar/22.22.4673

Wang J, Wang P (1980) Relationship between sea-level changes and climatic fluctuations in East China since late Pleistocene. *Acta Geographica Sinica*, **35**, 299–312 (in Chinese with English abstract).
doi: 10.11821/xb198004003

Wasylyk C, Schlumberger SE, Criqui-Filipe P, Wasylyk B (2002) Sp100 interacts with ETS-1 and stimulates its transcriptional activity. *Molecular and Cellular Biology*, **22**, 2687–2702.
doi: 10.1128/MCB.22.8.2687-2702.2002

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, New York, NY.

Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–893.
doi: 10.1086/429864

Wu C-I, Ting C-T (2004) Genes and speciation. *Nature Review Genetics*, **5**, 114–122.
doi: 10.1038/nrg1269

Xue D, Wang H, Zhang T, Gao Y, Zhang S, Xu F (2012) Morphological and genetic identification of the validity of the species *Atrina Chinensis* (Bivalvia: Pinnidae). *Journal of Shellfish Research*, **31**, 739–747.
doi: 10.2983/035.031.0318

Xue D, Wang H, Zhang T (2021) Phylogeography and taxonomic revision of the pen shell *Atrina pectinata* species complex in the South China Sea. *Frontiers in Marine Science*, **8**, 753553.
doi: 10.3389/fmars.2021.753553

Yokogawa K (1996) Genetic divergence in two forms of pen shell *Atrina pectinata*. *VENUS*, **55**, 25–39 (in Japanese with English abstract).
doi: 10.18941/venusjjm.55.1_25

Zhang Y, Zhang Y, Li J, Ziniu Yu Z (2017) Morphological and molecular evidence of hybridization events between two congeneric oysters, *Crassostrea hongkongensis* and *C. ariakensis*, in southern China. *Journal of Molluscan Studies*, **83**, 129–131.
doi: 10.1093/mollus/eyw035

Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J,

Salzberg SL (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Research*, **27**, 787–792.
doi: 10.1101/gr.213405.116

Zimin AV, Salzberg SL (2020) The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Computational Biology*, **16**, e1007981.
doi: 10.1371/journal.pcbi.1007981

TABLE 1: Population samples of pen shells

Abbr. area/sample name[*1]	Sample size	Year of collection	Shell length (mm; mean ± s.d.)
HKDT	45	2016	262.1 ± 20.0
HMKJ	35	2014, 2017[*3]	260.3 ± 8.9
KGWA	202 (183)[*2]	2015–2017	198.8 ± 30.8
SAGA	40	2009, 2018[*4]	168.6 ± 30.9[*5]
GOTO	33	2016	303.8 ± 24.6

[*1]For the geographical location of each population sample, see Figure 1.

[*2]Of the 202 specimens, 19 were omitted from statistical analyses (see text). Mean shell length for KGWA was calculated based on data from the remaining 183

[*3]Five specimens were sampled in 2014.

[*4]Because of the difficulty in collecting specimens in the SAGA area, 11 archival specimens, which were sampled in 2009 and kept frozen by Saga Prefectural Ariake Fisheries Research and Development Center, were included.

[*5]Shells of one specimen in SAGA were badly damaged and not available for measurment of shell length. Mean shell length for SAGA was calculated without the specimen.

TABLE 2: Summary statistics of nuclear polymorphisms estimated for each population sample

	Population sample					
	HKDT	HMKJ	KGL1	KGL2	SAGA	GOTO
Sample size	45	35	112	71	40	33
Mitochondrial lineage	mt-L1	mt-L1	mt-L1	mt-L2	mt-L2	mt-L2
Number of polymoriphic sites[*1]	581 (0.396)	507 (0.345)	1,467 (0.999)	1,450 (0.987)	799 (0.544)	471 (0.321)
Mean H_O (± s.d.)[*2]	0.237 (± 0.169)	0.255 (± 0.157)	0.162 (± 0.114)	0.168 (± 0.129)	0.184 (± 0.179)	0.261 (± 0.172)
Mean H_E (± s.d.)[*2]	0.243 (± 0.166)	0.273 (± 0.157)	0.230 (± 0.114)	0.218 (± 0.126)	0.196 (± 0.178)	0.266 (± 0.169)
Ho/H_E	0.975	0.934	0.704	0.771	0.939	0.981
Number of non-HWE sites[*3]	--	43 (0.085)	922 (0.628)	479 (0.330)	75 (0.094)	--

[*1]Proportion of variable sites, which was clculated by dividing the number of variable sites in each sample by the number of all sites (1,469), is given in parenthesis.

[*2]Based on variable sites in each population sample.

[*3] HWE: Hardy-Weinberg equilibrium. A significant deviation of genotype frequencies from expectations under HWE was determined at the uncorrected threshold probability of 0.05 in an exact test. The proportion of non-HWE sites (in parenthesis) was calculated by dividing the number of non-HWE sites by the number of all variable sites in each sample. Sites with significant HWE departure in HKDT and GOTO were removed during the process of SNP filtering (see text).

