



# Genome Structural Variation Landscape and Its Selection Signatures in the Fast-growing Strains of the Pacific Oyster, *Crassostrea gigas*

Zexin Jiao<sup>1</sup> · Yuan Tian<sup>1</sup> · Boyang Hu<sup>1</sup> · Qi Li<sup>1,2</sup> · Shikai Liu<sup>1,2</sup>

Received: 16 June 2021 / Accepted: 17 August 2021 / Published online: 8 September 2021  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

The Pacific oyster (*Crassostrea gigas*) genome is highly polymorphic and affluent in structural variations (SVs), a significant source of genetic variation underlying inter-individual differences. Here, we used two genome assemblies and 535 individuals of genome re-sequencing data to construct a comprehensive landscape of structural variations in the Pacific oyster. Through whole-genome alignment, 11,087 short SVs and 11,561 copy number variations (CNVs) were identified. While analysis of re-sequencing data revealed 511,170 short SVs and 979,486 CNVs, a total of 63,100 short SVs and 58,182 CNVs were identified in at least 20 samples and regarded as common variations. Based on the common short SVs, both *Fst* and *Pi* ratio statistical methods were employed to detect the selective sweeps between 20 oyster individuals from the fast-growing strain and 20 individuals from their corresponding wild population. A total of 514 overlapped regions (8.76 Mb), containing 746 candidate genes, were identified by both approaches, in addition with 103 genes within 61 common CNVs only detected in the fast-growing strains. The GO enrichment and KEGG pathway analysis indicated that the identified candidate genes were mostly associated with apical part of cell and were significantly enriched in several metabolism-related pathways, including tryptophan metabolism and histidine metabolism. This work provided a comprehensive landscape of SVs and revealed their responses to selection, which will be valuable for further investigations on genome evolution under selection in the oysters.

**Keywords** *Crassostrea gigas* · Structural variation · Artificial selection · Fast growth

## Introduction

Structural variations (SVs) are generally defined as the large-scale variations, which would alter chromosomal structure and provide the raw material for evolution (Hurles et al. 2008; Guan and Sung 2016). Originally, the lengths of SVs were limited at least 1,000 bp. With DNA sequencing becoming routine, the operational spectrum of SVs has been then widened to include much smaller events, the lengths of which are greater than 50 bp (Alkan et al. 2011; Kosugi et al. 2019). SVs are commonly classified into various types based on their structure features, including deletion (DEL),

insertion (INS), inversion (INV), duplication (DUP), translocation (BND), and copy number variation (CNV).

Compared to single nucleotide polymorphisms (SNPs), SVs can cause large-scale perturbations of *cis*-regulatory regions or directly alter the gene copy number to have greater influences on the gene expression and phenotypes (Weischenfeldt et al. 2013; Chiang et al. 2017; Alonge et al. 2020). However, SVs are more challenging to be detected in comparison with single nucleotide variations (Hurles et al. 2008). As a result, the studies of SVs lagged much behind SNPs. The rapid development of next-generation sequencing technology and reliable detection approaches make it possible to detect the full extent of SVs and genotype it routinely.

Numerous studies of SVs have been conducted in human (*Homo sapiens*) with the aim to identify the association of genomic SVs with genetic diseases (Stankiewicz and Lupski 2010; Vacic et al. 2011; MacDonald et al. 2014). In addition, to uncover the genetic molecular basis of important economic traits, numerous studies have also been undertaken in various domestic plants and animals. For example, in maize (*Zea*

Zexin Jiao and Yuan Tian contributed equally to this work.

✉ Shikai Liu  
liushk@ouc.edu.cn

<sup>1</sup> Key Laboratory of Mariculture (Ocean University of China), Ocean University of China Ministry of Education College of Fisheries, Qingdao 266003, China

<sup>2</sup> Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

mays), several SVs on chromosome 4 could affect the oil concentration and long-chain fatty acid composition, through regulating expressions of 16 functional genes (Yang et al. 2019). An important harvesting trait in tomato (*Lycopersicon esculentum*), jointless fruit pedicel, is originated from the mutations of four SVs (Alonge et al. 2020). In cattle (*Bos taurus*), 34 CNVs on 22 chromosomes were identified to be significantly associated with several milk production traits (Xu et al. 2014). In addition, during the process of sheep domestication, a few biological processes and traits were related with hundreds of CNVs, including follicular development and fertility, adipogenesis, wool production, milk production, oxygenated red blood cells, and spleen size (Li et al. 2020). This work revealed the critical and underexplored roles of SVs in genotype-to-phenotype relationships. However, despite their importance, SV landscapes are only systemically characterized in a few aquaculture species. Recently, thousands of high-confidence SVs are identified in 492 Atlantic salmon (*Salmo salar*), suggesting their roles in the genome evolution and genetic architecture of domestication traits (Bertolotti et al. 2020).

The Pacific oyster, *Crassostrea gigas*, is one of the most cultivated bivalve species, contributing significantly to global seafood production (Troost 2010; Zhao et al. 2012). Given its economic importance, numerous selective breeding programs of the Pacific oyster have been conducted over years with the aim to improve the economically important traits (Evans and Langdon 2006; Li et al. 2011; de Melo et al. 2016). In China, we have conducted the selective breeding program of the Pacific oyster since 2007, by constructing breeding base populations with the oysters collected from wild populations in Rushan (China), Miyagi (Japan), and Busan (South Korea). After generations of artificial selection for fast growth, several fast-growing strains have been produced with superior growth performance. Several studies informed by SNPs and other molecular markers have been conducted to investigate the genetic basis of the fast-growing traits of the Pacific oysters (Zhong et al. 2013; Jin et al. 2014; Kong et al. 2014; Wang and Li 2017). However, the large-scale SV landscapes of the Pacific oyster have not been systemically characterized and their potential association with growth remains largely unexplored.

In the present study, we performed whole-genome alignments and genome re-sequencing data analyses to identify genome-wide SVs and construct the first comprehensive SV landscape in the Pacific oyster. Selective sweeps were further detected to determine the SV differentiations between the fast-growing strains and their wild populations, providing insights into the potential role of SVs associated with growth in the Pacific oyster. This work provided the first comprehensive overview of SVs which would be valuable information for future investigations on genome evolution under selection in the oysters.

## Material and Methods

### Data Collection

Two sources of datasets were used for the detection of SVs, including genome assemblies and whole-genome re-sequencing data. Two genome assemblies (cgigas\_uk\_roslin\_v1 and ASM1103280v1) of the *C. gigas* were retrieved from NCBI genome database with the assembly accession number of GCA\_902806645.1 and GCA\_011032805.1, respectively. For whole-genome re-sequencing data, 150-bp paired-end short reads sequenced from 495 samples were retrieved from NCBI Sequence Read Archive (SRA) database (BioProject ID: PRJNA394055) with the detailed sample information described in the previous study (Li et al. 2018). In addition, whole-genome re-sequencing data of 40 samples were sequenced from the oysters from our selection breeding program (Li et al. 2011). Among which, 10 samples were randomly chosen from each of the fast-growing strains which have been successively selected for 10 generations and possessed superior growth advantages over wild oysters. We also sequenced 20 samples from the wild populations in Rushan and Miyagi (10 samples from each population). Adductor muscle tissues of the 6-month-old individuals were dissected and used for DNA extraction following a modified phenol–chloroform protocol (Li et al. 2006). DNA integrity and quantity were assessed using 1% agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). The DNA libraries were constructed with an average insert size of 350 bp. The 150-bp paired-end short reads were sequenced from the Illumina HiSeq X Ten platform with sequencing depth of 10×.

