



**HAL**  
open science

## Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*

Pierre-Alexandre Gagnaire, Jean-Baptiste Lamy, Florence Cornette, Serge Heurtebise, Lionel Degremont, Emilie Flahauw, Pierre Boudry, Nicolas Bierne, Sylvie Lapègue

### ► To cite this version:

Pierre-Alexandre Gagnaire, Jean-Baptiste Lamy, Florence Cornette, Serge Heurtebise, Lionel Degremont, et al.. Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biology and Evolution*, 2018, 10 (9), pp.2518-2534. 10.1093/gbe/evy194 . hal-02395878

**HAL Id: hal-02395878**

**<https://hal.science/hal-02395878>**

Submitted on 5 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*

Pierre-Alexandre Gagnaire<sup>1</sup>, Jean-Baptiste Lamy<sup>2</sup>, Florence Cornette<sup>2</sup>, Serge Heurtebise<sup>2</sup>, Lionel Dégremont<sup>2</sup>, Emilie Flahauw<sup>2</sup>, Pierre Boudry<sup>3</sup>, Nicolas Bierne<sup>1</sup>, and Sylvie Lapègue<sup>2,\*</sup>

<sup>1</sup>Institut des Sciences de l'Evolution, ISEM-CNRS, UMR5554, Montpellier, France

<sup>2</sup>Ifremer, SG2M-LGPM, Laboratoire de Génétique et Pathologie des Mollusques Marins, La Tremblade, France

<sup>3</sup>Ifremer, UMR LEMAR, Laboratoire des Sciences de l'Environnement Marin (UBO, CNRS, IRD, Ifremer), Plouzané, France

\*Corresponding author: E-mail: sylvie.lapegue@ifremer.fr.

**Accepted:** September 2, 2018

**Data deposition:** This project has been deposited at <http://dx.doi.org/10.12770/dbf64e8d-45dd-437f-b734-00b77606430a>.

## Abstract

The Pacific cupped oyster is genetically subdivided into two sister taxa, *Crassostrea gigas* and *Crassostrea angulata*, which are in contact in the north-western Pacific. The nature and origin of their genetic and taxonomic differentiation remains controversial due the lack of known reproductive barriers and the high degree of morphologic similarity. In particular, whether the presence of ecological and/or intrinsic isolating mechanisms contributes to species divergence is unknown. The recent co-introduction of both taxa into Europe offers a unique opportunity to test how genetic differentiation is maintained under new environmental and demographic conditions. We generated a pseudochromosome assembly of the Pacific oyster genome using a combination of BAC-end sequencing and scaffold anchoring to a new high-density linkage map. We characterized genome-wide differentiation between *C. angulata* and *C. gigas* in both their native and introduced ranges, and showed that gene flow between species has been facilitated by their recent co-introductions in Europe. Nevertheless, patterns of genomic divergence between species remain highly similar in Asia and Europe, suggesting that the environmental transition caused by the co-introduction of the two species did not affect the genomic architecture of their partial reproductive isolation. Increased genetic differentiation was preferentially found in regions of low recombination. Using historical demographic inference, we show that the heterogeneity of differentiation across the genome is well explained by a scenario whereby recent gene flow has eroded past differentiation at different rates across the genome after a period of geographical isolation. Our results thus support the view that low-recombining regions help in maintaining intrinsic genetic differences between the two species.

**Key words:** cupped oysters, genome assembly, species divergence, reproductive barriers, recombination rate.

## Introduction

Broadcast-spawning marine invertebrates usually display high genetic similarity between adjacent populations, a consequence of combining high dispersal ability and large population sizes. Many such species, however, appear to be genetically subdivided into sibling species, cryptic species-pairs, ecotypes, or partially reproductively isolated populations (Knowlton 1993; Palumbi 1994; Hellberg 2009; Bierne et al. 2011; Gagnaire et al. 2015; Kelley et al. 2016). The

ability to score divergence at the genomic level in nonmodel organisms is revealing an increasing number of cases of cryptic species subdivision in broadcast-spawning marine invertebrates (e.g., Westram et al. 2014; Ravinet et al. 2015; Fraïsse et al. 2016; Rose et al. 2018). These recent findings confirm the view that semipermeable barriers to gene flow between closely related taxa (Harrison and Larson 2014, 2016) are relatively frequent in the marine realm, and may explain short-distance genetic differentiation patterns

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

that are seemingly in contradiction with species' dispersal potential.

The existence of marine semi-isolated species pairs evolving in the "grey zone" of the speciation continuum (Roux et al. 2016), that is, before complete reproductive isolation, provides interesting opportunities to contribute to some highly debated questions in the field of speciation genomics. Speciation is a progressive process by which reproductive isolation barriers of various types progressively appear and combine their effects to reduce gene flow (Abbott et al. 2013). As long as speciation is not complete, diverging populations continue to evolve nonindependently because some regions of their genome can still be exchanged. This results in a dynamic architecture of divergence characterized by temporal changes in the number, genomic distribution and magnitude of the genetic differences between incipient species. One of the speciation genomics approaches investigates this architecture by characterizing heterogeneous genome divergence patterns, ultimately aiming at a detection of the loci directly involved in reproductive isolation (Feder et al. 2012). This strategy, however, has come up against a large diversity of evolutionary processes influencing the genomic landscape of species divergence (Yeaman et al. 2016; Ravinet et al. 2017; Wolf and Ellegren 2017), some of which are not directly linked to speciation (Noor and Bennett 2009; Nachman and Payseur 2012; Cruickshank and Hahn 2014; Burri 2017a). The field is now progressively moving toward a more mechanistic understanding of the evolutionary processes underlying heterogeneous genome divergence. However, the issue of distinguishing the impact of genetic barriers from the effect of confounding processes such as linked selection remains challenging (Burri 2017a). Because genetic divergence does not easily persist in the face of gene flow in the absence of genetic barriers (Bierne et al. 2013), high gene flow species such as broadcast-spawning marine invertebrates offer valuable study systems for disentangling the mechanisms at play during speciation.

Here, we investigated the existence and the type of genetic barriers between divergent lineages of the Pacific cupped oyster, which has been taxonomically subdivided into two sister species, *Crassostrea gigas* and *Crassostrea angulata*. The two species are presumed to be parapatrically distributed in their native range in the north-western Pacific, but the location of putative contact zones remains largely unknown. Whether *C. gigas* and *C. angulata* truly represent biological species, semi-isolated species or populations of the same species also remains unclear. The two taxa can be cross-fertilized in the laboratory to form viable and fertile offspring (Takeo and Sakai 1961; Huvet et al. 2001, 2002), and some authors have proposed that they should be considered as a single species (López-Flores et al. 2004; Reece et al. 2008). On the other hand, the finding of genetic differences between *C. angulata* and *C. gigas* lead other authors to conclude that they form different but genetically closely related species

(Boudry et al. 1998; Lapègue et al. 2004; Ren et al. 2010; Wang et al. 2014). One originality of the Pacific cupped oyster system is the co-introduction of both taxa into Europe. *Crassostrea angulata*, also called the Portuguese oyster, is presumed to have been nonvoluntarily introduced by merchant ships during the 16th century, probably from Taiwan (Boudry et al. 1998; Huvet et al. 2000) although the exact origins of introduced stocks remains unknown (Grade et al. 2016). Recent studies have shown that this species is widely distributed in Asian seas where it shows a high genetic diversity (Wang et al. 2010; Sekino and Yamashita 2013; Zhong, Li, Yu, et al. 2014; Hsiao et al. 2016), and also suggested a more complex history of transfers between Asia and Europe (Grade et al. 2016). The Pacific oyster, *C. gigas*, has been voluntarily introduced from Japan and British Columbia into Europe in the early 1970s, mainly to replace the Portuguese oyster in the French shellfish industry following a severe disease outbreak (Grizel and Héral 1991). Since then, the two species are in contact in southern Europe and therefore have the potential to exchange genes in a new environment (Huvet et al. 2004; Batista et al. 2017). Gene flow in the context of marine invasions has mainly been studied between native and nonindigenous lineages (Saarman and Pogson 2015; Viard et al. 2016) but rarely between two co-introduced genetic backgrounds in a new place. A notorious exception is the European green crab (*Carcinus maenas*) in the Northwest Atlantic (Darling et al. 2008). However, although the genome-wide genetic differentiation has been studied in the introduced range (Jeffery et al. 2017, 2018), it has not been compared with the differentiation observed in the native range to date.

In the present study, we first generated new genomic resources in the Pacific oyster to improve the species genome assembly and characterize chromosomal variation in recombination rate. Then, we tested the existence of genetic barriers between *C. angulata* and *C. gigas* by searching for genomic regions that remain differentiated in the presence of gene flow, accounting for the demographic divergence history of the species. We also evaluated whether ecological divergence driven by local adaptations is the main factor maintaining species divergence in the native range. We hypothesized that if this is the case, the different ecological conditions encountered by the two species in Europe would have reshaped the original genomic landscape of species divergence existing in the native Asian range. Finally, we attempted to relate genome-wide divergence patterns to underlying evolutionary processes including demography, selection and recombination constraints.

## Materials and Methods

### Biological Material for the Mapping Populations

Two F2 families (second generation of biparental crosses) were used for genetic map reconstruction in *C. gigas*. These

families were obtained by experimental breeding as part of the MOREST project (Boudry et al. 2008; Dégremont et al. 2010) by crossing families selected for resistance (*R*) or susceptibility (*S*) to summer mortality, which were subsequently found to have respectively higher and lower resistance to the herpesvirus OsHV-1 (Dégremont 2011). The F2-19 family was generated through biparental crossing between an *S* female and an *R* male to produce a F1 family from which a single sib-pair was randomly sampled to produce the F2 progeny. The F2-21 family was obtained under the same mating scheme but starting with an *R* female and a *S* male. A total of 293 and 282 F2 progenies were used for map reconstruction in family F2-19 and F2-21, respectively. For each individual, whole genomic DNA was isolated from gill tissue using the QIAamp DNA minikit (Qiagen). DNA was checked for quality by electrophoresis on agarose gel and then quantified using the Quant-iT PicoGreen dsDNA assay kit (Life Technologies).

### Genotyping Panel for Low-Density Map Construction

We genotyped F1 parents and their F2 progenies in both families (F2-19: 2 F1 and 293 F2; F2-21: 2 F1 and 282 F2) using a panel of 384 SNPs previously developed in *C. gigas* (Lapègue et al. 2014). SNP genotyping was performed using the Golden Gate assay and analyzed with the Genome Studio software (Illumina Inc.). In addition, 42 microsatellite markers were also genotyped according to published protocols (Li et al. 2003; Taris et al. 2005; Yamtich et al. 2005; Li et al. 2010; Sauvage et al. 2010) in order to include markers from previous generation linkage maps in *C. gigas* (Hubert and Hedgecock 2004; Sauvage et al. 2010).

### RAD Genotyping for High-Density Map Construction

We selected 106 progenies from each family as well as their four F1 parents for RAD library construction following the original protocol (Baird et al. 2008). Briefly, 1  $\mu$ g of genomic DNA from each individual was digested with the restriction enzyme *Sbf*I-HF (New England Biolabs), and then ligated to a P1 adapter labeled with a unique barcode. We used 16 barcodes of 5-bp and 16 barcodes of 6-bp long in our P1 adapters to build 32-plex libraries. Seven pools of 32 individuals were made by mixing individual DNA in equimolar proportions. Each pool was then sheared to a 350 pb average size using a Covaris S220 sonicator (KBIosciences), and size-selected on agarose gel to keep DNA fragments within the size range 300–700 pb. Each library was then submitted to end-repair, A-tailing and ligation to P2 adapter before PCR amplification for 18 cycles. Amplification products from six PCR replicates were pooled for each library, gel-purified after size selection and quantified on a 2100 Bioanalyzer using the High Sensitivity DNA kit (Agilent). Each library was sequenced on a separate lane of an Illumina HiSeq 2000 instrument by IntegraGen Inc. (France, Evry), using 100-bp single reads.

