Supplementary Materials for

The genetic legacy of a global marine invader

Erik E. Sotka*[1],Ryan B. Carnegie[2], James T. Carlton[3], Lucia Couceiro[4], Jeffrey A. Crooks[5], Hikaru Endo[6], Hilary Hayford[7], Masakazu Hori[8], Mitsunobu Kamiya[9], Gen Kanaya[10], Judith Kochmann[11], Kun-Seop Lee[12], Lauren Lees[13], Hannah Miller[1], Masahiro Nakaoka[14], Eric Pante[15], Jennifer L. Ruesink[16], Evangelina Schwindt[17], Åsa Strand[18], Richard B. Taylor[19], Ryuta Terada[6], Martin Thiel[20], Takefumi Yorisue[21,22], Danielle Zacherl[23], Allan E. Strand*[1]

* Corresponding author (SotkaE@cofc.edu; StrandA@cofc.edu)

The file includes:

Appendix 1. Library Preparation, Sequencing and Initial Processing

To generate libraries for RADseq (restriction-site associated DNA sequencing), we digested gDNA with two restriction enzymes, *Eco*RI and *Mse*I, and ligated adaptors containing unique 8 to 10 bp barcodes to the digested DNA of each individual. The products were then PCR amplified in two independent reactions with standard Illumina primers. All amplicons were pooled and shipped to the University of Texas Genomic Sequencing and Analysis Facility or the Tufts University Core Facility, which used Pippin Prep® to isolate the 300 – 450 bp fraction. This fraction was then single-read sequenced (100 or 150 basepairs) with Illumina Novaseq, HiSeq 2500 and/or HiSeq 4000 machines. We used custom scripts to demultiplex into sample-specific FASTQ-formatted files. When individuals were run within multiple libraries or sequencing runs, we combined all data per individual into a single FASTQ file per individual.

Reads were aligned to the *Magallana gigas* genome (1) with bwa mem (2) and samtools/bcftools 1.9 (3) using default settings. Reads that aligned to multiple sites in the genome (i.e., "XA" tag in sam-formatted files) were filtered out to attempt to control for paralogy. We then used angsd (4) on 738 samples to identify loci with a minimum allele frequency (MAF) greater than 1%, at least one read in 88% of individuals, minimum mapQ = 30 and minQ = 20. We then used angsd to filter out SNPs that were out of Hardy-Weinberg Equillibrium in 2 or more populations. This yielded a set of 738 individuals and genotype likelihoods (using the SAMtools model) at 298K SNPs, with an average of 11.1 reads per SNP-by-individual combination. We used bcftools to generate genotype calls using this SNP set, keeping SNPs that were called in >95% of individuals and with QUAL >500. We entered the vcf-formatted file using the R::vcfR library (5) and removed individuals with >5% of SNPs missing and loci with no genotype calls. This yielded 726 individuals and 7046 loci. These two datasets (genotype likelihoods and genotype calls) were then used for further analyses.

We aligned the first 500K reads from individual FASTQ files to a representative mitochondrial genome using bowtie2 (6). These reads were BLASTed against two mitochondrial loci that occurred within the mitochondrial genomes of 11 *Crassostrea* species uploaded to GenBank (NADH dehydrogenase subunit 4 at 13182-13267 and a non-coding region at 7985-8069 of *M. gigas* mitochondrial genome KJ855245). Nearly all samples had 98-100% match out of 85 bp to either or both *M. gigas* mitochondrial loci. The exceptions were *C. angulata* mitochondria within two individuals in Boursefranc, France (BOU), and *C. nippona* mitochondria in two samples from Iwaki City, Japan (SAM).

To confirm these patterns with a phylogenetic approach, we used bowtie2 to align our raw sequences to a *M. gigas* mitochondrial genome (NC_001276.1) and used bcftools 'mpileup' and 'call' to identify variants and bcftools 'consensus' to generate sequences at the NADH and a non-coding region for each individual. For the 606 individuals for which both regions sequenced, we assessed which haplotypes were unique and their frequencies with custom R scripts. We then aligned these unique haplotypes and the two mitochondrial loci for 8 species with muscle (Edgar 2004) and generated a neighbor-joining tree with K80 distances and 1000 bootstrap replicates using the R package, ape (7); Figure S2). All samples aligned with *Magallana gigas*, except for Haplotype 3, a *C. angulata* mitochondrial haplotype within the two individuals at BOU indicated by the BLAST search. Haplotype 10 is the most common haplotype (98% of 606 individuals).

We note that 20 individuals at a Japanese west coast population (MUN) had *M. gigas* mtDNA but contained reads poorly mapped (<10% of reads) to the *C gigas* genome, indicating introgression of gigas mtDNA into non-gigas nuclear genome. These individuals were removed from further analyses.

Appendix 2. Descriptions of the population genetics and evidence for vector hypotheses for 14 introduced species from Japan. Key: Hok = Hokkaido, Miy = Miyagi, Tok = Tokyo, Seto = Seto Inland Sea, Kag = Kagoshima.

Acanthogobius flavimanus (Chordata; yellowfin goby)

Population genetics dataset. We used mitochondrial haplotypes and their frequency reported in Table 2 of Hirase et al. (8). There were samples from five regions in Japan (Hok, Miy, Tok, Seto, Kag) and from western North America.

Vector hypotheses: Shipping. The yellowfin goby is thought to be introduced to San Francisco Bay in the early 1960s via ballast water transport of eggs or larvae with a "probable likelihood" or pipes of ships (see review in (9). See also https://invasions.si.edu/nemesis/calnemo/species_summary/171882

Batillaria attramentaria (Mollusca; Japanese False Cerith)

Population genetics dataset. We used mitochondrial haplotypes and their frequencies visualized in pie charts in Figure 1 of Miura et al. (9). There were samples from four regions in Japan (Miy, Tok, Seto, Kag) and from western North America.

Vector hypotheses: Oyster *Batillaria* was introduced with Japanese oysters to Washington; it was introduced to Bodega Harbor on scientists' boots or with scientific equipment; how it was introduced to the San Diego area is uncertain, but no oysters were transplanted to the known sites in the San Diego area. "In 1924, *Batillaria attramentaria* was first collected near transplanted Pacific Oysters (*Crassostrea gigas*) in Samish Bay, Washington (WA). It was spread to many West coast locations with oysters transplanted from Washington State." https://invasions.si.edu/nemesis/calnemo/species_summary/567272

Cercaria batillariae (HL1; Trematoda)

Population genetics dataset. See *Batillaria attramentaria* for sampling. Miura et al. (9) noted that "*No significant geographic genetic structure was found in either of the common introduced trematode species (HL1 and HL6) in Japanese populations, with almost all of the total diversity being distributed within populations*"

Vector hypotheses: Mixed or unknown. HL1 is one of three cryptic species of *Cercaria batillariae* that occurs in both Japan and western North America, and was genotyped by Miura et al. (9). Because HL1 had similar amounts of genetic diversity in non-native versus native populations, it was inferred to have been "*repeatedly introduced or originate from multiple source regions….We postulate that HL1 was repeatedly introduced to North America by migratory birds, which serve as its final hosts.*"

Cercaria batillariae (HL6; Trematoda).

Population genetics dataset. See *Batillaria attramentaria* for sampling. Miura et al. (9) noted that "*No significant geographic genetic structure was found in either of the common introduced trematode species (HL1 and HL6) in Japanese populations, with almost all of the total diversity being distributed within populations*"

Vector hypotheses: Oyster HL6 is one of three cryptic species of *Cercaria batillariae* that occurs in both Japan and western North America, and was genotyped by Miura et al. (9). Because HL6 had lower genetic diversity in non-native versus native populations, it was inferred to have been co-introduced with *B. attramentaria* and *M. gigas.*

Didemnum vexillum (Ascidiacea)

Population genetics dataset. We used the mitochondrial haplotype frequencies in Appendix Table 1 from (10) There were samples from three regions in Japan (Hok, Miy, Tok), western North America and Europe.

Vector hypotheses: Shipping Lambert (11) suggested that shipping was most likely. See summary at https://invasions.si.edu/nemesis/calnemo/species_summary/-334

Gracilaria vermiculophylla (Rhodophyta)

Population genetics dataset. We combined mitochondrial haplotype frequencies from five sources (12–16). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag), western North America and Europe.

Vector hypotheses: Mixed or unknown.

- Shipping: Western Europe
- Oyster: US Atlantic coast
- Shipping: Baja California
- Oyster: California (The first California site has no international shipping, but international shipping could have introduced it to another site on the west coast, and then coastal shipping to Monterey Bay (Elkhorn Slough); in addition, there were Japanese oyster introductions to Elkhorn Slough, but decades before this algae was first found, although it may have long overlooked.

Haminoea japonica (Mollusca; Japanese Bubble Snail)

Population genetics dataset We used mitochondrial haplotype frequencies from Table 1 of (17). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag), western North America and Europe.

Vector hypotheses: Oysters. The Japanese Bubble Snail "is associated with Pacific Oyster (*Crassostrea gigas*) and Japanese Littleneck (*Venerupis philippinarum*) aquaculture, the likeliest vectors for its introduction and spread." From https://invasions.si.edu/nemesis/calnemo/species_summary/567649

Hemigrapsus sanguineus (Crustacea; Asian shore crab).

Population genetics dataset. We generated a mitochondrial haplotype frequency table from two papers (18, 19). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag) and Europe.

Vector hypotheses: Shipping. Blakeslee et al (18) indicated that shipping was the most likely vector. See also "Ballast water is the most likely vector for its initial introduction" from https://invasions.si.edu/nemesis/species_summary/-2

Hemigrapsus takanoi (Mollusca; Asian brush-claw shore crab).

Population genetics dataset. We used mitochondrial haplotype frequencies in Supplemental Table 1 from (20). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag) and Europe.

Vector hypotheses: Shipping From Makino et al.(20)***: "****It is generally thought that European H. takanoi were likely introduced with Asian oysters and/or by shipping lines (e.g., Noel et al. 1997; Gollasch 1999) possibly via multiple independent introductions into French Atlantic, French British Channel, and the Netherland coast (Markert et al. 2014) […] This leads us to conclude that shipping lines were indeed effective vectors for the current H. takanoi populations in the Bay of Seine."*

Mutimo cylindricus (Phaeophyta)

Population genetics dataset. We used haplotype frequencies in Supplemental Table S9 from (21). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag) and western North America.

Vector hypotheses: Mixed or unknown Kogishi et al. (22) states "The mechanism for the introduction of *C. cylindrica* to California remains unknown."

Palaemon macrodactylus (Crustacea; Oriental Shrimp)

Population genetics dataset. We used the mitochondrial haplotype frequencies visualized in Figure 1 from (23). We have samples from three regions in Japan (Miy, Tok, Seto) and western North America.

Vector hypotheses: Shipping "The present study strongly supports the hypothesis that ballast water in international shipping is the original and main introduction vector of *P. macrodactylus* worldwide." (23) See also https://invasions.si.edu/nemesis/calnemo/species_summary/96450

Polydora hoplura (Polychaeta; shell-boring spionid)

Population genetics dataset. We pulled 16S rDNA haplotype frequencies from Figure 3 of (24). We have samples from three regions in Japan (Miy, Seto, Kag), western North America and Europe.

