Gene
Volume 406, Issues 1-2, 30 December 2007, Pages 13-22
doi:10.1016/j.gene.2007.05.011

Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*

C. Sauvage[1], N. Bierne[2], S. Lapègue[1*] and P. Boudry[1]

[1] Laboratoire de Génétique et Pathologie - IFREMER - La Tremblade, France
[2] Département de Biologie Intégrative, Institut des Sciences de l'Evolution de Montpellier UMR 5554 CNRS-UMII, Station Méditerranéenne de l'Environnement Littoral, Sète, France

* Correspondance :
Dr Sylvie Lapègue, LGP, Station IFREMER, 17390 La Tremblade, France
Tél : +33 (0)5 46 76 26 30 ; Fax : +33 (0)5 46 76 26 11
email : slapegue@ifremer.fr

Abstract:

DNA sequence polymorphism and codon usage bias were investigated in a set of 41 nuclear loci in the Pacific oyster *Crassostrea gigas*. Our results revealed a very high level of DNA polymorphism in oysters, in the order of magnitude of the highest levels reported in animals to date. A total of 290 single nucleotide polymorphisms (SNPs) were detected, 76 of which being localised in exons and 214 in non-coding regions. Average density of SNPs was estimated to be one SNP every 60 bp in coding regions and one every 40 bp in non-coding regions. Non-synonymous substitutions contributed substantially to the polymorphism observed in coding regions. The non-synonymous to silent diversity ratio was 0.16 on average, which is fairly higher to the ratio reported in other invertebrate species recognised to display large population sizes. Therefore, purifying selection does not appear to be as strong as it could have been expected for a species with a large effective population size. The level of non-synonymous diversity varied greatly from one gene to another, in accordance with varying selective constraints. We examined codon usage bias and its relationship with DNA polymorphism. The table of optimal codons was deduced from the analysis of an EST dataset, using EST counts as a rough assessment of gene expression. As recently observed in some other taxa, we found a strong and significant negative relationship between codon bias and non-synonymous diversity suggesting correlated selective constraints on synonymous and non-synonymous substitutions. Codon bias as measured by the frequency of optimal codons for expression might therefore provide a useful indicator of the level of constraint upon proteins in the oyster genome.

Keywords: SNP; Genetic diversity; Codon bias; *Crassostrea gigas*

Introduction

Single Nucleotide Polymorphisms (SNPs) are the most abundant sequence variations encountered in a genome (Cho, Mindrinos et al., 1999; Picoult-Newberg, Ideker et al., 1999; Griffin and Smith, 2000). Although moderately sparse in the human genome with one SNP per kb (Sachidanandam, Weissman et al., 2001), they sometimes reach high densities in some highly diverse species such as in some insects (e.g. one SNP every 125 bp in the *Aedes* mosquito genome, Morlais and Severson, 2003) or some crops (e.g. one SNP every 104 bp in the maize genome, Tenaillon, Sawkins et al., 2001). With the development of DNA-based marker assays and high-throughput genotyping technologies, SNPs have become markers of choice for large scale mapping and genotyping (Rafalski, 2002; Black, Baer et al., 2001). For example, it has had a great impact on the generation of genetic maps, for the analysis of genetic diversity, trait mapping and diagnostics. They are especially useful for association studies because of their high frequency in the genome, and they are genetically more stable than microsatellite markers. SNPs are therefore ideally suited for the generation of high-density genetic maps (Cho, Mindrinos et al., 1999; Nairz, Stocker et al., 2002) and have number of advantages for population genetics studies (Vignal, Milan et al., 2002). However, only few SNP markers have been developed in marine bivalves and more generally in marine invertebrates.

Marine molluscs are recognised to reveal one of the highest level of allozyme polymorphism within the animal kingdom (average heterozygosity 15-30%, Ward, Skibinski et al., 1992; Bazin, Glemin et al. 2006). The extreme heterozygosity of marine molluscs is at first best explained by large effective population sizes expected for species with high fecundities, extensive larval dispersal, dense populations and broad distribution. However, numerous studies have questioned the neutrality of allozyme variation in marine molluscs. Popular examples of direct selection on some allozyme loci come from the marine bivalves literature (Koehn, Newell et al., 1980; Karl and Avise, 1992; Riginos, Sukhdeo et al., 2002), although they have sometimes been criticized (McDonald, 1996; Bierne, Daguin et al., 2003). The observation that multilocus heterozygosity is frequently correlated with fitness-related traits in these species (review in David, 1998) was initially taken as evidence for overdominance at allozymes (Mitton, 1993). However, a consensus now emerged that the correlation is in fact best explained by the indirect effect of deleterious mutations on neutral marker variations (David, 1998), this neutral alternative introduced two requisites that have a bearing on our understanding of the population genetics of marine bivalves: (i) a high genetic load and (ii) the existence of particular population structures generating a strong variance in individual inbreeding (Bierne, Tsitrone et al., 2000). Both have received support. The genetic load has been quantified and was estimated to be extremely high (Bierne, Launey et al., 1998; Launey and Hedgeock, 2001). Several authors have challenged the idea that marine bivalves occur in large, homogeneous, randomly mating populations. Hedgecock's sweepstakes hypothesis proposes that large variation in reproductive success lead to effective population sizes several orders of magnitude below census numbers (Hedgecock, 1994). The slight, unpredictable but significant genetic differentiation observed at small spatial scales (Johnson and Black, 1984; David, Perdieu et al., 1997) is also consistent with the fragmentation of marine populations into small, transient reproductive groups. Finally, Bazin (2006) reported that mitochondrial diversity was not higher in marine molluscs than in other taxonomic groups in sharp contrasts with allozyme diversity. Bazin (2006) also reported a positive correlation between allozyme diversity and nuclear DNA diversity at a large taxonomic scale, suggesting that these two types of variation share the same information, namely population size. However, their meta-analysis nonetheless revealed that quantification of nuclear DNA diversity remains uncommon in many invertebrate groups, including molluscs, precluding the use of DNA polymorphism for comparison with mitochondrial diversity. Taking into consideration the doubts accumulated on the neutral status of allozymes and on the effective population size in marine bivalves, one might enquire thorough assessments of silent DNA polymorphism in these taxa.