We used the program *Stacks* (Catchen et al. 2011, 2013) to build loci from short-read sequences and determine individual genotypes. Raw sequence reads were quality filtered and demultiplexed using *process\_radtags.pl* before being trimmed to 95 bp. We explored different combinations of parameter values for the minimum stack depth (*-m*) and the maximum mismatch distance (*-M*) allowed between two stacks to be merged into a locus. We found that the combination *-m* 3 *-M* 7 represented the best compromise to avoid overmerging loci, while providing an average number of 2.3 SNPs per polymorphic RAD locus (supplementary fig. S1, Supplementary Material online) which is consistent with the high polymorphism rate in *C. gigas* (Sauvage et al. 2007; Zhang et al. 2012). Individual de novo stack formation was done with *ustacks* (*-m* 3 *-M* 7 *-r* *-d*). We then built a catalog of loci using all individuals from both families with *cstacks* (*-n* 7) and matched back all the samples against this catalog using *sstacks*. After this step, the two families were treated as two separate populations to produce a table of observed haplotypes at each locus in each family using *populations*.

We developed a Bayesian approach that uses information from progenies' genotypes to correct for missing and miscalled genotypes in parental samples. The probability of a given combination of parental genotypes ( $G_P$ ) conditional on the genotypes observed in their descendants ( $G_D$ ) is given by

$$P(G_P|G_D) = \frac{P(G_D|G_P) \times P(G_P)}{P(G_D)}$$

where the probability of the observed counts of progenies' genotypes ( $N_{AA}, N_{AB}, N_{AC}, N_{AD}, N_{BB}, N_{BC}, N_{BD}$ ) conditional on the genotypes of their parents is drawn from a multinomial distribution that takes different parameter values for each of the six alternative models of parental genotypes (i.e., crosses AA  $\times$  AA, AA  $\times$  AB, AB  $\times$  AB, AA  $\times$  BB, AB  $\times$  AC, and AB  $\times$  CD). Each RAD locus was treated independently using observed haplotypes to determine the best model of parental haplotype combination. When the actual haplotypes called in parents did not match the best model, a correction was applied to restore the most likely combination of parental haplotypes, taking read depth information into account. Data were finally exported in *JoinMap 4* format with population type CP (van Ooijen 2006). The R script for correcting missing and miscalled haplotypes in parental samples is available from the authors upon request.

### Genetic Map Construction

A new Pacific oyster linkage map was constructed in four successive steps: 1) First, a low-density map was built for each family using the SNP and microsatellite markers data set. These two maps were used to provide accurate ordering of markers, since both mapping populations comprised  $\sim$ 300 individuals with very few missing genotypes (0.5%). 2) In a



second step, we tried to reach a consensus order for the markers that were included in the two previous maps, in order to determine a set of anchor loci for each linkage group. 3) Third, we included RAD markers and determined the order of all loci in each mapping family after setting a fixed order for the anchor loci. 4) Finally, we integrated the two high-density linkage maps to produce a consensus genetic map for the Pacific oyster.

All linkage mapping analysis were performed using *JoinMap 4* (van Ooijen 2006). Markers showing significant segregation ratio distortion after Bonferroni correction ( $P < 0.05$ ) were removed from the analysis. Markers were grouped using an independence LOD threshold of 16 for the SNP and microsatellite marker data sets and a LOD threshold of 10 for the RAD marker data sets. Additional ungrouped markers were assigned at a LOD threshold of 5 using their strongest cross-link information. We used the regression mapping algorithm to build the maps using a recombination frequency threshold of 0.4, a minimal LOD score of 1 and a goodness-of-fit jump value of 5. The ordering of markers was optimized using the ripple function after each added locus. Map distances in centiMorgans (cM) were calculated from recombination frequencies using Kosambi's mapping function.

### Identification of Chimeric Scaffolds and Reassembly Using BAC-End Scaffolding

In order to detect chimeric scaffolds in the oyster\_v9 assembly (Zhang et al. 2012), the consensus sequences of the markers included in the new linkage map were blasted against the reference genome. An  $E$ -value threshold of  $10^{-30}$  and a minimum identity threshold of 90% were set to retain only highly significant matches. Assembly errors were identified by scaffolds anchored to more than a single linkage group. Chimeric scaffolds were subsequently splitted at all stretches of Ns connecting adjacent contigs.

In order to improve the scaffolding of the Pacific oyster genome, we sequenced 73,728 BAC clones of 150 kb average insert size (Gaffney 2008) from both ends using Sanger sequencing at the Genoscope facility (Evry, France). This resulted in 60,912 cleaned full-length (i.e.,  $>1,300$  bp) BAC-End sequences (BES) pairs. Each sequence from each pair was trimmed to only conserve 1,000 bp between positions 10 and 1010, and clipped into 19 evenly-spaced 100-mers overlapping by 50 bp. This clipping procedure aimed at constructing a high quality short-read paired-end data set from our BES data set.

Clipped BES were used for an additional round of contig extension and scaffolding with *SSPACE* (Boetzer et al. 2011). For each LG, we used unambiguously anchored scaffolds and contigs originating from splitted scaffolds matching to this LG. Clipped BES were aligned to scaffolds using *Bowtie2* (Langmead and Salzberg 2012), allowing at most 3

mismatches per 100-bp read. Scaffolding parameters to *SSPACE* were set to a minimum of 5 links (-k) to validate a new scaffold pair and a maximum link ratio of 0.7 (-a). Scaffolding was only permitted between scaffolds of at least 1,000 bp, using an allowed insert size range of 75–225 kb.

### Scaffold Anchoring to the Linkage Map

We searched for nonambiguous associations between the new set of extended scaffolds and markers included in the consensus linkage map using the same blasting procedure as prior to genome reassembly. We constructed pseudochromosomes by positioning scaffolds along each linkage group using *Harry Plotter* (Moretto et al. 2010). Because the genetic resolution of our consensus map was still limited by the size of our mapping populations ( $\sim 100$  individuals each), many scaffolds were anchored by a single marker or had too few markers to determine their orientation. Therefore, we did not determine scaffold orientation. Scaffolds that were anchored to multiple markers were positioned using the average cM value of their anchor loci. Unmapped scaffolds were placed in an artificial chromosome named UN.

### Local Recombination Rate Estimation

We used the R package *MareyMap v1.2* (Rezvoy et al. 2007) to estimate local recombination rates along the genome, by comparing the consensus linkage map and the physical map for each pseudochromosome. The relationship between genetic and physical distances was first visualized to remove outlier markers resulting from scaffold misplacement within pseudochromosomes. We then used the Loess interpolation method which estimates recombination rates by locally adjusting a 2nd degree polynomial curve, setting the span parameter value to 0.25.

### RAD Genotyping of Natural Populations

We sampled four wild populations of *C. gigas* and *C. angulata* from both their native Asian and introduced Atlantic areas. We used 24 individuals of *C. gigas* from both Japan (JAP, native) and French Marennes-Oléron Bay (LAF, introduced), and 24 individuals of *C. angulata* from both Taiwan (KEE, native) and Portugal (SAD, introduced). We prepared and sequenced four 24-plex RAD libraries using the same protocol as described above.

Cleaned demultiplexed reads were mapped to each newly assembled pseudochromosome using *Bowtie2* (Langmead and Salzberg 2012) with the very-sensitive option, allowing a maximum of seven mismatches per alignment. SNPs were called from aligned reads with *Stacks* using a minimum read depth of  $5\times$  per individual per allele (Catchen et al. 2011, 2013). The correction module *rxstacks* was used to re-evaluate individual genotypes and exclude low-quality variants with a cutoff log-likelihood value of  $-500$ . Only RAD

loci that were successfully genotyped in at least 80% of the samples in each population were retained for subsequent population genomic analyses.

### Population Genomic Analyses

We used *VCFTools* v0.1.11 (Danecek et al. 2011) to apply within-population filters in order to improve the quality of our SNP data set. We excluded loci showing >4 missing genotypes over 24 individuals, as well as markers showing departure from Hardy–Weinberg equilibrium within at least one population using a *P*-value cutoff of 0.01. Nucleotide diversity ( $\pi$ ), computed as the average number of pairwise differences, was estimated from retained loci for each population within 150 kb windows. Genetic differentiation between all possible pairs of populations was estimated SNP by SNP as well as in 150 kb windows using  $F_{ST}$  (Weir and Cockerham 1984). Within- and between-population components of genetic diversity were decomposed using a discriminant analysis of principal components (dAPC), in order to maximize genetic variation between populations while minimizing within-population variance. The dAPC analysis was performed with the R package *Adegenet* (Jombart 2008), using the global SNP data set containing the two populations from both species with only one randomly selected SNP per RAD locus.

In order to evaluate the influence of recombination rate variation on genetic differentiation and absolute genetic diversity between species, we tested for correlations across the entire genome using Spearman correlation test. Recombination rate values were averaged in 500 kb windows to increase the number of informative sites per window. We excluded windows with recombination rate values exceeding 10 cM/Mb (i.e., corresponding to the 99th percentile of the distribution of estimated recombination rate). This cutoff aimed at removing outlying recombination rate values estimated along chromosome IX (fig. 1). In addition, we calculated the average  $F_{ST}$  in 500 kb windows, as well as the between-species genetic diversity ( $\pi_B$ ) following (Aeschbacher et al. 2017), which is the equivalent of  $d_{XY}$  in the absence of phased sequence data. Both statistics were averaged between the Asian and European species pairs. Finally, we used a nonparametric quantile regression approach to test whether the 97.5th quantile of the distribution of  $F_{ST}$  and  $\pi_B$  were related to the local average recombination rate. The regression fits were computed at each of 10 equally spaced recombination intervals distributed over the range of recombination values. We used *Treemix* v1.12 (Pickrell and Pritchard 2012) to test for the presence of gene flow between diverged populations of oysters. This program represents the genetic relationships among populations by building a population tree, allowing for gene flow in addition to population splits by introducing migration edges between populations of the tree. The optimal number of migration edges was determined by comparing likelihood values of fitted models with

increasing numbers of migration edges. These models were fitted to the global SNP data set with only one randomly selected SNP per RAD locus to avoid linkage disequilibrium. We then used the  $f_3$  statistics (Reich et al. 2010) to further test whether the recipient populations detected by *Treemix* displayed evidence of admixture. The value  $f_3(X, Y, Z)$  is negative when the focal population (X) is not related to populations Y and Z by a simple tree, but rather appears to be a mixture of Y and Z. Therefore, more negative  $f_3$  values indicate stronger evidence for admixture. We applied this test on two separate data sets consisting of unlinked SNPs from either low- or high-recombination regions. A threshold value of 2.21 cM/Mb corresponding to the genome-wide median recombination rate was used to assign SNPs to low- and high-recombination categories. This partitioning of the genome aimed at further testing the influence of the local recombinational environment on gene flow.