TABLE 3: Genotype frequencies at SNP sites with a low rate of introgression

			Cline parameter[*3]			Genotype frequency[*4]						
Site ID	Contig ID[*1]	Pos[*2]	α	β	Genotype	HKDT	HMKJ	GOTO	SAGA	KGWA (all)	KGWA (KGL1)	KGWA (KGL2)
S0002	AJ_contig0002	2,380,891	-1.78*	2.17*	*A/A*	--	--	33	40	89	20	69
	L = 5,812,167				*A/G*	--	--	--	--	1	1	--
					G/G	45	35	--	--	93	91	2
S0033	AJ_contig0033	69,499	-1.66*	1.88*	*C/C*	--	--	33	40	88	21	67
	L = 2,341,375				*A/C*	1	--	--	--	3	1	2
					A/A	44	35	--	--	92	90	2
S0094	AJ_contig0094	775,917	0.27	2.05*	*T/T*	--	--	33	40	66	7	59
	L = 1,644,075				*A/T*	--	--	--	--	1	--	1
					A/A	45	35	--	--	116	105	11
S0112	AJ_contig0112	1,161,175	0.18	2.10*	*G/G*	--	--	33	40	66	6	60
	L = 1,488,041				*A/G*	--	--	--	--	2	2	--
					A/A	45	35	--	--	114	103	11
S0133	AJ_contig0133	214,353	0.18	1.96*	*C/C*	--	--	33	40	65	7	58
	L = 1,354,993				*A/C*	--	--	--	--	2	1	1
					A/A	45	35	--	--	116	104	12
S0227	AJ_contig0227	391,931	0.57	2.06*	*T/T*	--	--	33	39	65	7	58
	L = 993,056				*C/T*	--	--	--	--	1	--	1
					C/C	45	35	--	1	117	105	12
S0289	AJ_contig0289	448,758	0.59	2.01*	*A/A*	--	--	33	39	64	5	59
	L = 857,080				*A/G*	--	--	--	--	1	--	1
					G/G	45	35	--	1	118	107	11
S0361	AJ_contig0361	619,065	0.44	2.19*	*T/T*	--	--	31	29	65	7	58
	L = 728,703				*C/T*	--	--	1	--	--	--	--
					C/C	45	35	--	8	118	105	13
S0861	AJ_contig0861	170,852	0.59	2.07*	*T/T*	--	--	33	39	65	7	58
	L = 256,892				*G/T*	--	--	--	--	1	--	1
					G/G	45	35	--	1	117	105	12
S0939	AJ_contig0939	192,946	0.51	2.03*	*A/A*	--	--	33	40	63	6	57
	L = 209,638				*A/G*	--	--	--	--	2	1	1
					G/G	45	35	--	--	118	105	13

[*1]Contig IDs correspond to those in the *Atrina japonica* reference genome. The length of each contig (bases) is provided (*L*).

[*2]Position of each SNP site in the corresponding contig.

[*3]Values of cline parameters determined as an outlier in an admixed sample (KGWA) are denoted by adding an asterisk ($qN_{0.975}$ threshold; see text).

[*4]There was a missing genotype at the site S0112 in KGL1. Additionally, genotyping at the site S0361 failed in one specimen in GOTO and three in SAGA.

TABLE 4: Predicted genes associated with SNP sites with a low rate of introgression

Site ID	Contig ID[*1]	Within Gene?[*2]	Gene (nearest gene) position[*3]	Predicted gene	Accession	*E*-value
S0002	AJ_contig0002	No d = 12,767	2,393,658–2,400,125 (-) L_A = 283 (Full)	*ETS2*: Protein C-ets-2 (*Crassostrea gigas*)	XP_034317332.1	5.0×10^{-62}
S0033	AJ_contig0033	Yes	61,995–88,820 (-) L_A = 441	*EIF3B*: Eukaryotic translation initiation factor 3 subunit B (*Crassostrea gigas*)	XP_011432969.2	0.000
S0094	AJ_contig0094	Yes	744,691–800,335 (+) L_A = 741 (Full)	*ZBTB49*: Zinc finger and BTB domain-containing protein 49-like (*Crassostrea virginica*)	XP_022317845.1	0.000
S0112	AJ_contig0112	Yes	1,117,392–1,196,271 (+) L_A = 862 (Full)	Hypothetical predicted protein (*Mytilus galloprovincialis*)	VDH95928.1	0.000
S0133	AJ_contig0133	No d = 777	215,130–215,476 (-) L_A = 116	Uncharacterized protein LOC105328369 isoform X2 (*Crassostrea gigas*)	XP_011427518.2	1.0×10^{-6}
S0227	AJ_contig0227	Yes	384,297–403,225 (+) L_A = 477 (Full)	*TRMT6*: tRNA (adenine(58)-N(1))-methyltransferase non-catalytic subunit TRM6 (*Crassostrea gigas*)	XP_011415154.2	0.000
S0289	AJ_contig0289	Yes	440,432–452,583 (-) L_A = 383 (Full)	*UTR7*: UDP-galactose/UDP-glucose transporter 7-like (*Crassostrea gigas*)	XP_034314452.1	3.0×10^{-173}
S0361	AJ_contig0361	No d = 3,100	601,848–615,965 (+) L_A = 845	*SP3*: Transcription factor Sp3 (*Crassostrea gigas*)	XP_011417101.2	0.000
S0861	AJ_contig0861	Yes	153,542–189,946 (-) L_A = 564 (Full)	*TRAF3*: Tumor necrosis factor receptor-associated factor 3 (*Pinctada fucata*)	AFL03408.1	0.000
S0939	AJ_contig0939	No d = 3,515	172,015-189,431 (+) L_A = 427	*HDGFL2*: Hepatoma-derived growth factor-related protein 2-like isoform X2 (*Crassostrea virginica*)	XP_022289617.1	2.0×10^{-28}

