Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*

Pierre-Alexandre Gagnaire[1], Jean-Baptiste Lamy[2], Florence Cornette[2], Serge Heurtebise[2], Lionel Dégremont[2], Emilie Flahauw[2], Pierre Boudry[3], Nicolas Bierne[1], and Sylvie Lapègue[2,*]

[1] ISEM - CNRS, UMR5554, Institut des Sciences de l'Evolution, 34095 Montpellier, France

[2] Ifremer, SG2M-LGPMM, Laboratoire de Génétique et Pathologie des Mollusques Marins, 17390 La Tremblade, France

[3] Ifremer, UMR LEMAR, Laboratoire des Sciences de l'Environnement Marin, 29200 Plouzané, France

[*] Corresponding author (Sylvie.Lapegue@ifremer.fr)

Data deposition

Raw sequences data as well as genetic maps, genome-assembly, and SNP datasets have been deposed on 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 *C. 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 morphologic similarity. In particular, whether ecological and/or intrinsic isolating mechanisms participate to species divergence remains unknown. The recent co-introduction of both taxa into Europe offers a unique opportunity to test how genetic differentiation maintains under new environmental and demographic conditions. We generated a pseudo-chromosome 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, Bierne et al., 2011, Palumbi, 1994, Kelley et al., 2016, Hellberg, 2009, Gagnaire et al., 2015). The ability to score divergence at the genomic level in non-model organisms has recently revealed an increasing number of cases of cryptic species subdivision in broadcast-spawning marine invertebrates (e.g. Ravinet et al., 2015, Westram et al., 2014, Fraïsse et al., 2016, Rose et al., 2018). These recent findings confirm the view that semi-permeable barriers to gene flow between closely related taxa (Harrison & Larson, 2014, Harrison & Larson, 2016) are relatively frequent in the marine realm, and may explain short-distance genetic differentiation patterns 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 non-independently 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. The speciation genomics approach 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 (Ravinet et al., 2017, Wolf & Ellegren, 2017, Yeaman et al., 2016), some of which are not directly linked to speciation (Noor & Bennett, 2009, Nachman & Payseur, 2012, Cruickshank & Hahn, 2014, Burri, 2017). 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, 2017). Because genetic divergence does not easily maintain 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 *C. 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 of populations of the same species also remains unclear. The two taxa can be cross-fertilized in the laboratory to form viable and fertile offspring (Huvet et al., 2001, Huvet et al., 2002, Takeo & Sakai, 1961), and some authors have proposed that they should be considered as a single species (Reece et al., 2008, López-Flores et al., 2004). 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, Lapegue 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. *C. angulata*, also called the Portuguese oyster, is presumed to have been non-voluntarily 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 (Zhong et al., 2014b, Sekino & Yamashita, 2013, Wang et al., 2010, 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 & Héral, 1991). Since then, the two species are in contact in southern Europe and therefore have the potential to exchange genes in a new

98 environment (Batista, et al. 2017; Huvet, et al. 2004). Gene flow in the context of marine invasions has
99 mainly been studied between native and non-indigenous lineages (Saarman and Pogson 2015; Viard,
100 et al. 2016) but rarely between two co-introduced genetic backgrounds in a new place. A notorious
101 exception is the European green crab (*Carcinus maenas*) in the Northwest Atlantic (Darling, et al.
102 2008). However, although the genome-wide genetic differentiation has been studied in the introduced
103 range (Jeffery, et al. 2017; Jeffery, et al. 2018), it has not been compared with the differentiation
104 observed in the native range to date.

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

115

116 **Material and Methods**

117 ***Biological material for the mapping populations***

118 Two F2 families (second generation of biparental crosses) were used for genetic map reconstruction in
119 *C. gigas*. These families were obtained by experimental breeding as part of the MOREST project
120 (Boudry, et al. 2008; Dégremont, et al. 2010) by crossing families selected for resistance (*R*) or
121 susceptibility (*S*) to summer mortality, which were subsequently found to have respectively higher and
122 lower resistance to the herpesvirus OsHV-1 (Dégremont 2011). The F2-19 family was generated
123 through biparental crossing between an *S* female and an *R* male to produce a F1 family from which a
124 single sib-pair was randomly sampled to produce the F2 progeny. The F2-21 family was obtained
125 under the same mating scheme but starting with an *R* female and a *S* male. A total of 293 and 282 F2

126 progenies were used for map reconstruction in family F2-19 and F2-21, respectively. For each
127 individual, whole genomic DNA was isolated from muscle tissue using the QIAamp DNA minikit
128 (Qiagen). DNA was checked for quality by electrophoresis on agarose gel and then quantified using
129 the Quant-iT PicoGreen dsDNA assay kit (Life Technologies).
130
131 ***Genotyping panel for low-density map construction***
132 We genotyped F1 parents and their F2 progenies in both families (F2-19: 2 F1 and 293 F2; F2-21: 2
133 F1 and 282 F2) using a panel of 384 SNPs previously developed in *C. gigas* (Lapègue et al., 2014).
134 SNP genotyping was performed using the Golden Gate assay and analyzed with the Genome Studio
135 software (Illumina Inc). In addition, 42 microsatellite markers were also genotyped according to
136 published protocols (Li et al., 2003, Taris et al., 2005, Yamtich et al., 2005, Li et al., 2010, Sauvage et
137 al., 2010) in order to include markers from previous generation linkage maps in *C. gigas* (Hubert &
138 Hedgecock, 2004, Sauvage et al., 2010).
139
140 ***RAD genotyping for high-density map construction***
141 We selected 106 progenies from each family as well as their four F1 parents for RAD library
142 construction following the original protocol (Baird et al., 2008). Briefly, 1 µg of genomic DNA from
143 each individual was digested with the restriction enzyme *Sbf*I-HF (New England Biolabs), and then
144 ligated to a P1 adapter labeled with a unique barcode. We used 16 barcodes of 5-bp and 16 barcodes of
145 6-bp long in our P1 adapters to build 32-plex libraries. Seven pools of 32 individuals were made by
146 mixing individual DNA in equimolar proportions. Each pool was then sheared to a 350 pb average
147 size using a Covaris S220 sonicator (KBiosciences), and size-selected on agarose gel to keep DNA
148 fragments within the size range 300-700 pb. Each library was then submitted to end-repair, A-tailing
149 and ligation to P2 adapter before PCR amplification for 18 cycles. Amplification products from six
150 PCR replicates were pooled for each library, gel-purified after size selection and quantified on a 2100
151 Bioanalyzer using the High Sensitivity DNA kit (Agilent). Each library was sequenced on a separate
152 lane of an Illumina HiSeq 2000 instrument by IntegraGen Inc. (France, Evry), using 100-bp single
153 reads.

We used the program *Stacks* (Catchen et al., 2013, Catchen et al., 2011) 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 Figure S1) 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×AA, AA×AB, AB×AB, AA×BB, AB×AC and AB×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 approach for correcting missing and miscalled haplotypes in parental samples was coded into R.

Genetic map construction

A new Pacific oyster linkage map was constructed in four successive steps: (*i*) First, a low-density map was built for each family using the SNP and microsatellite markers dataset. These two maps were used to provide accurate ordering of markers, since both mapping populations comprised approximately 300 individuals with very few missing genotypes (0.5%). (*ii*) 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. (*iii*) 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. (*iv*) 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 datasets and a LOD threshold of 10 for the RAD marker datasets. 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

208 significant matches. Assembly errors were identified by scaffolds anchored to more than a single
209 linkage group. Chimeric scaffolds were subsequently splitted at all stretches of Ns connecting adjacent
210 contigs.

