



Genome Resources

De novo whole genome assembly of the globally invasive green shore crab *Carcinus maenas* (Linnaeus, 1758) via long-read Oxford Nanopore MinION sequencing

Jolanda K. Brons^{1,‡}, Thomas Hackl^{1,‡} , Riccardo Iacovelli^{2,3} , Kristina Haslinger², Sebastian Lequime¹ , Sancia E.T. van der Meij^{1,4,*}

¹Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands, ²Groningen Research Institute of Pharmacy, University of Groningen, Groningen, The Netherlands, ³Present address: VTT Technical Research Centre of Finland, Espoo, Finland, ⁴Naturalis Biodiversity Center, Leiden, The Netherlands

*Corresponding author: Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands.

Email: Sancia.van.der.Meij@rug.nl

‡These authors contributed equally to this work.

Abstract

Invasive species are reshaping aquatic ecosystems worldwide at an accelerating pace, with profound ecological and economic impacts. Many crustacean species have demonstrated invasive potential or are already well-established invaders. The green shore crab, *Carcinus maenas*, native to Europe and North Africa, is one of the most successful global marine invaders and is now present on six continents. Although the role of genomics in invasion science is increasingly recognized, genomic resources for brachyuran crabs remain limited, including the notable absence of a reference genome for *C. maenas*. Here we report on a *de novo* whole genome assembly of *C. maenas* via long-read Oxford Nanopore Technology sequencing. The assembly spans 1.09 Gbp across 21 887 scaffolds (NG50 = 13 Mbp) with a BUSCO completeness of 98.4%, providing a high-quality resource for future genomic analyses. We provide a detailed protocol for obtaining high-quality DNA to successfully sequence brachyuran crabs using a long-read approach. This new resource expands available genomic data for the species-rich infraorder Brachyura, and provides a valuable foundation for understanding the genetic factors underlying the global invasion success of *C. maenas*, supporting future research in marine invasion genomics.

Key words: Brachyura, Carcinidae, European green crab, invasion genomics

Introduction

Invasive species are transforming marine habitats worldwide (Molnar et al. 2008), and the rate at which species occur outside their native ranges has increased considerably over the past few centuries. This increase does not show any sign of saturation, indicating that current efforts to mitigate invasions are inadequate to keep up with the accelerating impacts of globalization (Seebens et al. 2017). While the drivers of biological invasion are increasingly global in nature, the impacts of invasions are mainly observed at local scales (Molnar et al. 2008). Some non-native species might provide benefits to an ecosystem (Schlaepfer 2018), while others can cause severe ecological or economic damage, making them invasive (Simberloff et al. 2013). The negative impacts of invasive species range from habitat destruction and disease transmission to displacement or even extinction of native species (Molnar et al. 2008).

Invasive species often possess traits linked to invasion success, including broad environmental tolerance, high reproductive output (r-strategists), phenotypic plasticity, effective dispersal, and competitive advantages (Davidson et al. 2011; Weir & Salice 2011). Additionally, invasive species may

gain an additional advantage by escaping natural enemies, although this effect should not be overstated (Colautti et al. 2004). The success of many invasive species may rely more on their capacity to adapt through natural selection than on general physiological tolerance or plasticity alone (Lee 2002). Whole-genome data are invaluable for advancing the field of invasion genomics (Jaspers et al. 2021) and offer powerful tools to enhance our understanding of biological invasions in aquatic systems (Tepolt 2015; Bourne et al. 2018; North et al. 2021). For example, comparative whole-genome sequencing in crustaceans demonstrated that many expanded gene families are involved in the environmental tolerance of the red swamp crayfish *Procambarus clarkii* (Girard, 1852) (Xu et al. 2021), whereas the invasion potential of the Chinese mitten crab *Eriocheir sinensis* (H. Milne Edwards, 1853) is largely attributed to its strong osmoregulatory capacity and high fertility (Cui et al. 2021). Despite their potential, these resources so far remain underexploited in efforts to predict invasive potential (Kołodziejczyk et al. 2025).

The list of the 100 most notorious globally invasive species in the Global Invasive Species Database, compiled by Lowe et al. (2000), includes nine aquatic invertebrates. Whole

Received on 21 May 2025; revised on 4 September 2025; accepted on 20 October 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of The American Genetic Association.