Here, we describe the characterization of SNPs in coding and non-coding sequences of the Pacific oyster *Crassostrea gigas*. We used primer sequences designed in a set of 41 ESTs and direct sequencing of PCR products. The validity of detected SNPs was

ascertained by multiple sampling of the same allele in a sample of related individuals. This sampling strategy allowed us to remove technical artefacts of PCR and sequencing that generate singleton substitutions. We describe levels of diversity observed in coding and non-coding regions. We also undertake an investigation of codon usage bias and its relationship with DNA polymorphism. It is now widely accepted that selection acts on synonymous codons to improve translation in many organisms (Ikemura, 1992; Akashi, 1994) but not always (Duret and Mouchiroud, 1999). Variation in the effectiveness of selection on synonymous codons between species is likely to reflect variation in effective population sizes rather than variation in coefficients of selection (Li, 1987; Bulmer, 1991; Akashi, 1995; Cutter, Baird et al., 2006). Indeed, genomes with the strongest variation in codon usage correspond to species recognised to have large effective population sizes such as bacteria, yeast or insects (Ikemura, 1982; Merkl, 2003; Akashi, 1995) while natural selection does not appear to play a role in Mammals (Urrutia and Hurst, 2001). Although selection on synonymous codon use has been known for more than two decades, it remains unclear whether codon usage primarily affects the elongation rate or the fidelity of protein synthesis (Akashi, 2001; Duret, 2002). However, the latter hypothesis has received evidences in *Drosophila melanogaster*, *Caenorhabditis elegans* and *Escherichia coli* (Akashi, 1994; Marais and Duret, 2001; Stoletzki and Eyre-Walker, 2007). A negative correlation between the rate of non-synonymous substitution and codon bias have been described in a number of species (Stoletzki and Eyre-Walker, 2007; Pal, Papp et al., 2001; Betancourt and Presgraves, 2002) and seems best explained by the effect of selection on the accuracy of translation (Bierne and Eyre-Walker, 2006; Stoletzki and Eyre-Walker, 2007). As a consequence, some authors have recently proposed that codon bias might sometimes be used as an indicator of the selective constraint acting on a protein (Stoletzki and Eyre-Walker, 2007; Plotkin, Dushoff et al., 2006) as others proposed to use codon bias as a measure of the level of gene expression (Sharp and Li, 1987; Karlin and Mrazek, 2000; Coghlan and Wolfe, 2000). We here provide results suggesting that codon bias could provide useful information about the level of constraint upon a gene in the oyster genome.

2. Material and Methods

2.1 Biological material and DNA extraction

Twenty four two year-old oysters (*Crassostrea gigas*) were used in this SNP discovery experiment. They were produced during the French MOREST program that studied the genetic basis of the summer mortality phenomenon as described in Degremont (2007).These 24 individuals belongs to the third stage of selection for the resistance to the summer mortality and were selected to initiate a genetic linkage mapping experiment to detect quantitative trait loci (QTL) of aquacultural interest. These oyster are related with each others as following (figure 1): there are two replicates of four groups composed of three individuals (one male and two females), which are brothers and sisters. They are related by a semi-brother relationship to individuals of the three other groups. All individuals that compose the eight groups are the progeny issued from a cross of two individuals randomly taken from eight pools (G2) of individuals, named A, O, G, J, K, Q, R and AC. These eight pools are themselves the progeny of an ancestral couple of oysters (G1) selected for their resistance to summer mortality phenomenon (Degremont, Ernande et al., 2007).

DNA was extracted from samples of mantle tissue following the Promega DNA Wizard cleanup Kit recommendations. The quality of DNA was first checked on a 1.5% agarose gel (15V/cm; 40 min) and secondly by quantification using an Eppendorf biospectrophotometer. This allowed to get 24 equal concentration of DNA (100µg/mL) prepared for the amplification step.

2.2 Primer Design and Amplification

Primers were designed using the Primer3 software (Rozen and Skaletsky, 2000) from Expressed Sequence Tags (EST) developed in the Pacific oyster and retrieved from the Genbank database (http://www.ncbi.nlm.nih.gov/). Few additional unpublished ESTs were

also added in our experiment (Huvet, Boulo and Renault, Pers. Com.). Primers produced fragments from 250 to 550 bp. PCR conditions were standardized by the use of a touchdown PCR protocol. The same reaction mix was used for each pair of primers and amplification was performed with three ranges of annealing temperatures (60 to 50°c, 65 to 55°c and 82 to 74°c). The mix was composed for each sample of 0.3U *Taq* polymerase (New England Biolabs), 10mM of provided buffer, 1mM $MgCl_2$, 2mM of dNTP (Eurogentec), 10µM of each primer and 100ng of genomic DNA, in a final reaction volume of 48µL. Pairs of primers with the same optimal range of annealing temperature were grouped and ran on a Perkin-Elmer ABI 2700 PCR machine (Applied Biosystems) as following: initial denaturation step for 5 min at 94°C, then 2 cycles of 1 min denaturation at 94°c, for every subsequent 2 cycles, the annealing temperature was decreased by 1 degree Celsius during 1 min, and an extension step at 72°c of 1 min. At the lowest annealing temperature (50, 55 or 74°c), 25 cycles at 94°c for 30s, 30s annealing and 1 min extension at 72°c were applied. A quantity of 5µL of PCR products was purified with 2µL of the ExoSAP-IT enzyme (Amersham Biosciences) to remove non-incorporated dNTPs and primers according to the manufacturer manual. Then, purified PCR products were sequenced in a single direction using the foward primer with the ABI Prism BigDye v3 Terminator Cycle sequencing Kit (Applied Biosystems) and the sequences were analysed on an ABI 3100 *Avant* genetic analyser (Applied Biosystems).

2.3 Nucleotide diversity

Sequences were edited, corrected by hand if needed and aligned using Seqscape v2.1 software (Applied Biosystems) using the KB basecaller algorithm. After a Blast homology search (http://www.ncbi.nlm.nih.gov/BLAST/), the sequences that did not correspond, at least partly to the EST used to design primers, were removed. Multiple sequences alignment was exported to BioEdit v7.0.5. (Hall, 1997) and aligned with the corresponding EST sequence in order to infer intronic regions. Open Reading Frames (ORF) were inferred by trying the three frames and finding the longest ORF together with the help of orthologous sequence from GenBank. Regions outside the ORF but matching the EST sequence were considered to be UTRs. However, UTRs were found for only 4 loci (Table 1) and the majority of non-coding DNA sequences described in the present study belongs to introns. We expected to find identical alleles by descent (i.e. the same sequence) several times because individuals of the sample are related. As a consequence, singleton substitutions were inferred to be technical artefacts arisen during PCR and sequencing steps. Although, SNPs were characterised by direct sequencing of PCR products, this sampling strategy allowed us to obtain a valid characterisation of these markers. SNPs were identified as transitions or transvertions for both coding and non-coding regions. For SNPs occurring in coding sequences, variations were classified as synonymous or non-synonymous changes. The analysis of genetic diversity was conducted with DnaSP 4.0 (Rozas and Rozas, 1995). We computed the average number of SNPs per site (Π) which is the number of polymorphic sites divided by the length of the sequence. Calculations were conducted independently for non-coding (Π_{nc}), synonymous (Π_s), and non synonymous (Π_{ns}) substitutions. In order to compute Π_s and Π_{ns}, we estimated the number of synonymous and non-synonymous sites in a coding sequence with the method of Nei and Gojobori (1986).

