

Genome assemblies of two species of porcelain crab, *Petrolisthes cinctipes* and *Petrolisthes manimaculis* (Anomura: Porcellanidae)

Pascal Angst ^{1,*}, Eric Dexter ¹, Jonathon H. Stillman ^{1,2,3,*}

¹Department of Environmental Sciences, Zoology, University of Basel, 4051 Basel, Switzerland

²Department of Biology, San Francisco State University, San Francisco, CA 94132, USA

³Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA

*Corresponding author: Email: pascal.angst@unibas.ch; *Corresponding author: Email: stillmaj@sfsu.edu

Crabs are a large subtaxon of the Arthropoda, the most diverse and species-rich metazoan group. Several outstanding questions remain regarding crab diversification, including about the genomic capacitors of physiological and morphological adaptation, that cannot be answered with available genomic resources. Physiologically and ecologically diverse Anomuran porcelain crabs offer a valuable model for investigating these questions and hence genomic resources of these crabs would be particularly useful. Here, we present the first two genome assemblies of congeneric and sympatric Anomuran porcelain crabs, *Petrolisthes cinctipes* and *Petrolisthes manimaculis* from different microhabitats. Pacific Biosciences high-fidelity sequencing led to genome assemblies of 1.5 and 0.9 Gb, with N50s of 706.7 and 218.9 Kb, respectively. Their assembly length difference can largely be attributed to the different levels of interspersed repeats in their assemblies: The larger genome of *P. cinctipes* has more repeats (1.12 Gb) than the smaller genome of *P. manimaculis* (0.54 Gb). For obtaining high-quality annotations of 44,543 and 40,315 protein-coding genes in *P. cinctipes* and *P. manimaculis*, respectively, we used RNA-seq as part of a larger annotation pipeline. Contrarily to the large-scale differences in repeat content, divergence levels between the two species as estimated from orthologous protein-coding genes are moderate. These two high-quality genome assemblies allow future studies to examine the role of environmental regulation of gene expression in the two focal species to better understand physiological response to climate change, and provide the foundation for studies in fine-scale genome evolution and diversification of crabs.

Keywords: genomics; pacBio HiFi; genome assembly; Arthropoda; Crustacea; Decapoda; Malacostraca

Introduction

Arthropoda is the largest and most diverse metazoan phylum (Thomas et al. 2020). Yet questions of genome evolution and diversification are limited to a relatively small number of clades (i.e. Diptera, Hymenoptera) for which a wealth of genome sequence data are available (Thomas et al. 2020). The crustaceans are one of the most diverse arthropod groups in terms of variation in morphology, habitat, and lifestyle, but also one of the most poorly represented arthropod groups in terms of whole genome sequence data. Of the six classes of Crustacea, the largest group, Malacostraca, contains about 40K species including crabs, shrimps, lobsters, crayfish, krill, amphipods, and isopods. Of the 3,750 complete Arthropod genomes currently available in the NCBI database of sequenced genomes, only 51 (~1%) are for Malacostracans. If we consider only the Malacostracan order Decapoda, we observe tremendous species richness and diversity. The Decapoda contains over 14K extant species, of which the majority are the nearly 9K species of crabs (6.5K species of Brachyura and 2.4K species of Anomura, which includes the king, hermit, porcelain, and galatheid crabs) (De Grave et al. 2009). Despite the species richness and diversity of crabs, there are only 13 crab

genome records in the NCBI database: 10 Brachyura, 5 of which are for the commercially important swimming crabs (Portunidae: *Portunus* and *Callinectes*), 4 of which are for the commercially important Chinese mitten crab (Varunidae: *Eriocheir*), and 1 spider crab (Majidae: *Chionecetes*), and 3 Anomura (1 Coenobitidae: *Birgus*, 2 Lithodidae: *Paralithodes*). Clearly, if we are to understand the genomic basis of the ecological, physiological, and taxonomic diversification of such large and diverse group of arthropods, there is a need to develop better genomic resources for the Malacostraca, and for crabs in particular.

In this study, we produced functionally annotated long-read genome assemblies for two species of Anomuran porcelain crabs: *P. cinctipes* and *P. manimaculis*. Porcelain crabs, family Porcellanidae, are a species-rich group that inhabit shallow coastal ecosystems throughout the temperate and tropical regions of the Pacific Rim and warm regions of the Western Atlantic (Kropp and Haig 1994; Stillman and Reeb 2001; Raso et al. 2005; Rodriguez et al. 2005; Naderloo et al. 2013; Werdinger and Hiller 2015; Limviriyakul et al. 2016; Diez and Lira 2017; DE Azevedo Ferreira and Anker 2021; Mantelatto et al. 2021). The largest genus of Porcellanidae is *Petrolisthes* (Haig 1960). In the eastern Pacific, there are approximately 50 species of *Petrolisthes* split into four

Received on 10 October 2023; accepted on 09 November 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

principal regions, the north temperate, the northern Gulf of California, the tropics, and the south temperate (Haig 1960; Stillman and Reeb 2001). Within each biogeographic region, species are distributed across vertical distribution gradients such that some species live solely in the intertidal zone and some are subtidal (Stillman and Somero 2000). As intertidal zone species are exposed to terrestrial conditions during low tide, they experience a wider range of environmental variation than subtidal zone species (Stillman 2002; Gunderson et al. 2019). Intertidal zone species possess physiological and morphological adaptations that allow them to survive the challenges of life out of water, including thermal variation and respiratory challenges (Stillman and Somero 1996; Gaitan-Espitia et al. 2014).

A molecular phylogenetic analysis of the eastern Pacific *Petrolisthes* indicated that there are two main subgenera or clades, which can be identified by the presence or absence of serrate saw-teeth on the meral segment of the chelae (Stillman and Reeb 2001). The clade possessing the serrate teeth is comprised mainly of species that live in tropical subtidal habitats; of the ~25 species in that clade only two species have radiated to a different habitat: *Petrolisthes armatus*, inhabits tropical intertidal habitats, and *Petrolisthes desmarestii*, inhabits temperate subtidal habitats (Stillman and Reeb 2001). Thus, that clade has not had much adaptive radiation. Additionally, from phylogenetic analyses, the speciation events in the serrate teeth clade are well resolved (Stillman and Reeb 2001). In contrast, the other clade has an unresolvable polytomy at the base of the phylogenetic tree, with species that have radiated into every possible habitat (temperate, tropical, intertidal, subtidal) and evolved additional life-history innovations (e.g. specific commensalism). Only more recent speciation events within that clade are phylogenetically resolvable, and include additional radiation into different vertical zones within a biogeographic region (Stillman and Reeb 2001).

In one of those subclades there are two sympatric species, *P. cinctipes* and *P. manimaculis* (Fig. 1), that share a common ancestor approximately 8–14 mya (Stillman and Reeb 2001), and live in different vertical zones on shores of the northeastern Pacific (Miller et al. 2013; Delmanowski and Tsukimura 2015; Armstrong and Stillman 2016; Delmanowski et al. 2017; Gunderson et al. 2017). These two species differ in their heat tolerance (Stillman and Somero 2000; Stillman 2002; Miller et al. 2013), and their responses to stress at the organismal (Wasson et al. 2002; Gunderson et al. 2017) and transcriptomic (Armstrong and Stillman 2016) levels. The genomic bases for the physiological differences between *P. cinctipes* and *P. manimaculis* are unknown.

Previous comparative studies of mitochondrial genomes have indicated that genome arrangements have likely played a strong role in the evolution of crabs (Wang et al. 2021; Zhang et al. 2021;

Sun et al. 2022), and gene arrangement is known to be essential for emergent properties of gene products in development such as HOX genes (Sun and Patel 2019) and in cancer (Heng and Heng 2021, 2022). Evidence for the extent to which arrangement of nuclear genes is in general involved in adaptive evolution does not yet exist for crabs, but has been observed in other taxa including bacteria (Kang et al. 2022), fungal pathogens (Gourlie et al. 2022; Ma et al. 2022), and domesticated yeast (Garcia-Rios and Guillamon 2022), and may represent a generalized aspect of genome evolution during adaptive radiation (Cao et al. 2022; Wang et al. 2022). By providing two high-quality porcelain crab genome assemblies generated using long-read high-fidelity (PacBio HiFi) whole genome sequencing along with RNA-seq data for both species for genome annotation, we set the stage for exploring the extent to which differences in the physiology and ecology of *P. cinctipes* and *P. manimaculis* are reflected in their genomes.

