Heliyon 7 (2021) e07291

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Genomic insights into the sessile life and biofouling of barnacles (Crustacea: Cirripedia)

Jack Chi-Ho Ip [a,b,c], Jian-Wen Qiu [a,b,c,*], Benny K.K. Chan [d,**]

[a] Department of Biology and Hong Kong Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Hong Kong Baptist University, Hong Kong
[b] Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China
[c] HKBU Institute of Research and Continuing Education, Virtual University Park, Shenzhen, China
[d] Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan

ARTICLE INFO

Keywords:
Barnacle
Biofouling
Evolution
Gene duplication
Genome
Underwater attachment

ABSTRACT

Members of the infraclass Cirripedia, commonly called barnacles, are unique among the subphylum Crustacea in that they exhibit a biphasic life cycle with a planktonic larval stage and a sessile adult stage. Understanding their unique sessile life and mechanisms of attachment are hampered by the lack of genomic resources. Here, we present a 746 Mb genome assembly of *Lepas anserifera* – the first sequenced stalked barnacle genome. We estimate that Cirripedia first arose ~495 million years ago (MYA) and further diversified since Mesozoic. A demographic analysis revealed remarkable population changes of the barnacle in relation to sea-level fluctuations in the last 2 MYA. Comparative genomic analyses revealed the expansion of a number of developmental related genes families in barnacle genomes, such as Br–C, PCP20 and Lola, which are potentially important for the evolution of metamorphosis, cuticle development and central nervous system. Phylogenetic analysis and tissue expression profiling showed the possible roles of gene duplication, functional diversification and co-option in shaping the genomic evolution of barnacles. Overall, our study provides not only a valuable draft genome for comparative genomic analysis of crustacean evolution, but also facilitates studies of biofouling control.

1. Introduction

Barnacles (Thecostraca: Cirripedia) are sessile crustaceans ubiquitously present in marine environments from the intertidal to the deep-sea, and from the tropics to the polar regions; some of them even live on other marine organisms such as whales and turtles (Chan et al., 2009; Buckeridge et al., 2018, 2019; Kim et al., 2020). The subclass Cirripedia is divided into three infraclasses that differ in morphology and way of life: Thoracica includes acorn and stalked barnacles with phosphatic or calcified shells (Chan and Høeg, 2015; Chan et al., 2021); Acrothoracica is characterized for not having shells but specifically burrowing into diverse calcareous substrates (Chan et al., 2014; Nielsen et al., 2016); Rhizocephala includes endoparasitic barnacles of other crustaceans (Anderson, 1994; Jung et al., 2019). Barnacles represent an excellent model system to study the evolution of body forms owing to their drastic changes from free-swimming planktonic naupliar and cyprid larval stages to sessile juveniles and adults with their forehead fixed to a solid substrate and their body placed upside down during the life cycle (Qiu et al., 1997; Qiu and Qian, 1999). Although many other crustaceans also possess the naupliar stage, other stages of the barnacle life cycle are unique. Unlike the body plan of many crustaceans with six thoracic segments and five abdominal segments, barnacles lack completely developed abdominal segments in the cyprid and sessile stages (Anderson, 1994). Among crustaceans, only barnacles have a cement gland that secrets a protein matrix for underwater adhesion, which is critical for permanent attachment (Liang et al., 2019).

The strong attachment of barnacles on underwater structures makes them an important group of biofouling organisms. Biofouling incurs economic loss by increase of fuel consumption on sailing vessels and damage of man-made submarine structures (Holm, 2012). Various antifouling coatings have been developed to control barnacle attachment, but many of them contain chemical ingredients that have serious toxic effects on non-target marine life (Qiu et al., 2005, 2008). For example, tributyltin, which had been widely used in antifouling paints before its ban by the International Convention on the Control of Harmful Anti-fouling Systems on Ships, is an endocrine disruptive chemical causing

* Corresponding author.
** Corresponding author.
E-mail addresses: qiujw@hkbu.edu.hk (J.-W. Qiu), chankk@gate.sinica.edu.tw (B.K.K. Chan).

https://doi.org/10.1016/j.heliyon.2021.e07291
Received 6 January 2021; Received in revised form 6 April 2021; Accepted 9 June 2021

abnormal shell formation in oysters and imposex in neogastropods (Chan et al., 2008a; Qiu et al., 2011; Yi et al., 2012; Ip et al., 2016).

There has been a substantial amount of research on the synthesis and properties of the barnacle cement, especially the roles of barnacle cement proteins (BCPs) in forming the larval adhesive plaque on solid substrates as well as its expansion in the adult stage (Zhong et al., 2014; Liang et al., 2019). Previous studies of BCPs have mainly relied on conventional biochemical methods including identifying their corresponding genes and determining their tissue-specific expressions in different life stages. Assembly of barnacle genomes will not only shed light on the molecular mechanisms of adhesion and evolution of this group of highly derived crustaceans, but also facilitate the investigation of underwater adhesives materials, as well as the development of environmental-friendly anti-fouling agents (Liang et al., 2019).

In this study, we aimed to sequence and assemble the genome of *Lepas anserifera* (Cirripedia, Lepadidae; Figure S1), a stalked fouling barnacle widely distributed in temperate and tropical seas, which grows in clusters attached to driftwood and other floating objects by its peduncle (Isaac and Moyse, 1990). Despite their important roles in fouling communities, no species from the family Lepadidae has a sequenced genome. Within the whole infraclass Cirripedia, only the acorn barnacle *Amphibalanus amphitrite* (Balanidae) has a sequenced genome (Kim et al., 2019). We conducted comparative genomic analyses with other pancrustacean lineages to understand barnacle genomic structures that are likely associated with their unique sessile life with particular emphasis on gene family expansion and contraction. We also produced tissue-specific transcriptomes and analyzed them in order to gain insight into the functions of candidate genes that are likely critical for adaptation in the sessile lifestyle including underwater adhesion.

2. Materials and methods

2.1. Sample collection, library construction, and sequencing

The stalked barnacles were collected by hand from floating buoys in Keelung City, Taiwan (25°08′31.0″N 121°48′08.7″E) in November 2018. They were transported to and cultured with their attached substrate for one week in the Biodiversity Research Center, Academia Sinica. An individual was selected for genome sequencing. To reduce the contamination by bacteria or algae, the shells of the barnacle were cleaned using a brush and rinsed with Milli-Q water before dissection, and the deshelled somatic tissues were rinsed using Milli-Q water. A part of the peduncle (stalk) was cut and fixed in pure ethanol for DNA extraction and genome sequencing. The peduncle, cirri and other tissues were immediately fixed in RNAlater (Invitrogen, CA, USA) for transcriptome sequencing. All tissue samples were stored in a freezer at -80 °C until use for DNA and RNA extraction.

Genomic DNA was extracted using the CTAB method (Porebski et al., 1997). DNA quality was evaluated and its quantity measured using agarose gel electrophoresis and a Qubit fluorometer (Thermo Fisher Scientific, MA, USA), respectively. High-quality DNA (OD 260/280 = 1.8–2.0, OD 260/230 = 2.0–2.2) was sent to Novogene Bioinformatics Technology (Beijing, China) for library preparation and whole genome sequencing using both the Illumina and PacBio platforms (Table S1). In brief, 1 μg DNA was used to construct an Illumina library with a 350-bp insert size, and sequenced on an Illumina NovaSeq sequencer to generate ~120 Gb paired-end reads with a read length of 150 bp. In addition, 10 μg DNA was used to construct a 20-kb SMRTbell library, and sequenced on a PacBio Sequel sequencer to generate ~18 Gb data. For RNA sequencing, total RNA was extracted from the barnacle tissues using TRIzol reagent (Thermo Fisher Scientific, MA, USA). The quality of the samples was determined with 1% agarose gel electrophoresis and the quantity of RNA was determined using a NanoDrop 2000c Spectrophotometer (ThermoFisher, MA, USA). The RNA samples were sent to Novogene Bioinformatics Technology (Beijing, China) for transcriptome library preparation using an Illumina NEBNext Ultra RNA Library Prep Kit (New England Biolabs, MA, USA) and sequenced using an Illumina NovaSeq sequencer to produce 150-bp paired-end reads (Table S1).

2.2. Genome assembly

Illumine short reads were trimmed to remove adaptors and low-quality reads (quality score <20, length <40 bp) using Trimmomatic v0.38 (Bolger et al., 2014). A kmer histogram was then generated with the clean Illumina data using Platanus v1.2.4 (Kajitani et al., 2019; settings: "-k 17 -s 10 -u 0.2 -t 29") to predict the genome size. GenomeScope (Vurture et al., 2017) was used to characterize the heterozygosity and contents of repetitive elements (Figure S2). FM-index Long Read Corrector (FMLRC) was used to correct the PacBio raw reads using Illumina reads (Wang et al., 2018; Wong et al., 2018). After trimming and error correction, a total of 788 million Illumina and 1.05 million PacBio reads (longest length = 120 kb, N50 = 18.2 kb) were retained (Table S1). Two bioinformatic approaches, including one based on the long-reads (PacBio data; wtdbg2 v2.5, Ruan and Li, 2020) and the other (hybrid one) based on both Illumina and PacBio data (MaSuRCA v3.4.1, Zimin et al., 2013) were applied to assemble the genome. A comparison of the assembly statistics showed that the wtdbg2 pipeline resulted in a better initial assembly, in term of genome size, length statistics and the percentage of complete Benchmarking Universal Single-Copy Orthologs (BUSCOs) (Table S2). To improve the initial assembly, the draft genome was further polished with PacBio and Illumina reads using the wtpoa-cns script implemented in wtdbg2. Due to the high heterozygosity rate of *Lepas* genome, Redundans (Pryszcz and Gabaldón, 2016) was performed to eliminate the haplotypic contigs. The curated contigs were scaffolded with PacBio data using SSPACE-LongRead v1.1 (Boetzer and Pirovano, 2014) and RNA-Seq data using P_RNA_scaffolder (Zhu et al., 2018), and then polished with wtpoa-cns. Possible contamination sequences (e.g., microbe, fungi, algae and virus) were removed by search using BLASTn against the NCBI database with an E-value threshold of 1e-20 and manual correction. The completeness of the final assembled sequences was assessed by analyzing the BUSCO scores (Simão et al., 2015) against the metazoan (metazoa_odb9) and arthropod (arthropoda_odb9) datasets.