211 In order to improve the scaffolding of the Pacific oyster genome, we sequenced 73,728 BAC clones
212 of 150 kb average insert size (Gaffney 2008) from both ends using Sanger sequencing at the
213 Genoscope facility (Evry, France). This resulted in 60,912 cleaned full-length (i.e. >1300 bp) BAC-
214 End sequences (BES) pairs. Each sequence from each pair was trimmed to only conserve 1000 bp
215 between positions 10 and 1010, and clipped into 19 evenly-spaced 100-mers overlapping by 50 bp.
216 This clipping procedure aimed at constructing a high quality short-read paired-end dataset from our
217 BES dataset.

218 Clipped BES were used for an additional round of contig extension and scaffolding with *SSPACE*
219 (Boetzer et al., 2011). For each LG, we used unambiguously anchored scaffolds and contigs
220 originating from splitted scaffolds matching to this LG. Clipped BES were aligned to scaffolds using
221 *Bowtie2* (Langmead & Salzberg, 2012), allowing at most 3 mismatches per 100-bp read. Scaffolding
222 parameters to *SSPACE* were set to a minimum of 5 links (-k) to validate a new scaffold pair and a
223 maximum link ratio of 0.7 (-a). Scaffolding was only permitted between scaffolds of at least 1000 bp,
224 using an allowed insert size range of 75-225 kb.

225

226 ***Scaffold anchoring to the linkage map***

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

236

237 ***Local recombination rate estimation***

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

244

245 ***RAD genotyping of natural populations***

246 We sampled four wild populations of *C. gigas* and *C. angulata* from both their native Asian and
247 introduced Atlantic areas. We used 24 individuals of *C. gigas* from both Japan (native) and French
248 Britany (introduced), and 24 individuals of *C. angulata* from both Taiwan (native) and Portugal
249 (introduced). We prepared and sequenced four 24-plex RAD libraries using the same protocol as
250 described above.

251 Cleaned demultiplexed reads were mapped to each newly assembled pseudo-chromosome using
252 *Bowtie2* (Langmead & Salzberg, 2012) with the very-sensitive option, allowing a maximum of 7
253 mismatches per alignment. SNPs were called from aligned reads with *Stacks* using a minimum read
254 depth of 5x per individual per allele (Catchen et al., 2011, Catchen et al., 2013). The correction
255 module *rxstacks* was used to re-evaluate individual genotypes and exclude low-quality variants with a
256 cutoff log-likelihood value of -500. Only RAD loci that were successfully genotyped in at least 80%
257 of the samples in each population were retained for subsequent population genomic analyses.

258

259 ***Population genomic analyses***

260 We used *VCFtools* v0.1.11 (Danecek et al., 2011) to apply within-population filters to exclude loci
261 showing more than 4 missing genotypes over 24 individuals, as well as markers showing departure to
262 Hardy-Weinberg equilibrium within at least one population using a p-value cutoff of 0.01. Nucleotide
263 diversity (π), 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 & 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 performed with the R package *Adegenet* (Jombart, 2008), using the global SNP dataset 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 between species, we used a nonparametric quantile regression approach. Recombination rate and F_{ST} values (averaged between the Asian and European species pair) were averaged in 500 kb windows to increase the number of informative sites per window. We excluded windows with recombination rate values exceeding 9 cM/Mb (i.e. corresponding to the 98th percentile of the distribution of estimated recombination rate). This cutoff aimed at removing outlying recombination rate values estimated along chromosome IX (Figure 1). The 97.5th quantile regression fit of the distribution of F_{ST} as a function of recombination rate was computed at each of 10 equally spaced recombination intervals distributed over the range of recombination values.

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 *δaδi* (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 migration rates: IM2*m*, AM2*m* and SC2*m* (Tine et al., 2014). These models offer simplified but more

292 realistic representations of the speciation process, by enabling loci to be either neutrally exchanged
293 between populations, or to have a reduced effective migration rate to account for the direct and
294 indirect effects of selection. Each model was fitted to the observed JAFS (singletons were masked)
295 using three successive optimization steps: hot simulated annealing, cold simulated annealing and
296 BFGS. Comparison among models was made using the Akaike Information Criterion (AIC).
297
298 **Results**
299 ***Construction of an integrated high-density linkage map***
300 The microsatellite and SNP datasets used for constructing the low-density linkage map contained 133
301 informative markers in family F2-19 (293 progeny, Table S1) and 137 informative markers in family
302 F2-21 (282 progeny, Table S2). Ten linkage groups were found in each family in agreement with the
303 haploid number of chromosomes of the species, and the two maps generated by the first round of
304 regression mapping respectively contained 98 and 105 mapped markers in family F19 and F21. The
305 integrated map obtained after identifying homologous LG pairs between families contained 136
306 markers successfully ordered after the first round of regression mapping. This sex-averaged consensus
307 low-density linkage map displayed a strong collinearity with the two maps from which it was derived
308 (Supplementary Figure S2, left panel), indicating that the ordering of markers was highly consistent
309 between the two families. We thus used it as an anchoring map to support the construction of the high-
310 density RAD-linkage map.
311 The average number of filtered RAD-Sequencing reads obtained per individual was similar
312 between family F2-19 (1.8×10^6) and F2-21 (1.9×10^6). After correcting for missing and miscalled
313 haplotypes in parental samples, the number of informative markers was 1278 in family F2-19 and 996
314 in family F2-21 (Table S3, Table S4). A total of 1855 markers were assigned to linkage groups (1231
315 in family F19 and 935 in family F21, 311 in common), and the number of markers per linkage groups
316 was highly positively correlated between families (R^2=0.71, Supplementary Figure S3). Using the
317 anchoring map to set fixed orders, the two RAD maps generated after three rounds of regression
318 mapping respectively contained 1023 and 705 mapped markers in family F19 and F21, respectively.
319 The final combination of these two maps resulted in a sex-averaged consensus map containing 1516

markers (Table 1, Table S5, and Supplementary Figure S2, 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 Figure S4).

Table 1. Summary statistics for the 10 linkage groups of the new C. gigas high-density linkage map and the reassembled physical genome.