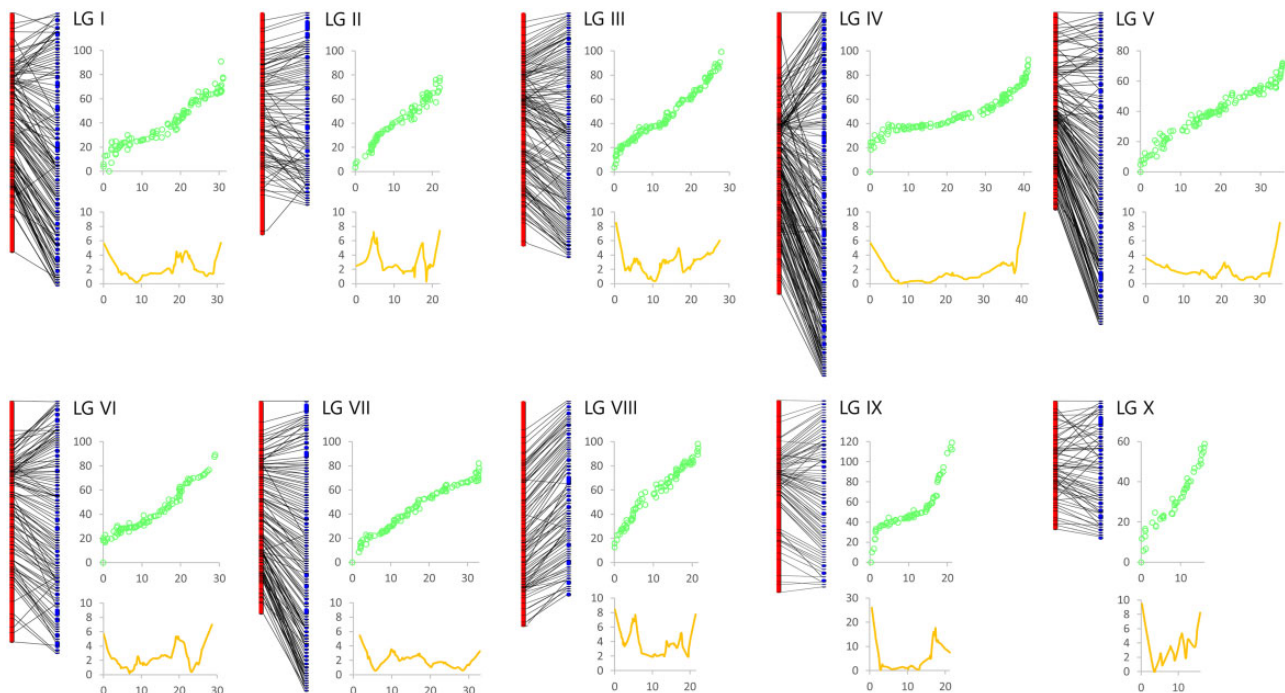
### Inference of the Demographic Divergence History

We inferred the demographic divergence history between *C. angulata* and *C. gigas* in both their native and introduced ranges using a version of the program *δadi* (Gutenkunst et al. 2009) modified by Tine et al. (2014). The observed joint allele frequency spectrum (JAFS) of each population pair was obtained by randomly selecting one SNP for each pair of RAD loci associated to the same restriction site in order to avoid the effect of linkage between SNPs, and then by projecting the data to 20 chromosomes per population to reduce the JAFS size. We first considered four classical models of divergence including models of strict isolation (SI), isolation-with-migration (IM), ancient migration (AM), and secondary contact (SC). We also considered extensions of the three divergence-with-gene-flow models assuming two categories of loci occurring in proportions *P* and 1–*P* and having different effective migration rates: IM2*m*, AM2*m*, and SC2*m* (Tine et al. 2014). These models offer simplified but more realistic representations of the speciation process, by enabling loci to be either neutrally exchanged between populations, or to have a reduced effective migration rate to account for the direct and indirect effects of selection. Each model was fitted to the observed JAFS (singletons were masked) using three successive optimization steps: Hot simulated annealing, cold simulated annealing and BFGS. Comparison among models was made using the Akaike Information Criterion (AIC).

## Results

### Construction of an Integrated High-Density Linkage Map

The microsatellite and SNP data sets used for constructing the low-density linkage map contained 133 informative markers in family F2-19 (293 progeny) and 137 informative markers in family F2-21 (282 progeny). Ten linkage groups were found in



**FIG. 1.**—Anchoring of newly edited scaffolds (blue) into pseudochromosomes using the markers included in the 10 linkage groups (red) of the new consensus linkage map. Comparison of the physical (x axis, Mb) and the genetic (y axis, cM) maps are provided for each linkage group (green points, outlier values removed), as well as the local recombination rate (orange line) estimated using the Loess interpolation method in *MareyMap*.

each family in agreement with the haploid number of chromosomes of the species, and the two maps generated by the first round of regression mapping respectively contained 98 and 105 mapped markers in family F19 and F21. The integrated map obtained after identifying homologous LG pairs between families contained 136 markers successfully ordered after the first round of regression mapping. This sex-averaged consensus low-density linkage map displayed a strong collinearity with the two maps from which it was derived ([supplementary fig. S2, Supplementary Material online, left panel](#)), indicating that the ordering of markers was highly consistent between the two families. We thus used it as an anchoring map to support the construction of the high-density RAD-linkage map.

The average number of filtered RAD-Sequencing reads obtained per individual was similar between family F2-19 ( $1.8 \times 10^6$ ) and F2-21 ( $1.9 \times 10^6$ ). After correcting for missing and miscalled haplotypes in parental samples, the number of informative markers was 1,278 in family F2-19 and 996 in family F2-21. A total of 1,855 markers were assigned to linkage groups (1,231 in family F19 and 935 in family F21, 311 in common), and the number of markers per linkage groups was highly positively correlated between families ( $R^2 = 0.71$ , [supplementary fig. S3, Supplementary Material online](#)). Using the anchoring map to set fixed orders, the two RAD maps generated after three rounds of regression mapping respectively contained 1,023 and 705 mapped markers in family F19 and F21, respectively. The final combination of these two maps

resulted in a sex-averaged consensus map LG containing 1,516 markers (table 1, and [supplementary fig. S2, Supplementary Material online, right panel](#)). The total map length was 965 cM and the average spacing between two neighboring markers was 0.64 cM. We compared this new linkage map to the Pacific oyster's second generation linkage map (table S7 from [Hedgecock et al. 2015](#)) using scaffolds from the oyster reference genome (oyster\_v9, [Zhang et al. 2012](#)) as intermediate. Using 278 pairs of markers colocalized to the same scaffolds, we found a good collinearity between the two maps, as illustrated by linkage group-wise correlations between recombination distances ( $0.35 < R^2 < 0.90$ , [supplementary fig. S4, Supplementary Material online](#)).

### Pseudochromosome Assembly

A majority (78.3%) of the markers included in the new consensus RAD linkage map matched to the oyster reference genome (oyster\_v9, [Zhang et al. 2012](#)). Among the 592 scaffolds that were matched with mapped markers, 327 (55.2%) contained a single mapped marker, 127 (21.5%) contained two markers, and 138 (23.3%) contained three to eleven markers ([supplementary table S1, Supplementary Material online](#)). We found 117 (44.2%) chimeric scaffolds mapping to different linkage groups among the 265 scaffolds containing at least two markers. Splitting chimeric scaffolds into smaller contigs at poly-N stretches decreased the assembly N50 size from 401 kb to 218 kb.

**Table 1**Summary Statistics for the 10 Linkage Groups of the New *C. gigas* High-density Linkage Map and the Reassembled Physical Genome

LG	Correspondence with Hedgecock et al. (2015)	Length (cM)	Nb of Markers F2-19	Nb of Markers F2-21	Total nb of Markers	Marker Spacing	Physical Length of Anchored Scaffolds	Recombination Rate (cM/Mb)	Recombination Rate Marey Map (cM/Mb)
I	1	90.92	86	88	158	0.58	31113268	2.92	2.61
II	4	94.23	89	56	117	0.81	21949604	4.29	3.65
III	5	99.17	166	69	229	0.43	27979720	3.54	3.40
IV	7	116.98	149	103	226	0.52	41226161	2.84	2.29
V	10	72.01	95	74	163	0.44	35636424	2.02	2.18
VI	3	98.87	143	60	132	0.75	28856184	3.43	2.71
VII	8 & 2	88.78	113	71	157	0.57	33248857	2.67	2.36
VIII	9	108.20	64	68	131	0.83	22447078	4.82	4.13
IX	6	136.51	94	58	124	1.10	21192597	6.44	5.47
X	6	58.89	24	58	79	0.75	16594032	3.55	3.65
ALL	–	964.55	1023	705	1516	0.64	280243925	3.44	3.05

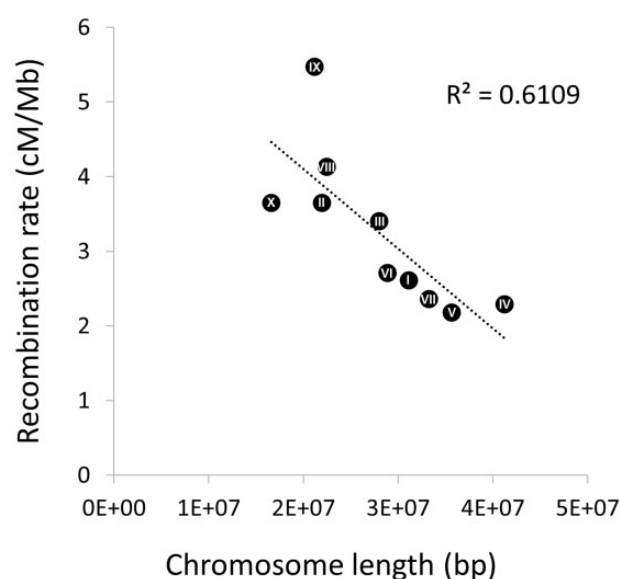
Contig extension and scaffolding using clipped BES allowed the merging of 4,162 contigs. The mean insert size of mapped BES was 116 kb, which was consistent with the mean insert size of the BAC end library. The resulting N50 of new scaffolds increased by 47% to 320 kb (supplementary table S2, Supplementary Material online). A total of 773 scaffolds from this new set of extended scaffolds were anchored to 1,161 loci of the consensus linkage map. Among them, 531 (68.7%) contained a single mapped marker, 153 (19.8%) contained two markers, and 89 (11.5%) contained three to seven markers (supplementary table S3, Supplementary Material online). The pseudochromosomes assembly obtained had a length of 280.2 Mb (fig. 1), representing 50.1% of the total length of the oyster\_v9 assembly (Zhang et al. 2012).

### Local Recombination Rate

The comparison between the pseudochromosome assembly and the consensus linkage map revealed variation in local recombination rate along chromosomes, with generally increased values toward linkage group extremities compared with central chromosomal regions (fig. 1). The average genome-wide recombination rate estimated with *MareyMap* was 3.05 cM/Mb (median = 2.21 cM/Mb), a value close to the total map length divided by the size of the physical assembly (3.44 cM/Mb). The real value, however, may be twice lower considering that the pseudochromosome assembly only represents 50% of the total genome size. The average chromosome-wide recombination rate (table 1) was negatively correlated with chromosome length ( $R^2 = 0.61$ , fig. 2), as expected under strong chiasma interference.

