Genomic Landscape of Mutational Biases in the Pacific Oyster *Crassostrea gigas*

Kai Song*

School of Mathematics and Statistics, Qingdao University, Shandong, China

*Corresponding author: E-mail: songkai1987@126.com.

Accepted: 23 July 2020

Abstract

Mutation is a driving force of evolution that has been shaped by natural selection and is universally biased. Previous studies determined genome-wide mutational patterns for several species and investigated the heterogeneity of mutational patterns at fine-scale levels. However, little evidence of the heterogeneity of mutation rates over large genomic regions was shown. Hence, the mutational patterns of different large-scale genomic regions and their association with selective pressures still need to be explored. As the second most species-rich animal phylum, little is known about the mutational patterns in Mollusca, especially oysters. In this study, the mutational bias patterns are characterized by using whole-genome resequencing data in the *Crassostrea gigas* genome. I studied the genome-wide relative rates of the pair mutations and found that the predominant mutation is GC -> AT, irrespective of the genomic regions. This analysis reveals that mutational biases were associated with gene expression levels across the *C. gigas* genome. Genes with higher expression levels and breadth expression patterns, longer coding length, and more exon numbers had relatively higher GC -> AT rates. I also found that genes with larger d*N*/d*S* values had relatively higher GC -> AT rates. This work represents the first comprehensive research on the mutational biases in Mollusca species. Here, I comprehensively investigated the relationships between mutational biases with some intrinsic genetic factors and evolutionary indicators and proposed that selective pressures are important forces shaping the mutational biases across the *C. gigas* genome.

Key words: mutational biases, Mollusca species, AT-biased mutational pattern.

Significance

Mutation is a driving force of evolution that has been shaped by natural selection and is universally biased. However, little evidence of the heterogeneity of mutation rates over large genomic regions was shown. Here, I comprehensively investigated the relationships between mutational biases with some intrinsic genetic factors and evolutionary indicators and proposed that selective pressures are important forces shaping the mutational biases across the *Crassostrea gigas* genome.

Introduction

Mutation occurs in all living beings and is a driving force of evolution that has been shaped by natural selection (Smith and Eyre-Walker 2002; Andolfatto 2005; Lynch et al. 2016; Katju and Bergthorsson 2019). However, mutation is not a completely random process because some types of mutations have higher probability than others. In previous studies, it has been shown that the mutational patterns observed in eukaryotes are universally AT-biased (Petrov and Hartl 1999; Haddrill and Charlesworth 2008; Denver et al. 2009; Lynch 2010; Ossowski et al. 2010), in particular because of the high rate of G/C to A/T transitions (Hershberg and Petrov 2010; Hildebrand et al. 2010). This pattern is also common in bacteria, but with few exceptions, such as *Rhodotorula toruloides* (Long et al. 2016), and *Deinococcus radiodurans* (Long et al. 2015). Previous studies determined the genome-wide

mutational patterns for each species and found that the heterogeneity of mutational patterns within the genome could be affected by several factors, such as local sequence context, coding versus noncoding sequences, or distance from replication origin in *Chlamydomonas reinhardtii*, *Bacillus subtilis*, *Escherichia coli*, and *Mesoplasma florum* (Ness et al. 2015; Sung et al. 2015). Different genomic regions experience different selective pressures; this explains the variety in the nucleotide diversity (Song, Li, Huang, et al. 2017; Song et al. 2018, 2019). Hence, the mutational patterns of different genomic regions and their association with selective pressures still need to been explored.

Previous studies investigating mutational biases mainly focused on bacterial, unicellular eukaryotic, and other models. However, mutational pattern studies using Mollusca have rarely been performed. Currently, high-throughput sequencing technology has been widely used for nonmodel organisms and is a powerful approach for genome assembly, making it more flexible and convenient of investigating the genetic and evolutionary patterns for these species. Mollusca is the second largest animal phylum with >100,000 recognized species, including many important food sources for humans, disease vectors, destructive invasive species, or aesthetic resources (Demaintenon 2008). The genomic resources of several molluscan species are available, such as oysters (Zhang et al. 2012), scallops (Li et al. 2017; Wang, Zhang, et al. 2017), abalone (Nam et al. 2017), limpets (Simakov et al. 2013), clams (Mun et al. 2017), and mussels (Sun et al. 2017). Thus, it is possible to investigate the evolutionary patterns at genome-wide levels of molluscan species; however, little is known about the patterns of mutational biases.

The diploid oyster *Crassostrea gigas* has been previously used as a model shellfish species for genetic analysis (Xu et al. 2017; Gagnaire et al. 2018; Guo et al. 2018; Li, Wang, et al. 2018; Wei et al. 2018; Yue et al. 2018) and stress-response mechanism studies (Zhang et al. 2015; Zhao et al. 2016; Wang X, Wang M, et al. 2017). The publication of the *C. gigas* genome allows genetic and evolutionary studies of this species to be performed at the genome-wide level (Zhang et al. 2012). The oyster genome has a relatively modest size of 559 Mb, a GC content of 33.4% and contig N50 of 19.4 kb. The assembly quality of the oyster genome was evaluated by the successful mapping of 99% of the BAC sequences and 98% of expressed sequence tags (EST), which reflect that the genome was sufficient for calling single-nucleotide polymorphisms (SNPs) and studying the pattern of mutational biases.

At the same time, comprehensive whole-genome resequencing data have been produced which make research on mutational patterns across the oyster genome feasible (Li L, Li A, et al. 2018). In this study, I quantified the patterns of mutational biases across the oyster genome and investigated their association with selective pressures.