Methods

Specimen collection, DNA extraction and sequencing

We produced de-novo genome assemblies *P. cinctipes* and *P. manimaculis* based upon Pacific Biosciences high-fidelity (PacBio HiFi) sequence data, which was then cross-validated and contaminant-filtered using independently generated cDNA library data, with additional 10x Genomics short-read data available from another study (J. Stillman, unpublished) for cross-validation in the case of *P. cinctipes*. PacBio sequencing was conducted using the HiFi method on gill tissues dissected from a single male crab specimen of each species collected near Fort Ross, California, USA (38.50421°N, 123.23152°W) on January 16, 2022 and frozen on liquid nitrogen and stored at –80 °C. Gill tissue was used because other tissue types did not provide suitable DNA for PacBio sequencing. Special attention was paid to bioinformatic filtering of non-target DNA, because gill likely had a high load of epi-microbiota. Frozen tissues were delivered to UC Davis Genome Center DNA Technologies Core for high molecular weight (HMW) DNA extraction that yielded a fragment size peak of 155 and 132 Kb for *P. manimaculis* and *P. cinctipes*, respectively. PacBio HiFi libraries were prepared from those samples and each library was sequenced across three SMRT cells on the PacBio Sequel II platform.

10x Genomics sequencing was conducted on claw muscle tissue dissected from a single specimen of *P. cinctipes*. The specimen was likely collected at the site near Fort Ross, CA, USA (as above), but collection date of the specimen and the sex of the specimen are unknown. The HMW DNA extraction from claw muscle tissue is less prone to contamination and yielded adequate DNA of

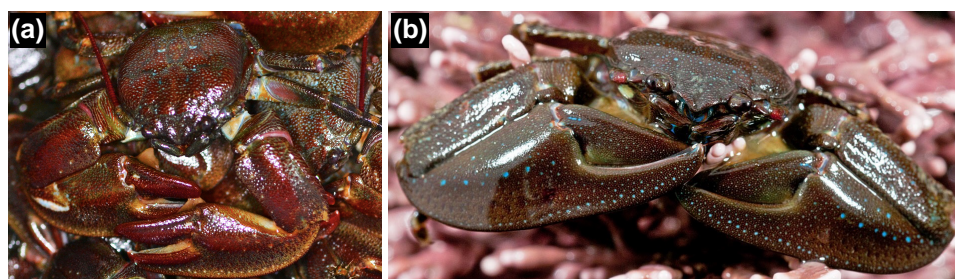


Fig. 1. *Petrolisthes cinctipes* (a) and *Petrolisthes manimaculis* (b). Identifying marks on *P. cinctipes* include red antennae, red spots on claws, and red mouthparts. Identifying marks on *P. manimaculis* include lines of blue spots on claws, blue mouthparts, and red spots on base of gray antennae. *P. cinctipes* photograph by Adam Paganini and *P. manimaculis* photograph by Steven Sharnoff; both used with permissions.

>40 Kb for 10× Genomics library construction. Libraries were size selected to 350–650 bp to maximize the quality of the paired end reads. Samples were sent to Novogene (Sacramento, CA, USA) for 150 bp PE sequencing on the Illumina HiSeqX10 platform.

RNA-seq data used in the analysis was obtained from Illumina 100 bp PE reads of cDNA libraries made from gill tissue of *P. cinctipes* and *P. manimaculis* as previously described (Armstrong and Stillman 2016). Additional transcriptomics data from ESTs of a cloned cDNA library of *P. cinctipes* are available (Tagmount et al. 2010), though not used in the present study.

Assembly

Our primary genome assemblies were created using hifiasm v.0.16.0-r369 (Cheng et al. 2021). As our tissue samples likely contained nontarget DNA from epibionts and associated microbiota, our primary assemblies were carefully filtered to remove nontarget contigs. Contig filtering was performed using BlobTools v.1.1.1 (Laetsch and Blaxter 2017), which combines information about GC

content, sequencing depth, and taxonomic classification to create a profile of each contig (Fig. 2). Based upon an iterative filtering process, we found that the following parameters removed all clearly nontarget contigs: GC content between 0.3 and 0.5 (five standard deviations from the mean of all unambiguously arthropod contigs) and sequencing depth between 0.33 and 3 times the average sequencing depth of all unambiguously arthropod contigs. We also removed circular contigs and contigs with strong sequence similarity to taxa outside the animal kingdom, except microsporidia, which tend to be wrongly annotated in reference databases likely because of tight host–parasite relations (intra-cellular parasitism). Sequencing depth for HiFi reads (required for BlobTools analysis) was calculated based on minimap2 map-hifi v.2.20-r1061 (Li 2018) mapping, while nucleotide alignments to reference databases (also required for BlobTools analysis) were done using blastn v.2.12.0+ (Camacho et al. 2009) and the NCBI database, as well as Diamond blastx -F 15 -b4 -c1 v.2.0.15.153 (Buchfink et al. 2021) and the UniProt database

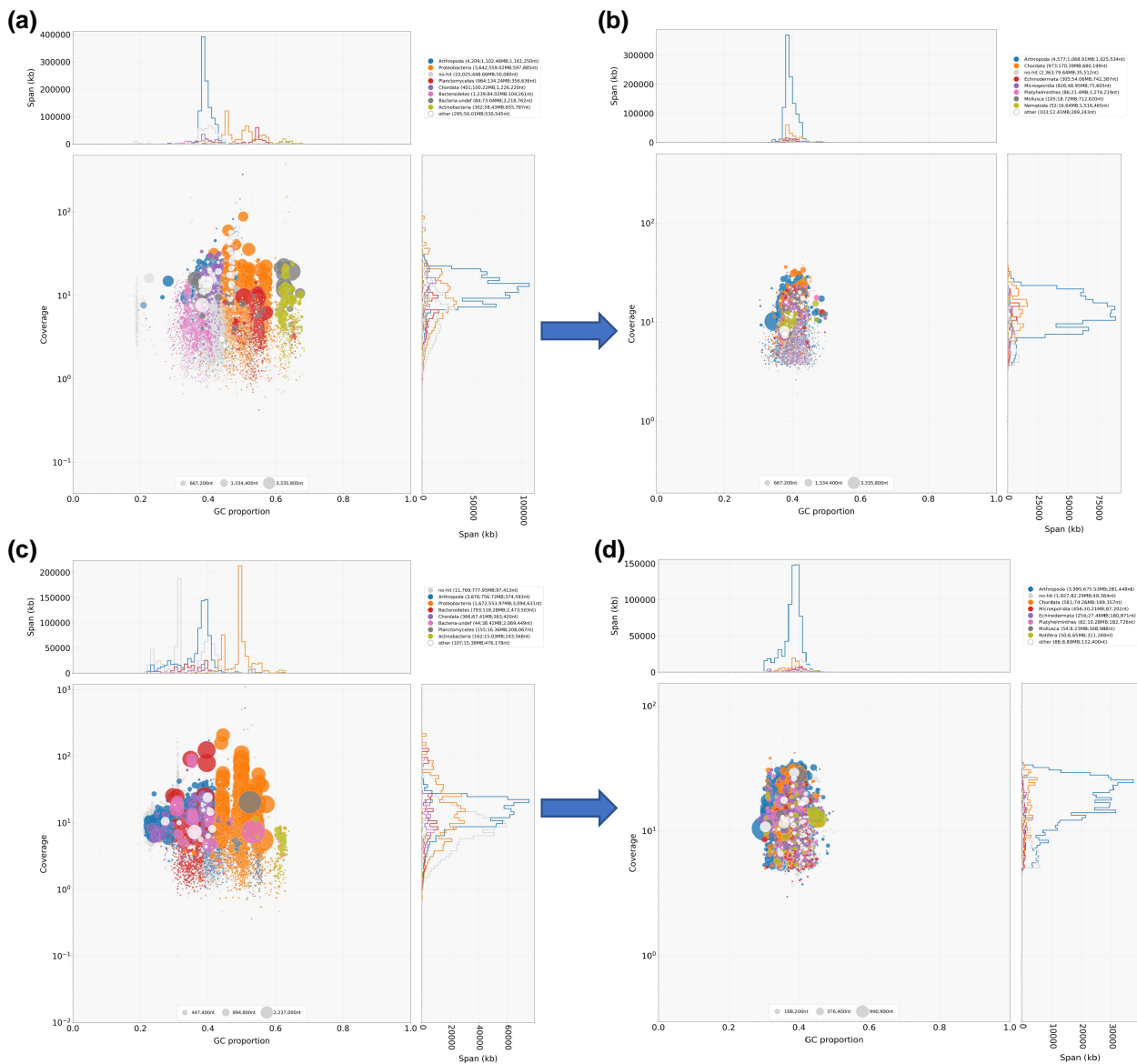


Fig. 2. Filtering of raw sequence data from (a, b) *P. cinctipes* and (c, d) *P. manimaculis* using BlobTools. For each species, the coverage and GC content are compared (main plot Y- and X-axes), and are also plotted against sequence span for (a, c) unfiltered sequence data and (b, d) filtered sequence data. For both species, the effectiveness of sequence filtering can be observed by the enrichment of data represented by the blue Arthropod color in plots of sequence span vs GC proportion (top plot) and in plots of coverage vs sequence span (right plot).

