Marine Genomics 39 (2018) 22–25

Contents lists available at ScienceDirect

Marine Genomics

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

Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*

Jianbo Yuan[a,b], Xiaojun Zhang[a,b,*], Chengzhang Liu[a], Yang Yu[a], Jiankai Wei[c], Fuhua Li[a,b], Jianhai Xiang[a,b,**]

[a] Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China
[b] Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China
[c] Ministry of Education Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

ARTICLE INFO

Keywords:
Penaeid shrimps
Genome
Assembly
Annotation

ABSTRACT

Penaeid shrimps are among the most economically important crustaceans, which provide an important global food source. These species exhibit complex body plans and novelties, such as segments and appendages, which render them interesting organisms for developmental biology study of crustaceans. However, limited genomic resources have been put forward for the researches of them. Here, we report the genome sequencing and draft assembly of two economically important penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*. A total of 132.86 Gb and 132.83 Gb sequencing data was obtained in the two shrimp species. The genome assembly, a total length of 1.94 Gb and 2.04 Gb in *M. japonicus* and *P. monodon*, respectively, covers more than 97% of coding regions. We further identified 626 Mb (34.96%) and 833 Mb (46.68%) repeats, 16,716 and 18,100 genes in these two genomes, respectively. We also identified Hox genes that are important to their body plans. These data will provide valuable resources for the study of selective breeding and some plastic biological characters of penaeid shrimps, including molting, lobstering, brooding eggs and sensitization in humans.

1. Introduction

Penaeid shrimps belong to Penaeidae, a family of marine crustaceans, which includes many economical important species, such as Pacific whiteleg shrimp *Litopenaeus vannamei*, kuruma prawn *Marsupenaeus japonicus* and the giant tiger prawn *Penaeus monodon* (Koyama et al., 2010; Wilson et al., 2000; Farfante and Kensley, 1997). These species are the subject of commercial fisheries, which makes them as the valuable internationally traded commodity in aquaculture (FAO, Yearbook of Fisheries Statistics Summary Tables, 2013). Penaeid shrimps exhibit complex body plans and novelties, such as segments, appendages and lateral line-like sense organs on the antennae (Farfante and Kensley, 1997), thus, the research of them may be important for developmental biology study of crustaceans. However, to our knowledge, except for the low coverage sequencing and draft assembly of *L. vannamei* (Yu et al., 2015), *Exopalaemon carinicauda* (Yuan et al., 2017), *Parhyale hawaiensis* (Kao et al., 2016), and *Neocaridina denticulata* (Kenny et al., 2014), none of the shrimp genomes has been ultimately completed because of the large genome size and highly repetitive sequences (Yu et al., 2015; Abdelrahman et al., 2017).

Here, we provide genome sequences of two penaeid shrimps, *M. japonicus* and *P. monodon*. We performed draft genome assemblies, gene structure and repetitive sequences predictions for these two species. These data can be used for comparative genomics analyses, and provide valuable resources for shrimp genetics and breeding.

2. Data description

2.1. Sample preparation and sequencing

The nomenclature of the two penaeid shrimps, *M. japonicus* and *P. monodon*, was referred to the (ITIS) database (https://www.itis.gov/) and previous researches (Koyama et al., 2010; Wilson et al., 2000; Farfante and Kensley, 1997). The DNA was extracted from muscle of male adults using a TIANamp Marine Animal DNA Kit (TIANGEN, Beijing, China) (Table 1). Two paired-end DNA libraries with insert size of 230 bp and 500 bp were constructed following the standard Illumina operating procedure (Illumina, San Diego, CA). The paired-end

Abbreviations: MIGS, Minimum Information about a Genome Sequence; CEGMA, core eukaryotic genes mapping approach; TGICL, TIGR Gene Indices clustering tools; NCBI, National Center for Biotechnology Information; TEs, transposable elements; SSR, simple sequence repeats; SNP, single-nucleotide polymorphisms; Indels, short insertion/deletion

* Correspondence to: X. Zhang, Institute of Oceanology, Chinese Academy of Sciences, 7, Nanhai Road, Qingdao 266071, China.
** Correspondence to: J. Xiang, Institute of Oceanology, Chinese Academy of Sciences, 7, Nanhai Road, Qingdao 266071, China.
E-mail addresses: xjzhang@qdio.ac.cn (X. Zhang), jhxiang@qdio.ac.cn (J. Xiang).

https://doi.org/10.1016/j.margen.2017.12.006
Received 29 September 2017; Received in revised form 16 December 2017; Accepted 16 December 2017
Available online 27 December 2017
1874-7787/

Table 1
General information of *M. japonicus* and *P. monodon*.

Items	Description
General feature of classification	
Investigation type	Eukaryote
Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Crustacea; Malacostraca; Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea; Penaeidae; Penaeus
Project name	Whole genome sequencing of *Marsupenaeus japonicus* and *Penaeus monodon*
Geographic location	*M. japonicus*: Nanning, Guangxi, China *P. monodon*: Shenzhen, Guangzhou, China
Latitude, longitude	*M. japonicus*: 21.83°N/108.29°E *P. monodon*: 22.35°N/114.18°E
Collection date	2015-07
Environment (biome)	Water body (ENVO:00000063)
Environment (material)	Sea water (ENVO: 00002149)
Sequencing method	Illumina HiSeq2500; Paired-end (2 × 150)
MIGS-specific mandatory descriptors	
Ploidy	Diploid
Number of replicons	*M. japonicas*: 2n = 86 chromosomes; *P. monodon*: 2n = 88 chromosomes
Estimated genome size	*M. japonicus*: 2.28 Gb *P. monodon*: 2.59 Gb
Reference of biomaterial	(Koyama et al., 2010; Wilson et al., 2000; Farfante and Kensley, 1997)
Assembly method	*De novo* assembly
Assembly program	SOAPdenovo2

Table 2
Summary of the genome assembly of two penaeid shrimp species.