Materials and Methods

Data Acquisition

Resequencing data from 40 wild *C. gigas* samples collected along China's coastline (Li L, Li A, et al. 2018) were downloaded from the National Center for Biotechnology Information (NCBI) website under the project number PRJNA394055 (supplementary table S1, Supplementary Material online). The transcriptome data of *C. gigas* were also downloaded from the NCBI under accession number GSE31012. The previously generated high-throughput bisulfite sequencing (BS-seq) data set of *C. gigas* from the male gametes, mantle, and gill (Gavery and Roberts 2013; Olson and Roberts 2014; Wang et al. 2014) was downloaded from NCBI (accession number SRX390346, GSE40302, and SRX32737).

Read Mapping and SNP Calling

The cleaned Next Generation Sequencing (NGS) reads were generated from raw sequencing reads using the NGS QC ToolKit version 2.3.3 (Patel and Jain 2012). The BWA software (Burrows–Wheeler Aligner) (Li and Durbin 2009) was used for mapping the clean reads from all samples to the oyster reference genome (GenBank accession number GCA_000297895.1) with the command "bwa aln -o 1 -e -1 -i 5." Then, the Sequence Alignment/MAP (SAM) files for each sample were generated using the command "bwa sampe" with default parameters.

Next, I filtered the low quality alignment reads using the following three steps: 1) filtering out the aligned reads mapped to multiple locations in the genome; 2) filtering out the aligned reads with more than five mismatches to the genome or mapping quality <20; and 3) removing the potential polymerase chain reaction (PCR) duplications using SAMtools with the command "rmdup."

After read alignment, SAMtools (Li et al. 2009) was used for calling SNPs for the 40 *C. gigas* samples using a Bayesian approach. To identify SNPs, the "mpileup" command in SAMtools was used with parameters "-C 50 -t DP -t SP -q 20 -ug."

Gene Expression Measurement

The RNA-seq data sets from eight different tissues and eleven different developmental stages were downloaded from Zhang et al. (2012). The eight tissues include: mantle, gill, adductor muscle, digestive gland, hemocyte, labial palp, female gonad, and male gonad. The eleven developmental stages include: eggs, two cells, four cells, morula, blastula, trochophore, D-shaped larva, umbo larva, pediveliger, spat,

and juvenile. Three software HISAT2 (Kim et al. 2015, 2019), StringTie (Pertea et al. 2015), and Ballgown (Pertea et al. 2016) were used for analyzing the RNA-seq data sets, according to the same pipelines used in the previous study (Song, Li, and Zhang 2017). The fragments per kilobase per million reads (FPKM) for each gene were obtained for the following analysis.

Methylated Site Identification

The BS-seq data from three different tissues of *C. gigas*, including male gametes (Olson and Roberts 2014), mantles (Wang et al. 2014), and gills (Gavery and Roberts 2013) and were analyzed with Bismark (Krueger and Andrews 2011). First, the "bismark_genome_preparation" command was used to prepare the reference genome. Second, the "bismark" command was used to map the BS-seq reads to the oyster genome (GenBank accession number GCA_000297895.1) with parameters "-multicore 12 -bowtie2." Then, the two commands, "bismark_methylation_extractor" and "coverage2cytosine," were used to extract methylation information and report the methylated status for each cytosine site with the default parameters.

Methylated cytosine sites (CpG, CHG, or CHH) are defined as cytosine sites with $\geq$10% methylated Cs and a coverage larger than five.

Orthologous Gene Identification

The evolutionary rates of the orthologous gene pairs (d*N*/d*S* ratio: d*N*, the number of nonsynonymous substitutions per nonsynonymous site; d*S*, the number of synonymous substitutions per synonymous site) between two species from the genus *Crassostrea* (*C. gigas* and *Crassostrea hongkongensis*) were obtained from Zhao et al. (2015).

Tajima's *D* Value Estimation

The software Variscan (version 2.0.3) (Vilella et al. 2005) was used for calculating the statistic of population genetics, Tajima's *D* (Tajima 1989), which tested the neutral assumption. The genes with <10 SNPs were excluded from the calculation of Tajima's *D*.

Calculating the Relative Rates of the Six Nucleotide Pair Mutations

I inferred the direction of mutation using the method described by Hildebrand et al. (2010). The minor mutation was used as a new mutation inferred using the allele frequencies; the mutations were discarded if they had more than two alleles or two alleles at equal frequency. Considering the unequal nucleotide content of the different genomic regions, the counts of the mutations from AT -> GC, CG, or TA were normalized by the number of AT sites at the considered genomic regions, and the counts of the mutations from GC -> TA, CG, or AT were normalized by the number of GC sites (supplementary table S2, Supplementary Material online). In this manner, the expected number of these mutation pairs can be determined under equal GC and AT content.

Table 1
Fraction and Relative Rates of Pair Mutations Per Site after Normalization

Substitutions	Fraction	Mutation Rate
Transitions		
GC -> AT	40.30%	0.032
AT -> GC	16.90%	0.013
Transversions		
AT -> TA	14.50%	0.011
GC -> TA	15.40%	0.012
AT -> CG	6.90%	0.0054
GC -> CG	6.00%	0.0048
Transition/transversion ratio	1.34	—

Results

Spectrums of Different Pair Mutations

In the present study, the whole-genome resequencing data from 40 wild *C. gigas* were aligned to the Pacific oyster reference genome. I obtained an average coverage of 20.3× per individual (supplementary table S1, Supplementary Material online) and identified 30 million SNP, of which 15 million belonged to gene regions and 2.4 million to coding regions.

I classified these mutations into six possible types (AT -> CG, AT -> TA, AT -> GC, GC -> TA, GC -> CG, and GC -> AT). Of these, the two pair mutations AT -> GC and GC -> AT were transitions whereas the other four pair mutations were transversion. The transition/transversion ratio was 1.34 (table 1). I investigated the spectrum of different pair mutations and found that different pair mutations had very similar spectrums irrespective of whether they were the transversion or transition type (fig. 1).