### SV Detection from Whole-Genome Alignment

Whole-genome alignment was performed between cgigas\_uk\_roslin\_v1 and ASM1103280v1 genome assemblies using nucmer software of MUMmer (v3.1) program with the parameters of “-maxmatch -l 100 -c 500” (Kurtz et al. 2004). Then, alignment block was filtered using delta-filter software with one alignment mode (−1). SVs were determined based on the filtered blocks using the Assemblytics software (Nattestad and Schatz 2016). The length of short SVs was limited between 50 and 1000 bp, while the length of CNVs ranged from 1000 to 100,000 bp.

### SV Detection from Whole-Genome Re-sequencing Data

The Fastp (v.20.0) software was employed to trim adaptor sequences and filter low-quality reads (quality score < 20 or length < 35), in order to obtain high-quality clean reads for

downstream analysis (Chen et al. 2018). The FastQC (v0.11.8) was used to assess the quality of clean reads (Kim et al. 2018). The clean reads from each of 535 samples were aligned to the *cgigas\_uk\_roslin\_v1* reference genome using BWA-mem (v0.7.17) with the default parameters (Li and Durbin 2009). Alignment results were sorted and converted to BAM files using SAMtools (v1.9) software (Li et al. 2009). The sequencing coverage and depth of each sample were estimated using Bamdst software (<https://github.com/shiquan/bamdst>).

The SVs were mainly classified into two categories based on the length, including short SVs (> 50 bp and ≤ 1000 bp) and CNVs (> 1000 bp). To obtain individual-specific short SVs, variations of 535 whole-genome re-sequencing samples were independently called using delly (v0.8.1) software with the recommended parameters (Rausch et al. 2012). CNVnator (v0.3.3) software, a read-depth based method, was used for CNV calling for each sample (Rausch et al. 2012). The CNV calls were then filtered with *P*-value less than 0.01, zero mapping quality (q0) less than 0.05, and size greater than 1 kb. The gene copy numbers of each region were determined using the “-genotype” option of CNVnator. The SVs of 535 individuals were merged using VCFtools (v0.1.17) (Danecek et al. 2011). The CNVs were aggregated into CNV regions based on at least 1-bp overlap. Short SVs and CNVs detected in at least 20 individuals and minor allele frequency greater than 0.05 were considered as common SVs in 535 samples. The RepeatMasker (v4.0.9) software was used to detect repeat sequences by aligning the *cgigas\_uk\_roslin\_v1* genome sequences to Repbase library (v 20181026). Based on the sequence characteristic, the repeat sequences were further classified into different types, including simple repeat, satellite, low complexity, retroelements, RC/Helitron, DNA transposons, and others.

### Population Differentiation Between Fast-growing Strains and Wild Populations

To investigate association of genome structural variations with the artificial selection of the Pacific oyster, *Pi* ratios and *Fst* values were calculated between fast-growing strains and their wild populations using VCFtools. The whole genome was scanned with the sliding windows of 20 kb with 10-kb step size. The empirical cutoffs for the candidate windows were set as bottom 5% and top 5% for *Pi* ratios, and top 5% for *Fst* values, respectively (Dennis et al. 2017; He et al. 2019; Bertolotti et al. 2020). The overlapping candidate windows of *Pi* ratios and *Fst* values were detected using BEDtools software (Quinlan and Hall 2010).

### Functional Analysis of Candidate Genes Associated with SVs Under Selection

The association of SVs with genes or functional elements was identified using Annovar software according to

*cgigas\_uk\_roslin\_v1* reference genome annotation (Wang et al. 2010). The Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of the function genes were annotated using eggNOG-Mapper (Huerta-Cepas et al. 2017). GO term and KEGG pathway enrichment analyses of candidate genes were carried out using clusterProfiler R package (Yu et al. 2012). GO terms and KEGG pathways with more than two enrichment genes or background genes were retained.

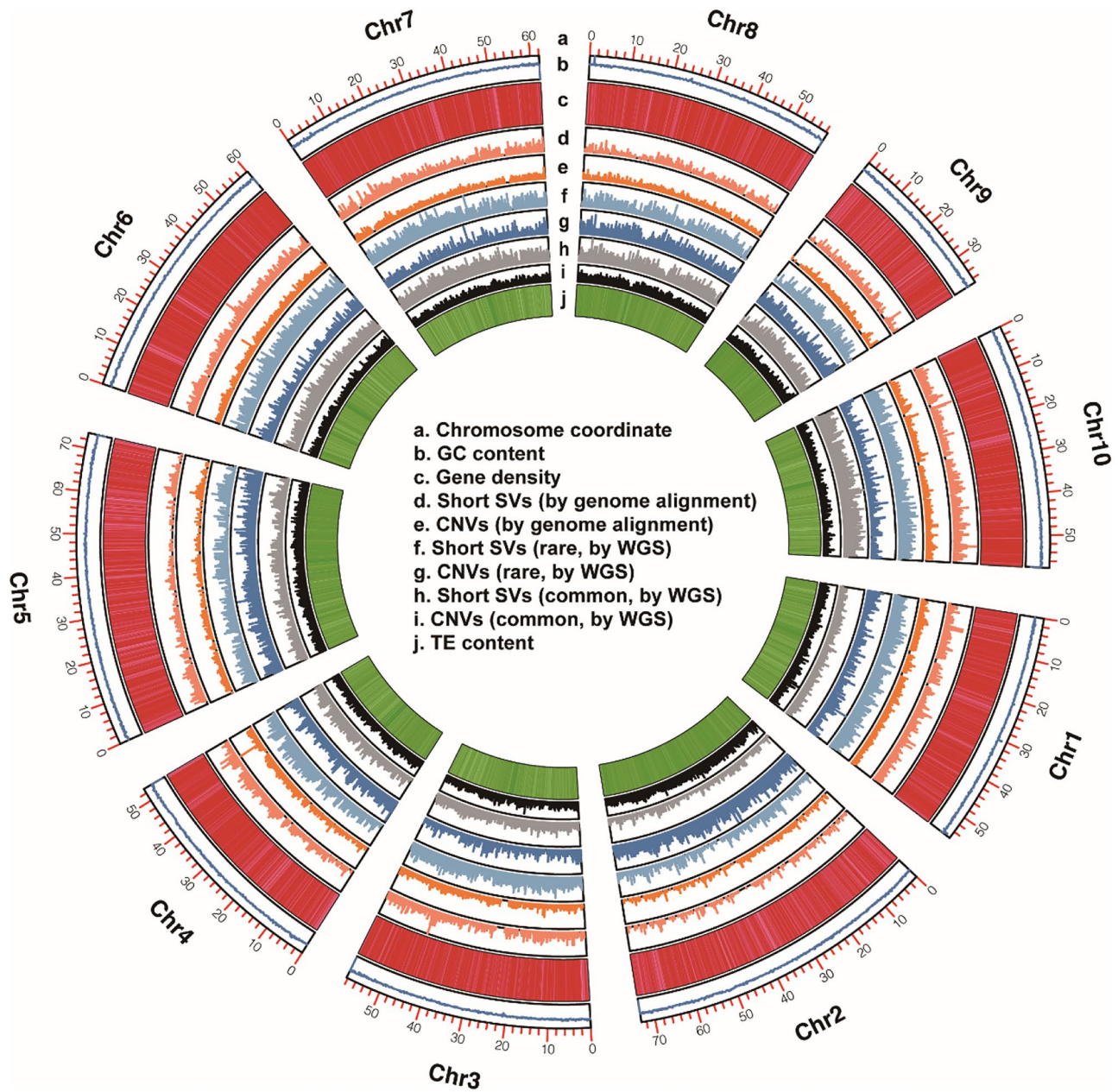
## Results

### Identification of SVs from Whole-Genome Alignment

Through the comparison of *cgigas\_uk\_roslin\_v1* with ASM1103280v1 genome assemblies, a total of 11,087 short SVs and 11,561 CNVs were identified across the chromosomes, including 2400 deletions, 3214 insertions, 9038 repeat contractions, 7394 repeat expansions, 21 tandem contractions, and 581 tandem expansions (Fig. 1a–e, Fig. 2a, Supplementary Table 1). There were distinct distribution patterns of short SVs across chromosomes with the density ranging from 6.51/Mb (NC\_047560.1) to 25.79/Mb (NC\_047568.1) (Supplementary Table 2). In contrast, the density of CNVs across chromosomes was between 10.13/Mb (NC\_047560.1) and 24.75/Mb (NC\_047568.1) (Supplementary Table 3). The frequencies of insertion and deletion in short SVs were decreased markedly with increased length, and the length of most short SVs (80%) varied from 50 to 647 bp (Fig. 2b). Similar pattern was observed in insertion, deletion, repeat expansion, and repeat contraction of CNVs, and 80% CNVs were with lengths ranging from 1 kb to 10,265 bp (Fig. 2b).