(The UniProt Consortium 2022). For *P. cinctipes*, we additionally filtered PacBio-based contigs based on independently generated 10× Genomics short-read data. We retained only contigs with mean 10× Genomics read depth between 1 and 286 (three times the mean whole-genome depth). 10× Genomics reads were mapped to the PacBio-based contigs using *bwa-mem2* v.2.2.1 (Vasimuddin et al. 2019). For *P. manimaculis*, we additionally removed a set of contigs with no taxonomic classification which were responsible for an odd peak in the GC distribution around 0.33, likely arising from an unknown epibiont that was not represented in our BLAST database (Fig. 2).

BlobTools analyses indicated that 20% of the primary assemblies' contigs had sequence similarity to Arthropoda, and the bulk of the sequences had either no strong similarity to any taxa ("No-hit") or were microbial in origin (Table 1, Fig. 2). Following filtration of the data with BlobTools, the overall contig number was reduced by approximately 60% for both species, and the fraction of contigs with sequence similarity to Arthropoda increased to ~50% for both species (Table 1). In fact, 75% or greater of the contigs in both species had similarity to either Arthropoda or had no similarity to any known taxa (Table 1, Fig. 2). Because further filtration of the data with BlobTools may have caused inadvertent discarding of *Petrolisthes* contigs, we kept all the contigs in the further analyses. Both final assemblies had similar proportions of taxonomic classification (Table 1).

All contig filtering was performed using *seqtk* v.1.3-r106 (<https://github.com/lh3/seqtk>), with before and after *kmer* distributions visualized using *jellyfish* v.2.2.10 (Marçais and Kingsford 2011) and *GenomeScope* v.2.0 (Ranallo-Benavidez et al. 2020). We used the default parameters for all bioinformatics tools if not mentioned otherwise.

Annotation

After removal of all identifiable nontarget contigs in the assemblies, we masked repeats with lower-case letters using *RepeatModeler*

Table 1. Assembly statistics before and after filtering the data using BlobTools.

| Prefiltering | | | Postfiltering | | |
|---------------------------------|--------------|-----|-----------------|--------------|-----|
| Taxon | # of contigs | % | Taxon | # of contigs | % |
| <i>Petrolisthes cinctipes</i> | | | | | |
| Arthropoda | 5,361 | 42 | Arthropoda | 4,577 | 72 |
| Proteobacteria | 3,824 | 22 | Chordata | 973 | 11 |
| Chordata | 1,175 | 7 | No-hit | 2,364 | 5 |
| No-hit | 5,097 | 6 | Echinodermata | 305 | 4 |
| Planctomycetes | 1,144 | 5 | Microsporidia | 826 | 3 |
| Bacteroidetes | 1,579 | 4 | Platyhelminthes | 86 | 1 |
| Bacteria-undef | 97 | 3 | Mollusca | 105 | 1 |
| Actinobacteria | 342 | 2 | Nematoda | 52 | 1 |
| Other | 2,542 | 9 | Other | 103 | <1 |
| Total | 21,161 | 100 | Total | 9,391 | 100 |
| <i>Petrolisthes manimaculis</i> | | | | | |
| Arthropoda | 5,214 | 35 | Arthropoda | 3,991 | 73 |
| Proteobacteria | 1,815 | 24 | No-hit | 1,926 | 9 |
| No-hit | 7,728 | 19 | Chordata | 581 | 8 |
| Bacteroidetes | 865 | 5 | Microsporidia | 456 | 3 |
| Chordata | 870 | 4 | Echinodermata | 254 | 3 |
| Microsporidia | 936 | 2 | Platyhelminthes | 82 | 1 |
| Echinodermata | 384 | 2 | Mollusca | 54 | 1 |
| Rotifera | 155 | 2 | Rotifera | 50 | 1 |
| Other | 977 | 6 | Other | 88 | <1 |
| Total | 18,944 | 100 | Total | 7,482 | 100 |

Percentages refer to the sequence content per taxon.

v.2.0.2, including the LTR pipeline (Flynn et al. 2020), and *RepeatMasker* v.4.1.2 (Smit et al. 2013). Masking the repeats using lower-case letters ensured that software for gene prediction was aware of them. We then performed quality and adapter trimming on the RNA-seq reads using *trim_galore* v.0.6.4_dev (<https://github.com/FelixKrueger/TrimGalore>) and *cutadapt* v.2.3 (Martin 2011) before mapping them to the respective species' genome using *HISAT2* v.2.2.1 (Kim et al. 2019). With the mapped reads, we trained *GeneMark-ES* v.4.62 (Brùna et al. 2020) and *AUGUSTUS* v.3.4.0 (Stanke et al. 2006, 2008) for gene prediction as implemented in *BRAKER* v.2.1.6 (Brùna et al. 2021). The resulting annotation files were converted to GFF3 files and to sequence files using *AGAT* v.1.0.0 (Dainat 2022). The resulting files were used for functional annotation with the combined evidence from *InterProScan* v.5.55_88.0 (Jones et al. 2014), *eggNOG-mapper* v.2.1.9 (Huerta-Cepas et al. 2019; Cantalapiedra et al. 2021), *Phobius* v.1.01 (Käll et al. 2004), and *SignalP* v.5.0b (Almagro Armenteros et al. 2019) as well as with comparisons to Pfam (Mistry et al. 2021), UniProt (The UniProt Consortium 2022), MEROPS (Rawlings et al. 2014), dbCAN (Yin et al. 2012) databases, and BUSCO (Manni et al. 2021) with the Arthropoda database (*arthropoda_odb10*) with *funannotate* 1.8.11 (Palmer and Stajich 2022). Annotation statistics were generated with *agat_sp_statistics.pl* v.1.0.0 (Dainat 2022).

Comparative genomics

We examined synteny between the two species in terms of sequence homology across large contigs and in terms of the order of orthologous gene pairs. For assessing sequence similarity on the contig level, we used *D-Genies* v.1.5.0 (Cabanettes and Klopp 2018). For gene level comparisons, we only used single copy orthologs inferred with *OrthoFinder* v.2.5.4 (Emms and Kelly 2019). Each information was used to identify 10 homologous contigs for in-depth sequence comparison (Supplementary Fig. 1). For visualization of the syntenic regions, we used *GENESPACE* v.1.1.7 (Lovell et al. 2022) and *gggenomes* v.0.9.5.9000 (Hackl et al. 2021) in R v.4.2.2 (R Core Team 2022). To estimate divergence between single copy orthologs, we aligned them using *prank* v.1.70427 (Löytynoja 2014) while using *seqinR* v.4.2-23 (Charif and Lobry 2007) for file handling. The alignment was followed by a masking step, in which poorly aligned sequences were excluded from downstream analysis. Sequence divergence was then calculated using *CodeML* of the *paml* v.4.9 package (Yang 2007). A phylogenetic tree based on available Anomuran and Brachyuran crab genomes was inferred using *IQ-TREE2* v.2.1.4-beta (Minh et al. 2020) as implemented in *funannotate*'s *compare* function with the spiny lobster, *Panulirus ornatus* (Veldsman et al. 2021), as an outgroup. For this, we performed 1,000 bootstrap replicates.

Results

PacBio HiFi sequencing resulted in a total throughput of 63.1 Gb for *P. cinctipes* (read N50: 12.9 Kb) and 81.4 Gb for *P. manimaculis* (read N50: 13.6 Kb), which we individually used for genome assembly of the two species. The assembled and filtered genome for *P. cinctipes* comprised 9.4K contigs with an assembly N50 of 707 Kb and a total length of 1.49 Gb (Table 2). The number of contigs (7.5K), assembly N50 (219 Kb), and total length (0.92 Gb) were all lower for *P. manimaculis* (Table 2). The genome assembly length of *P. cinctipes* was closer to the genome size estimate of a species of the same genus (~2.05 Gbp in *P. galathinus*; Rheinsmith et al. 1974). Despite the differences in overall sequence length, the two genomes had equivalent completeness with 94% and 95% complete BUSCOs in *P. cinctipes* and *P. manimaculis*, respectively (Table 2

Table 2 Sequencing and annotation statistics for two species of porcelain crab.

| | <i>Petrolisthes cinctipes</i> | <i>Petrolisthes manimaculis</i> |
|----------------------------------|-------------------------------|---------------------------------|
| Assembly | | |
| Total length (Gb) | 1.49 | 0.92 |
| GC content (%) | 39.63 | 38.23 |
| Contig N50 (Kb) | 706.73 | 218.94 |
| Contig number | 9391 | 7482 |
| BUSCO completeness score (%) | 91.7 | 92.3 |
| Annotation | | |
| Total length of repeats (Gb) | 1.12 | 0.54 |
| % Repeats | 75% | 59% |
| Number of protein-coding genes | 44,543 | 40,315 |
| Mean transcript length (bp) | 6,672 | 6,277 |
| Mean coding sequence length (bp) | 1,200 | 1,218 |
| Mean exon length (bp) | 299 | 275 |
| Mean intron length (bp) | 1,362 | 1,188 |
| Average exons per gene | 4.8 | 5.1 |
| BUSCO completeness score (%) | 94.3 | 95.1 |

Table 3. Different repeat categories found using RepeatModeler coupled with RepeatMasker.

| | <i>Petrolisthes cinctipes</i> | <i>Petrolisthes manimaculis</i> |
|-----------------|-------------------------------|---------------------------------|
| Retroelements | 203 | 100 |
| DNA transposons | 29 | 25 |
| Unclassified | 806 | 329 |
| Small RNA | 2 | 2 |
| Satellites | <1 | <1 |
| Simple repeats | 67 | 78 |
| Low complexity | 6 | 8 |
| Total | 1,115 | 543 |

All values are in Mb. Unclassified repeats could not be assigned to any category and might represent species-specific repeat families.

and [Supplementary Table 1](#)). Additionally, the total number of protein-coding genes identified was similar in the two species, with 45K and 40K for *P. cinctipes* and *P. manimaculis*, respectively ([Table 2](#)). Within protein coding genes, the mean transcript length, exon length, and exons per gene were also similar between the two species ([Table 2](#)).