LG	Correspondance 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 MareyMap (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

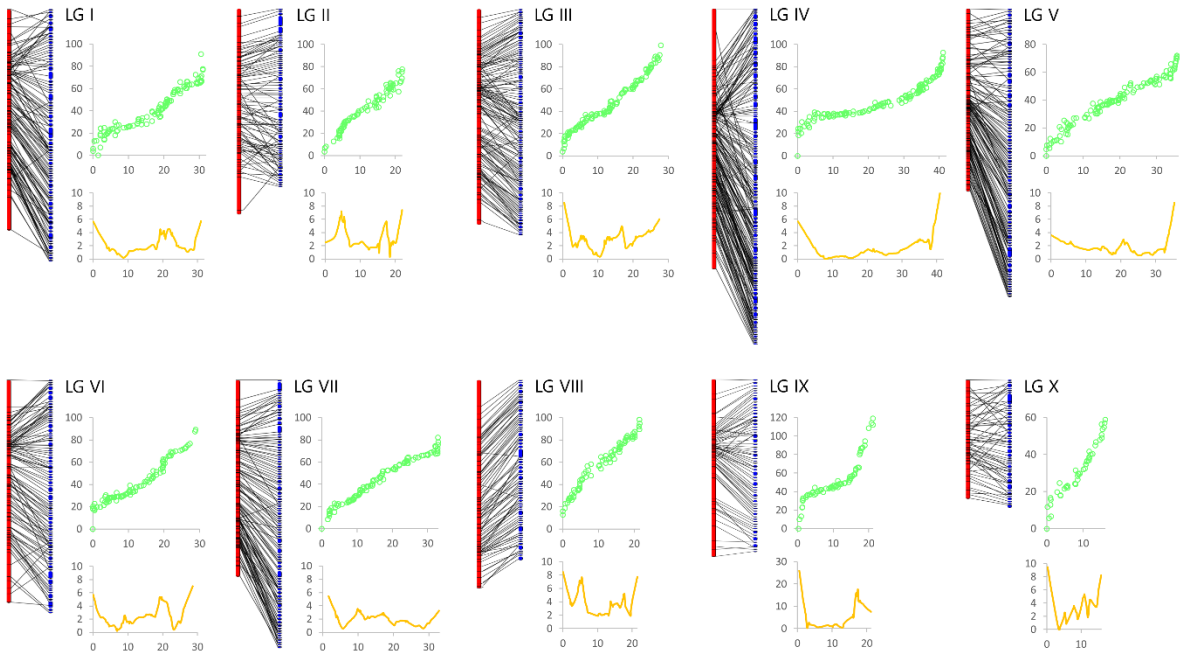
Pseudo-chromosome assembly

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 (Table S6). 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.

Contig extension and scaffolding using clipped BES allowed the merging of 4162 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 (Table S7). A total of 773 scaffolds from this new set of extended scaffolds were anchored to 1161 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 (Table S8). The pseudo-chromosomes assembly obtained had a length of 280.2 Mb (Figure 1), representing 50.1% of the total length of the oyster_v9 assembly (Zhang et al., 2012).

Fig. 1. Anchoring of newly edited scaffolds (blue) into pseudo-chromosomes 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*.

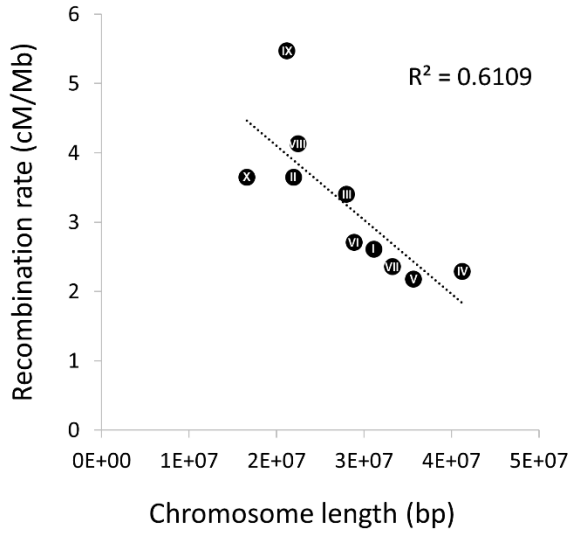


356 ***Local recombination rate***

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

365

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

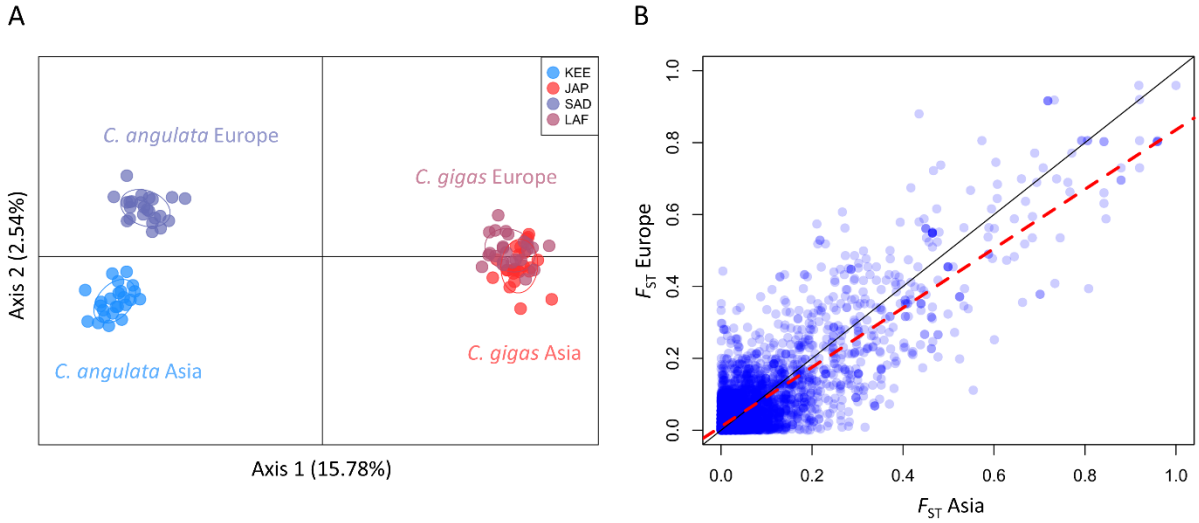


371

372

373 ***Genetic diversity and differentiation***

374 The average number of sequence reads retained per individual for calling genotypes was 3.1×10^6.
375 After filtering for missing genotypes, HWE and a minor allele frequency of 1%, we kept a total of
376 10,144 SNPs from 1,325 RAD loci for downstream genetic diversity analyses. Within-population
377 nucleotide diversity was elevated in both species and showed very similar levels between native and
378 introduced populations (*C. angulata*: $\pi_{KEE} = 0.0095$, $\pi_{SAD} = 0.0096$; *C. gigas*: $\pi_{JAP} = 0.0099$, $\pi_{LAF} =$
379 0.0101). The genome-wide averaged differentiation assessed by F_{ST} between native and introduced
380 populations from the same species was higher in *C. angulata* ($F_{ST\ KEE\text{-}SAD} = 0.0179$) compared to *C.*
381 *gigas* ($F_{ST\ JAP/LAF} = 0.0130$). The mean genetic differentiation between species was very similar
382 between Asia ($F_{ST\ KEE/JAP} = 0.0440$) and Europe ($F_{ST\ SAD/LAF} = 0.0459$).
383
384 **Fig. 3.** (**A**) dAPC plot of the four populations of Pacific cupped oysters assessed with 1,325 SNPs (one
385 SNP per RAD), with C. angulata from Asia (KEE, light blue) and Europe (SAD, dark blue), and C.
386 gigas from Asia (JAP, light red) and Europe (LAF, dark red). (**B**) Genetic parallelism in the level of
387 differentiation between species in Asia (x axis) and Europe (y axis) measured at 10,144 SNPs. The
388 black line shows the y = x equation line and the red dashed line corresponds to the linear regression
389 ($F_{ST\ Europe} = 0.83 \times F_{ST\ Asia} + 0.01$, $R^2 = 0.71$, $P < 10^{-10}$).



