Gene 667 (2018) 34–44

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Ancient horizontally transferred genes in the genome of California two-spot octopus, *Octopus bimaculoides*

Conghui Liu*, Bo Liu, Yan Zhang, Fan Jiang, Yuwei Ren, Shuqu Li, Hengchao Wang, Wei Fan*

Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China

ARTICLE INFO

Keywords:
Horizontally transferred genes
Octopus bimaculoides
Bacteria
Zn-metalloproteinase
Negative selection

ABSTRACT

Horizontal gene transfer (HGT), a mechanism that shares genetic material between the host and donor from separated offspring branches, has been described as a means of producing novel and beneficial phenotypes for the host organisms. However, in molluscs, the second most diverse group, the existence of HGT is still controversial. In the present study, 12 HGT genes were identified from California two-spot octopus *Octopus bimaculoides* based on a similarity search, phylogenetic construction, gene composition analysis and PCR (Polymerase Chain Reaction) validation. Based on the phylogenetic topologies, ten HGT genes were identified to have been transferred into the possible molluscan ancestor, possibly before its radiation. Furthermore, most of the donor organisms were predicted to be familiar bacteria in marine environments. These horizontally transferred genes were under a strong negative selection and could be transcribed in octopus functionally. The predicted biochemical functions of these genes include metabolism, neurotransmission, immune defense and tissue integrity. Seven Zn-metalloproteinases were validated as the main type of HGT genes in octopus with divergent motif composition, intron presence and phylogenetic relationship to the endogenous ones. Furthermore, the functions of Zn-metalloproteinase were predicted to be responsible for immune defense and tissue remolding. Three HGT genes were distributed mainly in the nervous system and were predicted to regulate the neurotransmission through glia–neuronal interactions. The results collectively indicated the existence of HGT in molluscs and its potential contribution to the evolution of octopus with regards to functional innovation and adaptability.

1. Introduction

The race between the microbial community and the host organisms has been depicted for centuries, with the understanding of host–microbe interaction also changing rapidly and tending to be more comprehensive (Casadevall and Pirofski, 1999; Dethlefsen et al., 2007). Apart from the previous knowledge that microbes act as primary aggressors for the host diseases, commensal microorganisms or mutualistic microorganisms were reported to benefit the host or both partners (Casadevall and Pirofski, 2000; Eckburg et al., 2005). At the genomic level, horizontal gene transfer (HGT), also known as lateral gene transfer, has been described as a mechanism that shares genetic material between the host and microbe from separated offspring branches to produce beneficial phenotypes (Keeling and Palmer, 2008; Soucy et al., 2015). Across normal mating barriers, HGT was long considered to be prevalent and ubiquitous in the evolution of prokaryotic genomes, especially bacteria and archaea (Boto, 2010). HGT provides the recipient with novel functions, such as ecological plasticity, pathogenic capacity or metabolic characteristics (Ochman et al., 2000). Compared with prokaryotes, relatively few HGT events have occurred in eukaryotes because of nuclear membrane and sexual reproduction (Andersson, 2005; Keeling and Palmer, 2008). However, accumulating evidence suggested that HGT events between prokaryotes and eukaryotes or even among eukaryotes appear more frequently than previously thought (Boto, 2014; Huang and Yue, 2013).

Due to frequent contact with prokaryotic DNA, multiple eukaryotes have acquired adaptively important traits through gene exchange, regardless of the evolutionary distance between the two species (Andersson et al., 2006; Richards et al., 2003). Among all the eukaryotes, HGT events within unicellular eukaryotes, especially parasitic protozoa, have been well-documented, compared with that of multicellular eukaryotes (Deitsch et al., 2001; Doolittle and Logsdon Jr, 1998). In multicellular eukaryotes, gene exchange is relatively difficult. However, foreign genes could enter the multicellular organism through

Abbreviations: HGT, horizontal gene transfer; EB, ethidium bromide; FPKM, fragments per kilobase of exon per million fragments mapped; d_N/d_S, nonsynonymous or synonymous substitutions; ML, maximum likelihood method; NJ, neighbor-joining method; BI, Bayesian inference method; D-AL, D-aminoacylase; PGS, endo alpha-1,4-polygalactosaminidase; SRL, serine racemase-like; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology

* Corresponding authors.
E-mail addresses: liuconghui@caas.cn (C. Liu), fanwei@caas.cn (W. Fan).

https://doi.org/10.1016/j.gene.2018.05.013
Received 25 November 2017; Received in revised form 10 April 2018; Accepted 2 May 2018
Available online 05 May 2018
0378-1119/

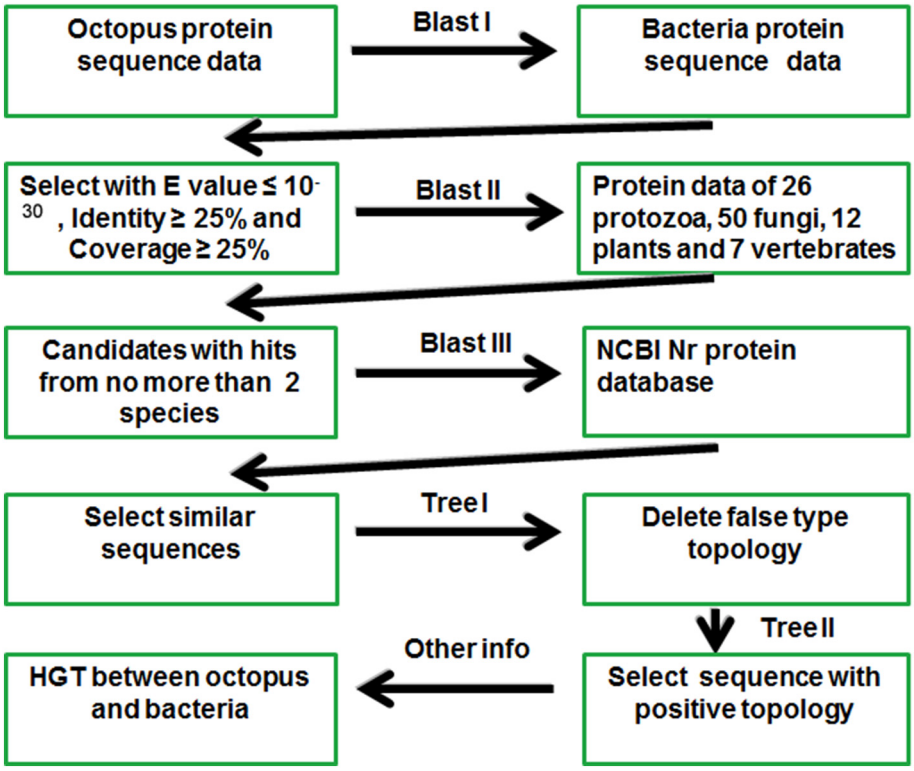

Fig. 1. A flowchart view of the HGT determination process. It was composed by three rounds of BLAST alignment and two rounds of phylogenetic analysis.

endosymbiosis (Timmis et al., 2004), feeding activity (Doolittle, 1998) or the weak-link model (Huang, 2013). Recently, HGT has been widely found in plants (Hannaert et al., 2003; Richards et al., 2006), invertebrates (Drezen et al., 2016) and vertebrates (Zhang et al., 2014). With the spread of inherited bacterial endosymbionts (Moran and Baumann, 2000), it is relatively easier for the invertebrate genome to obtain bacterial DNA because the intimacy of their relationships (Drezen et al., 2016). For example, a great number of acquired genes originated from microbes were revealed within the genome of arthropod (Hotopp et al., 2007), nematomorpha (Danchin et al., 2016; Danchin et al., 2010), rotifera (Eyres et al., 2015) and spongia (Conaco et al., 2016) so far. In molluscs, the second most diverse group, although an intriguing report evidenced the transfer from bacteriophages to bivalve genomes, there remain fragmental reports related to the HGT (Ren et al., 2017). In eastern emerald elysia, PCR experiments have suggested the presence of genes derived from the food source algae, *Vaucheria litorea*, but further transcriptome analyses revealed opposite results (Boto, 2014; Rumpho et al., 2008; Bhattacharya et al., 2013). Additionally, for marine invertebrates, the marine microbial population was reported to have significantly high genomic flexibility and HGT frequency (McDaniel et al., 2010), thus resulting in HGT also having been reported in shrimp (Yuan et al., 2013) and sponge (Conaco et al., 2016). For example, 14 genes with HGT events were identified from white shrimp, *Litopenaeus vannamei*. These originated from bacteria and fungi, functioning in energy metabolism and defense (Yuan et al., 2013). In the genome of the sponge, *Amphimedon queenslandica*, 227 horizontally acquired genes were sifted out and were found to contribute to sponge adaptability (Conaco et al., 2016). Thus far, a considerable amount of information about the HGT events and acquired genes has been described from vertebrates to invertebrates. However, information about the scale and significance of HGT is still deficient in marine invertebrates.