Comparing the sequence content of the two genome assemblies, the difference in length can largely be explained by differing level of repeat content ([Table 3](#)). Despite the presence of many repetitive regions, which might be species specific, we found that 36.47% of the *P. cinctipes* sequence had matches in *P. manimaculis* by whole genome sequence alignment ([Fig. 3a](#)). To be able to compare genetic regions that not only share sequence similarity but also share the same evolutionary origin, i.e. are homologous, we used alignments of single copy orthologs. Assessing divergence between single copy orthologs of the two species, we found a mean d_N value of 0.154 (95% confidence interval 0.148–0.160) and a mean d_N/d_S value of 0.266 (0.260–0.273) indicating a relatively low level of divergence. For a larger scale comparison of homologous sequence between the two species, we identified 10 contigs of at least 200 Kb in length in which the two species shared at least 17 single copy orthologs ([Supplementary Fig. 1](#) and [Fig. 3b](#)). An examination of those contig pairs indicated that genes were always in a similar order ([Fig. 4](#), top panel), though there were some differences between the genomic regions in terms of gene spacing ([Fig. 4](#), middle panels) and gene sequence ([Fig. 4](#), bottom panels). For example, in contig pair A, there is a region of insertion/deletion of approximately 150 Kb ([Fig. 4](#), middle panels). Homologous contig pairs were for the most part syntenic in their overlapping regions ([Table 4](#) and [Supplementary Fig. 2](#)), but in contig pair A, there was a nonsyntenic region in which none of the genes were shared between species ([Fig. 4](#)). Orthologs within these contig pairs showed very high sequence homology.

Focusing on the predicted gene functions obtained from funannotate, we found similar distributions in both *Petrolisthes* genomes

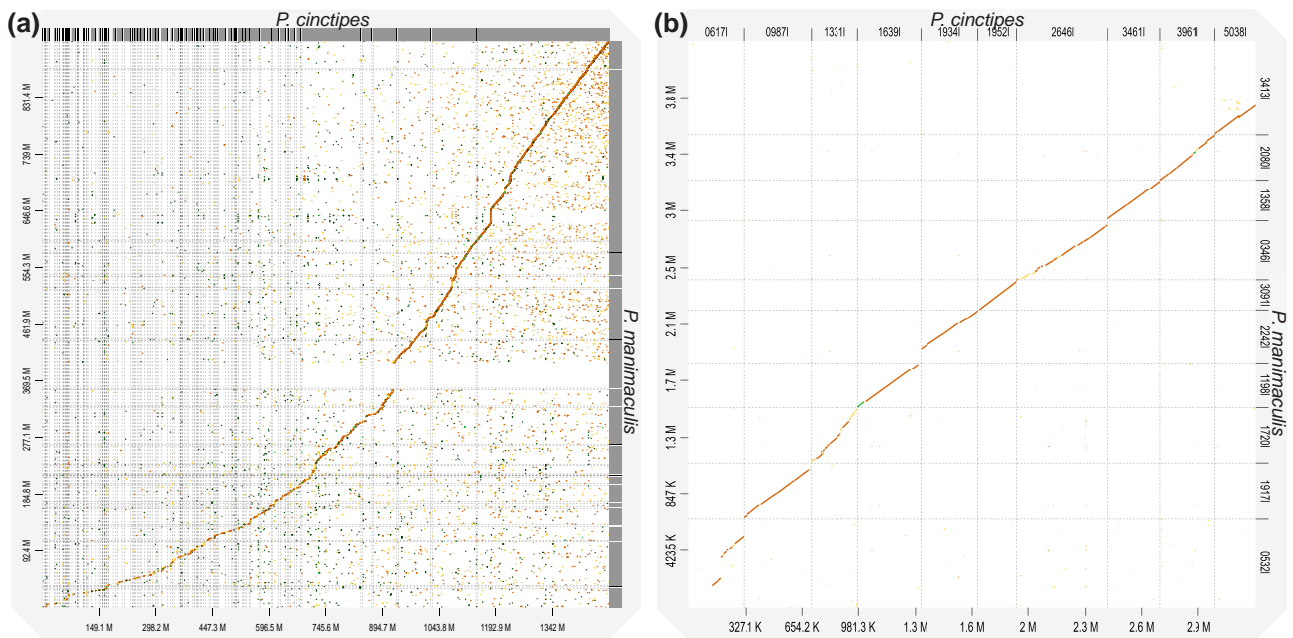


Fig. 3. Pairwise dotplots for (a) entire genome assemblies and (b) 10 selected homologous contigs (also see [Supplementary Fig. 1](#), [Table 4](#)). In both plots, contigs from *P. cinctipes* are on the X-axis and contigs from *P. manimaculis* are on the Y-axis. In panel a, what looks like a gap in the *P. cinctipes* assembly is an assortment of numerous small *P. manimaculis* contigs (clustered together by D-genies) not present in *P. cinctipes*. If the sorting and clustering would be done based on *P. manimaculis*, a gap-like pattern would appear in the *P. manimaculis* assembly.

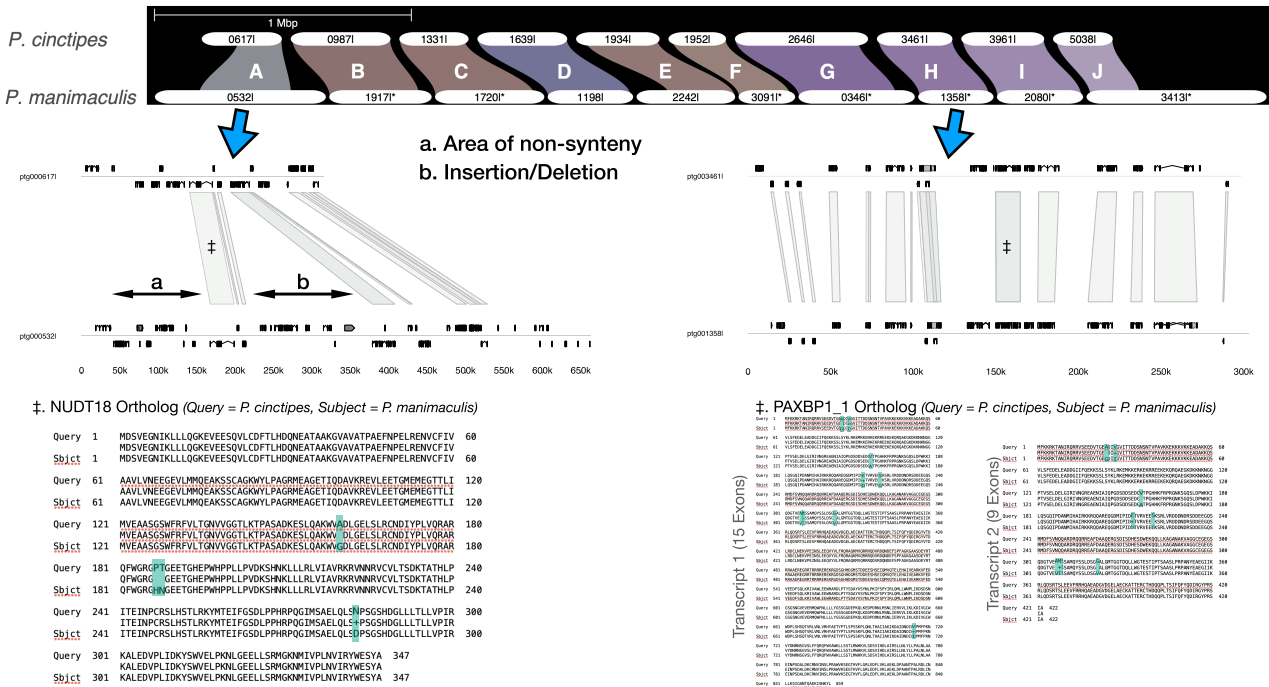


Fig. 4. Syntenic map of orthologous regions among *P. cinctipes* and *P. manimaculis* for contigs selected based on gene density and size (see Fig. 3, Table 4). Ribbons are named (“A” to “J”) and color coded by contig. The beginning of each contig name (ptg00) was trimmed and contigs with an asterisk were inverted to improve visibility. Gene order and spacing for two contig pairs (“A” and “H”) are provided, and gene sequence comparison for one of the larger genes within each contig pair illustrates interspecific differences in the genomes at structural (e.g. nonsynteny and indel in “A”) and sequence levels (blue highlighted amino acids in sequences). For more information about each contig pair, see Table 4, and for detailed figures of each contig pair, see Supplementary Fig. 2.