	M. japonicus		*P. monodon*	
	Contig	Scaffold	Contig	Scaffold
Number:	5,632,117	3,719,281	7,106,289	4,985,320
Total length (bp):	1,924,054,682	1,942,550,811	1,882,378,599	2,035,458,477
Longest (bp):	16,221	1,606,464	12,599	1,275,042
Shortest (bp):	100	100	100	100
N50 (bp):	416	937	301	786
N90 (bp):	159	189	138	144
> 2 kb:	154,376	97,798	118,142	74,634

sequencing was performed on the Illumina HiSeq2500 platform with read length of 150 bp. The raw sequencing data were trimmed to filter out low-quality data and adapter contaminates by using the NGS QC Toolkit with the parameters of "2 A-c 10" (Patel and Jain, 2012). Finally, we collected the clean data of the two penaeid shrimps (Table S1).

2.2. Estimation of genome size, polymorphism, and repetitiveness

Genome size was estimated based on the K-mer depth distribution according to previous researches (Li et al., 2010). A major peak was observed around K-mer depth of 45 and 47 in *M. japonicus* and *P. monodon*, respectively, which corresponds to homozygous regions (Fig. S1). Genome size was estimated to be 2.28 Gb and 2.59 Gb in *M. japonicus* and *P. monodon*, respectively, which was similar to the results (C-value of 2.83 pg and 2.53 pg) from Animal Genome Size Database (www.genomesize.com/). Besides, a high proportion of K-mers with depth higher than 200 × (47.73% and 50.92% in *M. japonicus* and *P. monodon*, respectively) indicated the presence of abundant repetitive sequences.

The sequence polymorphism rate was calculated based on single-nucleotide polymorphisms (SNPs) and short insertion/deletion (Indels) according to previous researches (Kao et al., 2016). Burrows-Wheeler Aligner (BWA) was used to measure the level of heterozygosity by aligning sequencing reads to the genome (Li and Durbin, 2010). SAMtools was used to call SNPs and Indels from the alignment results (Li et al., 2009). Finally, 2,969,278 SNPs and 637,450 Indels were detected in the *M. japonicus* genome, yielding a sequence polymorphism rate of 0.19%. Besides, a sequence polymorphism rate of 0.21% (3,562,719 SNPs and 711,744 Indels) was detected in the *P. monodon* genome.

2.3. Genome assembly

A *de novo* assembly procedure was performed on the clean reads using SOAPdenovo2 with the k value set from 31 to 99 (Luo et al., 2012). And the assembly was improved by using L_RNA_scaffolder, which can use long single-end RNA-seq reads to order, orient and combine genomic fragments into larger sequences (Xue et al., 2013). Finally, a total length of 1.94 Gb scaffolds with N50 length of 937 bp were produced in *M. japonicus*; and for *P. monodon*, 2.04 Gb scaffolds with N50 length of 786 bp were obtained (Table 2), which was comparable to that of *L. vannamei* (Yu et al., 2015) and *N. denticulata* (Kenny et al., 2014).

2.4. Estimation of genome completeness

We collected the transcriptome data of two shrimp species from NCBI SRA database (accession no. of *M. japonicus*: SRX2030618; and accession no. of *P. monodon*: SRX110649, SRX110651, SRX110652, SRX1333495, SRX1333568, SRX1333569, SRX1333570, SRX757561). The transcriptome data were assembled by Trinity (Haas et al., 2013), and removed isoforms by TGICL (Pertea et al., 2003). There were 80,444 and 89,473 unigenes assembled in *M. japonicus* and *P. monodon*, respectively (Supplementary materials 2 and 3, Table S2). We downloaded 2885 ESTs of *M. japonicus* and 424 complete genes of *P. monodon* from NCBI GenBank, and compared it with the *de novo* assembled unigenes. > 90% of these sequences were covered by unigenes,

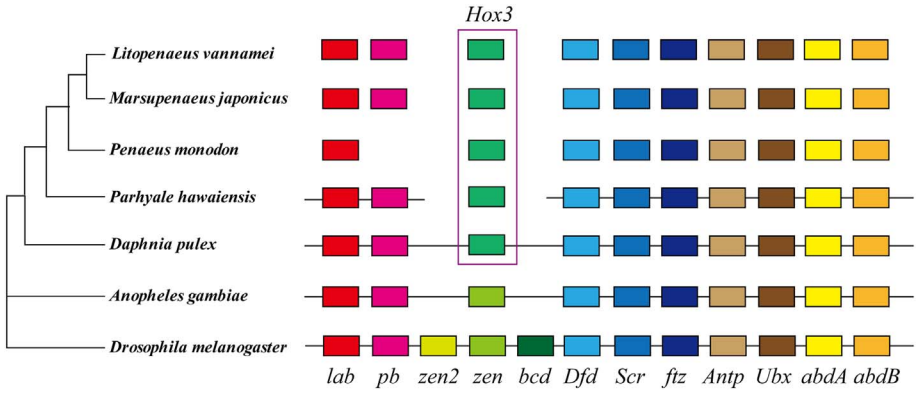

Fig. 1. Hox gene cluster of penaeid shrimps. The relevant information was referred and modified to (Yuan et al., 2017). The lines that connect each gene indicates they are synteny in the same scaffold. *zen2*, *zen* and *bcd* are three homologous genes in *D. melanogaster*, but only one homologous gene (*Hox3*) was identified in crustaceans.

and > 80% of them were covered by single unigene in half length, indicating the accuracy and completeness of these *de novo* assembled unigenes. For both assemblies, over 97% of unigenes could be covered by the genome, and over 82% of unigenes whose 50% length of the sequences could be covered by single scaffold (Table S3), which indicating highly genome completeness. Besides, it provide valuable resources for gene structure analysis.