2.4 Synonymous codon use

In order to investigate codon usage bias, the table of optimal codons was deduced from the analysis of an EST dataset using EST counts as a rough assessment of gene expression (Duret and Mouchiroud, 1999). The dataset was composed of 8800 EST sequences which were available for *C. gigas* in Genbank (http://www.ncbi.nlm.nih.gov/) and the Marine Genomics Europe Network of Excellence databases. Open reading frames (ORFs) were identified by choosing the longest possible translation into amino acid sequence. Sequences with an ORF smaller than 100 codons (300bp) were removed from the dataset to prevent the occurrence of wrong ORFs and to have enough codons to compute the frequency of synonymous codons. It is likely that some EST annotations are wrong as a result of artifactual frameshift substitutions in single-pass sequences. However, our procedure should result in removing a part of the coding region from the analysis rather than

misinferring UTRs as coding regions. We therefore argue that sequence quality problems could introduce a statistical noise but should not bias the analysis. We built clusters of ESTs corresponding to the same gene by using the algorithm developed by Bazin, Duret et al. (2005) to construct the Polymorphix database. Two sequences were clustered into the same family if they shared fragment longer than 300pb with similarity superior to 95%. Only the longest ORF of a cluster was used to compute tables of codon usage. The frequency of synonymous codon per amino acid was computed with the program CODONW (Peden, 1999). A synonymous codon was inferred as "optimal" when its frequency significantly increased with EST counts (Duret and Mouchiroud, 1999). This was tested by non-parametric Spearman's correlation in JMP v5.0© (SAS Institute Inc.). Once optimal codons were inferred, we used the frequency of optimal codons (Fop; Ikemura, 1985) as a measure of codon bias.

3. Results

3.1 Types and distribution of polymorphism

A total of 84 EST sequences were chosen for primer design. Fifteen of these ESTs were chosen on the basis of their putative function in relation to summer mortality. Forty one ESTs (51%) amplified a clear genomic DNA fragment. For 31 of the 41 amplifying loci (72%), the size of the fragment obtained was longer than the size expected from the EST sequence because of the presence of introns. Therefore, we suspect that the presence of several and/or very long introns could partly explain the rate of failure of PCR amplification.

Approximately, a total of 10.5 kb of the pacific oyster genome were sequenced with a total of 290 SNP markers identified in both coding and non-coding DNA. Around ¾ of SNPs (69%) were localised in non-coding regions (Table 1). The level and the repartition of this polymorphism in each locus is shown in the figure 2. We found only 4 monomorphic locus which represented only 5% of amplified loci. The average density of SNPs in coding regions was estimated to 1 every 60 bp and 1 every 40 bp in non-coding ones. An average number of 6.7 SNPs were detected in each amplicon with a minimum of 0 (monomorphic sequence) and a maximun of 30 SNPs within a 241 bp (1 SNP every 9bp) entirely non-coding sequence of the Integrin gene.

Our data show a global t_s/t_v ratio (1.3) in the Pacific oyster, which is similar to the ratio observed in *Drosophila* (1.5, Moriyama and Powell, 1996) or humans (1.4, Brookes, 1999). t_s/t_v was equal to 2.1 for synonymous changes while it was 0.9 and 1.2 for non-synonymous and non-coding changes respectively (exact *t* Fisher test, *t*=0.264). With twice more possibilities of transvertions than transitions, t_s/t_v should ideally be equal to ½, if substitutions appeared randomly in DNA. However, it is well-known that a mutational bias favours transition over transversion. Because of the structure of the genetic code, synonymous changes are more often transitions than transversions. As much of the diversity is composed of synonymous substitutions, the t_s/t_v ratio increases in coding sequences (t_s/t_v =1.34). The same phenomenon is observed in the mosquito *Aedes aegypti* where t_s/t_v in coding sequences is similar. (Morlais and Severson, 2003). The mutational bias for transition is therefore more accurately estimated from non-coding sequences and was estimated to be t_s/t_v =1.2 in *C. gigas*.

The number of SNPs per site was calculated for both coding and non-coding regions. Diversities were similar at synonymous (Πs=0.035) and non-coding positions (Πnc=0.038) and were not significantly different (per substitution *t* test = -0.55, P=0.29). Although the level of polymorphism in coding and non coding regions was positively correlated as expected for neutral substitutions sharing correlated genealogies, the correlation was not significant (Spearman's ρ= 0.3; P=0.16). As the two kinds of silent substitutions gave the same information, they have been combined into a single indice (Πsi). The non-synonymous to silent polymorphism ratio (Πn/Πsi) was 0.16 on average. However, Πn/Πsi varied greatly from one gene to another.

3.2 Codon Usage Bias

Table 2 presents Spearman's correlation coefficients between the frequency of synonymous codons per amino acid and EST counts. As an illustration, we explain the results obtained at the Phenylalanine amino acid. Phenylalanine is a two-fold degenerate codon, which means that two different codons (UUU and UUC) can code for this amino acid. The average frequency of this amino acid among all others is 0.05 in the Pacific oyster genome and it does not vary significantly with expression levels as measured by EST counts. Ideally, each of the two codons of the phenylalanine should be equally represented by a frequency of 0.5. However, as expected in a GC-poor genome, the UUU codon is over-represented on average in lowly expressed genes (~60% in the category of genes with a single EST). On the other hand, the frequency of the UUC codon increases with expression level and the situation is reversed in highly expressed genes for which the UUU codon is under-represented (~40% in the category of genes with 5 or more ESTs). The same trends were observed for other amino-acid: although -A and –U ending codons are more frequent on average as expected in a GC-poor genome, -G or -C ending optimal codons are more frequently observed (Table 2). As a consequence, GC-content at the third coding position is also correlated to expression levels (Spearman's ρ = 0.087, p<0.0001) although less than *Fop* (Spearman's ρ = 0.13, p<0.0001). As a consequence, codon usage is paradoxically less balanced in lowly expressed genes that are GC-poor than in highly expressed genes that reach an intermediate GC-content. This result highlights that codon usage bias is difficult to establish without an assessment of expression levels (Duret and Mouchiroud, 1999). In the case of the genome of *C. gigas*, a simple analysis of codon usage without expression data could have resulted in the inference that GC-poor genes are more biased and that -A and -U ending codons are optimal for translation while it is indeed the reverse –translational selection primarily favours -G or -C ending codons although rare in a GC-poor genome but the effectiveness of selection is not sufficient to result in a strong enrichment in GC-content which remains intermediate in biased genes. The results we obtained in *C. gigas* are similar to those recently reported in some GC-poor Nematode genomes in which translational selection nonetheless tends to favour -G or -C ending codons (Cutter, Baird et al., 2006).

We analysed whether genetic diversities would be correlated with codon bias. Synonymous diversity could have been expected to decrease with codon bias because of selection for codon usage. However, selection on synonymous substitutions is often too small to result in a detectable correlation (Bulmer, 1991; McVean and Charlesworth, 2000). In *Drosophila* for instance, although synonymous divergence decrease with codon bias (Sharp and Li, 1987) synonymous polymorphism does not vary with codon bias with a dataset size comparable to this study (Bierne and Eyre-Walker, 2006). On the other hand, we detected a significant correlation between non-synonymous diversity and *Fop* (Figure 3; Spearman's ρ= -0.47, P=0.005) or between Πn/Πsi and *Fop* (Spearman's ρ= -0.45, P=0.02).