Mutation is AT-Biased

Because the GC and AT content of different genomic regions was different, we calculated the relative rate of each of the six pair mutations by normalizing the current GC and AT content at the studied regions (Materials and Methods). First, I investigated the genome-wide relative rates of the pair mutations and found that the predominant mutation is GC -> AT irrespective of the genomic regions (fig. 2). The genome-wide relative rate of GC -> AT was 40.3%, whereas the rates were 39.2%, 46.6%, and 40.2% for the intergenic, coding and intronic regions, respectively. The coding regions had a significantly higher GC -> AT relative rates than other regions (Pearson χ^2 test, *P* value of $<10^{-16}$).

Because the methylation of cytosine sites can significantly increase the rates of cytosine-to-thymine (C-to-T) transitions, I

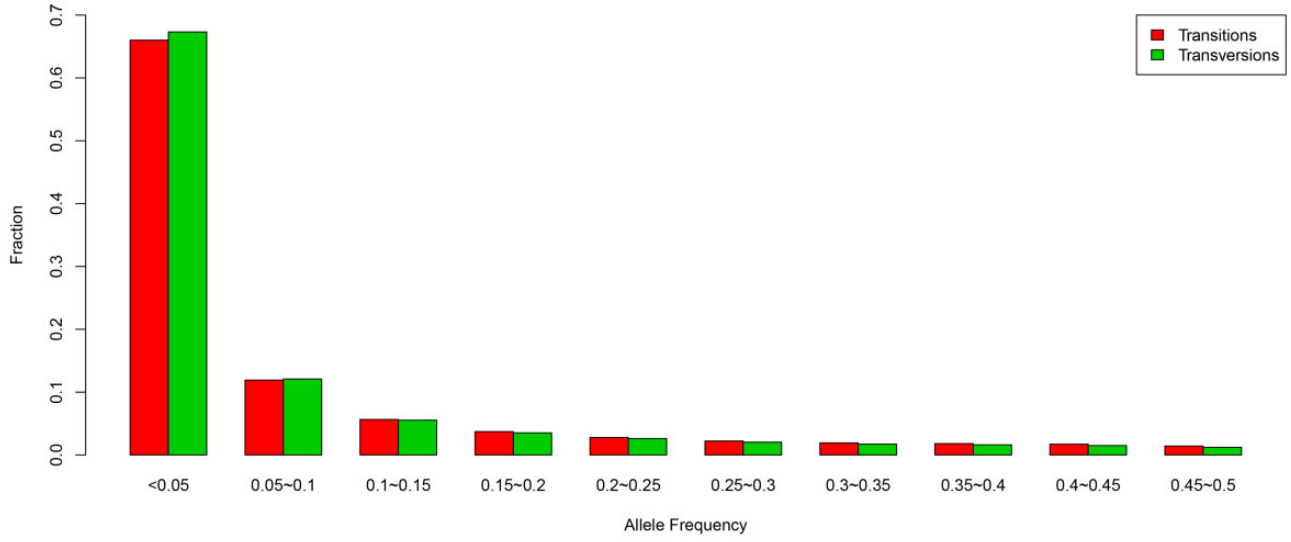

Fig. 1.—Allelic frequency distribution for the transition and transversion pair mutations. The relative proportion of the nucleotide pair mutations is shown in table 1. The allelic frequency distribution is calculated by normalizing in each of them.

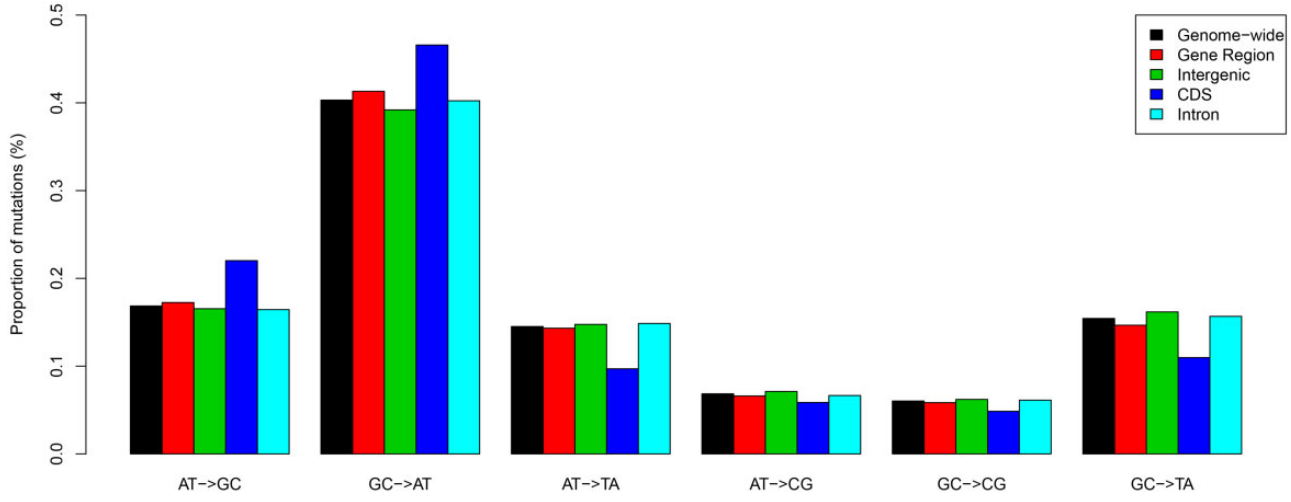

Fig. 2.—Relative rates of the six nucleotide pair mutations in different genomic regions. The most common mutation is always GC -> AT transitions for different genomic regions. The rates are normalized for the unequal nucleotide content of the different genomic regions (Materials and Methods). The relative rates estimated for each genomic region using all the nucleotide sites.

first analyzed the relative rates of the pair mutations of the methylated sites. Thus, I used the BS-seq data set from the three oyster tissues to identify the methylated sites. I found that the mutation rate of the methylated CG sites was 17.3% (table 2), which was much higher than that of the genome-wide CG sites (4.9%). Of the methylated cytosine sites, 88.1% of the pair mutations were GC -> AT which suggests that DNA methylation played an important role in shaping the mutational pattern of the methylated cytosine sites. However,