Besides, the genome completeness was also estimated by 248 conserved eukaryotic genes recovered by CEGMA 2.4 (Parra et al., 2007). About 82.66% and 87.10% core genes were covered by the two genomes, respectively (Table S3).

2.5. Phylogenetic analysis

To understand the phylogeny of penaeid shrimps, we performed phylogenetic analysis based on the 13 mitochondrial genes of crustaceans. The amino acid sequences were completely aligned using MUSCLE 3.6 (Edgar, 2004), and the maximum likelihood analysis was performed using PhyML for 1000 bootstraps with the substitution model of MtREV + I (Guindon and Gascuel, 2003). The topology of the phylogenetic tree was consistent with many previous researches (De Grave et al., 2015). Penaeoidea and Caridea were monophyletic and phylogenetically close with each other (Fig. S2). They were nested within Stomatopoda at the basal branch of decapods. *M. japonicus* and *P. monodon* were phylogenetically close to the last common ancestor of Penaeiodea.

2.6. Repetitive elements analysis

A local database of repetitive elements was constructed by RepeatModeler, and RepeatMasker was used to identify the transposable elements (TEs) by aligning the genome sequences against RepBase (RepBase21.04) and the local database (Tarailo-Graovac and Chen, 2009; Bao, 2015). A total of 626 Mb (34.96%) and 833 Mb (46.68%) repeats were identified in the genome of *M. japonicus* and *P. monodon*, respectively (Table S4). LINEs and DNA transposons were two major TEs among two shrimp genomes. LINEs are the most abundant TEs that account for 12.41% of the *P. monodon* genome, which support the view from E. de la Vega's research (de la Vega et al., 2007). RTE-RTE (1.29%) was the major LINEs in *M. japonicus*; RTE-BovB (4.96%) and LINE1 (2.03%) were two major types of LINEs in *P. monodon*. Penelope elements have been detected in the genomes of *P. monodon* and *M japonicus* that account for 0.82% and 0.65% of the genomes, respectively. For DNA transposons, En-Spm and Maverick were abundant in two genomes. Similar with that of *L. vannamei* and *E. carinicauda*, Gypsy and RTE-BovB were two abundant retrotransposons that commonly present in shrimp genomes (Yuan et al., 2017; Zhao et al., 2012). Besides, *M. japonicus* and *P. monodon* contain more LINE1 than that of *E. carinicauda*.

A great many simple sequence repeats (SSR), which account for about 10% of the genome, was also identified in the two genomes. The two genomes showed similar distribution pattern of SSR that dinucleotide SSR ((AG)n and (AC)n) were the most abundant (> 4% of genome), whereas (AT)n and (CG)n were merely detected in two genomes (Fig. S3). When comparing the two genomes, it seems *P. monodon* contains relative more trinucleotide SSR ((AAT)n, (AAG)n and (ATC)n) than that of *M. japonicus*, while *M. japonicus* contains relative more mononucleotide SSR ((A)n and (C)n) than that of *P. monodon*.

2.7. Gene predictions

A preliminary annotation of two shrimp genomes was constructed by three approaches, the homolog-based predictions by Genewise (version 2.2.0) (Birney et al., 2004), *de novo* predictions by Augustus (version 2.5.5) (Hmajoros et al., 2004), and transcriptome-based predictions by Tophat (version 2.0.8) (Trapnell et al., 2009). All gene evidences were combined by EVM into a weighted and non-redundant consensus of the gene structures (Ruiz-Trillo et al., 2008). Finally, we obtained 16,734 gene features in *M. japonicus* and 18,115 gene features in *P. monodon* (Supplementary materials 4 and 5, Table S3). Among these gene features, there are 4845 candidate orthologs showed pairwise best blast hit between two genomes.

2.8. Hox gene cluster

The Hox gene cluster of various species contains 10 conserved Hox genes, which play an important role in the development and morphology of eukaryotic organisms. When blasted genomes against the Hox genes of *D. pulex* and *P. hawaiensis*, we identified the Hox genes in the two genomes. In comparison to other crustaceans, the 10 Hox genes were almost present in the two genomes, except for *pb*, which has not been identified in the genomes of *P. monodon* (Fig. 1).

2.9. Horizontally transferred genes (HTGs)

White spot syndrome virus (WSSV) and infectious hypodermal and hematopoietic necrosis virus (IHHNV) are the two best-known viruses that can transmit horizontally among different kinds of shrimps and cause disastrous diseases in crustaceans (Soowannayan and Phanthura, 2011). Recently, homologous WSSV genes have been found in the genome of *M. japonicus* (a BAC clone Mj024A04) and *P. monodon* (a fosmid library) (Dang et al., 2010; Huang et al., 2011). It implies that horizontal gene transfer (HGT) might have happened between shrimps and WSSV. In order to clarify these HGT events, we excluded probable contaminate sequences that showed significant similarity (identity of 98%~100%) with WSSV and IHHNV sequences from NCBI nt database, and then compared the two shrimp genomes against the genome of WSSV (accession no. AF369029.2) and IHHNV (accession no. JN377975.1) *via* BLASTN analysis with E-value cutoff of 1E-05. However, none homologous genome regions have been detected except some short DNA segments (~30 bp). Then, we compared the genes of the two shrimp genomes against the genes of WSSV and IHHNV *via* BLASTX analysis. None of genes showed homologous to IHHNV genes. But 13 genes of *M. japonicus* and 15 genes of *P. monodon* showed homologous to WSSV genes with moderate similarity (21%–63% identities) (Table S5), which was similar to the results of Mj024A04. Whereas these genes showed higher similarity to the genes of other eukaryotic species (41%–97% identities).