### SV Landscape Constructed Based on Whole-Genome Re-sequencing

In order to construct the comprehensive SV landscape of the Pacific oyster, whole-genome re-sequencing data from a total of 535 individuals sampled from 28 populations were analyzed, containing 8,224.61G clean reads with an average sequencing depth of 22× (Fig. 3a; Supplementary Table 4). Through standard analysis pipeline and strict thresholds, 220,468 short SVs and 13,176 CNVs were identified from each sample on average (Supplementary Table 5). It was obvious that the number of short SVs from each sample was greater than the CNVs, of which DEL was the most frequent SV type (Fig. 3b). Then, the short SVs and CNVs of each sample were merged and classified into two categories: rare (identified in less than 20 samples) and common (identified in at least 20 samples) variant sets, respectively. The size of rare short SV

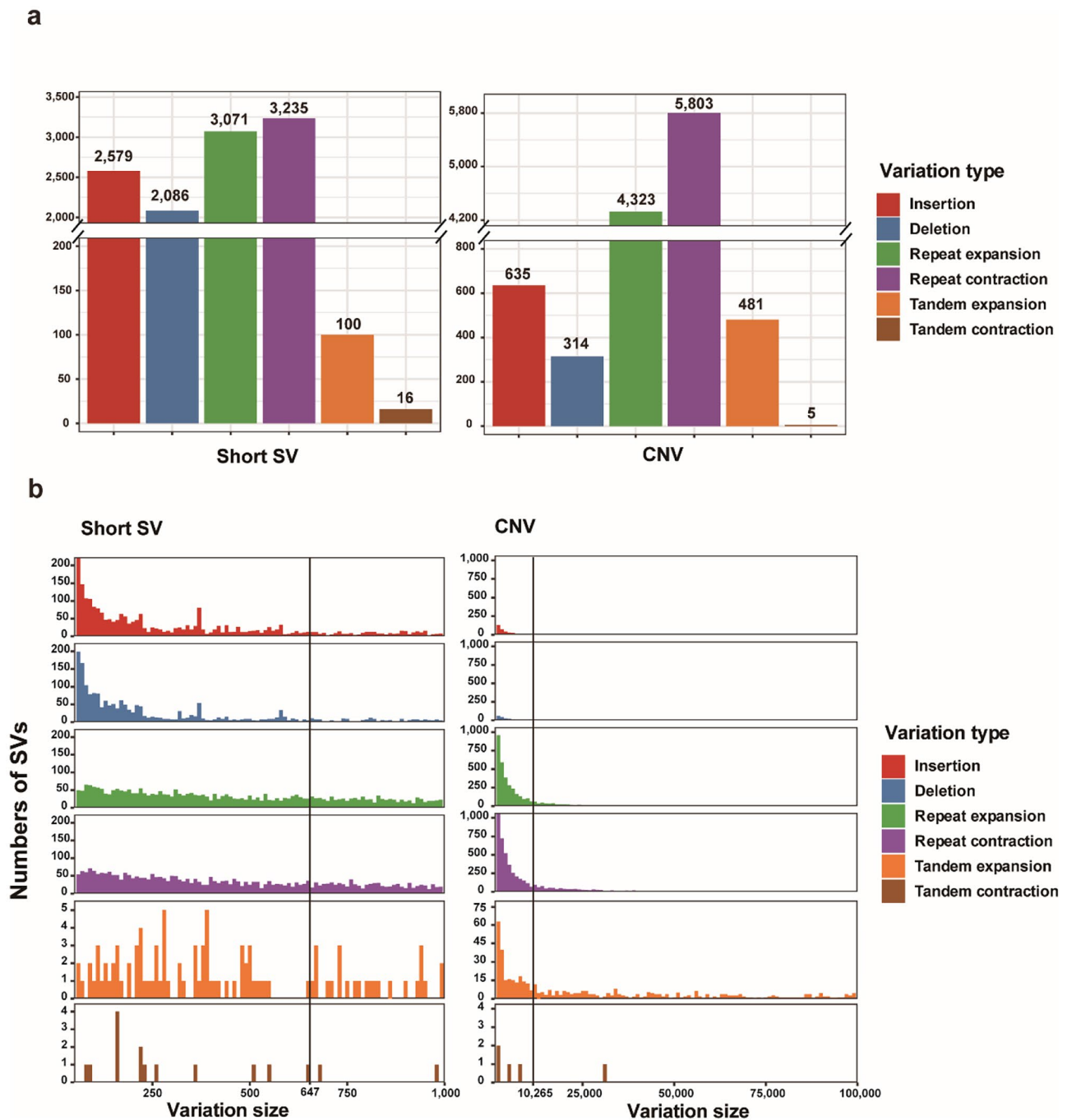


**Fig. 1** Genomic structural variation (SV) landscapes of the Pacific oyster. Tracks' information is listed as the following (outer to inner circles): **a** chromosome karyotype; **b** GC content; **c** gene density; **d** short SVs (length from 50 to 1000 bp); **e** CNVs (length over 1,000 bp); **f** rare short SVs (identified in less than 20 samples); **g** rare CNVs (identified in less than 20 samples); **h** common short SVs (identified in at least 20 samples); **i** common CNVs (identi-

fied in at least 20 samples); **j** TE content. The units of chromosome karyotype are set as 2 Mb, while the other information was calculated in 20-kb windows. Short SVs and CNVs (**d**, **e**) were obtained through the genome alignment between *cgigas\_uk\_roslin\_v1* (GCA\_902806645.1) and *ASM1103280v1* (GCA\_011032805.1). Rare and common short SVs and CNVs (**f**–**i**) were detected based on whole-genome re-sequencing data

set grew dramatically at the beginning, while the size of common set was shrunk rapidly (Fig. 3c). They gradually slowed down and approached the plateaus with the increased number of samples. Similar results were also observed in the rare and common CNV sets (Fig. 3d). The results demonstrated that a considerable proportion of short SVs and CNVs were