2.3. Genome annotation and gene prediction

The barnacle genome was annotated using MAKER v3.0 (Cantarel et al., 2008). Before gene prediction, repetitive contents in the genome were identified using RepeatMasker (http://www.repeatmasker.org/) with RepBase library v20150807 (Bao et al., 2015), and species-specific repeat libraries in RepeatModeler v1.0.8 (Smit and Hubley, 2008-2015). Divergences of transposable elements from the consensus sequences and insertion time were determined with the script parseRM.pl implemented in RepeatMasker with a mutation rate of 3.1×10^{9} per site per year (Nunez et al., 2020). To provide transcriptomic evidence, *de novo* and genome-guided transcriptomes of the stalked barnacle were assembled using Trinity v2.5.1 (Haas et al., 2013) based on multiple tissues RNAseq data, and then merged with PASA pipeline v2.2.0 (Haas et al., 2003). Crustacean protein sequences from Swiss-Prot database and protein sequences for other crustacean genomes were used as protein evidence. Augustus 3.1 (Stanke and Morgenstern, 2005) and SNAP (Korf, 2004) were used for *ab initio* gene prediction. Results from different gene predictors were integrated into a consensus weighted annotation with EVidenceModeler v1.1.1 (EVM; Haas et al., 2008). The predicted genes were functionally annotated using Diamond v0.9.24 BLASTp (Buchfink et al., 2015) against NCBI non-redundant (nr) and UniProtKB databases under the "more-sensitive" option and an E-value threshold of 1e-5. Gene functional annotation was conducted using eggNOG-mapper v2 (Huerta-Cepas et al., 2017) for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and InterProScan5 (Jones et al., 2014) for InterPro and Pfam.

2.4. Heterozygous SNP detection and estimation of population history

The pairwise sequentially Markovian coalescent (PSMC) analysis was used to estimate the history of effective population sizes (Li and Durbin, 2011). The Illumina sequences of *Lepas anserifera* and *Amphibalanus amphitrite* (SRR9595623) were aligned back to the respective reference genome using BWA-MEM v0.7.12-r1039 under default settings. Heterozygous variants were called with the pipeline of SAMtools mpileup v1.7 (-C 50) with bcftools call v1.2 and vcfutils.pl vcf2fq (-d 8, -D 100). The PSMC analysis was run under the following settings: -t 15 -r 5 -p 4+25*2+4+6 and 100 bootstrap replicates assessed variance of the simulated results. Effective population size was calculated using a mutation rate of 3.1×10^{-9} substitutions per site per year (adopted from the acorn barnacle *Semibalanus balanoides*; Nunez et al., 2020) and a generation time of 1 year.

2.5. Phylogenomic and gene family analyses

Orthologous groups among nine arthropod genomes [i.e., *Amphibalanus amphitrite* (GCA_009805615.1), *Daphnia pulex* (GCA_900092285.2), *Drosophila melanogaster* (GCA_000001215.4), *Lepas anserifera* (this study), *Parhyale hawaiensis* (GCA_001587735.2), *Penaeus vannamei* (GCF_003789085.1), *Portunus trituberculatus* (GCA_008373055.1), *Strigamia maritima* (GCA_000239455.1), *Tigriopus californicus* (GCA_007210705.1)], and two barnacle transcriptomes [i.e., *Semibalanus balanoides* (GHBL00000000.1) and *Pollicipes pollicipes* (GGJN00000000.1)] and species tree were identified using OrthoFinder v2.2.7 (Emms and Kelly, 2015) with Diamond v0.9.24 BLASTp under "more-sensitive" mode, >50% identity and an E-value threshold of 1e-10 selected. All single-copy genes were aligned using MUSCLE v3.8.31 (Edgar, 2004) and trimmed using TrimAL v1.3 (Capella-Gutiérrez et al., 2009). The aligned sequences were concatenated and used for phylogenetic analyses. A maximum-likelihood method (ML) phylogenetic analysis was conducted using IQ-TREE v1.6.9 (Nguyen et al., 2014), with the best model of LG+F+G4 selected by ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 bootstrapping replicates. A maximum parsimony (MP) analysis was conducted using PAUP4 (Swofford, 2001) with 1000 bootstrapping replicates. A Bayesian (BI) analysis was conducted using MrBayes v3.2.7 (Huelsenbeck and Ronquist, 2001) under the GTR+I+G model; four Markov chains were applied for 10 million generations, sampled every 1000 generations, and the first 25% discarded as the burn-in, respectively. MCMCtree implemented in PAML v4.8 (Yang, 2007) was used to estimate divergence times using the burn-in, sample frequency and number of samples as 10 million, 10,00 and 10,000, respectively. The node calibration among arthropods was based on the TIMETREE database (Hedges et al., 2006), i.e., 560–642 million year ago (MYA) for Arthropoda; 426–561 MYA for Pancrustacea; 467–494 MYA for Hexapoda. Based on the results of orthologous genes and estimated divergence times, gene family expansion and contraction were determined for each node using CAFÉ v4 (Han et al., 2013).

2.6. Transcriptomic analysis

Gene expression levels were normalized as Transcripts Per Kilobase Million (TPM) using Salmon v0.9.1 (Patro et al., 2017). One single RNA sample (n = 1) per tissue was prepared. Differentially expressed genes were determined using run_DE_analysis.pl script in Trinity with the edgeR method which utilizes statistical models of expected variation, i.e., Poisson and negative binomial distribution (Haas et al., 2013). Tissue-specific genes for the tissues were determined based on comparing expression levels across tissue types. Only genes with FDR <0.05 and TPM >5 were considered as overexpressed. Potential functions of these target genes were further assessed by GO enrichment analysis using GOseq v1.39.0 (Young et al., 2012).

Real-time PCR was employed to validate tissue expression patterns in twelve selected genes. Total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, MA, USA) and the first strand cDNA was synthesized by using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Massachusetts, USA). The primers were designed using the online NCBI Primer-BLAST tool (Table S11). Real-Time PCR was performed with KAPA SYBR® FAST qPCR Master Mix (2X) Kit (Kapa Biosystems, Massachusetts, USA) on StepOnePlus™ Real-Time PCR System (Applied Biosystems, Massachusetts, USA) through polymerase activation at 95 °C for 3 min, and annealing and extending at 60 °C for 20s with a total of 40 cycles. The specificity of primer pairs for the PCR amplification was checked by the melting curve method. Triplicates were applied for each gene, and the relative gene expression level was calculated based on the $2^{\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The cytochrome *b* gene was used as the internal standard gene as done in a previous study of barnacle gene expression (De Gregoris et al., 2009).

3. Results and discussion

3.1. General genome features

We generated approximately 118.2 Gb Illumina reads and 17.7 Gb PacBio reads (Table S1). Kmer analysis predicted the size of the *L. anserifera* genome to be 722 Mb (Figure S2A), which is relatively small among members of Thoracica (Cirripedia) whose genome sizes have been estimated (Figure S2B). The Illumina and PacBio sequences generated in this study corresponded to 162 and 24 times of the genome size, respectively. GenomeScope analysis of the Illumina reads predicted the heterozygosity and repetitive contents to be 4.21% and 33.8%, respectively (Figure S2A). Using a long-reads assembly approach coupled with correction using short reads, and elimination of haplotypic contigs with Redundance (Table S2), we produced a final haploid genome assembly of 746 Mb, which is very close to the estimated genome size. The final assembly consists of 2,220 contigs with an N50 of 0.67 Mb and GC content of 46.5% (Table 1). The contig size ranges from 10.1 kb to 3.16 Mb (Figure 1A). BUSCO analysis showed that the *L. anserifera* genome has 96.1% arthropod BUSCOs (94.6% complete +1.5% fragmented) and 95.0% metazoan BUSCOs (92.4% complete +2.6% fragmented). These *L. anserifera* BUSCO scores are comparable to those of the acorn barnacle *A. amphitrite*, and those of other pancrustaceans (Figure 1C; Table S3). The high quality of the *L. anserifera* assembly was also demonstrated by the high mapping rates of the Illumina DNA reads (95.2%), PacBio reads (98.3%) and *de novo* transcriptome (94.2%) (Table S4). A total of 28,040 protein-coding genes (PCGs) were predicted in the *L. anserifera* genome, with an average length of 1,517 bp. BUSCO analysis of the PCGs revealed 94.6% (84.8% complete +9.8% fragmented) arthropod BUSCOs and 94.5% (85.2% complete +9.3% fragmented) metazoan BUSCOs (Figure 1C; Table 1 and S3). Of the PCGs, 20,989 (74.9%) were found to be homologous to proteins in the NR/Uniport databases and 19,044 (67.9%) could be assigned functional domains using InterProScan5 (Table 1).

3.2. Repetitive sequences and genome size evolution

Repeat content analysis showed that 341.3 Mb (45.2%) of the *L. anserifera* genome assembly consists of transposable elements (TEs), whereas the *A. amphitrite* assembly comprises 272.7 Mb (44.7%) repetitive sequences (Table S5). These sequences, which amounted to 68 Mb more than *A. amphitrite*, accounted for most of the difference in size (48%) between the two assembled genomes. The two barnacles also differed in their TE composition, with *L. anserifera* containing more DNA transposons (7.1% vs 5.6% of total genome size) and long interspersed nuclear elements (LINEs, 5.7 vs 3.6%), and *A. amphitrite* having higher proportions of long terminal repeats (LTRs, 3.6% vs 2.0%) and unclassified TEs (28.0% vs 24.5%) (Figure 1B; Table S5). Since TEs have been shown to be a major contributor to the evolution of genomic size (Kidwell, 2002), we compared the TE divergence and insertion times. Higher levels of divergence were determined in most of the major TE classes (i.e.,

Table 1. Comparison of assembly and annotation statistics between the stalked barnacle genome assembly and the published acorn barnacle assembly.