Vector hypotheses: Mixed or unknown. "*The history of the discovery of P. hoplura around the world appears to be intimately linked to global shipping commencing in the mid-19th century, followed by the advent of the global movement of commercial shellfish (especially the Pacific oyster Magallana gigas) in the 20th century, interlaced with continued, complex dispersal by vessels and aquaculture.*" (24).
"Due to its shell-boring habits, it has been transported in the hull fouling of ships, and with transfers of bivalves, especially Pacific Oysters (Magallana = Crassostrea gigas)"
https://invasions.si.edu/nemesis/species_summary/-795

Ulva australis (Chlorophyta)

Population genetics dataset. We concatenated the atpI (chloroplast) and trnA (mtDNA) sequences for *Ulva pertusa* from (21); this species was synomynized to *Ulva australis* in (25). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag), western North America and Europe.

Vector hypotheses: Mixed or unknown. "a non-intentional introduction associated with aquaculture (probably associated with young oysters) or associated with maritime activities. *Ulva* spp. are frequent fouling species, and have also been reported from ballast waters of trans-ocean shipping." (21)

Undaria pinnatifida (Ochrophyta; Wakame)

Population genetics dataset. We used the mitochondrial haplotype frequencies from Table 3 of (26). We have samples from all five regions in Japan (Hok, Miy, Tok, Seto, Kag), western North America and Europe.

Vector hypotheses: Mixed or unknown.
Shipping: California (many Japanese oyster bays are also near many major shipping ports)
Oysters: Mediterranean coast of France
Deliberate: Atlantic coast of France; Known intentional introduction by France to Atlantic France
See https://invasions.si.edu/nemesis/calnemo/species_summary/-21

Appendix 3. Approximate Bayesian Computation methods

Alternative models of introduction were compared using Approximate Bayesian Computation (ABC). Briefly, we developed models summarizing the history of global Oyster movement and those summarizing shipping from Japan to introduced regions. We then simulated individual haplotypes of each introduced species under these alternatives. At the conclusion of each simulation replicate, we summarized the distribution of haplotypes with population genetic summary statistics intended to capture diversity both among and within populations. Next, we used random forests (27) to develop a model that describes the distribution of summary statistics across simulations under the different models of introduction and used the resulting random forest model to estimate the posterior probability of each introduction model by applying it to the empirical data for each species. Finally we calculated log Bayes factors (28) supporting the alternatives from the logarithm of the ratio of posteriors for oyster introduction models to shipping introduction models. We note our ABC approach did not tease apart primary and secondary introduction routes and instead inferred only primary introductions.

Simulation model (common to all introductions)

The structure of each simulation conducted for each species differed primarily in the choice of vector model. The matrices (i.e., Oyster models are in Table S3 and S4 and Shipping models are in Table S5 and S6) reflected the alternative introduction models we compared. Other than these differences in introduction model, simulations were parameterized by pulling values for the demographic parameters described below (illustrated in Figure S10 and detailed in Table S7.

Native range:

The native range topology represents the root of simulated coalescent trees. The upper portion of the tree in Figure S10 illustrates this fixed topology. Within each region in the native range sampled and unsampled populations are simulated. Sampled populations correspond to empirical data and are used to calculate summary statistics to compare to the empirical data to estimate posterior probabilities of models. Simulated, unsampled, "ghost populations" can share lineages with sampled populations through gene flow and can also serve as the source population for introductions. Though Figure S10 includes 5 populations per native range for visual clarity, we included enough unsampled populations in each simulation to result in 20 populations in each of the 5 native ranges. Four time parameters (*t0, t1, t2*, and *t3*) and two population size parameters (*Ne* and *ancestralNe*) determine the demography in the native range. In addition, migration was allowed under an island model within each region with a rate parameter, *nativeM*.

Introductions:

Each introduction into a non-native region assumed that there was an instantaneous colonization of the region by 20 populations and that these populations were introduced *tIntro* generations in the past. The source of these populations was determined by the introduction models described above. In addition, for each introduction model two scenarios were implemented. First, a source population for an introduction was chosen treating the columns in the introduction models described above as probabilities of a multinomial distribution (single source introduction).

Second, introductions occurred through a simultaneous mixture of lineages proportional to the values in the columns of the same introduction model (mixed source introduction).

Each introduced population had an effective size given by *NeIntro* and within each introduced region, populations could exchange migrants under an island model at a rate given by *introM.*

Figure S10 provides a simplified example of the demographic history for one possible simulation. In this case the introduction model was Oyster Admixture and single source introductions were used. Most demographic parameters are denoted on this figure. This history shows introductions into Europe and PNW from Miyagi. For this (hypothetical) species, there is only a single population sampled in Miyagi (the source population of the PNW introduction), though both introduced regions were sampled. The introductions from Miyagi also demonstrate that introductions can occur from populations that were not directly sampled (though other populations in the regions were, if the region was included in the model).

In addition the coalescence of introduced regions to particular lineages within native regions is determined by entries in the introduction models specified above, divided by the number of populations in the region. Finally there were two gene flow parameters (not illustrated), chosen from uniform priors, that specify connectivity among populations within each region. The two parameters correspond to 'm' in native and introduced populations respectively.

Simulation Priors:

The underlying demographic parameters that specified each simulation (other than the introduction model) were chosen from uniform priors. These parameters are outlined in Table S7. Parameters served as nuisance parameters and model comparisons were integrated across these distributions

Introduction model priors:

Flat prior approach:
An introduction model was chosen from eight possible introduction models with equal probability. They correspond to the 2 oyster models and the 2 shipping models. For each choice of introduction model both possible introduction dynamics could occur.

Historical prior approach:
The prior probability of choosing oyster vs shipping models was estimated through expert elicitation and examination of the literature. See Appendix 2 and Table S2 for details.
The prior probability of a particular mode of introduction (shipping vs oyster) was divided equally among the submodels.

Simulation implementation
The actual coalescent simulations were implemented in fastSimCoal 2.7 (29) wrapped by functions in an R package written by AES (https://github.com/stranda/testInvPath). testInvPath parameterized the simulations, controlled simulation runs via calls to the R package strataG, imported genotype data and calculated summary statistics.

To aid in reproducibility, the entire simulation environment including the coalescent simulator, R, and R packages is containerized and a docker image of the environment is available at: https://hub.docker.com/r/astrand/testinvpath.

Summary statistics

Population genetic summary statistics were used to convert individual genotypes into metrics that could be compared among replicate simulations and among simulations and empirical data. Broadly, they estimated species, regional and population-level diversity. In addition among population diversity was assessed through pairwise population and regional comparisons.

The number of summary statistics were variable among species because each species of interest had a different number and geographic distribution of populations. Table S8 outlines the statistics calculated for each species.

Model posterior probability estimation

We estimated the posterior probability of the "correct" introduction model for each species using random forests (27). These calculations were implemented in the R package 'abcrf'(30). For the results presented in the main text, we combined all shipping models together as a single shipping model and likewise we combined all oyster models together as a single oyster model. This gave us the ability to estimate the posterior probability of shipping versus oyster for each species. Inspection of the summary statistics in PCA space for each of the submodels (see next section) indicate that this is a reasonable approach for the species and samples under consideration.

Simulation model evaluation

One approach to assessing whether a simulation model can produce haplotypes and resulting summary statistics that are consistent with empirical data for a species is to perform PCA on the concatenated simulated *and* empirical summary statistics. We can then plot the simulations and empirical data in PC space and assess visually whether the empirical data could be produced by the simulation model. An additional benefit of this evaluation is that it also provides a visual assessment of the identifiability of alternative models. Figure S11 illustrates this approach for two species, one that is consistent with shipping vectors and one consistent with oyster transport. In both cases, the empirical data sits well within the space explored by simulations. Furthermore there is good separation between simulations of oyster and shipping introductions, though sub-models within these two categories are indistinguishable. Finally, in the case of *P. macrodactylus* the empirical data resembles simulated shipping introductions whereas the converse is true for *P. hoplura*.

As elucidated in the main text, not all species showed such clear distinction among models. Figure S12 illustrates the same ordination for simulations of *Acanthogobius flavimanus*. In this example, simulations still generate summary statistics that are consistent with the empirical data, but the simulation outputs for the different introduction models are essentially indistinguishable.

The ability of the approach employed to distinguish among introduction models can also be assessed through examination of confusion matrices associated with each species. Table S9 represents confusion matrices with equal prior probability, while Table S10 represents confusion matrices assuming prior probabilities for each introduction model (S2). The overall performance of the classifier is assessed by the P4 statistic (28).

To explore whether the inference of a vector is consistent across oyster models, we compared the posterior probabilities of a shipping vector (0 = oyster; 1 = shipping) using the admixture oyster model (y-axis) and ML oyster model (x-axis), where each point represents one of 14 introduced species (Figure S13). The two oyster models differ largely in the relative importance of Tokyo as a source (Table S3-S4). The Pearson's correlation coefficient was 0.860 ($n = 14$, $p < 0.001$), indicating that our interpretation is largely robust across the two oyster models. In the case of the species outlier *Undaria* ("Up"), there is a change in the inference when you use one model over the other. In Figure 3 and Table S2, we collapsed the two submodels, and inferred a shipping vector for *Undaria*. But, the genetic variation does not fit either model well for Undaria. More broadly, when *Undaria* is removed, the correlation between admixture and ML models is 0.988 ($n = 13$; $p<0.001$). Thus, the two oyster models do not differ in their inference, except in the case of *Undaria*.

Table S1. Metadata for populations. N = number of sequenced individuals.

