

# 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

Estelle Harrang,\* Sylvie Lapègue,\* Benjamin Morga,\* and Nicolas Bière<sup>†,\*1</sup>

\*Ifremer, Laboratoire de génétique et pathologie, 17390 La Tremblade, France, <sup>†</sup>Université Montpellier 2, 34095 Montpellier cedex 5, France, and <sup>‡</sup>CNRS - Institut des Sciences de l'Évolution, UMR5554, Station Méditerranéenne de l'Environnement Littoral, 34200 Sète, France

**ABSTRACT** Marine bivalves show among the greatest allozyme diversity ever reported in Eukaryotes, putting them historically at the heart of the neutralist–selectionist controversy on the maintenance of genetic variation. Although it is now acknowledged that this high diversity is most probably a simple consequence of a large population size, convincing support for this explanation would require a rigorous assessment of the silent nucleotide diversity in natural populations of marine bivalves, which has not yet been done. This study investigated DNA sequence polymorphism in a set of 37 nuclear loci in wild samples of the flat oyster *Ostrea edulis*. Silent diversity was found to be only moderate (0.7%), and there was no departure from demographic equilibrium under the Wright-Fisher model, suggesting that the effective population size might not be as large as might have been expected. In accordance with allozyme heterozygosity, nonsynonymous diversity was comparatively very high (0.3%), so that the nonsynonymous to silent diversity ratio reached a value rarely observed in any other organism. We estimated that one-quarter of amino acid-changing mutations behave as neutral in *O. edulis*, and as many as one-third are sufficiently weakly selected to segregate at low frequency in the polymorphism. Finally, we inferred that one oyster is expected to carry more than 4800 non-neutral alleles (or  $4.2 \text{ cM}^{-1}$ ). We conclude that a high load of segregating non-neutral amino-acid polymorphisms contributes to high protein diversity in *O. edulis*. The high fecundity of marine bivalves together with an unpredictable and highly variable success of reproduction and recruitment (sweepstakes reproduction) might produce a greater decoupling between  $N_e$  and  $N$  than in other organisms with lower fecundities, and we suggest this could explain why a higher segregating load could be maintained for a given silent mutation effective size.

## KEYWORDS

nucleotide  
polymorphism  
marine bivalve  
deleterious  
mutations  
genetic load  
*Ostrea edulis*

Marine bivalves are recognized as being among the most polymorphic of all animal species. First observed with allozyme markers (Ward *et al.*

2000; Bazin *et al.* 2006), this high polymorphism now appears to be confirmed at the nucleotide level (Sauvage *et al.* 2007; Arias *et al.* 2009; Zhang and Guo 2010). However, the cause of the extreme heterozygosity of marine bivalves has remained controversial, and its investigation has been hampered by several factors.

First, the diversity of functional markers (*e.g.*, allozymes or non-synonymous polymorphisms) depends on the concomitant effects of population size and selective constraints on these markers, whereas the diversity of nonfunctional markers (*e.g.*, synonymous or noncoding polymorphisms) only depends on the population size (Kimura 1983). Ideally, genetic diversity should only be estimated using nonfunctional polymorphisms (*i.e.*, silent diversity).

Second, although silent diversity has now been evaluated in a few bivalve species, these studies were designed to discover single-nucleotide polymorphisms (SNPs) and used individuals sampled from hatcheries

Copyright © 2013 Harrang *et al.*

doi: 10.1534/g3.112.005181

Manuscript received October 19, 2012; accepted for publication December 15, 2012

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at <http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.112.005181/-/DC1>

Sequence data from this article have been deposited with the GenBank Data Libraries under accession nos. JN680816–JN680855.

<sup>1</sup>Corresponding author: Institut des Sciences de l'Évolution, Station Méditerranéenne de l'Environnement Littoral, 2, rue des Chantiers, 34200 Sète, France. E-mail: n-biere@univ-montp2.fr

rather than the wild, precluding an inference of the frequency of mutations in natural populations. Furthermore, the use of SNP density as a measure of diversity does not provide an accurate statistic for rigorous comparison among species and can be misleading.

Third, extreme polymorphism is sometimes prone to selectionist interpretations. This was true at the allozyme time, during which numerous studies questioned neutrality of these markers in marine mollusks (Koehn and Shumway 1982; Koehn 1990; Karl and Avise 1992; Mitton 1993, 1997; Riginos *et al.* 2002). More recently this has been reiterated in some surveys of DNA sequence polymorphism of protein-coding genes (Moy *et al.* 2008; Wharam *et al.* 2008; Costa *et al.* 2009; Parisi *et al.* 2009). It is therefore crucial to build a clearer picture by evaluating the genomic average of diversity of various categories of substitutions from a large panel of genes in different species of bivalves. Only knowledge of this kind would allow us to confidently say that the diversity of a specific protein was extreme.

Finally, and perhaps most importantly, large variation in reproductive success owing to a skewed offspring distribution can lead to effective population sizes several orders of magnitude below census numbers and the effective size of marine populations might not be as large as was first anticipated (Hedegcock 1994; Hedegcock and Pudovkin 2011). A skewed offspring distribution can also complicate the interpretation of descriptive statistics of genetic variation as this creates deviation from the standard Wright-Fisher model and the Kingman's coalescent (Sargsyan and Wakeley 2008, Eldon and Wakeley 2009, Der *et al.* 2012). All these caveats may have fueled the doubt cast on the simplest hypothesis: that the extreme protein diversity of marine mollusks is a direct consequence of their large population size. The present paper addresses this doubt.

We studied DNA sequence polymorphism in the flat oyster *Ostrea edulis* by direct resequencing of 37 loci identified from expressed sequence tag (EST) libraries (Morga *et al.* 2011, 2012). Samples were taken from different natural populations throughout the distribution area of *O. edulis* in Europe. The aim of this study was first to obtain a new estimate of silent diversity based on wild samples in an additional species of bivalve and second to verify whether purifying selection could be relaxed in marine bivalves, as suggested by Sauvage *et al.* (2007), by further investigating the potentially non-neutral nature of segregating nonsynonymous polymorphisms.

## MATERIALS AND METHODS

### Sampling and Molecular Protocols

We used 16 oysters from four different natural populations collected on the Atlantic and Mediterranean coasts, in Italy, Greece and France. Genomic DNA was extracted from samples of gill tissue using the Wizard DNA Clean-Up System (Promega) according to the manufacturer's recommendations. Quality and concentration were assessed on a 1% agarose gel and by using an Eppendorf BioPhotometer. Concentrations were equilibrated to 100 ng.μL<sup>-1</sup> per DNA sample.

Primers were designed from 114 EST sequences by using the online software Primer3 (Rozen and Skaletsky 2000). Expected fragment lengths in the absence of introns varied between 160 bp and 660 bp depending on the EST used. The polymerase chain reaction (PCR) protocol was the same as in Sauvage *et al.* (2007). Purified PCR products were then sequenced with both the forward and the reverse primers, using the ABI Prism BigDye v3 Terminator Cycle sequencing Kit (Applied Biosystems) in an ABI 3130 genetic analyzer (Applied Biosystems). A second PCR amplification and a second sequencing reaction allowed the removal of most of the artifactual mutations inherently produced during the experiments.