As one of the most characteristic molluscs, the California two-spot octopus, *Octopus bimaculoides*, shows the remarkable and largest nervous system among invertebrates in addition to striking morphological innovations (Hanlon and Messenger, 1996; Young, 1971). Recently, the genome database of *O. bimaculoides* has been released and 33,638 protein-coding genes have been reported (Albertin et al., 2015). Notably, this octopus is known to harbor endosymbionts of bacteria (Small and McFall-Ngai, 1999) and a number of prokaryote and eukaryote parasites, including even human pathogens (Abollo et al., 2001; Fidalgo et al., 2000; Fraj and Duce, 1998). The intimate communication between this octopus and microbes influences the development and growth status of the octopus (Malham et al., 2002) in addition to possibly enhancing the chance to horizontally transfer the genetic materials during evolutionary history (Beiko et al., 2005). The investigation of HGT may lead to a better understanding of the scale and significance of HGT between bacteria and eukaryotes as well as the contribution of HGT to the development process of specialized body plan of the octopus. The present study validates a number of horizontal transferred genes from California two-spot octopus with the main objectives: (1) to discover the existence of HGT from bacteria to molluscs; (2) to investigate their mRNA tissue distributions of horizontal transferred genes and survey their contribution to the adaption of octopus; and (3) to survey their characteristics in the evolution of molluscs.

2. Materials and methods

2.1. Determination of HGT based on BLAST search and phylogenetic analysis

O. bimaculoides genome sequences were downloaded from the National Center for Biotechnology Information (NCBI, v2_0 version, GCA_001194135.1) and 33,638 protein-coding genes were employed in the present study (Albertin et al., 2015). The bacteria sequences of 2,774 species were also collected from the NCBI ftp site (Supplementary File 1). Additionally, genome sequences of 26 protozoa, 50 fungi, 12 plants and 7 vertebrates were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.jp/kegg/) database (Supplementary File 2).

The HGT determination process was performed according to a previous report with modifications (Li et al., 2011), composed by three rounds of BLAST alignment and two rounds of phylogenetic analysis (Fig. 1). The BLASTP search was performed to detect similar protein sequences between *O. bimaculoides* and the local database constructed by bacteria with an E value ≤ 10^{-30}, coverage value ≥ 25% and identity value ≥ 25%. Following this, the BLASTP program with the same threshold was employed to estimate the distribution spectrum of sifted similar genes in 26 protozoa, 50 fungi, 12 plants and 7 vertebrates. The candidate genes with similar genes from 2 or more species were rejected. Following this, the sifted genes were adopted to BLASTP research against NCBI non-redundant (NR) protein database with an E value ≤ 10^{-3}, coverage value ≥ 30% and identity value ≥ 30%. Phylogenetic analysis was composed with two steps. We used MUSCLE 3.8.31 (http://www.drive5.com/muscle/) and FastTree (http://www.microbesonline.org/fasttree/) in the first step to construct a Maximum likelihood (ML) tree. After this, CLUSTALX 2.0 (http://www.clustal.org) and MEGA 7.0 (http://www.mega.co.nz) were used based on genes selected in the first step for the NJ and ML trees reconstruction. The phylogenetic trees were select based on the phylogenetic topology patterns reported by Stanhope (Stanhope et al., 2001). After the second tree construction analysis, octopus genes with explicit topologies of HGT type were considered as the candidate sequences (Table 1).

2.2. Detection of HGT donor and codon usage bias

Based on the NJ and ML trees constructed in the second round, the bacterial species joining into a cluster with the HGT genes was determined as the candidate donor. If two or more species of bacteria were clustered with the candidate genes, the bacterium with the top BLAST score was considered as the donor.

The correspondence analysis of codon usage bias, GC rates and scaffold length was carried out to measure the degree of adaptation in the octopus HGT genes and the predicted bacteria donors. Codon usage analysis was performed using CodonW (http://codonw.sourceforge.net), and a primary orthogonal axis representing the greatest variation within the data was employed in the correspondence analysis (Supplementary File 3).

2.3. PCR validation of HGT genes

Adult octopuses were collected from a local market in Shenzhen, Guangdong Province, China, and maintained in aerated fresh seawater at 20 ± 2 °C for a week before processing. Before sampling, the octopuses were washed by the sterile sea water and incubated in 75% alcohol for 1 min. Total RNA was isolated from octopus hepatopancreas using Trizol reagent (TaKaRa) following its protocol. The first strand cDNA synthesis was carried out based on Promega M-MLV RT Usage information using the DNase I (Promega)-treated total RNA as a template and oligo (dT)-adaptor as the primer. The reaction was performed at 42 °C for 1 h, terminated by heating at 95 °C for 5 min. The cDNA sequence fragments of HGT genes were cloned by PCR with primers (Supplementary File 4). Following this, after detection by agarose gel electrophoresis, the PCR products were sequenced.

2.4. Digital gene expression profiling analysis

Twelve transcriptomes of *O. bimaculoides* were collected in this study from the NCBI SRA database, including the ovary (Ova, SRX1044303), testes (Tes, SRX1044305), viscera (Vis, SRX1045423), posterior salivary gland (PSG, SRX1045409), suckers (Suc, SRX1045411), skin (Ski, SRX1045410), developmental stage 15 (Sta, SRX1045413; Boletzky, 1989), retina (Ret, SRX1045414), optic lobe (OL, SRX1045415), supra-esophageal brain (Sup, SRX1045432), sub-esophageal brain (Sub, SRX1045435) and axial nerve cord (ANC, SRX1045418). Clean reads were mapped on the genome, using TopHat 2.1.1 (Kim et al., 2013). Following this, mapped reads were calibrated and quantified through Cufflinks 2.2.1 (Trapnell et al., 2010). FPKM values (fragments per kilobase of exon per million fragments mapped) were calculated to measure the expression level of each transcript. The expression level was displayed by heatmap plotted by R's package ggplot2 (Wickham, 2011).

2.5. Gene ontology analysis

Gene ontology (GO) analysis was adopted to provide the functional annotation of the sifted HGT genes (Ashburner et al., 2000). The octopus HGT genes were aligned to proteins in the UniProt database using BLASTP with an e-value of 1×10^{-5}. Blast2GO (Conesa et al., 2005) was used to obtain Gene Ontology (GO) annotations based on the alignment result and WEGO (Ye et al., 2006) was employed to plot the GO functional classification of all HGT genes.

2.6. d_N/d_S ratios analysis

In the phylogenetic analysis, a portion of octopus genes converged with other molluscs and were nested within the bacteria. These HGT genes were deemed to be transferred to the octopus or its ancestor.

Table 1
Candidate HGT (horizontal gene transfer) genes in *O. bimaculoides*.

Name	Protein ID	Location	Len	Annotation	Con Mol	Con Bac	Predicted donor	Relationship	Phylum	Figure
ZnMP1	XP_014767445.1	NW_014656736.1	632	Hemagglutinin	Multiple	Single	*Moritella viscosa*	Marine pathogen	Proteobacteria	Fig. S1
ZnMP2	XP_014774680.1	NW_014666138.1	631	Hemagglutinin	Multiple	Multiple	*Moritella viscosa*	Marine pathogen	Proteobacteria	Fig. S1
ZnMP3	XP_014776931.1	NW_014670794.1	521	Hemagglutinin	Multiple	Single	*Shewanella marina*	Marine bacteria	Proteobacteria	Fig. S1
ZnMP4	XP_014776937.1	NW_014670794.1	612	Hemagglutinin	Multiple	Single	*Shewanella marina*	Marine bacteria	Proteobacteria	Fig.S1
ZnMP5	XP_014781458.1	NW_014682613.1	653	Hemagglutinin	Multiple	Single	*Moritella viscosa*	Marine pathogen	Proteobacteria	Fig. S1
ZnMP6	XP_014782361.1	NW_014686460.1	657	Hemagglutinin	Multiple	Single	*Moritella viscosa*	Marine pathogen	Proteobacteria	Fig. S1
ZnMP7	XP_014783435.1	NW_014692135.1	712	Hemagglutinin	Multiple	Single	*Moritella viscosa*	Marine pathogen	Proteobacteria	Fig. S1
ACS	XP_014779995.1	NW_014678550.1	541	Acyl-CoA synthetase	None	Multiple	*Rhodococcus erythropolis*	Marine bacteria	Actinobacteria	Fig. S2
D-AL	XP_014783568.1	NW_014692782.1	665	D-Aminoacylase	Multiple	Single	*Planctomyces maris 148,843,622*	Marine bacteria	Planctomycetes	Fig. 2
PGS	XP_014784751.1	NW_014699049.1	272	Endo alpha-1,4 polygalactosaminidase	Multiple	Multiple	*Desulfobulbus japonicus*	Marine bacteria	Proteobacteria	Fig. S3
SRL	XP_014788506.1	NW_014739388.1	362	Serine racemase-like	Multiple	Multiple	*Gemmatimonadetes bacterium 575461228*	–	Gemmatimonadetes	Fig. S4
UCP	XP_014790670.1	NW_014655863.1	181	Uncharacterized protein	None	Multiple	*Planctomyces brasiliensis*	Hypersaline lake	Planctomycetes	Fig. S5

Note: Protein ID: the NCBI ID of the HGT genes. Len: the amino acid numbers of the HGT genes. Con Mol/Bac: more than one mollusc/bacteria were identified converged with HGT genes in phylogenetic analysis.