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.

genome data is currently available for six of these species. Three out of six species are molluscs: the golden apple snail (*Pomacea canaliculata* Lamarck, 1822) (Lu et al. 2024), the Mediterranean mussel (*Mytilus galloprovincialis*; Lamarck, 1819) (Han et al. 2024), and the zebra mussel (*Dreissena polymorpha* Pallas, 1771) (McCartney et al. 2022). Whole genome data is furthermore available for an echinoderm, the Northern Pacific seastar (*Asterias amurensis* Lutken, 1871) (Wang et al. 2023); a ctenophore, *Mnemiopsis leidyi* (A. Agassiz, 1865) (Ryan et al. 2013); and a crustacean, the Chinese mitten crab (*E. sinensis*) (Cui et al. 2021). Whole genome data is so far missing for three invasive aquatic species: the fishhook waterflea (*Cercopagis pengoi* Ostroumov, 1891), the Asian clam (*Potamocorbula amurensis* Schrenck, 1861), and the green shore crab (*Carcinus maenas* Linnaeus, 1758).

Carcinus maenas is native to Europe and northwest Africa. It is one of the world's most widespread invasive marine species, now recorded from six continents (Carlton and Cohen 2003; GBIF 2025; Supplementary Fig. 1). Despite its global distribution and ecological impact, no reference genome is currently available for this species. As a generalist predator with a broad diet, *C. maenas* poses serious threats to native biodiversity across its introduced range (Ens et al. 2022). Due to its adaptability and resilience, *C. maenas* is likely to continue expanding into new ecosystems and may trigger secondary invasion waves in areas it already inhabits (Jeffery et al. 2017; Frederich and Lancaster 2024—and references therein). The lack of genomic data for *C. maenas* highlights a wider gap in genomic resources for decapod crustaceans (Yuan et al. 2023). Obtaining whole decapod genomes is particularly challenging due to their large size and high content of repetitive DNA (Van Quyen et al. 2020; Cui et al. 2021; Rutz et al. 2023). The infraorder Brachyura (true crabs) accounts for almost half of the total diversity in Decapoda, and this group has not only successfully conquered the marine realm, from the deep sea to the intertidal, but is also successful in a multitude of freshwater and terrestrial environments (De Grave et al. 2023). Despite the diversity of this group, with close to 8,000 species, and the presence of several notorious invaders and economically important species, only 10 reference genomes across five families are available for brachyuran crabs (NCBI 2025), with seven more on the way (Wang et al. 2025).

In this study, we present a whole genome assembly of *C. maenas* (family Carcinidae MacLeay 1838), based on a specimen from within its native range, using long-read sequencing (Oxford Nanopore Technologies (ONT), Oxford, UK). Nanopore technology is a third-generation sequencing method that produces long reads of consistent quality and is recognized as an affordable option for de novo sequencing, even for complex genomes (de Lannoy et al. 2017). Although this technology has significant potential to advance genomic research on non-model organisms like aquatic invertebrates, its application is often hindered by challenges in extracting and purifying genomic DNA from many species (Daniels et al. 2023). To address this, we provide a detailed, step-by-step protocol for obtaining sufficient DNA for Nanopore sequencing and genome assembly, along with a practical guide to sequencing brachyuran crabs using a MinION.

Material and methods

Specimen collection

Preliminary test runs were performed using an adult male *C. maenas* collected on 26 April 2023 from the marina of

Schiermonnikoog, the Netherlands (53.470427, 6.166608). The specimen (Supplementary Fig. 2A) was preserved in 96% ethanol and stored at -20°C . Data from these initial runs contributed to the final genome assembly. The whole genome assembly was based on a second adult male *C. maenas* (Supplementary Fig. 2B), collected on 11 April 2024 from the harbor of Lauwersoog, the Netherlands (53.410545, 6.208381). This specimen was preserved in DESS solution containing 20% dimethyl sulfoxide, 0.25 M ethylenediaminetetraacetic acid (EDTA), and saturated sodium chloride (NaCl), then stored at -20°C until DNA extraction (Oosting et al. 2020). Both crabs remain vouchered at the University of Groningen.

Preliminary MinION runs and method optimization

During initial MinION test runs using muscle tissue from the Schiermonnikoog specimen, consistent pore blockage occurred approximately 2 h into sequencing. To address this, we tested two preservation fluids, optimized DNA extraction protocols, employed two short fragment removal (SFR) kits and a whole genome amplification (WGA) to eliminate potential contaminants responsible for the blockages.