### Analysis of DNA sequence polymorphism

Sequence alignment was performed with ClustalW via the BioEdit interface (Hall 1999). Verification by eye and multiple alignments of sequences obtained from different PCR and sequencing reactions for the same individual were used to correct for artifactual mutations. All sites with two simultaneous fluorescent signals were carefully individually inspected and considered to be heterozygous sites if the intensity of the least intense peak was greater than 30% of the intensity of the most intense peak. The cDNA was used to identify intronic regions. A BLAST homology search was performed on GenBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify orthologous sequences which, together with ORF-Finder (Rombel *et al.* 2002), enabled the annotation of open reading frames (ORFs) and untranslated regions (UTRs). Annotations were further investigated and confirmed with Artemis (Rutherford *et al.* 2009). As we used direct sequencing, sequences stopped at the first heterozygous indel, sometimes producing sequences of variable lengths (L). For this reason, some sequences were too short for the analysis, and we obtained variable sample sizes. The validity of each SNP was checked individually on nucleotide sequences and sequence alignments.

SNPs were classified as synonymous and nonsynonymous in coding regions and as noncoding in intronic and UTR regions. Synonymous and noncoding SNPs were classified as silent. Genetic diversity was analyzed with DnaSP v5.10 (Librado and Rozas 2009) and double checked with homemade programs of the Montpellier Bioinformatics Biodiversity platform. We computed the nucleotide diversity ( $\pi$ , Nei 1987) and Watterson's theta ( $\theta$ , Watterson 1975) for each site category: synonymous ( $\pi_s, \theta_s$ ), nonsynonymous ( $\pi_n, \theta_n$ ), noncoding ( $\pi_{nc}, \theta_{nc}$ ), and silent ( $\pi_{si}, \theta_{si}$ ). Departure from the expectations of mutation/drift equilibrium under the Wright-Fisher model was evaluated with silent sites, using Tajima's D statistic (Tajima 1989), which is a standardized measure of the difference between the two diversity estimators ( $\pi$  and  $\theta$ ). Genetic structure, as measured by  $F_{ST}$  statistics (Wright 1951), was analyzed with Arlequin v3.5 (Excoffier *et al.* 2005) among all populations, and between Mediterranean and Atlantic populations. ORFs were used to compute the GC content at the third codon position (GC3) and to measure the effective number of codons (ENC) with DnaSP v5.10 (Librado and Rozas 2009).

### Estimating the fractions of nonsynonymous mutations that are neutral, weakly selected, or strongly deleterious

We used the method of Fay *et al.* (2001) to partition amino acid mutations into three categories: neutral, slightly selected, and strongly deleterious. The format of our dataset did not allow us to estimate a more detailed distribution of fitness effects (*e.g.*, Eyre-Walker *et al.* 2006). Although simpler, the method of Fay *et al.* (2001) does not rely on a specific population genetic model (*e.g.*, the Wright-Fisher model), and in the spirit of the McDonald-Kreitman approach (1991), it capitalizes on the fact that neutral and potentially selected sites are interspersed with each other throughout a gene. Such sites are therefore expected to have the same evolutionary history and sampling (*i.e.*, a shared genealogy). A possible deviation from the standard coalescent, as could be expected in marine species with a skewed offspring distribution (Eldon and Wakeley 2009), is not expected to bias the method, which simply assumes shared genealogies for the two categories of mutations whatever their shapes (*e.g.*, standard, multifurcating, star-like...). The neutral class,  $f_0$ , was estimated from common SNPs; the slightly selected class,  $f_1$ , accounted for the excess of low frequency amino acid SNPs; and the strongly deleterious class,  $f_2$ , was deduced from the synonymous diversity as those nonsynonymous

SNPs that were lacking from the sample. SNPs were classified into frequency categories: those with a minor variant frequency less than 0.1, those with a frequency between 0.1 and 0.2, and those with a frequency greater than 0.2. In addition, following the idea of Fay *et al.* (2001) that deleterious mutations are not expected to spread on a wide geographical scale, we also classified SNPs as shared or non-shared between populations or between the Atlantic Ocean and the Mediterranean Sea. The comparison of the  $\theta_n/\theta_{si}$  ratio among categories allowed  $f_0$ ,  $f_1$ , and  $f_2$  to be estimated (Fay *et al.* 2001).

The number of nonsynonymous SNPs expected if they were all neutral was estimated from the observed number of silent SNPs and used to calculate the proportion of non-neutral nonsynonymous SNPs in the sample. Estimating the number of heterozygote non-neutral amino-acid mutations an individual carries would require knowledge of the frequency of these mutations in the population. An underestimation and thus conservative estimate can nonetheless be obtained following Fay *et al.* (2001) by assuming non-neutral nonsynonymous alleles are present only once in the sample.

## RESULTS

### DNA polymorphism in *O. edulis*

From the 114 loci that we attempted to amplify, 13 presented no amplification signal, which could be explained by the presence of an intron within the priming sequences or a long intron between the two primers, and 47 presented a multibanded pattern. Among the 54 loci that produced a single-banded pattern on agarose gel, 10 produced unreadable electropherograms (overlapping peaks), two were composed of very large introns (>1000 bp), which precluded a forward/reverse sequencing double check, and two could not be aligned with their corresponding cDNA, indicating a lack of specificity of the primers or PCR conditions. Forty loci produced usable sequence sets and were long enough to be annotated and deposited in the GenBank database (accession nos. JN680816 to JN680855). Among these, 23 were composed of one to three large introns, with an individual length from 92 to 983 bp and a total length per locus from 92 to more than 1250 bp. The primers are listed in Supporting Information, Table S1 and a polymorphism analysis of the 40 loci is shown in Table S2. Three loci were mitochondrial (JN680816, JN680817, JN680826) and were discarded from subsequent analyses due to the specific characteristics of mitochondria (uniparental inheritance, nonrecombining, greater selective constraint on mitochondrial than nuclear proteins), as is usual practice in similar studies, including those with which we compare our data in the Discussion. A total of 16,525 base pairs of the European flat oyster genome were successfully sequenced and analyzed (37 nuclear loci, Table S2). Twenty-six of the 37 loci contained at least one noncoding region (intron and/or UTR), and one locus was entirely noncoding. A total of 283 SNPs were identified (Table S3) of which 28% were nonsynonymous, 12% were synonymous, and 60% were noncoding. Only one locus was monomorphic. An average density of 1 SNP every 76 bp was observed in coding regions and 1 every 47 bp in noncoding regions, which are comparable densities to those observed in the cupped oyster *Crassostrea gigas* (Sauvage *et al.* 2007).