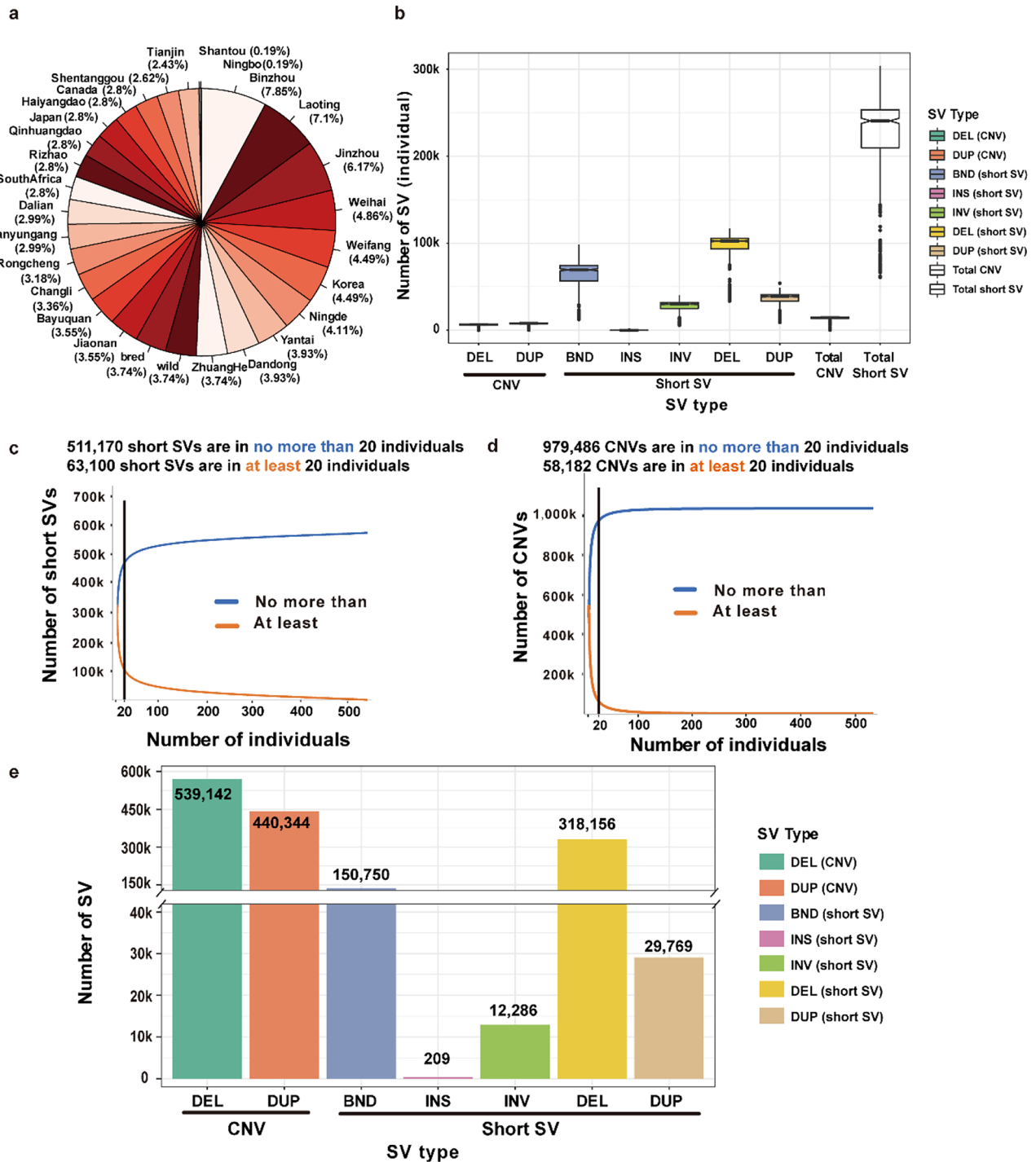
specific to some individuals or limited number of individuals. Together, 511,170 rare short SVs and 979,486 rare CNVs were identified, which were composed of 318,156 DEL (short SVs), 150,750 BND (short SVs), 29,769 DUP (short SVs), 12,286 INV (short SVs), 209 INS (short SVs), 539,142 DEL (CNVs), and 440,344 DUP (CNVs) (Fig. 3e).



**Fig. 2** The statistic of SVs detected by genome alignment. **a** The number of different short SVs and CNVs types. **b** The length distribution of different short SV and CNV types

A total of 63,100 short SVs and 58,182 CNVs were regarded as common variations among populations, including 30,009 DEL, 2,074 DUP, 725 INS, 613 INV, and 29,679 BND in short SV and 30,987 DEL and 27,195 DUP in CNV (Fig. 4a). The proportion of common BND was higher than that of the other short SV types, and DEL was the major type in common CNVs (Fig. 4a). These common short SVs and CNVs were annotated

across the cgigas\_uk\_roslin\_v1 reference genome to evaluate the potential functional impact on the genes. Around half of short SVs (50.4%, 31,806) were distributed in intronic regions, while only 1.5% (927) short SVs were related to exons. In contrast, a majority of CNVs (68.8%, 40,033) were overlapped with one or more exons. A total of 6977 (12.0%) and 5611 (9.6%) CNVs were located in intergenic and 5-kb upstream or downstream



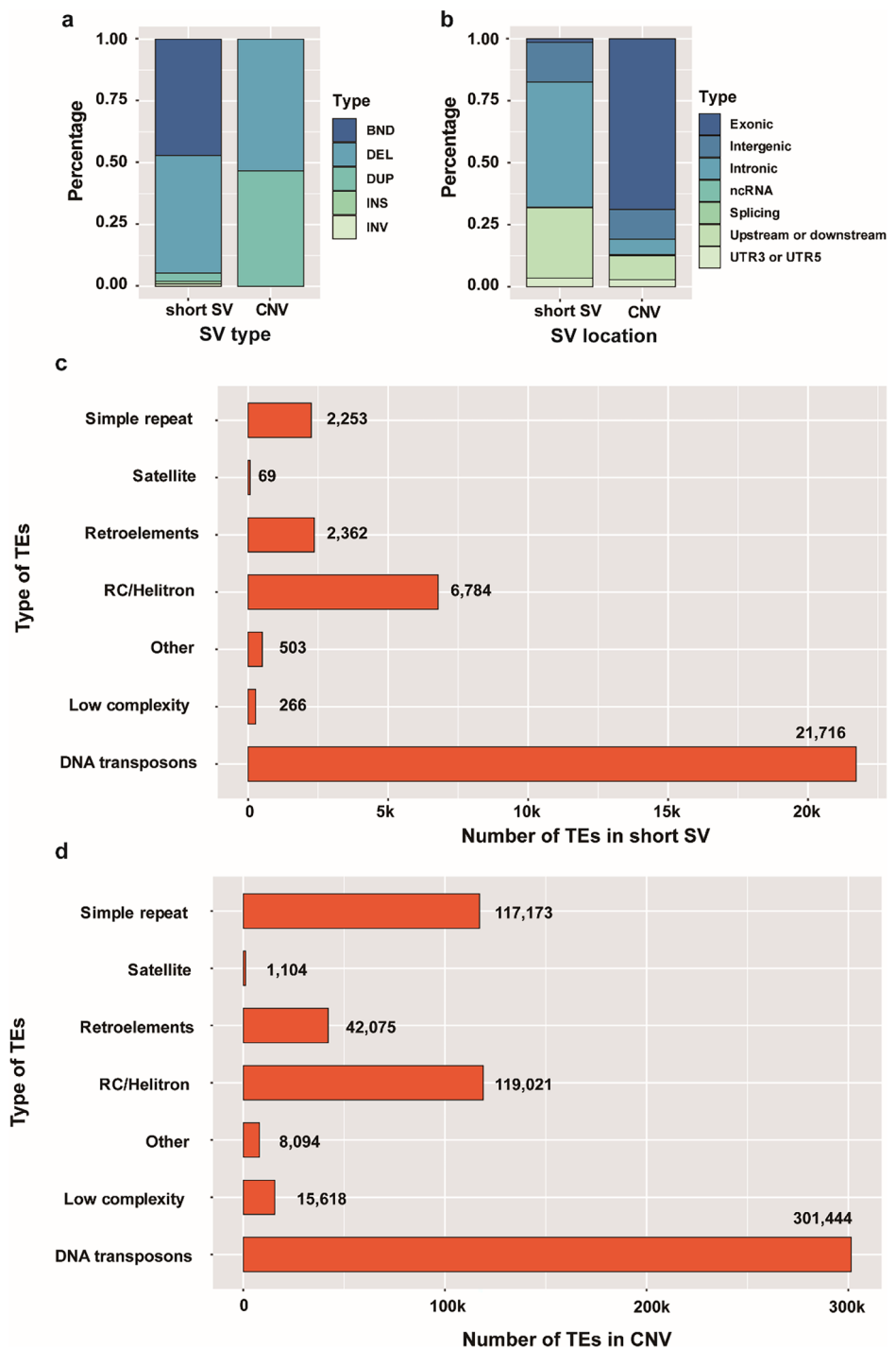
**Fig. 3** The information of SVs detected from the genotyping of whole-genome re-sequencing data. **a** The sample number and population information. **b** Number of different SV types. SV types include translocation (BND), deletion (DEL), duplication (DUP), insertion (INS), and inversion (INV). **c** The number of rare and common short