[*1]Contig IDs correspond to those in the *Atrina japonica* reference genome.

[*2]For each intergenic SNP site, the physical distance (bases) to the nearest gene is provided (d).

[*3]Position of corresponding gene in the contig of the *A. japonica* reference genome (+, sense; -, antisense). The sequence length of translated amino acids is given (L_A). Full: genes in which both the 5'UTR and 3'UTR sequences were available.

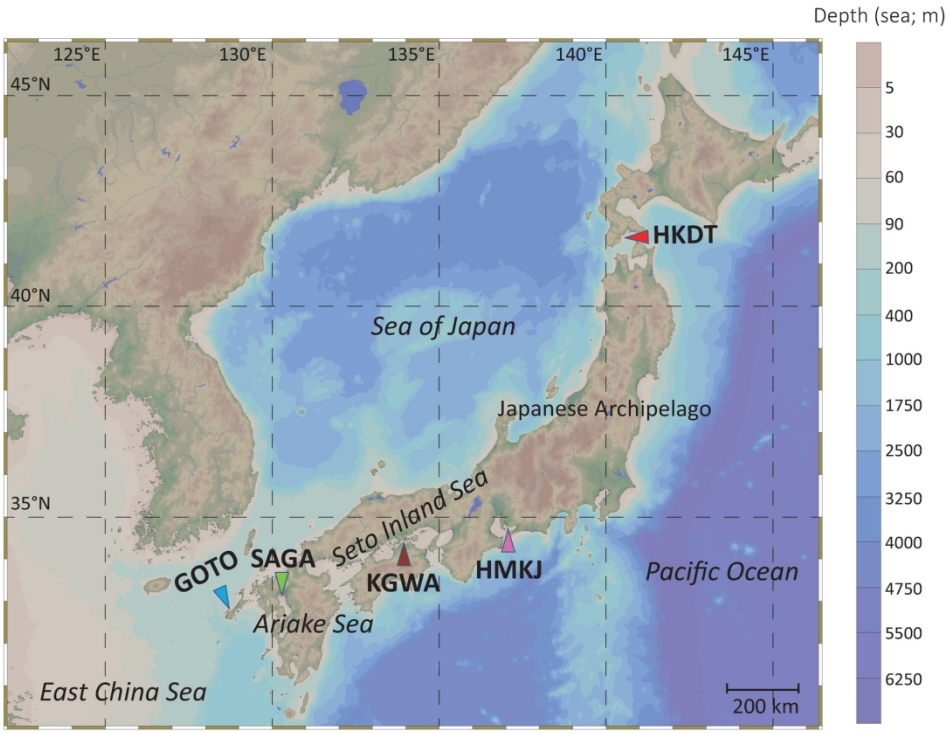

FIGURE 1. Geographical locations of sampling sites.

Triangles with abbreviated area/sample names indicate sampling sites. The map was drawn with Ocean Data View version 5.6.2 (https://odv.awi.de/; last accessed May 11, 2022).

Figure 1

174x173mm (300 x 300 DPI)

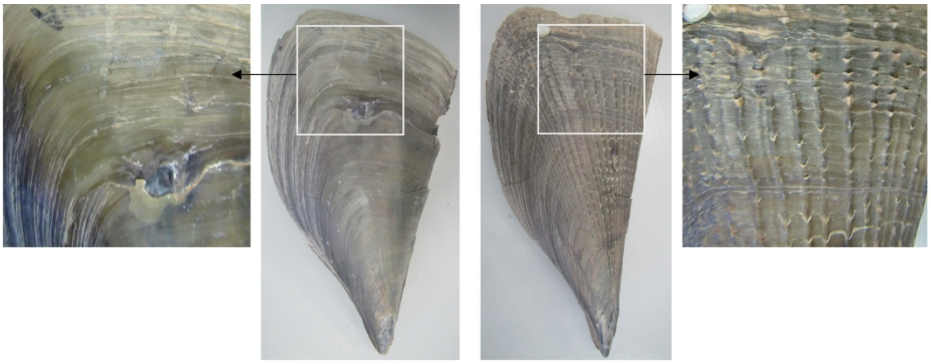

FIGURE 2. Shells of *Atrina pectinata* non-scaly and scaly forms.
For each morphotype, the outer right valve is presented (shell length: non-scaly form, 280 mm; and scaly form, 265 mm). The present study presumes the non-scaly and scaly forms as separate species (non-scaly form, *Atrina japonica*; and scaly form, *Atrina lischkeana*). The presence of fine ribs and spines characterizes the shells of *A. lischkeana*.