(Fig. 5), which was expected for species of the same genus. The combined study of the crabs gene arrangement with their expression level will hopefully provide insight into their adaptive evolution. Using additional Anomuran and Brachyuran crab species’ genomes, available from the NCBI database (Table 5), we conducted a phylogenetic analysis. We identified 114 single-copy orthologs and generated a maximum-likelihood tree which supports the phylogenetic placement of the Porcellanidae within the Anomura separate from the Lithodidae (*Paralithodes camtschaticus*) and Coenobitidae (*Birgus latro*) (Fig. 6) (Wolfe et al. 2021).

Discussion

Crabs are an exceptionally species-rich and diverse taxon (De Grave et al. 2009; Wang et al. 2021), whose evolution might be driven by ecology, physiology, and gene rearrangements (Tang et al. 2021; Veldsman et al. 2021; Wang et al. 2021). Available nuclear genomic resources of crabs are sparse but needed for comparative genomic approaches, which would allow investigating the evolutionary role of these factors in crabs. Here, we present two Anomuran crab genome assemblies, *P. cinctipes* and *P. manimaculis*, the first ones of the Anomuran porcelain crabs, family Porcellanidae. We found differences in genome size, genome structure, and gene sequence in homologous regions of the genomes. The largest differences between the two species include a larger genome size and a higher repeat content of *P. cinctipes*. Together, these findings suggest that minor differences in coding regions reflect just a part of the different evolutionary trajectories of the two species when considered with larger scale structure of the two species’ genomes.

Though there are about 9K crab species, genome sequence data are sparse: Only few nuclear genome assemblies ($n = 13$) are available, mainly for four species of Brachyuran ($n = 10$) crabs and three species of Anomuran ($n = 3$) crabs (Table 5). The genomes of Brachyuran crabs have been assembled to chromosome-level in Portunidae [*Callinectes sapidus* “blue crab” (Bachvaroff et al. 2021), *Portunus trituberculatus* “swimming crab” (Tang et al. 2020)] and Varunidae (*Eriocheir sinensis* “Chinese mitten crab”; Cui et al. 2021), as well as to a nonchromosome level in Majidae [*Chionectes opilio* “snow crab” (NCBI database; Assembly name: ASM1658430v1; GenBank assembly accession: GCA_016584305.1; Bioproject accession: PRJNA602365)]. While not a chromosome-level assembly, the genome of a third Portunid crab has been sequenced and had a similar genome size and other characteristics to the other Portunid crabs (*Charybdis japonica* “Asian paddle crab”; Liu et al. 2022). Only one Anomuran crab species has been assembled to chromosome-level (Lithodidae: *Paralithodes platypus* “blue king crab”; Tang et al. 2021) and there are nonchromosomal genome assemblies for two additional species (Lithodidae: *Paralithodes camtschaticus* “red king crab” and Coenobitidae: *Birgus latro* “coconut crab”; Veldsman et al. 2021). Comparing our genome assemblies to the available ones, the GC content in all assemblies is close to 40%. Our assemblies feature the highest completeness measured as BUSCO score and the highest number of genes. Assembly lengths of our porcelain crab assemblies are more similar to the Brachyuran crabs than to the other Anomuran crabs, which have about four times longer assemblies. Our porcelain crab assemblies, however, have a repeat content amount that is more similar to other Anomuran crabs than to Brachyuran crabs. These differences might reflect differential evolution among crab species (Iannucci et al. 2022) but might

Table 4. Interspecific synteny analysis of contigs from *P. cinctipes* (Cinc) and *P. manimaculis* (Mani) selected on the basis of length and number of single copy orthologous genes (orthologs).

| Contig pair | Cinc_Contig | Mani_Contig | Length of overlap (Kb) | Nr. shared single copy orthologs | Nr. Cinc nonsingle copy orthologs | Nr. Mani nonsingle copy orthologs | IDs Cinc nonsingle copy orthologs [Pcinc_v1.9_] | IDs Mani nonsingle copy orthologs [Pmani_v1.7_] |
|-------------|-------------|-------------|------------------------|----------------------------------|-----------------------------------|-----------------------------------|---|--|
| A | ptg000617l | ptg000532l | 244 | 8 | 2 | 14 | g782, g783 | g3404, g3405, g3406, g3407, g3408, g3409, g3410, g3411, g3413, g3416, 4266_g, g3417, g3418, g3419 |
| B | ptg000987l | ptg001917l | 365 | 23 | 13 | 10 | 10939_g, g42538, g42539, g42540, g42544, 10944_g, g42551, g42552, 10956_g, g42558, g42560, g42564, g42565 | g29145, 15539_g, g29148, g29152, 15545_g, 15547_g, g29154, g29160, 15562_g, g29171 |
| C | ptg001331l | ptg001720l | 275 | 19 | 2 | 13 | g1285, g1286 | g19593, g19595, g19596, g19608, g19609, g19613, g19614, g19615, g19616, g19617, 14126_g, g19618, g19619 |
| D | ptg001639l | ptg001198l | 334 | 22 | 6 | 13 | g34991, g34993, g35004, g35005, g35010, g35013 | g16436, g16437, g16438, g16440, g16444, 10076_g, g16448, 10078_g, 10082_g, g16450, g16456, g16458, g16463 |
| E | ptg001934l | ptg002242l | 239 | 18 | 9 | 2 | g3817, g3820, g3821, g3825, g3828, 18677_g, g3830, g3832, g3835 | g31995, g31999 |
| F | ptg001952l | ptg003091l | 204 | 42 | 7 | 14 | g17612, g17632, 18771_g, g17646, g17651, 18794_g, g17656 | g14228, g14237, g14239, g14241, g14242, g14243, g14244, g14246, g14256, g14258, g14259, g14276, g14280, g14281 |
| G | ptg002646l | ptg000346l | 409 | 19 | 11 | 3 | g16185, g16186, g16188, g16191, g16192, g16194, g16195, g16196, g16197, g16198, g16207 | g4887, g4891, g4900 |
| H | ptg003461l | ptg001358l | 270 | 17 | 9 | 3 | g39945, 28259_g, 28263_g, g39954, g39955, g39956, g39958, g39963, g39964 | g28965, g28967, 11234_g |
| I | ptg003961l | ptg002080l | 321 | 18 | 7 | 6 | g22530, g22531, g22534, 31088_g, 31090_g, 31091_g, g22544 | g8956, g8957, g8969, g8970, 16629_g, g8973 |
| J | ptg005038l | ptg003413l | 188 | 29 | 11 | 9 | g29107, g29109, g29122, g29126, g29127, g29128, g29130, g29134, g29135, g29136, g29137 | g1785, g1789, g1790, g1791, g1792, g1796, g1797, g1802, g1815 |

Ortholog IDs starting with “g” were predicted by AGUSTUS gene predictor software, and ortholog IDs ending with “_g” were predicted by GeneMark.hmm gene predictor software (see methods for details). Please see the full GFF file for additional details on each ortholog. Contig pairs refer to Fig. 4.

also partially arise from sequencing artifacts owing to the different sequencing technologies used for the assemblies.

Previous studies on the two *Petrolisthes* species studied here have found that their responses to stress differ at the transcriptomic level (Armstrong and Stillman 2016). Given that gene arrangements have been suggested to be involved in the mitochondrial evolution of crabs (Wang et al. 2021; Zhang et al. 2021; Sun et al. 2022), the question arises of the extent to which differences in the physiology and ecology of the two crab species is reflected in their genomes. A common hypothesis is that under stress, there are more genetic rearrangements (Heng and Heng 2021, 2022). The high-quality genome assemblies of the two porcelain crab species presented here can be combined with existing knowledge of their ecological, physiological, and transcriptional differences to provide a maximally integrative investigation of their adaptive evolution and understanding of the mechanisms driving their physiological differences. For example, the absolute

and relative location of differentially expressed genes in the two species can be compared, which allows inferences about their mobility level as compared to nondifferentially expressed genes. Such investigations should consider the genome size differences in the here-generated assemblies. The larger genome *P. cinctipes* features more repetitive regions (Table 3) and a higher number of duplicated BUSCO genes (Supplementary Table 1), which could indicate different performance of the applied software for assembly or biological differences. In the latter, relaxed selection could enable the proliferation of repetitive elements and gene duplication in *P. cinctipes*. The generated genome assemblies add to an ever-increasing number of available crab genomes, improving the potential for deeper insights into the evolution of the genomes and the diverse traits in Anomuran and Brachyuran Decapod crustaceans.