4. Discussion

4.1 High Nucleotide diversity

The level of DNA polymorphism observed in this study is one of the highest ever observed to date, with one SNP every 40 bp in non-coding regions. To our knowledge, the highest levels of DNA polymorphism reported in the animal kingdom were found in the nematod *Caenorhabditis remanei* (Cutter, Baird et al., 2006) and the sea squirts *Ciona savignyi* (Small, Brudno et al., 2007) with one SNP every 20 bp. Insect species are also recognised to often be highly polymorphic. The density of SNPs was reported to be one every 50bp in *Drosophila* (Shapiro, Huang et al., 2007) and one every 125 bp in Mosquitoes (Morlais and Severson, 2003). Our results therefore confirm previous reports on allozyme diversities which revealed that marine invertebrates are amongst the most diverse animal species (Ward, Skibinski et al., 1992; Solé-Cava and Thorpe, 1991). Furthermore, DNA sequence polymorphism provides the opportunity to compare different categories of substitutions on which selection is expected to act differently. Synonymous and non-coding

(silent) substitutions that do not result in a change in the protein can be used as a neutral reference to infer the level of purifying selection that act on non-synonymous (replacement) substitutions. A theoretical expectation is that selection should more efficiently remove deleterious mutations in species with large population sizes (Kimura, 1983). The non-synonymous to silent polymorphism ratio ($\Pi n/\Pi si$) should therefore be inversely correlated to the effective population size. In addition, unlike divergence data, polymorphism data should not be affected by adaptive evolution (McDonald and Kreitman, 1991) and $\Pi n/\Pi si$ is expected to be a good measure of the proportion of non-synonymous substitutions that are effectively neutral (Smith and Eyre-Walker, 2002). We found that non-synonymous substitutions contributed substantially to the polymorphism observed in *C. gigas*: $\Pi n/\Pi si$=16%. This value is much lower than in humans (0.22, Bustamante, Fledel-Alon et al., 2005), but it is fairly higher to what has been reported in *Drosophila* (0.12, Shapiro, Huang et al., 2007) and much higher than in *Caenorhabditis remanei* (0.09, Cutter, Baird et al., 2006) and *Ciona savignyi* (0.07, Small, Brudno et al., 2007). Therefore, purifying selection does not appear to be as strong as it could have been expected for a species with a large effective population size. However, this result at the molecular level concurs with the observation of a high genetic load in this species (Launey and Hedgecock, 2001) and would require further attention. Unfortunately, without any data on the frequency of substitutions in natural populations, it is difficult to investigate further the potentially deleterious nature of segregating non-synonymous polymorphisms (e.g. Eyre-Walker, Woolfit et al., 2006).

Some polymorphisms that affect biological functions can occur outside of coding regions, in promoters or other regulating regions. Non-coding DNA contains important sequences for regulatory functions and can be constrained to some extent. Detectable levels of constraint on non-coding DNA have already been reported in *Drosophila* where non-coding diversity is lower than synonymous diversity (Moriyama and Powell, 1996); (Andolfatto, 2005). The non-coding sequences we studied were in the close vicinity of genes, mostly in introns. However, contrary to what was reported in *Drosophila* we observed a similar level of diversity between the two kinds of silent substitutions. We conclude that synonymous and non-coding substitutions are equally (un)constrained in the oyster genome.

Recombination might be an interesting parameter to investigate further in the future in order to reconcile the possible discrepancy observed between high diversity and relaxed purifying selection. Indeed, intriguing results have been reported in marine bivalves. Guo and Allen (1996) reported the existence of a single recombinational hot spot in the proximal region of each chromosome arm in *Mulinia lateralis*. Hubert and Hedgecock (2004) detected chromosomal rearrangements in *C. gigas* and suggested that blocks of the genome may retain linkage disequilibria for longer than might be expected in putatively large, well-mixed populations.

4.2 Pattern of Codon Usage Bias

The pattern of codon usage within a genome results from a balance among selection, random genetic drift and mutation (Bulmer, 1991). Our results in *Crassostrea gigas* first suggest that a mutational bias tends to enrich the genome in A and T because the genome is GC-poor on average (GC3=43%). Secondly, using a moderately large EST database (8800 ESTs) we nonetheless were able to easily detect synonymous codons that significantly increase with expression levels measured with EST counts (Duret and Mouchiroud, 1999). Ideally, one may verify that optimal codons correspond to the most abundant tRNAs in the cell (Ikemura, 1985). Unfortunately, tRNA concentrations in the cell or tRNA gene copy numbers are not available for *C. gigas* to date and we had to deal with expression levels alone. In addition, it should be noted here that expression levels inferred from EST data seem to be affected by gene length (Munoz, Bogarad et al., 2004). It can potentially be a problem because gene length and codon bias are correlated in a number of species, positively in Prokaryotes (Eyre-Walker, 1996) but negatively in Eukaryotes (Moriyama and Powell, 1998; Duret and Mouchiroud, 1999). Although puzzling the correlation is not

necessarily the result of a direct link of causality (Duret and Mouchiroud, 1999). It seems difficult to imagine that a gene length effect could bias the inference of optimal codons for two reasons: (i) Longer genes are more energetically costly to translate. It seems expected that they could be underrepresented in the category of highly expressed genes. The fact that expression levels measured with EST data miss the correlation does not imply that other variables such as codon bias cannot be studied with EST counts. (ii) Although a significant correlation between codon bias and gene length has been reported in some genomes, the correlation is always small, accounting for <5% of the total variance in codon bias (Moriyama and Powell, 1998).

Most of these codons that we inferred as being optimal for translation were -G or -C ending codons. As a consequence selection for translational efficiency tends to enrich highly expressed genes in G and C (GC3=49% for genes with five or more ESTs) when compared to lowly expressed genes (GC3=43% for genes with one EST). A difference of 6% in GC-content (or 5% in Fop) between highly and lowly expressed genes may be considered as small. However, EST count is a very rough assessment of expression level and the inherent statistical noise of an analysis of a moderately small EST dataset prevents from using this difference as a reliable estimate of selection intensity. Nonetheless, using a much larger EST database (114665 ESTs) in *Drosophila melanogaster* (data from Hey and Kliman, 2002), one may remark that genes with one EST are already GC-rich (GC3=64%) and that highly express genes are not tremendously richer in GC (GC3=70% for genes with more than 50 ESTs). Conversely, the simple fact that a significant effect is detected with a small EST dataset could have been taken as evidence that translational selection is strong in *C. gigas*.

We observed a strong and significant negative correlation between the frequency of optimal codons and non-synonymous diversity. Several hypotheses have been proposed to explain the correlation but an agreement recently emerged to propose that this is a consequence of correlated selective constraints on synonymous and non-synonymous substitutions (Bierne and Eyre-Walker, 2006; Stoletzki, Welch et al., 2005) possibly as a consequence of selection for translational robustness (Drummond, Bloom et al., 2005). Indeed, Akashi (1994), Marais and Duret (2004) and Stoletzki and Eyre Walker (2007) provided evidence for selection on translational accuracy in *Drosophila, C. elegans* and *E. coli* respectively. Codon usage bias in *C. gigas* could therefore be due, at least partly, to selection for translational accuracy: a synonymous codon tends to be replaced by a more preferred one to reduce the fitness cost of a misincorporation (minimize the costs of proofreading and/or the cost of producing incorrect proteins). However, the confirmation of translational selection for accuracy in *C. gigas* would require further tests (see Stoletzki and Eyre-Walker, 2007). Whatever the exact cause of the correlation, our results nonetheless provide evidence that codon bias as measured by the frequency of optimal codons for expression might be used as a useful indicator of the selective constraint acting on a protein in the oyster genome (Stoletzki, Welch et al., 2005).