Table 2
The Mutation Rates for Methylated and Unmethylated Cytosines

	Number	Mutation Rate	Fraction of GC -> AT
Methylated cytosine	6,268,531	0.173	88.1%
Methylated CpG	6,205,846	0.173	88.1%
Methylated CHG	5,014	0.124	82.5%
Methylated CHH	57,671	0.112	81.2%
Unmethylated cytosine	158,692,819	0.049	—

only ~4% of the cytosine sites in the Pacific oyster genome were identified as methylated using the BS-seq data from three different tissues; thus, I investigated the relative rates of other pair mutations for cytosine sites to exclude these methylated sites. The predominant mutation is also GC -> AT for unmethylated sites. The genome-wide relative rate of GC -> AT was 38.2%, a little lower than the previous estimation with methylated sites. The rates were 38.1%, 43.2%, and 37.9% for the intergenic, coding, and intronic regions, respectively. Therefore, the results showed that, for unmethylated cytosines, other factors also influence the predominant GC -> AT mutations.

Relationship between Mutational Biases and Gene Expression Patterns

I investigated the extent of mutational biases for genes with different expression patterns across the *C. gigas* genome (fig. 3). The total number of genes was divided into four different groups using the quartile of their expression levels, from the bottom 25% to the top 25%. As shown in figure 3*a*, the lowest 25% expressed genes had a relative GC -> AT rate of 42.0% which was significantly lower than that of the other three groups of genes (*t* test, *P* value of $<10^{-16}$). The highest 25% expressed genes had the highest rate of 45.5%.

I also examined whether the genes with different temporal or spatial expression patterns had different relative mutation rates. The relative GC -> AT rates observed in genes that were not expressed in any tissue or developmental stage were lower than those observed in other genes expressed in more than one tissue or developmental stage (fig. 3*b* and *c*). The relative GC -> AT rates of genes with expression in all tissues and developmental stages were higher than those of other genes.

Relationship between Mutational Biases and Gene Length

Coding sequence (CDS) length is an important factor affecting protein evolution. In the present study, the CDS length of each gene was extracted from the oyster genome and their association with the relative mutation rates was investigated. The total genes were divided into four different groups using the quartile of their CDS length, from the bottom 25% to the top 25%. The relative GC -> AT rates of genes with the shortest length were lower than those of other genes, whereas the relative rates of genes with the longest length were the highest (fig. 4*a*).

In the Pacific oyster genome, a large fraction of genes has more than two exons (~83% of the genes). Therefore, the relationships between exon number and intronic length with relative mutation rates were also investigated in the present study. First, the total genes were divided into three groups based on the number of exons they contained: one, two, or more than two. As shown in figure 4*b*, the relative rates of genes with more than two exons were significantly and relatively higher than those of the other genes, whereas the difference between genes with one exon and two exons was not significant.

Additionally, I also found that intronic length influenced the relative mutation rates. The total intronic regions were divided into four groups according to their length using the quantile from the top 25% to the bottom 25%. The relative GC -> AT rates of the intronic regions with the shortest length were the lowest among the different intron groups. The relative rates were the highest in intronic regions with the longest length (fig. 4*c*).

Relationship between Mutational Biases and Selective Pressures

To determine the relationship between mutational biases and selective pressures, I used 11,409 genes (~40% of the total genes, 11,409/28,000) that have orthologs between two species in the *Crassostrea* genus, *C. gigas* and *C. hongkongensis* (Zhao et al. 2015). First, I compared the relative GC -> AT rates of the coding regions between the gene groups with and without orthologs. As shown in figure 5*a*, the genes with orthologs had a significantly higher level of relative GC -> AT rates than those without orthologs.

Then, I used the 11,409 genes with orthologs to investigate the relationship between mutational biases and selective pressures. I divided these genes into four different groups based on the d*N*/d*S* values between these two species, from the top 25% to the bottom 25%. As shown in figure 5*b*, genes with the lowest d*N*/d*S* values had the lowest levels of relative GC -> AT rates among the four gene groups, whereas those with the largest d*N*/d*S* values had the highest relative rates.

I calculated the Tajima's *D* (Tajima 1989) value for each gene using the polymorphic sites. The Tajima's *D* values distribution was strong negatively skewed mostly because of the high proportion of low-frequency SNPs in the *C. gigas* genome (supplementary fig. S1, Supplementary Material online). In addition, this phenomenon was also observed in other species such *Arabidopsis thaliana* (Nordborg et al. 2005) and *Medicago truncatula* (Branca et al. 2011), which can be explained by purifying selection. Then, I divided these genes into three different groups based on the Tajima's *D* values, from the top 10%, the middle portion, and the bottom 10%. The genes with lowest Tajima's *D* values were considered to be influenced by purifying selection. The relative GC -> AT rates were not different between the genes with the top 10% and bottom 10% Tajima's *D* values. The relative rates among these three groups were not significantly different (*t* test, *P* value > 0.05).