We also analyzed genome-wide HGT events in the two shrimp genomes following the methods of previous researches (Yuan et al., 2013). We identified 16 candidate HTGs between two shrimp genomes (Table S6). However, these sequences could also be contaminating bacterial sequences, that need further confirmation. There are 13 candidate HTGs shared by two shrimp genomes except for *Ankp*, *CTC*, and *mtkA*. Among the 14 HTGs identified in *L. vannamei* (Yuan et al., 2013), there are 8 genes of *M. japonicus* and 9 genes of *P. monodon* were detected, which indicates these genes may be transferred from a bacterial source to the ancestors of penaeid shrimps.

We also identified nuclear mitochondrial DNA segments (NUMTs) that horizontally transferred from mitochondrial genome to nuclear genome according to previous researches (Yuan et al., 2017). Finally, 90 NUMTs (total length of 73,207 bp) and 26 NUMTs (total length of 38,633 bp) were identified in the genome of *M. japonicus* and *P. monodon*, respectively.

3. Data availability

All the sequencing reads were deposited in the NCBI SRA database under the accession number of SRR5620465-SRR5620468. Draft assemblies were deposited in the GenBank under the accession number of NIUS00000000 and NIUR00000000.

Conflicts of interest

None.

Authors' contributions

JX, FL and XZ conceived the study. XZ, YY, and JW collected the samples, isolated DNA and performed sequencing. FL and CL provided advices about the analysis. JY performed the analysis and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by National Natural Science Foundation of China (41506189, 31672632, 41376165), China Agriculture Research system-48 (CARS-48), and The Scientific and Technological Innovation Project Financially Supported by Qingdao National Laboratory for Marine Science and Technology (2015ASKJ02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2017.12.006.

References

Abdelrahman, H., ElHady, M., Alcivar-Warren, A., Allen, S., Al-Tobasei, R., Bao, L., Beck, B., Blackburn, H., Bosworth, B., Buchanan, J., Chappell, J., Daniels, W., Dong, S., Dunham, R., Durland, E., Elaswad, A., Gomez-Chiarri, M., Gosh, K., Guo, X., Hackett, P., Hanson, T., Hedgecock, D., Howard, T., Holland, L., Jackson, M., Jin, Y., Kahlil, K., Kocher, T., Leeds, T., Li, N., Lindsey, L., Liu, S., Liu, Z., Martin, K., Novriadi, R., Odin, R., Palti, Y., Peatman, E., Proestou, D., Qin, G., Reading, B., Rexroad, C., Roberts, S., Salem, M., Severin, A., Shi, H., Shoemaker, C., Stiles, S., Tan, S., Tang, K.F., Thongda, W., Tiersch, T., Tomasso, J., Prabowo, W.T., Vallejo, R., van der Steen, H., Vo, K., Waldbieser, G., Wang, H., Wang, X., Xiang, J., Yang, Y., Yant, R., Yuan, Z., Zeng, Q., Zhou, T., 2017. Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. BMC Genomics 18, 191.

Bao, W., 2015. Non-LTR retrotransposons from the shrimp genome. Repbase Rep. 15, 1592.

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res. 14, 988.

Dang, L.T., Koyama, T., Shitara, A., Kondo, H., Aoki, T., Hirono, I., 2010. Involvement of WSSV-shrimp homologs in WSSV infectivity in kuruma shrimp: *Marsupenaeus japonicus*. Antivir. Res. 88, 217–226.

De Grave, S., Chan, T.Y., Chu, K.H., Yang, C.H., Landeira, J.M., 2015. Phylogenetics reveals the crustacean order Amphionidacea to be larval shrimps (Decapoda: Caridea). Sci. Rep. 5, 17464.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

FAO, Yearbook of Fisheries Statistics Summary Tables, 2013. Food and Agriculture Organization of the UnitedNations (FAO), Rome, Italy. 2015.

Farfante, I.P., Kensley, B., 1997. Penaeoid and Sergestoid Shrimps and Prawns of the World. Keys and Diagnoses for the Families and Genera. Memories du Museum National D'Historie Naturelle, Paris, France.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.

Hmajoros, W., Pertea, M., Lsalzberg, S., 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879.

Huang, S.W., Lin, Y.Y., You, E.M., Liu, T.T., Shu, H.Y., Wu, K.M., Tsai, S.F., Lo, C.F., Kou, G.H., Ma, G.C., Chen, M., Wu, D.Y., Aoki, T., Hirono, I., Yu, H.T., 2011. Fosmid library end sequencing reveals a rarely known genome structure of marine shrimp *Penaeus monodon*. BMC Genomics 12.

Kao D, Lai AG, Stamataki E, Rosic S, Konstantinides N, Jarvis E, Di Donfrancesco A, Pouchkina-Stantcheva N, Semon M, Grillo M, Bruce H, Kumar S, Siwanowicz I, Le A, Lemire A, Extavour C, Browne W, Wolff C, A, M, The genome of the crustacean *Parhyale hawaiensis*: a model for animal development, regeneration, immunity and lignocellulose digestion, bioRxiv, (2016).

Kenny, N.J., Sin, Y.W., Shen, X., Zhe, Q., Wang, W., Chan, T.F., Tobe, S.S., Shimeld, S.M., Chu, K.H., Hui, J.H.L., 2014. Genomic sequence and experimental tractability of a new decapod shrimp model, *Neocaridina denticulata*. Mar. Drugs 12, 1419–1437.