SVs. Rare SVs are identified in less than 20 samples, while the common are present in at least 20 samples. **d** The number of rare and common CNVs. Rare SVs are identified in less than 20 samples, while the common are present in at least 20 samples. **e** The numbers of different types of rare short SVs and CNVs

regions, respectively (Fig. 4b). In addition, we investigated the distribution of short SVs and CNVs across repeat sequences. Nearly 46.7% of common short SVs (29,460/63,100) and 88.4%

of CNVs (51,413/58,182) overlapped with the repeat elements, most of which are DNA transposon (50.1%), RC/Helitron (19.7%), and simple repeat (19.2%) (Fig. 4c, d).

**Fig. 4** The common SVs detected from the genotyping of 535 whole-genome re-sequencing samples. **a** The percentages of different short SV and CNV types. **b** Genomic annotation of short SVs and CNVs. **c** Distribution of repeat types in short SVs. **d** Distribution of repeat types in CNVs



## Identification of Selective Sweeps Under Artificial Selection

Common short SVs detected from the 20 oysters of the fast-growing strains and 20 individuals of the wild population were further analyzed to identify the selective sweeps underlying artificial selection. Both *F<sub>st</sub>* and *Pi* ratio statistical methods were employed to determine

population differentiation by comparing allele frequency and nucleotide diversity between artificially selected strains and wild population (Fig. 5a, b). Based on the empirical threshold of top 5% *F<sub>st</sub>* value ( $F_{st} > 0.1825$ ), 1373 windows were identified by *F<sub>st</sub>*, covering 23.05 Mb for 3.6% genome and containing 2022 functional genes. Meanwhile, a total of 3992 windows with top or bottom 5% of *Pi* ratio ( $Pi \text{ ratio} > 2.5858$  or  $Pi \text{ ratio} < 0.3286$ ) were

also identified, which covered 66.75 Mb (10.3% genome) and harboring 5534 genes. A total of 514 genomic regions (8.76 Mb) were identified by both approaches (green and blue dots in Fig. 5c), containing 746 candidate genes (Fig. 5d). The detailed information of the candidate regions and genes was provided in Supplementary Table 6. Notably, 61 common CNVs were specifically identified from the fast-growing strains rather than their wild populations. These CNVs contained 103 genes, which were further investigated for their potential association with growth trait (Supplementary Table 7).

### Functional Analysis of Candidate Genes from SVs Under Selection

The 746 genes identified from short SVs under selection and 103 genes identified from CNVs specifically detected in fast-growing strains, in total 843 genes, were subject to functional analysis with GO and KEGG pathway enrichment. The results of GO enrichment analysis revealed 294 GO terms with  $P < 0.05$  and the top 20 enriched GO terms, ranked by  $P$ -value, are shown in Fig. 6a. Among which, the most significantly enriched term was apical part of cell (GO:0045177). The KEGG enrichment analysis indicated that the most significantly enriched pathway was tryptophan metabolism (ko00380) (Fig. 6b), followed by histidine metabolism (ko00340) and vitamin digestion and absorption (ko04977). Biosynthesis of secondary metabolites pathway (ko01110) contained the most enriched genes and the highest enrich factor was detected in ascorbate and aldarate metabolism (ko00053). In addition, the specific biology functions of the candidate genes were also further investigated according to published studies and were classified into 10 functional groups (Supplementary Table 8), including tissue morphogenesis, organic compound metabolism, cell cycle, cellular component, ion and amino acid transport, neurogenesis and nerve-impulse transmission, protein modification, RNA processing, signal transduction, and immune response. The detailed information of all the GO and KEGG terms is provided in Supplementary Table 9.

### Discussion

Genome structural variations account for a major portion of genomic variations in an organism, which play critical roles in biological functions. In comparison with single nucleotide variations that are relatively well studied, genome structural variations remain largely unexplored due to the limitation of detection approach (Bertolotti et al. 2020; Qi et al. 2021). In the present study,

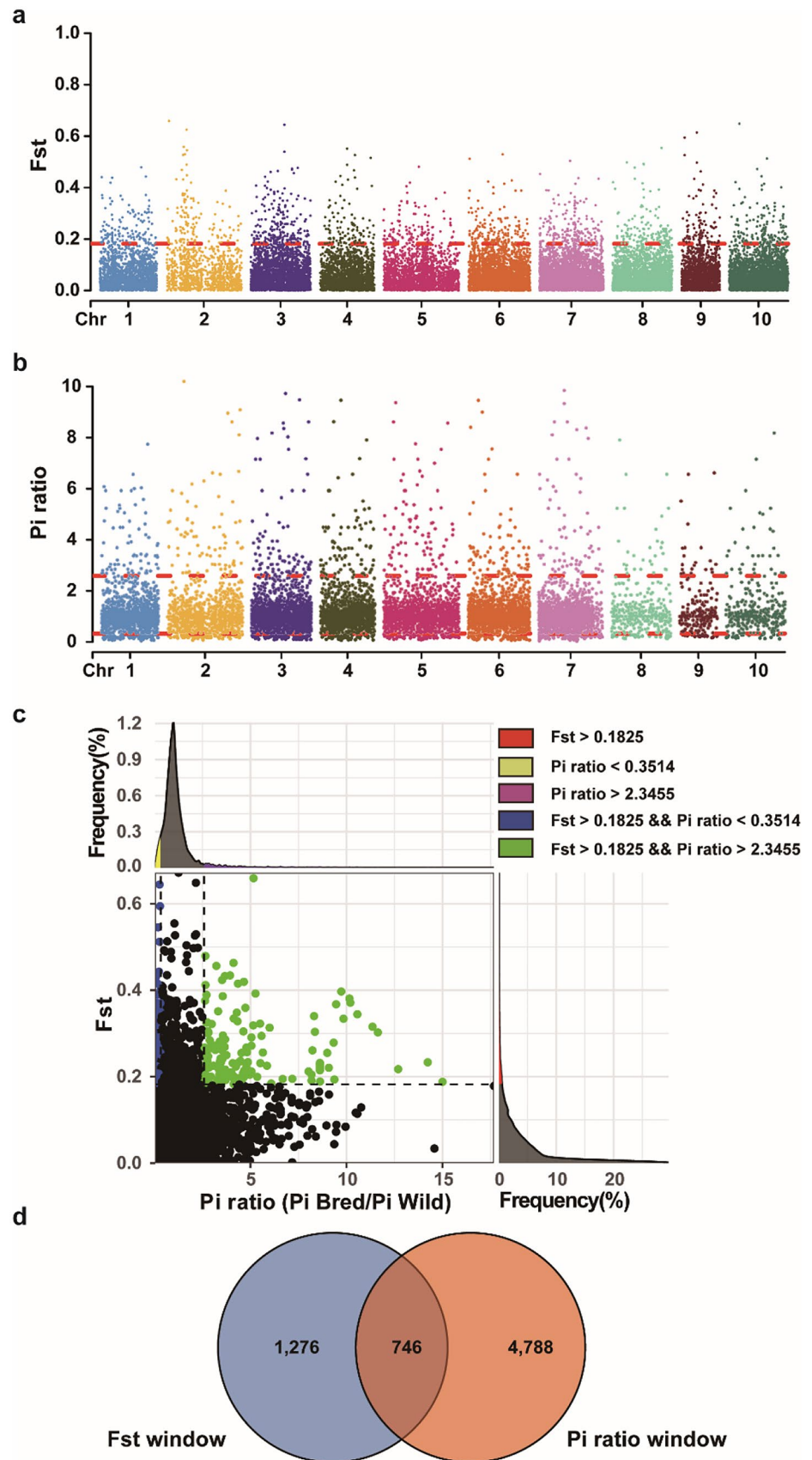
we constructed the first comprehensive genome structural variation landscape in the Pacific oyster by performing alignment of whole-genome assemblies and analyses of whole-genome re-sequencing data. We further detected the selection signatures of the SV landscape in the fast-growing strains which have undergone 10 generations of artificial selection for growth. This work provided valuable information for further investigations on genome evolution under selection in the oysters.