### Genetic Diversity and Differentiation

The average number of sequence reads retained per individual for calling genotypes was  $3.1 \times 10^6$ . After filtering for



**Fig. 2.**—Negative correlation between the length of newly assembled pseudochromosomes and the average chromosome-wide recombination rate assessed with *MareyMap*. Chromosome lengths and raw estimates of recombination rate are based on the physical chromosome lengths effectively covered by scaffolds. Rescaled values taking into account the presence of unmapped scaffolds can be obtained by multiplying (dividing) chromosome length (recombination rate) values by two (see text).

missing genotypes, HWE and a minor allele frequency of 1%, we kept a total of 10,144 SNPs from 1,325 RAD loci for downstream genetic diversity analyses. Within-population nucleotide diversity was elevated in both species and showed very similar levels between native and introduced populations (*C. angulata*:  $\pi_{KEE} = 0.0095$ ,  $\pi_{SAD} = 0.0096$ ; *C. gigas*:  $\pi_{JAP} = 0.0099$ ,  $\pi_{LAF} = 0.0101$ ). The genome-wide averaged differentiation assessed by  $F_{ST}$  between native and introduced populations from the same species was higher in *C. angulata* ( $F_{ST\ KEE-SAD} = 0.0179$ ) compared with *C. gigas* ( $F_{ST\ JAP/LAF} = 0.0130$ ). The mean genetic differentiation between



species was very similar between Asia ( $F_{ST}^{KEE/JAP} = 0.0440$ ) and Europe ( $F_{ST}^{SAD/LAF} = 0.0459$ ).

The between-population genetic structure revealed by the dAPC (fig. 3A) separated the two species along the first PC axis (explaining 15.78% of total variance) and the Asian and European populations of each species along the second axis (2.54% of total variance). The European populations of *C. angulata* (SAD) and *C. gigas* (LAF), were slightly shifted toward the center of the first PC axis, indicating an increased genetic similarity of the two species in Europe.

Genetic differentiation between species was highly heterogeneously distributed across the genome, with regions of elevated differentiation alternating with genetically undifferentiated regions in most chromosomes (fig. 4). The maximum  $F_{ST}$  value between *C. gigas* and *C. angulata* was 1 in Asia and 0.96 in Europe. Between-species genetic differentiation at individual SNPs was highly positively correlated between Asia and Europe, although it was on average lower in Europe (fig. 3B).

Chromosomal variations in average genetic differentiation ( $F_{ST}$ ) and between-species diversity ( $\pi_B$ ) were both related to the local recombination rate. More precisely, we found that  $F_{ST}$  and  $\pi_B$  were respectively negatively (fig. 5A, Spearman's  $\rho = -0.076^*$ ) and positively (fig. 5B, Spearman's  $\rho = 0.088^*$ ) related to the local recombination rate. Moreover, the within-species nucleotide diversity was significantly positively correlated with recombination (supplementary fig. S5, Supplementary Material online). Despite showing inverted relationships at the genome-wide scale, both  $F_{ST}$  and  $\pi_B$  showed decreasing trends in the maximal amount of divergence (measured by the 97.5th quantile of each distribution) with increasing local recombination rate (fig. 5). This

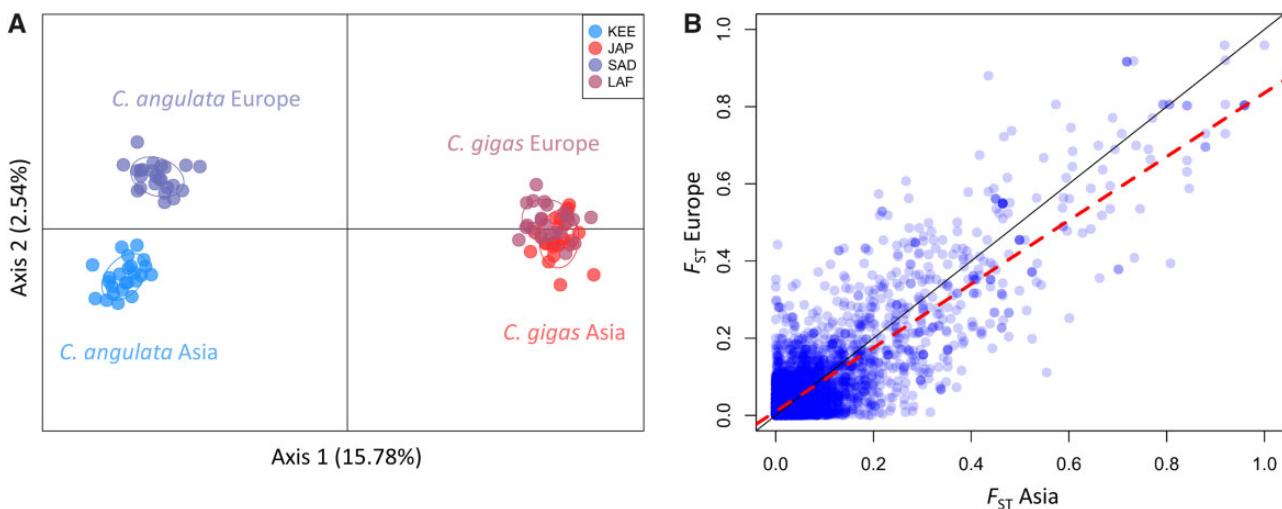
result supports that the highest values of both relative and absolute divergence preferentially map to low-recombining regions of the oyster genome.

### Inference of Gene Flow

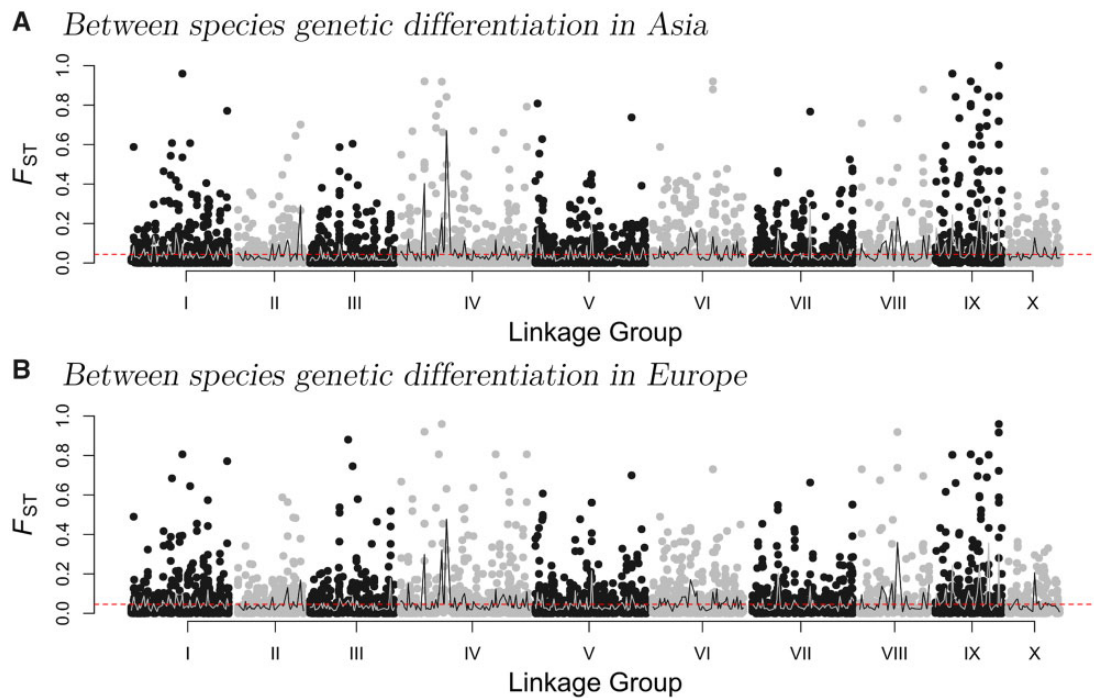
The population tree estimated with *Treemix* closely mirrored the genetic relationships evidenced with the dAPC (fig. 6A). Two migration edges had to be introduced for improving the model fit, afterwards the likelihood did not improve by adding further migration events. Evidence was found for asymmetrical gene flow from *C. angulata* to *C. gigas* in both Asia and Europe. Nevertheless, a stronger intensity of migration was inferred in Europe (i.e., migration weight was 0.146 in Europe against 0.122 in Asia).

The only significantly negative  $f_3$  scores calculated over all possible three-population tests were obtained using the European *C. gigas* population (LAF) as the focal population. Estimates for the LAF population were more negative when considering the reference *C. angulata* population from Europe (SAD) instead of Asia (KEE), thus supporting increased gene flow in Europe (fig. 6B). In contrast, no negative value was obtained for the Asian *C. gigas* population (JAP), which confirms the *Treemix* result that the JAP population is less introgressed by *C. angulata* than the European population of LAF.

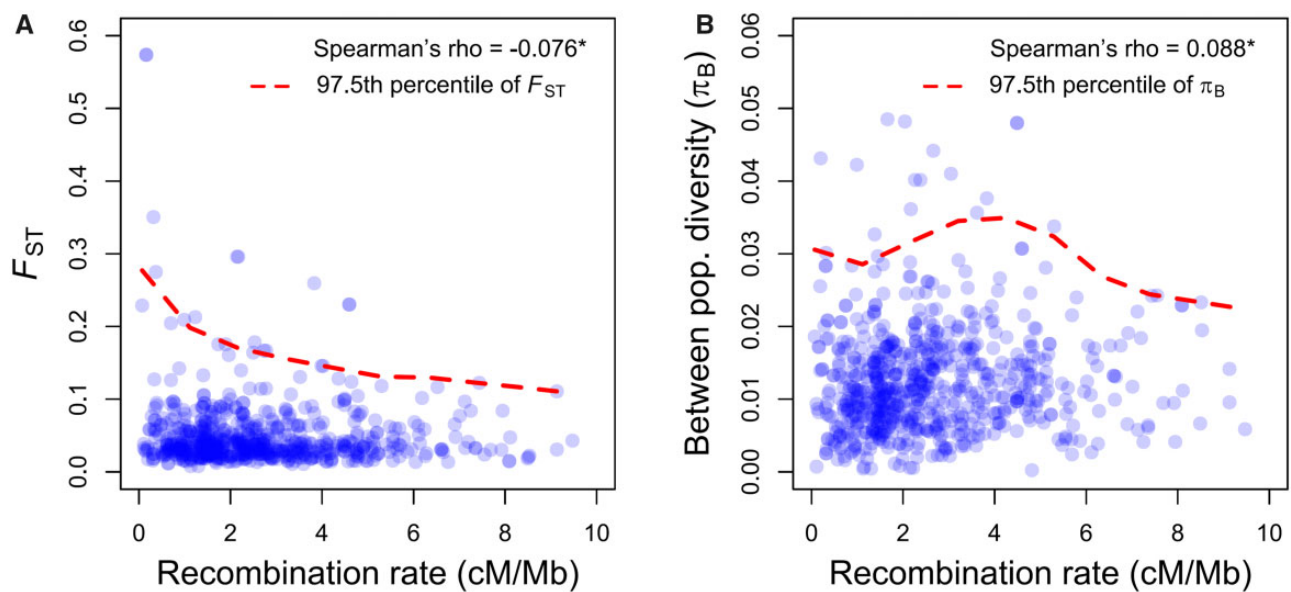
We found significant contrasts between  $f_3$  estimates obtained from high- versus low-recombining regions of the genome. Evidence for genetic admixture in the European *C. gigas* population (LAF) was only detected using SNPs located in the 50% of the genome associated with the highest recombination rates (fig. 6B). These results thus indicate that between-species gene flow is reduced in low- compared with



**Fig. 3.**—(A) dAPC plot of the four populations of Pacific cupped oysters assessed with 1,325 SNPs (one SNP per RAD), with *Crassostrea angulata* from Asia (KEE, light blue) and Europe (SAD, dark blue), and *Crassostrea gigas* from Asia (JAP, light red) and Europe (LAF, dark red). (B) Genetic parallelism in the level of differentiation between species in Asia (x axis) and Europe (y axis) measured at 10,144 SNPs. The black line shows the  $y = x$  equation line and the red dashed line corresponds to the linear regression ( $F_{ST}^{Europe} = 0.83 \times F_{ST}^{Asia} + 0.01$ ,  $R^2 = 0.71$ ,  $P < 10^{-10}$ ).