	Lepas anserifera	*Amphibalanus amphitrite*
A. Genome assembly		
Estimated genome size (Mb)	722.3	481
Assembly total length (Mb)	756.4	613.4
Repeat content (Mb)	342.1 (45.2%)	274.1 (44.7%)
GC content (%)	46.5	49.8
Unknown Sequences (%)	0	0.46
No. of contig (scaffold)*	2,220	3,308 (2,240)
N50 (Mb)	0.67	0.31 (0.46)
Average length (Mb)	0.34	0.18 (0.27)
Max. length (Mb)	3.16	1.45 (2.22)
Min. length (kb)	10.1	1.7 (1.9)
No. of contig >10 Kb	2,220	3,168 (2,184)
BUSCO arthropoda#	C:94.6%, F:1.5%, M:3.9%	C:94.1%, F:2.3%, M:3.6%
BUSCO metazoa#	C:92.4%, F:2.6%, M:5.0%	C:93.4%, F:1.1%, M:5.5%
B. Gene prediction		
No. of genes	28,040	28,182
Average gene length (bp)	1,517	1,628
Average no. of exons per gene	4.84	5.79
Average exon length (bp)	313.6	280
BUSCO arthropoda#	C:84.8%, F:9.8%, M:5.4%	C:86.6%, F:6.6%, M:6.8%
BUSCO metazoa#	C:85.2%, F:9.3%, M:5.5%	C:85.6%, F:5.2%, M:9.2%
C. Annotation		
With hits	22,132	22,867
NCBI Nr	20,705	–
UniProtKB	20,706	–
eggNOG	18,141	–
InterPro	18,990	–
Pfam	16,449	–
GO	10,538	–
KEGG	11,073	–
COG	18,146	–

* *Lepas anserifera* genome only contains contigs, and *Amphibalanus amphitrite* genome contains both contigs and scaffolds.
\# C, number of complete BUSCOs; F, number of fragmented BUSCOs; M, number of missing BUSCOs.

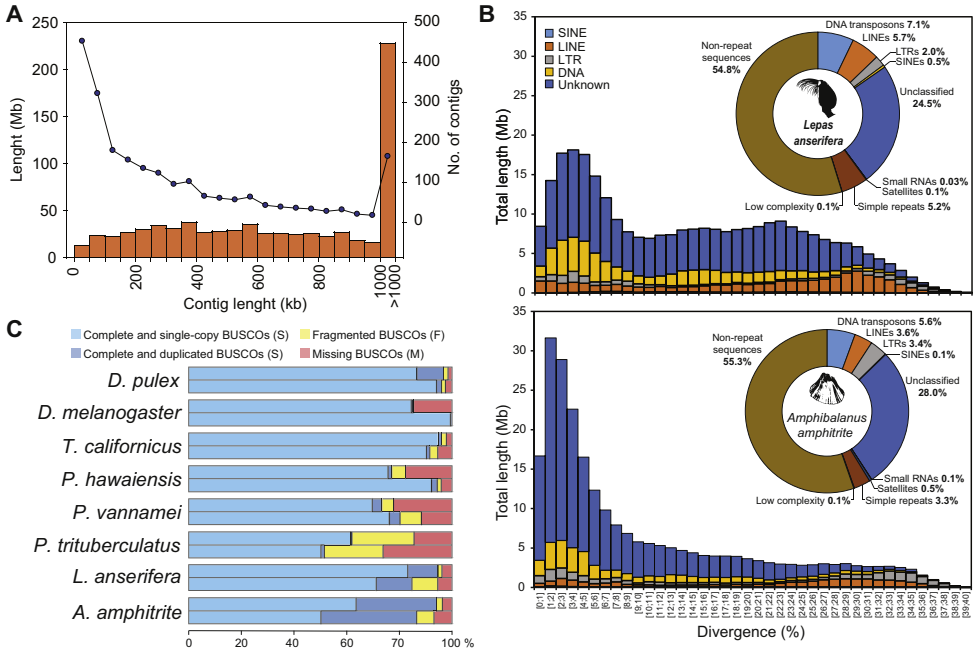

Figure 1. Genome characteristics of *Lepas anserifera*. (A) Length distributions of the assembled contigs. (B) Composition of repetitive sequences (doughnut plots) and transposable element (TE) landscape (histogram) of *L. anserifera* and *Amphibalanus amphitrite*. Percentage of divergence from consensus (by bins of 1%) is used as a proxy for age: the aged TEs have accumulated mutations with higher percentage of divergence; young TEs show little divergence from consensus. (C) Arthropoda BUSCO scores (top bar: genome; bottom bar: protein) of eight pancrustaceans.

LINEs and DNA transposons) of *L. anserifera* when compared with *A. amphitrite* (Figure 1B). TEs insertions were active in *L. anserifera* since ~100 million year ago (MYA) (~30% divergence), whereas the burst of TEs in *A. amphitrite* occurred in ~50 MYA (~15% divergence). These results indicate that the expansion of TEs in acorn and stalked barnacles occurred after the divergence of these two major barnacle lineages (Figure 3C), and such expansion may have contributed to the differences in their genome sizes.

3.3. Genomic variation and population history

We identified 2,976,254 and 2,106,553 heterozygous single-nucleotide polymorphisms (SNPs) in the *L. anserifera* and *A. amphitrite* genome, respectively. With these heterozygous SNPs, the demographic history of the two barnacle species from 2 MYA to 10 thousand year ago (KYA) was reconstructed using the Pairwise Sequentially Markovian Coalescent (PSMC) model (Figure 2). The population size of *L. anserifera* was estimated to be larger than that of *A. amphitrite*. The population histories of the two barnacle species showed a remarkable relation to the eustatic sea-level fluctuation. A drastic decline in effective population size was apparent ~120 KYA, coinciding well with the inception of the Last Glacial Period (115 KYA – 12 KYA). For coastal marine species (including barnacles), the rapid cooling and expansion of glaciers caused a drastic reduction in the sea level and led to a decrease in available coastal habitat, resulting in fragmented populations, genetic bottleneck and population restructuring (Ludt and Rocha, 2015; Tsang et al., 2012). In addition, since adult *L. anserifera* usually attaches to floating objects and are transported by ocean currents which may enhance their genetic connectivity (Buckeridge, 2012; Schiffer and Herbig, 2016), the genetic variation and population size of *L. anserifera* was estimated to be larger than those of the *A. amphitrite*, which usually inhabits rocky shores.

3.4. Phylogenetic relationships and divergence times

Phylogenetic relationships of pancrustaceans were constructed by including the two available barnacle genomes, six other pancrustacean genomes and two published barnacle transcriptomes with high completeness based on BUSCO analysis (Table S3). *Strigamia maritima* (Myriapoda) was used as the outgroup. Orthologous analysis identified that the *L. anserifera* assembly contained 10,358 gene families, in which 7,662 are shared with other crustaceans, and 2,669 are unique *to L. anserifera* (Figure 3A). The phylogenomic trees, reconstructed using Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian Inference (BI) methods based on 171 single-copy orthologs with 79,437 amino acid positions, consistently recovered with strong support *L. anserifera* as an early diverging Cirripedia. It is sister to a clade including *A. amphitrite*, *Pollicipes pollicipes* and *Semibalanus balanoides*. Together, these four species form the Cirripedia clade, which is sister to the Malacostraca clade. Cirripedia and Malacostraca form a clade that is sister to a clade comprising Copepoda and Hexapoda, although this conclusion is not fully supported in ML (bootstrap value 86%) or MP (bootstrap value 85%) analysis (Figure 3B). Furthermore, our analysis recovered Branchiopoda as sister to a clade comprising Cirripedia, Malacostrea, Copepoda and Hexapoda, but again this conclusion is not fully supported by ML or MP analysis.

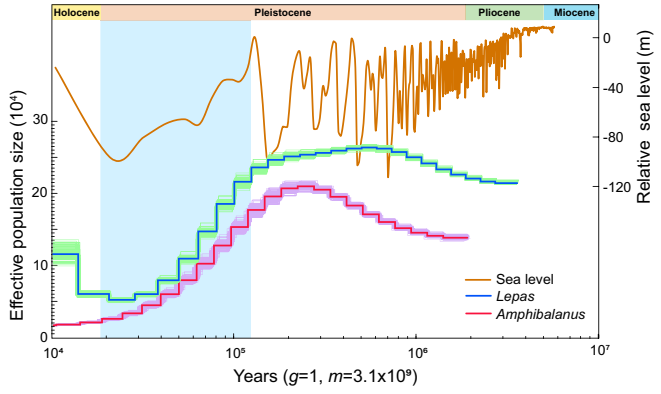

Figure 2. Demographic history inferred from barnacle genomes. Pairwise sequentially Markovian coalescent (PSMC) analysis was conducted using variant data from the sequenced individual. The blue and pink line represents the population size changes in *Lepas anserifera* and *Amphibalanus amphitrite*, respectively. The green and light purple lines, around the blue and pink lines, are the PSMC estimates on 100 rounds of bootstrapped sequences. The orange line denotes the fluctuation of the global sea level relative to the present day (de Boer et al., 2014). The light blue shading represents the Last Glacial Period in the Late Pleistocene. g, generation time; μ, mutation rate.

Despite the use of genome-scale data, previous studies did not fully resolve the phylogenetic relationships of the major groups of pancrustaceans, but suggested the two clades: i) Malacostraca, Copepoda, and Cirripedia; and ii) Branchiopoda, Remipedia, and Hexapoda (von Reumont et al., 2012; Lozano-Fernandez et al., 2019). To evaluate our tree topology, we performed the Approximately Unbiased (AU) test for all 105 possibilities among Branchiopoda, Copepoda, Hexapoda, Malacostraca and Cirripedia with the outgroup of Myriapoda. The AU topology test rejected the sister relationship between Copepoda and (Malacostraca, Cirripedia) at a significance level of 0.05. The statistical results supported the sister group relationship between Malacostraca and Cirripedia, but the phylogenetic positions of Branchiopoda, Copepoda and Hexapoda remain unsolved (Table S6). The divergence time estimation based on all single-copy orthologs showed that the Pancrustacea emerged between the Ediacaran and the Cambrian, and Cirripedia split with Malacostraca at the end of the Cambrian (~495 MYA). Cirripedia became diversified in Mesozoic, and *Lepas* separated from other groups of barnacles 154 MYA (95% confidence interval of 110–212 MYA; Figure 3C). Our divergence time estimation of Cirripedia (Sessilia—Pedunculata) is different from previous estimation of >500 MYA based on 18S genes (Pérez-Losada et al., 2004), but provides strong statistical support for the result of ~220 MYA in a multiple-gene analysis (Linse et al., 2013). Within the Cirripedia clade, our result showed that *Lepas* is sister to the clade containing *Pollicipes*, *Amphibalanus* and *Semibalanus*.