Pop	N	Population description	Country	Region	Native / Introduced	CollectionMonth	CollectionYear	Latitude	Longitude	Collector
AKK	27	Akkeshi	Japan	Hokkaido	Native	August	2017	43.02131	144.836622	Yorisue/Nakaoka
ALB	19	Alamitos Bay LA County - 2nd street	USA	California	Introduced	September	2017	33.746239	-118.135526	Zacherl
BAN	23	Banzu Tidalflat in Tokyo Bay (Kisarazu city)	Japan	Chiba	Native	June	2017	35.41246	139.902977	Kanaya
BBH	18	Bahia Blanca Harbor (BB)	Argentina		Introduced	October	2017	-38.78497	-62.29695	Schwindt
BOI	20	Waikare Inlet Bay of Islands (from Oyster World)	New Zealand		Introduced	September	2017	-35.31	174.18	Taylor
BOU	7	Boursefranc-le-Chapus	France		Introduced	July	2017	45.85504	-1.170636	Pante and Emmanuel Dubillot
CHI	18	Estero, Tongoy, Coquimbo	Chile		Introduced	October	2019	-30.244331	-71.491032	Martin Thiel & Jim Carlton
CLE	18	Clevedon (from Clevedon Coast Oysters)	New Zealand		Introduced	October	2017	-36.94	175.11	Taylor
ESN	9	Esnandes (Aiguillon Bay Pertuis Breton)	France		Introduced	July	2017	46.24123	-1.219623	Pante and Emmanuel Dubillot
ESP	20	Espasante_Galicia	Spain		Introduced	August	2017	43.72127	-7.814205	Couceiro
GAN	10	Gandario	Spain		Introduced	August	2017	43.34252	-8.23345	Couceiro
GOS	13	GoseongBay	Korea		Native	June	2017	34.9555	128.3257	Lee
GRC	10	Grand Caribe San Diego Bay	USA	California	Introduced	September	2017	32.62674	-117.129703	Zacherl
GWG	25	Gwanyang	Korea		Native	June	2017	34.94519	127.77932	Lee
HIR	24	Oono Strait in Hiroshima Bay (Hiroshima)	Japan	Seto Inland Sea	Native	November	2021	34.27547	132.266519	Hori
HOD	10	HoodCanal_HammaHammaRiver_NorthShore	USA	Washington State	Introduced	June	2017	47.54666	-123.041383	Ruesink
KOJ	26	Koje Bay	Korea		Native	June	2017	34.7991	128.5829	Lee
LDY	19	Ladysmith British Columbia	Canada	British Columbia	Introduced	August	2017	48.99065	-123.808072	Ruesink
LHE	15	L, L'Herbe (Arcachon Bay)	France		Introduced	July	2017	44.68641	-1.233772	Thomas Lacoue-Labarthe
LOS	16	Lough Swilly	Ireland		Introduced	November	2008	55.0206	-7.577	Kochmann
MAH	18	Dyers Creek Mahurangi (from Matakana Oysters)	New Zealand		Introduced	September	2017	-36.46	174.71	Taylor
MAI	4	Maizuru	Japan	Sea of Japan	Native	October	2021	35.515338	135.336001	Yorisue
MAL	24	Matsukawaura Lagoon (A branch Bay of Sendai Bay)	Japan	Fukushima	Native	May	2017	37.82314	140.954988	Kanaya
MAT	27	Matsushima Bay (A branch Bay of Sendai Bay)	Japan	Miyagi	Native	July	2017	38.35186	141.058982	Kanaya
MON	29	Nishi Mone Bay (A branch Bay of Kesennuma Bay)	Japan	Miyagi	Native	May	2017	38.90008	141.624736	Kanaya

NEW	19	NewportBay Orange County - 15th street	USA	California	Introduced	September	2017	33.60813	-117.92009	Zacherl
OHK	15	Ohkamo River in Izu Peninsura (Shimoda city)	Japan	Shizuoka	Native	August	2017	34.65528	138.918478	Kanaya
OKA	23	Hinase Bay (Okayama)	Japan	Seto Inland Sea	Native	November	2021	34.45628	133.896859	Hori
OTA	27	Oita	Japan	Seto Inland Sea	Native	November	2021	33.62537	131.193653	Hori
PES	16	Los Penasquitos Lagoon Coast hwy	USA	California	Introduced	October	2017	32.93318	-117.259885	Crooks
QB2	20	Quilcene - Hood Canal	USA	Washington State	Introduced	September	2021	47.78238	-122.856315	Heyford
SAM	25	Same River (Iwaki City)	Japan	Fukushima	Native	July	2017	36.90936	140.816924	Kanaya
SAR	26	Lake Saroma	Japan	Hokkaido	Native	July	2017	44.12084	143.966724	Yorisue/Nakaoka
SBB	16	San Blas Bay (SB)	Argentina	Argentina	Introduced	October	2017	-40.5394	-62.252666	Schwindt
SVA	18	Svallhagen	Sweden		Introduced		2013	58.8684	11.1551	Asa Strand
TJE	11	TJE Boca Rio	USA	California	Introduced	October	2017	32.5595	-117.129165	Crooks
TRO	16	Tromlingene	Norway		Introduced		2013	58.4748	8.9067	Asa Strand
WAD	19	Wadden Sea	Denmark		Introduced	July	2005	55.1859	8.6222	Asa Strand
WLB	12	WillapaBay_Nahcotta	USA	Washington State	Introduced	May	2017	46.46874	-124.023696	Ruesink
YOJ	1	Yojiro Kagoshima	Japan	Kagoshima	Native	August	2017	31.55863	130.562861	Terada/Endo
YUR	13	Yura_Sumoto_HyogoPrefecture	Japan	Seto Inland Sea	Native	May	2017	34.27394	134.953667	Kamiya

Table S2. The hypothesized and inferred vectors for 14 Japanese species that were introduced into western North America (wNA), Europe, or both. Species are sorted from strongest to weakest support for an oyster vs. shipping vector (See Figure 3). Historical prior comes from a summary of the literature (see Appendix 2). Number of populations (N) within regions and the mean number of individuals per population are shown. Abbreviations for Regions in Japan: Hok = Hokkaido, Miy = Miyagi, Tok = Tokyo, Seto = Seto Inland Sea, Kag = Kagoshima. BF for models with different priors are indicated, as well as the accuracy of the random forest model for the genetic data.

Species	Taxonomy	Historical Prior	Ref.	Genetic inference	PopGen Marker	PopGen Ref.	Regions in Japan	Native (N)	wNA (N)	Europe (N)	Individuals (mean)	BF-HistPrior	BF-FlatPrior	RF-Accuracy
Polydora hoplura	Polychaeta	Mixed or Unknown	(24)	Oyster	mtDNA	(24)	Miy, Tok, Seto, Kag	6	2	2	2.5	1.834	2.607	0.5
Haminoea japonica	Mollusc	Oyster	(31)	Oyster	mtDNA	(31)	Hok, Miy, Tok, Seto, Kag	15	2	3	5.2	1.402	1.272	0.648
Batillaria attramentaria	Mollusc	Oyster	(9)	Oyster	mtDNA	(9)	Miy, Tok, Seto, Kag	14	4	0	10.0	1.105	0.531	0.852
Cercaria batillaria (HL6)	Trematode	Oyster	(9)	Oyster	mtDNA	(9)	Miy, Tok, Seto, Kag	14	4	0	25.0	0.892	-0.309	0.675
Hemigrapsus takanoi	Mollusc	Shipping	(20)	Oyster	mtDNA	(20)	Hok, Miy, Tok, Seto, Kag	51	0	8	10.4	0.703	0.895	0.461
Gracilaria_vermiculophylla	Alga	Mixed or Unknown	(12–16)	Oyster	mtDNA	(12–16)	Hok, Miy, Tok, Seto, Kag	67	30	20	5.2	0.683	0.687	0.692
Ulva pertusa	Alga	Mixed or Unknown	(21)	Unknown	mtDNA; chl	(21)	Hok, Miy, Tok, Seto, Kag	59	3	4	1.8	-0.157	-0.175	0.589
Acanthogobius flavimanus	Chordate	Shipping	(9)	Unknown	mtDNA	(8)	Hok, Miy, Tok, Seto, Kag	21	2	0	26.3	-0.322	0.001	0.565
Mutimo cylindricus	Alga	Shipping	(22)	Unknown	mtDNA	(21)	Hok, Miy, Tok, Seto, Kag	15	1	0	6.1	-0.383	0.087	0.725
Undaria pinnatifida	Alga	Mixed or Unknown	(26)	Shipping	mtDNA	(26)	Hok, Miy, Tok, Seto, Kag	35	3	2	3.3	-0.547	0.43	0.766
Cercaria batillaria (HL1)	Trematode	Mixed or Unknown	(9)	Shipping	mtDNA	(9)	Miy, Tok, Seto, Kag	14	4	0	13.5	-0.556	-0.634	0.681
Hemigrapsus sanguineus	Crustacea	Shipping	(18)	Shipping	mtDNA	(18, 19)	Hok, Miy, Tok, Seto, Kag	26	0	1	16.1	-0.602	-0.28	0.499
Didemnum vexillum	Ascidian	Shipping	(11)	Shipping	mtDNA	(10)	Hok, Miy, Tok	8	17	9	6.3	-0.891	1.137	0.885
Palaemon macrodactylus	Crustacea	Shipping	(23)	Shipping	mtDNA	(23)	Miy, Tok, Seto	6	3	5	17.0	-0.947	-0.813	0.167

Table S3: Relative contributions of native sources (rows) to introduced regions (columns) via admixture analysis (Figure S8). The entries in this table were estimated through visual inspection of the admixture analysis conducted on *M. gigas*. Each column has been normalized to sum to 1.0. In the case of species that were only introduced into a subset of these regions, or for which native range information is absent, the relevant columns and rows were removed and columns renormalized if necessary.

	PNW	California	EU
hok	0.5	0.0	0.3
hon	0.5	0.2	0.7
tok	0.0	0.4	0.0
sea	0.0	0.4	0.0
kag	0.0	0.0	0.0

Table S4: Relative contributions of native sources (rows) to introduced regions (columns) via machine learning. The entries in this table were estimated using random forest assignment of *M. gigas* introduced populations to native regions. Each column has been normalized to sum to 1.0. In the case of species that were only introduced into a subset of these regions, or for which native range information is absent, the relevant columns and rows were removed and columns renormalized if necessary.

	PNW	California	EU
hok	0.033	0.000	0.031
hon	0.967	0.493	0.969
tok	0.000	0.000	0.000
sea	0.000	0.507	0.000
kag	0.000	0.000	0.000

Table S5: Relative contributions of native sources (rows) to introduced regions (columns) via the 30-day voyage dataset. These entries were estimated from voyages originating in Japan and travelling to the introduced regions. In this case, voyages that exceed 30 days are excluded. Each column has been normalized to sum to 1.0. In the case of species that were only introduced into a subset of these regions, or for which native range information is absent, the relevant columns and rows were removed and columns renormalized if necessary.

	EU	PNW	California
hok	0.000	0.343	0.111
kag	0.125	0.086	0.111
hon	0.000	0.086	0.037
sea	0.500	0.171	0.315
tok	0.375	0.314	0.426

Table S6: Relative contributions of native sources (rows) to introduced regions (columns) via the 60-day voyage dataset. These entries were estimated from voyages originating in Japan and travelling to the introduced regions. In this case, voyages that exceed 60 days are excluded. Each column has been normalized to sum to 1.0. In the case of species that were only introduced into a subset of these regions, or for which native range information is absent, the relevant columns and rows were removed and columns renormalized if necessary.

	EU	PNW	California
hok	0.104	0.359	0.120
kag	0.188	0.078	0.133
hon	0.000	0.047	0.024
sea	0.417	0.172	0.301
tok	0.292	0.344	0.422

Table S7 Prior distributions for each demographic parameter. Each is chosen from a uniform distribution with the specified lower and upper bounds:

Parameter	lowerBound	upperBound
tIntro	2	200 (generations)
t0	tIntro+1	10000 (gen)
t1	t0+10	50000 (gen)
t2	t1+1	50000 (gen)
t3	max(t1,t2)+1	100000(gen)
NeIntro	5	5000
Ne	1000	7000
AncestralNeCoef	1	400 (coefficient to multiply Ne to determine ancestral size)
nativeM	0	0.01
introM	0	0.01
mutation (mu)	0	5x10-3 (mutations per position)

Table S8. Outline of summary statistics and their estimation for each species simulated. Columns correspond to the hierarchical level of diversity estimated. Rows correspond to specific statistics. These statistics were estimated using three R packages: strataG v2.50 (32), pegas v1.3 (33), and adegenet v2.1.10 (34). In table cells, the functions used for estimating that statistic for a particular level of hierarchy are given.