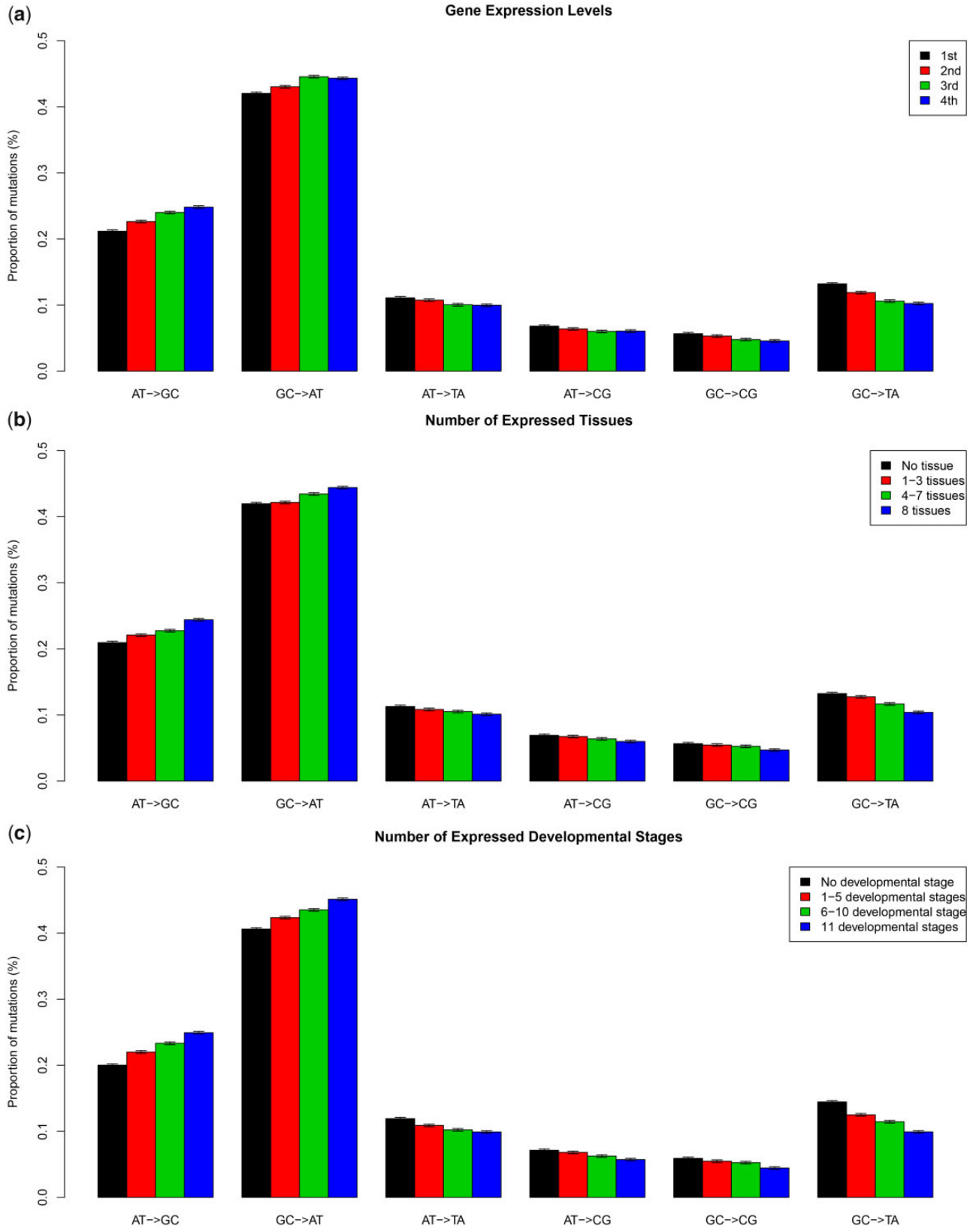

Fig. 3.—Relationship between mutational biases and gene expression patterns. (*a*) Total genes divided into four groups based on their expression levels from first (the bottom 25%) to fourth (the top 25%). The relative rates of the six nucleotide pair mutations were estimated for each group (*b*, *c*). Total genes divided into groups based on the number of tissues and developmental stages they are expressed in.