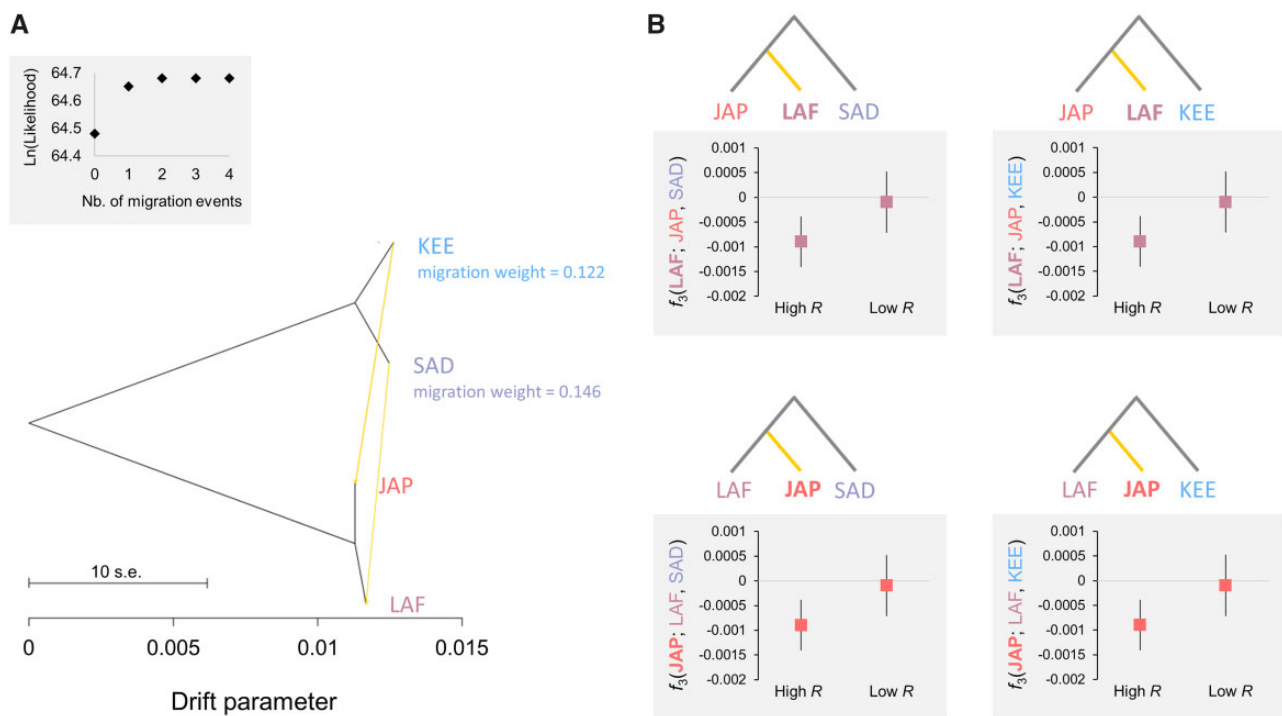
**FIG. 4.**—Genomic landscape of between-species genetic differentiation ( $F_{ST}$ ) measured at 10,144 SNPs (alternatively represented by black and grey points for odd and even chromosome numbers for convenience) along the ten Pacific oyster linkage groups. The red dotted line shows the genome-wide average  $F_{ST}$  and the grey line shows the local average  $F_{ST}$  calculated in 150 kb windows. Genetic differentiation between *Crassostrea angulata* and *Crassostrea gigas* is shown (A) in Asia and (B) in Europe.



**FIG. 5.**—Distribution of between-species genetic differentiation ( $F_{ST}$ ) and diversity ( $\pi_B$ ) as a function of the local recombination rate assessed with *MareyMap*. Each point represents an estimate of  $F_{ST}$  (A) or  $\pi_B$  (B) and recombination rate averaged in a 500 kb window. Outlier recombination rate values exceeding 10 cM/Mb were excluded. The red dashed line represents the nonparametric 97.5th quantile regression fit of  $F_{ST}$  or  $\pi_B$  as a function of the local recombination rate.

high-recombining regions. Moreover, the contrast between  $f_3(\text{LAF}; \text{gigas}, \text{angulata})$  and  $f_3(\text{JAP}; \text{gigas}, \text{angulata})$  was always amplified in high-recombination rate regions. This result

suggests that the different introgression rates experienced by LAF and JAP populations (fig. 6A) mainly concern the regions with the highest recombination rates.



**Fig. 6.**—Analysis of gene flow between oyster populations. (A) *Treemix* analysis showing the population tree with two migration edges from *Crassostrea angulata* to *Crassostrea gigas* populations in Asia and Europe. The upper left panel shows model likelihood as a function of the number of migration edges considered in the model. (B) Results of  $f_3$  calculations obtained using LAF and JAP as focal populations, on two similar size data sets containing independent SNPs from high- (i.e.,  $R > 2.21$  cM/Mb) or low- (i.e.,  $R < 2.21$  cM/Mb) recombining regions.

### Divergence History

The demographic history of divergence between *C. angulata* and *C. gigas* inferred with *δaδi* showed evidence for a SC in both Asia and Europe (table 2). The JAFS presented in figure 7 show a high proportion of shared polymorphisms at high frequency along the diagonal, which is a characteristic footprint of secondary introgression that is not expected under a SI model (Alcala et al. 2016; Roux et al. 2016) and explain the good support for the SC model. Moreover, varying rates of introgression among loci (i.e., the SC2*m* model) was strongly supported for both native and introduced species pairs, since reduced introgression explains the presence of SNPs in the upper right and lower left corners of the spectrum. The SC2*m* model received a good statistical support compared with the six other alternative models (Akaike weights  $> 0.95$ ), and its goodness-of-fit showed almost no trend in the distribution of residuals for both species pairs (fig. 7).

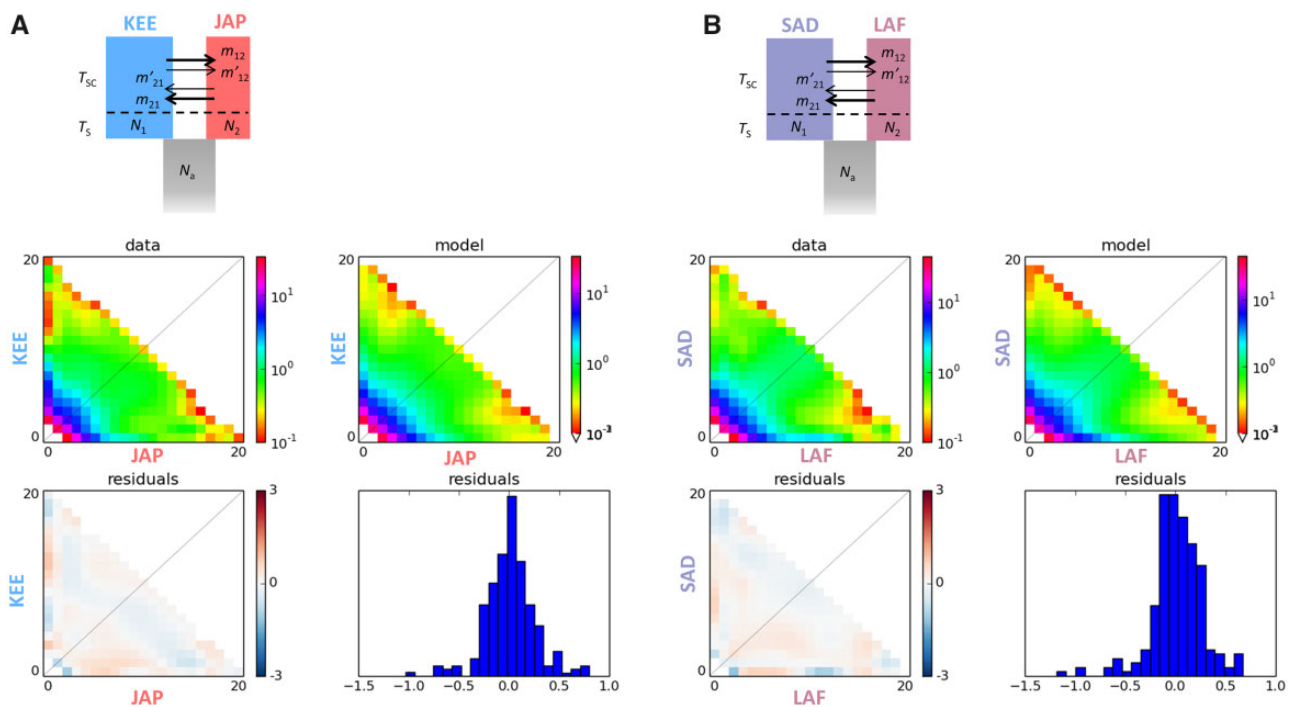
Therefore, the semipermeable species-barrier model, in which introgression occurs at variable rates across the genome since SC, explains well the observed data. In the neutrally introgressing fraction of the genome (i.e., 55–63% of the genome), gene flow was found to occur in both directions but at slightly higher rates from *C. angulata* to *C. gigas* (i.e., the  $N_e m$  product indicates 12.2–16.7 migrants per generation from *C. gigas* to *C. angulata* and 16.9–17.3 migrants per

generation from *C. angulata* to *C. gigas*, table 2). In contrast, there was a 6–11-fold decrease in gene flow ( $m/m'$ , table 2) in the remaining fraction of the genome, due to reduced effective migration rates in both directions. Time parameters indicated that the duration of isolation without gene flow was 11–21 times longer than the duration of secondary gene flow ( $T_S/T_{SC}$ , table 2). These results support that the allopatric divergence period was long enough to enable differentiation prior to SC.

### Discussion

#### An Improved Scaffolding of the Pacific Oyster Genome

Our new high-density linkage map in *C. gigas* adds to a series of first-generation linkage maps (Hubert and Hedgecock 2004; Li and Guo 2004; Hubert et al. 2009; Sauvage et al. 2010; Plough and Hedgecock 2011; Guo et al. 2012; Zhong, Li, Guo, et al. 2014) and more recent second-generation maps (Hedgecock et al. 2015; Wang et al. 2016) that were produced in the Pacific oyster. Compared with the two most recent linkage maps, this map offers a slightly improved resolution, with an average marker-spacing of 0.64 cM compared with 1 cM in Hedgecock et al. (2015) and 0.8 cM in Wang et al. (2016). We paid attention to minimize the influence of the potentially widespread transmission ratio



**FIG. 7.**—Demographic divergence history of *Crassostrea angulata* and *Crassostrea gigas* in Asia (A) and Europe (B). The best selected model in both species pairs was the secondary contact model with two categories of loci experiencing different rates of introgression (SC2m) and occurring in proportions  $P$  (for freely introgressing loci) and  $1 - P$  (for loci with reduced effective migration rates) across the genome. The model parameters are described in the legend to table 2. Each plot shows the observed JAFS (upper left), the JAFS obtained under the maximum-likelihood SC2m model (upper right), the spectrum of residuals with over (red) and under (blue) predicted SNPs (lower left), and the distribution of residuals (lower right).

distortions observed across most of the 10 linkage groups (Launey and Hedgecock 2001; Plough and Hedgecock 2011; Plough 2012; Hedgecock et al. 2015), by discarding distorted markers prior to map reconstruction. Our map, however, shows a good collinearity with the map produced by Hedgecock et al. (2015) that did not exclude distorted markers. This suggests that distortions of segregation ratios do not substantially affect the ordering of markers in *C. gigas* genetic maps, and further support that Hedgecock's consensus linkage map and ours provide a reasonable assessment of marker order and genetic distances across the Pacific oyster genome.