Figure 2

167x122mm (300 x 300 DPI)

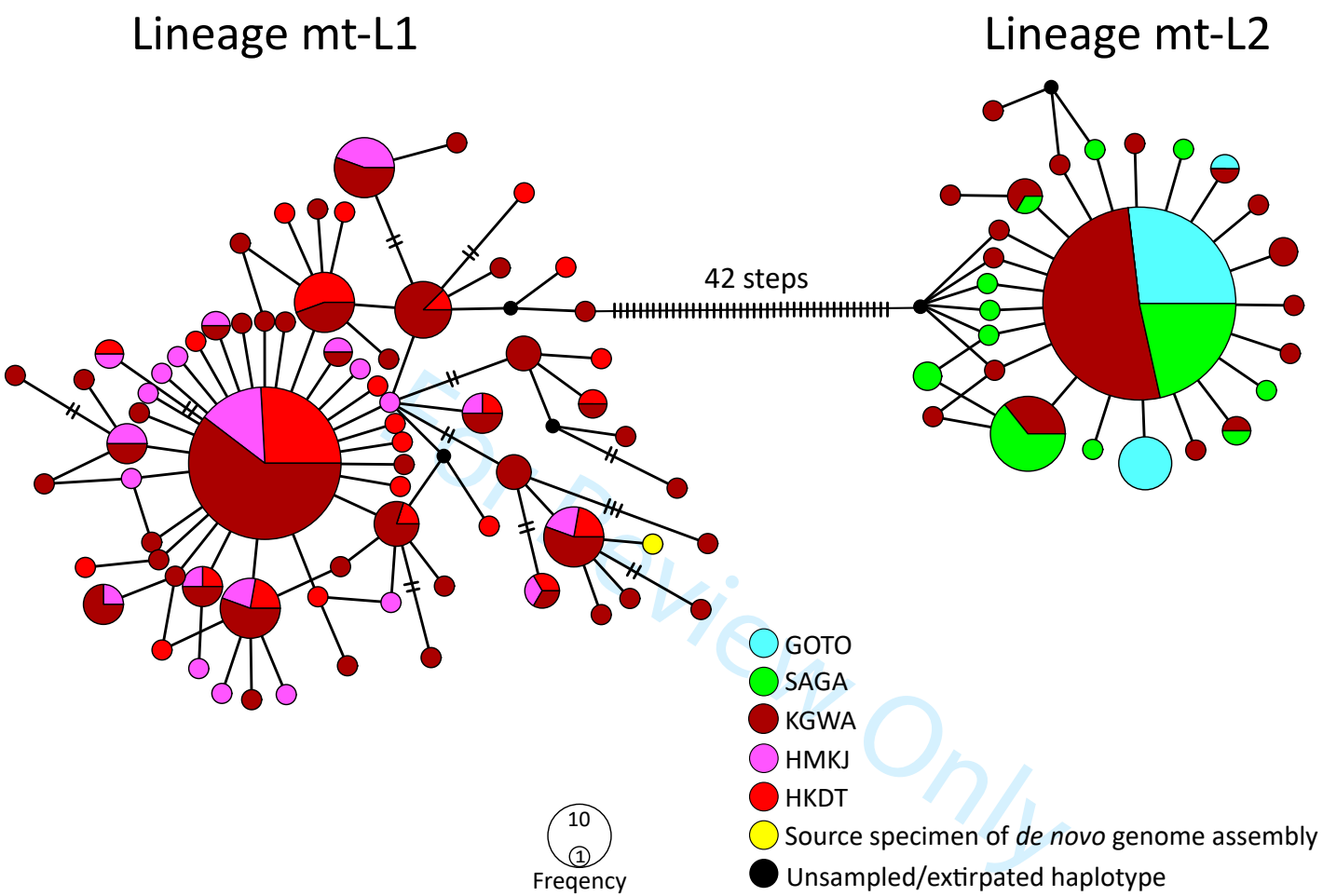

FIGURE 3. Haplotype network constructed based on the nucleotide sequences of the mitochondrial cytochrome *c* oxidase subunit I gene.

Each circle represents a haplotype, and the size of each circle reflects the frequency of the haplotype. The number of bars on each branch corresponds to the number of mutational steps between haplotypes (no bar is given for single-step mutations).