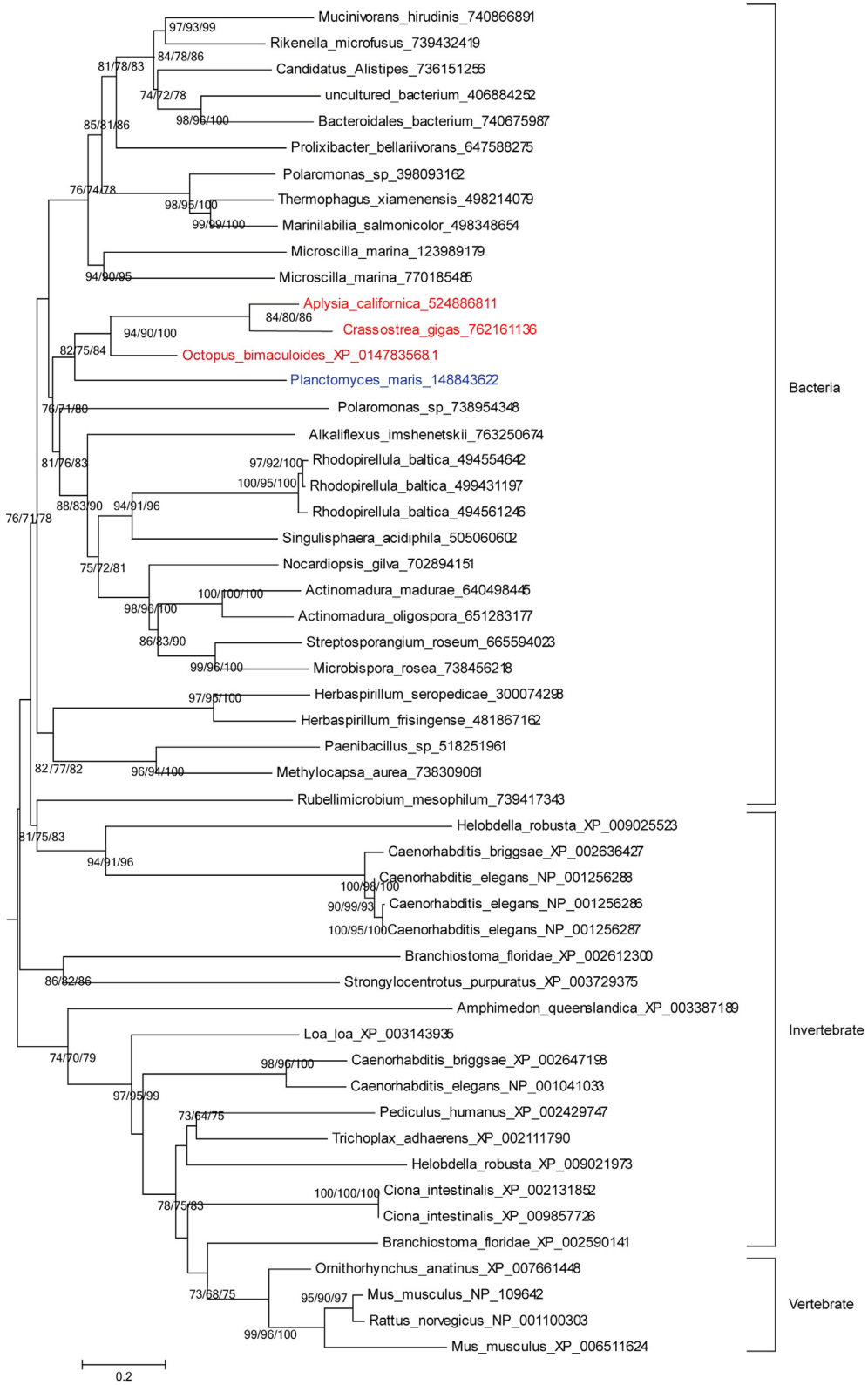

Fig. 2. Phylogenetic analysis of the octopus D-AL and its homologs. The phylogenetic tree was constructed using Mega 7.0 to perform ML (Maximum likelihood) analysis. The support values of ML, NJ (Neighbor-joining) and BI (Bayesian-inference) analysis are displayed beside each node. The homolog from the predicted bacterial donor is indicated in blue and the HGT genes are indicated in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

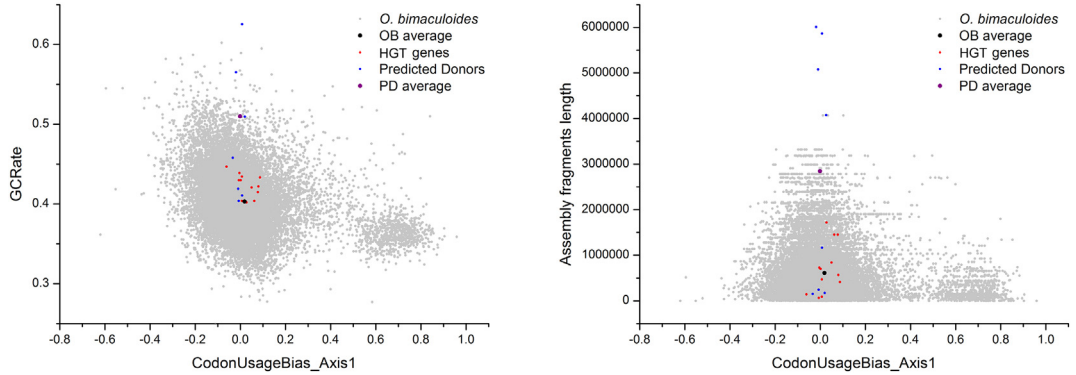

Fig. 3. Compositional traits comparison of HGT genes and host genes. Correspondence analysis was employed to compare the codon usage and GC (Guanine and Cytosine) content among the sifted HGT genes, host genome and the predicted donor genomes. Scatterplots show: (A) the GC content of protein coding genes (GC%) and (B) the size of the assembly fragments on which the gene located is plotted against primary codon usage. Candidate HGT genes are indicated in red; the rate of predicted donor genomes is displayed in blue; the average rate of predicted donor genomes is displayed in purple; other octopus genes are shown in light grey and the average rate of octopus genes are indicated in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Thus, the synonymous and nonsynonymous substitutions between orthologous pairs were calculated by KaKs_Calculator to identify whether there is natural selection acting on these genes (Zhang et al., 2006). The transcriptomic reads of *O. variabilis* (SRR1507221), *L. pealei* (SRR1725172) and *S. japonica* (SRR2889752) were downloaded from NCBI database. After being assembled by Trinity software (Grabherr et al., 2011), 48,957, 47,746 and 52,911 unigenes were obtained respectively. Furthermore, orthologous gene clusters were assigned from OrthoFinder with its default parameters of four species to identify gene families (Emms and Kelly, 2015). The homolog search of HGT genes was performed against the transcriptome unigenes of *O. variabilis*, *L. pealei* and *S. japonica*. The homologs were aligned using ClustalX2 and the conserved sites were extracted using Gblocks (Castresana, 2002). Following this, the d_N/d_S for each orthologous pair were calculated using YN algorithm with KaKs_Calculator.

2.7. Zn-metalloproteinase family analysis

The protein domain annotation of *O. bimaculoides* was identified by Interproscan (Zdobnov and Apweiler, 2001). Following this, Zn-metalloproteinase domain-contained genes were sifted out for phylogenetic analysis based on the annotation results. The protein sequences of the Zn-metalloproteinase family were aligned by the MUSCLE program and the phylogenetic tree was constructed by maximum likelihood (ML) using PhyML v3.060 (Guindon et al., 2009) with default parameters. Additionally, using Multiple EM for the Motif Elicitation (MEME) suite, the conserved motifs of the gene family was analyzed (Bailey et al., 2010).

2.8. Statistical analysis

Results are presented as mean ± standard error of means (SEM). Prior to statistical analysis, all data were tested for normality of distribution using the Kolmogorov-Smirnov test. Differences between control and high temperature were analyzed by Student's *t*-test for independent samples. Analysis was performed using SPSS 18.0 for Windows (SPSS, Michigan Avenue, Chicago, IL, USA).