Koyama, T., Asakawa, S., Katagiri, T., Shimizu, A., Fagutao, F.F., Mavichak, R., Santos, M.D., Fuji, K., Sakamoto, T., Kitakado, T., Kondo, H., Shimizu, N., Aoki, T., Hirono, I., 2010. Hyper-expansion of large DNA segments in the genome of kuruma shrimp, *Marsupenaeus japonicus*. BMC Genomics 11, 141.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, R.Q., Fan, W., Tian, G., Zhu, H.M., He, L., Cai, J., Huang, Q.F., Cai, Q.L., Li, B., Bai, Y.Q., Zhang, Z.H., Zhang, Y.P., Wang, W., Li, J., Wei, F.W., Li, H., Jian, M., Li, J.W., Zhang, Z.L., Nielsen, R., Li, D.W., Gu, W.J., Yang, Z.T., Xuan, Z.L., Ryder, O.A., Leung, F.C.C., Zhou, Y., Cao, J.J., Sun, X., Fu, Y.G., Fang, X.D., Guo, X.S., Wang, B., Hou, R., Shen, F.J., Mu, B., Ni, P.X., Lin, R.M., Qian, W.B., Wang, G.D., Yu, C., Nie, W.H., Wang, J.H., Wu, Z.G., Liang, H.Q., Min, J.M., Wu, Q., Cheng, S.F., Ruan, J., Wang, M.W., Shi, Z.B., Wen, M., Liu, B.H., Ren, X.L., Zheng, H.S., Dong, D., Cook, K., Shan, G., Zhang, H., Kosiol, C., Xie, X.Y., Lu, Z.H., Zheng, H.C., Li, Y.R., Steiner, C.C., Lam, T.T.Y., Lin, S.Y., Zhang, Q.H., Li, G.Q., Tian, J., Gong, T.M., Liu, H.D., Zhang, D.J., Fang, L., Ye, C., Zhang, J.B., Hu, W.B., Xu, A.L., Ren, Y.Y., Zhang, G.J., Bruford, M.W., Li, Q.B., Ma, L.J., Guo, Y.R., An, N., Hu, Y.J., Zheng, Y., Shi, Y.Y., Li, Z.Q., Liu, Q., Chen, Y.L., Zhao, J., Qu, N., Zhao, S.C., Tian, F., Wang, X.L., Wang, H.Y., Xu, L.Z., Liu, X., Vinar, T., Wang, Y.J., Lam, T.W., Yiu, S.M., Liu, S.P., Zhang, H.M., Li, D.S., Huang, Y., Wang, X., Yang, G.H., Jiang, Z., Wang, J.Y., Qin, N., Li, L., Li, J.X., Bolund, L., Kristiansen, K., Wong, G.K.S., Olson, M., Zhang, X.Q., Li, S.G., Yang, H.M., Wang, J., Wang, J., 2010. The sequence and de novo assembly of the giant panda genome. Nature 463, 311–317.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1, 18.

Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genornes. Bioinformatics 23, 1061–1067.

Patel, R.K., Jain, M., 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7, e30619.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., Gene Indices, T.I.G.R., 2003. Clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652.

Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., Lang, B.F., 2008. A phylogenomic investigation into the origin of metazoa. Mol. Biol. Evol. 25, 664–672.

Soowannayan, C., Phanthura, M., 2011. Horizontal transmission of white spot syndrome virus (WSSV) between red claw crayfish (*Cherax quadricarinatus*) and the giant tiger shrimp (*Penaeus monodon*). Aquaculture 319, 5–10.

M. Tarailo-Graovac, N. Chen, 2009, Using RepeatMasker to identify repetitive elements in genomic sequences, Curr. Protoc. Bioinformatics, Andreas D. Baxevanis, et al., Chapter 4 Unit 4 10.

Trapnell, C., Pachter, L., Lsalzberg, S., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105.

de la Vega, E., Degnan, B.M., Hall, M.R., Wilson, K.J., 2007. Differential expression of immune-related genes and transposable elements in black tiger shrimp (*Penaeus monodon*) exposed to a range of environmental stressors. Fish Shellfish Immunol. 23, 1072–1088.

Wilson, K., Cahill, V., Ballment, E., Benzie, J., 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? Mol. Biol. Evol. 17, 863–874.

Xue, W., Li, J.T., Zhu, Y.P., Hou, G.Y., Kong, X.F., Kuang, Y.Y., Sun, X.W., 2013. L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics 14, 604.

Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., Huang, L., Zheng, H., Xiang, J., 2015. Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. Sci. Rep. 5, 15612.

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.

Yuan, J.B., Gao, Y., Zhang, X.J., Wei, J.K., Liu, C.Z., Li, F.H., Xiang, J.H., 2017. Genome sequences of marine shrimp *Exopalaemon carinicauda* Holthuis provide insights into genome size evolution of Caridea. Mar. Drugs 15.

Zhao, C., Zhang, X.J., Liu, C.Z., Huan, P., Li, F.H., Xiang, J.H., Huang, C., 2012. BAC end sequencing of Pacific white shrimp *Litopenaeus vannamei*: a glimpse into the genome of Penaeid shrimp. Chin. J. Oceanol. Limnol. 30, 456–470.

Table S1. Summary of sequencing data of two penaeid shrimp species.

	Insert size/bp	Lib	Raw data/G	Clean data/G	*Coverage/X
M. japonicus	230	DES00237	44.53	43.7	19.16
	230	DES00238	39.6	38.92	17.07
	500	DES00234	48.73	44.88	19.68
	Total		**132.86**	**127.5**	**55.92**
P. monodon	230	DES00239	45.09	44.29	17.10
	230	DES00240	42.15	41.39	15.98
	500	DES00235	45.59	41.62	16.07
	Total		**132.83**	**127.3**	**49.15**

* The sequencing coverage was estimated according to the estimated genome size: *M. japonicus* (2.28Gb) and *P. monodon* (2.59Gb).

Table S2. Transcriptome assembly of two shrimp species*.

	M. japonicus	***P. monodon***
Unigene number:	80,444	89,473
Total length:	65,334,630	97,948,701
Longest:	15,301	21,917
Shortest:	201	201
N50:	1,537	2,315
N90:	297	373
Mean length:	812	1,094

* "Total length" indicates the total length of unigenes. "N50" indicates the unigene length such that 50% of the *de novo* assembled sequences lies in unigenes of this size or larger. "N90" indicates the unigene length such that 90% of the *de novo* assembled sequences lies in unigenes of this size or larger.

Table S3. Summary of unigenes and core genes coverage, and annotated gene features.