Comparative genomics yield most valuable insights when applied to completest possible genome assemblies of highest

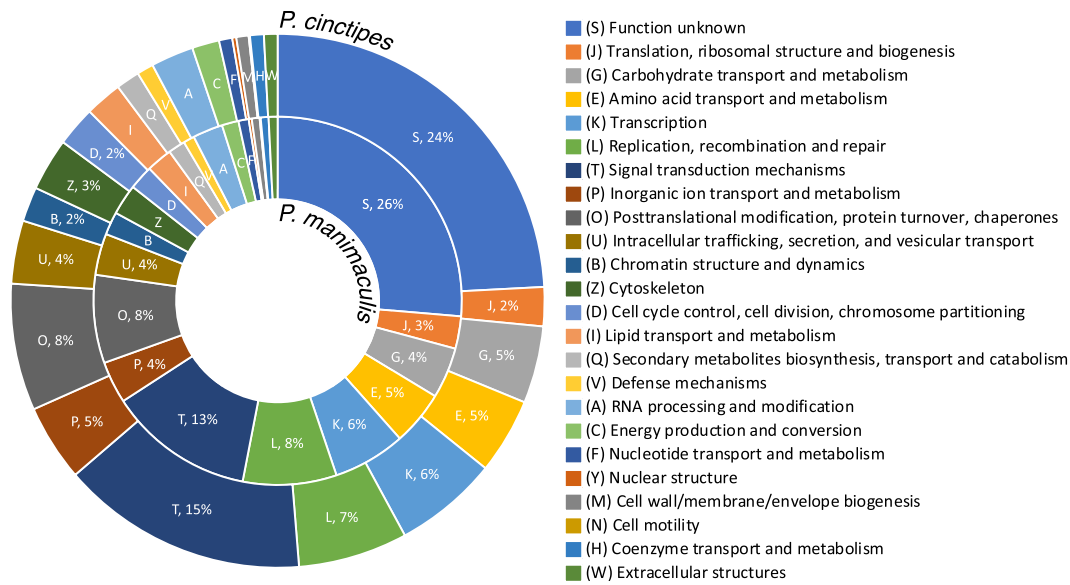


Fig. 5. Predicted functions of gene sets based on clusters of orthologous genes (COGs). Depicted are the distributions of the predicted gene function categories in *P. cinctipes* (outer ring) and *P. manimaculis* (inner ring). These distributions are similar. The legend's order reflects the clockwise order of the functional categories in the graph. Categories which consist of only a few genes are not labeled with percentages to increase visibility.

Table 5. Summary of Brachyuran and Anomuran genome sequencing projects.

| Species/(Infraorder) | Genome Size (Gb) | Repeat % | GC Content % | Protein coding genes (1,000 s) | Complete BUSCO % | Reference |
|---|------------------|----------|--------------|--------------------------------|------------------|--------------------------|
| Brachyura | | | | | | |
| <i>Callinectes sapidus</i> (Portunidae) | 1.1 | 36 | 40 | 25 | 93 | Bachvaroff et al. (2021) |
| <i>Portunus trituberculatus</i> (Portunidae) | 1.0 | 54 | 41 | 17 | 95 | Tang et al. (2020) |
| <i>Eriocheir sinensis</i> (Varunidae) | 1.6 | 45 | 41 | 28 | 92 | Cui et al. (2021) |
| <i>Chionoecetes opilio</i> (Majidae) | 2 | NA | 42 | 22 | NA | NCBI GCA_016584305.1 |
| Anomura | | | | | | |
| <i>Paralithodes platypus</i> (Lithodidae) | 4.8 | 78 | 42 | 28 | 77 | Tang et al. (2021) |
| <i>Paralithodes camtschaticus</i> (Lithodidae) | 7.3 | 68 | 41 | 29 | 90 | Veldsman et al. (2021) |
| <i>Birgus latro</i> (Coenobitidae) | 6.2 | 24 | 42 | 24 | 90 | Veldsman et al. (2021) |
| <i>Petrolisthes cinctipes</i> (Porcellanidae) | 1.5 | 75 | 40 | 45 | 94 | This Study |
| <i>Petrolisthes manimaculis</i> (Porcellanidae) | 0.9 | 59 | 38 | 40 | 95 | This Study |

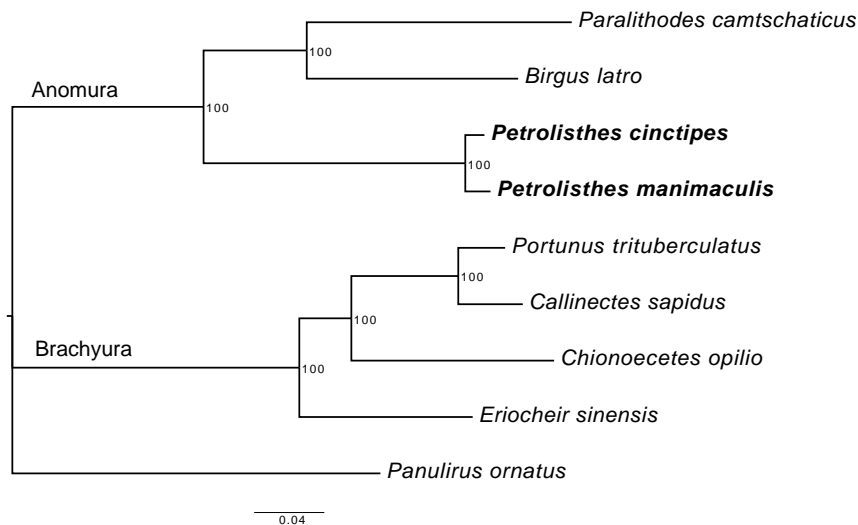


Fig. 6. Maximum-likelihood tree of available Brachyuran and Anomuran genomes. The two *Petrolisthes* species are less related to the other two Anomuran crab species than these two other species to each other. The spiny lobster, *P. ornatus* (Palinuridae), was used as an outgroup. Node labels represent bootstrap values.

possible quality. Our genome assemblies were generated based on the most reliable available sequencing technology for de novo genome sequencing, but there are options which might improve their contiguity further. For example, with the available resources, it is possible to use the sequence information from one species to scaffold the genome assembly of the other species and vice versa, because of their moderate sequence divergence. This approach, however, would only improve the assemblies to a small degree and at the same time might lead to wrong sequence links in cases of genomic rearrangements. Other possibilities for scaffolding and therefore improving the assemblies would involve the generation of additional data using Oxford Nanopore (Price et al. 2023; Salson et al. 2023) and Hi-C (Bracewell et al. 2023) sequencing, both of which are established scaffolding approaches. Furthermore, even though we applied state-of-the-art methodology to identify (non-)focal DNA sequence in our assemblies, future work should focus on generating higher quality DNA from less contaminant-prone tissue, like muscle tissue. This would reduce uncertainty in the identification of (non-)focal DNA sequence, i.e. reduce the number of nonfocal contigs still included in the genome assemblies. Using approaches such as those would be a next step toward generating the first-ever chromosome-level assembly for a porcelain crab.

Data availability

Raw data is deposited at the NCBI SRA database, and the assembled genomes as well as the predicted sets of protein sequences are available at the NCBI GenBank database (BioProject ID: PRJNA1002960) and at <https://doi.org/10.6084/m9.figshare.23823531>. Analysis scripts are deposited at https://github.com/pascalangst/Petrolisthes_assemblies.

Supplemental material available at G3 online.

Acknowledgments

We thank Jeremah Ets-Hokin for his help in specimen collection and dissection. We thank Marlon Henseler and John Lovell for their help in data visualization. Computationally demanding calculations were performed at sciCORE (<https://scicore.unibas.ch/>) scientific computing center at University of Basel.

Funding

This work was supported in part by the National Science Foundation (NSF) grant BIO-IOS-1558159 to JHS. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Conflicts of interest

The authors declare no conflicts of interest.