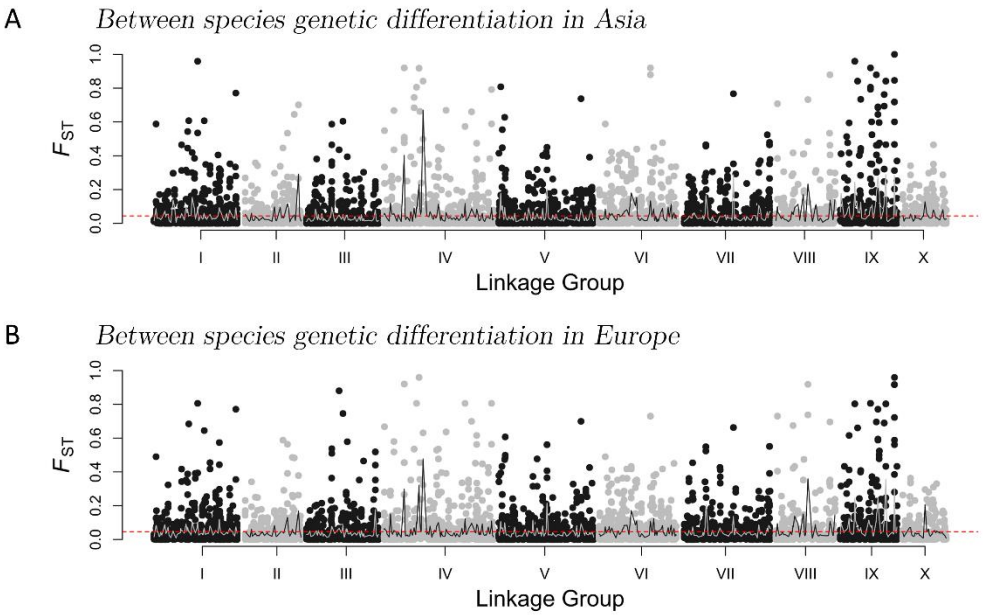
390
391
392

393 The between-population genetic structure revealed by the dAPC (Figure 3A) separated the two
394 species along the first PC axis (explaining 15.78% of total variance) and the Asian and European
395 populations of each species along the second axis (2.54% of total variance). The European populations
396 of *C. angulata*, and to a lesser extent *C. gigas*, were slightly shifted toward the center of the first PC
397 axis, indicating an increased genetic similarity of the two species in Europe.

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

403

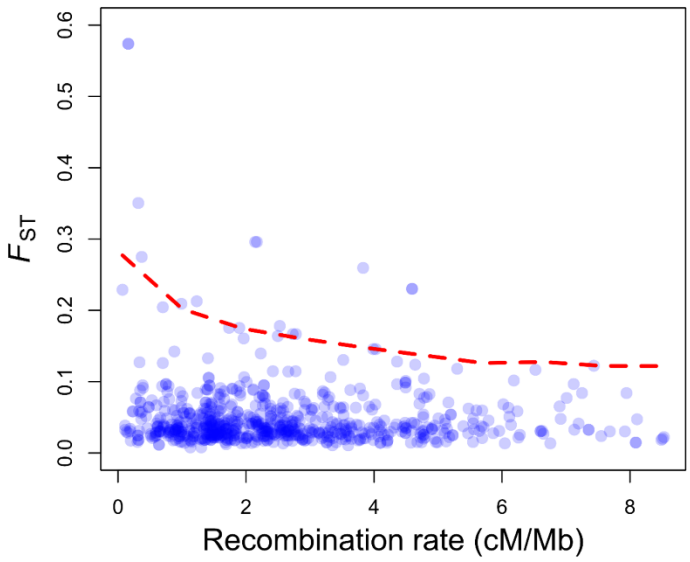
404 **Fig. 4.** Genomic landscape of between-species genetic differentiation (F_{ST}) measured at 10,144 SNPs
405 (alternatively represented by black and grey points for odd and even chromosome numbers for
406 convenience) along the ten Pacific oyster linkage groups. The red dotted line shows the genome-wide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



409

410

411 Chromosomal variation in average genetic differentiation between species was related to the local
412 recombination rate. More precisely, we found a decreasing trend in the maximal amount of
413 differentiation (measured by the 97.5th quantile of the distribution) with increasing local recombination
414 rate (Figure 5). Moreover, within-species nucleotide diversity was significantly positively correlated
415 with the local recombination rate (Supplementary Figure S5).
416
417 **Fig. 5.** Distribution of between-species genetic differentiation (F_{ST}) as a function of the local
418 recombination rate assessed with *MareyMap*. Each point represents an estimate of F_{ST} and
419 recombination rate averaged in a 500 kb window. Outlier recombination rate values exceeding 9
420 cM/Mb were excluded. The red dashed line represents the nonparametric 97.5th quantile regression fit
421 of F_{ST} as a function of the local recombination rate.



422
423
424 ***Divergence history***
425 The demographic history of divergence between *C. angulata* and *C. gigas* inferred with *δaδi* showed
426 evidence for a secondary contact in both Asia and Europe (Table 2). The JAFS presented in Figure 6
427 show a high proportion of shared polymorphisms at high frequency (in the upper right and lower left
428 corners) which is a characteristic footprint of secondary introgression that is not expected under a
429 strict isolation 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. This model received a good statistical support compared to 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 (Figure 6).

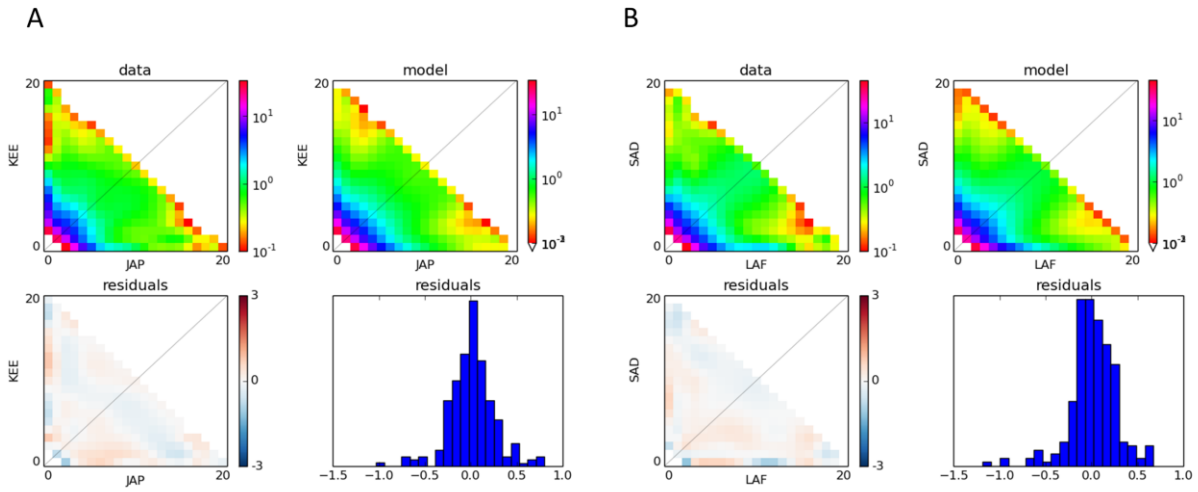
Table 2. Results of model fitting for seven alternative models of divergence between *C. angulata* and *C. gigas* in Asia and Europe. In the order of appearance in the table: the model fitted, the maximum likelihood estimate over 20 independent runs after 3 successive optimisation steps: simulated annealing hot, cold and BFGS, the Akaike Information Criterion, the difference in AIC with the best model (SC2*m*), 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_S), 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}).