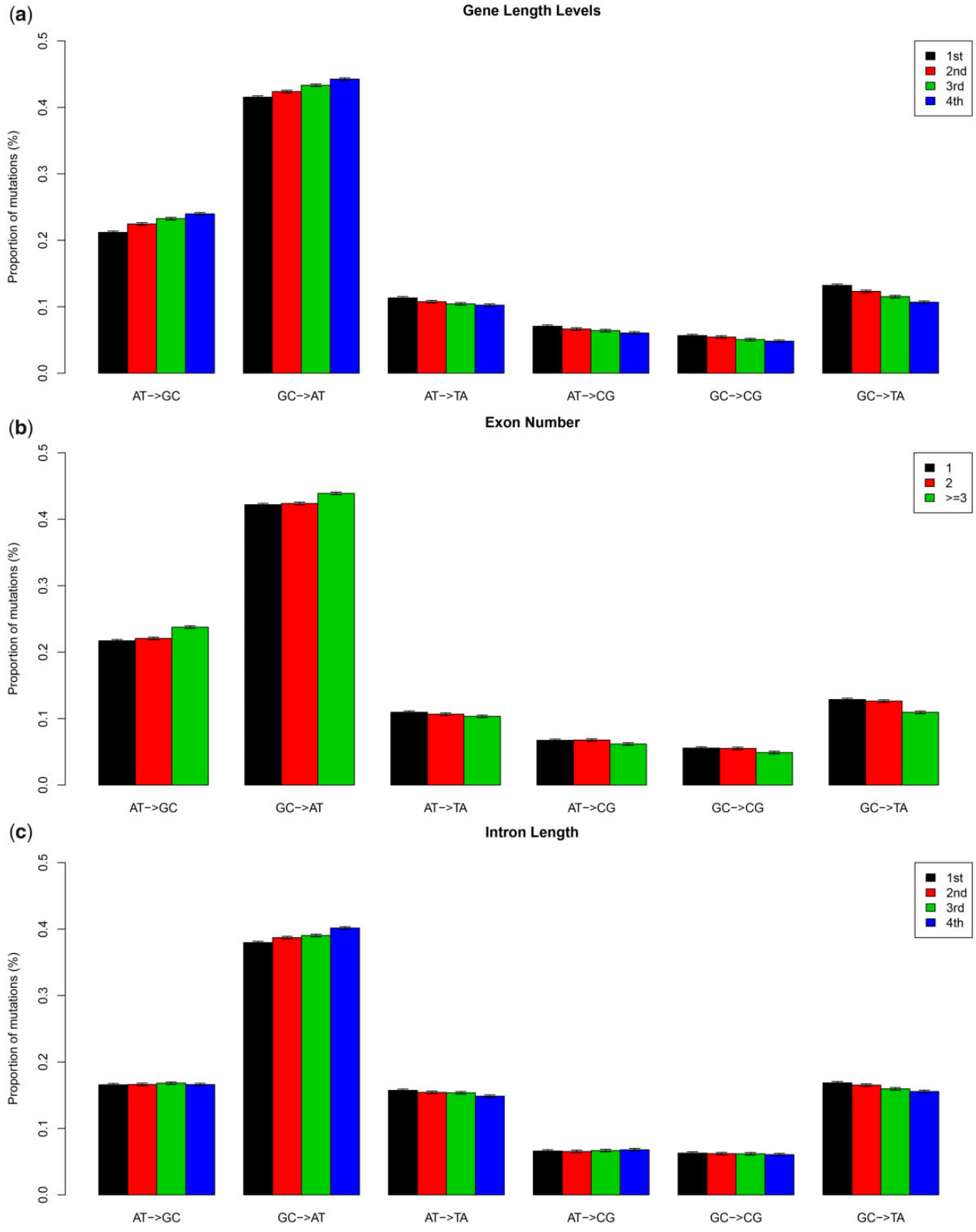

Fig. 4.—Relationship of mutational biases with gene length, exon number, and intronic length. (*a*) Total genes divided into four groups based on their length (coding sequences length) from the first (the bottom 25%) to fourth (the top 25%). The relative rates of the six nucleotide pair mutations were estimated for each group. (*b*) Total genes divided into three groups based on their exon number. (*c*) Total intronic regions divided into four groups based on their length.

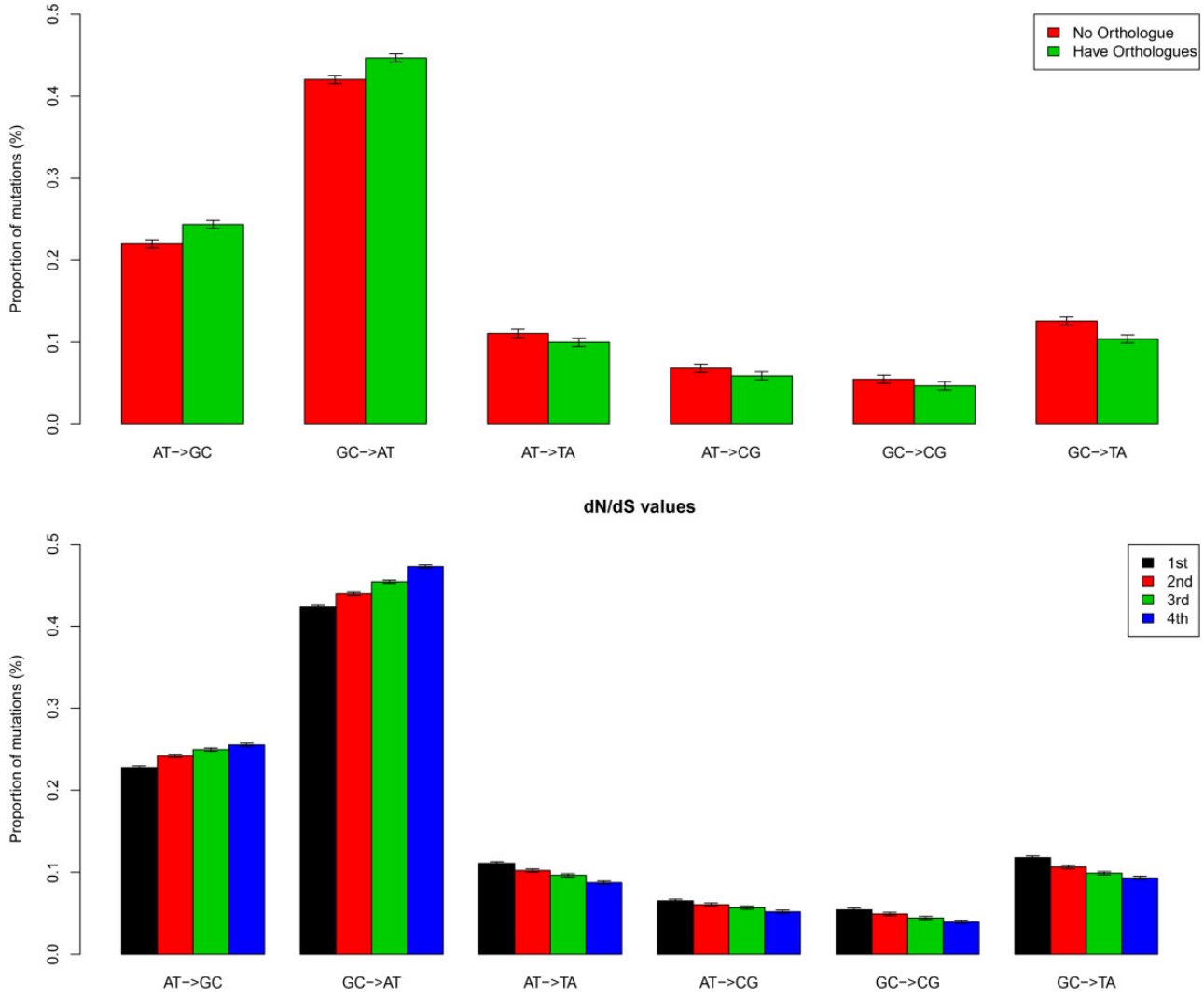

Fig. 5.—Relationship of mutational biases with evolutionary patterns. (*a*) Total genes divided into two groups based on whether they had orthologs or not compared with *Crassostrea hongkongensis* genes. (*b*) The genes with orthologs divided into four groups based on the d*N*/d*S* values.