The average nucleotide diversity at synonymous sites was  $\pi_s = 0.006$  (range, 0–0.08) and the average Watterson's theta of  $\theta_s = 0.005$ . The average nucleotide diversity at non-coding sites was  $\pi_{nc} = 0.005$  (range, 0–0.017) with an average Watterson's theta of  $\theta_{nc} = 0.006$ . Synonymous and noncoding diversities were not significantly different from each other (nucleotide diversity  $t = 0.97$ ;  $df = 36$ ;  $P = 0.34$ ; Watterson's theta:  $t = 0.29$ ;  $df = 36$ ;  $P = 0.77$ ), so we combined the two categories of polymorphism into a single category

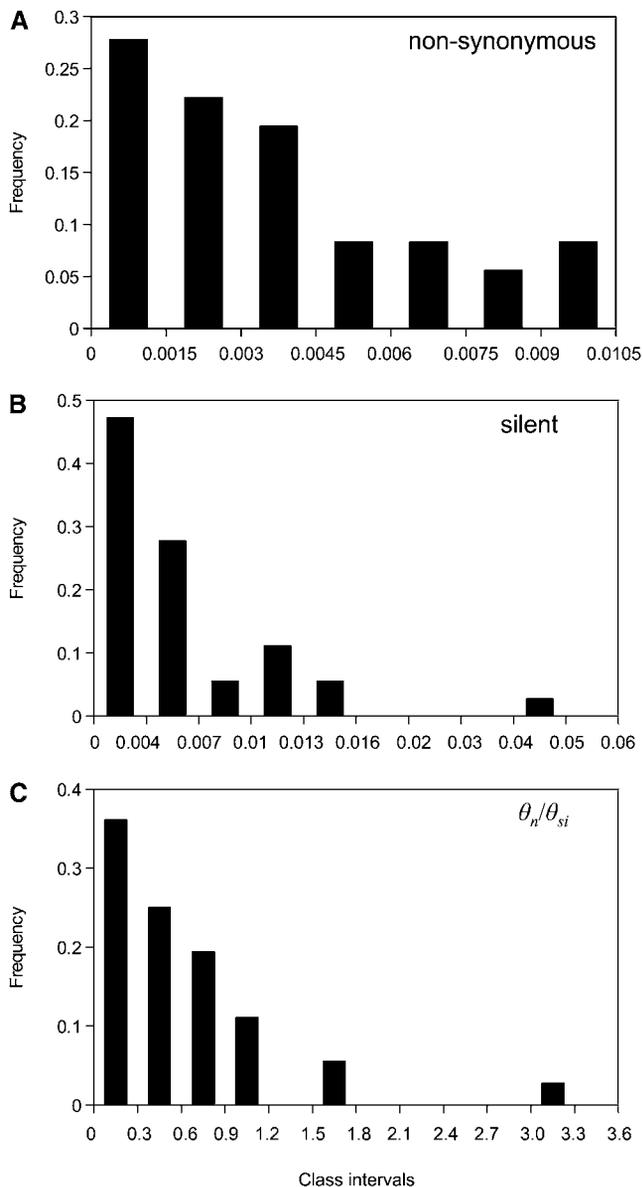
(silent mutations). The average nucleotide diversity at silent sites was  $\pi_{si} = 0.0067$  (range, 0–0.08) with an average Watterson's theta of  $\theta_{si} = 0.0065$ . The similarity between the two estimators ( $\pi_{si}$  and  $\theta_{si}$ ) suggests an absence of any strong deviation from the expectation of demographic equilibrium under a Wright-Fisher model.

The average nucleotide diversity at nonsynonymous sites was  $\pi_n = 0.0025$  (range, 0–0.014) and was lower than the Watterson's theta, which was  $\theta_n = 0.0035$ . This illustrates that nonsynonymous polymorphisms tend to segregate at low frequency. The nonsynonymous to silent diversity ratios were  $\pi_n/\pi_{si} = 0.38$  and  $\theta_n/\theta_{si} = 0.56$ . To be conservative in our conclusions, and to display a genome-wide trend without the influence of specific loci that might cause downward  $\theta_{si}$  or upward  $\theta_n/\theta_{si}$  bias, we plotted the distributions of  $\theta_n$ ,  $\theta_{si}$ , and  $\theta_n/\theta_{si}$  across loci (Figure 1). Two loci with a high  $\theta_n$  (JN680851 and JN680855) were removed from further analysis as they could have artificially increased our estimates of the genetic load, although the diversity estimates were virtually unchanged ( $\pi_n = 0.0025$ ;  $\pi_{si} = 0.0071$ ;  $\pi_n/\pi_{si} = 0.36$ ;  $\theta_n = 0.0034$ ;  $\theta_{si} = 0.0066$ ;  $\theta_n/\theta_{si} = 0.52$ ). One locus showed a surprisingly high silent diversity. This locus was also the only one exhibiting a significant positive Tajima's D, which would suggest it could be under balancing selection or that we might have amplified paralogous sequences. Including this locus in the analysis makes our conclusion of a low silent diversity conservative.

Figure 1 suggests that if we have a biased representation of the genetic diversity of the flat oyster genome, it would be caused by global bias in the choice of the loci analyzed. To check for such a bias, we verified whether the results were affected by the functional categories of the gene surveyed and by codon usage, which can be used as a proxy for expression levels as well as selective constraints on the protein (Stoletzki *et al.* 2005; Bierne and Eyre-Walker 2006; Sauvage *et al.* 2007). Genes were clustered into eight categories according to their putative biological function predicted by NCBI according to Gene Ontology (Figure S1). The distribution of  $\theta_n/\theta_{si}$  across the eight functional categories is presented in supplementary Figure S2. A similarly high  $\theta_n/\theta_{si}$  was estimated for every functional group, including ribosomal proteins, which are known to be highly constrained in other species. Representativeness of the genome diversity was also checked by computing GC3 and ENC on our set of 37 loci and on the full EST database (Figure S3). Codon bias proved to be low (high value of ENC) and GC3 moderately low. The distribution of the two measures was not significantly different between the two datasets. We also tested whether codon biases were correlated with diversity indices. Neither  $\theta_s$  nor  $\theta_{nc}$  or  $\theta_n$  were significantly correlated with ENC ( $\theta_s$ :  $r = -0.12$ ,  $P = 0.49$ ,  $\theta_{nc}$ :  $r = -0.04$ ,  $P = 0.79$ ,  $\theta_n$ :  $r = 0.13$ ,  $P = 0.45$ ). Finally, we verified if some errors in the annotation of the ORFs could have biased the results by removing every loci with an annotated UTR, and the results were not profoundly modified ( $\pi_n = 0.0023$ ,  $\theta_n = 0.003$ ,  $\pi_n/\pi_{si} = 0.31$ , and  $\theta_n/\theta_{si} = 0.48$ ).

Tajima's D was computed with silent mutations for each polymorphic locus. Negative values were obtained for 24 loci (ranged from  $-2.01$  to  $-0.15$ ) and positive values for 10 loci (ranged from 0.12 to 2.43). Among these values, none were statistically different from zero after a correction for multiple testing. The average Tajima's D was  $D = -0.38$  with the 37 loci dataset and  $D = -0.5$  with the 35 loci dataset.

Distribution of  $F_{ST}$  values for nonsynonymous and silent SNPs between all populations, and between Mediterranean and Atlantic populations, are presented on Figure S4. Although a large variance is observed among loci due to the small size of our samples, this analysis shows that the genetic differentiation was low overall and did not differ between the two categories of mutations.



**Figure 1** Distribution of Watterson's theta (A) at nonsynonymous sites ( $\theta_n$ ) and (B) at silent sites ( $\theta_{si}$ ) and (C) distribution of the non-synonymous-to-silent diversity ratio ( $\theta_n/\theta_{si}$ ) across the 37 loci.