3.5. Dynamic evolution of Hox genes in crustaceans

The homeobox (*Hox*) genes are conserved transcription factors that play crucial roles in the development and segmentation of the anterior–posterior axis in metazoans (Sun et al., 2019; Ip et al., 2021). The ancestral arthropod has been hypothesized to possess ten *Hox* genes (eight with homeotic function and two *Hox*-like genes) (Grenier et al., 1997). To reveal the *Hox* cluster in Cirripedia, a BLASTp search of arthropod *Hox* genes was performed for the two barnacle genomes. The *L. anserifera* genome assembly encodes nine *Hox* genes in five contigs including *lab*, *pb*, *Hox3*, *Dfd*, *Scr*, *Diva*, *Antp*, *Ubx* and *Abd-B* (Figure 3D). Among them, *Diva*, the homologue of arthropod *ftz*, is located between *Scr* and *Antp* (contig669). Neither the *L. anserifera* assembly nor the *A. amphitrite* assembly has *Abd-A*. The missing of this gene was further confirmed by a phylogenetic analysis of the pancrustacean bithorax complex (Figure S3). In Cirripedia, the abdominal segments are reduced (Anderson, 1994). Previous studies of *Hox* genes in three cirripedes using molecular cloning and *in situ* hybridization revealed the expression of *Abd-B*, but not *Abd-A* – a gene required for the development of anterior abdomen in other non-barnacle crustaceans (Mouchel-Vielh et al., 1998; Blin et al., 2003). Previous studies of Rhizocephalan *Sacculina* barnacles, which has no distinct abdominal structures in adults but has an extensive root system, revealed no expression of *Abd-A* related *Hox* and *para-Hox* genes (*Antp*, *Ubx* and *caudal*) in the vestigial abdomen during development (Rabet et al., 2001; Mouchel-Vielh et al., 2002; Blin et al., 2003). However, Ascothoracida, the sister-group of Cirripedia possessing a

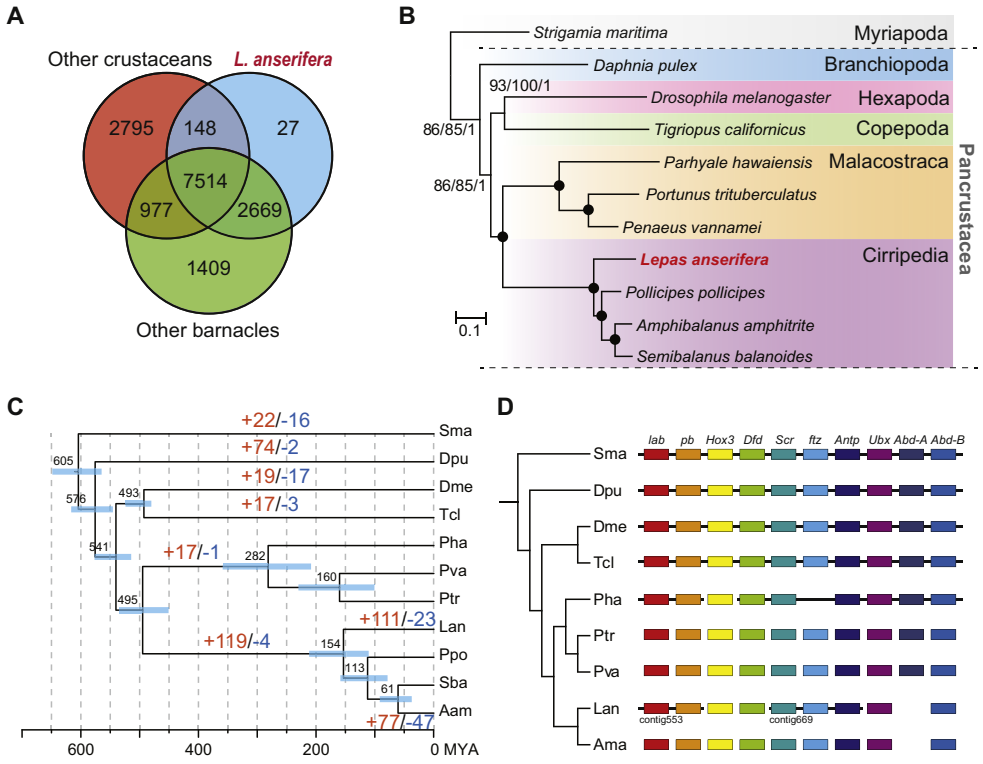

Figure 3. *Lepas anserifera* gene contents, and phylogenetic relationship, divergence times and Hox genes of pancrustaceans. (A) Venn diagram of the orthologous gene families among *L. anserifera*, barnacles and other crustaceans. (B) Phylogeny of arthropods inferred from 171 one-to-one orthologous genes (79,437 amino acid positions with 1.0–24.6% gaps among species) using ML, MP and BI methods with support values on nodes, respectively. Black circles on nodes indicate 100% bootstrap support from all three analyses. (C) Estimated divergence time in million year ago (MYA) among 11 pancrustaceans. Numbers on the branch show the lineage-specific expanded/contracted families. (D) Hox clusters among nine assembled genomes. The first letter of the genus name and the first two letters of the species name are used in abbreviations, e.g., Lan for *Lepas anserifera*.

well-developed abdomen, does have an *Abd-A*. Therefore, *Abd-A* might have been lost during the evolution of cirripedes, and this may be related to the conspicuous change in the body plan (reduction of the abdomen) as a result of their adoption of sessile life (Deutsch and Mouchel-Vielh, 2003).

In arthropods, *ftz* has been implicated in the early development (Hughes and Kaufman, 2002), but its function and expression profiles are highly divergent in different crustacean lineages (Gibson, 2000). For instance, in the water flea *Daphnia pulex ftz* is expressed in the posterior mandibular segment (Papillon and Telford, 2007), while the *ftz* homolog is missing in the amphipod *Parhyale hawaiensis* (Serano et al., 2016). In Cirripedia, *Diva* (the homologue of *ftz*) has lost its *Hox*-like expression pattern, and it is expressed solely in the central nervous system (CNS), suggesting its function in neuronal differentiation (Mouchel-Vielh et al., 2002). Unlike other crustaceans, the barnacle cyprid has a well-developed CNS and sense organs. This larva detects settlement cues and coordinates the substrate exploratory behaviors using a pair of antennules equipped with terminal and subterminal sensory setae (Harrison and Sandeman, 1999; Chan et al., 2018). During metamorphosis, the CNS system is largely reconstructed in adult barnacles with a high diversity of sensory setae on the cirri for detection of prey and water currents (Chan et al., 2008b). Therefore, the evolution of *Diva* and its CNS specific transcriptional profile might have facilitated the development of CNS – a key adaptation of barnacles to the sessile adult life.

3.6. Gene family expansion provides insight into environmental adaptation

Reciprocal BLAST analysis of *L. anserifera* and *A. amphitrite* gene sets identified >85% genes with homologs and 10,212 one-to-one orthologs in these two species (Table S7). As changes in gene family size are one of the primary drivers of evolution and adaptation (Demuth and Hahn, 2009), we compared gene family sizes between cirripedes and other pancrustaceans. To ensure the reliability of the analysis, the genome of *Portunus trituberculatus* with relatively low BUSCO values and two *de novo* transcriptomes were excluded. Results showed that the Cirripedia clade had much more expanded (119) than contracted (4) gene families (Figure 3C). The *L. anserifera* genome had 111 expanded and 23 contracted gene families. The *A. amphitrite* had 77 expanded and 47 contracted gene families (Figure 3C).

The expanded gene families unique to each of the two barnacles differ in their enriched functional categories (Table S8). In the stalked barnacle, they mainly include neurogenesis, such as cerebellar granule cell differentiation, synapse assembly and neuromuscular process; whereas in the acorn barnacle the primary functions are protein metabolism, such as proteolysis and protein deubiquitination. These differences likely reflect adaptations of the two barnacles to their way of life: the enhanced proteolytic capability in *A. amphitrite* may be advantageous for cementation and attachment under strong waves in the intertidal and shallow subtidal areas (Dickinson et al., 2009); whereas in *L. anserifera*, which are attached to floating objects, enhanced neurogenesis may help them better sense and respond to the dynamic environmental conditions in the water column.

To understand the evolution of Cirripedia, we examined the 119 gene families that are expanded in both the acorn and stalk barnacle genomes. Among them are gene families associated with nervous system development, settlement and biofouling – key traits required for their successful sessile lifestyle (Table S9). One of the most expanded families is broad-complex core protein isoforms 1/2/3/4/5-like (Br–C), which is a transcription factor playing a critical role in ecdysone and juvenile hormone (JH) responsive processes such as molting and metamorphosis in insects (Bayer et al., 1996; Reza et al., 2004). In holometabolans, Br–C is involved in JH signaling, and knockdown of Br–C resulted in failure of metamorphosis from larva to adult (Konopova and Jindra, 2008). Little information of Br–C function is available in crustaceans, except that its regulation of Br–C resulted in the action of methyl farnesoate that functionally resembles JH in crustaceans (Cheong et al., 2015), and it was involved in larval metamorphosis (Sin et al., 2015). A previous phylogenetic study of arthropods discovered that Br–C was specifically gained in the ancestral pancrustaceans and it possibly coevolved with the JH pathway for regulation of metamorphosis (Qu et al., 2015). A total of 80 and 63 Br–C genes were found in *L. anserifera* and *A. amphitrite*, respectively. In contrast, only zero to four copies of Br–C were found in

each of the other pancrustaceans. Phylogenetic analysis revealed only one origin of Br–C in pancrustaceans (Figure 4A), and in the Cirripedia lineage it has undergone massive duplication, with remarkable tandem duplications of Br–C observed on contig3. There are 19 copies of Br–C in contig3 within a 191-kb length in the *L. anserifera* genome (Figure 4A). To further examine the massive expansion of Br–C in cirripeds, we preformed the ortholog search in two barnacle transcriptomes and identified more than 20 paralogs of Br–C in each species (Figure S4), suggesting its expansion in all cirripeds. During the life cycle of barnacles, there are two major metamorphosis events – one from nauplius to cypris and the other cypris to juvenile. Larval settlement involves the attachment of the cypris antennules onto the substratum by a cement, followed by ecdysis of the cyprid exoskeleton and metamorphosis (Høeg et al., 2012; Wong et al., 2018). Consistent with its key function in metamorphosis, the expansion of Br–C genes may facilitate the metamorphosis and settlement, enabling the drastic change from the larval to juvenile body plan in barnacles. The roles of Br–C genes should be investigated in the future by comparative gene expression analyses at different developmental stages.