	Approach to estimation				
Statistic	**Species overall**	**Within Population diversity**	**Among population diversity**	**Within Region**	**Among Region**
Nucleotide Diversity	mean of strataG function "nucleotideDiversity" across all sequences	strataG function "nucleotideDiversity applied to all sequences in each population		strataG function "nucleotideDiversity" averaged within each region	
Nei's nucleotide distance (da)		strataG function nucleotideDivergence among sequences (average da within population)	pairwise da metric among populations estimated using strataG nucleotideDivergence		
Haplotype diversity	pegas function hap.div	pegas function hap.div applied to each population		pegas function hap.div applied to each region	
PhiST	strataG function overallTest()				strataG function "pairWiseTest"
Population structure theta Weir and Cockerham 1983			pairwise implemented in adegenet "dist.genpop"		

Table S9: Confusion matrices for random forest models assuming an equal prior probability for each introduction model. For each species, shipping and oyster transport submodels are combined. The entries in each of the first two columns represent the classification of simulations into shipping and oyster transport, respectively. The rows correspond to the model under which the simulation was generated (shipping versus oysters). The third column shows the rate at which simulations were misclassified. Low classification errors are another indicator of model identifiability and high error indicates an inability to distinguish models. P4 ranges from 0-1 with higher values indicating better ability to discriminate among introduction models.

	Ship	Oyster	Classification error	P4
Polydora hoplura				
Ship	49688	397	0.008	0.986
Oyster	977	49023	0.020	
Haminoea japonica				
Ship	48742	1408	0.028	0.983
Oyster	302	49868	0.006	
Batillaria attramentaria				
Ship	94577	5423	0.054	0.972
Oyster	243	99757	0.002	
Cercaria HL6				
Ship	33585	3995	0.106	0.944
Oyster	228	37341	0.006	
Hemigrapsus takanoi				
Ship	46898	3121	0.062	0.960
Oyster	874	49151	0.017	
Gracillaria vermiculophylla				
Ship	36327	1119	0.030	0.982

	Oyster	230	37198	0.006	
Ulva pertusa					
	Ship	36338	1029	0.028	0.984
	Oyster	161	37418	0.004	
Acanthogobius flavimanus					
	Ship	32465	17718	0.353	0.641
	Oyster	18169	31581	0.365	
Mutimo cylinidricus					
	Ship	37448	12469	0.250	0.672
	Oyster	19973	29971	0.400	
Undaria pinnatifida					
	Ship	47247	2833	0.057	0.969
	Oyster	273	49873	0.005	
Cercaria HL1					
	Ship	44561	5353	0.107	0.944
	Oyster	267	49884	0.005	
Hemigrapsus sanguineus					
	Ship	46335	3654	0.073	0.959
	Oyster	466	49680	0.009	
Didemnum vexillum					
	Ship	45925	4120	0.082	0.955
	Oyster	406	49877	0.008	
Palaemon_macrodactylus					

Ship	49999	269	0.005	0.990
Oyster	725	49186	0.015	

Table S10: Confusion matrices for random forest models assuming prior probabilities for each introduction model (Table S2). Otherwise, this table follows Table S9.

	Shipping	Oyster	Classification error	P4
Polydora hoplura				
Ship	89852	45	0.001	0.989
Oyster	351	9731	0.035	
Haminoea japonica				
Ship	9616	401	0.040	0.987
Oyster	45	89625	0.001	
Batillaria attramentaria				
Ship	18869	1131	0.057	0.984
Oyster	7	179993	0.000	
Cercaria HL6				
Ship	6586	865	0.116	0.965
Oyster	13	67653	0.000	
Hemigrapsus takanoi				
Ship	88223	1986	0.022	0.900
Oyster	1472	8530	0.147	
Gracillaria vermiculophylla				
Ship	36136	1094	0.029	0.983
Oyster	210	37254	0.029	
Ulva pertusa				
Ship	36392	1030	0.028	0.984
Oyster	153	37474	0.004	

Acanthogobius flavimanus				
Ship	89874	187	0.002	0.048
Oyster	9889	126	0.987	
Mutimo cylinidricus				
Ship	88298	1798	0.020	0.433
Oyster	8081	1927	0.807	
Undaria pinnatifida				
Ship	47252	2946	0.059	0.968
Oyster	263	49778	0.005	
Cercaria HL1				
Ship	44671	5485	0.109	0.942
Oyster	288	49487	0.006	
Hemigrapsus sanguineus				
Ship	86012	3831	0.043	0.872
Oyster	1012	9149	0.100	
Didemnum vexillum				
Ship	88480	1460	0.016	0.898
Oyster	1908	8095	0.191	
Palaemon_macrodactylus				
Ship	90104	23	0.000	0.993
Oyster	241	9820	0.024	

Figure S1. Map of populations in native and non-native range. See Table S1 for metadata.

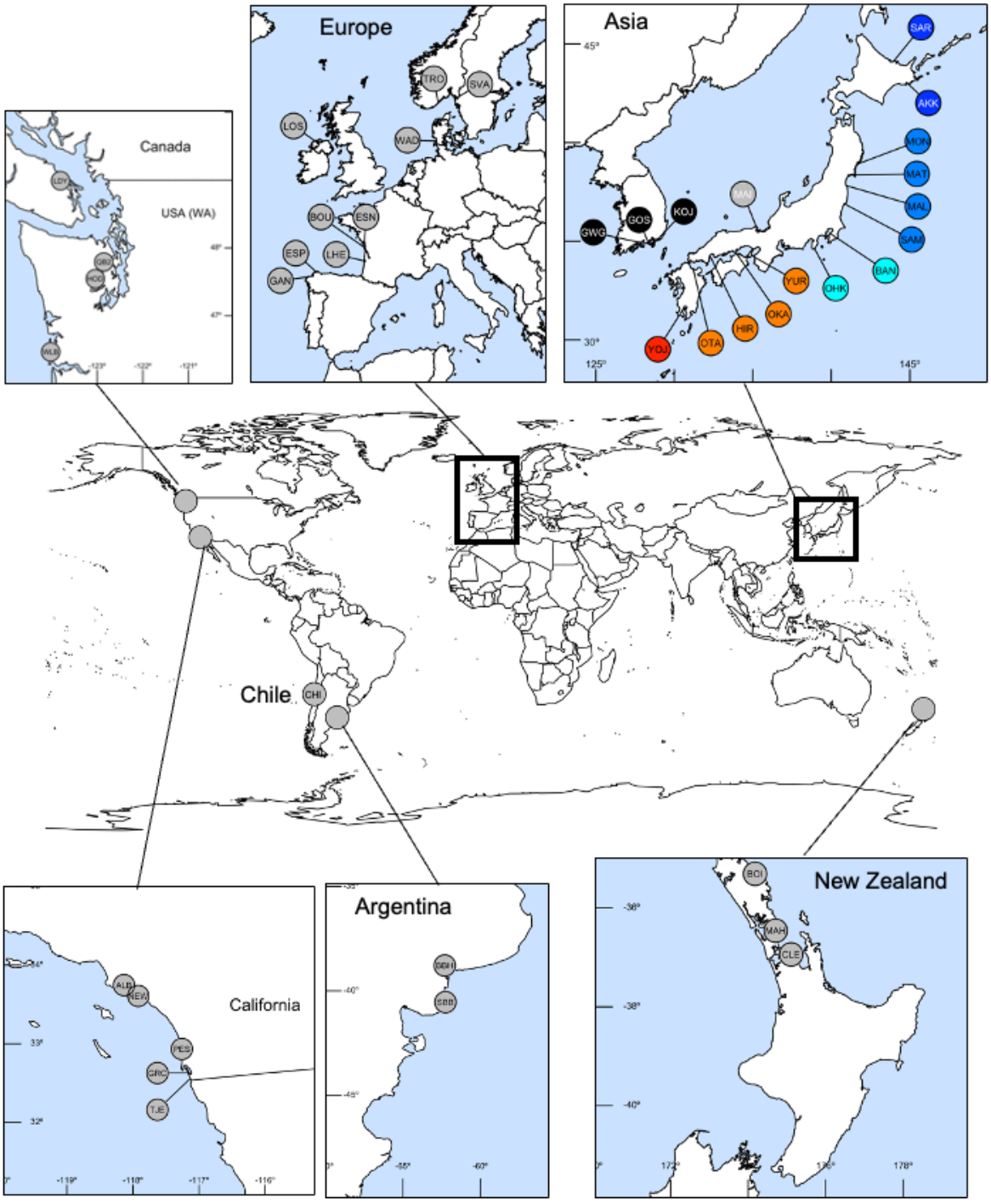
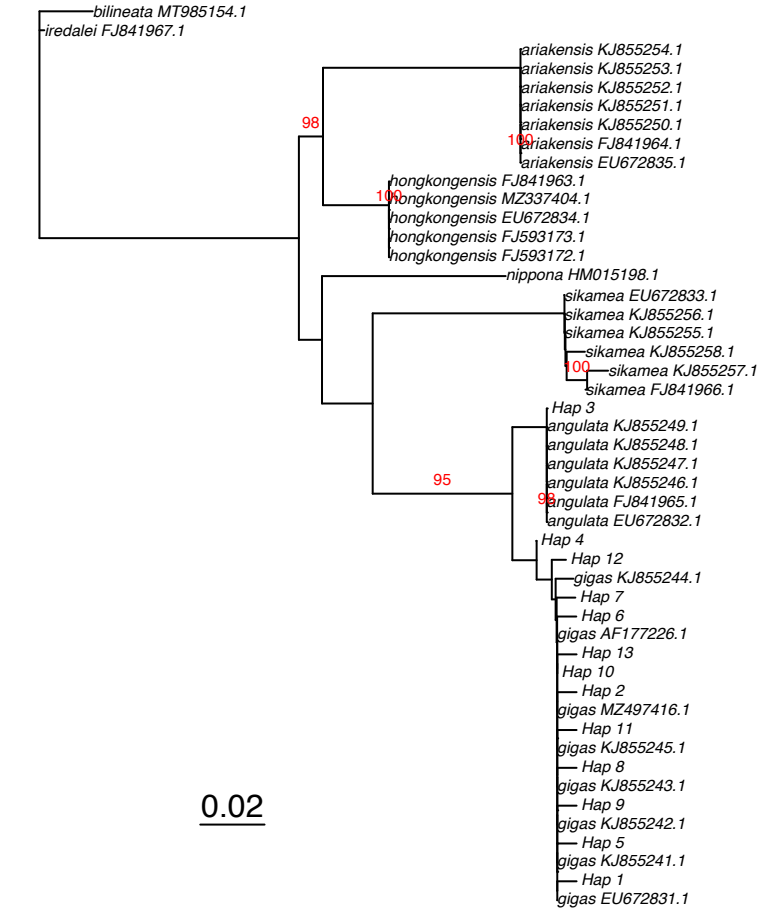


Figure S2. A neighbor-joining tree of 176 bp from two mitochondrial loci (NADH and non-coding region) in 8 *Crassostrea / Magallana* species and 13 unique haplotypes found in our dataset. Red numbers indicate the bootstrap support above 90% from 1000 replicates. Bar indicates branch length of 2% (K80 distance). Haplotype 10 is the most common haplotype (98% of 606 individuals).