### Selective constraint and the load of segregating non-neutral mutations

Figure 2A shows the  $\theta_n/\theta_{si}$  ratio for various frequency categories of SNPs, and Figure 2B shows the allele frequency spectrum of nonsynonymous and silent mutations, as well as the expectations of the mutation/drift equilibrium under the Wright-Fisher model. This shows that: (1) the frequency spectrum of silent SNPs did not depart from the neutral expectation at mutation/drift equilibrium under the Wright-Fisher model; (2) there is an excess of nonsynonymous mutations segregating at low frequency, suggesting that a large fraction of the amino acid changing SNPs are slightly selected; (3) the  $\theta_n/\theta_{si}$  ratio decreases when rare variants or when population-specific SNPs are removed, and it converges at a value of 23%, which can be used as an estimate of  $f_0$  (Fay *et al.* 2001), the fraction of amino acid-changing mutations that behave as neutral (or at least as silent mutations do). We then estimated that 44% of amino acid mutations are strongly deleterious and not

found in the sample, and 33% are slightly selected and segregate at low frequency.

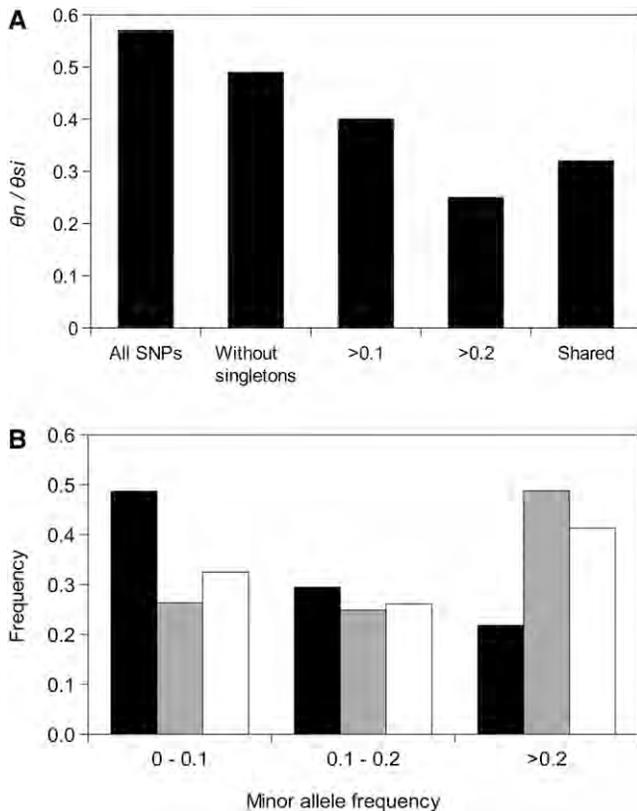
We observed 68 nonsynonymous SNPs in the sample (35 loci dataset) whereas 110 would have been expected from the number of silent SNPs if nonsynonymous mutations were all neutral. From the calculation above, a fraction  $f_0 = 0.23$  of nonsynonymous mutations are estimated to be neutral, which allowed us to infer that 25 of the 68 observed nonsynonymous SNPs were expected to be neutral and the remaining 43 were expected to be non-neutral in our sample of 16 oysters. By assuming non-neutral nonsynonymous alleles are present only once in the sample (Fay *et al.* 2001), we could conservatively estimate (*i.e.*, underestimate) that the number of non-neutral non-synonymous heterozygote mutations carried by an individual in the 35 loci surveyed was 3.2. Assuming there are  $10^7$  nonsynonymous sites in the oyster genome, we inferred that one diploid genome can be expected to carry more than ~4800 non-neutral mutations (or 2400 non-neutral mutations per haploid genome). Assuming a genetic map of 575 cM (Lallias *et al.* 2007) we therefore deduced the density of non-neutral mutations in the oyster genome to be  $4.2 \text{ cM}^{-1}$ .

### DISCUSSION

The explanation for the extreme protein heterozygosity of marine bivalves has been hotly debated in the past, at the heart of the neutralist–selectionist controversy, but the question has never been answered definitively. With the switch to DNA based molecular techniques to monitor genetic diversity, the time-lag to collect within-species genomic data in nonmodel organisms has led interest on the issue to dissipate. The simplest explanation—that a high diversity is the result of a large population size—has persisted as the valid interpretation. This was indeed the most logical explanation for the extreme diversity observed in *Ciona* sea squirts (Small *et al.* 2007; Tsagkogeorga *et al.* 2012), another genus of ocean-dwelling broadcast spawners. As a high density of SNPs has been observed in the first marine bivalves studied (Sauvage *et al.* 2007; Li *et al.* 2009; Zhang and Guo 2010; Kim *et al.* 2011), the same explanation was proposed (Sauvage *et al.* 2007), but perhaps too hastily.

#### Moderate silent diversity in a marine bivalve

In the present study, the density of SNPs was comparable with those observed in other bivalves (1 SNP every 50 bp). However, we were able to obtain estimates of nucleotide diversity to use in a rigorous comparison with published results on other species. The silent nucleotide diversity estimated in *Ostrea edulis* ( $\pi_{si} = 0.007$ ) was not extreme and appears rather moderate when compared with the champions of DNA polymorphism such as some *Caenorhabditis* nematodes [*C. remanei*:  $\pi_s = 0.045$  (Cutter *et al.* 2006), *Caenorhabditis* sp. 5:  $\pi_s = 0.068$  (Wang *et al.* 2010)] or *Ciona* sea squirts [*C. savignyi*:  $\pi_s = 0.08$  (Small *et al.* 2007) *C. intestinalis* sp. B:  $\pi_s = 0.057$  (Tsagkogeorga *et al.* 2012)]. The simplest explanation is that the effective population size of *O. edulis* is not as large as might have been expected, which could be due to an intrinsically low effective size that remained roughly constant for a long period, or the consequence of a departure from demographic equilibrium (*i.e.*, a population bottleneck). At first sight, the absence of departure from mutation/drift equilibrium under the Wright-Fisher model (Figure 2B) would suggest the former hypothesis to be more supported and that the moderately low diversity of *O. edulis* cannot be simply attributed to a demographic effect. However, one might suspect that the assumptions of the Wright-Fisher model do not apply to highly fecund marine organisms (Eldon and Wakeley 2006, 2009; Sargsyan and Wakeley 2008). Non-Wright-Fisher models that consider multifurcating genealogical processes with or without simultaneous multiple mergers (Eldon and Wakeley 2006, 2009; Sargsyan



**Figure 2** (A) Nonsynonymous-to-silent diversity ratio ( $\theta_n/\theta_{si}$ ) in the *O. edulis* nuclear genome. All SNPs: full dataset; without singletons: when singletons are removed from the dataset; >0.1: SNPs with a minor variant frequency greater 0.1; >0.2: SNPs with a minor variant frequency greater than 0.2.; shared: SNPs shared between at least two populations from the Atlantic or the Mediterranean Sea. (B) Allele frequency spectrum of nonsynonymous (dark), silent (gray), and expected under the neutral Wright-Fisher model (light). 0–0.1: SNPs with a minor variant frequency below 0.1; 0.1–0.2: SNPs with a minor variant frequency between 0.1 and 0.2; >0.2: SNPs with a minor variant frequency greater than 0.2. 35 loci dataset.