In addition, cypris are highly mobile and equipped with organs for environmental sensing. We found an expansion of the longitudinals lacking protein-like (*Lola*) gene families in the *L. anserifera* genome (39 copies) and the *A. Amphitrite* genome (29 copies), compared to only two to seven copies in each of the other five compared pancrustacean genomes. Phylogenetic analysis of these Lola genes revealed several duplication events in the Cirripedia clade (Figure 4B). Lola was first identified in *Drosophila* as a broad-complex, tramtrack and bric-à-brac - zinc finger (BTB-ZF) family transcription factor regulating nervous system development. In *Drosophila* embryos, Lola regulated the axon tract development in the CNS, and axon guidance and growth in the peripheral nervous system (Giniger et al., 1994; Crowner et al., 2002). Although little is known about the function of this protein in crustaceans, previous phylogenetic analysis revealed that Lola is a new arthropod specific BTB protein that is only present in pancrustaceans (Quijano et al., 2016). A previous study of *A. amphitrite* cypris revealed a complex connection between peripheral sense organs and the CNS by a large number (2,000) of neurons, and the antennular nerve is the major peripheral nervous system (Harrison and Sandeman, 1999). Among the four antennular segments, the third segment contains an attachment disc with dense villi, and the fourth segment contains setae for detection of the physical and chemical properties of the substrate surface (Chan et al., 2018). Given the common origin of insect and crustacean Lola and the conserved nervous system features between insects and crustaceans (Osorio et al., 1997), the expansion of Lola genes and their expression in barnacles during different life stages are consistent with their purported roles in response to settlement cues and environmental signaling.

Unlike other pancrustaceans which grow by replacing the whole chitinous exoskeleton through periodic molting, cirripedes retain the outer and armored shell and grow by shedding only the exoskeleton between the soft tissues and the attachment base (Burden et al., 2014). The pupal cuticle protein 20-like (PCP20) homolog may have contributed to the unique growth pattern in cirripedes (Skinner et al., 1992). PCPs are proteins known to play an important role in the molting of various insects (Chihara et al., 1982; Nakato et al., 1990). For instance, in the silkworm *Bombyx mori*, it is synthesized in the epidemics during the metamorphosis from larvae to pupae, and then secreted into the pupal cuticle, with a peak of abundance before molting (Nakato et al., 1990). Under the search criteria of sequence identity >50% and E-value of 1e-10, we found 3 and 20 copies of PCP20 in *L. anserifera* and *A. amphitrite*, respectively, but no homolog of this gene in other pancrustaceans. Using more relaxed criteria (i.e., query coverage >30%, identity >30%, E-value of 1e-5), we found 90 PCP20 genes in *L. anserifera*, 231 in *A. amphitrite*, and 16 to 128

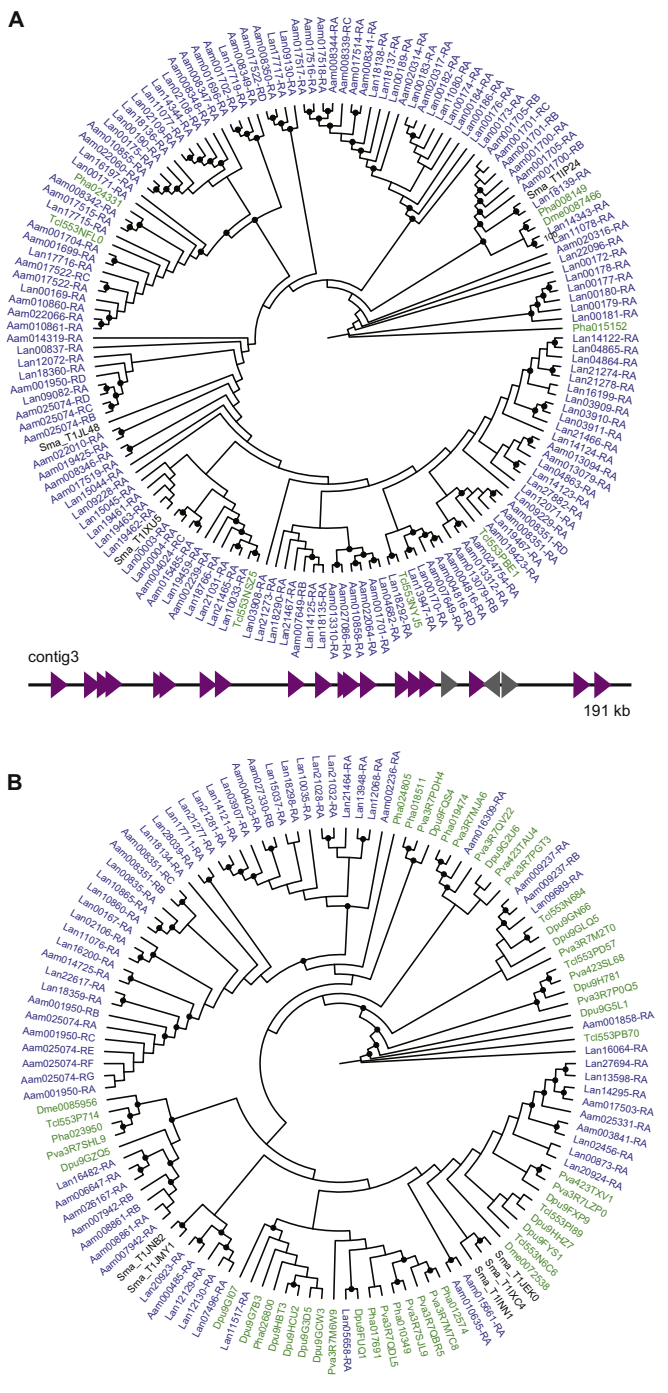

Figure 4. Maximum likelihood trees of expanded (A) broad-complex core protein isoforms 1/2/3/4/5-like (Br–C) and (B) longitudinals lacking protein-like (Lola). (A) Br–C with 80 in *Lepas anserifera* and 63 in *Amphibalanus amphitrite* compared to 0–4 copies in other pancrustaceans (substitution model: Blosum62+G4), in which 19 copies (purple color) located in *L. anserifera* scaffold3 with three non-Br-C genes (gray color). (B) Lola with 29 in *L. anserifera* and 30 in *A. amphitrite* compared to 2–17 copies in other pancrustaceans (substitution model: LG+G4). Expression data of *L. anserifera* is shown in Figure S5. The first letter of the genus name and the first two letters of the species name are used in abbreviations, e.g., Lan for *Lepas anserifera*. Nodes with ≥70 bootstrap support are labelled with a black dot. Gene ID labeled of barnacles are in blue, other pancrustaceans in green and outgroup in black color.

genes in each of the other pancrustaceans. Phylogenetic analysis revealed lineage-specific expansion of PCP20 in Cirripedia (Figure S6). The duplication of PCP20 and their active transcription (Figure S5) are consistent with the roles of these proteins as novel molting proteins uniquely found in barnacles.

3.7. Tissue-specific expression profiles and functions

To gain insight into tissue-specific functions, we analyzed the 16,356 genes expressed (TPM >1) in the peduncle (stalk), 17,861 genes in the cirri and 16,922 genes in the somatic body part (body). Among them, 13,804 genes were shared among the three tissues, which included many house-keeping genes such as elongation factor 1, ribosomal proteins and actin (Table S10). There were 879, 364 and 714 tissue-specific genes (TSGs) in the peduncle, cirri and body, respectively (Figure 5A and 5B). As our RNA-Seq analysis doses not have biological replication, the gene expression pattern of selected DEGs was further confirmed with real-time PCR (Figure 5C; Table S11). GO enrichment analyses were performed on the TSGs to gain insight into their biological functions (Figure 5D; Table S12). Scatterplots were also used to show tissue enriched biological processes. The flexible peduncle is armed with powerful longitudinal muscles, cement glands and covered by an unarmored cuticle (Power et al., 2010). Attachment of the peduncle to the substrate is maintained through continuous secretion from the cement gland (Anderson, 1994; Power et al., 2010). The enrichment of muscle development and biosynthetic related GO terms (e.g., muscle structure development and regulation of cellular macromolecule biosynthetic process) is consistent with the anatomy of the peduncle with strong muscle structure, and its role in secretion of adhesives for attachment to the substrate. The cirri were substantially enriched in cuticle development and metabolic pathways, which are related to its chitin-based exoskeleton and its role in filter-feeding. The somatic body parts which contained multiple organs, including the digestive, reproductive and nervous systems were enriched in an array of biological functions, such as innate immunity, metabolic process, response to stress and reproduction.

3.8. Adaptation to sessile life by settlement inducing protein complex and cement proteins

Apart from gene family expansion, formation of new genes also provides genetic novelty. Previous studies have reported two settlement related proteins in barnacles, i.e., the settlement inducing protein complex (SIPC) and cement protein (CP). SIPC, a large glycoprotein acting as a pheromone, is critical for larval settlement and substance/barnacle species recognition (Dreanno et al., 2006a, 2007). Barnacle CP is used as a glue for settlement of the cyprids and permanent underwater attachment of the adults (Liang et al., 2019). Multiple adult CPs have been identified and named based on protein masses (i.e., CP19k, CP43k, CP52k and CP100k) (Power et al., 2010). Our phylogenetic analysis of pancrustacean SIPCs revealed a specific barnacle SIPC clade, including one *L. anserifera* gene and six *A. amphitrite* genes; these barnacle SIPCs are clustered with reported barnacle SIPCs (Figure 6A; also see Yorisue et al., 2012). These SIPC genes exhibited tissue-specific expression profiles (Figure 6C). Apart from its role in larvae settlement, previous study of SIPC in *Balanus* barnacle found its transcription in different developmental stages and tissues, and showed that it functioned in cuticle secretion (Dreanno et al., 2006b). Therefore, the phylogeny and tissue expression of SIPCs indicated that this gene has duplicated many times during the evolution of barnacles, and different paralogs of this gene may have co-opted into different organs with functions that are related to settlement and cuticle formation.