Species pair	Model	anneal. hot	anneal. cold	BFGS	AIC	Δ_i	w_i	θ	N_1	N_2	m_{12}	m_{21}	m'_{12}	m'_{21}	T_S	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	-
	IM2*m*	-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
	AM2*m*	-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
	SC2*m*	**-218.2059**	**-218.2059**	**-207.5581**	**433.1162**	**0**	**0.9963**	**56.5407**	**1.5600**	**1.5353**	**10.6952**	**11.2826**	**1.4966**	**1.5031**	**1.1291**	**-**	**0.0536**	**0.6336**
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	-
	IM2*m*	-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
	AM2*m*	-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
	SC2*m*	**-233.321**	**-233.321**	**-226.910**	**471.821**	**0**	**0.9566**	**85.6599**	**1.1751**	**1.3148**	**10.3971**	**12.8181**	**0.9340**	**2.0132**	**0.6982**	**-**	**0.0644**	**0.5527**

Therefore, the semi-permeable species-barrier model, in which introgression occurs at variable rates across the genome since secondary contact, explains well the observed data. In the neutrally introgressing fraction of the genome (*i.e.* 55 to 63% of the genome), gene flow was found to occur in both directions at closely similar rates corresponding to 12 to 18 migrants per generation (N_em, Table 2). By contrast, there was a 6 to 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 to 21 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 secondary contact.

Fig. 6. Demographic divergence history of *C. angulata* and *C. 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 (SC2*m*) and occurring in proportions *P* (for freely introgressing loci) and 1-*P* (for loci with reduced effective migration rates) across the genome. Each plot shows the observed JAFS (upper left), the JAFS obtained under the maximum-likelihood SC2*m* model (upper right), the spectrum of residuals with over (red) and under (blue) predicted SNPs (lower left), and the distribution of residuals (lower right).



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 & Hedgecock, 2004, Sauvage et al., 2010, Hubert et al., 2009, Plough & Hedgecock, 2011, Guo et al., 2012, Li & Guo, 2004, Zhong et al., 2014a) and more recent second-generation maps

(Hedgecock et al., 2015, Wang et al., 2016) that were produced in the Pacific oyster. Compared to the two most recent linkage maps, this map offers a slightly improved resolution, with an average marker-spacing of 0.64 cM compared to 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 distortions observed across most of the 10 linkage groups (Plough, 2012, Plough & Hedgecock, 2011, Launey & Hedgecock, 2001, 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 to 218 kb), but the rescaffolding step using BES increased the scaffold N50 again by 47% (320 kb). Due to 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 pseudo-chromosomes 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 (890-1084 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 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 & Roesti, 2017), 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. 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 and/or crossover interference.

The recombinational 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 & 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 to 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. In addition, other invertebrate species studied so far tend to have smaller genomes than the oyster genome. 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 (Boudry, et al. 2002; Harrang, et al. 2013; Hedgecock 1994; Hedgecock and Pudovkin 2011; 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 (*i*) determine the extent and variation in the level of differentiation within and between *C. angulata* and *C. gigas* across the genome, (*ii*) test the existence

555 of differences in the genomic landscape of differentiation between native and introduced species pairs,
556 (*iii*) infer the history of gene flow during divergence, and finally (*iv*) relate interspecies divergence to
557 underlying evolutionary processes including demography, selection and genomic constraints.

558 Our results reveal that the European introduction of *C. angulata* in the 16th century, and more
559 recently of *C. gigas* in the 70's, did not lead to a reduction in genetic diversity for both species
560 compared to source populations in their native range. This observation is rather common in the
561 literature of marine invasions, and has been attributed to the combined effects of multiple
562 introductions and high propagule pressure (Viard et al., 2016). Another possibility is that the initiation
563 of the demographic wave of invasion can simply not occur without a sufficiently large number of
564 individuals in species with a strong Allele effect like sessile broadcast spawners. In other words a
565 successful marine introduction cannot happen without a high number of funders. We found a relatively
566 low background level of differentiation between *C. gigas* and *C. angulata* (i.e. genome-wide averaged
567 F_{ST} = 0.045), which is consistent with previous estimates based on microsatellite loci (Huvet et al.,
568 2004, Huvet et al., 2000). However, differentiation was highly heterogeneous across the genome in
569 both Asian and European populations, with peaks of elevated F_{ST} values being found in most
570 chromosomes, sometimes even reaching differential allelic fixation in Asia. These 'genomic islands of
571 differentiation' are commonly observed in genome scans for divergence between closely related
572 species pairs (Wolf & Ellegren, 2017), and multiple mechanisms have been proposed to explain their
573 formation (Cruickshank & Hahn, 2014, Burri et al., 2015, Ravinet et al., 2017, Yeaman et al., 2016).
574 They broadly fall in two categories: (*i*) mechanisms of differential gene flow involving the existence
575 of reproductive isolation and/or local adaptation loci acting as genetic barriers to interspecific gene
576 flow (Barton & Bengtsson, 1986, Feder & Nosil, 2010), and (*ii*) heterogeneous divergence
577 mechanisms due to variation in the rate of lineage sorting across the genome (Burri et al., 2015,
578 Cruickshank & Hahn, 2014, Burri, 2017).

579 In order to test whether specific adaptations to different habitats participate to genetic barriers, we
580 hypothesized that the different ecological conditions encountered by oysters in Europe would either
581 result in relaxed divergent selection pressures on genes involved in local adaptation in Asia, or in new
582 selective constraints targeting different subsets of genes. In both cases, we expect that the new

583 environmental conditions in Europe would promote a remodeling of the genomic landscape of species
584 divergence and hence reduce the extent of parallelism in genetic differentiation between species-pairs
585 from native and introduced populations. Contrary to that prediction, we found a remarkably high
586 degree of divergence parallelism indicating that the genetic architecture of reproductive isolation
587 between *C. angulata* and *C. gigas* has not been reinforced, nor significantly weakened following the
588 co-introduction of the two species in Europe. Nevertheless, a slightly lower interspecific
589 differentiation was found in Europe. This reduction was not driven by the most differentiated loci (Fig.
590 3B), and therefore rather indicated increased gene flow across the whole genome than relaxed
591 divergent selection on barrier loci in the novel habitat. Moreover, the result of the PCA suggests that
592 more pronounced gene flow is ongoing from *C. gigas* to *C. angulata* than in the opposite direction in
593 Europe, as revealed by the more pronounced shift of *C. angulata* samples from Portugal towards the
594 central part of the first PCA axis compared to *C. gigas* samples from Brittany (Fig. 3A). This
595 asymmetrical introgression pattern seems consistent with a demographic imbalance due to a higher
596 abundance of *C. gigas* in European farms. Overall, these results suggest that the particular
597 demographic conditions imposed by aquaculture in southern Europe have facilitated opportunities for
598 genetic interactions between species, without modifying the nature of the species boundary.