		Unigenes	Matched unigenes	90% in one scaffold	50% in one scaffold	CEGMA complete	CEGMA partial	Annotated genes
M. japonicus	Number	80,444	79,343	35,208	66,281	111	205	16,734
	Percent		98.63%	43.77%	82.39%	44.76%*	82.66%*	
P. monodon	Number	89,473	86,980	40,478	73,513	115	216	18,115
	Percent		97.21%	45.24%	82.16%	46.37%*	87.10%*	

* The percentage of CEGMA complete and partial was calculated as the rate of (matched CEGMA genes)/(248 core eukaryotic genes).

Table S4. Summary of repetitive sequences*.

		M. japonicus	*P. monodon*
Genome size:		1,793,469,144	1,785,683,792
GC level:		39.19%	41.18%
Repeat total length:		626,997,403	833,561,962
Repeat percent:		34.96%	46.68%
SINEs:		0.03%	1.44%
LINEs:		4.75%	12.41%
	RTE-BovB	0.41%	4.96%
	RTE-RTE	1.29%	0.60%
	LINE1	0.73%	2.03%
	Penelope	0.65%	0.82%
LTR elements:		1.14%	2.15%
	Gypsy	0.82%	1.77%
	Copia	0.04%	0.03%
DNA elements:		5.66%	2.01%
	En-Spm	1.07%	0.84%
	Maverick	1.78%	0.01%
Unclassified:		7.19%	11.84%
Total interspersed repeats:		18.77%	29.84%
Small RNA:		0.05%	0.05%
Satellites:		0.35%	0.12%
Simple repeats:		9.79%	10.90%
Low complexity:		6.28%	5.89%

* SINEs indicates short interspersed transposable elements; LINEs indicates long interspersed transposable elements; LTR elements indicates long terminal repeat; TEs indicates transposable elements.

Table S5. Summary of genes showed homologous to WSSV genes in two shrimp genomes*.

Gene ID	WSSV gene			Best hit gene			
	ID	Identity	E-value	GI list	Species	Identity	E-value
Mjap_1664_1549	ALN66462.1_7	63.29	4.00E-113	240129500	*Litopenaeus vannamei*	97.23	1.00E-168
Mjap_1966_2164	ALN66504.1_29	58.79	2.00E-105	585704556	*Elephantulus edwardii*	81.35	4.00E-154
Mjap_22641_881	ALN66444.1_1	38.32	3.00E-14	328723513	*Acyrthosiphon pisum*	55.49	1.00E-20
Mjap_29340_2037	ALN66504.1_29	54.75	1.00E-106	664717565	*Equus przewalskii*	74.05	4.00E-153
Mjap_31502_3196	ALN66570.1_70	39.49	2.00E-37	443697237	*Capitella teleta*	57.84	6.00E-62
Mjap_34513_2636	ALN66444.1_1	37.67	2.00E-40	646716259	*Zootermopsis nevadensis*	57.98	4.00E-176
Mjap_35504_3383	ALN66444.1_1	31.74	7.00E-14	568249260	*Anopheles darlingi*	58.62	9.00E-88
Mjap_37164_911	ALN66570.1_70	36.13	5.00E-27	675388265	*Stegodyphus mimosarum*	71.82	6.00E-71
Mjap_37926_1279	ALN66491.1_24	23.42	5.00E-11	410509310	*Litopenaeus vannamei*	75.82	4.00E-158
Mjap_38402_5887	ALN66491.1_24	21.33	8.00E-11	1067092364	*Hyalella azteca*	41.32	4.00E-178
Mjap_39931_4726	ALN66444.1_1	45.9	4.00E-11	1067085238	*Hyalella azteca*	53.13	4.00E-176
Mjap_41276_379	ALN66444.1_1	43.61	2.00E-12	919282941	*Mytilus coruscus*	52.01	2.00E-23
Mjap_42218_519	ALN66444.1_1	40.68	1.00E-24	542189173	*Oreochromis niloticus*	51.75	3.00E-27
Pmo_949_4732	ALN66444.1_1	40.07	7.00E-44	665798419	*Microplitis demolitor*	48.69	1.00E-93
Pmo_1839_3731	ALN66444.1_1	37.82	2.00E-19	568249260	*Anopheles darlingi*	56.64	4.00E-83
Pmo_3937_6421	ALN66444.1_1	35.12	1.00E-16	1067093956	*Hyalella azteca*	75.52	8.00E-116
Pmo_6177_1837	ALN66570.1_70	41.43	8.00E-42	443697237	*Capitella teleta*	55.56	1.00E-61
Pmo_18422_1708	ALN66504.1_29	54.67	7.00E-108	585704556	*Elephantulus edwardii*	75.2	2.00E-156
Pmo_29223_1142	ALN66462.1_7	63.29	5.00E-113	240129500	*Litopenaeus vannamei*	96.89	7.00E-167
Pmo_30303_1170	ALN66444.1_1	40.99	2.00E-13	642114770	*Oncorhynchus mykiss*	55.26	9.00E-17
Pmo_35976_2008	ALN66505.1_30	33.03	3.00E-12	260891838	*Zootermopsis nevadensis*	45.45	2.00E-77
Pmo_36313_2036	ALN66444.1_1	33.1	2.00E-12	1067113109	*Hyalella azteca*	62.45	2.00E-122
Pmo_43083_4907	ALN66444.1_1	45.39	4.00E-11	1067067763	*Hyalella azteca*	69.88	4.00E-114
Pmo_45056_1987	ALN66444.1_1	36.63	2.00E-13	1042307446	*Ictalurus punctatus*	44.56	2.00E-26
Pmo_50960_1104	ALN66512.1_34	29.41	4.00E-14	1242848202	*Crassostrea virginica*	39.28	3.00E-25
Pmo_54143_5600	ALN66444.1_1	34.51	8.00E-11	1067085238	*Hyalella azteca*	52.71	0
Pmo_57884_4353	ALN66570.1_70	35.48	4.00E-22	646700882	*Zootermopsis nevadensis*	43.31	0
Pmo_61447_5295	ALN66444.1_1	36.83	2.00E-95	926614132	Limulus polyphemus	69.78	5.00E-110

* The first column is the list of genes showed homologous to WSSV genes in two shrimp genomes. Gene ID head with "Mjap" are the genes of *M. japonicus*, and gene ID head with "Pmo" are the genes of *P. monodon*. The column in gray background is the homologous WSSV genes, and the column in green background is the best hit genes in BLAST results against nr database.