Literature cited

- Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 37(4):420–423. doi:10.1038/s41587-019-0036-z.
- Armstrong EJ, Stillman JH. 2016. Construction and characterization of two novel transcriptome assemblies in the congeneric porcelain crabs *Petrolisthes cinctipes* and *P. Manimaculis*. *Integr Comp Biol.* 56(6):1092–1102. doi:10.1093/icb/icw043.
- Bachvaroff TR, McDonald RC, Plough LV, Chung JS. 2021. Chromosome-level genome assembly of the blue crab, *Callinectes sapidus*. *G3 (Bethesda)*. 11(9):jkab212. doi:10.1093/g3journal/jkab212.
- Bracewell RR, Stillman JH, Dahlhoff EP, Smeds E, Chatla K, Bachtrog D, Williams C, Rank NE. 2023. A chromosome-scale genome assembly and evaluation of mtDNA variation in the willow leaf beetle *Chrysomela aeneicollis*. *G3 (Bethesda)*. 13(7):jkad106. doi:10.1093/g3journal/jkad106.
- Brūna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform.* 3(1):lqaa108. doi:10.1093/nargab/lqaa108.
- Brūna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform.* 2(2):lqaa026. doi:10.1093/nargab/lqaa026.
- Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods.* 18(4):366–368. doi:10.1038/s41592-021-01101-x.
- Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 6:e4958. doi:10.7717/peerj.4958.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 10(1):421. doi:10.1186/1471-2105-10-421.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 38(12):5825–5829. doi:10.1093/molbev/msab293.
- Cao S, Brandis G, Huseby DL, Hughes D. 2022. Positive selection during niche adaptation results in large-scale and irreversible rearrangement of chromosomal gene order in Bacteria. *Mol Biol Evol.* 39(4):msac069. doi:10.1093/molbev/msac069.
- Charif D, Lobry JR. 2007. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*. Berlin, Heidelberg: Springer. p. 207–232.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 18(2):170–175. doi:10.1038/s41592-020-01056-5.
- Cui ZX, Liu Y, Yuan JB, Zhang XJ, Ventura T, Ma KY, Sun S, Song C, Zhan D, Yang Y, et al. 2021. The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. *Nat Commun.* 12(1):2395. doi:10.1038/s41467-021-22604-3.
- Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. Zenodo. <https://www.doi.org/10.5281/zenodo.3552717>.
- DE Azevedo Ferreira LA, Anker A. 2021. An annotated and illustrated checklist of the porcelain crabs of Panama (Decapoda: Anomura). *Zootaxa.* 5045(1):1–154. doi:10.11646/zootaxa.5045.1.1.
- De Grave S, Pentcheff ND, Ahyong ST, Chan TY, Crandall KA, Dworschak PC, Felder DL, Feldmann RM, Franssen CHJM, Goulding LYD, et al. 2009. A classification of living and fossil genera of Decapod Crustaceans. *Raffles Bull Zool (Suppl 21)*:1–109. <https://api.semanticscholar.org/CorpusID:80915884>.

- Delmanowski RM, Brooks CL, Salas H, Tsukimura B. 2017. Reproductive life history of *Petrolisthes cinctipes* (Randall, 1840) and *P-manimaculis* Glassell, 1945 (Decapoda: Anomura: Porcellanidae), with the development of an enzyme-linked immunosorbant assay (ELISA) for the determination of hemolymph levels of vitellogenin. *J Crust Biol.* 37(3):315–322. doi:10.1093/jcbiol/rux017.
- Delmanowski RM, Tsukimura B. 2015. Characterization of vitellins from *Petrolisthes cinctipes* and *Petrolisthes manimaculis* and the development of a compatible ELISA. *Integr Comp Biol.* 55:E244. doi:10.1093/jcbiol/rux017.
- Diez YL, Lira C. 2017. Systematics and biogeography of Cuban porcelain crabs (Decapoda: Anomura: Porcellanidae). *Zootaxa.* 4216(5): zootaxa.4216.5.2. doi:10.11646/zootaxa.4216.5.2.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20(1):238. doi:10.1186/s13059-019-1832-y.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA.* 117(17): 9451–9457. doi:10.1073/pnas.1921046117.
- Gaitan-Espitia JD, Bacigalupe LD, Opitz T, Lagos NA, Timmermann T, Lardies MA. 2014. Geographic variation in thermal physiological performance of the intertidal crab *Petrolisthes violaceus* along a latitudinal gradient. *J Exp Biol.* 217(Pt. 24):4379–4386. doi:10.1242/jeb.108217.
- Garcia-Rios E, Guillamon JM. 2022. Genomic adaptations of *Saccharomyces* genus to Wine Niche. *Microorganisms.* 10(9): 1811. doi:10.3390/microorganisms10091811.
- Gourlie R, McDonald M, Hafez M, Ortega-Polo R, Low KE, Abbott DW, Strelkov SE, Daayf F, Aboukhaddour R. 2022. The pangenome of the wheat pathogen *pyrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors ToxA and ToxB. *BMC Biol.* 20(1):239. doi:10.1186/s12915-022-01433-w.
- Gunderson AR, Abegaz M, Ceja AY, Lam EK, Souther BF, Boyer K, King EE, You Mak KT, Tsukimura B, Stillman JH. 2019. Hot rocks and not-so-hot rocks on the seashore: patterns and body-size dependent consequences of microclimatic variation in intertidal zone boulder habitat. *Integr Org Biol.* 1(1):obz024. doi:10.1093/iob/obz024.
- Gunderson AR, King EE, Boyer K, Tsukimura B, Stillman JH. 2017. Species as stressors: heterospecific interactions and the cellular stress response under global change. *Integr Comp Biol.* 57(1): 90–102. doi:10.1093/icb/ix019.
- Hackl T, Duponchel S, Barenhoff K, Weinmann A, Fischer MG. 2021. Virophages and retrotransposons colonize the genomes of a heterotrophic flagellate. *eLife.* 10:e72674. doi:10.7554/eLife.72674.
- Haig J. 1960. The Porcellanidae (Crustacea Anomura) of the Eastern Pacific. Los Angeles: University of Southern California Press.
- Heng J, Heng HH. 2021. Two-phased evolution: genome chaos-mediated information creation and maintenance. *Progr Biophys Mol Biol.* 165:29–42. doi:10.1016/j.pbiomolbio.2021.04.003.
- Heng J, Heng HH. 2022. Genome chaos, information creation, and cancer emergence: searching for new frameworks on the 50th anniversary of the “war on cancer”. *Genes (Basel).* 13(1):101. doi:10.3390/genes13010101.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47(D1):D309–D314. doi:10.1093/nar/gky1085.
- Iannucci A, Saha A, Cannicci S, Bellucci A, Cheng CLY, Ng KH, Fratini S. 2022. Ecological, physiological and life-history traits correlate with genome sizes in decapod crustaceans. *Front Ecol Evol.* 10: 930888. doi:10.3389/fevo.2022.930888.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 30(9):1236–1240. doi:10.1093/bioinformatics/btu031.
- Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol.* 338(5):1027–1036. doi:10.1016/j.jmb.2004.03.016.
- Kang M, Lim JY, Kim J, Hwang I, Goo E. 2022. Influence of genomic structural variations and nutritional conditions on the emergence of quorum sensing-dependent gene regulation defects in *Burkholderia glumae*. *Front Microbiol.* 13:950600. doi:10.3389/fmicb.2022.950600.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 37(8):907–915. doi:10.1038/s41587-019-0201-4.
- Kropp RK, Haig J. 1994. *Petrolisthes-Extremus*, a new porcelain crab (Decapoda, Anomura, Porcellanidae) from the Indo-West Pacific. *Proc Biol Soc Wash.* 107:312–317.
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res.* 6:1287 doi:10.12688/f1000research.12232.1.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 34(18):3094–3100. doi:10.1093/bioinformatics/bty191.
- Limviriyakul P, Tseng LC, Hwang JS, Shih TW. 2016. Anomuran and brachyuran symbiotic crabs in coastal areas between the southern Ryukyu arc and the coral triangle. *Zool Stud.* 55:e7. doi:10.6620/ZS.2016.55-07.
- Liu M, Ge S, Bhandari S, Fan CL, Jiao Y, Gai C, Wang Y, Liu H. 2022. Genome characterization and comparative analysis among three swimming crab species. *Front Mar Sci.* 9:895119. doi:10.3389/fmars.2022.895119.
- Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife.* 11:e78526. doi:10.7554/eLife.78526.
- Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. *Multiple Sequence Alignment Methods*. Totowa (NJ): Humana Press. p. 155–170.
- Ma WD, Yang J, Ding JQ, Zhao WS, Peng YL, Bhaduria V. 2022. Gapless reference genome assembly of *Didymella glomerata*, a new fungal pathogen of maize causing *Didymella* leaf blight. *Front Plant Sci.* 13:1022819. doi:10.3389/fpls.2022.1022819.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 38(10): 4647–4654. doi:10.1093/molbev/msab199.
- Mantelatto FL, Miranda I, Vera-Silva AL, Negri M, Buranelli RC, Terossi M, Magalhães T, Costa RC, Zara FJ, Castilho AL. 2021. Checklist of decapod crustaceans from the coast of the Sao Paulo state (Brazil) supported by integrative molecular and morphological data: IV. Infraorder Anomura: Superfamilies Chirostyloidea, Galattheoidea, Hippoidea and Paguroidea. *Zootaxa.* 4965(3):558–600. doi:10.11646/zootaxa.4965.3.9.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 27(6): 764–770. doi:10.1093/bioinformatics/btr011.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17(1):10–12. doi:10.14806/ej.17.1.200.