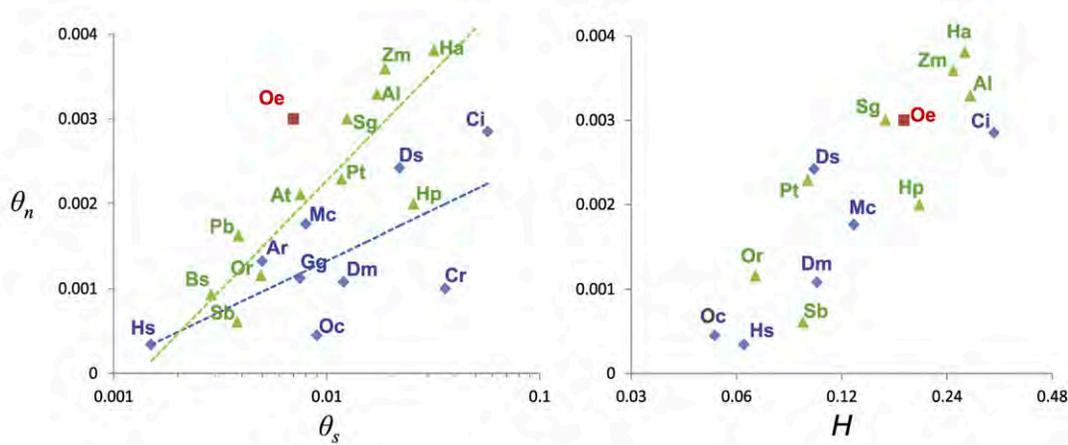
and Wakeley 2008) might be more relevant to the *Ostrea* system, for which the hypothesis of sweepstakes reproductive success has already received support (Hedgecock *et al.* 2007). Under sweepstakes reproductive processes, Tajima's D is expected to be negative at steady state (Sargsyan and Wakeley 2008). The fact that we observed it to be close to zero might therefore suggest that the population indeed departs from equilibrium. Another possibility could be that genetic differentiation among the geographically distant sampling locations we used could have contributed to affect the allele frequency spectrum (Städler *et al.* 2009; Cutter *et al.* 2012). However, *O. edulis* is known to be only slightly substructured, with an isolation-by-distance pattern (Launey *et al.* 2001). Not only we did not find evidence for population subdivision in our data but, under the hypothesis that there was indeed subdivision, the average silent diversity would have been inflated, while we found it to be low. Finally, selection on codon usage and functional noncoding DNA could have contributed to reducing silent diversity. We observed a similar level of diversity with synonymous and noncoding mutations, although synonymous diversity is usually lower than noncoding diversity in species in which synonymous selection is active (Zeng and Charlesworth 2010). Furthermore, we did not detect any correlation between diversity indices and codon bias. These results suggest that selection for

codon usage is not very effective in *O. edulis* and corroborates the hypothesis of a low  $N_e$  on the long term. To summarize, *O. edulis* is not extremely polymorphic at silent sites and this is probably a consequence of its effective population size being smaller than expected from census size and previous measures of allozyme variability. Alternative hypotheses cannot be definitively ruled out, however, but most of them are expected to equally affect silent and nonsynonymous mutations.

### A high load of non-neutral amino acid polymorphisms

Although silent diversity was not extreme in *O. edulis*, a high level of nonsynonymous diversity was nonetheless observed ( $\pi_n = 0.0024$ ). High amino acid diversity is not restricted to a few extremely polymorphic proteins but appeared to be more a general feature of the whole collection of genes analyzed (Figure 1B). It is worth emphasizing that none of the examples of protein diversity previously reported to be extremely high (Moy *et al.* 2008; Wharam *et al.* 2008; Costa *et al.* 2009; Parisi *et al.* 2009) would appear as outliers if compared to the distribution we obtained in Figure 1B. Using the meta-analysis of Gossmann *et al.* (2010) in plants and some published results on nuclear DNA in animals (Fay *et al.* 2001; Bierne and Eyre-Walker 2004; Cutter 2008; Axelsson and Ellegren 2009; Halligan *et al.* 2010; Carneiro *et al.* 2012; Gagnaire *et al.* 2012; Tsagkogeorga *et al.* 2012), we computed  $\theta_s$  and  $\theta_n$  for a variety of species, shown in Figure 3.  $\theta_n$  seems greater than what could have been expected from silent diversity in *O. edulis* (Figure 3). On the other hand, this result supports high protein heterozygosity in marine bivalves. From the literature, we obtained the average allozyme heterozygosity ( $H$ ) of all the possible species of Figure 3 (Schmidtke and Engel 1980; O'Brien *et al.* 1983; Nevo *et al.* 1984; Singh and Rhomberg 1987; Morden *et al.* 1989; Rajora and Dancik 1992; Saavedra *et al.* 1995; Weller *et al.* 1996; Cronn *et al.* 1997; Awasthi *et al.* 1998; Gao *et al.* 2000; Ansell *et al.* 2010) and plotted  $\theta_n$  against  $H$ . This time the flat oyster estimates fall right where they should be, in the group of species with the highest  $\theta_n$  and  $H$  values (Figure 3). Therefore, a discrepancy between silent and amino-acid diversities seems to exist, as already noted by Sauvage *et al.* (2007) in the cupped oyster *Crassostrea gigas*. As a consequence, the synonymous to nonsynonymous diversity ratio ( $\theta_n/\theta_s$ ) observed in *O. edulis* was very high ( $\theta_n/\theta_s = 0.56$ ), higher than any value we have seen in the literature to date. The explanation could be that the selective constraint is low (many amino acid mutations are neutral) and/or that there is a high segregating load in this species (many slightly selected mutations segregate in the populations). Partitioning nonsynonymous and silent SNPs according to their frequency, we estimated that 23% of amino acid changing mutations behave as neutral in *O. edulis*, which is similar to the estimates obtained in humans and the common fruit fly [20% and 24%, respectively (Fay *et al.* 2001; Shapiro *et al.* 2007)]. However, as much as 33% of nonsynonymous mutations are sufficiently weakly selected to segregate at low frequency in the polymorphism, which is much greater than usually reported. For instance, only 20% and 17% of amino acid changing mutations have been estimated to be mildly deleterious mutations in humans and flies, respectively. The high protein diversity of *O. edulis* is therefore partly due to a high load of segregating, weakly selected mutations. These mutations are often assumed to be slightly deleterious (Fay *et al.* 2001; Eyre-Walker *et al.* 2006), but any type of selection capable of generating an excess of low frequency amino acid variants can be invoked, such as balancing, weakly positive, or local selection, which may be more widespread than generally thought (Mitton 1997). Genetic incompatibilities (*i.e.*, negative epistatic interactions) are also often forgotten in population genomics (Fay 2011), even though they can accumulate in

**Oe** *Ostrea edulis*  
**Ar** *Anguilla rostrata*  
**Cr** *Caenorhabditis remanei*  
**Ci** *Ciona intestinalis*  
**Dm** *Drosophila melanogaster*  
**Ds** *Drosophila simulans*  
**Gg** *Gallus gallus*  
**Hs** *Homo sapiens*  
**Mc** *Mus castaneus*  
**Oc** *Oryctolagus cuniculus*  
**Al** *Arabidopsis lyrata*  
**At** *Arabidopsis thaliana*  
**Bs** *Boechea stricta*  
**Ha** *Helianthus annuus*  
**Hp** *Helianthus petiolaris*  
**Or** *Oryza rufipogon*  
**Pt** *Populus balsamifera*  
**Pb** *Populus tremula*  
**Sb** *Sorghum bicolor*  
**Sg** *Schiedea globosa*  
**Zm** *Zea mays*



**Figure 3** Nonsynonymous diversity ( $\theta_n$ ) in plant (green points) and animal species (red and blue points) as a function of synonymous diversity ( $\theta_s$ , left panel) and as a function of allozyme heterozygosity ( $H$ , right panel). References used are listed in Table S4.

cryptic tension zones and contribute to the species-wide diversity (Bierne *et al.* 2011).