Discussion

In the present study, the patterns of mutational biases were characterized in different genomic regions for the *C. gigas* genome using resequencing data. The genome-wide mutational bias rates were quantitatively estimated, and their association with some intrinsic genetic factors and evolutionary indicators were investigated. Previous studies mainly focused on mutational biases occurring in bacterial species (Hershberg and Petrov 2010; Hildebrand et al. 2010; Long et al. 2015) and other model organisms (Lynch et al. 2008; Denver et al. 2009; Keightley et al. 2009; Ossowski et al. 2010; Long et al. 2016). However, although Mollusca is the second most species-rich animal phylum, little is known about their mutational bias patterns. Therefore, in the present study, I used a globally distributed Mollusca—the Pacific oyster—as a representative to investigate the patterns of mutational biases.

DNA methylation is an important epigenetic modification in eukaryotic genomes that could significantly increase the rates of C-T mutations. Although the genome-wide methylated cytosine sites estimated in my study were only ~4% in the *C. gigas* genome, the influence of these sites on the mutational patterns was considerable in this organism. In addition to this, I also analyzed the patterns of mutational biases by excluding the methylated cytosine which showed that other factors also influence the mutational biases, such as the oxidation of the guanine, which is also a common factor increasing the GC to AT mutation rate.

In previous studies, the dominant mutation was GC -> AT and other pair mutations rates were much lower in bacteria and *Plasmodium falciparum* (Hershberg and Petrov 2010; Hildebrand et al. 2010; Long et al. 2014, 2015, 2016; Hamilton et al. 2016). However, in the present study, I found that, in the Pacific oyster, there was a dominant mutation of GC -> AT with a relative rate of around 40%; this rate was more

balanced than other relative mutation rates. The relative rates for mutation pairs AT -> GC, AT -> TA, and GC -> TA were all around 15%.

The rates of mutational biases estimated at the genome-wide level could not reflect the heterogeneity of the mutational patterns within the genome. The natural selective pressures were different on different parts of the genome which were observed in many species (Begun et al. 2007; Halligan et al. 2010). Thus, the patterns of mutational biases should be heterogeneous across different genomic regions and would be influenced by their function and evolutionary characteristics, which had not been well investigated. In the present study, I investigated the association between the patterns of mutational biases and some genomic factors, such as gene expression patterns, gene length, and gene evolutionary patterns. In my previous study, I found that both purifying selection (Song et al. 2018) and positive selection (Song et al. 2019) were strongly driving the evolutionary patterns of genes with higher expression levels or breadth expression patterns in *C. gigas*. The change in mutation rate between gene expression categories is common, and several mechanisms can explain these variations both in the sense of increasing or decreasing mutation rate (Hanawalt and Spivak 2008; Jinks-Robertson and Bhagwat 2014; Belfield et al. 2018), such as transcription-coupled repair (TCR), and DNA mismatch repair (MMR) mechanisms. In the present study, except for the variation in mutation rates in genes with different expressions, I also found that the pattern of mutational biases was associated with expression levels of which the underlying causes still need to be explored. The process of GC-biased gene conversion, a by-product of recombination, has been shown to influence GC composition variation and promote the segregation and fixation of deleterious AT to GC mutations within and between genomes in many lineages across the eukaryotic tree of life (Duret and Galtier 2009; Pessia et al. 2012) which could be an explanation for the heterogeneity of mutational patterns within the genome. The GC to AT bias is significantly stronger in coding regions than other regions which maybe a consequence of the codon bias toward GC rich codon, because the observed mutation pattern is not exactly the spontaneous mutation pattern due to selection and drift that had the time to occur in the population.

I also used another indicator, Tajima's *D* value, to classify the genes into three different groups. I believed that the genes with the lower Tajima's *D* values experienced stronger purifying selection. I found no difference between the genes with the lowest and highest Tajima's *D* values, suggesting that purifying selection played a weak role in shaping the mutational pattern in the *C. gigas* genome.

The study of mutational patterns was very scarce in Mollusca which is the animal phylum with the second most species. Therefore, the present study provided some information about the mutational patterns of *C. gigas* and represents the first such study in Mollusca. Several factors could affect the heterogeneity of mutational patterns within the genome, such as local sequence context, or distance from replication origin. However, as an indicator affected by natural selection, the associations of mutational biases and gene expression patterns, gene length, exon number and selective pressures are still unexplored in bacteria and other model organisms. The findings from the present study effectively complement previous research in other organisms. Altogether, this study provides evidence that the proportion of GC -> AT mutations increases with expression level and gene length whereas it decreases with purifying selection acting on amino-acid composition.

Supplementary Material

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

Acknowledgment

This study was supported by the National Natural Science Foundation of China (Grant No. 11701546).

Literature Cited

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437(7062):1149–1152.

Begun DJ, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5(11):e310.

Belfield EJ, et al. 2018. DNA mismatch repair preferentially protects genes from mutation. Genome Res. 28(1):66–74.

Branca A, et al. 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. Proc Natl Acad Sci USA. 108(42):E864–E870.

Demaintenon MJ. 2008. Phylogeny and evolution of the Mollusca. Bull Mar Sci-Miami. 83:435–437.

Denver DR, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proc Natl Acad Sci USA. 106(38):16310–16314.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genom Hum Genet. 10(1):285–311.

Gagnaire P-A, et al. 2018. Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. Genome Biol Evol. 10(9):2518–2534.

Gavery MR, Roberts SB. 2013. Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. PeerJ 1:e215.

Guo X, Li C, Wang H, Xu Z. 2018. Diversity and evolution of living oysters. J Shellfish Res. 37(4):755–772.

Haddrill PR, Charlesworth B. 2008. Non-neutral processes drive the nucleotide composition of non-coding sequences in *Drosophila*. Biol Lett. 4(4):438–441.