Phylogenetic analysis showed that the CPs are clustered by their protein masses (Figure 6B), and there are long branch lengths (substitution) between these CP clades, which indicated episodes of historical duplication and diversification of this gene in Cirripedia. Among pancrustaceans, only *Daphnia* possesses one homolog of CP - a protein called Filaggrin involved in filament-aggregation, epidermal barrier formation against antimicrobial attack and damage by UV radiation (Harding et al., 2013). Our transcriptome analysis showed that the CPs were specifically expressed in the peduncle (Figure 6C) where the cement gland is located, indicating its function in permanent attachment. Thus, our results indicated that barnacle CPs may have been co-opted from a crustacean gene

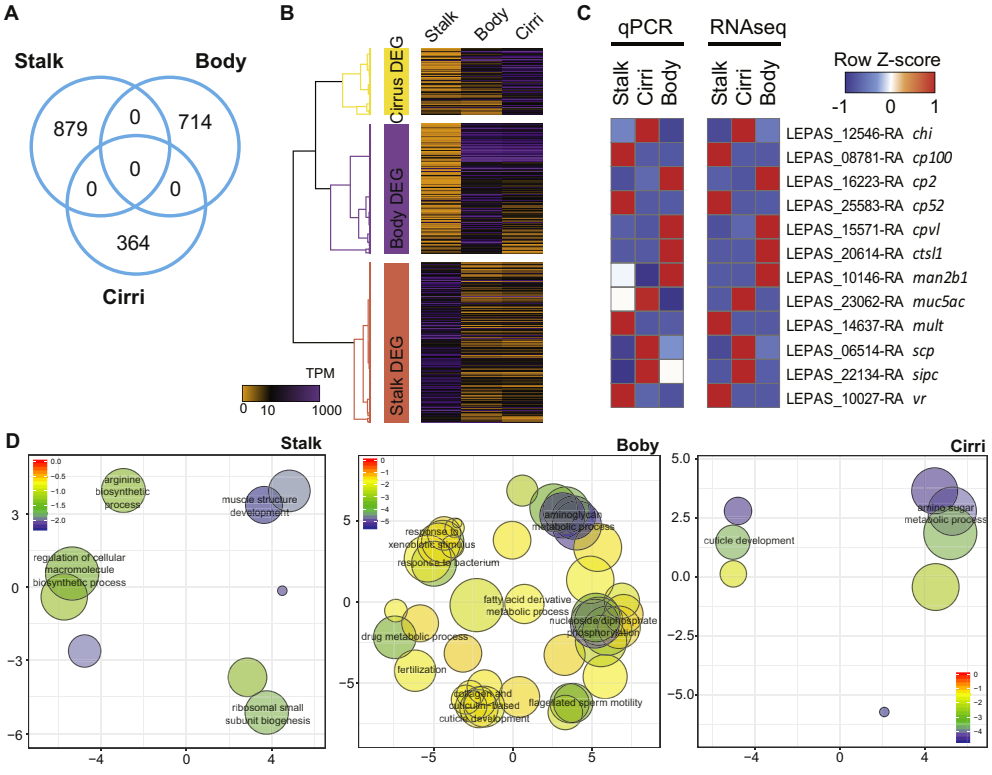

Figure 5. Tissue expression profile in *Lepas anserifera*. (A) Venn diagram of tissue-specific genes (TSGs). (B) Heatmaps showing expression of all TSGs with the hierarchical clustering. (C) Real-time PCR analysis. Relative fold changes of selected DEGs in stalk, cirri and other body part of *L. anserifera* were examined (data in Table S11). The heatmap shows the Z-scores (n = 3 for qPCR; n = 1 for RNA-Seq, in each tissue). (D) Biological processes (GO enrichment analysis, Table S12) enriched in TSGs. In the scatterplots, each bubble indicates a significantly enriched term (FDR <0.05) in a two-dimensional space derived by applying multidimensional scaling to a matrix of the GO terms' semantic similarities (Supek et al., 2011). Bubble color represents the Log10 FDR value. The circle size indicates the frequency of the GO term in the UniPort database (more general terms are represented by larger size bubbles).

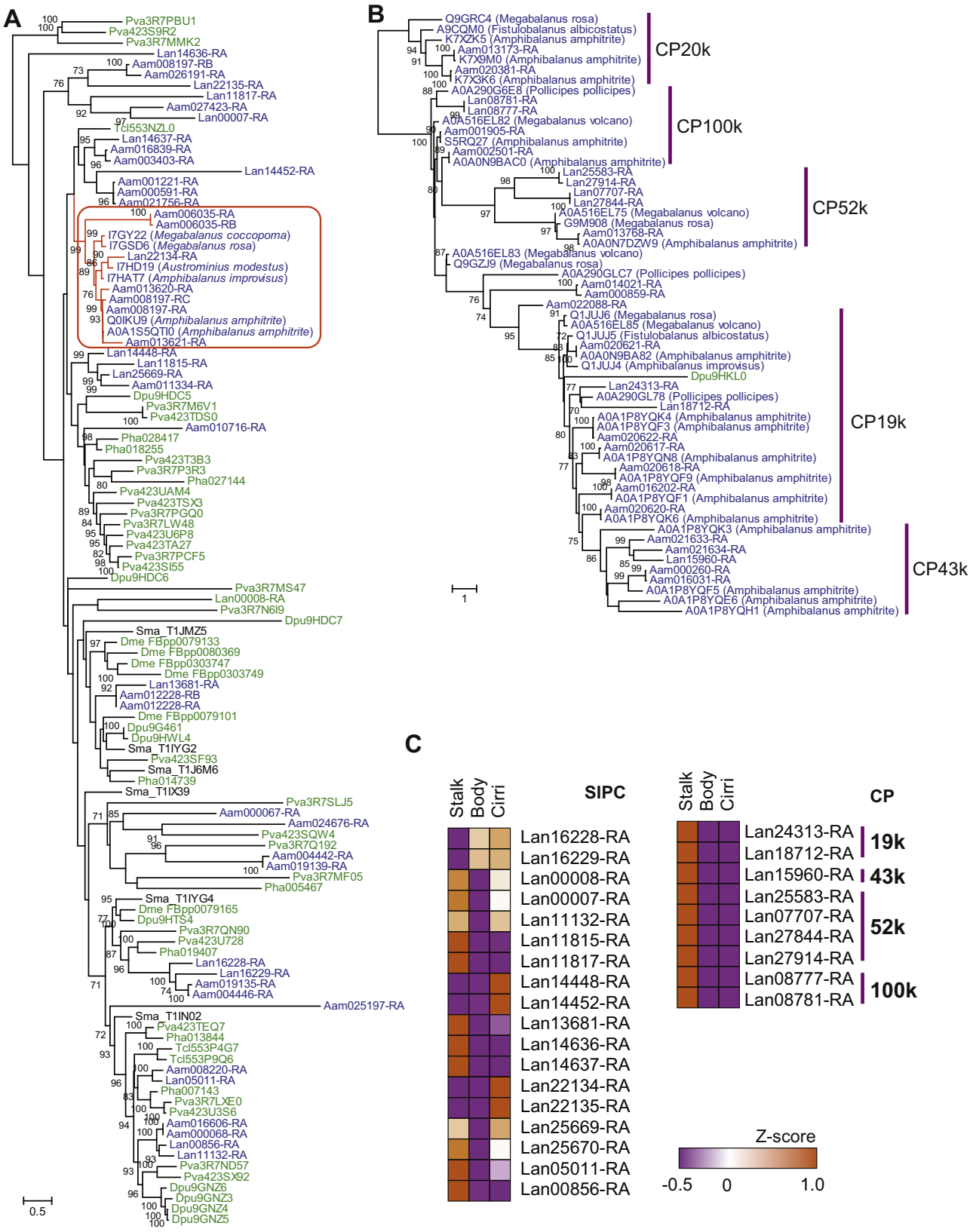

Figure 6. Maximum likelihood trees of (A) settlement inducing protein complex (SIPC) and (B) cement protein (CP) in pancrustacean genomes and barnacle sequences downloaded from UniProt. (A) SIPC with 18 in *Lepas anserifera* and 29 in *Amphibalanus amphitrite* compared to 3–29 copies in other pancrustaceans (substitution model: WAG+F+G4). Barnacle-specific clade highlighted in orange color. (B) CP with 9 in *L. anserifera* and 18 in *A. amphitrite*, but only one homolog in *Daphnia* (substitution model: WAG+F+G4). (C) Tissue expression data of SIPC and CP in *L. anserifera*. The first letter of the genus name and the first two letters of the species name are used in abbreviations, e.g., Lan for *Lepas anserifera*. Nodes with ≥70 bootstrap support are labelled with a black dot. Gene ID labeled by barnacles in blue, other pancrustaceans in green and outgroup in black color.

with multiple functions including cuticle secretion and immunity. This hypothesis can be tested using phylogenetic methods when more genomes from barnacles and their relatives become available.

4. Conclusions

In summary, we have provided a comprehensive genomic resource and established a phylogenetic framework for understanding the molecular adaptations of cirripedes to the sessile lifestyle. Gene family analysis revealed genetic mechanisms for environmental sensing, CNS development and cuticle formation. The formation of SIPCs and CPs as new genes and their duplications may have enabled barnacle underwater attachment and survival in habitats with strong wave action. Our results will be valuable for the development of antifouling strategy against barnacle settlement, and underwater adhesives that mimic the functions of barnacle cement proteins. Moreover, given the general lack of genomic resources in pancrustaceans, the *L. anserifera* genome will facilitate comparative studies aiming to understand the evolution of this large group of arthropods.

Declarations

Author contribution statement

Jack Chi-Ho Ip: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Jian-Wen Qiu & Benny K. K. Chan: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Jian-Wen Qiu was supported by the National Key R&D Program, Ministry of Science and Technology, China (2018YFC0310005), Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0409), Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (L20190005), General Research Fund from the University Grants Committee of Hong Kong Special Administrative Region (12302917), and State Key Laboratory of Environmental and Biological Analysis, Hong Kong Baptist University (SKLP_1920_P04).