Ackowledgements

The authors are very indebted to Erick Bazin for the procedure allowing clustering EST sequences in gene families and thank two anonymous referees for their constructive remarks. This work was partly funded by the European project Aquafirst (Combined genetic and functional genomic approaches for stress and disease resistance marker assisted selection in fish and shellfish, FP6), the Marine Genomics Europe Network of Excellence (FP6), the French national programme GIS “Génomique Marine” (The Bivalvomix Project, coord. N. Bierne), the Bureau des Ressources Génétiques (BRG) and the Region Poitou-Charentes (CPER). The authors would like to thank also Nicole Faury, Tristan Renault, Viviane Boulo and Arnaud Huvet who provided us some of the primer sequences used in this paper.

References

Akashi, H., 1994. Synonymous Codon Usage in *Drosophila melanogaster*: Natural Selection and Translational Accuracy. Genetics. 136, 927-935.

Akashi, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. Genetics. 139, 1067-76.

Akashi, H., 2001. Gene expression and molecular evolution. Curr Opin Genet Dev. 11, 660-6.

Andolfatto, P., 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature. 437, 1149-52.

Bazin, E., Duret, L., Penel, S. and Galtier, N., 2005. Polymorphix: a sequence polymorphism database. Nucl. Acids Res. 33, 481-484.

Bazin, E., Glemin, S. and Galtier, N., 2006. Population size does not influence mitochondrial genetic diversity in animals. Science. 312, 570-2.

Betancourt, A.J. and Presgraves, D.C., 2002. Linkage limits the power of natural selection in Drosophila. Proc Natl Acad Sci U S A. 99, 13616-20.

Bierne, N., Launey, S., Naciri-Graven, Y. and Bonhomme, F., 1998. Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. Genetics. 148, 1893-906.

Bierne, N., Tsitrone, A. and David, P., 2000. An inbreeding model of associative overdominance during a population bottleneck. Genetics. 155, 1981-90.

Bierne, N., Daguin, C., Bonhomme, F., David, P. and Borsa, P. (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a Mytilus hybrid zone. Mol Ecol 12(9): 2505-10.

Bierne, N. and Eyre-Walker, A., 2006. Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. J Evol Biol. 19, 1-11.

Black, W.C.t., Baer, C.F., Antolin, M.F. and DuTeau, N.M., 2001. Population genomics: genome-wide sampling of insect populations. Annu Rev Entomol. 46, 441-69.

Brookes, A.J., 1999. The essence of SNPs. Gene. 234, 177-86.

Bulmer, M., 1991. The Selection-Mutation-Drift Theory of Synonymous Codon Usage. Genetics. 129, 897-907.

Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M. and Clark, A.G., 2005. Natural selection on protein-coding genes in the human genome. Nature. 437, 1153-7.

Cho, R.J., et al. 1999 Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genetics. 23, 203-7.

Coghlan, A. and Wolfe, K.H., 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast. 16, 1131-45.

Cutter, A.D., Baird, S.E. and Charlesworth, D., 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics. 174, 901-13.

David, P., 1998. Heterozygosity-fitness correlations: new perspectives on old problems. Heredity. 80 (5), 531-7.

David, P., Perdieu, M.A., Pernot, A.F. and Jarne, P., 1997. Spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. Evolution. 51, 1318-1322.

Dégremont, L., Ernande, B., Bedier, E. and Boudry, P., 2007. Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. Aquaculture. 262, 41-53.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O. and Arnold, F.H., 2005. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 102, 14338-43.

Duret, L., 2002. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev. 12, 640-9.

Duret, L. and Mouchiroud, D., 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc Natl Acad Sci U S A. 96, 4482-7.

Eyre-Walker, A., 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? Mol Biol Evol. 13, 864-72.

Eyre-Walker, A., Woolfit, M. and Phelps, T., 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics. 173, 891-900.

Griffin, T.J. and Smith, L.M., 2000. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. Trends Biotechnol. 18, 77-84.

Guo, X. and Allen, S.K., Jr., 1996. Complete interference and nonrandom distribution of meiotic crossover in a mollusc, *Mulinia lateralis* (Say). Biol Bull. 191, 145-8.

Hall, T., 1997: Bioedit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Hedgecock, D., 1994: Does variance in reproductive success limit effective population sizes of marine organisms? In: Beaumont, A.R. (Ed.), Genetics and Evolution of Aquatic Organisms. Chapman and Hall, London, pp. 122-134.
Hey, J. and Kliman, R.M., 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. Genetics. 160, 595-608.
Hubert, S. and Hedgecock, D., 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics. 168, 351-62.
Ikemura, T., 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J Mol Biol. 158, 573-97.
Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 2, 13-34.
Ikemura, T., 1992: Correlation between codon usage and tRNA content in microorganisms, in Transfer RNA in Protein Synthesis. In: Hatfiel, D.L., Lee, B.J. and Pirtle, R.M. (Eds.). CRC, Boca Raton, FL, pp. 87-111.
Johnson, M.S. and Black, R., 1984. Pattern beneath chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. Evolution. 38, 1371-1383.
Karl, S.A. and Avise, J.C., 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science. 256, 100-2.
Karlin, S. and Mrazek, J., 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J Bacteriol. 182, 5238-50.
Kimura, M., 1983. The neutral theory of molecular evolution. New York, Cambridge University Press.
Koehn, R.K., Newell, R.I. and Immermann, F., 1980. Maintenance of an aminopeptidase allele frequency cline by natural selection. Proc Natl Acad Sci U S A. 77, 5385-9.
Launey, S. and Hedgecock, D., 2001. High genetic load in the Pacific oyster Crassostrea gigas. Genetics. 159, 255-65.
Li, W.H., 1987. Models of nearly neutral mutations with particular implications for non-random usage of synonymous codons. J Mol Evol. 24, 337-45.
Marais, G., Domazet-Lošo, T., Tautz, D. and Charlesworth, B., 2004. Correlated Evolution of Synonymous and Nonsynonymous Sites in *Drosophila*. Journal of Molecular Evolution. 59, 771-779.
Marais, G. and Duret, L., 2001. Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. J Mol Evol. 52, 275-80.
McDonald, J.H., 1996. Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol Biol Evol. 13, 253-60.
McDonald, J.H. and Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351, 652-654.
McVean, G.A.T. and Charlesworth, B., 2000. The Effects of Hill-Robertson Interference Between Weakly Selected Mutations on Patterns of Molecular Evolution and Variation. Genetics. 155, 929-944.
Merkl, R., 2003. A survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. J Mol Evol. 57, 453-66.
Mitton, J. B., 1993. Theory and data pertinent to the relationship between heterozygosity and fitness in The Natural History of Inbreeding and Outbreeding—Theoritical and Empirical Perspectives. N. W. THORNHILL. Chicago, University of Chicago Press. pp 17-41.
Moriyama, E.N. and Powell, J.R., 1996. Intraspecific nuclear DNA variation in *Drosophila*. Mol Biol Evol. 13, 261-77.
Moriyama, E.N. and Powell, J.R., 1998. Gene length and codon usage bias in *Drosophila melanogaster, Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. 26, 3188-93.
Morlais, I. and Severson, D.W., 2003. Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti*. Insect Mol Biol. 12, 631-639.
Munoz, E.T., Bogarad, L.D. and Deem, M.W., 2004. Microarray and EST database estimates of mRNA expression levels differ: the protein length versus expression curve for C. elegans. BMC Genomics. 5, 30.
Nairz, K., Stocker, H., Schindelholz, B. and Hafen, E., 2002. High-resolution SNP mapping by denaturing HPLC. Proc Natl Acad Sci U S A. 99, 10575-80.
Nei, M. and Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3, 418-26.
Pal, C., Papp, B. and Hurst, L.D., 2001. Highly expressed genes in yeast evolve slowly. Genetics. 158, 927-31.