599 The finding of unaltered genomic differentiation patterns in the introduced species-pair despite
600 increased gene flow suggests either that the contact is too recent to observe significant effects, or that
601 historical and architectural genomic features have played major role in shaping the genomic landscape
602 of species divergence between Pacific cupped oysters. Moreover, heterogeneous genome divergence is
603 predicted by the semipermeable species boundary model, in which neutral introgression is only
604 permitted in genomic regions unlinked to reproductive isolation barriers (Harrison & Larson, 2014,
605 Harrison & Larson, 2016). Under this model, elevated F_{ST} values above the background level indicate
606 the location of genomic regions that are resistant to introgression due to the presence of reproductive
607 isolation loci (i.e. speciation genes). The existence and the possible origin of a semi-permeable barrier
608 to gene flow between *C. gigas* and *C. angulata* was addressed through the comparison of various
609 theoretical models offering simplified representations of contrasted evolutionary scenarios. Our results
610 indicated that contemporary differentiation patterns likely result from a long period of allopatric

611 divergence followed by a recent secondary contact with different rates of introgression among loci.
612 This model consistently outperformed alternative scenarios in both native and introduced ranges,
613 predicting observed differentiation patterns with very small residual errors (Figure 6). Inferred model
614 parameters capturing semi-permeability indicated that the fraction of the genome experiencing
615 reduced introgression rates amounts to 37-45 %, consistent with the view that the species barrier is
616 still largely permeable to gene flow. Similar estimates have been obtained in other marine species-
617 pairs analyzed with the same approach (Tine et al., 2014, Le Moan et al., 2016, Rougemont et al.,
618 2017). In all these cases post-glacial secondary contacts between allopatrically diverged lineages have
619 been inferred. The northwestern Pacific region is known to harbor multiple cases of marine species
620 subdivided into divergent lineages that are supposed to have originated in separate glacial refugia,
621 most likely corresponding to the current seas of Japan and China (reviewed in Ni et al., 2014). Our
622 study provides the first genome-wide view of the evolutionary consequences of past sea-level
623 regression in this region, and adds to previous studies suggesting their important role in initiating
624 speciation in northwestern Pacific taxa (e.g. Shen et al., 2011).

625 In addition to providing evidence for partial reproductive isolation between *C. gigas* and *C.*
626 *angulata*, our results also bring new empirical support for the role of recombination rate variation in
627 shaping the genomic landscape of species divergence. A recent study in the European sea bass has
628 shown that low-recombining regions experiencing accelerated rates of lineage sorting during allopatric
629 phases are preferentially involved in the barrier to gene flow during secondary contact (Duranton et
630 al., 2017). The negative relationship between differentiation and recombination in Pacific cupped
631 oysters is consistent with the idea that foreign alleles are more efficiently removed by selection after
632 introgression when recombination is locally reduced (Schumer, et al. 2017). This interpretation also
633 suggests that the sites under selection are widespread across the genome, although not individually
634 under strong selection (Aeschbacher et al., 2017). Evidence for this type of genetic architecture has
635 recently been evidenced from hybridization studies in a number of animal and plant species (Simon et
636 al., 2017).
637

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 secondary contact, we show that these semi-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.

Conflict of Interest

All the authors declare no conflict of interest concerning the data presented here.

Acknowledgements

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 Hélène 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 plateform Maladies Métaboliques et Infectieuses.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W. & Zinner, D. 2013. Hybridization and speciation. Journal of Evolutionary Biology 26: 229-246.

Aeschbacher, S., Selby, J. P., Willis, J. H. & Coop, G. 2017. Population-genomic inference of the strength and timing of selection against gene flow. Proceedings of the National Academy of Sciences 114: 7061-7066.

Alcala, N., Jensen, J. D., Telenti, A. & Vuilleumier, S. 2016. The genomic signature of population reconnection following isolation: from theory to HIV. G3: Genes, Genomes, Genetics 6: 107-120.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: e3376.

Barton, N. H. & Bengtsson, B. O. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57: 357-376.

Batista F.M., Fonseca V.G., Ruano F., Boudry P. 2017. Asynchrony in settlement time between the closely related oysters Crassostrea angulata and C. gigas in Ria Formosa lagoon (Portugal). Marine Biology 164: 110.

Berner, D. & Roesti, M. 2017. Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. Molecular Ecology 26: 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. Molecular Ecology 20: 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. Current Zoology 59: 72-86.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579.

Boudry, P., Heurtebise, S., Collet, B., Cornette, F. & Gérard, A. 1998. Differentiation between populations of the Portuguese oyster, Crassostrea angulata (Lamark) and the Pacific oyster, Crassostrea gigas (Thunberg), revealed by mtDNA RFLP analysis. Journal of experimental marine biology and ecology 226: 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: 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 Morest Project. Versailles, France: Quae Editions: 153-196.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S. & Garamszegi, L. Z. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome research: gr. 196485.115.

Burri, R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection. Evolution Letters 1: 118-131.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. 2011. Stacks: Building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics 1: 171-182.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. 2013. Stacks: an analysis tool set for population genomics. Molecular ecology 22: 3124–3140.

Charlesworth, B., Morgan, M. & Charlesworth, D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303.