3. Results

3.1. Twelve HGT genes were determined in O. bimaculoides through BLAST search and phylogenetic analysis

Three steps of BLAST search and two steps of phylogenetic analysis were employed to identify HGT events in 33,638 proteins of *O. bimaculoides*. After the first blast alignment against bacteria sequences, 3,287 candidates showed with an E value $\leq 10^{-30}$, coverage value ≥ 25% and identity value ≥ 25%. After being aligned against protozoa, fungi, plants and vertebrates, the candidate genes with similar genes from 2 or more species were rejected. 1,087 candidate genes were left and send to the third round blast search against NR database. Following this, 137 genes were sifted out through the first step of tree construction. Finally, with the second step of phylogenetic screening, 12 genes were nested within bacteria homologs and identified as HGT genes transferred from the bacteria to the octopus (Table 1). After Interproscan identification, seven genes were annotated as metalloproteinases, including XP_014767445.1 (ZnMP1), XP_014774680.1 (ZnMP2), XP_014776931.1 (ZnMP3), XP_014776937.1 (ZnMP4), XP_014781458.1 (ZnMP5), XP_014782361.1 (ZnMP6) and XP_014783435.1 (ZnMP7). Four genes were identified as enzymes related with carbohydrate and amino acid metabolism. XP_014779995.1 (ACS), XP_014783568.1 (D-AL), XP_014784751.1 (PGS), XP_014788506.1 (SRL) and XP_014790670.1 (UCP) were annotated as acyl-CoA synthetase, D-aminoacylase, endo alpha-1,4 polygalactosaminidase, serine racemase-like and uncharacterized protein, respectively. The sifted candidate genes showed high similarity to the bacterial best-hit sequences (E-values ranged from 1.04E − 99 to 7.01E − 42 and identity values ranged from 69% to 98%). Apart from ACS and UCP, the other 10 HGT candidate genes converged with homologs from other molluscs, before being nested within bacteria homologs (Fig. 2 and Supplementary Figs. 1–5). For example, D-AL and the homologs from oyster (*Crassostrea gigas*) and sea hare (*Aplysia californica*) were bunched together into the mollusc branch surrounded by bacteria (Fig. 2). Additionally, ZnMP1, ZnMP2, ZnMP3, ZnMP4, ZnMP5, ZnMP6, ZnMP7 and D-AL were directly clustered with the probable donor homologs, respectively (Fig. 2 and Supplementary Fig. 1). While the other five genes clustered with a group of bacteria, thus the homolog with the highest BLAST score was deemed as the donor (Supplementary Figs. 2–5).

3.2. Compositional traits comparison of HGT genes and host genes

Unusual codon usage or GC (Guanine and Cytosine) content was depicted as an important criterion for HGT events (Lawrence and Ochman, 1997). In this study, we performed correspondence analysis to compare the codon usage and GC content among the sifted HGT genes, host genome and the predicted donor genomes. CodonW was employed to investigate the codon usage bias and the genes of *O. bimaculoides* were divided into two clusters. Plotting codon usage bias against the GC rate showed that the candidate HGT genes fell into the major cluster of

the genes of *O. bimaculoides* (Fig. 3). The GC rates of HGT genes ranged from 40.2% to 44.7%. The rates were lower than the donor (average GC rate = 51.0%) and 11 of those were higher than the average rate of host (average GC rate = 40.3%) (Fig. 3A). The codon usage values of HGT genes ranged from −0.0611 to 0.0872 and nine genes showed higher codon usage values than the donor (−0.000778) (Fig. 3A). Similarly, we observed that HGT genes were mostly located on assembly fragments (ranging from 60k to 1717k) that were smaller than the donor (2843k) (Fig. 3B). Additionally, the HGT genes located on the large fragments showed codon usage values and GC rates close to the host, such as ZnMP5 (1717k, 0.0283, and 40.2%), ZnMP4 (1449k, 0.0620, and 40.4%) and ZnMP3 (1449k, 0.0782 and 41.5%). In comparison, the HGT genes located on the small fragments showed codon usage values and GC rates close to the donor, such as ZnMP1 (60k, −0.00592 and 43.0%) and ZnMP7 (88k, 0.00721, and 43.4%).

3.3. mRNA expression and distribution of HGT genes

To validate the existence and expression of the HGT genes, PCR assay was performed to clone the cDNA fragments of twelve sifted HGT genes, with the cDNA synthesis from the hepatopancreas mRNA of octopus used as a template. The specificity of PCR results was evaluated with agarose gel electrophoresis with ethidium bromide (EB) staining. Twelve distinct bands were revealed with length ranging from 250 bp to 550 bp, which is consistent with the predicted length of the HGT cDNA fragments (Fig. 4). Following this, after being extracted from agarose gel, the PCR products were sequenced to further validate the expression of the HGT genes (Supplementary File 5).

Twelve HGT genes showed differential expression among 12 transcriptomes (Fig. 5). PGS displayed the highest expression level (FPKM values ranged from 0.11406 to 137.77) at all 12 samples, while SRL was expressed at the lowest level (FPKM values ranged from 0 to 1.11986). Seven HGT genes annotated as Zn-metalloproteinase largely expressed in the tissue of skin and suckers, while ZnMP2, ZnMP3, ZnMP4 and ZnMP5 also displayed an expression pattern in the nervous system. ACS seemed to be mainly expressed in the tissue of viscera. PGS, SRL and UCP displayed a pattern that had a significantly higher expression level in the tissue of nervous system compared with that in the other stages.

3.4. GO (gene ontology) annotation of HGT genes

GO assignments were used to classify the functions of the HGT genes. Based on sequence homology, the sifted HGT genes can be assigned at least one GO term. Furthermore, GO terms were used to classify the sequences in terms of their involvement as cellular components, in molecular functions and in biological processes (Fig. 6). In total, 11 of the 12 genes were clustered in three assignments. Among these genes, 6 were categorized as “cellular component”, 10 as “molecular function” and 6 as “biological process”. The most enriched gene ontology term was metabolic catalytic, including protein, ketone, carbohydrate, lipid and nitrogen compound metabolic processes. Additionally, ZnMP1 and ZnMP2 were predicted to function in symbiosis, encompassing mutualism through parasitism.

3.5. d_N/d_S (the nonsynonymous substitutions per site to synonymous substitutions per site) ratios analysis of HGT genes

In order to detect whether HGT genes have been exposed to selective pressure after the octopus diverged from its ancestor, transcriptomes of three phylogenetically close species with *O. bimaculoides* (*Octopus variabilis*, *Loligo pealei* and *Sepiella japonica*) were assembled. A total of 30 pairs of ortholog genes were obtained through comparison of 10 HGT genes converging with other molluscs in the phylogenetic analysis and the unigenes of three close species. The statistical results indicated that synonymous substitutions were relatively high (d_S is around 3.5). Among all 30 pairs of orthologs, the ratios of the nonsynonymous substitutions per site to synonymous substitutions per site (d_N/d_S) were found to be significantly lower than 1 (Table 2). In particular, aside from ZnMP1, ZnMP2 and ZnMP1, the other seven genes showed d_N/d_S values lower than 0.1, compared with the close species.

3.6. Phylogenetic and structural analysis of Zn-metalloproteinases from O. bimaculoides

In the domain analysis, seven HGT genes were predicted as Zn-metalloproteinases. In order to obtain more information about the divergent characteristics between the horizontally acquired and endogenous Zn-metalloproteinases, 16 sequences containing the Zn-metalloproteinase domain was identified through a genome-wide search in

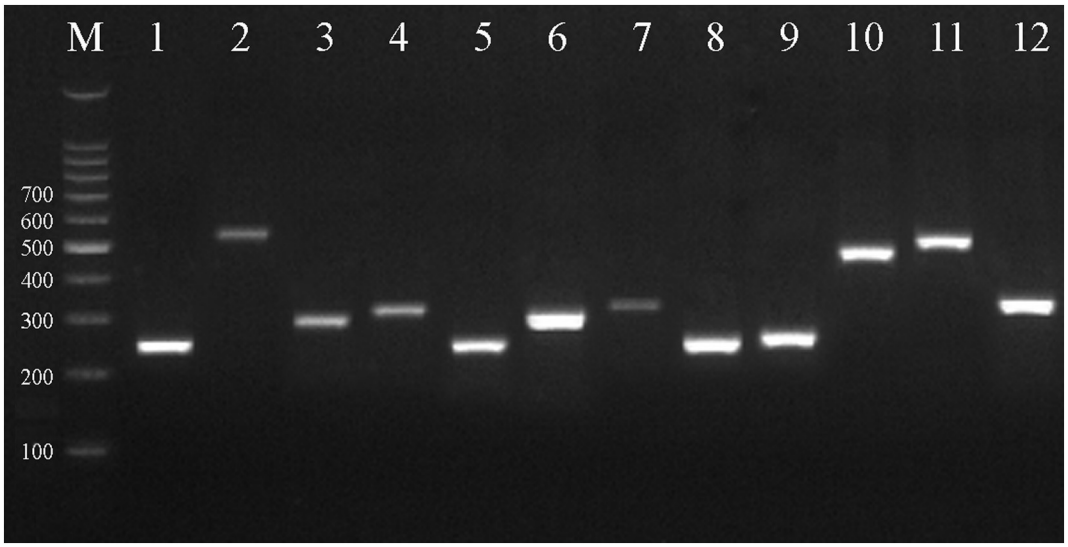

Fig. 4. PCR validation of the HGT genes, with amplified fragments shown on 1% agarose gel electrophoresis with ethidium bromide staining. Identical length of specific PCR products was identified. M: Marker 100 bp DNA ladder; Lane 1–12: XP_014767445.1 (ZnMP1), XP_014774680.1 (ZnMP2), XP_014776931.1 (ZnMP3), XP_014776937.1 (ZnMP4), XP_014779995.1 (ACS), XP_014781458.1 (ZnMP5), XP_014782361.1 (ZnMP6), XP_014783435.1 (ZnMP7), XP_014783568.1 (D-AL), XP_014784751.1 (PGS), XP_014788506.1 (SRL) and XP_014790670.1 (UCP), respectively. A detailed primers and nuclear sequences content is shown in Supplementary Files 3 and 4.