Peden, J.F., 1999. CodonW: Analysis of Codon Usage, University of Nottingham.
Picoult-Newberg, L., Ideker, T.E., Pohl, M.G., Taylor, S.L., Donaldson, M.A., Nickerson, D.A. and Boyce-Jacino, M., 1999. Mining SNPs from EST databases. Genome Res. 9, 167-74.
Plotkin, J.B., Dushoff, J., Desai, M.M. and Fraser, H.B., 2006. Estimating selection pressures from limited comparative data. Mol Biol Evol. 23, 1457-9.
Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 5, 94-100.
Riginos, C., Sukhdeo, K. and Cunningham, C.W., 2002. Evidence for selection at multiple allozyme loci across a mussel hybrid zone. Mol Biol Evol. 19, 347-51.
Rozas, J. and Rozas, R., 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput Appl Biosci. 11, 621-5.
Rozen, S. and Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132, 365-86.
Sachidanandam, R., et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 409, 928-33.
Shapiro, J.A., Huang, W., Zhang, C., Hubisz, M.J., Lu, J., Turissini, D.A., Fang, S., Wang, H.Y., Hudson, R.R., Nielsen, R., Chen, Z. and Wu, C.I., 2007. Adaptive genic evolution in the *Drosophila* genomes. Proc Natl Acad Sci U S A. 104, 2271-6.
Sharp, P.M. and Li, W.H., 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol Biol Evol. 4, 222-30.
Small, K.S., Brudno, M., Hill, M.M. and Sidow, A., 2007. Extreme genomic variation in a natural population. Proc Natl Acad Sci U S A. 104, 5698-703.
Smith, N.G. and Eyre-Walker, A., 2002. Adaptive protein evolution in *Drosophila*. Nature. 415, 1022-4.
Solé-Cava, A.M. and Thorpe, J.P., 1991. High levels of genetic variation in natural populations of marine lower invertebrates. Biological Journal of the Linnean Society. 44, 65-80.
Stoletzki, N. and Eyre-Walker, A., 2007. Synonymous Codon Usage in *Escherichia coli*: Selection for Translational Accuracy. Mol Biol Evol. 24, 374-81.
Stoletzki, N., Welch, J., Hermisson, J. and Eyre-Walker, A., 2005. A Dissection of Volatility in Yeast. Mol Biol Evol. 22, 2022-2026.
Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F. and Gaut, B.S., 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays L*.). Proc Natl Acad Sci U S A. 98, 9161-6.
Urrutia, A.O. and Hurst, L.D., 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics. 159, 1191-9.
Vignal, A., Milan, D., SanCristobal, M. and Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution. 34, 275-305.
Ward, R.D., Skibinski, D.O.F. and Woodward, M., 1992. Protein heterozygosity, protein structure, and taxonomic differentiation. Evol Biol. 26, 73-159.

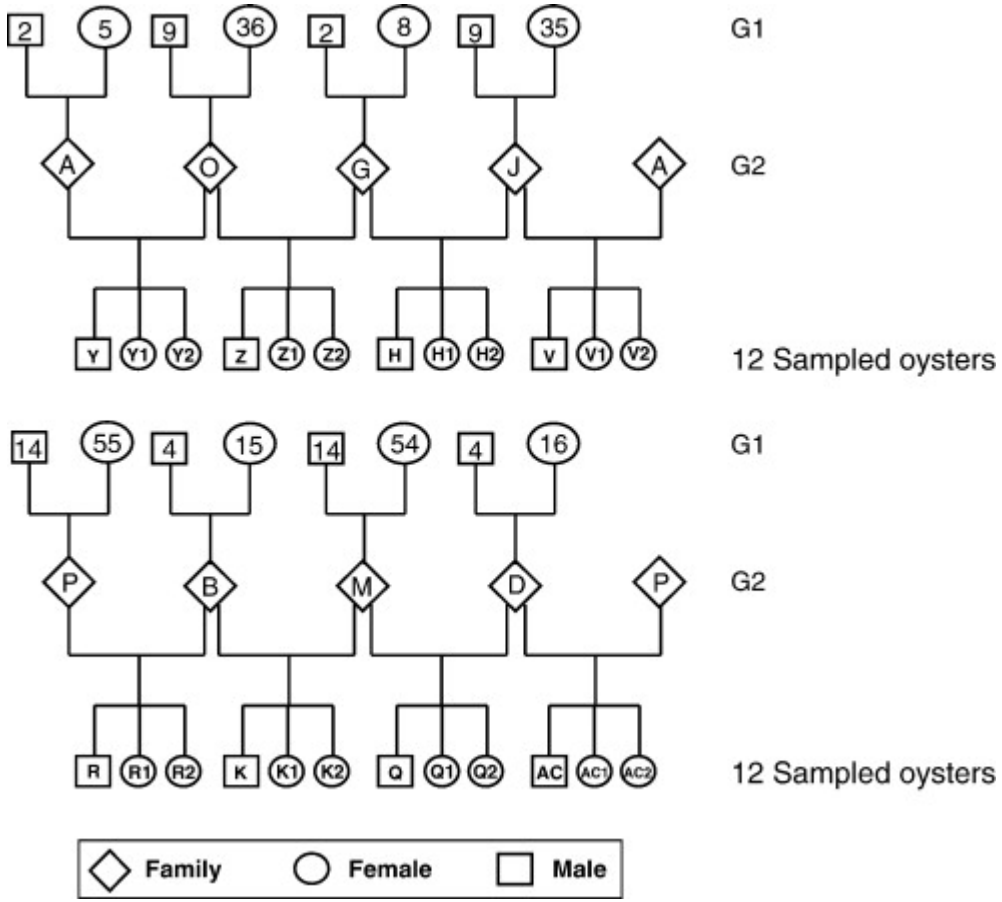

Figure 1: Relationship between the oysters used in our analysis of polymorphism and codon usage bias