- Miller NA, Paganini AW, Stillman JH. 2013. Differential thermal tolerance and energetic trajectories during ontogeny in porcelain crabs, genus *Petrolisthes*. *J Therm Biol.* 38(2):79–85. doi:10.1016/j.jtherbio.2012.11.005.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37(5):1530–1534. doi:10.1093/molbev/msaa015.
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49(D1):D412–D419. doi:10.1093/nar/gkaa913.
- Naderloo R, Türkay M, Sari A. 2013. Intertidal habitats and decapod (Crustacea) diversity of Qeshm island, a biodiversity hotspot within the Persian gulf. *Mar Biodivers.* 43(4):445–462. doi:10.1007/s12526-013-0174-3.
- Palmer J, Stajich J. 2022. nextgenusfs/funcannotate: funcannotate. Zenodo. <https://zenodo.org/record/2604804>.
- Price RJ, Davik J, Fernandez Fernandez F, Bates HJ, Lynn S, Nellist CF, Buti M, Røen D, Šurbanovski N, Alsheikh M, et al. 2023. Chromosome-scale genome sequence assemblies of the 'Autumn Bliss' and 'Malling Jewel' cultivars of the highly heterozygous red raspberry (*Rubus idaeus* L.) derived from long-read Oxford nanopore sequence data. *PLoS One.* 18(5):e0285756. doi:10.1371/journal.pone.0285756.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 11(1):1432. doi:10.1038/s41467-020-14998-3.
- Raso JEC, Martínez-Iglesias JC, Manjon-Cabeza ME. 2005. Crustacean decapods from cardenas bay and the macrolagoon of the sabana-camaguey archipelago (northern coast of Cuba). *Cah Biol Mar.* 46(1):43–55.
- Rawlings ND, Waller M, Barrett AJ, Bateman A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42(D1):D503–D509. doi:10.1093/nar/gkt953.
- R Core Team. 2022. R: The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rheinsmith EL, Hinegardner R, Bachmann K. 1974. Nuclear DNA amounts in crustacea. *Comp Biochem Physiol B Comp Biochem.* 48(3):343–348. doi:10.1016/0305-0491(74)90269-7.
- Rodríguez IT, Hernández G, Felder DL. 2005. Review of the western atlantic porcellanidae (Crustacea: Decapoda: Anomura) with new records, systematic observations, and comments on biogeography. *Caribb J Sci.* 41(3):544–582.
- Salson M, Orjuela J, Mariac C, Zekraoui L, Couderc M, Arribat S, Rodde N, Faye A, Kane NA, Tranchant-Dubreuil C, et al. 2023. An improved assembly of the pearl millet reference genome using Oxford nanopore long reads and optical mapping. *G3 (Bethesda).* 13(5):jkad051. doi:10.1093/g3journal/jkad051.
- Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 24(5):637–644. doi:10.1093/bioinformatics/btn013.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* 7(1):62. doi:10.1186/1471-2105-7-62.
- Stillman JH. 2002. Causes and consequences of thermal tolerance limits in rocky intertidal porcelain crabs, genus *Petrolisthes*. *Integr Comp Biol.* 42(4):790–796. doi:10.1093/icb/42.4.790.
- Stillman JH, Reeb CA. 2001. Molecular phylogeny of eastern pacific porcelain crabs, genera *Petrolisthes* and *Pachycheles*, based on the mtDNA 16S rDNA sequence: phylogeographic and systematic implications. *Mol Phylogenet Evol.* 19(2):236–245. doi:10.1006/mpev.2001.0924.
- Stillman J, Somero G. 1996. Adaptation to temperature stress and aerial exposure in congeneric species of intertidal porcelain crabs (genus *Petrolisthes*): correlation of physiology, biochemistry and morphology with vertical distribution. *J Exp Biol.* 199(8):1845–1855. doi:10.1242/jeb.199.8.1845.
- Stillman JH, Somero GN. 2000. A comparative analysis of the upper thermal tolerance limits of eastern Pacific porcelain crabs, genus *Petrolisthes*: influences of latitude, vertical zonation, acclimation, and phylogeny. *Physiol Biochem Zool.* 73(2):200–208. doi:10.1086/316738.
- Sun SE, Jiang W, Yuan ZM, Sha ZL. 2022. Mitogenomes provide insights into the evolution of thoracotremata (Brachyura: Eubrachyura). *Front Mar Sci.* 9. doi:10.3389/fmars.2022.818738.
- Sun DA, Patel NH. 2019. The amphipod crustacean *Parhyale hawaiiensis*: an emerging comparative model of arthropod development, evolution, and regeneration. *Wiley Interdiscip Rev Dev Biol.* 8(5):e355. doi:10.1002/wdev.355.
- Tagmount A, Wang M, Lindquist E, Tanaka Y, Teranishi KS, Sunagawa S, Wong M, Stillman JH. 2010. The porcelain crab transcriptome and PCAD, the porcelain crab microarray and sequence database. *PLoS One.* 5(2):e9327. doi:10.1371/journal.pone.0009327.
- Tang B, Wang Z, Liu Q, Wang Z, Ren Y, Guo H, Qi T, Li Y, Zhang H, Jiang S, et al. 2021. Chromosome-level genome assembly of parathodes platypus provides insights into evolution and adaptation of king crabs. *Mol Ecol Resour.* 21(2):511–525. doi:10.1111/1755-0998.13266.
- Tang B, Zhang D, Li H, Jiang S, Zhang H, Xuan F, Ge B, Wang Z, Liu Y, Sha Z, et al. 2020. Chromosome-level genome assembly reveals the unique genome evolution of the swimming crab (*Portunus trituberculatus*). *Gigascience.* 9(1):giz161. doi:10.1093/gigascience/giz161.
- The UniProt Consortium. 2022. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51(D1):D523–D531. doi:10.1093/nar/gkac1052.
- Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub NA, Batterham P, Bellair M, et al. 2020. Gene content evolution in the arthropods. *Genome Biol.* 21(1):15. doi:10.1186/s13059-019-1925-7.
- Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for Multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). p. 314–324.
- Veldsman WP, Ma KY, Hui JHL, Chan TF, Baeza JA, Qin J, Chu KH. 2021. Comparative genomics of the coconut crab and other decapod crustaceans: exploring the molecular basis of terrestrial adaptation. *BMC Genom.* 22(313). doi:10.1186/s12864-021-07636-9.
- Wang B, Lv R, Zhang Z, Yang C, Xun H, Liu B, Gong L. 2022. Homoeologous exchange enables rapid evolution of tolerance to salinity and hyper-osmotic stresses in a synthetic allotetraploid wheat. *J Exp Bot.* 73(22):7488–7502. doi:10.1093/jxb/erac355.
- Wang Q, Wang J, Wu Q, Xu X, Wang P, Wang Z. 2021. Insights into the evolution of *Brachyura* (Crustacea: Decapoda) from mitochondrial sequences and gene order rearrangements. *Int J Biol Macromol.* 170:717–727. doi:10.1016/j.ijbiomac.2020.12.210.
- Wasson K, Lyon BE, Knope M. 2002. Hair-trigger autotomy in porcelain crabs is a highly effective escape strategy. *Behav Ecol.* 13(4):481–486. doi:10.1093/beheco/13.4.481.

- Werdning B, Hiller A. 2015. Description of a new species of *Petrolisthes* in the Indo-West Pacific with a redefinition of *P. hastatus* Stimpson, 1858 and resurrection of *P. inermis* (Heller, 1862) (Crustacea, Anomura, Porcellanidae). *Zookeys*. 516:95–108. doi:[10.3897/zookeys.516.9923](https://doi.org/10.3897/zookeys.516.9923).
- Wolfe JM, Luque J, Bracken-Grissom HD. 2021. How to become a crab: phenotypic constraints on a recurring body plan. *BioEssays*. 43(5): 2100020. doi:[10.1002/bies.202100020](https://doi.org/10.1002/bies.202100020).
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591. doi:[10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088).
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 40(W1):W445–W451. doi:[10.1093/nar/gks479](https://doi.org/10.1093/nar/gks479).
- Zhang Y, Meng L, Wei L, Lu X, Liu B, Liu L, Lü Z, Gao Y, Gong L. 2021. Different gene rearrangements of the genus *Dardanus* (Anomura: Diogenidae) and insights into the phylogeny of Paguroidea. *Sci Rep*. 11(1):21833. doi:[10.1038/s41598-021-01338-8](https://doi.org/10.1038/s41598-021-01338-8).

Editor: K. Vogel