### How can we explain the discrepancy between protein and silent diversities?

Our results suggest that *O. edulis* might have greater protein diversity than expected from its effective size based on silent mutations. We can first suspect that the recent demography of *O. edulis* could be responsible for the high segregating load observed. When a population departs from demographic equilibrium, the allele frequency distribution of selected mutations is likely to be shifted more strongly than that of neutral mutations. Eyre-Walker *et al.* (2006) showed that this effect can be strong when the population is expanding while there is little effect in bottlenecked or admixed populations. The effect increases with the deviation, so that it is inevitably accompanied by a deviation of the synonymous frequency spectra (Eyre-Walker *et al.* 2006). We observed neither a departure from the Wright-Fisher model at equilibrium nor evidence of population subdivision in the *O. edulis* data. We conclude that demography should not explain the results obtained under the hypothesis of a Wright-Fisher functioning of *O. edulis* populations. If *O. edulis* really departs from demographic equilibrium, it would mean that it also departs from a Wright-Fisher model which is precisely the hypothesis we will propose hereafter.

The results obtained in *O. edulis* revitalize the hypothesis that the biological and ecological characteristics of marine invertebrates make their population genetics somewhat different from most other animal species, while making them more similar to these of highly fecund plants [see Williams' "elm-oyster" model (Williams 1975; Launey and Hedgecock 2001)]. Many marine invertebrates share the characteristics of some plants, like trees: they are abundant, relatively long-lived and, maybe most importantly, highly fecund. Interestingly, Figure 3 suggests that the slope of the correlation between  $\theta_s$  and  $\theta_n$  tends to be slightly greater in plants and *O. edulis* than in animals. In oysters, adults produce millions of small eggs that develop into planktonic larvae that suffer high mortalities (Type III survivorship). Marine invertebrates undergo a large variance in reproductive success and generally have only fractions of the genetic diversity expected from their census size [see Hedgecock's hypothesis of sweepstakes reproductive success (Hedgecock 1994; Hedgecock and Pudovkin 2011)]. In the same way that a large variance in reproductive success can impact the neutral coalescence process (Eldon and Wakeley 2006, 2009; Sargsyan and Wakeley 2008), it might also impact the behavior of

selected mutations (Der *et al.* 2012). This does not mean that our estimate of the segregating load is flawed, as it was free from specific population genetics, but that the strong decoupling between  $N_e$  and  $N$  in non-Wright-Fisher populations with skewed offspring distribution could affect non-neutral mutations in an unpredicted manner. Theory predicts that the load of segregating deleterious mutations can deviate from the expectation based on a single Wright-Fisher population and become less dependent on the population size in subdivided populations (Glémin *et al.* 2003), and this deviation can become even stronger when the strength of selection vary spatially (Roze 2012) because of variation in the stressfulness of the environment or in source-sink metapopulation systems (Ronce and Kirkpatrick 2001).

The probability of fixation of a favorable mutation has recently been shown to be considerably increased when offspring distribution is skewed (Der *et al.* 2012). One can also speculate that deleterious mutations could remain polymorphic for longer periods as they could sometimes enter the population at a high frequency, "sweepstake" in the progeny of a lucky winner and therefore taking more time to be purged than in a standard Wright-Fisher population. Investigating the effect of skewed offspring distribution on the evolution and diversity of deleterious mutations might be an interesting line for future research. In any case, one is left with the fact that high fecundity is a prerequisite for tolerating a high load of segregating selected mutations (Williams 1975). We should also emphasize here that the fecundity of *Ciona* sea squirts is two orders of magnitude lower than the fecundity of the flat oyster (Petersen and Svane 1995), which might explain the high  $\theta_s$  and low  $\theta_n/\theta_s$  ratio observed in these species. Quantifying the proportion of neutral and slightly selected amino acid mutations in other extremely fecund organisms will help to test the potential impact of fecundity and skewed offspring distribution on protein polymorphism and its relation with silent diversity.

### Comparing estimates of the genetic load obtained from molecular data and from segregation distortions in inbred progenies

The study of segregation distortion in inbred progenies has revealed a large number of strongly deleterious recessive mutations (~15 per genome on average) in the flat oyster *O. edulis* (Bierne *et al.* 1998) and cupped oyster *C. gigas* (Launey and Hedgecock 2001; Plough and Hedgecock 2011). In these experiments, a molecular marker typically maps ~40 cM of the genome. According to our estimate, ~170 non-neutral mutations are expected on this map length, although we have

no idea of their true effect on viability. It is possible that only one mutation has a strong effect on fitness while the other mutations only have a negligible effect at the scale of the one generation used in these lab experiments. However, it is also possible that, assuming a single viability QTL at a given chromosomal position in mapping experiments is an unrealistic assumption that should deserve further examination in the future. The broad genomic distribution of distorted ratios (Lallias *et al.* 2007; Plough and Hedgecock 2011) might suggest a greater density of deleterious alleles than estimated, each with a lesser effect on fitness. One interpretation often neglected in QTL mapping is that of an abundance of mutations with small effects rather than a single mutation with a large effect, although this is highly plausible (Rockman 2011). Indeed, simulation studies of the infinitesimal model have shown that chance spatial clustering of infinitesimals caused by nonuniform recombination rates and gene densities can easily be confounded with a small number of large effect QTL (Noor *et al.* 2001). Our molecular estimate of a high genomic density of non-neutral amino-acid alleles in the *O. edulis* genome can be added to a long list of observations about the genetics of bivalve mollusks that started with abundant reports of heterozygosity-fitness correlations (Zouros 1987; Szulkin *et al.* 2010) and segregation distortion in pair crosses (Foltz 1986; Launey and Hedgecock 2001), which suggest bivalve genomes are heavily loaded by slightly selected polymorphisms. The architecture of this high genetic load will need to be further characterized and understood in order to correctly interpret the population genetics and evolution of these animals.

## ACKNOWLEDGMENTS

We would like to thank Toni Gossmann for sharing his dataset, John Theodorou for supplying the sample from Greece, Luca Bargelloni and Pierre Boudry for the Italian samples, the French network Repamo (REseau PATHologie des MOLLusques) for the French samples, Helen McCombie-Boudry for corrections of the manuscript, and two reviewers for helpful comments. This research was funded by the European Transnational Seafare project (Sustainable and Environmentally Friendly Aquaculture for the Atlantic Region of Europe) and co-funded by Électricité Réseau Distribution France, the Atlantic Area Transnational Program (2007–2013) and Agence National de la Recherche (Hi-Flo project ANR-08-BLAN-0334-01). This is article 2013-002 of Institut des Sciences de l'Évolution de Montpellier.