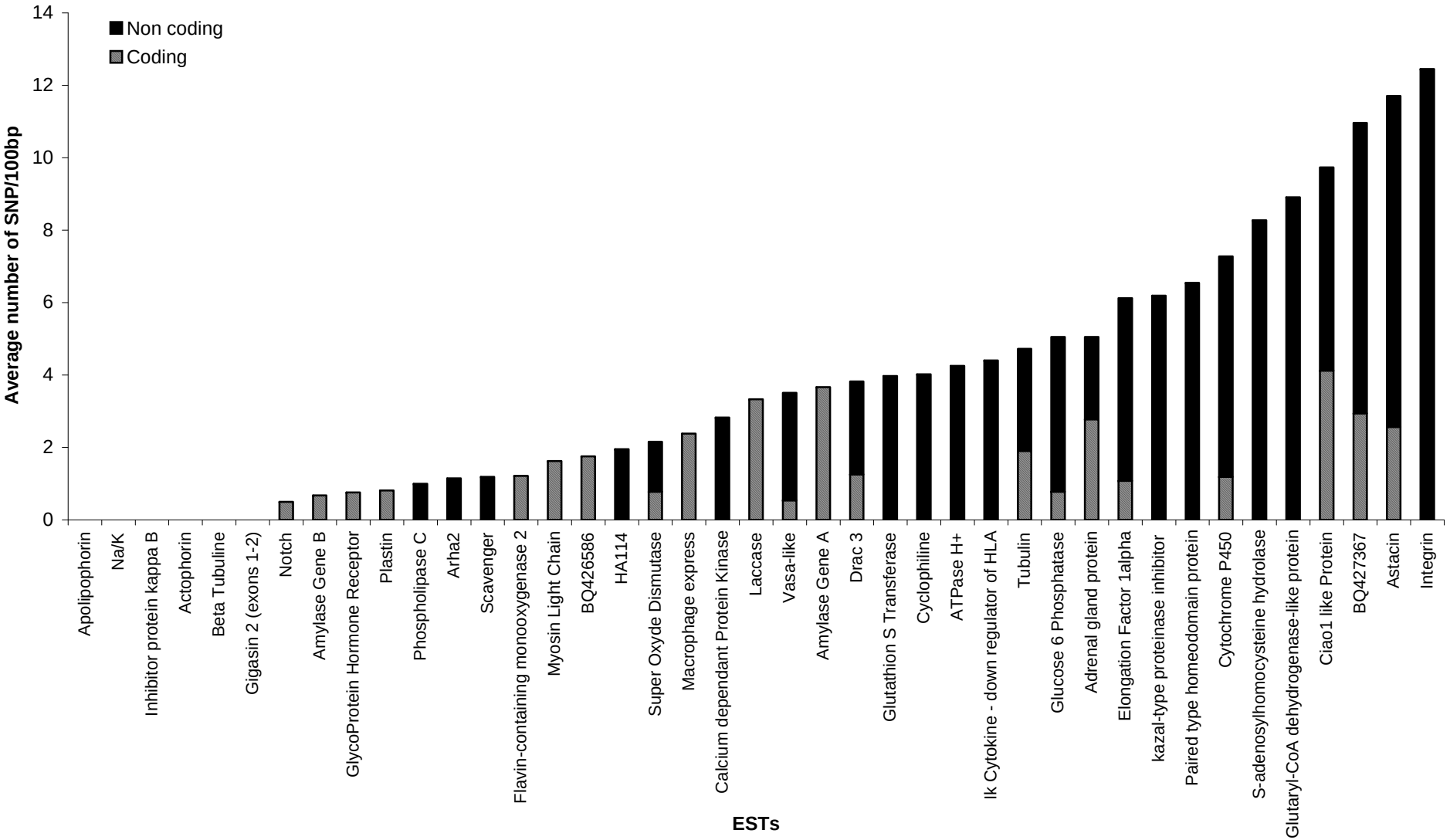

Figure 2: Average level of polymorphism detected in coding and non-coding parts of ESTs.

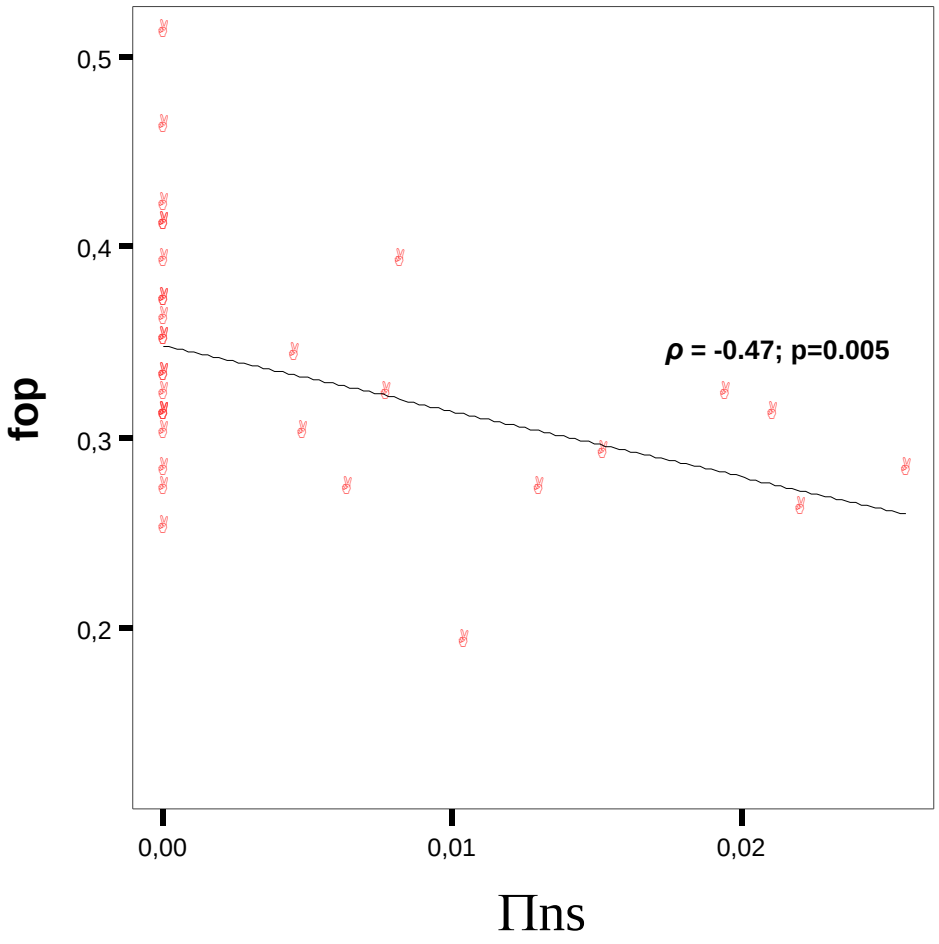

Figure 3: Significant correlation between non-synonymous diversity and codon bias