Benny K. K. Chan is supported by a grant in Senior Investigator Award, Academia Sinica, Taiwan (AS-IA-105-L03).

Data availability statement

The genomic DNA and RNA sequences have been deposited in NCBI Sequence Read Archive under Bioproject No. PRJNA678024. The genome assemblies and annotations are available from the Figshare DOI:10.6084/m9.figshare.12831926.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e07291.

Acknowledgements

We are grateful to thank the Department of Biology at the Hong Kong Baptist University for its support and usage of the facilities. We would also like to thank Dr. Pei-Chen Tsai for sampling assistance.

References

Anderson, D.T., 1994. Barnacles: Structure, Function, Development and Evolution. Springer Science & Business Media.

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6, 11.

Bayer, C.A., Holley, B., Fristrom, J.W., 1996. A switch in broad-complex zinc-finger isoform expression is regulated post transcriptionally during the metamorphosis of *Drosophila* imaginal discs. Dev. Biol. 177, 1–14.

Blin, M., Rabet, N., Deutsch, J.S., Mouchel-Vielh, E., 2003. Possible implication of Hox genes Abdominal-B and abdominal-A in the specification of genital and abdominal segments in cirripedes. Dev. Gene. Evol. 213, 90–96.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Boetzer, M., Pirovano, W., 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinf. 15, 211.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

Buckeridge, J.S., 2012. Opportunism and the resilience of barnacles (Cirripedia: Thoracica) to environmental change. Integr. Zool. 7, 137–146.

Buckeridge, J.S., Chan, B.K.K., Lee, S.W., 2018. Accumulations of fossils of the whale barnacle *Coronula bifida* Bronn, 1831 (Thoracica: Coronulidae) provides evidence of a late Pliocene cetacean migration route through the Straits of Taiwan. Zool. Stud. 57, e54.

Buckeridge, J.S., Chan, B.K.K., Lin, J.P., 2019. Paleontological studies of whale barnacles in Taiwan reveal new cetacean migration routes in the Western Pacific since the Miocene. Zool. Stud. 58, e39.

Burden, D.K., Spillmann, C.M., Everett, R.K., Barlow, D.E., Orihuela, B., Deschamps, J.R., Fears, K.P., Rittschof, D., Wahl, K.J., 2014. Growth and development of the barnacle *Amphibalanus amphitrite*: time and spatially resolved structure and chemistry of the base plate. Biofouling 30, 799–812.

Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18, 188–196.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Chan, B.K.K., Dreyer, N., Gale, A.S., Glenner, H., Ewers-Saucedo, C., Perez-Losada, M., Kolbasov, G.A., Crandall, K.A., Hoeg, J.T., 2021. The evolutionary diversity of barnacles, with an updated classification of fossil and living forms. Zool. J. Linn. Soc. zlaa160.

Chan, K.M., Leung, K.M.Y., Cheung, K.C., Wong, M.H., Qiu, J.W., 2008a. Seasonal changes in imposex and tissue burden of butyltin compounds in *Thais clavigera* populations along the coastal area of Mirs Bay, China. Mar. Pollut. Bull. 57, 645–651.

Chan, B.K.K., Garm, A., Høeg, J.T., 2008b. Setal morphology and cirral setation of thoracican barnacle cirri: adaptations and implications for thoracican evolution. J. Zool. 275, 294–306.

Chan, B.K.K., Prabowo, R.E., Lee, K.S., 2009. Crustacean Fauna of Taiwan: Barnacles, Volume I - Cirripedia: Thoracica Excluding the Pyrgomatidae and Acastinae. National Taiwan Ocean University Press.

Chan, B.K.K., Hsieh, W.P., Kolbasov, G.A., 2014. Crustacean Fauna of Taiwan: Barnacles, Vol III- Cirripedia: Acrothoracica. Academia Sinica Press.

Chan, B.K.K., Høeg, J.T., 2015. Diversity of lifestyles, sexual systems, and larval development patterns in sessile crustaceans. In: Lifestyles and Feeding Biology, the Natural History of the Crustacea, 2. Oxford University Press, pp. 14–34.

Chan, B.K.K., Sari, A., Høeg, J.T., 2018. Cirripede cypris antennules: How much structural variation exists among balanomorphan species from hard-bottom habitats? Biol. Bull. 233 (2), 135–143.

Cheong, S.P., Huang, J., Bendena, W.G., Tobe, S.S., Hui, J.H., 2015. Evolution of ecdysis and metamorphosis in arthropods: the rise of regulation of juvenile hormone. Integr. Comp. Biol. 55, 878–890.

Chihara, C.J., Silvert, D.J., Fristrom, J.W., 1982. The cuticle proteins of *Drosophila melanogaster*: stage specificity. Dev. Biol. 89, 379–388.

Crowner, D., Madden, K., Goeke, S., Giniger, E., 2002. Lola regulates midline crossing of CNS axons in *Drosophila*. Development 129, 1317–1325.

de Boer, B., Lourens, L.J., Van De Wal, R.S., 2014. Persistent 400,000-year variability of Antarctic ice volume and the carbon cycle is revealed throughout the Plio-Pleistocene. Nat. Commun. 5, 2999.

De Gregoris, T.B., Borra, M., Biffali, E., Bekel, T., Burgess, J.G., Kirby, R.R., Clare, A.S., 2009. Construction of an adult barnacle (*Balanus amphitrite*) cDNA library and selection of reference genes for quantitative RT-PCR studies. BMC Mol. Biol. 10, 1–12.

Demuth, J.P., Hahn, M.W., 2009. The life and death of gene families. Bioessays 31, 29–39.

Deutsch, J.S., Mouchel-Vielh, E., 2003. Hox genes and the crustacean body plan. Bioessays 25, 878–887.

Dickinson, G.H., Vega, I.E., Wahl, K.J., Orihuela, B., Beyley, V., Rodriguez, E.N., Everett, R.K., Bonaventura, J., Rittschof, D., 2009. Barnacle cement: a polymerization model based on evolutionary concepts. J. Exp. Biol. 212, 3499–3510.

Dreanno, C., Matsumura, K., Dohmae, N., Takio, K., Hirota, H., Kirby, R.R., Clare, A.S., 2006a. An α2-macroglobulin-like protein is the cue to gregarious settlement of the barnacle *Balanus amphitrite*. Proc. Natl. Acad. Sci. USA 103, 14396–14401.

Dreanno, C., Kirby, R.R., Clare, A.S., 2006b. Locating the barnacle settlement pheromone: spatial and ontogenetic expression of the settlement-inducing protein complex of *Balanus amphitrite*. Proc. Royal Soc. B. 273, 2721–2728.