.

Figure S3. Population averages of mapping rate to *Magallana gigas* genome as a function of region. Native = Japan and Korea; noEurope = Ireland, Denmark, Norway, Sweden; soEurope = Spain, France; PNW = Canada, Washington State USA; soCalifornia = southern California USA. ANOVA $F_{7,33}$=0.843 p = 0.560. See Table S1 for population codes.

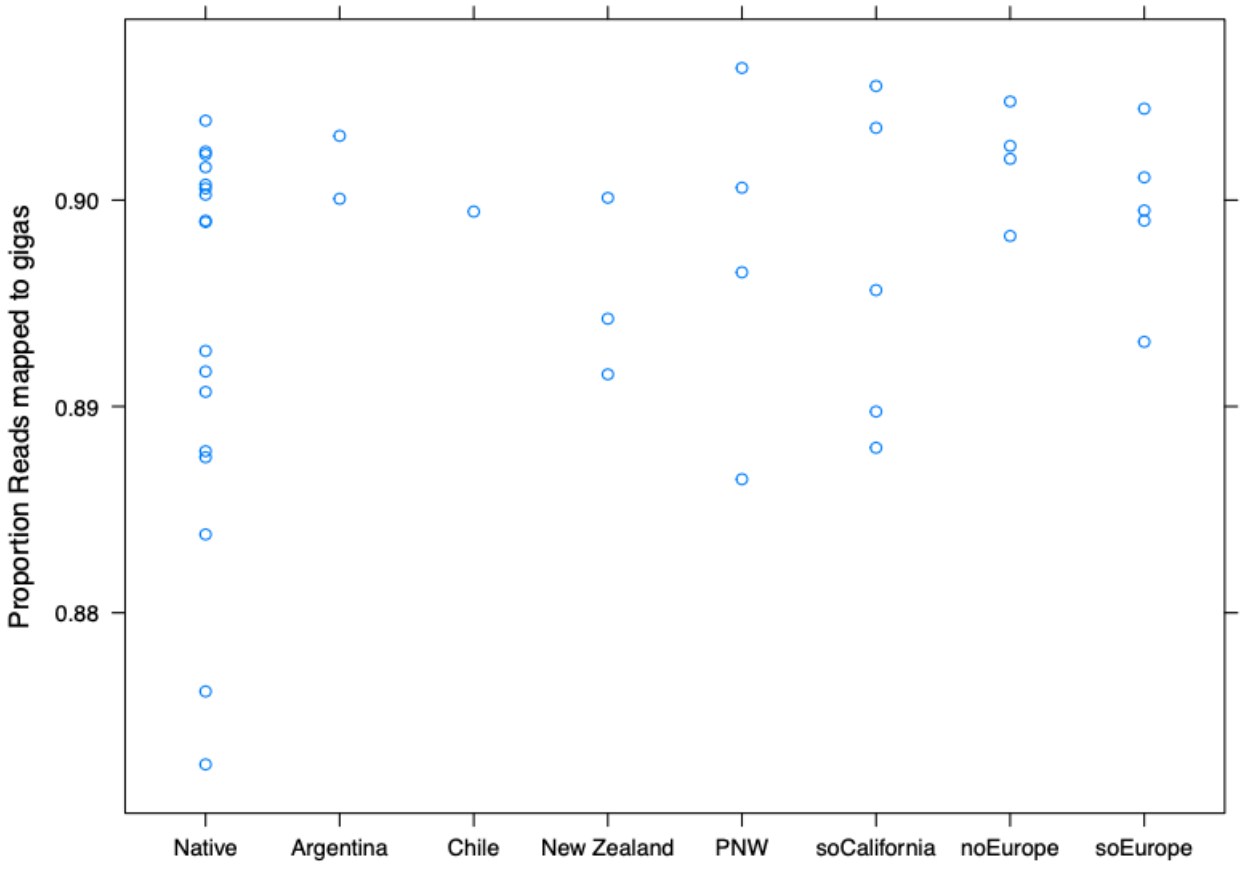


Figure S4. Distribution of PC1 of Korea and Japan samples using the PCA in Figure S4. See Figure 1b for map of populations and color codes for regions. YoJ not included because of low sample size (n=1)

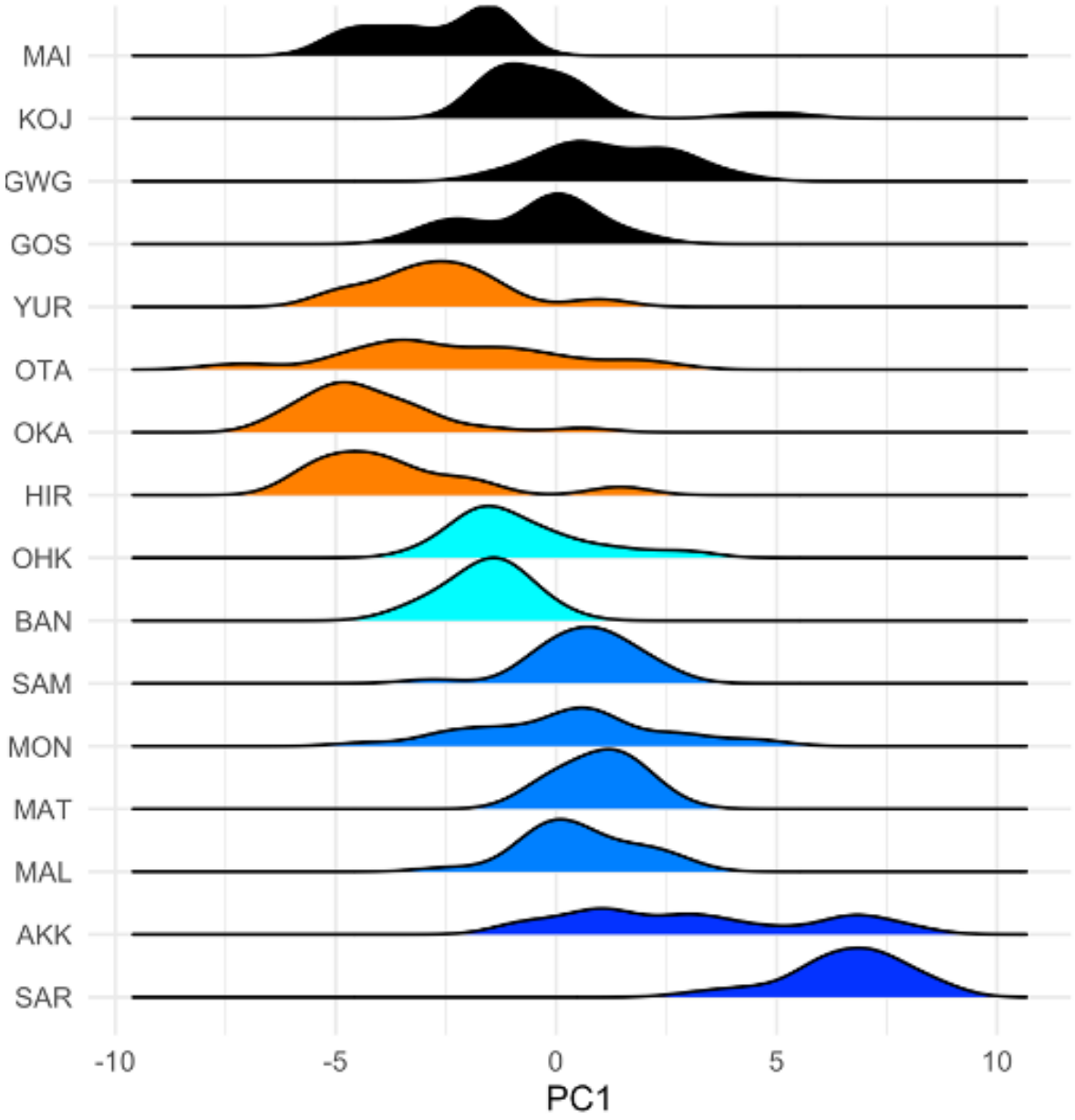


Figure S5. Principal components analysis (PCA) of 726 oysters using 7046 loci. Population codes are in Table S1

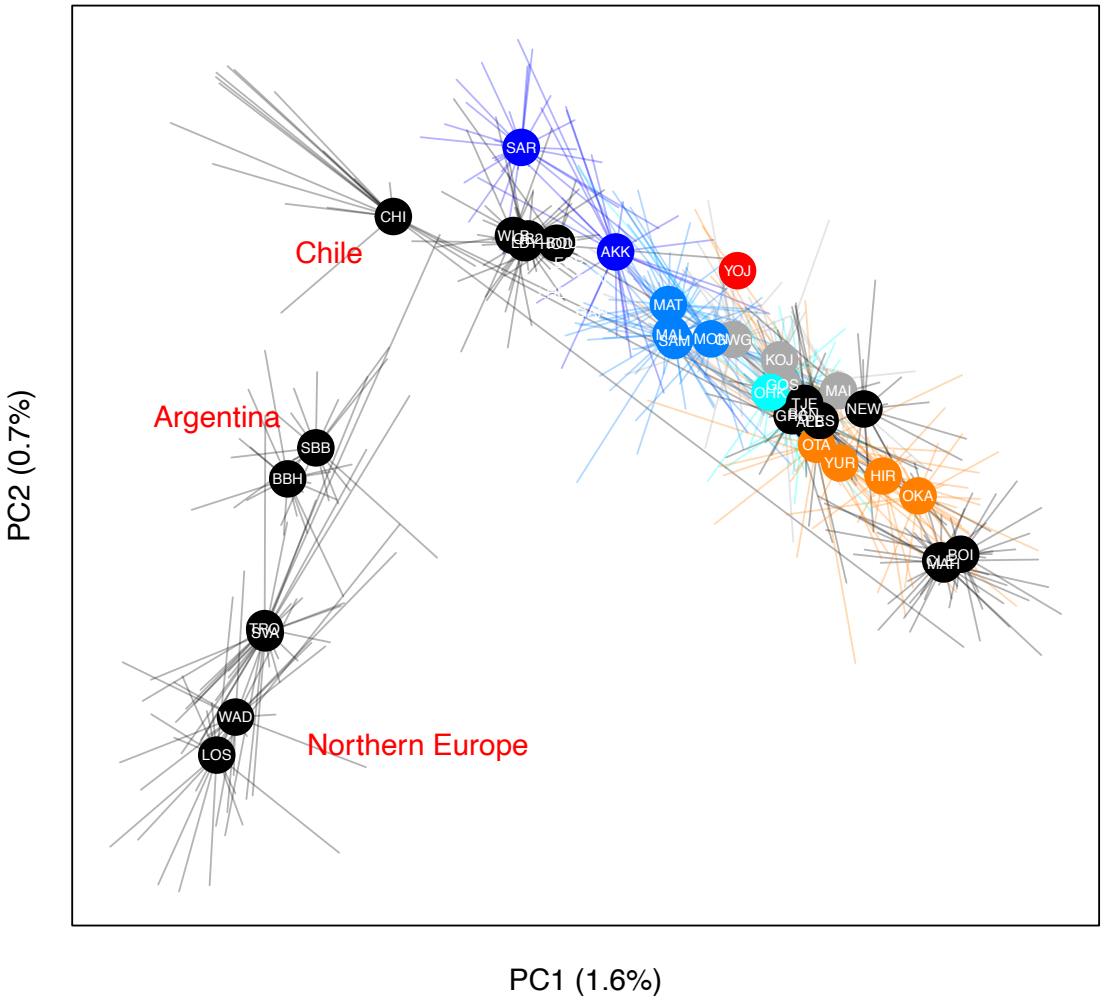


Figure S6. Pairwise genetic differentiation between native and non-native populations. Population codes are in Table S1. We calculated Φ_{ST} for all populations except YOJ and MAI; these latter two had low sample size (n<5).