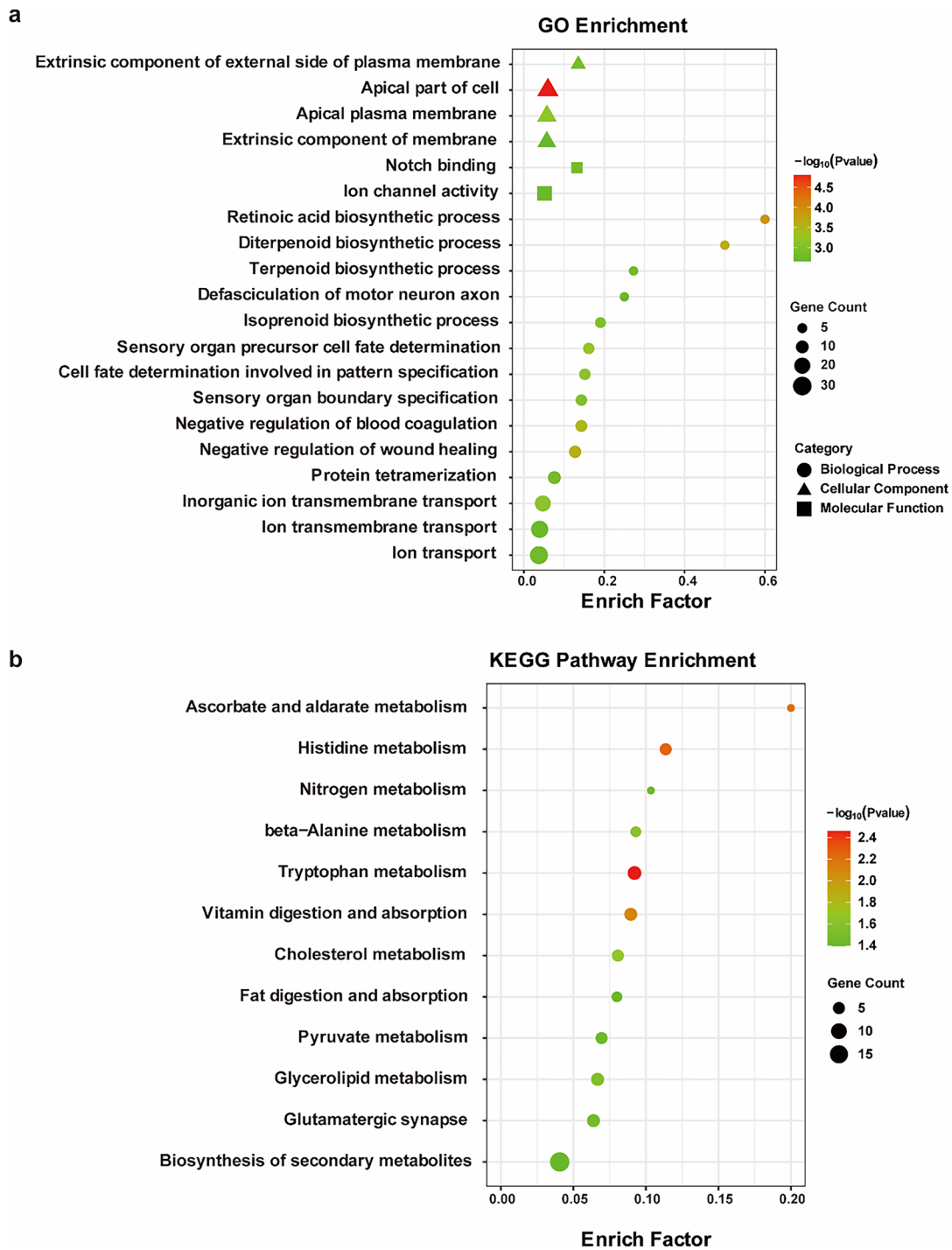
SVs are important source of genetic variations underlying important domestication traits (Vlad et al. 2010; Dorshorst et al. 2015; Dharmayanthi et al. 2017; Duan et al. 2017; Simam et al. 2018; Chakraborty et al. 2019; Liu et al. 2019). Many SVs have been identified and characterized to be closely associated with some domestication or artificial breeding traits, such as berry color of grapevine (*Vitis vinifera* ssp. *sativa*) (Zhou et al. 2019) and dietary shifts between grey wolf (*Canis lupus*) and dhole (*Cuon alpinus*) (Wang et al. 2019). However, there are few studies of SVs in aquatic organisms. The Pacific oyster with great value in economics has been successfully selected for over ten generations and obtained superior growth performance. Hence, it represents an excellent model to investigate the genetic basis of SVs in artificial breeding in the Pacific oyster.

The SVs are usually rare and with low frequency in the population. In this work, we performed selective sweep analysis based on the common SVs to ensure the reliability. In total, 843 genes were identified that were associated with short SVs under selection or related to CNVs that were detected only in the fast-growing strains in comparison with the wild populations. The GO enrichment analysis revealed that the candidate genes were significantly associated with the apical part of cell term. In GO resource and annotation, it is defined as the region of a polarized cell that forms a tip or is distal to a base with key genes such as *5'-AMP-activated protein kinase subunit beta-2* (LOC105344372), *angiominin* (LOC105333766), *fibroblast growth factor receptor 2* (LOC105321229), and *regulator of G-protein signaling 12* (LOC105343750). However, the specific role of apical part of cell and its association with growth requires future investigations. KEGG enrichment analysis demonstrated that the candidate genes were enriched in several metabolism-related pathways, such as tryptophan metabolism, histidine metabolism, vitamin digestion and absorption, and ascorbate and aldarate metabolism. Tryptophan is an indispensable and essential dietary amino acid for the regulation of growth and immune response in animals and plants (Walton et al. 1984; Le Floc'h et al. 2011; Fukuwatari and Shibata 2013; Hiruma et al. 2013). Histidine is another important amino acid and plays a critical role in growth and development



**Fig. 5** Selective sweep analysis based on short SVs. Manhattan plots showing  $F_{st}$  (a) and Pi ratio (b) values calculated based on short SVs between fast-growing strains and their corresponding wild populations. The red dotted lines represented the significant thresholds for  $F_{st}$  ( $>0.1825$ , top 5%) and Pi ratio ( $>2.5858$ , top 5% or  $<0.3286$ , bottom 5%) values. c Selective sweep regions selected by both  $F_{st}$  value and Pi ratio tests. d Venn graph showing the candidate genes within selective sweep regions





**Fig. 6** Functional enrichment of candidate genes. **a** The GO enrichment analysis of candidate genes. The top 20 enriched GO terms, ranked by *P*-value, were displayed in **a**. Biological process, molecular

function, and cellular component are visualized by different shapes. **b** KEGG pathway enrichment analysis of candidate genes

of animals and plants (Ingle 2011; Powell et al. 2011; Brosnan and Brosnan 2020). Vitamin digestion and absorption pathway is closely associated with the vitamin metabolism in organisms. Also, ascorbate and aldarate metabolism could be directly related to the biosynthesis,

recycling, and degradation of vitamin C (Linster and Van Schaftingen 2007). It is well documented that vitamins have great impact on the growth rate and stress resistance of aquatic animals (Sealey and Gatlin 2002; Kumari and Sahoo 2005; Dawood and Koshio 2018). Therefore, how

these candidates could influence the growth performance of the Pacific oyster through the regulation of amino acid and vitamin metabolism deserves future investigation.