Dreanno, C., Kirby, R.R., Clare, A.S., 2007. Involvement of the barnacle settlement-inducing protein complex (SIPC) in species recognition at settlement. J. Exp. Mar. Biol. Ecol. 351, 276–282.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.
Gibson, G., 2000. Evolution: Hox genes and the cellared wine principle. Curr. Biol. 10, R452–R455.
Grenier, J.K., Garber, T.L., Warren, R., Whitington, P.M., Carroll, S., 1997. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr. Biol. 7, 547–553.
Giniger, E., Tietje, K., Jan, L.Y., Jan, Y.N., 1994. *Lola* encodes a putative transcription factor required for axon growth and guidance in *Drosophila*. Development 120, 1385–1398.
Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr., R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L., White, O., 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31, 5654–5666.
Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9, R7.
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.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.
Han, M.V., Thomas, G.W., Lugo-Martinez, J., Hahn, M.W., 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. 30, 1987–1997.
Harding, C., Aho, S., Bosko, C., 2013. Filaggrin–revisited. Int. J. Cosmet. Sci. 35, 412–423.
Harrison, P.J., Sandeman, D.C., 1999. Morphology of the nervous system of the barnacle cypris larva (*Balanus amphitrite* Darwin) revealed by light and electron microscopy. Biol. Bull. 197, 144–158.
Hedges, S.B., Dudley, J., Kumar, S., 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971–2972.
Høeg, J.T., Maruzzo, D., Okano, K., Glenner, H., Chan, B.K.K., 2012. Metamorphosis in balanomorphan, pedunculated, and parasitic barnacles: a video-based analysis. Integr. Comp. Biol. 52, 337–347.
Holm, E.R., 2012. Barnacles and Biofouling. Oxford University Press.
Hughes, C.L., Kaufman, T.C., 2002. Hox genes and the evolution of the arthropod body plan. Evol. Dev. 4, 459–499.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.
Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122.
Isaac, M.J., Moyse, J., 1990. Crustacea I: Entomostraca. In: Hayward, P.J., Ryland, J.S. (Eds.), The marine fauna of the British Isles and north-west Europe. Clarendon Press, pp. 322–361.
Ip, J.C.H., Leung, P.T., Ho, K.K., Qiu, J.W., Leung, K.M., 2016. *De novo* transcriptome assembly of the marine gastropod *Reishia clavigera* for supporting toxic mechanism studies. Aquat. Toxicol. 178, 39–48.
Ip, J.C.-H., Xu, T., Sun, J., Li, R., Chen, C., Lan, Y., Han, Z., Zhang, H., Wei, J., Wang, H., et al., 2021. Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. Mol. Biol. Evol. 38, 502–518.
Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.
Jung, J., Yoshida, R., Kim, W., 2019. Diversity of parasitic peltogastrid barnacles (Crustacea: Cirripedia: Rhizocephala) on hermit crabs in Korea. Zool. Stud. 58, 33.
Kajitani, R., Yoshimura, D., Okuno, M., Minakuchi, Y., Kagoshima, H., Fujiyama, A., Kubokawa, K., Kohara, Y., Toyoda, A., Itoh, T., 2019. Platanus-allee is a *de novo* haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat. Commun. 10, 1–15.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587.
Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115, 49–63.
Kim, J.H., Kim, H., Kim, H., Chan, B.K., Kang, S., Kim, W., 2019. Draft genome assembly of a fouling barnacle, *Amphibalanus amphitrite* (Darwin, 1854): the first reference genome for Thecostraca. Front. Ecol. Evol. 7, 465.
Kim, H.K., Chan, B.K., Kang, C.B., Kim, H.W., Kim, W., 2020. How do whale barnacles live on their hosts? Functional morphology and mating-group sizes of *Coronula diadema* (Linnaeus, 1767) and *Conchoderma auritum* (Linnaeus, 1767) (Cirripedia: Thoracicalcarea). J. Crustac Biol. ruaa075.
Konopova, B., Jindra, M., 2008. Broad-Complex acts downstream of Met in juvenile hormone signaling to coordinate primitive holometabolan metamorphosis. Development 135, 559–568.
Korf, I., 2004. Gene finding in novel genomes. BMC Bioinf. 5, 59.
Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496.
Liang, C., Strickland, J., Ye, Z., Wu, W., Hu, B., Rittschof, D., 2019. Biochemistry of barnacle adhesion: an updated review. Front. Mar. Sci. 6, 565.
Linse, K., Jackson, J.A., Fitzcharles, E., Sands, C.J., Buckeridge, J.S., 2013. Phylogenetic position of Antarctic scalpelliformes (Crustacea: Cirripedia: Thoracica). Deep Sea Res. Oceanogr. Res. Pap. 73, 99–116.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods 25, 402–408.
Lozano-Fernandez, J., Giacomelli, M., Fleming, J.F., Chen, A., Vinther, J., Thomsen, P.F., Glenner, H., Palero, F., Legg, D.A., Iliffe, T.M., 2019. Pancrustacean evolution illuminated by taxon-rich genomic-scale data sets with an expanded remipede sampling. Genome Biol. Evol. 11, 2055–2070.
Ludt, W.B., Rocha, L.A., 2015. Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. J. Biogeogr. 42, 25–38.
Mouchel-Vielh, E., Rigolot, C., Gibert, J.M., Deutsch, J.S., 1998. Molecules and the body plan: the Hox genes of cirripedes (Crustacea). Mol. Phylogenet. Evol. 9, 382–389.
Mouchel-Vielh, E., Blin, M., Rigolot, C., Deutsch, J.S., 2002. Expression of a homologue of the fushi tarazu (*ftz*) gene in a cirripede crustacean. Evol. Dev. 4, 76–85.
Nakato, H., Toriyama, M., Izumi, S., Tomino, S., 1990. Structure and expression of mRNA for a pupal cuticle protein of the silkworm, *Bombyx mori*. Insect Biochem. 20, 667–678.
Nielsen, S.K.D., Hoeg, J.T., Yusa, Y., 2016. Host relation, size and reproduction in the burrowing barnacle *Trypetesa lampas* (Hancock) (Crustacea Cirripedia Acrothoracica). Zool. Stud. 55, 14.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Nunez, J.C., Flight, P.A., Neil, K.B., Rong, S., Eriksson, L.A., Ferranti, D.A., Rosenblad, M.A., Blomberg, A., Rand, D.M., 2020. Footprints of natural selection at the mannose-6-phosphate isomerase locus in barnacles. Proc. Natl. Acad. Sci. U.S.A. 117, 5376–5385.
Osorio, D., Bacon, J.P., Whitington, P.M., 1997. The evolution of arthropod nervous systems: modern insects and crustaceans diverged from an ancestor over 500 million years ago, but their neural circuitry retains many common features. Am. Sci. 85, 244–253.
Papillon, D., Telford, M.J., 2007. Evolution of *Hox3* and *ftz* in arthropods: insights from the crustacean *Daphnia pulex*. Dev. Gene. Evol. 217, 315–322.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417.
Pérez-Losada, M., Høeg, J.T., Crandall, K.A., 2004. Unraveling the evolutionary radiation of the thoracican barnacles using molecular and morphological evidence: a comparison of several divergence time estimation approaches. Syst. Biol. 53, 278–298.
Porebski, S., Bailey, L.G., Baum, B.R., 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rep. 15, 8–15.
Power, A.M., Klepal, W., Zheden, V., Jonker, J., McEvilly, P., von Byern, J., 2010. Mechanisms of adhesion in adult barnacles. In: Biological Adhesive Systems. Springer, pp. 153–168.
Pryszcz, L.P., Gabaldón, T., 2016. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 44 e113-e113.
Qiu, J.W., Gosselin, L.A., Qian, P.Y., 1997. Effects of short-term variation in food availability on larval development in the barnacle *Balanus amphitrite amphitrite*. Mar. Ecol. Prog. Ser. 161, 83–91.
Qiu, J.W., Qian, P.Y., 1999. Tolerance of the barnacle *Balanus amphitrite amphitrite* to salinity and temperature stress: effects of previous experience. Mar. Ecol. Prog. Ser. 188, 123–132.
Qiu, J.W., Thiyagarajan, V., Cheung, S., Qian, P.Y., 2005. Toxic effects of copper on larval development of the barnacle *Balanus amphitrite*. Mar. Pollut. Bull. 51, 688–693.
Qiu, J.W., Hung, O.S., Qian, P.Y., 2008. An improved barnacle attachment inhibition assay. Biofouling 24, 259–266.
Qiu, J.W., Chan, K.M., Leung, K.M.Y., 2011. Seasonal variations of imposex indices and butyltin concentrations in the rock shell *Thais clavigera* collected from Hong Kong waters. Mar. Pollut. Bull. 63 (5-12), 482–488.
Quijano, J.C., Wisotzkey, R.G., Tran, N.L., Huang, Y., Stinchfield, M.J., Haerry, T.E., Shimmi, O., Newfeld, S.J., 2016. *Lolal* is an evolutionarily new epigenetic regulator of *dpp* transcription during dorsal–ventral axis formation. Mol. Biol. Evol. 33, 2621–2632.
Qu, Z., Kenny, N.J., Lam, H.M., Chan, T.F., Chu, K.H., Bendena, W.G., Tobe, S.S., Hui, J.H.L., 2015. How did arthropod sesquiterpenoids and ecdysteroids arise? Comparison of hormonal pathway genes in noninsect arthropod genomes. Genome Biol. Evol. 7, 1951–1959.
Rabet, N., Gibert, J.M., Quéinnec, É., Deutsch, J.S., Mouchel-Vielh, E., 2001. The caudal gene of the barnacle *Sacculina carcini* is not expressed in its vestigial abdomen. Dev. Gene. Evol. 211, 172–178.
Reza, A., Kanamori, Y., Shinoda, T., Shimura, S., Mita, K., Nakahara, Y., Kiuchi, M., Kamimura, M., 2004. Hormonal control of a metamorphosis-specific transcriptional factor Broad-Complex in silkworm. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 139, 753–761.
Ruan, J., Li, H., 2020. Fast and accurate long-read assembly with wtdbg2. Nat. Methods 17, 155–158.
Schiffer, P.H., Herbig, H.G., 2016. Endorsing Darwin: global biogeography of the epipelagic goose barnacles *Lepas* spp. (Cirripedia, Lepadomorpha) proves cryptic speciation. Zool. J. Linn. Soc. 177, 507–525.
Serano, J.M., Martin, A., Liubicich, D.M., Jarvis, E., Bruce, H.S., La, K., Browne, W.E., Grimwood, J., Patel, N.H., 2016. Comprehensive analysis of Hox gene expression in the amphipod crustacean *Parhyale hawaiensis*. Dev. Biol. 409, 297–309.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sin, Y.W., Kenny, N.J., Qu, Z., Chan, K.W., Chan, K.W., Cheong, S.P., Leung, R.W., Chan, T.F., Bendena, W.G., Chu, K.H., 2015. Identification of putative ecdysteroid and juvenile hormone pathway genes in the shrimp *Neocaridina denticulata*. Gen. Comp. Endocrinol. 214, 167–176.

Skinner, D.M., Kumari, S.S., O'brien, J.J., 1992. Proteins of the crustacean exoskeleton. Am. Zool. 32, 470–484.

Smit, A.F., Hubley, R.. 2008-2015. RepeatModeler open-1.0. Available from. http://www.repeatmasker.org.

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467.

Sun, J., Mu, H., Ip, J.C.H., Li, R., Xu, T., Accorsi, A., Sánchez Alvarado, A., Ross, E., Lan, Y., Sun, Y., et al., 2019. Signatures of divergence, invasiveness, and Terrestrialization revealed by four apple snail genomes. Mol. Biol. Evol. 36, 1507–1520.

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PloS One 6, e21800.

Swofford, D.L., 2001. Paup*: Phylogenetic Analysis Using Parsimony (And Other Methods) 4.0. B5.

Tsang, L.M., Wu, T.H., Shih, H.T., Williams, G.A., Chu, K.H., Chan, B.K.K., 2012. Genetic and morphological differentiation of the Indo-West Pacific intertidal barnacle Chthamalus malayensis. Integr. Comp. Biol. 52 (3), 388–409.

von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'Ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T.M., Stamatakis, A., 2012. Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. Mol. Biol. Evol. 29, 1031–1045.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204.

Wang, J.R., Holt, J., McMillan, L., Jones, C.D., 2018. FMLRC: hybrid long read error correction using an FM-index. BMC Bioinf. 19, 50.

Wong, Y.H., Ozaki, N., Zhang, W.P., Sun, J., Yoshimura, E., Oguro-Okano, M., Nogata, Y., Lin, H.C., Chan, B.K., Qian, P.Y., 2018. Identification of barnacle shell proteins by transcriptome and proteomic approaches. In: Biomineralization. Springer, pp. 105–112.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yi, A.X., Leung, K.M., Lam, M.H., Lee, J.S., Giesy, J.P., 2012. Review of measured concentrations of triphenyltin compounds in marine ecosystems and meta-analysis of their risks to humans and the environment. Chemosphere 89, 1015–1025.

Yorisue, T., Matsumura, K., Hirota, H., Dohmae, N., Kojima, S., 2012. Possible molecular mechanisms of species recognition by barnacle larvae inferred from multi-specific sequencing analysis of proteinaceous settlement-inducing pheromone. Biofouling 28, 605–611.

Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2012. Goseq: gene ontology testing for RNA-seq datasets. R Bioconductor.

Zhong, C., Gurry, T., Cheng, A.A., Downey, J., Deng, Z., Stultz, C.M., Lu, T.K., 2014. Strong underwater adhesives made by self-assembling multi-protein nanofibres. Nat. Nanotechnol. 9, 858.

Zhu, B.H., Xiao, J., Xue, W., Xu, G.C., Sun, M.Y., Li, J.T., 2018. P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. BMC Genom. 19, 175.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A., 2013. The MaSuRCA genome assembler. Bioinformatics 29, 2669–2677.