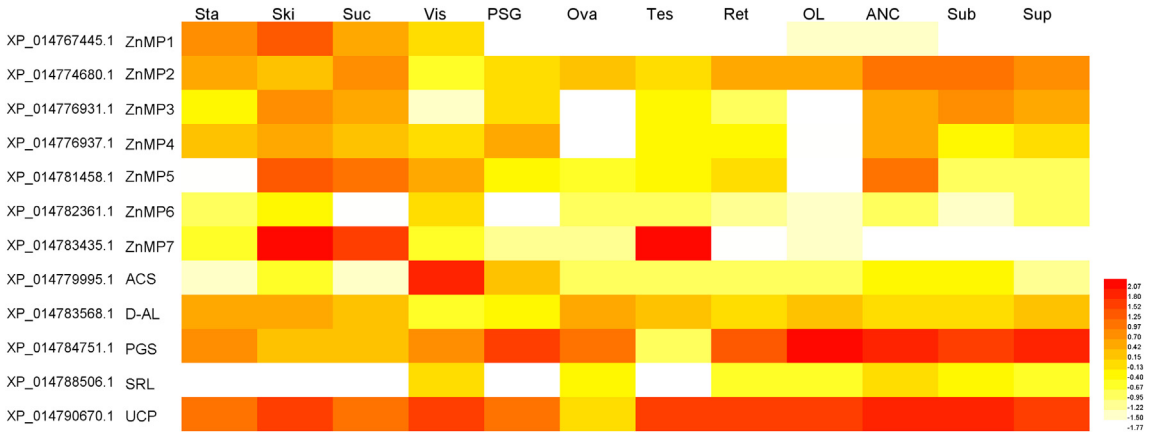

Fig. 5. Expression patterns of *O. bimaculoides* HTGs based on transcriptome analysis. Numbers in the bottom right show the values of FPKM (fragments per kilobase of exon per million fragments mapped) of the transcriptome analysis. Ova, Tes, Vis, PSG, Suc, Ski, Sta, Ret, OL, Sup, Sub and ANC represent ova, testes, viscera, posterior salivary gland, suckers, skin, developmental stage 15, retina, optic lobe, supra-esophageal brain, sub-esophageal brain and axial nerve cord, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

O. bimaculoides. A phylogenetic tree of 16 Zn-metalloproteinases was constructed by maximum likelihood method (ML) (Fig. 7). The Zn-metalloproteinase proteins acquired from HGT were clustered together. The phylogenetic classification was found to be consistent with the motif locations between the HGT and endogenous genes. In this study, 7 types of motifs (Supplementary Fig. 6) were detected. The motifs were similar in proteins of the horizontally acquired Zn-metalloproteinases, but differed to endogenous proteins. In particular, for ZnMP1, ZnMP2 and ZnMP5–7, the type, order and numbers of motifs were similar. For the other two HGT Zn-metalloproteinases, motifs 1, 3–7 were conserved, while motif 2 seemed to be missing in ZnMP3. Additionally, in ZnMP4, a repetitive motif 1 presented at the tail of the deduced protein. In comparison, for the endogenous Zn-metalloproteinases, no more than two motifs were found, which was significantly different to that of the HGT ones. Gene structures of the HGT and endogenous Zn-metalloproteinases also showed significant differences. The HGT Zn-metalloproteinases showed divergent characteristics from endogenous ones with a shorter span in the genome as well as fewer exons and introns.

4. Discussion

As one of the most important forces in the evolution of organism beyond the parent and offspring relationship, HGT has been demonstrated as a mechanism of allowing the host to acquire novel traits from species of unlimited evolutionary distance independently of reproduction (Beiko et al., 2005). Accumulating studies have reported the scale and function of HGT in animals (Crisp et al., 2015; Keeling and Palmer, 2008), while the information about the significance of HGT in molluscs, the second most diverse animal groups, is almost blank (Ponder and Lindberg, 2008). In the present research, 12 HGT genes were identified from California two-spot octopus, *O. bimaculoides*. These HGT genes could be translated with metabolically catalytic activity, which might provide novel traits in addition to contributing to the adaption and evolution of octopus in a similar way to HGT events reported in other animals.

The occurrence of HGT events has been reported to be much more common than what we used to consider (Soucy et al., 2015). In arthropods and nematodes, novel acquired functions from bacteria are well depicted (Nikoh et al., 2008; Werren et al., 2010). Molluscs are highly diverse and second only to arthropods in numbers (Chapman,

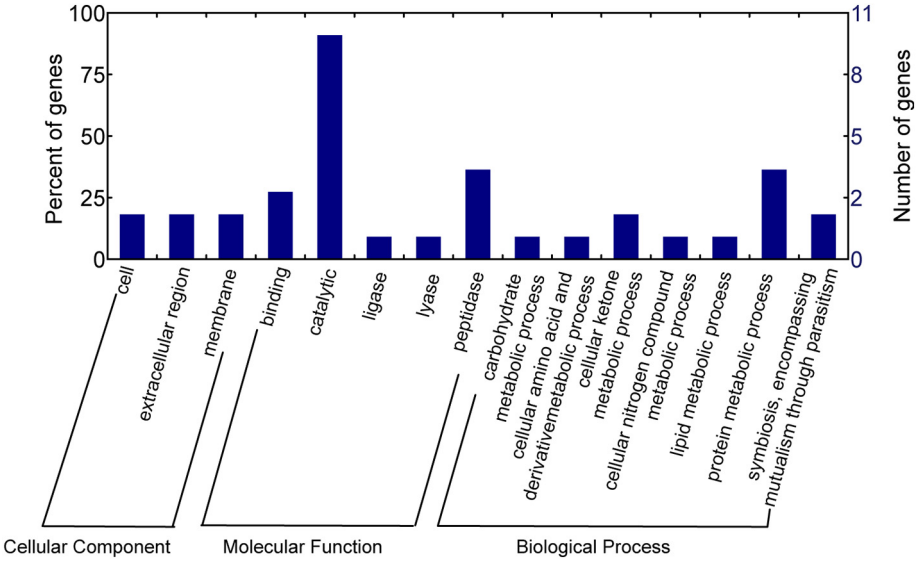

Fig. 6. Gene ontology (GO) annotation of the HGT candidates. The sifted genes were annotated within cellular components, molecular functions and biological processes, respectively.

Table 2
The d_N/d_S (the nonsynonymous substitutions per site to synonymous substitutions per site) values of HGT genes.

O. bimaculoides	*S. japonica*			*O. variabilis*			*L. pealei*		
HGT genes	Length	d_N/d_S	P-value	Length	d_N/d_S	P-value	Length	d_N/d_S	P-value
ZnMP1	1836	0.16340	5.68E−28	300	0.286174	0.000111	1827	0.162499	2.86E−37
ZnMP2	1836	0.15172	1.16E−36	309	0.265522	8.17E−07	1845	0.166448	1.64E−34
ZnMP3	528	0.20046	8.73E−15	1521	0.090349	2.08E−87	1518	0.093141	9.81E−77
ZnMP4	936	0.26476	1.53E−11	1815	0.086727	2.27E−79	1815	0.089306	8.83E−77
ZnMP5	1932	0.09300	4.56E−91	420	0.264747	1.34E−05	1914	0.092825	1.71E−92
ZnMP6	1434	0.09387	3.90E−82	1860	0.126172	3.37E−58	474	0.160401	5.23E−23
ZnMP7	1920	0.16209	2.44E−31	285	0.411138	0.005319	1932	0.172655	8.69E−27
D-AL	669	0.26127	8.33E−07	1926	0.077641	5.23E−87	423	0.157961	1.39E−27
PGS	813	0.16579	5.74E−05	345	0.535183	0.008312	324	0.021844	1.14E−24
SRL	669	0.18912	1.33E−22	420	0.277314	5.23E−08	978	0.049616	1.70E−69

Note: the d_N/d_S for each orthologous pair (*O. bimaculoides* with *S. japonica*, *O. variabilis* or *L. pealei*) were calculated.