The classifications of candidate genes related to SV indicate that the biological traits appear to be very complex, whose regulation is involved in diverse biological processes. These SV variations of the candidate genes may explain the differences between fast-growing strains and wild populations of oysters. It will be useful to combine SV with the functional genes to investigate SVs' function. Tissue morphogenesis could play an essential role in growth trait. A lot of genes related to tissue morphogenesis were found in high differential SV windows. For example, *titin* (LOC105328178) is essential in the temporal and spatial control of the assembly of striated muscles during myofibrillogenesis (Mayans et al. 1998). A total of 11 *multiple epidermal growth factor-like domains protein 10 (MEGF10)* genes lie within the selective sweep regions or located in specific CNVs of fast-growing strains. *MEGF10* is able to interact with Notch1 via their respective intracellular domains, playing vital roles in myogenesis (Takayama et al. 2016; Saha et al. 2017). As reported, *fibrillins* are important components of microfibril networks, which could interact with members of the *TGF- $\beta$*  growth factors family to take part in the tissue morphogenesis (Charbonneau et al. 2004; Gansner et al. 2008; Sengle et al. 2008; Ono et al. 2009). A total of four isoforms of *fibrillins* were identified, all having an evolutionarily conserved domain organization (Gansner et al. 2008; Jensen and Handford 2016). Three of which, including *fibrillin-1*, *fibrillin-2*, and *fibrillin-3*, have been regarded as candidates that could be associated with growth in the Pacific oysters.

The other classifications are organic compound metabolism and ion and amino acid transport which show the strong connection between SVs and biological traits. The protein encoded by *5'-AMP-activated protein kinase subunit beta-2* gene could influence the activity of AMP-activated protein kinase, an energy sensor protein kinase that plays a key role in regulating cellular energy metabolism by changing the rates of glucose uptake and fatty acid oxidation (Dyck et al. 1996; Winder and Thomson 2007). *Solute carrier family 15 member 5 (SLC15A5)*, a sodium-coupled citrate transporter, could import citrate from the circulation into cells. Recently, emerging evidence suggested the importance of *SLC15A5* in energy homeostasis, which could facilitate the utilization of circulating citrate for the generation of metabolic energy and for the synthesis of fatty acids and cholesterol (Inoue et al. 2002; Hardies et al. 2015; Li et al. 2017). In addition, a large number of genes related to immune response were also included in the candidate list, suggesting that immunity of fast-growing strains of the Pacific oysters is also largely shaped by artificial selection.

## Conclusion

In the present study, we constructed the first comprehensive landscape of genome structural variations in the Pacific oyster. Further analysis of the SVs that have been affected by artificial selection were performed. The *Fst* and *Pi* ratio tests revealed that 514 genomic regions (8.76 Mb), containing 746 candidate genes, were under artificial selection and could be associated with growth traits. Another 103 candidate genes were identified from the 61 common CNVs that were only present in the fast-growing strains. Functional analysis of the total 843 candidates revealed its enrichment in apical part of cell term and several metabolism-related pathways, including tryptophan metabolism and histidine metabolism. Taken together, this work provided a comprehensive landscape of SVs and revealed their responses to selection, which will be valuable for further investigations on genome evolution under selection in the oysters.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10126-021-10060-5>.

**Funding** This work was supported by the grants from the National Natural Science Foundation of China (Nos. 31802293, 41976098, and 31741122) and the Young Talent Program of Ocean University of China (No. 201812013).

## Declarations

**Conflict of Interest** The authors declare no competing interests.

## References

- Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363–376
- Alonge M, Wang X, Benoit M et al (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182:145–161.e23
- Bertolotti AC, Layer RM, Gundappa MK et al (2020) The structural variation landscape in 492 Atlantic salmon genomes. *Nat Commun* 11:5176
- Brosnan ME, Brosnan JT (2020) Histidine metabolism and function. *J Nutr* 150:2570S–2575S
- Chakraborty M, Emerson JJ, Macdonald SJ, Long AD (2019) Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat Commun* 10:4872
- Charbonneau NL, Ono RN, Corson GM et al (2004) Fine tuning of growth factor signals depends on fibrillin microfibril networks. *Birth Defects Res Part C Embryo Today Rev* 72:37–50
- Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890
- Chiang C, Scott AJ, Davis JR et al (2017) The impact of structural variation on human gene expression. *Nat Genet* 49:692–699
- Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158