First, we evaluated two different DNA extractions methods, the Modified Qiagen G/20 Tip DNA Extraction and Powersoil Pro DNA Extraction.

Modified Qiagen G/20 tip DNA extraction

Approximately 400 mg of muscle tissue from the specimens' pereopods was finely minced. The tissue was suspended in 800 μl of 1x phosphate-buffered saline (PBS, Fisher Scientific, Hampton, USA) and centrifuged at $4,000 \times g$ for 1 min at room temperature. The supernatant was discarded, and the washing step was repeated. The pellet was then resuspended in a solution of 16 μl RNase (Qiagen, Hilden, Germany) and 2 ml G2 buffer (Genomic DNA Buffer Set, Qiagen, Hilden, Germany). Subsequently, 250 μl of proteinase K (>600 mAU/ml, Qiagen, Hilden, Germany) was added, and the mixture was incubated at 50°C for 18 h.

After incubation, the mixture was centrifuged at $4,000 \times g$ for 5 min at room temperature. The supernatant was transferred to a new tube, and 1 ml of InhibitEX buffer (Qiagen, Hilden, Germany) was added to remove contaminants (Boughattas et al. 2021). The mixture was incubated at room temperature for 5 min and then centrifuged again at $4,000 \times g$ for 5 min. The resulting supernatant was used for genomic DNA extraction.

Genomic DNA from *C. maenas* was extracted using the "Isolation of Genomic DNA from Blood, Cultured Cells, Tissue, Yeast, or Bacteria Using Genomic-tips" protocol outlined in the Qiagen Genomic DNA Handbook, with the following slight modifications. A Qiagen genomic-tip 20/G (Qiagen Genomic-tip 20/G Kit, Qiagen, Hilden, Germany) was equilibrated with 1.0 ml of QBT buffer (Genomic DNA Buffer Set) and allowed to empty via gravity flow. The *C. maenas* sample was then applied to the prepared genomic-tip 20/G. After the sample passed through, the genomic tip was washed three times with 1.0 ml of QC buffer (Genomic DNA Buffer Set).

The genomic DNA was eluted with 2.0 ml of pre-warmed (50°C) QF buffer (Genomic DNA Buffer Set) and precipitated by adding 0.7 volumes (1.4 ml) of room-temperature (15°C to 25°C) isopropanol (Sigma-Aldrich, Saint Louis, USA). The mixture was gently inverted several times and incubated at room temperature for 5 min. It was then centrifuged at

5,000 × g for 15 min at 4°C. The supernatant was discarded, and the resulting pellet was washed with 1 ml of cold 70% ethanol (4°C) and centrifuged again at 5,000 × g for 10 min at 4°C. After removing the supernatant, the pellet was air-dried for 10 min and resuspended in 60 µl of pre-warmed (50°C) Tris-EDTA (TE) buffer (Sigma-Aldrich, Saint Louis, USA). The resuspension was incubated at 37°C for 1 h. DNA concentrations were measured using a Qubit Flex Fluorometer (Fisher Scientific, Waltham, USA). The A260/280 and A260/230 ratios were also assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, USA).

Powersoil pro DNA extraction

The crab specimen was subsampled again, and approximately 400 mg of pereopod muscle tissue was finely chopped. The tissue was combined with 800 µl of 1x PBS (Fisher Scientific, Hampton, USA) and centrifuged at 4,000 × g for 1 min at room temperature. The supernatant was discarded, and the washing step was repeated twice. DNA extraction was performed using the Powersoil Pro DNA Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, with the following modifications: 0.25 g of Zirconia beads (Biospec Products, Bartlesville, USA) were added, and the vortexing step (step 2, 10 min) was replaced with two 60-s bead-beating cycles at 6.0 m/s using a FastPrep-24™ 5G bead beater (MP Biomedicals, Santa Ana, USA). DNA quality and quantity were assessed as described above.

Next, an SFR step was applied to both DNA extraction methods to enhance the mean read length and investigate pore blockage caused by small reads.