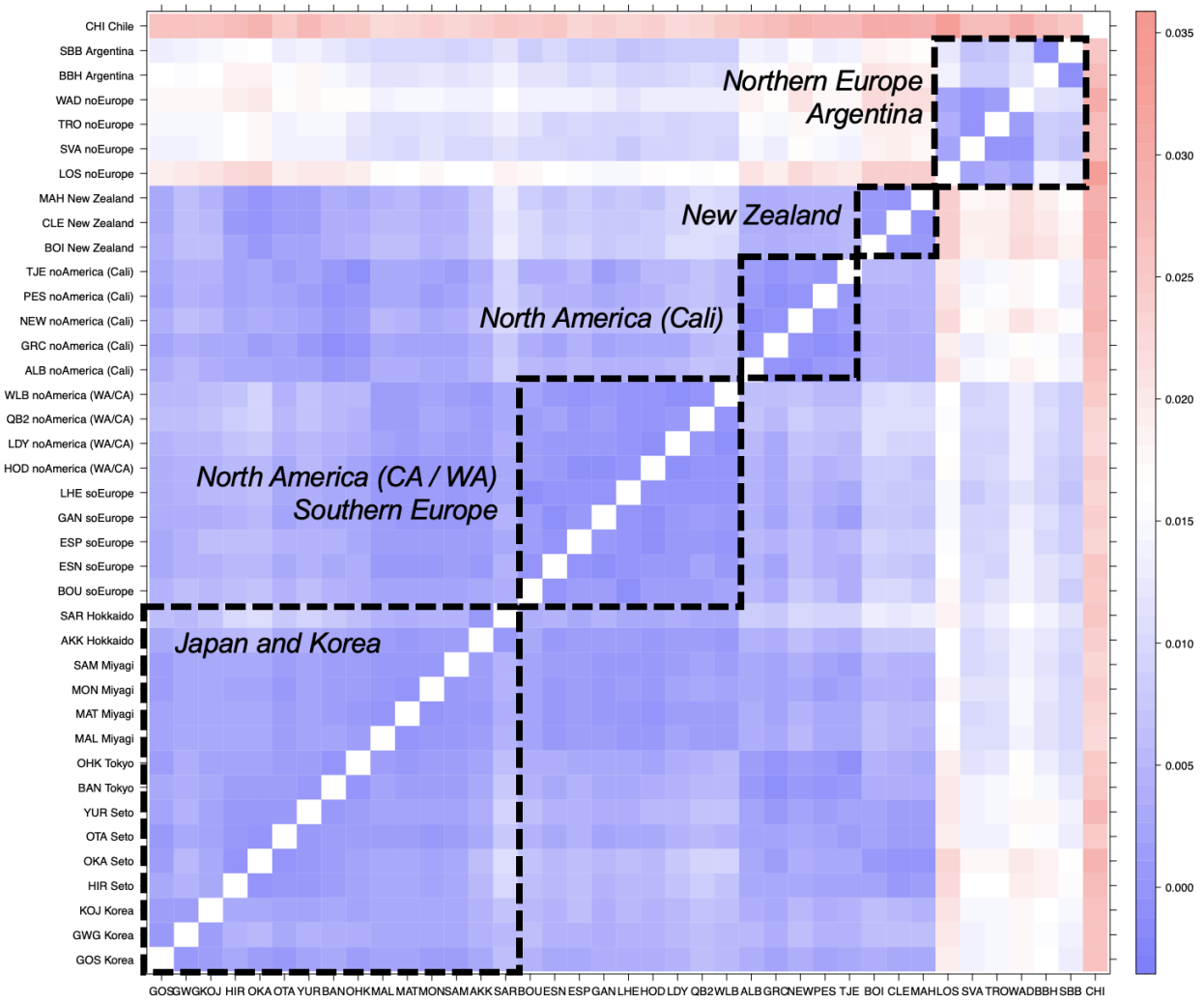


Figure S7. Simulations of the genetic differentiation (as measured by Φ_{ST}) that results from aquacultural breeding. Breeding pairs (x-axis) were generated between populations (e.g., OHK-MON, OKA-OHK; indicated with circles) and within populations (e.g., OHK, OKA, SAM; indicated with triangles) in the Miyagi region. Differentiation between these populations and the original populations are reported. Population codes are in Table S1. Each panel represents years since aquacultured samples were initiated (i.e., 1 – 20 years). Red line indicates $\Phi_{ST} = 0.01$, which was the mean divergence between Miyagi and northern European populations.

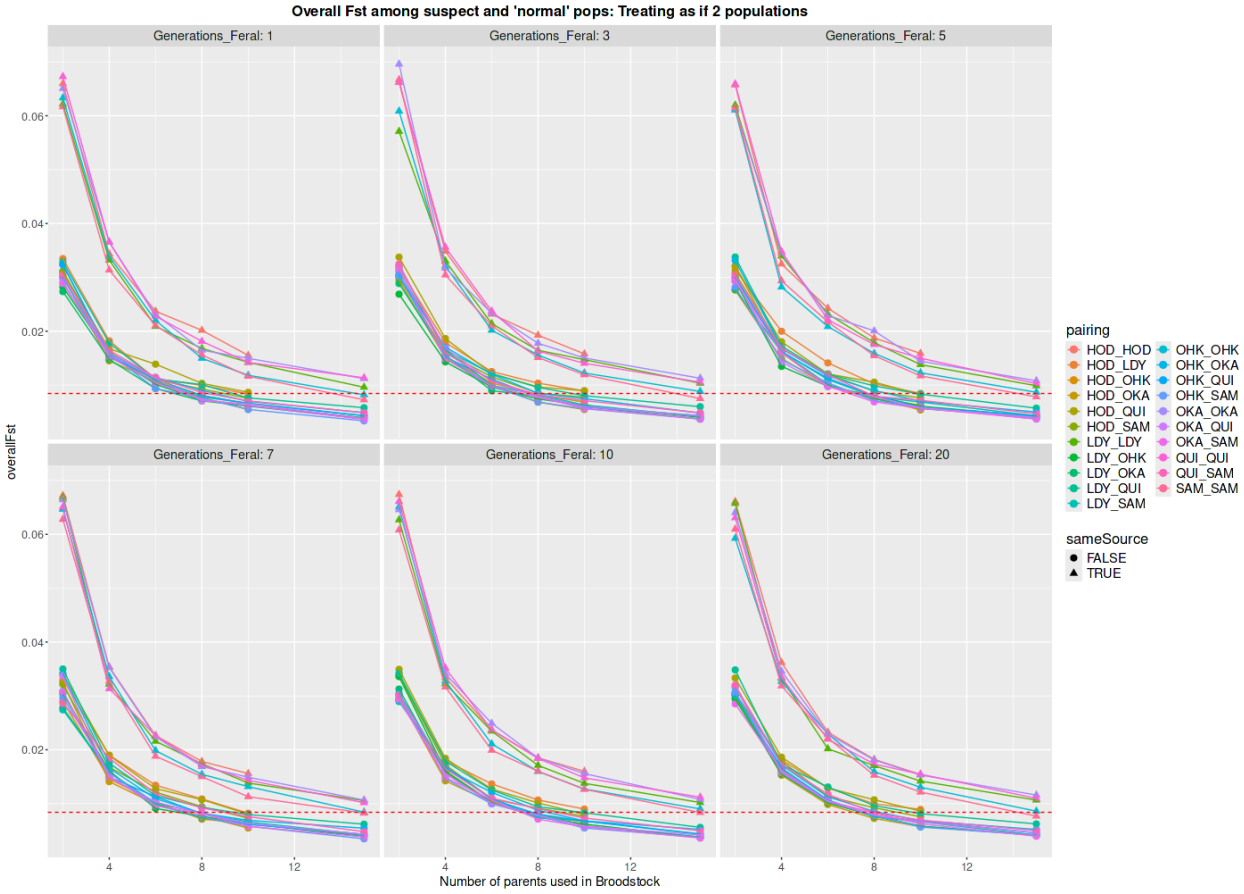


Figure S8. Inferred genomic admixture levels using K = 2-8 using genotype likelihoods Log-likelihood plotted in middle panel and double-prime log-likelihood is in the lower panel. Segments represent the 5% to 95% quantiles for 50 independent runs at each K. Population codes are in Table S1. “NonSource” indicates South Korea and western Japan populations.