720 Coop, G. 2016. Does linked selection explain the narrow range of genetic diversity across species?
721 bioRxiv: 042598.
722 Corbett-Detig, R. B., Hartl, D. L. & Sackton, T. B. 2015. Natural selection constrains neutral diversity
723 across a wide range of species. PLoS Biology 13: e1002112.
724 Cruickshank, T. E. & Hahn, M. W. 2014. Reanalysis suggests that genomic islands of speciation are
725 due to reduced diversity, not reduced gene flow. Molecular Ecology 23: 3133-3157.
726 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E.,
727 Lunter, G., Marth, G. T. & Sherry, S. T. 2011. The variant call format and VCFtools.
728 Bioinformatics 27: 2156-2158.
729 Darling J.A., Bagley M.J., Roman J., Tepolt C.K., Geller J.B. 2008. Genetic patterns across multiple
730 introductions of the globally invasive crab genus Carcinus. Molecular Ecology 17: 4992-5007.
731 Dégremont, L., Bédier, E. & Boudry, P. 2010. Summer mortality of hatchery-produced Pacific oyster
732 spat (Crassostrea gigas). II. Response to selection for survival and its influence on growth and
733 yield. Aquaculture 299: 21-29.
734 Dégremont L. 2011. Evidence of herpesvirus (OsHV-1) resistance in juvenile Crassostrea gigas
735 selected for high resistance to the summer mortality phenomenon. Aquaculture 317: 94-98.
736 Dukić, M., Berner, D., Roesti, M., Haag, C. R. & Ebert, D. 2016. A high-density genetic map reveals
737 variation in recombination rate across the genome of Daphnia magna. BMC genetics 17: 137.
738 Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F. & Gagnaire, P.-A. 2017. The origin
739 and remolding of genomic islands of differentiation in the European sea bass. bioRxiv:
740 223750.
741 Feder, J. L. & Nosil, P. 2010. The Efficacy of Divergence Hitchhiking in Generating Genomic Islands
742 during Ecological Speciation. Evolution 64: 1729-1747.
743 Feder, J. L., Egan, S. P. & Nosil, P. 2012. The genomics of speciation-with-gene-flow. Trends in
744 Genetics 28: 342-350.
745 Fraïsse, C., Belkhir, K., Welch, J. J. & Bierne, N. 2016. Local interspecies introgression is the main
746 cause of extreme levels of intraspecific differentiation in mussels. Molecular Ecology 25: 269-
747 286.
748 Gaffney P.M. Editor. Journal of Shellfish Research. 2008.
749 Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S. &
750 Bierne, N. 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine
751 populations in the genomic era. Evolutionary Applications 8: 769-786.
752 Grade, A., Chairi, H., Lallias, D., Power, D. M., Ruano, F., Leitão, A., Drago, T., King, J. W., Boudry,
753 P. & Batista, F. M. 2016. New insights about the introduction of the Portuguese oyster,
754 Crassostrea angulata, into the North East Atlantic from Asia based on a highly polymorphic
755 mitochondrial region. Aquatic Living Resources 29: 404.
756 Grizel, H. & Héral, M. 1991. Introduction into France of the Japanese oyster (Crassostrea gigas). ICES
757 Journal of Marine Science 47: 399-403.
758 Guo, X., Li, Q., Wang, Q. Z. & Kong, L. F. 2012. Genetic mapping and QTL analysis of growth-
759 related traits in the Pacific oyster. Marine Biotechnology 14: 218-226.
760 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. 2009. Inferring the Joint
761 Demographic History of Multiple Populations from Multidimensional SNP Frequency Data.
762 PLoS Genet 5: e1000695.
763 Harrang, E., Lapègue, S., Morga, B. & Bierne, N. 2013. A high load of non-neutral amino-acid
764 polymorphisms explains high protein diversity despite moderate effective population size in a
765 marine bivalve with sweepstakes reproduction. G3: Genes, Genomes, Genetics 3: 333-341.
766 Harrison, R. G. & Larson, E. L. 2014. Hybridization, Introgression, and the Nature of Species
767 Boundaries. Journal of Heredity 105: 795-809.
768 Harrison, R. G. & Larson, E. L. 2016. Heterogeneous genome divergence, differential introgression,
769 and the origin and structure of hybrid zones. Molecular ecology 25: 2454–2466.
770 Hedgecock D. 1994. Does variance in reproductive success limit effective population sizes of marine
771 organisms? In: Beaumont A, editor. Genetics and evolution of aquatic organisms. London,
772 UK.: Chapman & Hall. p. 122−134.
773 Hedgecock, D. & Pudovkin, A. I. 2011. Sweepstakes reproductive success in highly fecund marine
774 fish and shellfish: a review and commentary. Bulletin of Marine Science 87: 971-1002.

775 Hedgecock, D., Shin, G., Gracey, A. Y., Van Den Berg, D. & Samanta, M. P. 2015. Second-
776 generation linkage maps for the pacific oyster Crassostrea gigas reveal errors in assembly of
777 genome scaffolds. G3: Genes, Genomes, Genetics 5: 2007-2019.
778 Hellberg, M. E. 2009. Gene Flow and Isolation among Populations of Marine Animals. Annual
779 Review of Ecology, Evolution, and Systematics 40: 291-310.
780 Hsiao, S.-T., Chuang, S.-C., Chen, K.-S., Ho, P.-H., Wu, C.-L. & Chen, C. A. 2016. DNA barcoding
781 reveals that the common cupped oyster in Taiwan is the Portuguese oyster Crassostrea
782 angulata (Ostreoida; Ostreidae), not C. gigas. Scientific reports 6: 34057.
783 Hubert, S. & Hedgecock, D. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster
784 Crassostrea gigas. Genetics 168: 351-362.
785 Hubert, S., Cognard, E. & Hedgecock, D. 2009. Centromere mapping in triploid families of the Pacific
786 oyster Crassostrea gigas (Thunberg). Aquaculture 288: 172-183.
787 Huvet, A., Lapegue, S., Magoulas, A. & Boudry, P. 2000. Mitochondrial and nuclear DNA
788 phylogeography of Crassostrea angulata, the Portuguese oyster endangered in Europe.
789 Conservation Genetics 1: 251-262.
790 Huvet, A., Balabaud, K., Bierne, N. & Boudry, P. 2001. Microsatellite analysis of 6-hour-old embryos
791 reveals no preferential intraspecific fertilization between cupped oysters Crassostrea gigas and
792 Crassostrea angulata. Marine Biotechnology 3: 448-453.
793 Huvet, A., Gérard, A., Ledu, C., Phélipot, P., Heurtebise, S. & Boudry, P. 2002. Is fertility of hybrids
794 enough to conclude that the two oysters Crassostrea gigas and Crassostrea angulata are the
795 same species? Aquatic Living Resources 15: 45-52.
796 Huvet, A., Fabioux, C., Mccombie, H., Lapegue, S. & Boudry, P. 2004. Natural hybridization between
797 genetically differentiated populations of Crassostrea gigas and C. angulata highlighted by
798 sequence variation in flanking regions of a microsatellite locus. Marine Ecology Progress
799 Series 272: 141-152.
800 Jeffery N., DiBacco C., Wringe B., Stanley R., Hamilton L., Ravindran P., Bradbury I. 2017. Genomic
801 evidence of hybridization between two independent invasions of European green crab
802 (Carcinus maenas) in the Northwest Atlantic. Heredity 119: 154.
803 Jeffery N.W., Bradbury I.R., Stanley R.R., Wringe B.F., Van Wyngaarden M., Lowen J.B., McKenzie
804 C.H., Matheson K., Sargent P.S., DiBacco C. 2018. Genomewide evidence of environmentally
805 mediated secondary contact of European green crab (Carcinus maenas) lineages in eastern
806 North America. Evolutionary Applications.
807 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
808 Bioinformatics 24: 1403−1405.
809 Kaback, D. B., Guacci, V., Barber, D. & Mahon, J. W. 1992. Chromosome size-dependent control of
810 meiotic recombination. Science 256: 228-232.
811 Kelley, J. L., Brown, A. P., Therkildsen, N. O. & Foote, A. D. 2016. The life aquatic: advances in
812 marine vertebrate genomics. Nature Reviews Genetics 17: 523.
813 Knowlton, N. 1993. Sibling species in the sea. Annual review of ecology and systematics 24: 189-216.
814 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar,
815 K., Doyle, M. & FitzHugh, W. 2001. Initial sequencing and analysis of the human genome.
816 Nature 409: 860-921.
817 Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:
818 357-359.
819 Lapegue, S., Batista, F., Heurtebise, S., Yu, Z. & Boudry, P. 2004. Evidence for the presence of the
820 Portuguese oyster, Crassostrea angulata, in northern China. Journal of Shellfish Research 23:
821 759-763.
822 Lapegue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., Ballenghien, M.,
823 Genestout, L., Barbotte, L. & Mahla, R. 2014. Development of SNP-genotyping arrays in two
824 shellfish species. Molecular ecology resources 14: 820-830.
825 Launey, S. & Hedgecock, D. 2001. High genetic load in the Pacific oyster Crassostrea gigas. Genetics
826 159: 255-265.
827 Le Moan, A., Gagnaire, P. A. & Bonhomme, F. 2016. Parallel genetic divergence among coastal-
828 marine ecotype pairs of European anchovy explained by differential introgression after
829 secondary contact. Molecular Ecology 25: 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. Molecular Ecology Notes 3: 228-232.

Li, L. & Guo, X. 2004. AFLP-based genetic linkage maps of the Pacific oyster Crassostrea gigas Thunberg. Marine Biotechnology 6: 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: 234-239.