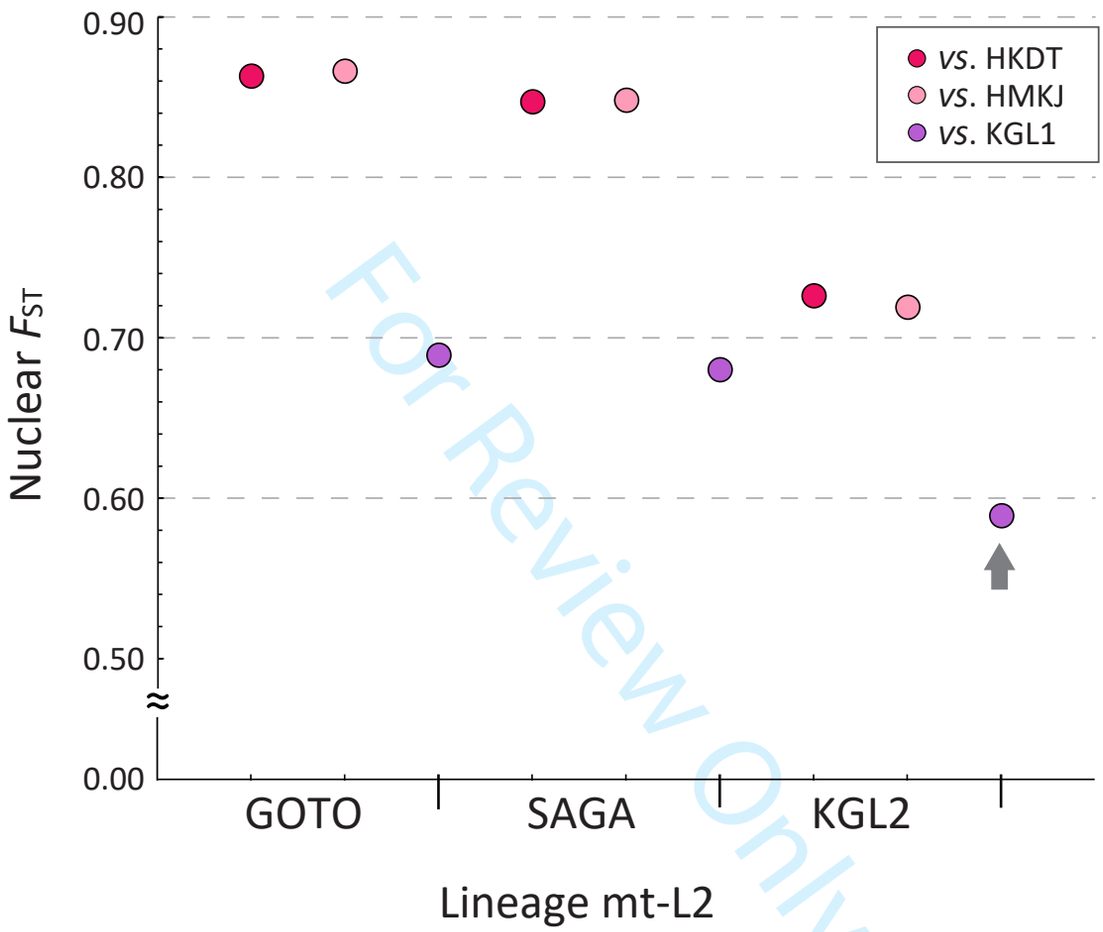

FIGURE 4. Nuclear F_{ST} estimated for pairs of population samples between mitochondrial lineages.

Vertical axis, nuclear F_{ST} (ncF_{ST}). Horizontal axis, population samples of mitochondrial lineage mt-L2. Values of ncF_{ST} estimated for each mt-L2 sample against mt-L1 samples (HKDT, HMKJ, and KGL1) are plotted. A gray arrow indicates ncF_{ST} between sympatric population samples.