Table S6. Candidate HTGs in two shrimp genomes*.

Gene	Types	*M. japonicus* gene	*P. monodon* gene	GI	Tophit species	Function
Cata	B→S	Mjap_178_2645	Pmo_63882_1335	744393345	*Escherichia coli*	chloramphenicol acetyltransferase
Dhfr	B→S	Mjap_3158_2360	Pmo_47455_1433	515574968	*Enterovibrio calviensis*	dihydrofolate reductase
Omtp	B→S	Mjap_13193_836	Pmo_54783_1210	340552665	*Collimonas fungivorans*	O-methyltransferase family protein
Stat	B→S	Mjap_20999_1558	Pmo_43517_1473	331678336	*Shigella flexneri*	streptomycin 3"-adenylyltransferase
SDRp	B→S	Mjap_24184_919	Pmo_1342_888	186471973	*Burkholderia phymatum*	short-chain dehydrogenase/reductase SDR
RpC2	B→S	Mjap_33470_1305	Pmo_40404_2892	333019710	*Escherichia coli*	repressor protein C2
Acsf	B→S	Mjap_37669_2705	Pmo_46622_2085	67983253	*Desulfotomaculum kuznetsovii*	acetyl-coenzyme A synthetase family protein
Deha	F→S	Mjap_40283_1422	Pmo_50270_2545	163788026	*Debaryomyces hansenii*	DEHA2A03014p
Ankp	F→S		Pmo_59462_2958	260060705	*Grosmannia clavigera*	ankyrin repeat-containing protein
Hypo1	F→S	Mjap_32263_4787	Pmo_15399_287	751744249	*Oidiodendron maius*	hypothetical protein
TRZ2	B→S	Mjap_178_2645	Pmo_66854_839	501454092	*Neisseria gonorrhoeae*	TEM-1 beta lactamase
nlpC	B→S	Mjap_37245_1001	Pmo_5611_274	308114711	*Vibrio parahaemolyticus*	Cell wall-associated hydrolase
Hypo2	B→S	Mjap_38535_517	Pmo_4536_644	78033430	*Magnetospirillum gryphiswaldense*	conserved hypothetical protein
memp	B→S	Mjap_18214_542	Pmo_63606_306	636794747	*Burkholderia sp.*	putative membrane protein
CTC	B→S		Pmo_50459_1902	753950203	*Azoarcus sp.*	creatininase
mtkA	B→S		Pmo_86767_218	489695728	*Methylobacterium extorquens*	malate--CoA ligase subunit beta

* B→S indicates HGT from Bacteria to shrimps or its ancestor; F→S indicates HGT from Fungi to shrimps or its ancestor.

Figures

Figure S1. K-mer distribution of the sequencing data with the K-mer size of 17. K=17 represents the chosen length of substrings.