2009). In the present study, HGT events are identified in octopus based on rigorous BLAST alignment and phylogenetic analysis. Twelve genes result in phylogenetic incongruences, which show glaring conflict in the phylogeny with the nested-in bacteria homologs. Phylogenetic incongruences have been deemed as the gold standard for identifying HGT events (Keeling and Palmer, 2008). Close evolutionary distances are revealed between 12 candidates and the bacteria homologs, respectively, instead of those from the close species, suggesting that the sifted genes could be acquired from the bacteria horizontally. Evolutionary close organisms often have comparable GC content and codon usage manner (Ochman et al., 2000). The 12 sifted HGT genes showed GC rates and codon usage values between the average value of the donor bacteria and that of octopus, contributing to the possibility of a horizontal origin instead of the vertical one. Due to the false assembly, the sifted bacterial DNA in an animal genome is not sufficient to imply that the target genes are horizontally transferred (Hotopp, 2011). Thus, the transcriptomic information and PCR results could further convince the existence of the candidates in the octopus genome as well as excluding the factors of contamination and false assembly. For multicellular eukaryotes, the high frequency of contact provides the facilities and enhances the possibility of HGT events (Robinson et al., 2010). The octopus was proved to harbor plentiful bacteria and the marine environment preserves high frequency of HGT (Jiang and Paul, 1998; Malham et al., 2002). Furthermore, large-scale genomic rearrangements were found in octopus apart from RNA editing (Albertin et al., 2015). In agreement of the reports, most of the predicted donors of 12 HGT genes identified in octopus are familiar bacteria and pathogens in marine environments, demonstrating the theoretical basis of HGT events. For example, *Moritella viscosa* was depicted as a usual pathogen causing mortality (Urakawa, 2014). Certainly, because the HGT in octopus might be ancient, so we cannot exclude the possibility that *M. viscosa* is the closest known phylogenetic match of another marine ancestor. The sequence similarity, phylogenetic incongruence, transcriptome information, PCR results and the possibility contacting bacteria of *O. bimaculoides* collectively present the evidence of HGT existing in molluscs. Additionally, seven HGT Zn-metalloproteinases present divergent motif composition, intron presence and phylogenetic relationship to the endogenous ones, providing further evidences to validate the HGT.

Bacteria, unicellular eukaryotes and multicellular organisms share diverse mechanisms and features in gene content exchanges (Soucy et al., 2015). While, in this process, it displays the same characteristic that the sequence composition (oligonucleotide composition and codon usage) of anciently transferred genes often ameliorate to the host genome (Lawrence and Ochman, 1997). The GC rates and codon usage values of identified genes from octopus are distributed between that of the donor and host, indicating that these genes were integrated into the octopus genome for a long period of time. Most of the short fragmental scaffolds or contigs result from assembly of repetitive sequences or transportable elements, which was closely related to HGT (Keeling and Palmer, 2008). HGT genes with codon usage values and GC rates similar to the host are found on bigger fragments, suggesting those genes were

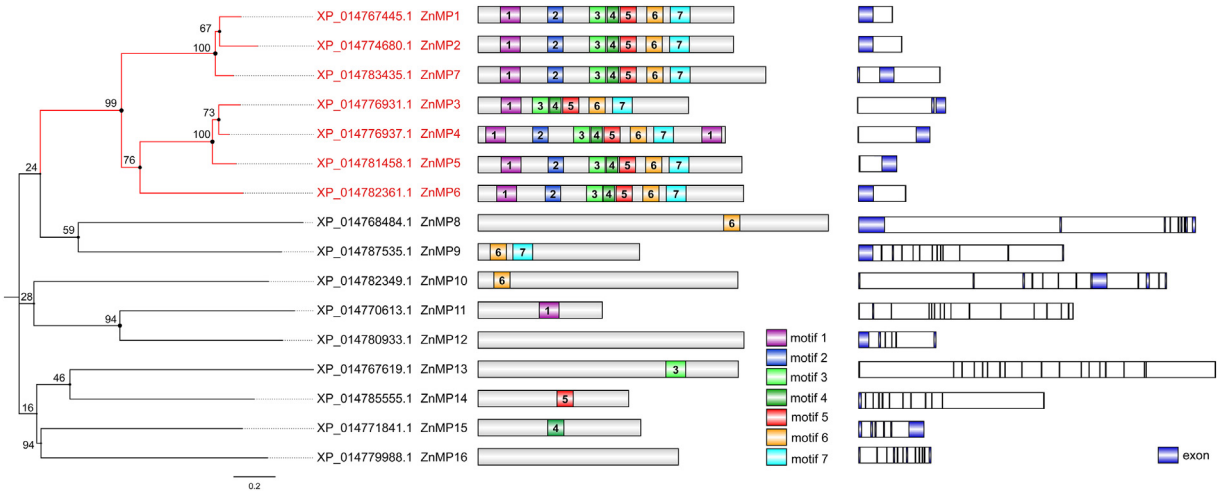

Fig. 7. An analytical view of the HGT Zn-metalloproteinases and the endogenous homologs. The protein sequences of Zn-metalloproteinase family were aligned by the MUSCLE program and the phylogenetic tree was constructed by maximum likelihood (ML). The colored boxes represent the motifs and exons in the protein. A total of 7 types of motifs were found in 16 Zn-metalloproteinases, as indicated in the table on the right-hand side. A detailed motif introduction is shown in Supplementary Fig. 6.

absorbed by octopus for a relatively longer time. In contrast, genes located on shorter fragments exhibit usage values and GC rates close to the donor, such as ZnMP1 and ZnMP7, suggesting a relatively recent transfer. A similar conclusion has also been shown in sponges (Conaco et al., 2016). The results of phylogenetic analysis could also present the transfer time and evolutionary process (Lapierre and Gogarten, 2009). ACS and UCP were nested within bacteria homologs without any other converged molluscs. ZnMP1 and ZnMP7 converged with oysters, while the others were converged with more than one mollusc. Based on the present completeness of public database, genes that had converged with more than one mollusc might be introduced into molluscan ancestor before radiation. Additionally, two flanking sequences, ZnMP3 and ZnMP4, were found to be located on the same scaffold NW_014670794.1, indicating that potential tandem duplications could exist in the evolutionary process of octopus HGT genes. After the transfer, natural selection (either purifying or diversifying) potentially shaped the genetic variation of the HGT genes, although there has been little attention paid to this (Lopezleal et al., 2016). In octopus, strong negative or purified selection is detected as the main form of selection acting on the HGT genes, which indicates the acquired genes are purged from the genome. Mutations on the negatively selected genes at the protein levels are usually depicted as deleterious effects (Peters et al., 2012), so HGT genes identified in this article could possess definite and important functions in the evolution of octopus after divergence from its ancestor.

Accumulating evidence has revealed that HGT is not only an important evolutionary force, but also could hold novel and biological significance for the host (Boto, 2014; Gogarten and Townsend, 2005). Indeed, the existence of HGT is not sufficient to infer that the acquired genes are functional. However, in the present study, the negative selection, PCR results and tissue distribution suggest the HGT genes could perform biological functions in octopus. The same finding has also been reported, with bacteriophages originated lysozyme genes found to be used by bivalves to combat bacteria (Ren et al., 2017). The important functions of HGT genes could be revealed through the maintenance provided by negative selection, similar to previously reported GTA genes (Lang et al., 2012). HGT genes could be transcribed with universal expression patterns in all detected tissues. The highest expression level of ACS is detected in visceral areas, which contains a variety of metabolites, indicating its function in cellular metabolism (Cheney et al., 2016). Additionally, the highest expression pattern of PGS, SRL and UCP in the tissue of nervous system shows the HGT might also contribute to the remarkable and largest nervous system in the octopus. Through GO analysis, the HGT genes are also annotated with the function of metabolism and defense process. ACS is annotated as acyl-CoA synthetase, functioning in the conversion of acetyl-CoA to acetate with the production of one ATP (Sánchez et al., 2000). In amoeba, acyl-CoA synthetase is reported to be horizontally acquired from bacteria, enabling the parasites to survive in anaerobic environments (Field et al., 2000), with ACS assumed to play the same role to improve the energy supplement of octopus. D-Amino acid residues are widespread in the natural antibiotics of microorganisms (Essack, 2001), with D-aminoacylase (D-AL XP_014783568.1) contributing to utilization of the D-amino acid from the bacteria. As excitatory amino acids, D-amino acids are known to mediate synaptic excitation (Moloney, 1998), thus D-AL might have a role in the nervous system. Additionally, endo alpha-1,4 polygalactosaminidase (PGS, XP_014784751.1) and serine racemase-like (SRL, XP_014788506.1) also act on D-amino acids and its derivates. PGS is involved in the endohydrolysis of the poly form (D-galactosamine) into D-galactosamine, while N-acetylgalactosamine is necessary for intercellular communication, especially in the sensory nerve structures of animals (Takamiya et al., 1996; Tamura et al., 1988). Serine racemase enriches the animal brain and catalyzes the formation of D-serine from L-serine in the regulation of neurotransmission through glia-neuronal interactions (Wolosker et al., 1999). Taken together, these findings indicate that the HGT plays a potential role in the innovations or improvements of the digestive and nervous system in *O. bimaculoides*.