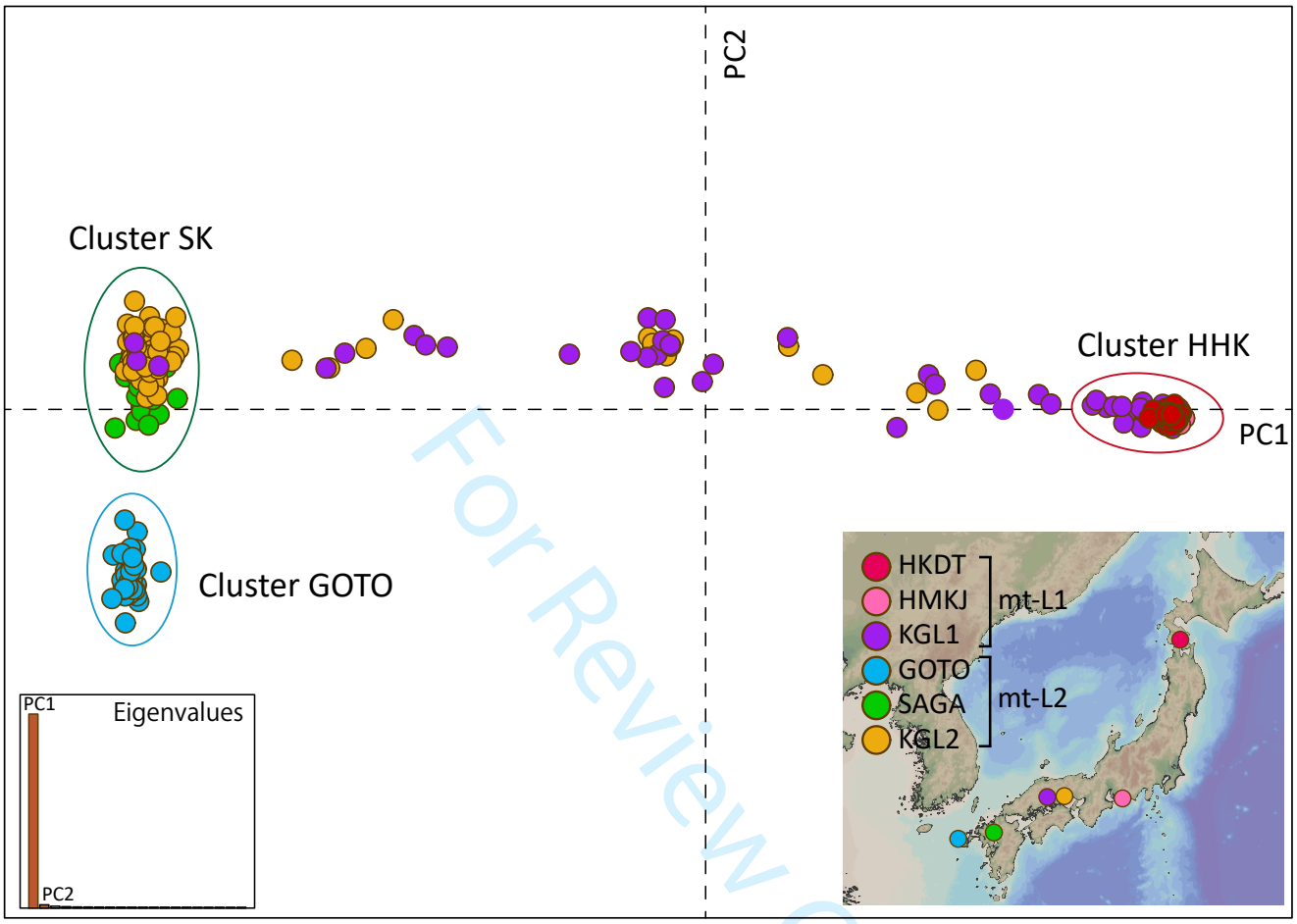

FIGURE 5. Planar plotting of specimens based on principal component analysis for nuclear SNP genotypes.

The first component (PC1; horizontal axis) explained 61.8% of total variances, and the second component (PC2; vertical axis), 1.0%. According to the configurations of specimens, three conspicuous clusters were defined (clusters HHK, SK, and GOTO). Note that all the HMKJ specimens are tightly clumped together in the cluster HHK, thereby being invisible in the figure.