Consistent with what has been previously reported by Hedgecock et al. (2015), we found a high incidence of chimeric scaffolds (38.5% and 44.2%, respectively) in the oyster\_v9 assembly (Zhang et al. 2012; <http://www.oysterdb.com>). Splitting scaffolds to resolve mapping conflicts to different linkage groups decreased the scaffold N50 by 45.6% (401–218 kb), but the resc scaffolding step using BES increased the scaffold N50 again by 47% (320 kb). Because of the limited density of markers in our linkage map, this new assembly version (available at <ftp://ftp.ifremer.fr/ifremer/dataref/bioinfo/>) is not complete and tends to exclude especially the shortest scaffolds present in the oyster\_v9 assembly (Zhang et al. 2012). Therefore, only 50.1% (280.2 Mb) of the oyster\_v9

assembly could be anchored into 10 pseudochromosomes based on the information contained in our new linkage map. This underlines the need for an improved genome assembly in *C. gigas* using new information from higher density linkage maps and long read sequencing data. Such improvements are mandatory for all functional analyses relying on a complete genome description and a thorough annotation of gene sequences, such as genome-wide association studies (GWAS). However, for the purpose of this study, a partial chromosome-level assembly is sufficient to explore broad-scale genomic variation in genetic differentiation between the two studied sister taxa and its possible relationship with recombination rate variation.

### Recombination Rate Variation across the Genome

The total length of our sex-averaged consensus map (965 cM) is intermediate to previous estimates that were reported from the two existing high-density linkage maps in the Pacific oyster (890–1,084 cM, Hedgecock et al. 2015, respectively, Wang et al. 2016). Considering the genome-size estimate of 559 Mb (Zhang et al. 2012), this corresponds to an average recombination rate of 1.73 cM per Mb, a value within the range of average recombination rate values obtained across



**Table 2** Results of Model Fitting for Seven Alternative Models of Divergence between *C. angulata* and *C. gigas* in Asia and Europe

Species Pair	Model	Anneal. Hot	Anneal. Cold	BFGS	AIC	$\Delta_i$	$w_i$	$\theta$	$N_1$	$N_2$	$m_{12}$	$m_{21}$	$m'_{12}$	$m'_{21}$	$T_5$	$T_{AM}$	$T_{SC}$	P	
ASIA	SI	-247.3844	-245.0567	-244.2167	494.4335	61.3173	0.0000	94.0960	0.9363	19.8905	-	-	-	-	0.2422	-	-	-	-
	IM	-232.0492	-231.6256	-229.6122	469.2244	36.1082	0.0000	50.7159	1.1160	2.1658	1.4404	0.6423	-	-	1.7087	-	-	-	-
	AM	-246.4434	-239.7242	-229.6158	471.2316	38.1154	0.0000	53.1106	1.0452	2.1069	1.5443	0.6500	-	-	1.5652	0.0000	-	-	-
	SC	-227.7188	-227.7188	-223.1180	458.2361	25.1198	0.0000	61.5886	1.2126	1.6412	2.1765	1.5932	-	-	0.8178	-	0.1899	-	-
	IM2m	-217.3343	-217.3343	-214.4709	444.9419	11.8257	0.0027	35.8268	2.2285	2.4318	3.7818	4.0944	0.2835	0.2487	3.0381	-	-	-	0.5496
	AM2m	-233.2967	-219.8171	-214.4997	446.9993	13.8831	0.0010	30.2063	2.7356	2.7870	2.4416	3.3503	0.2287	0.2133	3.8450	0.0001	-	-	0.5603
	<b>SC2m</b>	<b>-218.2059</b>	<b>-218.2059</b>	<b>-207.5581</b>	<b>433.1162</b>	<b>0</b>	<b>0.9963</b>	<b>56.5407</b>	<b>1.5600</b>	<b>1.5353</b>	<b>10.6952</b>	<b>11.2826</b>	<b>1.4966</b>	<b>1.5031</b>	<b>1.1291</b>	-	-	<b>0.0536</b>	<b>0.6336</b>
EUROPE	SI	-266.1601	-266.1601	-266.1370	538.2740	66.4534	0.0000	119.5436	0.7832	19.9447	-	-	-	-	0.2127	-	-	-	-
	IM	-251.9601	-251.9601	-250.9902	511.9804	40.1598	0.0000	78.0597	1.4969	1.0110	0.6372	1.8961	-	-	1.1803	-	-	-	-
	AM	-254.7446	-254.3147	-251.0059	514.0118	42.1913	0.0000	72.6025	1.6028	1.0537	0.5769	1.8227	-	-	1.4241	0.0000	-	-	-
	SC	-258.8205	-249.8626	-245.4758	502.9515	31.1310	0.0000	87.7852	1.6884	0.8123	0.7821	3.8227	-	-	0.5498	-	0.1525	-	-
	IM2m	-239.2627	-239.2627	-231.3043	478.6086	6.7880	0.0321	61.9879	1.3816	1.8837	5.3928	3.1806	0.2426	0.4237	1.8432	-	-	-	0.6025
	AM2m	-244.92139	-244.92139	-231.3516	480.7032	8.8826	0.0113	65.9174	1.2964	1.7676	5.9929	3.3784	0.2603	0.4515	1.6631	0.0004	-	-	0.6007
	<b>SC2m</b>	<b>-233.321</b>	<b>-233.321</b>	<b>-226.910</b>	<b>471.821</b>	<b>0</b>	<b>0.9566</b>	<b>85.6599</b>	<b>1.1751</b>	<b>1.3148</b>	<b>10.3971</b>	<b>12.8181</b>	<b>0.9340</b>	<b>2.0132</b>	<b>0.6982</b>	-	-	<b>0.0644</b>	<b>0.5527</b>

NOTE.—In the order of appearance in the table: the model fitted, the maximum likelihood estimate over 20 independent runs after three successive optimization steps: simulated annealing hot, cold and BFGS, the Akaike Information Criterion, the difference in AIC with the best model (SC2m), Akaike weight, and theta parameter for the ancestral population before split. Following are the inferred values for the model parameters (scaled by theta): the effective size of *C. angulata* ( $N_1$ ) and *C. gigas* ( $N_2$ ) populations, the migration rates for the two categories of loci in the genome (category 1:  $m_{12}$  and  $m_{21}$ , category 2:  $m'_{12}$  and  $m'_{21}$ ), the duration of the split ( $T_5$ ), of ancestral migration ( $T_{AM}$ ) and secondary contact ( $T_{SC}$ ) episodes, and the proportion (P) of the genome falling within category 1 (experiencing migration rates  $m_{12}$  and  $m_{21}$ ).

a wide diversity of animal species (Corbett-Detig et al. 2015; Stapley et al. 2017).

The chromosome-averaged recombination rate shows a 2.5-fold variation among linkage groups and is negatively correlated to chromosome length. Such a pattern has already been observed in several species including yeast (Kaback et al. 1992), human (Lander et al. 2001), stickleback (Roesti et al. 2013), and daphnia (Dukić et al. 2016), and has been commonly attributed to positive crossover interference. This phenomenon happens during meiosis, when the formation of an initial recombination event reduces the probability that an additional recombination event occurs nearby on the chromosome. If the mechanism of interference is still elusive, it is often assumed that a fixed number of crossovers happens per chromosome (between 1 and 2), making large chromosomes experiencing less recombination events per nucleotide position than smaller ones. Our results thus provide new empirical support for the existence of crossover interference in *C. gigas*.

We also revealed large-scale (i.e., megabase-scale) variation in local recombination rate within chromosomes. The general pattern we observed was a reduction of crossover rate in chromosome centers relative to peripheries. Similar patterns are widespread in animals, plants and fungi (reviewed in Berner and Roesti 2017; Haenel et al. 2018), and can be attributed to different but not mutually exclusive mechanisms. Firstly, a reduction in recombination near the centromere of metacentric chromosomes (Nachman 2002) can be hypothesized, possibly due to the presence of highly condensed heterochromatin. Secondly, crossover interference, which tends to produce evenly and widely spaced crossovers when multiple crossovers occur on the same chromosome, may increase the chance that multiple recombination events affect chromosomal extremities. Thirdly, heterochiasmy (i.e., different recombination rates and landscapes between sexes) could also explain these patterns (Stapley et al. 2017), especially if one sex tends to recombine preferentially in chromosomal extremities. Finally, a telomere-guided initiation of homology search before recombination could also explain an increased frequency of crossovers near chromosome peripheries. In the Pacific oyster, the comparison of male and female maps did not support the hypothesis of heterochiasmy (Hedgecock et al. 2015), so the observed recombination rate variation within chromosomes could be rather explained by centromere location, crossover interference and/or telomere-guided initiation of recombination.

The recombination environment is known to affect the impact of natural selection on linked neutral diversity across a wide range of species, in which low-recombining regions generally display reduced levels of nucleotide diversity (Corbett-Detig et al. 2015). As expected under the joint effect of background selection (Charlesworth et al. 1993) and hitchhiking (Maynard Smith and Haigh 1974) on linked neutral diversity, we found a positive relationship between the local

recombination rate and nucleotide diversity across the Pacific oyster genome. However, the magnitude of this effect was rather small compared with what could be expected from comparative estimates in other invertebrate species supposed to have large population sizes. Corbett-Detig et al. (2015) argued that the effect of natural selection on the reduction of linked neutral variation was stronger in species with large census sizes, but a reanalysis by Coop (2016) showed that this conclusion was not supported by the data. Oysters are usually perceived as species with large effective population sizes and this is corroborated by a medium to high nucleotide diversity (Sauvage et al. 2007; Zhang et al. 2012). However, several empirical evidences from molecular evolution and population genetics studies have also shown that oysters have among the highest segregating loads of deleterious mutations observed in marine invertebrates (Sauvage et al. 2007; Plough 2016). This may be due to a high variance in reproductive success (sweepstake effect) and population size fluctuations (Hedgcock 1994; Boudry et al. 2002; Hedgcock and Pudovkin 2011; Harrang et al. 2013; Plough 2016). We are lacking theoretical predictions on the effect of linked selection in a species with skewed offspring distribution, but the effect is likely more genome-wide during favorable sweepstake events than it is in the standard Wright–Fisher model.

### Genome-Wide Diversity Patterns within and between Pacific Oyster Species

The core objectives of this study were to 1) determine the extent and variation in the level of differentiation within and between *C. angulata* and *C. gigas* across the genome, 2) test the existence of differences in the genomic landscape of differentiation between native and introduced species pairs, 3) infer the history of gene flow during divergence, and finally 4) relate interspecies divergence to underlying evolutionary processes including demography, selection and genomic constraints.