SFR

This process removed short DNA fragments up to 10 kb (SFR-A) or 25 kb (SFR-B). The SFR-A solution was prepared with 4% PVP 360,000, 1.2 M KCl, and 20 mM Tris-HCl (pH 8), while the SFR-B solution contained 3% PVP 360,000, 1.2 M NaCl, and 20 mM Tris-HCl (pH 8) (Jones et al. 2021). Sixty µl of the *C. maenas* genomic DNA sample was mixed with 60 µl of either SFR-A or SFR-B solution, and the mixture was homogenized by gently tapping the tube. For SFR-B, the mixture was first incubated for 1 h at 50°C. Both SFR-A and SFR-B samples were then centrifuged at 10,000 × g for 30 min at room temperature. The supernatant was carefully removed, and 200 µl of cold 70% ethanol was added, followed by centrifugation at 10,000 × g for 2 min at room temperature. This ethanol-washing step was repeated three times. The pellet was subsequently dried for 10 min at 50°C and resuspended in 60 µl of pre-warmed (37°C) TE buffer for 20 min at 37°C. DNA quality and quantity were assessed as described above.

As the extraction/purification combinations did not resolve the pore blockage, whole-genome purification was performed as a final attempt.

WGA

To address persistent pore blockage in the MinION flow cell during sequencing, several approaches were tested to remove potential contaminants, including alternative DNA extraction methods, incorporation of an InhibitEX step, and SFR (as previously described). Despite these efforts, the blockage issue persisted. WGA was therefore performed on the modified G/20 tip DNA extraction + SFR-A sample using the REPLI-g Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions, in an effort to eliminate the

source of the blockage. DNA concentrations and purity ratios were again assessed using a Qubit Flex Fluorometer and a NanoDrop spectrophotometer.

Library preparation and MinION sequencing

Genomic DNA libraries for *C. maenas* were constructed from DNA using the aforementioned extraction/purification combinations and WGA approach (see [Supplementary Table 1](#) for details), with library preparation performed using the SQK-LSK114 Ligation Sequencing Kit V14 (ONT). Library preparation started with 1 µg of *C. maenas* DNA, quantified using a Qubit Flex Fluorometer. Following the “Ligation Sequencing gDNA V14” protocol provided by ONT, the process included a DNA repair and end-prep, adapter ligation, and clean-up step. DNA concentrations were measured after each step using the Qubit Flex Fluorometer. The sequencing flow cell (R10.4.1, FLO-MIN114, ONT) was primed and loaded with 20 to 60 ng of *C. maenas* library per run ([Supplementary Table 1](#)). Sequencing was performed on a MinION device (ONT), basecalling and demultiplexing were performed with MinKNOW/Dorado at “super-accurate basecalling, 400 bps”.

Pore blockage and sequencing optimization

Both the modified G/20 tip and Powersoil Pro DNA extraction methods, combined with either SFR-A or SFR-B or the WGA sample (modified G/20 tip DNA extraction + SFR-A kit), produced sufficient DNA for sequencing. The SFR kits (SFR-A and SFR-B) effectively reduced the number of small reads in the samples. However, none of the tested approaches, including WGA, were able to prevent pore blockage, which consistently occurred approximately 2 h into sequencing.

Interestingly, a pore scan revealed that some pores reopened after blockage, and a substantial increase in active pores was observed following the use of the Flow Cell Wash Kit (ONT), which contains DNase. This suggests that pore blockage may not be caused by contaminants, but by structural properties of the DNA itself. To mitigate blockage, we implemented a sequencing protocol that included a pore scan every 30 min during the run and a flow cell wash every 2 h ([Supplementary Table 1](#)).

Genome assembly and curation

The read data was assembled with Flye v2.9.2 (Kolmogorov et al. 2019) specifying high-quality Nanopore reads as input (“-nano-hq”), and the assembly was polished with medaka v1.11.3 (<https://github.com/nanoporetech/medaka>).

Due to difficulties in the sequencing process and the resulting limited amount of sequencing data, we generated two initial assemblies. The first assembly used only the data from the Lauwersoog specimen, while the second combined data from this specimen with an additional 10% of reads from the Schiermonnikoog specimen obtained during the test phase. Although combining data from multiple specimens is not generally recommended, since it can introduce small sequence and structural variations that may negatively impact assembly quality, comparison of the statistics of the two assemblies indicated that the combined assembly had better contiguity and completeness. Based on this, we chose to proceed with the combined assembly for all subsequent analyses.