Zn-metalloproteinase is a dominant group in the sifted HGT genes and is regarded to be functional through the analyses of natural selection and tissue distribution. Seven HGT genes annotated as Zn-metalloproteinases mainly expressed in skin and suckers, which is suspected to be related with the frequency of contact with pathogens. In mollusc, skin is the organ that coats the internal organs and labial palps are responsible for the fodder intake (Yang et al., 2011). The high expression level in mantle and labial palps implies that Zn-metalloproteinase might play a role at the first line of the invader defense. Following this, the phylogenetic and structural analysis of Zn-metalloproteinases reveal that the horizontally transferred genes possess a distinctly different motif organization compared to the endogenous ones, indicating a novel function divergent from the endogenous Zn-metalloproteinases. The HGT Zn-metalloproteinases share closer phylogenetic relationships and more similar motif structures with each other, while broad diversification is shown among the endogenous genes. Being consistent with the uniform expression pattern, the seven HGT Zn-metalloproteinases might partially or completely overlap in function. According to the GO analysis, the HGT Zn-metalloproteinases are also annotated as M4 metalloproteinase, which is frequently found on the membrane of both Gram-negative and Gram-positive eubacteria (Rawlings et al., 2002). This family is reported to be closely related to the pathogenicity, with numerous pathogens having acquired it for the successful invasion of host tissue (Miyoshi et al., 1998). As a pathogenic factor responsible for the bacterial infections, metalloproteinase can degrade various immune defense proteins (Adekoya and Sylte, 2009). Therefore, the HGT metalloproteinase is suspected as a powerful weapon to provide recipient organisms with defense against non-symbiotic bacteria or other organisms (Conaco et al., 2016). Furthermore, the variation on N terminus of HGT Zn-metalloproteinase may contribute to the recognition with broad substrate specificity, because the motif on the N-terminus could be better used in recognition and binding on the surface of substrates (Barclay et al., 1997; Kessler et al., 1998). Furthermore, metalloproteinases are also reported to be responsible for tissue integrity in both primitive and complex multicellular animals (Ziegler et al., 2002). Thus, the HGT Zn-metalloproteinase may provide novel capacity of normal growth and wound repair. The shared metalloproteinases have been shown to exist in both the bacteria and several host animals, such as sponges (Conaco et al., 2016) and sea urchins (Souza and Brentani, 1993). Metalloproteinases could also play roles in developmental and remodeling processes, with both sponges and sea urchins having important evolutionary placements (Sodergren et al., 2006; Srivastava et al., 2010). Therefore, we suppose that the HGT Zn-metalloproteinases are responsible for the evolution of the octopus, especially the morphological innovations.

5. Conclusions

In conclusion, we report the HGT events that were transferred from the bacteria to mollusc. From this, 12 HGT genes were identified from the California two-spot octopus based on similarity search, phylogenetic construction, gene composition analysis and PCR validation. The HGT genes could be ubiquitously translated in the tissues detected in this article. Strong negative selection was detected as the main form of selection acting on the HGT genes. Furthermore, the annotation indicates that HGT is associated with metabolism, intercellular communication, immune defense and tissue integrity. Thus, our results provide insights into understanding the biological significance of HGT events with regards to the evolution of octopus in nutrition supplements, neuronal interactions, immune defense and morphological development.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2018.05.013.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The authors are grateful to the Editor for polishing the language of our paper and the two anonymous reviewers for their comments that greatly improved the manuscript. This research was supported by Shenzhen Science and Technology program (JCYJ20150630165133395), Agricultural Science and Technology Innovation Program (ASTIP) of Chinese Academy of Agricultural Sciences (CAAS), National Key Research and Development Program of China (2016YFC1200600), and Fund of Key Laboratory of Shenzhen (ZDSYS20141118170111640).

References

Abollo, E., Gestal, C., Pascual, S., 2001. *Anisakis* infestation in marine fish and cephalopods from Galician waters: an updated perspective. Parasitol. Res. 87, 492–499.
Adekoya, O.A., Sylte, I., 2009. The thermolysin family (M4) of enzymes: therapeutic and biotechnological potential. Chem. Biol. Drug Des. 73, 7–16.
Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224.
Andersson, J.O., 2005. Lateral gene transfer in eukaryotes. Cell. Mol. Life Sci. 62, 1182–1197.
Andersson, J.O., Hirt, R.P., Foster, P.G., Roger, A.J., 2006. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. BMC Evol. Biol. 6, 27.
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.
Bailey, T.L., Boden, M., Whitington, T., Machanick, P., 2010. The value of position-specific priors in motif discovery using MEME. BMC Bioinf. 11, 179.
Barclay, A.N., Brown, M.H., Law, S.A.K.A., McKnight, A.J., Tomlinson, M.G., van der Merwe, P.A., 1997. The Leucocyte Antigen Factsbook. Academic Press.
Beiko, R.G., Harlow, T.J., Ragan, M.A., 2005. Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 14332–14337.
Bhattacharya, D., Pelletreau, K.N., Price, D.C., Sarver, K.E., Rumpho, M.E., 2013. Genome analysis of *Elysia chlorotica* egg DNA provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusc. Mol. Biol. Evol. 30, 1843–1852.
Boletzky, S., 1989. Recent studies on spawning, embryonic development, and hatching in the Cephalopoda. Adv. Mar. Biol. 25, 85–115.
Boto, L., 2010. Horizontal gene transfer in evolution: facts and challenges. Proc. R. Soc. Lond. B Biol. Sci. 277, 819–827.
Boto, L., 2014. Horizontal gene transfer in the acquisition of novel traits by metazoans. Proc. Biol. Sci. 20132450.
Casadevall, A., Pirofski, L.-A., 1999. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. Infect. Immun. 67, 3703–3713.
Casadevall, A., Pirofski, L.-A., 2000. Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. Infect. Immun. 68, 6511–6518.
Castresana, J., 2002. Gblocks Server v. 0.91 b. Institut de Biologia Evolutiva (CSIC-UPF).
Chapman, A.D., 2009. Numbers of Living Species in Australia and the World.
Cheney, K.L., White, A., Mudianta, I.W., Winters, A.E., Quezada, M., Capon, R.J., Mollo, E., Garson, M.J., 2016. Choose your weaponry: selective storage of a single toxic compound, Latrunculin A, by closely related nudibranch molluscs. PLoS One 11, e0145134.
Conaco, C., Tsoulfas, P., Sakarya, O., Dolan, A., Werren, J., Kosik, K.S., 2016. Detection of prokaryotic genes in the *Amphimedon queenslandica* genome. PLoS One 11, e0151092.
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Crisp, A., Boschetti, C., Perry, M.J., Tunnacliffe, A., Micklem, G., 2015. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol. 16, 50.
Danchin, E.G., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P.M., Henrissat, B., Abad, P., 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc. Natl. Acad. Sci. 107, 17651–17656.
Danchin, E.G., Guzeeva, E.A., Mantelin, S., Berepiki, A., Jones, J.T., 2016. Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. Mol. Biol. Evol. 33, 1571–1579.
Deitsch, K.W., Driskill, C.L., Wellems, T.E., 2001. Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Res. 29, 850–853.
Dethlefsen, L., McFall-Ngai, M., Relman, D.A., 2007. An ecological and evolutionary perspective on human–microbe mutualism and disease. Nature 449, 811–818.
Doolittle, W.F., 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 14, 307–311.
Doolittle, W.F., Logsdon Jr, J.M., 1998. Archaeal genomics: do archaea have a mixed heritage? Curr. Biol. 8, R209–R211.
Drezen, J.-M., Gauthier, J., Josse, T., Bézier, A., Herniou, E., Huguet, E., 2016. Foreign DNA acquisition by invertebrate genomes. J. Invertebr. Pathol. 14, 157168.
Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., Relman, D.A., 2005. Diversity of the human intestinal microbial flora. Science 308, 1635–1638.
Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.
Essack, S.Y., 2001. The development of β-lactam antibiotics in response to the evolution of β-lactamases. Pharm. Res. 18, 1391–1399.
Eyres, I., Boschetti, C., Crisp, A., Smith, T.P., Fontaneto, D., Tunnacliffe, A., Barraclough, T.G., 2015. Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. BMC Biol. 13, 90.
Fidalgo, S., Wang, Q., Riley, T., 2000. Comparison of methods for detection of *Erysipelothrix* spp. and their distribution in some Australasian seafoods. Appl. Environ. Microbiol. 66, 2066–2070.
Field, J., Rosenthal, B., Samuelson, J., 2000. Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. Mol. Microbiol. 38, 446–455.
Fraj, L.J., Duce, G.F., 1998. Measures for allergen avoidance in asthma. Arch. Bronconeumol. 35, 345–356.
Gogarten, J.P., Townsend, J.P., 2005. Horizontal gene transfer, genome innovation and evolution. Nat. Rev. Microbiol. 3, 679–687.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644.
Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V., Gascuel, O., 2009. PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. Mol. Biol. Evol. 9, 384–385.
Hanlon, R.T., Messenger, J.B., 1996. Cephalopod Behaviour. Cambridge University Press ISBN 0-521-42083-0.
Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J.-P., Rigden, D.J., Michels, P.A., Opperdoes, F.R., 2003. Plant-like traits associated with metabolism of Trypanosoma parasites. Proc. Natl. Acad. Sci. 100, 1067–1071.
Hotopp, J.C.D., 2011. Horizontal gene transfer between bacteria and animals. Trends Genet. 27, 157–163.
Hotopp, J.C.D., Clark, M.E., Oliveira, D.C., Foster, J.M., Fischer, P., Torres, M.C.M., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317, 1753–1756.
Huang, J., 2013. Horizontal gene transfer in eukaryotes: the weak-link model. BioEssays 35, 868–875.
Huang, J., Yue, J., 2013. Horizontal gene transfer in the evolution of photosynthetic eukaryotes. J. Syst. Evol. 51, 13–29.
Jiang, S.C., Paul, J.H., 1998. Significance of lysogeny in the marine environment: studies with isolates and a model of lysogenic phage production. Microb. Ecol. 35, 235–243.
Keeling, P.J., Palmer, J.D., 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9, 605–618.
Kessler, E., Safrin, M., Gustin, J.K., Ohman, D.E., 1998. Elastase and the LasA protease of *Pseudomonas aeruginosa* are secreted with their propeptides. J. Biol. Chem. 273, 30225–30231.
Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.
Lang, A.S., Zhaxybayeva, O., Beatty, J.T., 2012. Gene transfer agents: phage-like elements of genetic exchange. Nat. Rev. Microbiol. 10, 472–482.
Lapierre, P., Gogarten, J.P., 2009. Estimating the size of the bacterial pan-genome. Trends Genet. 25, 107–110.
Lawrence, J.G., Ochman, H., 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44, 383–397.
Li, Z.-W., Shen, Y.-H., Xiang, Z.-H., Zhang, Z., 2011. Pathogen-origin horizontally transferred genes contribute to the evolution of Lepidopteran insects. BMC Evol. Biol. 11, 356.
Lopezleal, G., Cevallos, M.A., Castilloramirez, S., 2016. Evolution of a sigma factor: an all-in-one of gene duplication, horizontal gene transfer, purifying selection, and promoter differentiation. Front. Microbiol. 7, 581.
Malham, S.K., Lacoste, A., Gélébart, F., Cueff, A., Poulet, S.A., 2002. A first insight into stress-induced neuroendocrine and immune changes in the octopus *Eledone cirrhosa*. Aquat. Living Resour. 15, 187–192.
McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B., Paul, J.H., 2010. High frequency of horizontal gene transfer in the oceans. Science 330, 50.
Miyoshi, S.-i., Nakazawa, H., Kawata, K., Tomochika, K.-i., Tobe, K., Shinoda, S., 1998. Characterization of the hemorrhagic reaction caused by *Vibrio vulnificus* metalloprotease, a member of the thermolysin family. Infect. Immun. 66, 4851–4855.
Moloney, M., 1998. Direct versatile route to conformationally constrained glutamate analogues. Chem. Commun. 461–462.
Moran, N.A., Baumann, P., 2000. Bacterial endosymbionts in animals. Curr. Opin. Microbiol. 3, 270–275.
Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M., Fukatsu, T., 2008. *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. Genome Res. 18, 272–280.
Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299–304.
Peters, A.E., Bavishi, A., Cho, H., Choudhary, M., 2012. Evolutionary constraints and expression analysis of gene duplications in *Rhodobacter sphaeroides* 2.4.1. BMC Res.