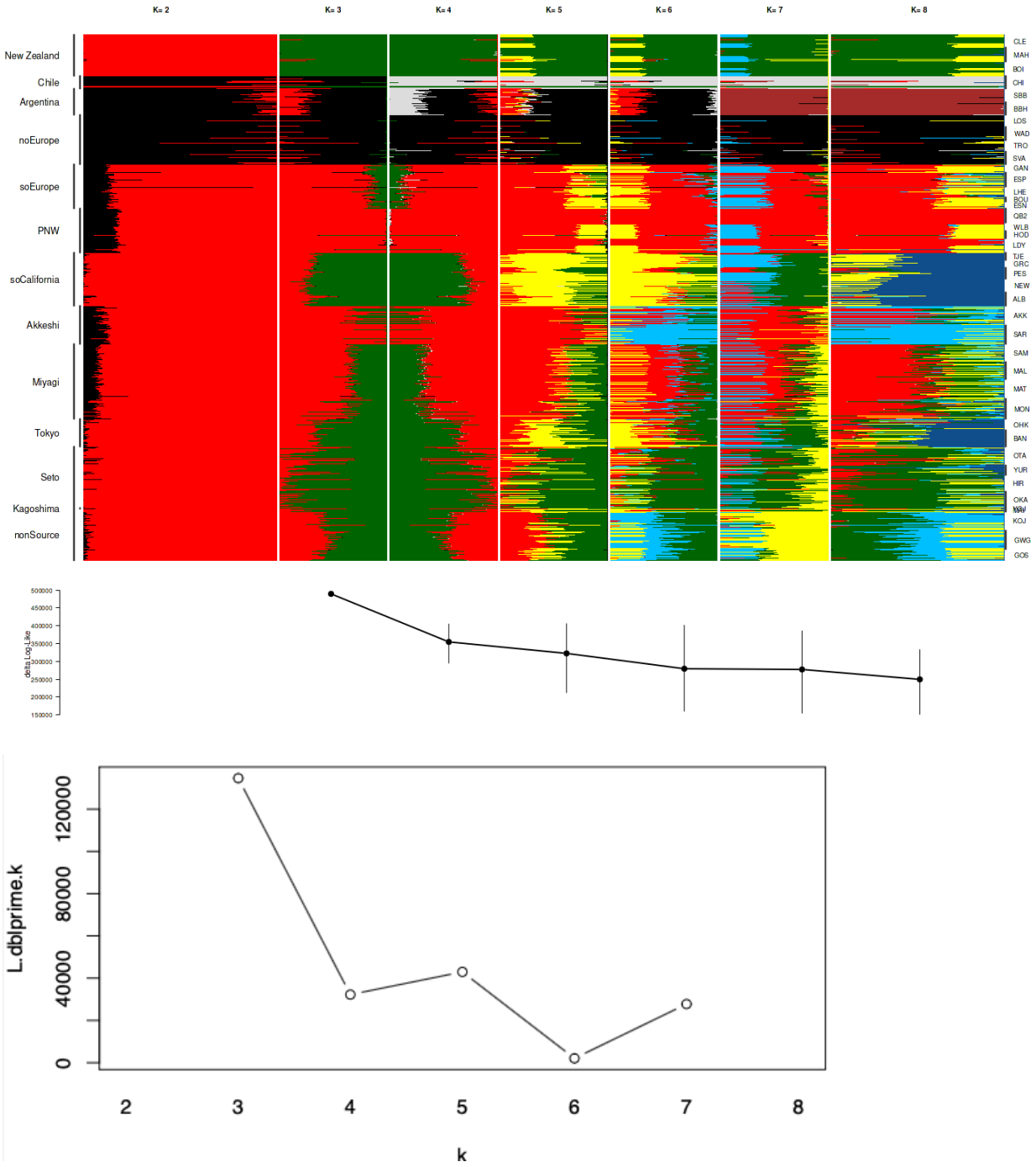


Figure S9. Sea surface temperature (SST) correlates with expected heterozygosity (Hs) in the native range (Native r = 0.725, df = 13, p = 0.002) even after the admixed Akkeshi population (AKK) was removed (r = 0.714, df = 12; p = 0.004). SST weakly correlates with Hs in the introduced range (r = 0.344, df = 22; p = 0.100). When regions with an aquacultural history were removed, the pattern was slightly stronger (r = 0.472, df = 15, p = 0.056) but not significant at alpha = 0.05. Population colors reflect native regions in Figure 1B and circles reflect non-native populations.

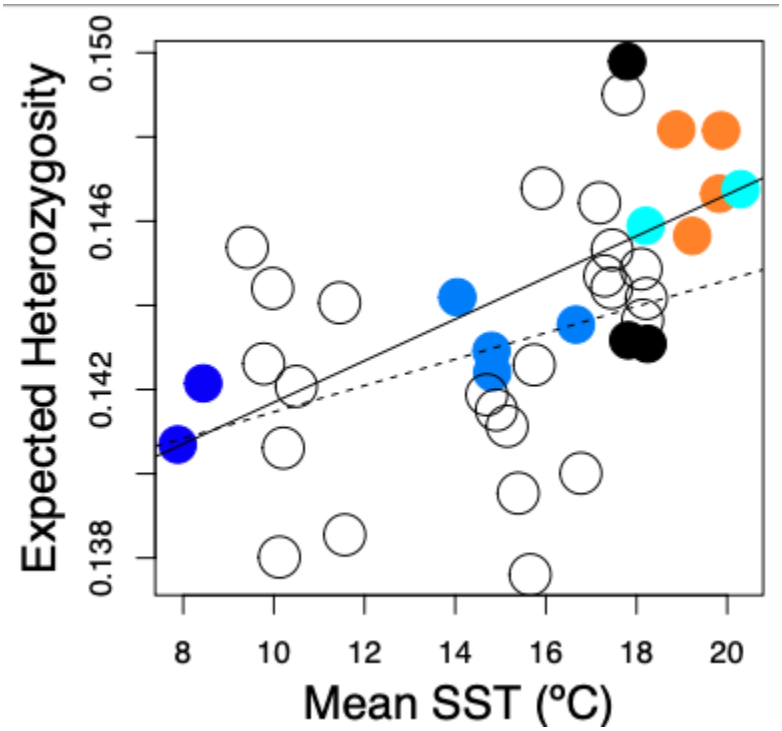


Figure S10. One possible configuration of history among demes used to simulate genotypes and summary statistics. On the tree, circle symbols at the tips correspond to native populations; the upper row of labels identify specific regions. Triangle symbols denote introduced populations and the lower row of labels identifies introduced regions. Filled symbols are sampled from each simulation and correspond to locations also sampled in the empirical data. Open symbols represent simulated, unsampled "ghost populations." Horizontal dashed lines correspond to parameters that identify in the past where lineages fuse into fewer populations. The color of the edges, red, black, and green, correspond to parameters for effective population size (Ne) in introduced populations, native populations, and the ancestral population, respectively. The sampled populations are consistent among simulations, as is the topology among native regions. Otherwise, times and population sizes are chosen from uniform distributions for each simulation.

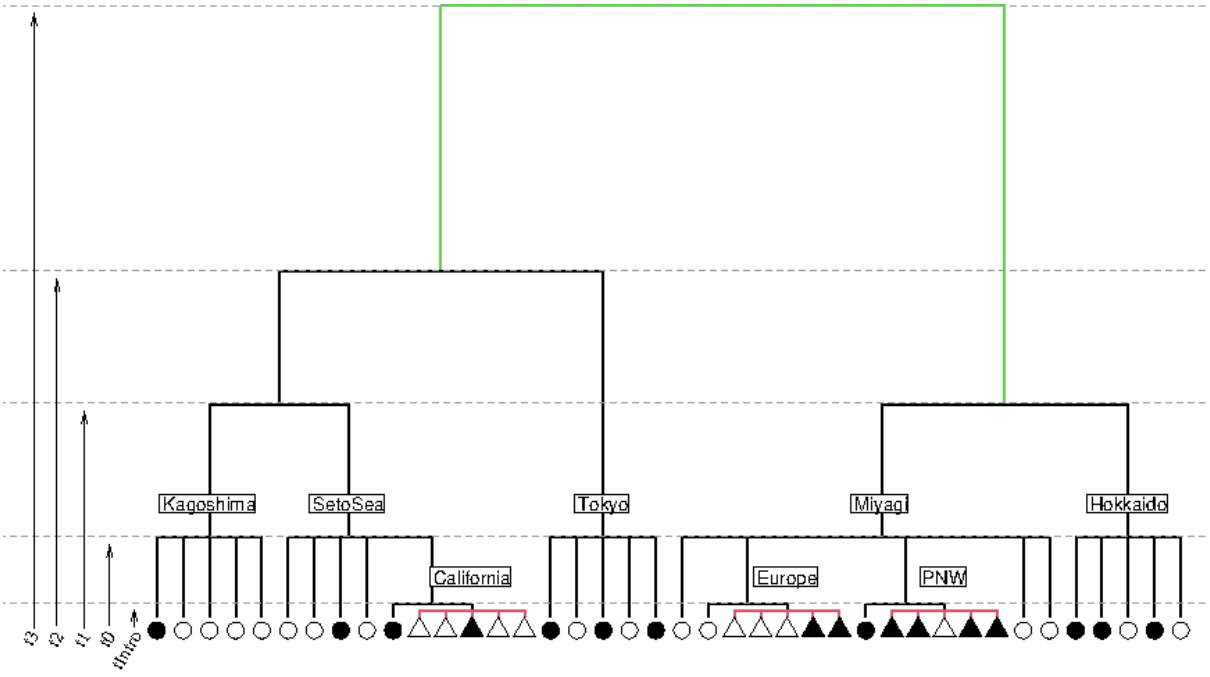


Figure S11. Ordination of simulated and empirical summary statistics for two introduced species, *Palaemon macrodactylus* and *Polydora hoplura*, respectively. The PC scores for summary statistics for the observed haplotypes in each species are indicated with the large magenta triangle with a black dot. The remaining points represent individual simulation replicates assuming equal prior probability of each model. The two max voyage duration shipping models are indicated in red and green for 30 and 60 days, respectively. The two alternatives for oyster introductions are indicated by dark and light blue. Simulations that allowed simultaneous introduction from all possible sources in each replicate are indicated with triangles and simulations that randomly chose a source based on its relative probability are indicated with circles.

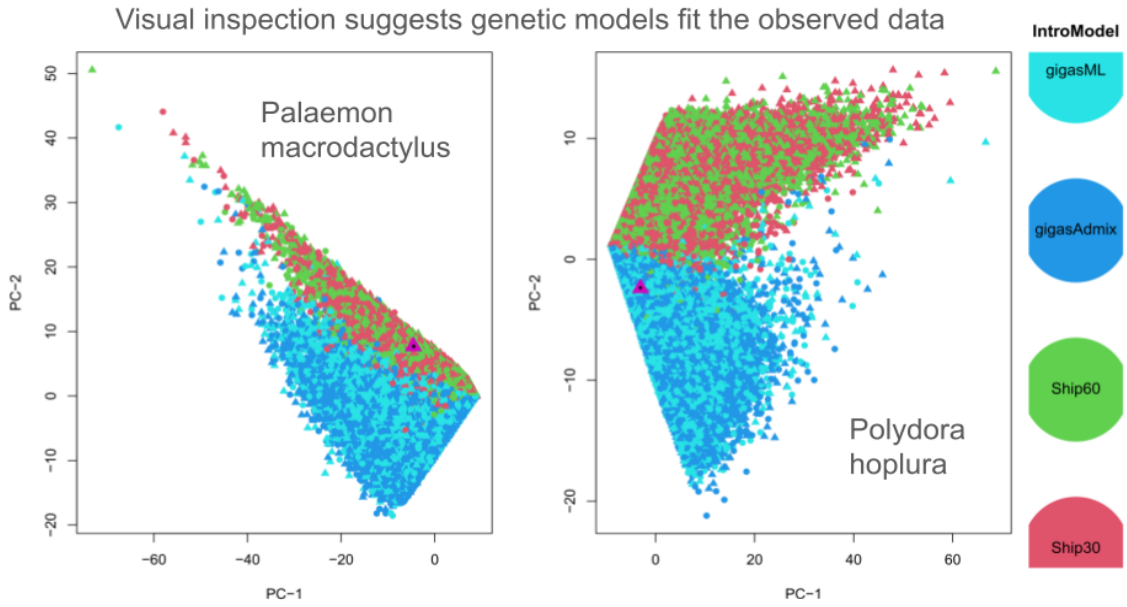


Figure S12. Ordination of simulated and empirical summary statistics for *Acanthogobius flavimanus*. The PC scores for summary statistics for the observed haplotypes are indicated with the large magenta triangle with a black dot. The remaining points represent individual simulation replicates assuming equal prior probability of each model. The two max voyage duration shipping models are indicated in red and green for 30 and 60 days, respectively. The two alternatives for oyster introductions are indicated by dark and light blue. Simulations that allowed simultaneous introduction from all possible sources in each replicate are indicated with triangles and simulations that randomly chose a source based on its relative probability are indicated with circles.