Table 1. Nucleotide polymorphism in *Crassostrea gigas* nuclear genes

Genes	Accession Number	Pn	Ps	Pnc	Type of non-coding sequence	Ln (bp)	Ls (bp)	L nc (bp)	L cds	Πs	Πn	Πnc	Πsi	Πn/Πsi	Fop	GC3
Drac 3	BQ427023	1	1	2	Intron	122	37	78	325	0,027	0,008	0,026	0,026	0,314	0,39	0,71
Elongation Factor 1alpha	AB122066	0	1	5	Intron	73	20	99	1386	0,050	0	0,051	0,050	0	0,42	0,48
Adrenal gland protein	CK172341	0	4	5	Intron	114	30	220	714	0,133	0,000	0,023	0,036	0,000	0,33	0,52
Astacin	AF075683	2	1	16	Intron	91	26	175	752	0,038	0,022	0,091	0,085	0,260	0,26	0,39
Cytochrome P450	AF075692	0	1	7	Intron	63	21	115	151	0,048	0	0,061	0,059	0	0,35	0,6
Tubulin	AB185494	2	0	4	Intron	78	27	142	533	0	0,026	0,028	0,024	1,083	0,28	0,46
GlycoProtein Hormone Receptor	AJ549813	1	1	0	UTR	207	56	67	3279	0,018	0,005	0,000	0,008	0,594	0,3	0,49
Glutaryl-CoA dehydrogenase-like protein	AJ563484	0	0	18	UTR	10	32	202	216	0	0	0,089	0,077	0	0,3	0,41
Glutathion S Transferase	AJ577235	0	0	11	Intron	70	23	277	537	0	0	0,040	0,037	0	0,37	0,53
Amylase Gene B	AJ496603	1	1	0	Intron	224	73	88	1557	0,014	0,004	0,000	0,006	0,719	0,34	0,49
Apolipophorin	CF369184	0	0	0	Intron	265	74	88	562	0	0	0,000	0		0,31	0,39
Ciao1 like Protein	AY339886	2	6	6	Intron	154	40	107	564	0,150	0,013	0,056	0,082	0,159	0,27	0,38
Vasa-like	AY423380	0	1	9	Intron	138	48	303	2274	0,021	0	0,030	0,028	0	0,35	0,43
ATPase H+	AY551099	0	0	6	Intron	61	17	141	183	0	0	0,043	0,038	0	0,37	0,64
BQ427367	BQ427367	0	3	15	Intron	79	23	187	489	0,130	0	0,080	0,086	0	0,36	0,36
S-adenosylhomocysteine hydrolase	BQ427368	0	0	12	Intron	97	32	145	470	0	0	0,083	0,068	0	0,28	0,52
Super Oxyde Dismutase	AY551094	0	1	2	Intron	97	32	145	576	0,031	0	0,014	0,017	0	0,41	0,52
Glucose 6 Phosphatase	AM076951	0	1	5	Intron	100	29	117	130	0,034	0	0,043	0,041	0	0,51	0,58
Ik Cytokine - down regulator of HLA		0	0	11	Intron	44	7	250	312	0	0	0,044	0,043	0	0,31	0,59
Na/K	AJ563804	0	0	0	Intron	79	20	98	100	0	0	0,000	0		0,41	0,52
Phospholipase C		0	0	2	Intron	76	23	201	343	0	0	0,010	0,009	0	0,32	0,55
Calcium dependant Protein Kinase	AY713401	0	0	5	Intron	131	34	177	744	0	0	0,028	0,024	0	0,31	0,51
Inhibitor protein kappa B	DQ250326	0	0	0	Intron	42	15	166	1086	0	0	0,000	0		0,27	0,41
Plastin	AF075690	1	0			96	27		177	0	0,010		0		0,19	0,36
Actophorin	AF075694	0	0			81	21		201	0	0		0		0,25	0,35
Myosin Light Chain	AJ563458	2	0			95	28		477	0	0,021		0		0,31	0,54
Amylase Gene A	AJ496597	4	6			206	67		1560	0,090	0,019		0,090	0,217	0,32	0,51
Beta Tubuline	AY713400	0	0			118	38		1128	0	0		0		0,46	0,54
BQ426586	BQ426586	2	1			132	39		279	0,026	0,015		0,026	0,591	0,29	0,49

Laccase		0	4			92	28		321	0,143	0		0,143	0	0,33	0,5
Macrophage express	AAR82936	1	3			131	37		471	0,081	0,008		0,081	0,094	0,32	0,52
Notch		1	0			159	42		310	0	0,006		0		0,27	0,37
Flavin-containing monooxygenase 2	AJ585074	0	2			130	35		1365	0,057	0		0,057	0	0,39	0,61
Integrin	BQ426737			30	UTR			241	234			0,124	0,124		0,37	0,53
Gigasin 2 (exons 1-2)	AJ582630			0	Intron			162	83			0,000	0		0,12	0,21
Paired type homeodomain protein	AY187692			11	Intron			168	96			0,065	0,065		0,39	0,67
Cyclophiline	AY551095			8	Intron			199	492			0,040	0,040		0,37	0,5
Scavenger	CX069350			2	Intron			168	246			0,012	0,012		0,14	0,29
HA114				5	Intron			256	360			0,020	0,020		0,38	0,53
Kazal-type proteinase inhibitor	CB617337			25	UTR			404	395			0,062	0,062		0,21	0,35
Arha2	CD526834			3	Intron			261	579			0,011	0,011		0,42	0,6

Pn : Number of non-synonymous polymorphic sites in coding DNA
Ps : Number of synonymous polymorphic sites in coding DNA
Pnc : Number of polymorphic sites in non-coding DNA
Ln : Number of non-synonymous sites in coding DNA
Ls : Number of synonymous sites in coding DNA
Lcds : Length of the longest coding sequence available (bp)
Lnc : Length of the non-coding sequence (bp)
Πn : Number of non-synonymous SNPs per non-synonymous sites, Pn/Ln
Πs : : Number of synonymous SNPs per synonymous sites, Ps/Ls
Πnc : Number of non-coding SNPs per non-coding sites, Pnc/Lnc
Πsi : Number of silent SNPs per silent sites, (Ps+Pnc)/(Ls+Lnc)
Fop: Frequency of optimal codon
GC3: Percentage of G and C content at third position of codon

Table 2: table of optimal codons deduced from the analysis of the EST dataset

Fourfold degenerate codons				Twofold degenerate codons				Six and threefold degenerate codons			
aa	Codon	Spearman Correlation	Probability	aa	Codon	Spearman Correlation	Probability	aa	Codon	Spearman Correlation	Probability
Thr	ACU	-0,02	NS	Lys	AAA			Arg	AGA	-0,03	NS
	ACC	**0,04**	0,012		**AAG**	**0,06**	<.001		AGG	0,03	NS
	ACA	-0,01	NS	Asn	AAU				**CGU**	**0,04**	0,012
	ACG	-0,02	NS		**AAC**	**0,04**	0,01		**CGC**	**0,05**	0,007
Pro	CCU	0,01	NS	Gln	CAA				CGA	0,01	NS
	CCC	**0,04**	0,023		**CAG**	**0,07**	<.001		CGG	-0,01	NS
	CCA	-0,02	NS	His	CAU			Leu	CUU	-0,02	NS
	CCG	-0,03	NS		**CAC**	**0,03**	0,023		**CUC**	**0,05**	0,002
Ala	**GCU**	**0,04**	0,011	Glu	GAA				CUA	-0,03	NS
	GCC	**0,05**	0,003		GAG	0,02	NS		**CUG**	**0,06**	0,001
	GCA	-0,05	0,006	Asp	GAU				UUA	-0,05	0,003
	GCG	-0,02	NS		GAC	0,03	NS		UUG	0,01	NS
Gly	GGU	-0,02	NS	Tyr	UAU			Ser	AGU	-0,02	NS
	GGC	0,01	NS		UAC	0,02	NS		AGC	0,01	NS
	GGA	**0,05**	0,008	Cys	UGU				UCU	0	NS
	GGG	-0,02	NS		UGC	0,01	NS		**UCC**	**0,05**	0,005
Val	GUU	-0,03	NS	Phe	UUU				UCA	-0,01	NS
	GUC	**0,05**	0,002		**UUC**	**0,07**	<.001		UCG	-0,03	NS
	GUA	-0,05	0,002					Ile	AUU	0	NS
	GUG	**0,04**	0,011						**AUC**	**0,1**	<.001
									AUA	-0,1	<.001