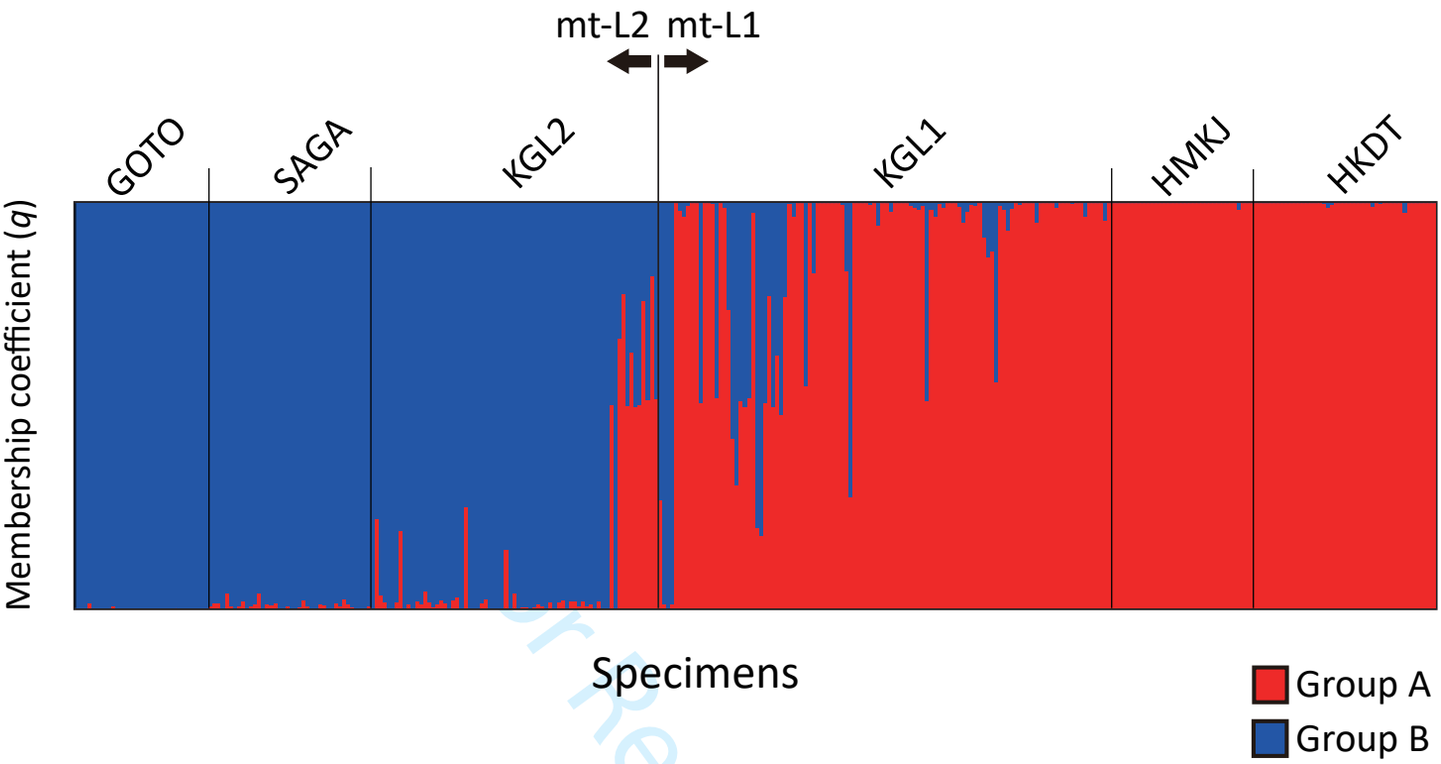

FIGURE 6. Genomic admixture estiamted based on Bayesian model-based individual clustering. Distributions of *q* values within specimens at *K* (number of clusters) of two are given. Group A (red-colored) was defined as a group of *Atrina japonica*, and group B (dark blue), as a group of *Atrina lischkeana* (see text).

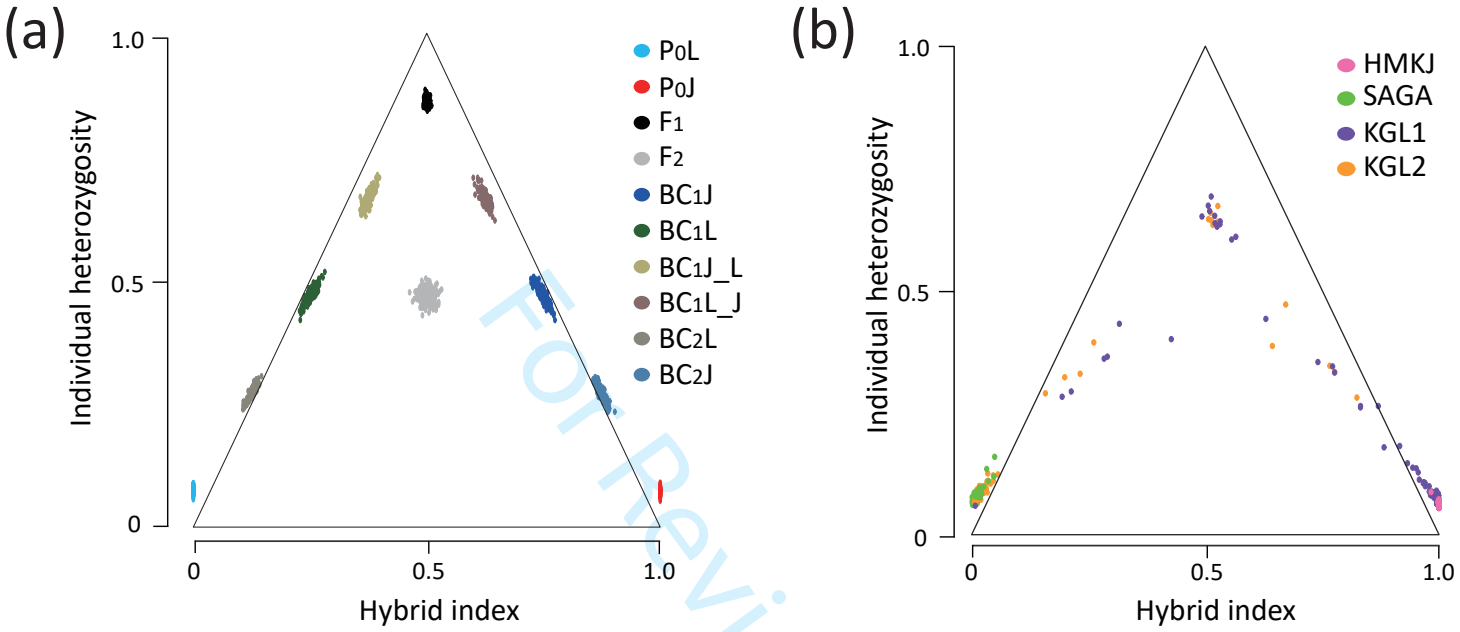

FIGURE 6. Triangle plots showing relationships between individual heterozygosity and hybrid index. Vertical axis, individual heterozygosity (H_{ind}). Horizontal axis, hybrid index (*HI*). Both H_{ind} and *HI* were estimated with 1,066 ancestry-infromative SNP sites (see text). Pure *Atrina lischkeana* and *Atrina iaponica* are presumed to have an *HI* of 0.000 and 1.000, respectively. (a) plot for simulated individuals of pure species (P_0L, *A. lischkeana*; and P_0J, *A. japonica*) and early generations of hybrids (eight hybrid classes; see text). For each class, 200 simulated individuals are plotted. (b) plot for specimens from four population samples (HMKJ, SAGA, KGL1, and KGL2), which had the possibility of containing hybrids.