- Dawood MAO, Koshio S (2018) Vitamin C supplementation to optimize growth, health and stress resistance in aquatic animals. *Rev Aquac* 10:334–350
- de Melo CMR, Durland E, Langdon C (2016) Improvements in desirable traits of the Pacific oyster, *Crassostrea gigas*, as a result of five generations of selection on the West Coast, USA. *Aquaculture* 460:105–115
- Dennis MY, Harshman L, Nelson BJ et al (2017) The evolution and population diversity of human-specific segmental duplications. *Nat Ecol Evol* 1:69
- Dharmayanthi AB, Terai Y, Sulandari S et al (2017) The origin and evolution of fibromelanosis in domesticated chickens: genomic comparison of Indonesian Cemani and Chinese Silkie breeds. *PLoS One* 12:e0173147
- Dorshorst B, Harun-Or-Rashid M, Bagherpoor AJ et al (2015) A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two Duplex-comb phenotypes. *PLoS Genet* 11:e1004947
- Duan P, Xu J, Zeng D et al (2017) Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. *Mol Plant* 10:685–694
- Dyck JRB, Gao G, Widmer J et al (1996) Regulation of 5'-AMP-activated protein kinase activity by the noncatalytic  $\beta$  and  $\gamma$  subunits. *J Biol Chem* 271:17798–17803
- Evans S, Langdon C (2006) Direct and indirect responses to selection on individual body weight in the Pacific oyster (*Crassostrea gigas*). *Aquaculture* 261:546–555
- Fukuwatari T, Shibata K (2013) Nutritional aspect of tryptophan metabolism. *Int J Tryptophan Res* 6:3–8
- Gansner JM, Madsen EC, Mecham RP, Gitlin JD (2008) Essential role for fibrillin-2 in zebrafish notochord and vascular morphogenesis. *Dev Dyn* 237:2844–2861
- Guan P, Sung WK (2016) Structural variation detection using next-generation sequencing data: a comparative technical review. *Methods* 102:36–49
- Hardies K, De Kovel CGF, Weckhuysen S et al (2015) Recessive mutations in SLC13A5 result in a loss of citrate transport and cause neonatal epilepsy, developmental delay and teeth hypoplasia. *Brain* 138:3238–3250
- He Y, Luo X, Zhou B et al (2019) Long-read assembly of the Chinese rhesus macaque genome and identification of ape-specific structural variants. *Nat Commun* 10:4233
- Hiruma K, Fukunaga S, Bednarek P et al (2013) Glutathione and tryptophan metabolism are required for Arabidopsis immunity during the hypersensitive response to hemibiotrophs. *Proc Natl Acad Sci USA* 110:9589–9594
- Huerta-Cepas J, Forslund K, Coelho LP et al (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34:2115–2122
- Hurles ME, Dermitzakis ET, Tyler-Smith C (2008) The functional impact of structural variation in humans. *Trends Genet* 24:238–245
- Ingle RA (2011) Histidine biosynthesis. *Arabidopsis Book* 9:e0141
- Inoue K, Zhuang L, Ganapathy V (2002) Human Na<sup>+</sup>-coupled citrate transporter: primary structure, genomic organization, and transport function. *Biochem Biophys Res Commun* 299:465–471
- Jensen SA, Handford PA (2016) New insights into the structure, assembly and biological roles of 10–12 nm connective tissue microfibrils from fibrillin-1 studies. *Biochem J* 473:827–838
- Jin YL, Kong LF, Yu H, Li Q (2014) Development, inheritance and evaluation of 55 novel single nucleotide polymorphism markers for parentage assignment in the Pacific oyster (*Crassostrea gigas*). *Genes and Genomics* 36:129–141
- Kim T, Seo HD, Hennighausen L et al (2018) Octopus-toolkit: a workflow to automate mining of public epigenomic and transcriptomic next-generation sequencing data. *Nucleic Acids Res* 46:e53
- Kong L, Bai J, Li Q (2014) Comparative assessment of genomic SSR, EST-SSR and EST-SNP markers for evaluation of the genetic diversity of wild and cultured Pacific oyster *Crassostrea gigas* Thunberg. *Aquaculture* 420–421:S85–S91
- Kosugi S, Momozawa Y, Liu X et al (2019) Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol* 20:117
- Kumari J, Sahoo PK (2005) High dietary vitamin C affects growth, non-specific immune responses and disease resistance in Asian catfish, *Clarias batrachus*. *Mol Cell Biochem* 280:25–33
- Kurtz S, Phillippy A, Delcher AL et al (2004) Versatile and open software for comparing large genomes *Genome Biol* 5:R12
- Le Floch N, Otten W, Merlot E, (2011) Tryptophan metabolism, from nutrition to potential therapeutic applications. *Amino Acids* 41:1195–1205
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
- Li L, Li A, Song K et al (2018) Divergence and plasticity shape adaptive potential of the Pacific oyster. *Nat Ecol Evol* 2:1751–1760
- Li Q, Wang Q, Liu S, Kong L (2011) Selection response and realized heritability for growth in three stocks of the Pacific oyster *Crassostrea gigas*. *Fish Sci* 77:643–648
- Li Q, Yu H, Yu R (2006) Genetic variability assessed by microsatellites in cultured populations of the Pacific oyster (*Crassostrea gigas*) in China. *Aquaculture* 259:95–102
- Li X, Yang J, Shen M et al (2020) Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun* 11:2815
- Li Z, Li D, Choi EY et al (2017) Silencing of solute carrier family 13 member 5 disrupts energy homeostasis and inhibits proliferation of human hepatocarcinoma cells. *J Biol Chem* 292:13890–13901
- Linster CL, Van Schaftingen E (2007) Vitamin C: biosynthesis, recycling and degradation in mammals. *FEBS J* 274:1–22
- Liu M, Zhou Y, Rosen BD et al (2019) Diversity of copy number variation in the worldwide goat population. *Heredity (edinh)* 122:636–646
- MacDonald JR, Ziman R, Yuen RKC et al (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42:986–992
- Mayans O, van der Ven PFM, Wilm M et al (1998) Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* 395:863–869
- Nattestad M, Schatz MC (2016) Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* 32:3021–3023
- Ono RN, Sengle G, Charbonneau NL et al (2009) Latent transforming growth factor  $\beta$ -binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites. *J Biol Chem* 284:16872–16881
- Powell S, Bidner TD, Payne RL, Southern LL (2011) Growth performance of 20- to 50-kilogram pigs fed low-crude-protein diets supplemented with histidine, cystine, glycine, glutamic acid, or arginine. *J Anim Sci* 89:3643–3650
- Qi H, Li L, Zhang G (2021) Construction of a chromosome-level genome and variation map for the Pacific oyster *Crassostrea gigas*. *Mol Ecol Resour* 21:1670–1685
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842
- Rausch T, Zichner T, Schlattl A et al (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28:333–339
- Saha M, Mitsuhashi S, Jones MD et al (2017) Consequences of MEGF10 deficiency on myoblast function and Notch1 interactions. *Hum Mol Genet* 26:2984–3000

- Sealey WM, Gatlin DM (2002) Dietary vitamin C and vitamin E interact to influence growth and tissue composition of juvenile hybrid striped bass (*Morone chrysops* ♀ × *M. saxatilis* ♂) but have limited effects on immune responses 1,2. *J Nutr* 132:748–755
- Sengle G, Charbonneau NL, Ono RN et al (2008) Targeting of bone morphogenetic protein growth factor complexes to fibrillin. *J Biol Chem* 283:13874–13888
- Simam J, Rono M, Ngoi J et al (2018) Gene copy number variation in natural populations of *Plasmodium falciparum* in Eastern Africa. *BMC Genomics* 19:372
- Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. *Annu Rev Med* 61:437–455
- Takayama K, Mitsuhashi S, Shin JY et al (2016) Japanese multiple epidermal growth factor 10 (MEGF10) myopathy with novel mutations: a phenotype–genotype correlation. *Neuromuscul Disord* 26:604–609
- Troost K (2010) Causes and effects of a highly successful marine invasion: case-study of the introduced Pacific oyster *Crassostrea gigas* in continental NW European estuaries. *J Sea Res* 64:145–165
- Vacic V, McCarthy S, Malhotra D et al (2011) Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* 471:499–503
- Vlad D, Rappaport F, Simon M, Loudet O (2010) Gene transposition causing natural variation for growth in *Arabidopsis thaliana*. *PLoS Genet* 6:e1000945
- Walton MJ, Coloso RM, Cowey CB et al (1984) The effects of dietary tryptophan levels on growth and metabolism of rainbow trout (*Salmo gairdneri*). *Br J Nutr* 51:279–287
- Wang GD, Shao XJ, Bai B et al (2019) Structural variation during dog domestication: insights from gray wolf and dhole genomes. *Nat Sci Rev* 6:110–122. <https://doi.org/10.1093/nsr/nwy076>
- Wang J, Li Q (2017) Characterization of novel EST-SNP markers and their association analysis with growth-related traits in the Pacific oyster *Crassostrea gigas*. *Aquac Int* 25:1707–1719
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164
- Weischenfeldt J, Symmons O, Spitz F, Korbel JO (2013) Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 14:125–138
- Winder WW, Thomson DM (2007) Cellular energy sensing and signaling by AMP-activated protein kinase. *Cell Biochem Biophys* 47:332–347
- Xu L, Cole JB, Bickhart DM et al (2014) Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics* 15:683
- Yang N, Liu J, Gao Q et al (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat Genet* 51:1052–1059
- Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi A J Integr Biol* 16:284–287
- Zhao X, Yu H, Kong L, Li Q (2012) Transcriptomic responses to salinity stress in the Pacific oyster *Crassostrea gigas*. *PLoS One* 7:e46244
- Zhong X, Li Q, Yu H, Kong L (2013) Development and validation of single-nucleotide polymorphism markers in the Pacific Oyster, *Crassostrea gigas*, using high-resolution melting analysis. *J World Aquac Soc* 44:455–465
- Zhou Y, Minio A, Massonnet M et al (2019) The population genetics of structural variants in grapevine domestication. *Nat Plants* 5:965–979

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.