Notes 5, 192.
Ponder, W., Lindberg, D.R., 2008. Phylogeny and Evolution of the Mollusca. Univ. of California Press.
Rawlings, N.D., O'brien, E., Barrett, A.J., 2002. MEROPS: the protease database. Nucleic Acids Res. 30, 343–346.
Ren, Q., Wang, C., Jin, M., Lan, J., Ye, T., Hui, K., et al., 2017. Co-option of bacteriophage lysozyme genes by bivalve genomes. Open Biol. 7, 160285.
Richards, T.A., Hirt, R.P., Williams, B.A., Embley, T.M., 2003. Horizontal gene transfer and the evolution of parasitic protozoa. Protist 154, 17.
Richards, T.A., Dacks, J.B., Campbell, S.A., Blanchard, J.L., Foster, P.G., McLeod, R., Roberts, C.W., 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryot. Cell 5, 1517–1531.
Robinson, C.J., Bohannan, B.J., Young, V.B., 2010. From structure to function: the ecology of host-associated microbial communities. Microbiol. Mol. Biol. Rev. 74, 453–476.
Rumpho, M.E., Worful, J.M., Lee, J., Kannan, K., Tyler, M.S., Bhattacharya, D., et al., 2008. Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug *Elysia chlorotica*. Proc. Natl. Acad. Sci. 105, 17867–17871.
Sánchez, L.B., Galperin, M.Y., Müller, M., 2000. Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (nucleoside diphosphate-forming). J. Biol. Chem. 275, 5794–5803.
Small, A.L., McFall-Ngai, M.J., 1999. Halide peroxidase in tissues that interact with bacteria in the host squid *Euprymna scolopes*. J. Cell. Biochem. 72, 445–457.
Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science 314, 941–952.
Soucy, S.M., Huang, J., Gogarten, J.P., 2015. Horizontal gene transfer: building the web of life. Nat. Rev. Genet. 16, 472–482.
Souza, S.J., Brentani, R., 1993. Sequence homology between a bacterial metalloproteinase and eukaryotic matrix metalloproteinases. J. Mol. Evol. 36, 596–598.
Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726.
Stanhope, M.J., Lupas, A., Italia, M.J., Koretke, K.K., Volker, C., Brown, J.R., 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature 411, 940–944.
Takamiya, K., Yamamoto, A., Furukawa, K., Yamashiro, S., Shin, M., Okada, M., Fukumoto, S., Haraguchi, M., Takeda, N., Fujimura, K., 1996. Mice with disrupted GM2/GD2 synthase gene lack complex gangliosides but exhibit only subtle defects in their nervous system. Proc. Natl. Acad. Sci. 93, 10662–10667.
Tamura, J.-i., Takagi, H., Kadowaki, K., 1988. Purification and some properties of the endo α-1,4 polygalactosaminidase from *Pseudomonas* sp. Agric. Biol. Chem. 52, 2475–2484.
Timmis, J.N., Ayliffe, M.A., Huang, C.Y., Martin, W., 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5, 123–135.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515.
Urakawa, H., 2014. The family Moritellaceae. In: The Prokaryotes. Springer, pp. 477–489.
Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Group, N.G.W., 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327, 343–348.
Wickham, H., 2011. ggplot2. Wiley Interdiscip. Rev. Comput. Stat. 3, 180–185.
Wolosker, H., Blackshaw, S., Snyder, S.H., 1999. Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission. Proc. Natl. Acad. Sci. 96, 13409–13414.
Yang, J., Wang, L., Zhang, H., Qiu, L., Wang, H., Song, L., 2011. C-type lectin in *Chlamys farreri* (CfLec-1) mediating immune recognition and opsonization. PLoS One 6, e17089.
Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34, W293–W297.
Young, J.Z., 1971. Anatomy of the Nervous System of *Octopus vulgaris*.
Yuan, J.-B., Zhang, X.-J., Liu, C.-Z., Wei, J.-K., Li, F.-H., Xiang, J.-H., 2013. Horizontally transferred genes in the genome of Pacific white shrimp, *Litopenaeus vannamei*. BMC Evol. Biol. 13, 165.
Zdobnov, E.M., Apweiler, R., 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.
Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G.K.-S., Yu, J., 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4, 259–263.
Zhang, H.-H., Feschotte, C., Han, M.-J., Zhang, Z., 2014. Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals. Genome Biol. Evol. 6, 1375–1386.
Ziegler, G., Paynter, K., Fisher, D., 2002. Matrix metalloproteinase-like activity from hemocytes of the eastern oyster, *Crassostrea virginica*. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 131, 361–370.