## LITERATURE CITED

- Ansell, S. W., H. K. Stenøien, M. Grundmann, H. Schneider, A. Hemp *et al.*, 2010 Population structure and historical biogeography of European *Arabidopsis lyrata*. *Heredity* 105: 543–553.
- Arias, A., R. Freire, P. Boudry, S. Heurtebise, J. Mendez *et al.*, 2009 Single nucleotide polymorphism for population studies in the scallops *Aequipecten opercularis* and *Mimachlamys varia*. *Conserv. Genet.* 10: 1491–1495.
- Awasthi, M., K. V. Bhat, and R. K. Anand, 1998 Genetic heterogeneity in the Indian *Mus musculus*. *Biochem. Genet.* 36: 245–258.
- Axelsson, E., and H. Ellegren, 2009 Quantification of adaptive evolution of genes expressed in avian brain and the population size effect on the efficacy of selection. *Mol. Biol. Evol.* 26: 1073–1079.
- Bazin, E., S. Glémin, and N. Galtier, 2006 Population size does not influence mitochondrial genetic diversity in animals. *Science* 312: 570–572.
- Bierne, N., and A. Eyre-Walker, 2004 The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol. Biol. Evol.* 21: 1350–1360.
- Bierne, N., and A. Eyre-Walker, 2006 Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. *J. Evol. Biol.* 19: 1–11.
- Bierne, N., S. Launey, Y. Naciri-Graven, and F. Bonhomme, 1998 Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* 148: 1893–1906.
- Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David, 2011 The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20: 2044–2072.
- Cameiro, M., F. W. Albert, J. Melo-Ferreira, N. Galtier, P. Gayral *et al.*, 2012 Evidence for widespread positive and purifying selection across the European rabbit (*Oryctolagus cuniculus*) genome. *Mol. Biol. Evol.* 29: 1837–1849.
- Costa, M. M., S. Dios, J. Alonso-Gutierrez, A. Romero, B. Novoa *et al.*, 2009 Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). *Dev. Comp. Immunol.* 33: 162–170.
- Cronn, R., M. Brothers, K. Klier, P. K. Bretting, and J. F. Wendel, 1997 Allozyme variation in domesticated annual sunflower and its wild relatives. *Theor. Appl. Genet.* 95: 532–545.
- Cutter, A. D., 2008 Multilocus patterns of polymorphism and selection across the X chromosome of *Caenorhabditis remanei*. *Genetics* 178: 1661–1672.
- Cutter, A. D., S. E. Baird, and D. Charlesworth, 2006 High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics* 174: 901–913.
- Cutter, A. D., G.-X. Wang, H. Ai, and Y. Peng, 2012 Influence of finite-sites mutation, population subdivision and sampling schemes on patterns of nucleotide polymorphism for species with molecular hyperdiversity. *Mol. Ecol.* 21: 1345–1359.
- Der, R., C. Epstein, and J. B. Plotkin, 2012 Dynamics of neutral and selected alleles when the offspring distribution is skewed. *Genetics* 191: 1331–1344.
- Eldon, B., and J. Wakeley, 2006 Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* 172: 2621–2633.
- Eldon, B., and J. Wakeley, 2009 Coalescence times and F(ST) under a skewed offspring distribution among individuals in a population. *Genetics* 181: 615–629.
- Excoffier, L., G. Laval, and S. Schneider, 2005 Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47–50.
- Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173: 891–900.
- Fay, J. C., 2011 Weighing the evidence for adaptation at the molecular level. *Trends Genet.* 27: 343–349.
- Fay, J. C., G. J. Wyckoff, and C. I. Wu, 2001 Positive and negative selection on the human genome. *Genetics* 158: 1227–1234.
- Foltz, D. W., 1986 Segregation and linkage studies of allozyme loci in pair crosses of the oyster *Crassostrea virginica*. *Biochem. Genet.* 24: 941–956.
- Gagnaire, P.-A., E. Normandeau, and L. Bernatchez, 2012 Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. *Mol. Biol. Evol.* 29: 2909–2919.
- Gao, L.-Z., B. A. Schaal, C. H. Zhang, J. Z. Jia, and Y. S. Dong, 2000 Allozyme variation and population genetic structure of common wild rice *Oryza rufipogon* Griff. in China. *Theor. Appl. Genet.* 101: 494–502.
- Glémin, S., J. Ronfort, and T. Bataillon, 2003 Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics* 165: 2193–2212.
- Gossmann, T. I., B. H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon *et al.*, 2010 Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* 27: 1822–1832.
- Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
- Halligan, D. L., F. Oliver, A. Eyre-Walker, B. Harr, and P. D. Keightley, 2010 Evidence for pervasive adaptive protein evolution in wild mice. *PLoS Genet.* 6: e1000825.
- Hedgecock, D., 1994 Does variance in reproductive success limit effective population sizes of marine organisms? pp. 122–134 in *Genetics and Evolution of Aquatic Organisms*, edited by A. R. Beaumont. Chapman & Hall, London.
- Hedgecock, D., and A. I. Pudovkin, 2011 Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull. Mar. Sci.* 87: 971–1002.