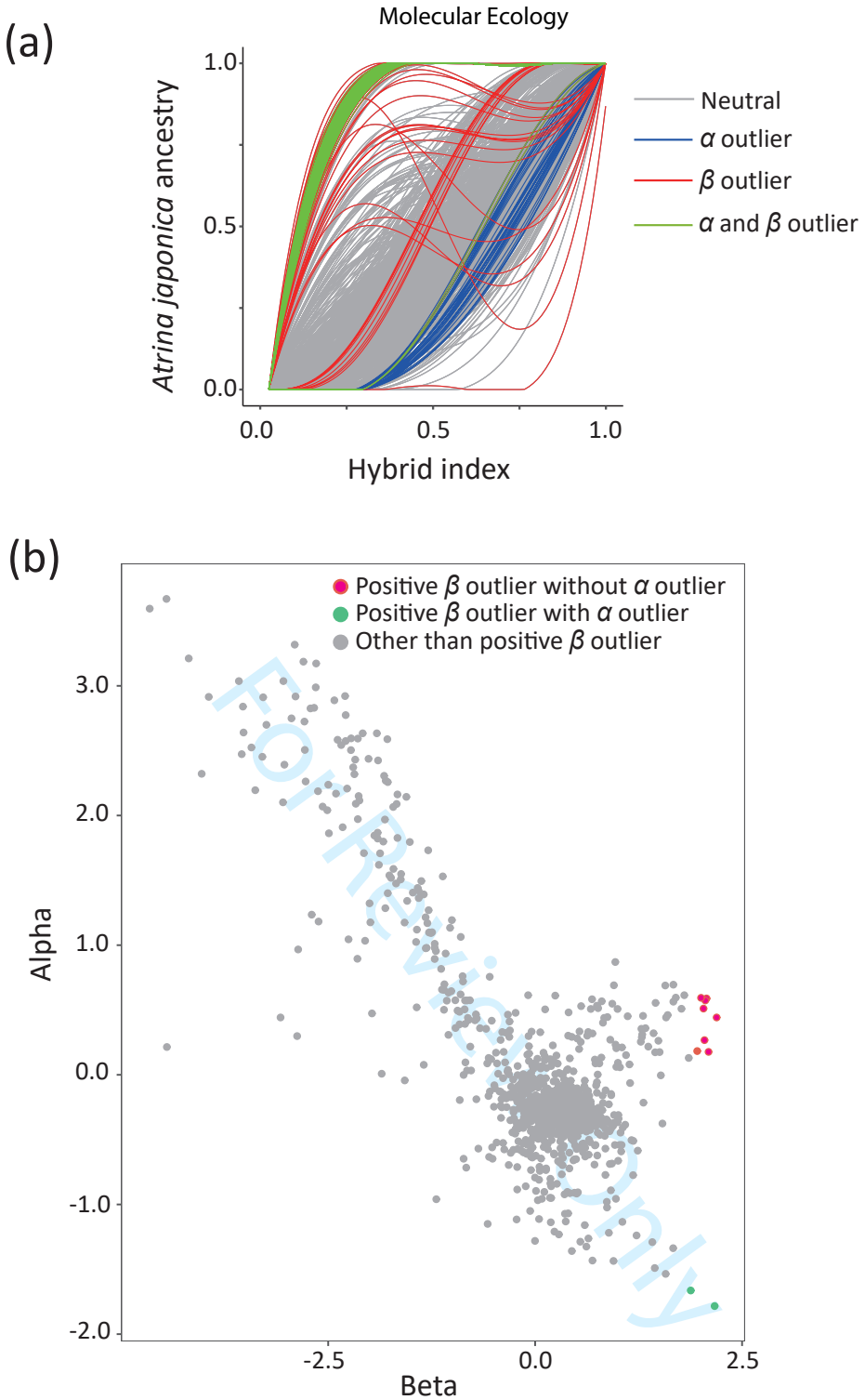

FIGURE 8. Representations of the results of Bayesian genomic cline analysis.

(a) genomic clines at 1,066 ancestry-informative SNP sites (see text) as represented by the probability of *Atrina Japonica* ancestry (vertical axis) and hybrid index (horizontal axis). Classifications of clines are indicated by different colors as follows: clines for sites at which their cline parameters (α and β) were consistent with the neutral expectations (gray); those for sites that had a significantly high or low estimate of α (α outlier; blue) or β (β outlier; orange); and those for sites with both α and β outliers (green). The figure was drawn with ClineHelpR (Martin *et al*., 2021). (b) distribution of cline parameters. Values of α are plotted along the vertical axis, and β values, along the horizontal axis. Ten sites with positive β outlier were indicated by colored dots as follows: those that met the neutral expectation concerning α (red); and those with α outlier (light green). Gray dots represent SNP sites without positive β outlier.