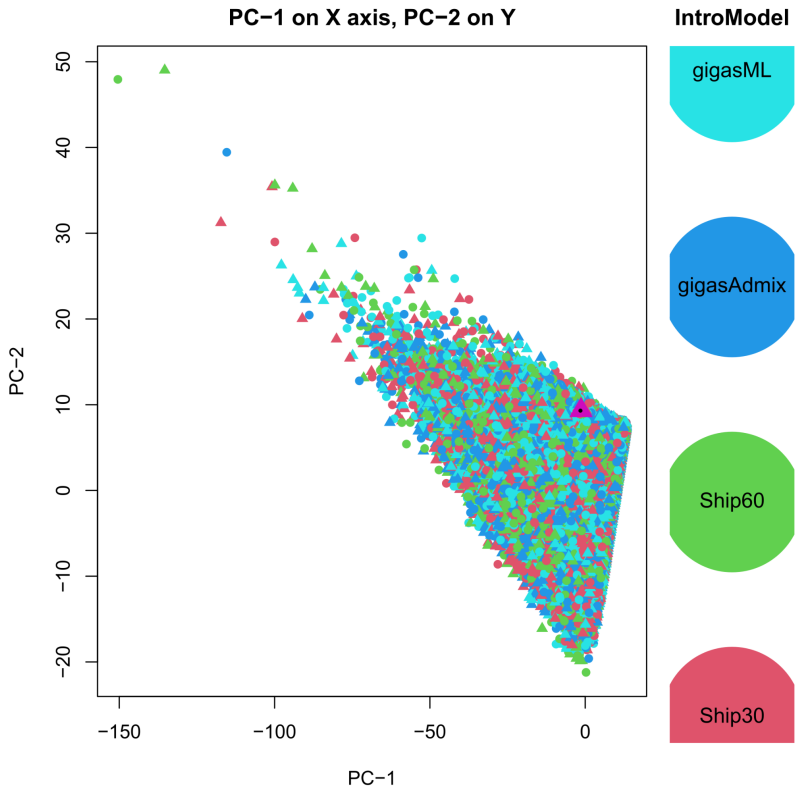
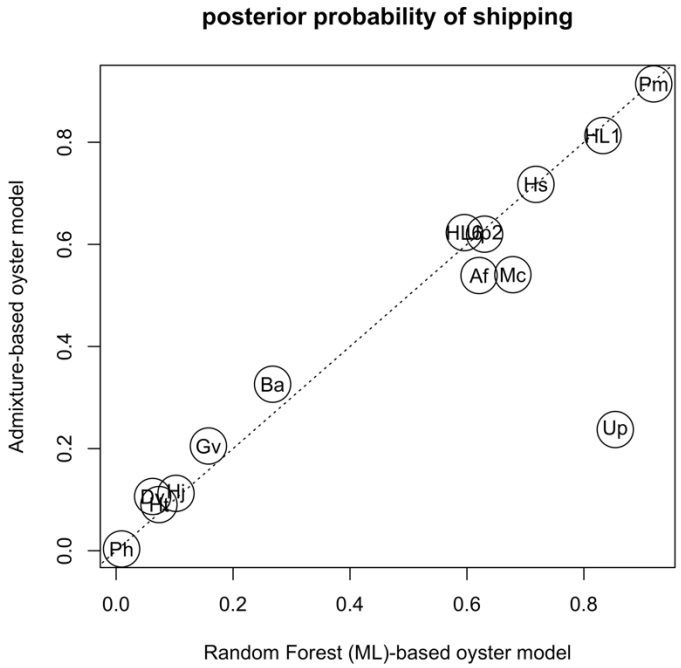


Figure S13. Posterior probability of a shipping vector (0 = oyster; 1 = shipping) from the admixture-based and ML-based oyster model. Key: Polydora hoplura (Ph), Haminoea japonica (Hj), Batillaria attramentaria (Ba), Cercaria batillaria (HL6), Hemigrapsus takanoi (Ht), Gracilaria vermiculophylla (Gv), Ulva pertusa (Up), Acanthogobius flavimanus (Af), Mutimo cylindricus (Mc), Undaria pinnatifida (Up), Cercaria batillaria (HL1), Hemigrapsus sanguineus (Hs), Ulva pertusa (Up2), Didemnum vexillum (Dv), Palaemon macrodactylus (Pm). The line denotes 1:1 parity among the estimates.



Bibliography – Supplemental Materials

1. C. Peñaloza, *et al.*, A chromosome-level genome assembly for the Pacific oyster (Crassostrea gigas). *Gigascience* **10**, 1–28 (2021).

2. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

3. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).

4. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356–356 (2014).

5. B. J. Knaus, N. J. Grünwald, vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).

6. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

7. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

8. S. Hirase, *et al.*, Phylogeography of the yellowfin goby Acanthogobius flavimanus in native and non-native distributions. *Mar. Biol.* **164**, 106–106 (2017).

9. O. Miura, M. E. Torchin, A. M. Kuris, R. F. Hechinger, S. Chiba, Introduced cryptic species of parasites exhibit different invasion pathways. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19818–19823 (2006).

10. L. Stefaniak, *et al.*, Determining the native region of the putatively invasive ascidian Didemnum vexillum Kott, 2002. *J. Exp. Mar. Bio. Ecol.* **422–423**, 64–71 (2012).

11. G. Lambert, Adventures of a sea squirt sleuth: unraveling the identity of Didemnum vexillum, a global ascidian invader. *Aquat. Invasions* **4**, 5–28 (2009).

12. S. Y. Kim, F. Weinberger, S. M. Boo, GENETIC DATA HINT AT A COMMON DONOR REGION FOR INVASIVE ATLANTIC AND PACIFIC POPULATIONS OF GRACILARIA VERMICULOPHYLLA (GRACILARIALES, RHODOPHYTA)1. *J. Phycol.* **46**, 1346–1349 (2010).

13. S. A. Krueger-Hadfield, *et al.*, When invaders go unnoticed: The case of*Gracilaria vermiculophylla*in the British isles. *Cryptogam. Algol.* **38**, 379–400 (2017).

14. S. A. Krueger-Hadfield, *et al.*, Genetic identification of source and likely vector of a widespread marine invader. *Ecol. Evol.* **7**, 4432–4447 (2017).

15. S. A. Krueger-Hadfield, *et al.*, Everywhere you look, everywhere you go, there's an estuary invaded by the red seaweed Gracilaria vermiculophylla (Ohmi) Papenfuss, 1967. *Bioinvasions Record* **7**, 343–355 (2018).

16. S. A. Krueger-Hadfield, *et al.*, Intraspecific diversity and genetic structure in the widespread macroalga Agarophyton vermiculophyllum. *J. Phycol.* 1–8 (2021).

17. D. Hanson, Y. Hirano, Á. Valdés, Population genetics of Haminoea (Haloa) japonica Pilsbry, 1895, a widespread non-indigenous sea slug (Mollusca: Opisthobranchia) in North America and Europe. *Biol. Invasions* **15**, 395–406 (2013).

18. A. M. H. Blakeslee, *et al.*, Reconstructing the Invasion History of the Asian shorecrab, Hemigrapsus sanguineus (De Haan 1835) in the Western Atlantic. *Mar. Biol.* **164**, 1–19 (2017).

19. M. J. Raupach, *et al.*, The Application of DNA Barcodes for the Identification of Marine Crustaceans from the North Sea and Adjacent Regions. *PLoS One* **10**, e0139421 (2015).

20. W. Makino, *et al.*, Evidence of multiple introductions and genetic admixture of the Asian brush-clawed shore crab Hemigrapsus takanoi (Decapoda: Brachyura: Varunidae) along the Northern European coast. *Biol. Invasions* **20**, 825–842 (2018).

21. T. Hanyuda, G. I. Hansen, H. Kawai, Genetic identification of macroalgal species on Japanese tsunami marine debris and genetic comparisons with their wild populations. *Mar. Pollut. Bull.* **132**, 74–81 (2018).

22. K. Kogishi, T. Kitayama, K. A. Miller, T. Hanyuda, H. Kawai, Phylogeography of Cutleria cylindrica (cutleriales, phaeophyceae) in northeastern asia, and the identity of an introduced population in California. *J. Phycol.* **46**, 553–558 (2010).

23. C. Lejeusne, *et al.*, High genetic diversity and absence of founder effects in a worldwide aquatic invader. *Sci. Rep.* **4** (2014).

24. V. I. Radashevsky, *et al.*, Searching for a home port in a polyvectic world: Molecular analysis and global biogeography of the marine worm Polydora hoplura (Annelida: Spionidae). *Biology (Basel)* **12** (2023).

25. P. G. Sauriau, *et al.*, Multiple genetic marker analysis challenges the introduction history of Ulva australis (Ulvales, Chlorophyta) on French coasts. *Eur. J. Phycol.* **00**, 455–467 (2021).

26. S. Uwai, *et al.*, Genetic diversity in Undaria pinnatifida (Laminariales, Phaeophyceae) deduced from mitochondria genes – origins and succession of introduced populations. *Phycologia* **45**, 687–695 (2006).

27. P. Pudlo, *et al.*, Reliable ABC model choice via random forests. *Bioinformatics* **32**, 859–866 (2016).

28. M. Sitarz, Extending F1 metric, probabilistic approach. *Adv Artif Intell Mach Learn* **abs/2210.11997**, 1025–1038 (2022).

29. L. Excoffier, *et al.*, Fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).

30. J.-M. Marin, L. Raynal, P. Pudlo, C. P. Robert, A. Estoup, abcrf: Approximate Bayesian Computation via Random Forests. [Preprint] (2022). Available at: https://CRAN.R-project.org/package=abcrf.

31. D. Hanson, *et al.*, Slipping through the Cracks: The Taxonomic Impediment Conceals the Origin and Dispersal of Haminoea japonica, an Invasive Species with Impacts to Human Health. *PLoS One* **8**, 1–12 (2013).

32. F. I. Archer, P. E. Adams, B. B. Schneiders, stratag: An r package for manipulating, summarizing and analysing population genetic data. *Mol. Ecol. Resour.* **17**, 5–11 (2017).

33. E. Paradis, pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).

34. T. Jombart, I. Ahmed, adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).