- Hedgecock, D., S. Launey, A. I. Pudovkin, Y. Naciri, S. Lapègue *et al.*, 2007 Small effective number of parents (N-b) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. *Mar. Biol.* 150: 1173–1182.
- Karl, S. A., and J. C. Avise, 1992 Balancing selection at allozyme loci in oysters - implications from nuclear RFLPs. *Science* 256: 100–102.
- Kim, W. J., H. Jung, and P. M. Gaffney, 2011 Development of type I genetic markers from expressed sequence tags in highly polymorphic species. *Mar. Biotechnol.* (NY) 13: 127–132.
- Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.
- Koehn, R. K., 1990 Heterozygosity and growth in marine bivalves—comments on the paper by Zouros, Romero-Dorey, and Mallet (1988). *Evolution* 44: 213–216.
- Koehn, R. K., and S. E. Shumway, 1982 A genetic/physiological explanation for differential growth-rate among individuals of the American Oyster, *Crassostrea virginica* (Gmelin). *Mar. Biol. Lett.* 3: 35–42.
- Lallias, D., A. R. Beaumont, C. S. Haley, P. Boudry, S. Heurtebise *et al.*, 2007 A first-generation genetic linkage map of the European flat oyster *Ostrea edulis* (L.) based on AFLP and microsatellite markers. *Anim. Genet.* 38: 560–568.
- Launey, S., and D. Hedgecock, 2001 High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159: 255–265.
- Launey, S., M. Barre, A. Gérard, and Y. Naciri-Graven, 2001 Population bottleneck and effective size in *Bonamia ostreae*-resistant populations of *Ostrea edulis* as inferred by microsatellite markers. *Genet. Res.* 78: 259–270.
- Li, R. H., Q. Li, and L. F. Kong, 2009 Characterization of expressed sequence tag-derived single-nucleotide polymorphisms in the bay scallop *Argopecten irradians irradians*. *Fish. Sci.* 75: 1389–1400.
- Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Mitton, J. B., 1993 Theory and data pertinent to the relationship between heterozygosity and fitness, pp. 17–41 in *The Natural History of Inbreeding and Outbreeding—Theoretical and Empirical Perspectives*, edited by N. W. Thornhill. University of Chicago Press, Chicago.
- Mitton, J. B., 1997 *Selection in Natural Populations*, Oxford University Press, New York.
- Morden, C. W., J. F. Doebley, and K. F. Schertz, 1989 Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). *Am. J. Bot.* 76: 247–255.
- Morga, B., I. Arzul, N. Faury, A. Segarra, B. Chollet *et al.*, 2011 Molecular responses of *Ostrea edulis* haemocytes to an in vitro infection with *Bonamia ostreae*. *Dev. Comp. Immunol.* 35: 323–333.
- Morga, B., T. Renault, and I. Arzul, 2012 New insights in flat oyster *Ostrea edulis* resistance against the parasite *Bonamia ostreae*. *Fish Shellfish Immunol.* 32: 958–968.
- Moy, G. W., S. A. Springer, S. L. Adams, W. J. Swanson, and V. D. Vacquier, 2008 Extraordinary intraspecific diversity in oyster sperm bindin. *Proc. Natl. Acad. Sci. USA* 105: 1993–1998.
- Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nevo, E., A. Beiles, and R. Ben-Shlomo, 1984 The evolutionary significance of genetic diversity: Ecological, demographic and life history correlates, pp. 13–213 in *Evolutionary Dynamics of Genetic Diversity. Lecture Notes in Biomathematics*, Vol. 53, edited by G. S. Mani. Springer-Verlag, Berlin.
- Noor, M. A., A. L. Cunningham, and J. C. Larkin, 2001 Consequences of recombination rate variation on quantitative trait locus mapping studies. Simulations based on the *Drosophila melanogaster* genome. *Genetics* 159: 581–588.
- O'Brien, S. J., D. E. Wildt, D. Goldman, C. R. Merrill, and M. Bush, 1983 The cheetah is depauperate in genetic variation. *Science* 221: 459–462.
- Parisi, M.-G., H. Li, M. N. Toubiana, N. Parrinello, M. Cammarata *et al.*, 2009 Polymorphism of mytilin B mRNA is not translated into mature peptide. *Mol. Immunol.* 46: 384–392.
- Petersen, J. K., and I. Svane, 1995 Larval dispersal in the ascidian *Ciona intestinalis* (l): evidence for a closed population. *J. Exp. Mar. Biol. Ecol.* 186: 89–102.
- Plough, L. V., and D. Hedgecock, 2011 Quantitative trait locus analysis of stage-specific inbreeding depression in the Pacific oyster *Crassostrea gigas*. *Genetics* 189: 1473–1486.
- Rajora, O. P., and B. P. Dancik, 1992 Genetic characterization and relationships of *Populus alba*, *P. tremula*, and *P. x canadensis*, and their clones. *Theor. Appl. Genet.* 84: 291–298.
- Riginos, C., K. Sukhdeo, and C. W. Cunningham, 2002 Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Mol. Biol. Evol.* 19: 347–351.
- Rockman, M. V., 2011 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66: 1–17.
- Rombel, I. T., K. F. Sykes, S. Rayner, and S. A. Johnston, 2002 ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282: 33–41.
- Ronce, O., and M. Kirkpatrick, 2001 When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution* 55: 1520–1531.
- Roze, D., 2012 Spatial heterogeneity in the strength of selection against deleterious alleles and the mutation load. *Heredity* 109: 137–145.
- Rozen, S., and H. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365–386.
- Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice *et al.*, 2009 Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–945.
- Saavedra, C., C. Zapata, and G. Alvarez, 1995 Geographical patterns of variability at allozyme loci in the European oyster *Ostrea edulis*. *Mar. Biol.* 122: 95–104.
- Sargsyan, O., and J. Wakeley, 2008 A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor. Popul. Biol.* 74: 104–114.
- Sauvage, C., N. Bierne, S. Lapègue, and P. Boudry, 2007 Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* 406: 13–22.
- Schmidtke, J., and W. Engel, 1980 Gene diversity in tunicate populations. *Biochem. Genet.* 18: 503–508.
- Shapiro, J. A., W. Huang, C. Zhang, M. J. Hubisz, J. Lu *et al.*, 2007 Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. USA* 104: 2271–2276.
- Singh, R. S., and L. R. Rhomberg, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. *Genetics* 117: 255–271.
- Small, K. S., M. Brudno, M. M. Hill, and A. Sidow, 2007 Extreme genomic variation in a natural population. *Proc. Natl. Acad. Sci. USA* 104: 5698–5703.
- Städler, T., B. Haubold, C. Merino, W. Stephan, and P. Pfaffelhuber, 2009 The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182: 205–216.
- Stoletzki, N., J. Welch, J. Hermisson, and A. Eyre-Walker, 2005 A dissection of volatility in yeast. *Mol. Biol. Evol.* 22: 2022–2026.
- Szulkin, M., N. Bierne, and P. David, 2010 Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* 64: 1202–1217.
- Tajima, F., 1989 Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Tsagkogeorga, G., V. Cahais, and N. Galtier, 2012 The population genomics of a fast evolver: high levels of diversity, functional constraint and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol. Evol.* 4:740–749.
- Wang, G.-X., S. Ren, Y. Ren, H. Ai, and A. D. Cutter, 2010 Extremely high molecular diversity within the East Asian nematode *Caenorhabditis* sp. 5. *Mol. Ecol.* 19: 5022–5029.
- Ward, R. D., L. J. English, D. J. McGoldrick, G. B. Maguire, J. A. Nell *et al.*, 2000 Genetic improvement of the Pacific oyster *Crassostrea gigas* (Thunberg) in Australia. *Aquacult. Res.* 31: 35–44.
- Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7: 256–276.
- Weller, S. G., A. K. Sakai, and C. Straub, 1996 Allozyme diversity and genetic identity in *Schideea* and *Alsinidendron* (Caryophyllaceae: *Alsinoidaeae*) in the Hawaiian Islands. *Evolution* 50: 23–34.
- Wharam, S. D., T. J. Wardill, V. Goddard, K. M. Donald, H. Parry *et al.*, 2008 A Leucine aminopeptidase gene of the Pacific Oyster *Crassostrea gigas* exhibits an unusually high level of sequence variation, predicted to affect structure, and hence activity, of the enzyme. *J. Shellfish Res.* 27: 1143–1154.

- Williams, G. C., 1975 *Sex and Evolution*. Princeton University Press, Princeton, NJ.
- Wright, S., 1951 The genetical structure of populations. *Ann. Eugen.* 15: 323–354.
- Zeng, K., and B. Charlesworth, 2010 Studying patterns of recent evolution at synonymous sites and intronic sites in *Drosophila melanogaster*. *J. Mol. Evol.* 70: 116–128.
- Zhang, L., and X. Guo, 2010 Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. *Aquaculture* 302: 124–129.
- Zouros, E., 1987 On the relation between heterozygosity and heterosis: an evaluation of the evidence from marine mollusks. *Isozymes Curr. Top. Biol. Med. Res.* 15: 255–270.

Communicating editor: J. M. Comeron