Our results reveal that the European introduction of *C. angulata* in the 16th century, and more recently of *C. gigas* in the 1970s, did not lead to a reduction in genetic diversity for both species compared with source populations in their native range. This observation is rather common in the literature of marine invasions, and has been attributed to the combined effects of multiple introductions and high propagule pressure (Viard et al. 2016). Another possibility is that the initiation of the demographic wave of invasion can simply not occur without a sufficiently large number of individuals in species with a strong Allee effect like sessile broadcast spawners. In other words, a successful marine introduction cannot happen without a high number of founders. We found a relatively low background level of differentiation between *C. gigas* and *C. angulata* (i.e., genome-wide averaged  $F_{ST} = 0.045$ ), which is consistent with previous estimates based on microsatellite loci (Huvet et al. 2000, 2004).

However, differentiation was highly heterogeneous across the genome in both Asian and European populations, with peaks of elevated  $F_{ST}$  values being found in most chromosomes, sometimes even reaching differential allelic fixation in Asia. These “genomic islands of differentiation” are commonly observed in genome scans for divergence between closely related species pairs (Wolf and Ellegren 2017), and multiple mechanisms have been proposed to explain their formation (Cruickshank and Hahn 2014; Burri et al. 2015; Yeaman et al. 2016; Ravinet et al. 2017). They broadly fall in two categories: 1) mechanisms of differential gene flow involving the existence of reproductive isolation and/or local adaptation loci acting as genetic barriers to interspecific gene flow (Barton and Bengtsson 1986; Feder and Nosil 2010), and 2) heterogeneous divergence mechanisms due to variation in the rate of lineage sorting across the genome through linked selection (Cruickshank and Hahn 2014; Burri et al. 2015; Burri 2017a). In accordance with the hypothesis of linked selection, genetic differentiation ( $F_{ST}$ ) and absolute divergence ( $\pi_B$ ) were respectively negatively and positively correlated with the local recombination rate. Concomitantly, however, the highest values of both relative and absolute divergence were preferentially observed in low-recombining regions, which is not predicted under linked selection alone (Burri 2017a,b). Using simulations, Duranton et al. (2018) recently showed that these observations are consistent with the existence of genetic barriers to interspecific gene flow, in addition to the action of linked selection. Therefore, genomic islands between Pacific oyster species may also involve both mechanisms.

In order to test whether specific adaptations to different habitats participate in genetic barriers, we hypothesized that the different ecological conditions encountered by oysters in Europe would either result in relaxed divergent selection pressures on genes involved in local adaptation in Asia, or in new selective constraints targeting different subsets of genes. In both cases, we expect that the new environmental conditions in Europe would promote a remodeling of the genomic landscape of species divergence and hence reduce the extent of parallelism in genetic differentiation between species-pairs from native and introduced populations. Contrary to that prediction, we found a remarkably high degree of divergence parallelism indicating that the genetic architecture of reproductive isolation between *C. angulata* and *C. gigas* has not been reinforced, nor significantly weakened following the co-introduction of the two species in Europe. Nevertheless, a slightly lower interspecific differentiation was found in Europe. This reduction was not driven by the most differentiated loci (figs. 3B and 6B), and therefore rather indicated increased gene flow in the low-differentiated fraction of the genome than relaxed divergent selection on barrier loci in the novel habitat. Moreover, the *Treemix* and three-population  $f_3$  tests suggest that more pronounced gene flow is ongoing from *C. angulata* to *C. gigas* than in the opposite direction, especially in Europe (fig. 6). This asymmetrical

introgression pattern seems consistent with the more recent introduction of *C. gigas*, which progressive spread over Europe has possibly favored the incorporation of genetic material from *C. angulata* through repeated backcrossing. Overall, these results suggest that the particular demographic conditions imposed by the co-introduction of the two species in southern Europe have facilitated opportunities for genetic interactions, without modifying the nature of the species boundary.

The finding of unaltered genomic differentiation patterns in the introduced species-pair despite increased gene flow suggests either that the contact is too recent to observe large effects, or that historical and architectural genomic features have played a major role in shaping the genomic landscape of species divergence between Pacific cupped oysters. The observed pattern of heterogeneous genome divergence is predicted by the semipermeable species boundary model, in which neutral introgression is only permitted in genomic regions unlinked to reproductive isolation barriers (Harrison and Larson 2014, 2016). Under this model, elevated  $F_{ST}$  values above the background level indicate the location of genomic regions that are resistant to introgression due to the presence of reproductive isolation loci (i.e., speciation genes). The existence and the possible origin of a semipermeable barrier to gene flow between *C. gigas* and *C. angulata* was further addressed through the comparison of various theoretical models offering simplified representations of contrasted evolutionary scenarios. Our results indicated that contemporary differentiation patterns likely result from a long period of allopatric divergence followed by a recent SC with different rates of introgression among loci. This model consistently outperformed alternative scenarios in both native and introduced ranges, predicting observed differentiation patterns with very small residual errors (fig. 7). Inferred model parameters capturing semipermeability indicated that the fraction of the genome experiencing reduced introgression rates amounts to 37–45%, consistent with the view that the species barrier is still largely permeable to gene flow. Similar estimates have been obtained in other marine species-pairs analyzed with the same approach (Tine et al. 2014; Le Moan et al. 2016; Rougemont et al. 2017). In all these cases, postglacial SCs between allopatrically diverged lineages have been inferred. The northwestern Pacific region is known to harbor multiple cases of marine species subdivided into divergent lineages that are supposed to have originated in separate glacial refugia, most likely corresponding to the current seas of Japan and China (reviewed in Ni et al. 2014). Our study provides the first genome-wide view of the evolutionary consequences of past sea-level regression in this region, and adds to previous studies suggesting their important role in initiating speciation in northwestern Pacific taxa (e.g., Shen et al. 2011).

In addition to providing evidence for partial reproductive isolation between *C. gigas* and *C. angulata*, our results also bring new empirical support for the role of recombination rate

variation in shaping the genomic landscape of species divergence. A recent study in the European sea bass has shown that low-recombining regions not only experience accelerated rates of lineage sorting during allopatric phases, but are also preferentially involved in the barrier to gene flow during SCs (Duranton et al. 2018). Here, the three-population  $f_3$  tests revealed that the increase in gene flow favored in the European context has mostly affected highly recombining regions, which likely explains the negative relationship observed between differentiation and recombination. This finding is consistent with the idea that foreign alleles are more efficiently removed by selection following introgression when recombination is locally reduced (Schumer et al. 2018). This interpretation also suggests that the sites under selection are widespread across the genome, although not individually under strong selection (Aeschbacher et al. 2017). This type of distribution of individual locus effect sizes has recently been evidenced from hybridization studies in a number of animal and plant species (Simon et al. 2018).

## Conclusion

Our results shed new light on the existence of reproductive isolation barriers between the two cupped oysters *C. angulata* and *C. gigas*. By providing empirical evidence for heterogeneous divergence patterns attributable to reduced introgression in low-recombining regions since SC, we show that these semi-isolated species are still evolving in the so-called “speciation grey zone” (Roux et al. 2016). Moreover, the finding of strong divergence parallelism between species-pairs from native and introduced areas suggests that the genomic architecture of reproductive isolation is not primarily determined by ecological divergence driven by local adaptations in the native range. Our study thus implies the existence of intrinsic genetic differences between the two species, which will be the focus of future investigations based on experimental crosses.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This work was supported by Interreg SUDOE “Aquagenet” (SOE2/P1/E287) and ANR “GAMETOGENES” (ANR-08-GENM041) projects, as well as from GIS-IBISA and Ifremer. We would like to thank Brieg Pontreau for his help in microsatellite genotyping, the hatchery team of Ifremer La Tremblade for its help in the production of the biological material, and H el ene Holota from Inserm UMR 1090 TAGC for her kind assistance during the sonication of RAD libraries. We are also grateful to Patrick Gaffney for providing BAC clones

and Pascal Favrel for managing BAC-end sequencing with the Genoscope. Golden Gate SNP genotyping was performed at the Genome Transcriptome Facility of Bordeaux (grants from Conseil Régional d'Aquitaine n°20030304002FA and 20040305003FA, from the European Union FEDER n°2003227 and from the Investissement d'Avenir ANR-10-EQPX-16-01). Sequencing was performed by the Genoscope, Integragen, and the CNRS plateforme Maladies Métaboliques et Infectieuses.

## Literature Cited

- Abbott R, et al. 2013. Hybridization and speciation. *J Evol Biol.* 26(2):229–246.
- Aeschbacher S, Selby JP, Willis JH, Coop G. 2017. Population-genomic inference of the strength and timing of selection against gene flow. *Proc Natl Acad Sci U S A.* 114(27):7061–7066.
- Alcala N, Jensen JD, Telenti A, Vuilleumier S. 2016. The genomic signature of population reconnection following isolation: from theory to HIV. *G3: Genes Genomes Genet.* 6: 107–120.
- Baird NA, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10):e3376.
- Barton NH, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. *Heredity* 57(3):357–376.
- Batista FM, Fonseca VG, Ruano F, Boudry P. 2017. Asynchrony in settlement time between the closely related oysters *Crassostrea angulata* and *C. gigas* in Ria Formosa lagoon (Portugal). *Mar Biol.* 164(5):110.
- Berner D, Roesti M. 2017. Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. *Mol Ecol.* 26(22):6351–6369.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol.* 20(10):2044–2072.
- Bierne N, Gagnaire P-A, David P. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Curr Zool.* 59(1):72–86.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.
- Boudry P, Heurtebise S, Collet B, Cornette F, Gérard A. 1998. Differentiation between populations of the Portuguese oyster, *Crassostrea angulata* (Lamarck) and the Pacific oyster, *Crassostrea gigas* (Thunberg), revealed by mtDNA RFLP analysis. *J Exp Mar Biol Ecol.* 226(2):279–291.
- Boudry P, Collet B, Cornette F, Hervouet V, Bonhomme F. 2002. High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* 204(3–4):283–296.
- Boudry P, Dégremont L, Haffray P. 2008. The genetic basis of summer mortality in Pacific oyster spat and potential for improving survival by selective breeding in France. Summer mortality of Pacific oyster-The Mores Project. Versailles, France: Quae Editions: 153–196.
- Burri R, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* 25:1656–1665.
- Burri R. 2017a. Interpreting differentiation landscapes in the light of long-term linked selection. *Evol Lett.* 1(3):118–131.
- Burri R. 2017b. Dissecting differentiation landscapes: a linked selection's perspective. *J Evol Biol.* 30(8):1501–1505.
- 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 Genet.* 1: 171–182.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 22(11):3124–3140.
- Charlesworth B, Morgan M, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289–1303.
- Coop G. 2016. Does linked selection explain the narrow range of genetic diversity across species? bioRxiv 042598 8 pages.
- Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* 13(4):e1002112.
- Cruikshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* 23:3133–3157.
- Danecek P, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Darling JA, Bagley MJ, Roman J, Tepolt CK, Geller JB. 2008. Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Mol Ecol.* 17(23):4992–5007.
- Dégremont L, Bédier E, Boudry P. 2010. Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). II. Response to selection for survival and its influence on growth and yield. *Aquaculture* 299(1–4):21–29.
- Dégremont L. 2011. Evidence of herpesvirus (OsHV-1) resistance in juvenile *Crassostrea gigas* selected for high resistance to the summer mortality phenomenon. *Aquaculture* 317(1–4):94–98.
- Dukić M, Berner D, Roesti M, Haag CR, Ebert D. 2016. A high-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. *BMC Genet.* 17(1):137.
- Duranton M, et al. 2018. The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nat Commun.* 9(1):2518.
- Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64(6):1729–1747.
- Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* 28(7):342–350.
- Fraïsse C, Belkhir K, Welch JJ, Bierne N. 2016. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol Ecol.* 25(1):269–286.
- Gaffney P. 2008. A BAC-based physical map for the Pacific oyster genome. *J Shellfish Res.* 27: 1009–1009.
- Gagnaire P-A, et al. 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol Appl.* 8: 769–786.
- Grade A, et al. 2016. New insights about the introduction of the Portuguese oyster, *Crassostrea angulata*, into the North East Atlantic from Asia based on a highly polymorphic mitochondrial region. *Aquat Liv Resour.* 29(4):404.
- Grizel H, Héral M. 1991. Introduction into France of the Japanese oyster (*Crassostrea gigas*). *ICES J Mar Sci.* 47(3):399–403.
- Guo X, Li Q, Wang QZ, Kong LF. 2012. Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Mar Biotechnol.* 14(2):218–226.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5(10):e1000695.
- Haenel Q, Laurentino TG, Roesti M, Berner D. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol Ecol.* 27:2477–2497.
- Harrang E, Lapègue S, Morga B, Bierne N. 2013. A high load of non-neutral amino-acid polymorphisms explains high protein diversity despite moderate effective population size in a marine bivalve with sweepstakes reproduction. *G3: Genes Genomes Genet.* 3:333–341.



- Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. *J Heredity*. 105(S1):795–809.
- Harrison RG, Larson EL. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol Ecol*. 25(11):2454–2466.
- Hedgecock D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? In: Beaumont A, editor. *Genetics and evolution of aquatic organisms*. London (United Kingdom): Chapman & Hall. p. 122–134.
- Hedgecock D, Pudovkin AI. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull Mar Sci*. 87(4):971–1002.
- Hedgecock D, Shin G, Gracey AY, Den Berg DV, Samanta MP. 2015. Second-generation linkage maps for the Pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. *G3: Genes Genomes Genet*. 5: 2007–2019.
- Hellberg ME. 2009. Gene flow and isolation among populations of marine animals. *Annu Rev Ecol Syst*. 40(1):291–310.
- Hsiao S-T, et al. 2016. DNA barcoding reveals that the common cupped oyster in Taiwan is the Portuguese oyster *Crassostrea angulata* (Ostreoida; Ostreidae), not *C. gigas*. *Sci Rep*. 6(1):34057.
- Hubert S, Hedgecock D. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* 168(1):351–362.
- Hubert S, Cognard E, Hedgecock D. 2009. Centromere mapping in triploid families of the Pacific oyster *Crassostrea gigas* (Thunberg). *Aquaculture* 288(3–4):172–183.
- Huvet A, Lapegue S, Magoulas A, Boudry P. 2000. Mitochondrial and nuclear DNA phylogeography of *Crassostrea angulata*, the Portuguese oyster endangered in Europe. *Conservation Genet*. 1:251–262.
- Huvet A, Balabaud K, Bierne N, Boudry P. 2001. Microsatellite analysis of 6-hour-old embryos reveals no preferential intraspecific fertilization between cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Mar Biotechnol*. 3(5):448–453.
- Huvet A, et al. 2002. Is fertility of hybrids enough to conclude that the two oysters *Crassostrea gigas* and *Crassostrea angulata* are the same species? *Aquat Liv Resour*. 15(1):45–52.
- 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. *Mar Ecol Prog Ser*. 272:141–152.
- Jeffery N, et al. 2017. Genomic evidence of hybridization between two independent invasions of European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Heredity* 119(3):154.
- Jeffery NW, et al. 2018. Genomewide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America. *Evol Appl*. 11: 869–882.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403–1405.
- Kaback DB, Guacci V, Barber D, Mahon JW. 1992. Chromosome size-dependent control of meiotic recombination. *Science* 256(5054):228–232.
- Kelley JL, Brown AP, Therkildsen NO, Foote AD. 2016. The life aquatic: advances in marine vertebrate genomics. *Nat Rev Genet*. 17(9):523.
- Knowlton N. 1993. Sibling species in the sea. *Annu Rev Ecol Syst*. 24(1):189–216.
- Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.
- Lapègue S, Batista F, Heurtebise S, Yu Z, Boudry P. 2004. Evidence for the presence of the Portuguese oyster, *Crassostrea angulata*, in northern China. *J Shellfish Res*. 23:759–763.
- Lapègue S, et al. 2014. Development of SNP-genotyping arrays in two shellfish species. *Mol Ecol Resour*. 14(4):820–830.
- Launey S, Hedgecock D. 2001. High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159(1):255–265.
- Le Moan A, Gagnaire PA, Bonhomme F. 2016. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol Ecol*. 25(13):3187–3202.
- Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D. 2003. Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Mol Ecol Notes*. 3(2):228–232.
- Li L, Guo X. 2004. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Mar Biotechnol*. 6(1):26–36.
- Li R, Li Q, Cornette F, Dégremont L, Lapègue S. 2010. Development of four EST-SSR multiplex PCRs in the Pacific oyster (*Crassostrea gigas*) and their validation in parentage assignment. *Aquaculture* 310(1–2):234–239.
- López-Flores I, et al. 2004. The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene* 339:181–188.
- Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res*. 23(01):23–35.
- Moretto M, Cestaro A, Troglio M, Costa F, Velasco R. (2010) Harry Plotter: a user friendly program to visualize genome and genetic map features. In: ECCB10, 9th European Conference Computational Biology. Ghent, Belgium.
- Nachman MW. 2002. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev*. 12(6):657–663.
- Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc B: Biol Sci*. 367(1587):409–421.
- Ni G, Li QI, Kong L, Yu H. 2014. Comparative phylogeography in marginal seas of the northwestern Pacific. *Mol Ecol*. 23(3):534–548.
- Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103(6):439–444.
- Palumbi SR. 1994. Genetic divergence, reproductive isolation, and marine speciation. *Annu Rev Ecol Syst*. 25(1):547–572.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 8(11):e1002967.
- Plough LV, Hedgecock D. 2011. Quantitative trait locus analysis of stage-specific inbreeding depression in the Pacific oyster *Crassostrea gigas*. *Genetics* 189(4):1473–1486.
- Plough LV. 2012. Environmental stress increases selection against and dominance of deleterious mutations in inbred families of the Pacific oyster *Crassostrea gigas*. *Mol Ecol*. 21(16):3974–3987.
- Plough LV. 2016. Genetic load in marine animals: a review. *Curr Zool*. 62(6):567–579.
- Ravinet M, et al. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol*. 30(8):1450–1477.
- Ravinet M, et al. 2015. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol Ecol*. 27:285–307.
- Reece KS, Cordes JF, Stubbs JB, Hudson KL, Francis EA. 2008. Molecular phylogenies help resolve taxonomic confusion with Asian *Crassostrea* oyster species. *Mar Biol*. 153(4):709–721.
- Reich D, et al. 2010. Unusual conservation of mitochondrial gene order in *Crassostrea* oysters: evidence for recent speciation in Asia. *BMC Evol Biol*. 10(1):394.

- Ren J, et al. 2010. Unusual conservation of mitochondrial gene order in *Crassostrea* oysters: evidence for recent speciation in Asia. *Evolutionary Biology*. 10:394, 14.
- Rezvoy C, Charif D, Guéguen L, Marais GA. 2007. MareyMap: an R-based tool with graphical interface for estimating recombination rates. *Bioinformatics* 23(16):2188–2189.
- Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome—patterns and consequences. *Mol Ecol*. 22(11):3014–3027.
- Rose NH, Bay RA, Morikawa MK, Palumbi SR. 2018. Polygenic evolution drives species divergence and climate adaptation in corals. *Evolution* 72(1):82–94.
- Rougemont Q, et al. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol*. 26(1):142–162.
- Roux C, et al. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol*. 14(12): e2000234.
- Saarman NP, Pogson GH. 2015. Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Mol Ecol*. 24(18):4723–4738.
- Sauvage C, Bierne N, Lapegue S, Boudry P. 2007. Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* 406(1–2):13–22.
- Sauvage C, et al. 2010. QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). *Anim Genet*. 41(4):390–399.
- Schumer M, et al. 2018. Natural selection interacts with the local recombination rate to shape the evolution of hybrid genomes. *Science* 10.1126/science.aar3684, 10.
- Sekino M, Yamashita H. 2013. Mitochondrial DNA barcoding for Okinawan oysters: a cryptic population of the Portuguese oyster *Crassostrea angulata* in Japanese waters. *Fish Sci*. 79(1):61–76.
- Shen K-N, Jamandre BW, Hsu C-C, Tzeng W-N, Durand J-D. 2011. Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globally-distributed flathead mullet *Mugil cephalus*. *BMC Evol Biol*. 11(1):83.
- Simon A, Bierne N, Welch JJ. 2018. Coadapted genomes and selection on hybrids: fisher's geometric model explains a variety of empirical patterns. *Evolution Letters* doi:10.1002/evl3.66, 27.
- Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc B*. 372(1736):20160455.
- Takeo I, Sakai S. 1961. Study of breeding of Japanese oyster, *Crassostrea gigas*. *Tohoku J Agric Res* 12: 125–171.
- Taris N, Baron S, Sharbel T, Sauvage C, Boudry P. 2005. A combined microsatellite multiplexing and boiling DNA extraction method for high-throughput parentage analyses in the Pacific oyster (*Crassostrea gigas*). *Aquacult Res*. 36(5):516–518.
- Tine M, et al. 2014. The European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun*. 5:5770.
- van Ooijen JW. 2006. JoinMap® 4, software for the calculation of genetic linkage maps in experimental populations. Wageningen (Netherlands): Kyazma B.V.
- Viard F, David P, Darling JA. 2016. Marine invasions enter the genomic era: three lessons from the past, and the way forward. *Curr Zool*. 62(6):629–642.
- Wang H, Qian L, Liu X, Zhang G, Guo X. 2010. Classification of a common cupped oyster from southern China. *J Shellfish Res*. 29(4):857–866.
- Wang J, Li L, Zhang G. 2016. A high-density SNP genetic linkage map and QTL analysis of growth-related traits in a hybrid family of oysters (*Crassostrea gigas* × *Crassostrea angulata*) using genotyping-by-sequencing. *G3: Genes|Genomes|Genet*. 6:1417.
- Wang J, Xu F, Li L, Zhang G. 2014. A new identification method for five species of oysters in genus *Crassostrea* from China based on high-resolution melting analysis. *Chin J Oceanol Limnol*. 32(2): 419–425.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.
- Westram A, et al. 2014. Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol Ecol*. 23(18):4603–4616.
- Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet*. 18(2):87–100.
- Yamitch J, Voigt ML, Li G, Hedgecock D. 2005. Eight microsatellite loci for the Pacific oyster *Crassostrea gigas*. *Anim Genet*. 36: 524–526
- Yeaman S, Aeschbacher S, Bürger R. 2016. The evolution of genomic islands by increased establishment probability of linked alleles. *Mol Ecol*. 25(11):2542–2558.
- Zhang G, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490(7418):49–54.
- Zhong X, Li Q, Guo X, et al. 2014. QTL mapping for glycogen content and shell pigmentation in the Pacific oyster *Crassostrea gigas* using microsatellites and SNPs. *Aquacult Int*. 22(6):1877–1889.
- Zhong X, Li Q, Yu H, et al. 2014. SNP mining in *Crassostrea gigas* EST data: transferability to four other *Crassostrea* species, phylogenetic inferences and outlier SNPs under selection. *PLoS One* 9(9): e108256.

Associate editor: Liliana Milani