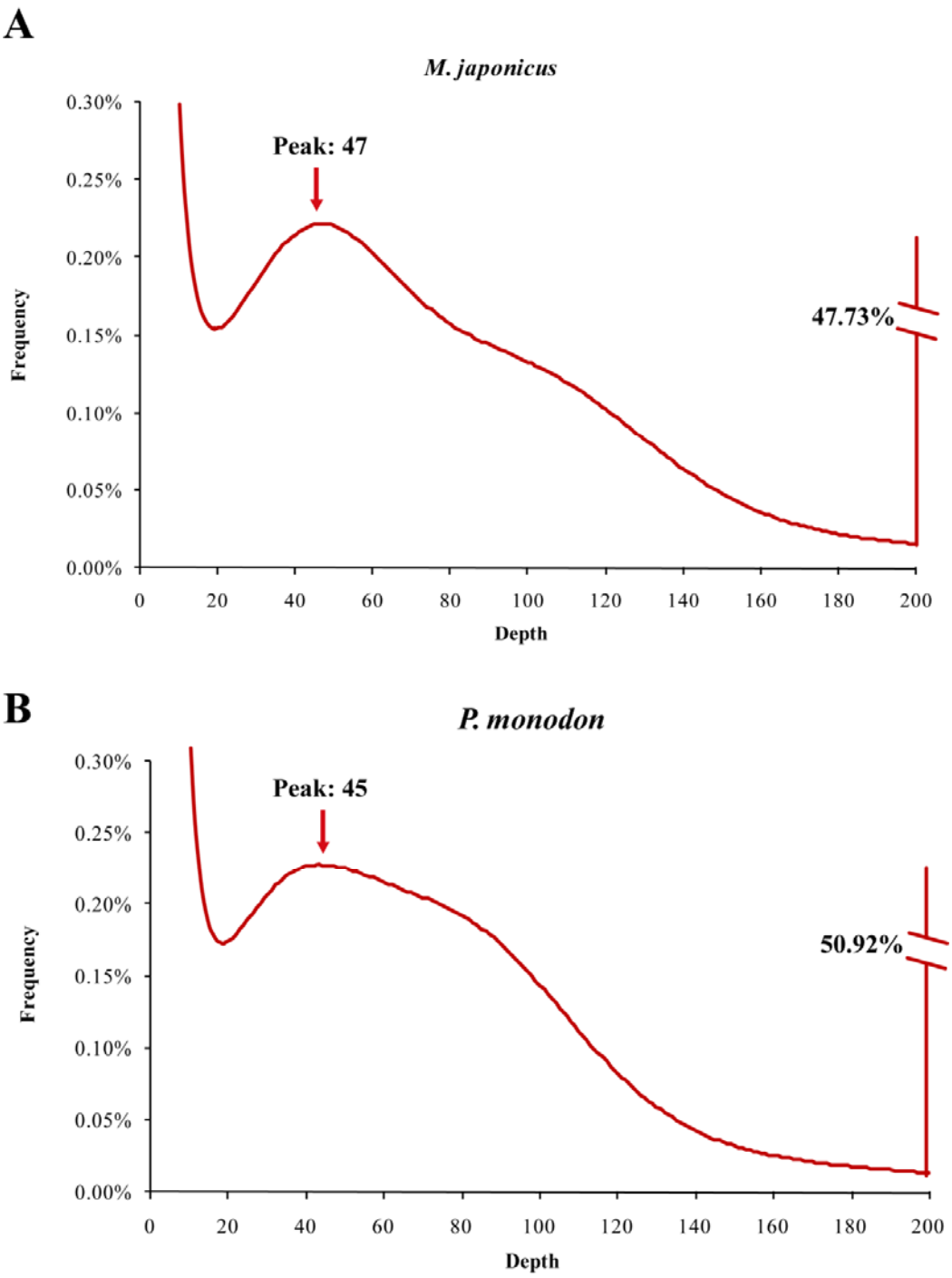

Figure S2. Phylogenetic tree of crustaceans. The phylogenetic tree was constructed based on the 13 mitochondrial genes (3720 amino acid sites) of crustaceans. the maximum likelihood analysis was performed using PhyML for 1000 bootstraps with the substitution model of MtREV + I. The bootstrap values of maximum likelihood analysis are displayed beside each node.

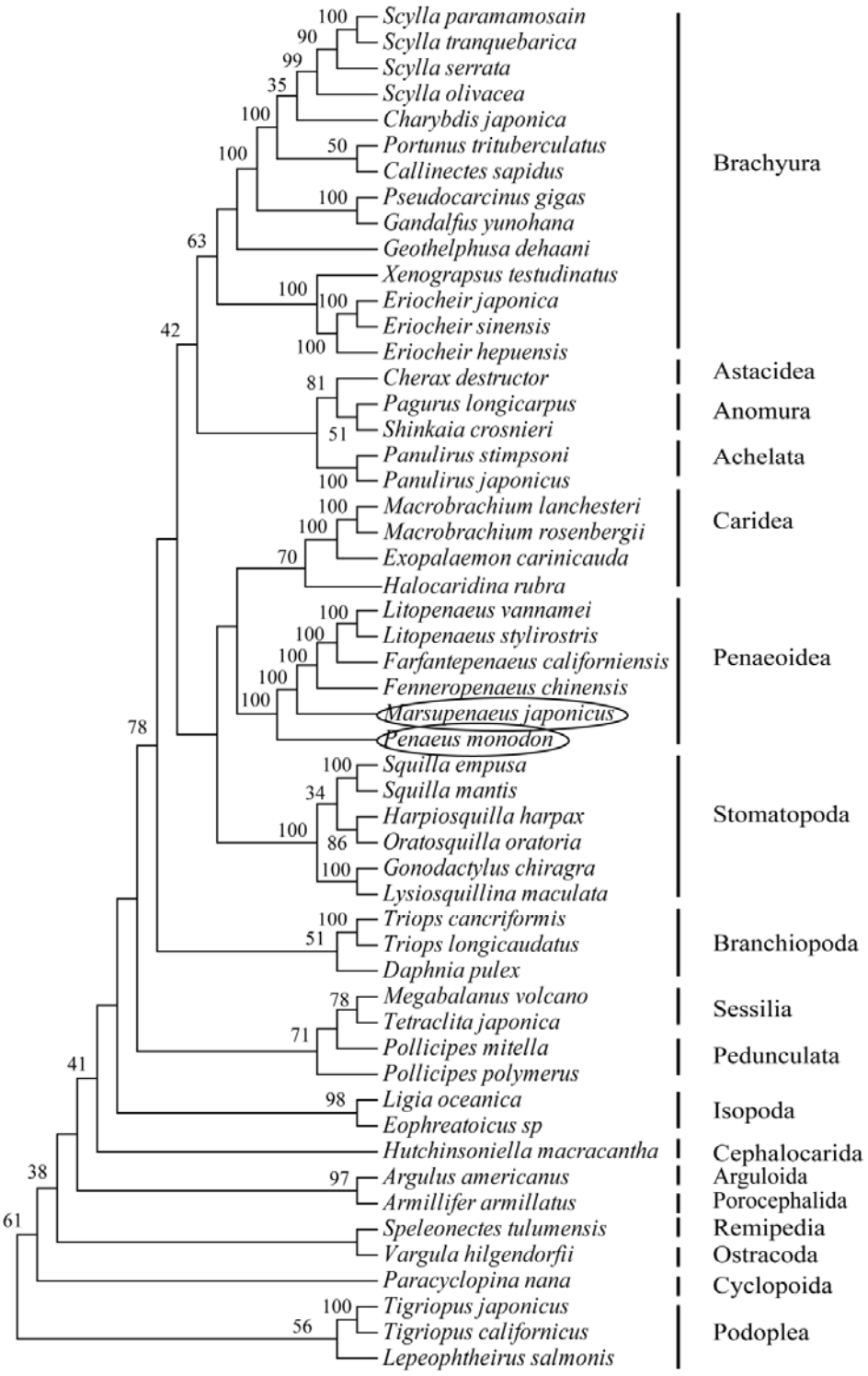

Figure S3. The distribution of SSR in two genomes. The lateral axis indicates the percentage of SSR in the full length of genomes.