Halligan DL, Oliver F, Eyre-Walker A, Harr B, Keightley PD. 2010. Evidence for pervasive adaptive protein evolution in wild mice. PLoS Genet. 6(1):e1000825.

Hamilton WL, et al. 2016. Extreme mutation bias and high AT content in Plasmodium falciparum. Nucleic Acids Res. 45:1889–1901.

Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol. 9(12):958–970.

Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6(9):e1001115.
Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 6(9):e1001107.
Jinks-Robertson S, Bhagwat AS. 2014. Transcription-associated mutagenesis. Annu Rev Genet. 48(1):341–359.
Katju V, Bergthorsson U. 2019. Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. Genome Biol Evol. 11(1):136–165.
Keightley PD, et al. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19(7):1195–1201.
Kim D, Landmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 12(4):357–U121.
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37(8):907–915.
Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11):1571–1572.
Li C, Wang J, et al. 2018. Construction of a high-density genetic map and fine QTL mapping for growth and nutritional traits of *Crassostrea gigas*. BMC Genomics 19(1):626.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.
Li L, Li A, et al. 2018. Divergence and plasticity shape adaptive potential of the Pacific oyster. Nat Ecol Evol. 2(11):1751–1760.
Li YL, et al. 2017. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nat Commun. 8(1):1721.
Long H, Behringer MG, Williams E, Te R, Lynch M. 2016. Similar mutation rates but highly diverse mutation spectra in ascomycete and basidiomycete yeasts. Genome Biol Evol. 8(12):3815–3821.
Long H, et al. 2015. Background mutational features of the radiation-resistant bacterium *Deinococcus radiodurans*. Mol Biol Evol. 32(9):2383–2392.
Long H, et al. 2014. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. Genome Biol Evol. 7(1):262–271.
Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci USA. 107(3):961–968.
Lynch M, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA. 105(27):9272–9277.
Lynch M, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 17(11):704–714.
Mun S, et al. 2017. The whole-genome and transcriptome of the Manila clam (*Ruditapes philippinarum*). Genome Biol Evol. 9(6):1487–1498.
Nam B-H, et al. 2017. Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. Gigascience 6(5):1–8.
Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD. 2015. Extensive de novo mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. Genome Res. 25(11):1739–1749.
Nordborg M, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 3(7):e196.
Olson CE, Roberts SB. 2014. Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes. Front Physiol. 5:224.
Ossowski S, et al. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327(5961):92–94.
Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7(2):e30619.
Pertea M, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 33(3):290–295.
Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 11(9):1650–1667.
Pessia E, et al. 2012. Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol Evol. 4(7):675–682.
Petrov DA, Hartl DL. 1999. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. Proc Natl Acad Sci USA. 96(4):1475–1479.
Simakov O, et al. 2013. Insights into bilaterian evolution from three spiralian genomes. Nature 493(7433):526–531.
Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. Nature 415(6875):1022–1024.
Song K, Li L, Zhang G. 2018. Relationship among intron length, gene expression, and nucleotide diversity in the Pacific Oyster *Crassostrea gigas*. Mar Biotechnol. 20(5):676–679.
Song K, Li L, Zhang GF. 2017. Bias and correction in RNA-seq data for marine species. Mar Biotechnol. 19(5):541–550.
Song K, Li YX, Huang BY, Li L, Zhang GF. 2017. Genetic and evolutionary patterns of innate immune genes in the Pacific oyster *Crassostrea gigas*. Dev Compar Immunol. 77:17–22.
Song K, Wen S, Zhang G. 2019. Adaptive evolution patterns in the pacific oyster *Crassostrea gigas*. Mar Biotechnol. 21(5):614–622.
Sun J, et al. 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nat Ecol Evol. 1(5):0121.
Sung W, et al. 2015. Asymmetric context-dependent mutation patterns revealed through mutation–accumulation experiments. Mol Biol Evol. 32(7):1672–1683.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595.
Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. 2005. VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics 21(11):2791–2793.
Wang S, Zhang J, et al. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol. 1:0120.
Wang X, et al. 2014. Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation. BMC Genomics 15(1):1119.
Wang X, Wang M, et al. 2017. A carbonic anhydrase serves as an important acid-base regulator in pacific oyster *Crassostrea gigas* exposed to elevated CO_2: implication for physiological responses of mollusk to ocean acidification. Mar Biotechnol. 19(1):22–35.
Wei L, et al. 2018. The molecular differentiation of anatomically paired left and right mantles of the Pacific oyster *Crassostrea gigas*. Mar Biotechnol. 20(4):425–411.
Xu L, Li Q, Yu H, Kong L. 2017. Estimates of heritability for growth and shell color traits and their genetic correlations in the black shell strain of pacific oyster *Crassostrea gigas*. Mar Biotechnol. 19(5):421–429.
Yue C, Li Q, Yu H. 2018. Gonad transcriptome analysis of the pacific oyster *Crassostrea gigas* identifies potential genes regulating the sex determination and differentiation process. Mar Biotechnol. 20(2):206–219.
Zhang GF, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490(7418):49–54.
Zhang Y, et al. 2015. Proteomic basis of stress responses in the gills of the pacific oyster *Crassostrea gigas*. J Proteome Res. 14(1):304–317.
Zhao X, Yu H, Kong L, Li Q. 2016. Gene co-expression network analysis reveals the correlation patterns among genes in euryhaline adaptation of *Crassostrea gigas*. Mar Biotechnol. 18(5):535–544.
Zhao X, Yu H, Kong L, Liu S, Li Q, 2015. Comparative transcriptome analysis of two oysters, *Crassostrea gigas* and *Crassostrea hongkongensis* provides insights into adaptation to hypo-osmotic conditions (vol 9, e111915, 2014). PLoS One 10:e111915.

Associate editor: Gwenael Piganeau