López-Flores, I., de la Herrán, R., Garrido-Ramos, M. A., Boudry, P., Ruiz-Rejón, C. & Ruiz-Rejón, M. 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: 23-35.

Moretto, M., Cestaro, A., Troggio, 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. pp., Ghent, Belgium.

Nachman, M. W. 2002. Variation in recombination rate across the genome: evidence and implications. Current opinion in genetics & development 12: 657-663.

Nachman, M. W. & Payseur, B. A. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B: Biological Sciences 367: 409-421.

Ni, G., Li, Q. I., Kong, L. & Yu, H. 2014. Comparative phylogeography in marginal seas of the northwestern Pacific. Molecular Ecology 23: 534-548.

Noor, M. A. F. & Bennett, S. M. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103: 439-444.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics: 547-572.

Plough, L. V. & Hedgecock, D. 2011. Quantitative trait locus analysis of stage-specific inbreeding depression in the Pacific oyster Crassostrea gigas. Genetics 189: 1473-1486.

Plough, L. V. 2012. Environmental stress increases selection against and dominance of deleterious mutations in inbred families of the Pacific oyster Crassostrea gigas. Molecular Ecology: 21:3974-3987.

Plough, L. V. 2016. Genetic load in marine animals: a review. Current Zoology 62: 567-579.

Ravinet, M., Faria, R., Butlin, R., Galindo, J., Bierne, N., Rafajlović, M., Noor, M., Mehlig, B. & Westram, A. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. Journal of evolutionary biology 30: 1450-1477.

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C. & Panova, M. 2015. Shared and nonshared genomic divergence in parallel ecotypes of Littorina saxatilis at a local scale. Molecular ecology. 27: 285-307.

Reece, K. S., Cordes, J. F., Stubbs, J. B., Hudson, K. L. & Francis, E. A. 2008. Molecular phylogenies help resolve taxonomic confusion with Asian Crassostrea oyster species. Marine Biology 153: 709-721.

Ren, J., Liu, X., Jiang, F., Guo, X. & Liu, B. 2010. Unusual conservation of mitochondrial gene order in Crassostrea oysters: evidence for recent speciation in Asia. BMC Evolutionary Biology 10: 394.

Rezvoy, C., Charif, D., Guéguen, L. & Marais, G. A. 2007. MareyMap: an R-based tool with graphical interface for estimating recombination rates. Bioinformatics 23: 2188-2189.

Roesti, M., Moser, D. & Berner, D. 2013. Recombination in the threespine stickleback genome—patterns and consequences. Molecular ecology 22: 3014-3027.

Rose, N. H., Bay, R. A., Morikawa, M. K. & Palumbi, S. R. 2018. Polygenic evolution drives species divergence and climate adaptation in corals. Evolution 72: 82-94.

Rougemont, Q., Gagnaire, P. A., Perrier, C., Genthon, C., Besnard, A. L., Launey, S. & Evanno, G. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular ecology 26: 142–162.

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N. & Bierne, N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS biology 14: e2000234.

Saarman N.P., Pogson G.H. 2015. Introgression between invasive and native blue mussels (genus Mytilus) in the central California hybrid zone. Molecular Ecology 24: 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: 13-22.

Sauvage, C., Boudry, P., De Koning, D. J., Haley, C. S., Heurtebise, S. & Lapègue, S. 2010. QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (Crassostrea gigas). Animal genetics 41: 390-399.

Schumer M., Xu C., Powell D., Durvasula A., Skov L., Holland C., Sankararaman S., Andolfatto P., Rosenthal G., Przeworski M. 2017. Natural selection interacts with the local recombination rate to shape the evolution of hybrid genomes. bioRxiv: 212407.

Sekino, M. & Yamashita, H. 2013. Mitochondrial DNA barcoding for Okinawan oysters: a cryptic population of the Portuguese oyster Crassostrea angulata in Japanese waters. Fisheries science 79: 61-76.

Shen, K.-N., Jamandre, B. W., 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 Evolutionary Biology 11: 83.

Simon, A., Bierne, N. & Welch, J. J. 2017. Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. bioRxiv: 237925.

Stapley, J., Feulner, P. G., Johnston, S. E., Santure, A. W. & Smadja, C. M. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Phil. Trans. R. Soc. B 372: 20160455.

Takeo, I. & Sakai, S. 1961. Study of breeding of Japanese oyster, Crassostrea gigas. Tohoku journal of agricultural research 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). Aquaculture Research 36: 516-518.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., R., D., K., S., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., Canario, A. V. M. & Reinhardt, R. 2014. The European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nature Communications 5.

van Ooijen, J. W. (2006) JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. pp., Kyazma B.V., Wageningen, Netherlands.

Viard, F., David, P. & Darling, J. A. 2016. Marine invasions enter the genomic era: three lessons from the past, and the way forward. Current zoology 62: 629-642.

Wang, H., Qian, L., Liu, X., Zhang, G. & Guo, X. 2010. Classification of a common cupped oyster from southern China. Journal of Shellfish Research 29: 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|Genetics 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. Chinese journal of oceanology and limnology 32: 419-425.

Weir, B. S. & Cockerham, C. C. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358-1370.

Westram, A., Galindo, J., Alm Rosenblad, M., Grahame, J., Panova, M. & Butlin, R. 2014. Do the same genes underlie parallel phenotypic divergence in different Littorina saxatilis populations? Molecular ecology 23: 4603-4616.

Wolf, J. B. & Ellegren, H. 2017. Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics 18: 87–100.

Yamtich, J., Voigt, M. L., Li, G. & Hedgecock, D. 2005. Eight microsatellite loci for the Pacific oyster Crassostrea gigas. Animal genetics 36: 524-526.

938 Yeaman, S., Aeschbacher, S. & Bürger, R. 2016. The evolution of genomic islands by increased
939 establishment probability of linked alleles. Molecular Ecology. 25: 2542-2558.
940 Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong,
941 Z., Que, H., Xie, Y., Holland, P. W. H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng,
942 C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A.,
943 Hedgecock, D., Xu, Z., Liu, Y., Domazet-Loso, T., Du, Y., Sun, X., Zhang, S., Liu, B.,
944 Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng,
945 Z., Li, Y., Li, N., Wang, J., Chen, M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang,
946 H., Du, X., Chen, L., Yang, M., Gaffney, P. M., Wang, S., Luo, L., She, Z., Ming, Y., Huang,
947 W., Zhang, S., Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, J., Wang, Q., Steinberg,
948 C. E. W., Wang, H., Li, N., Qian, L., Zhang, G., Li, Y., Yang, H., Liu, X., Wang, J., Yin, Y. &
949 Wang, J. 2012. The oyster genome reveals stress adaptation and complexity of shell
950 formation. Nature 490: 49–54.
951 Zhong, X., Li, Q., Guo, X., Yu, H. & Kong, L. 2014a. QTL mapping for glycogen content and shell
952 pigmentation in the Pacific oyster Crassostrea gigas using microsatellites and SNPs.
953 Aquaculture international 22: 1877-1889.
954 Zhong, X., Li, Q., Yu, H. & Kong, L. 2014b. SNP mining in Crassostrea gigas EST data:
955 transferability to four other Crassostrea species, phylogenetic inferences and outlier SNPs
956 under selection. PloS one 9: e108256.