Because a complete mitochondrial genome was not detectable in our draft assembly, we used a targeted approach to assemble it. Mitochondria of crabs have a lower GC content

than the nuclear genome, and mitochondrial DNA is present in high abundance in muscle tissue. We therefore extracted reads from our data with a GC content below 35% and a minimum length of 10 kbp and generated an assembly from a 10% subsample of the resulting reads to adjust the coverage to a level suited for assembly. Using mitochondrial-encoded cytochrome c oxidase subunit I protein sequences obtained from Interpro (IPR000883) as a marker (Blum et al. 2025), we identified a mitochondrial candidate contig using diamond blastx v2.19 (Buchfink et al. 2021) in the target assembly. Whole-genome self-alignment of the candidate sequence revealed that its palindromic structure was an artifact of the assembly process. We pruned the duplicated region to obtain the final mitochondrial sequence. Using this curated sequence as a reference, we removed remaining mitochondrial fragments from the whole genome draft based on minimap2 v2.28 (Li 2018) mappings before adding the curated mitochondrial sequence to the whole genome assembly.

To optimize our assembly in terms of contiguity, we used homology-based reference scaffolding to organize smaller contigs into larger scaffolds. Specifically, we used the homology scaffolder RagTag (Alonge et al. 2022), leveraging the highly contiguous genome assembly of the mud crab *Scylla paramamosain* (Estampador 1950 as reference) (RefSeq accession: GCF_035594125.1).

Assembly statistics and quality assessment

To assess potential contamination of the assembly, we employed two complementary approaches. First, we ran the deep learning classifier DeepMicroClass v1.0.3 (Hou et al. 2024) for a broad classification of sequences into eukaryotes, prokaryotes, and viruses. Second, we used MMSeg2 homology-based “easy-taxonomy” workflow (Mirdita et al. 2021) with UniRef50 (The UniProt Consortium 2025) as database to assign each contig a last common ancestor classification based on sequence homology.

To estimate assembly completeness, we used Compleasm v0.2.6 (Huang and Li 2023) with BUSCO (Tegenfeldt et al. 2025) marker gene sets and automatic lineage detection. To identify low-complexity regions within the sequences, we employed tantan v26 (Frith 2011). Additionally, we gathered basic assembly statistics—such as sequence length and base composition—using SeqKit (Shen et al. 2024).

Results

To obtain high-quality, high-molecular-weight DNA suitable for long-read sequencing of brachyuran crabs, we evaluated two extraction protocols: the Modified Qiagen Genomic-tip 20/G (G/20) method and the DNeasy Powersoil Pro Kit. Each was tested in combination with either of two SFR kits (SFR-A or SFR-B). Despite yielding high-quality DNA in all cases, none of these extraction/purification combinations prevented Nanopore sequencing pore blockage, which consistently occurred approximately 2 h after run initiation. To address this, we applied WGA to a *C. maenas* sample extracted with the Modified G/20 method and SFR-A. However, this approach also failed to mitigate the pore blockage issue. While pore blockage remained a limiting factor, all extraction and purification strategies tested (including WGA) produced DNA of sufficient quality for successful sequencing. Each method generated high-quality reads, indicating that these protocols are suitable for sequencing brachyuran crabs

with long-read platforms. To reduce pore loss and improve data yield, we incorporated periodic pore scans (every 30 min) and applied a flow cell wash after each run. This approach partially restored pore activity, enhancing overall sequencing throughput.

Based on existing measurements of *C. maenas* cellular DNA contents (1.24 pg and 1.07 pg; Gregory 2025), we estimated its genome size to be approximately 1.13 Gbp. We successfully generated 11.1 Gbp of *C. maenas* DNA sequence data (Lauwersoog specimen) using five MinION flow cells, achieving an average read N50 of 6.11 kbp. This dataset was supplemented with an additional 1.2 Gbp of data generated from the initial test sequencing runs on the Schiermonnikoog specimen (see Supplementary Table 1 for details). The whole genome of *C. maenas* was assembled de novo and polished, generating an initial contig assembly of 965 Mbp, comprising 34,569 contigs and an NG50 of 54 kbp. We further scaffolded our draft contigs based on homology to the mud crab (*S. paramamosain*) and curated the resulting scaffolds to ensure high quality (see Material and Methods). We obtained a final genome assembly of 1.09 Gbp in size, comprising 21,887 scaffolds and an NG50 of 13 Mbp (i.e. more than 50% of the genome is contained in scaffolds larger than 13 Mbp).

A closer examination of the number and size distribution of the assembled sequences reveals a bimodal pattern (Fig. 1a). Approximately 80% of the genome is represented by chromosome-scale scaffolds, while the remainder consists predominantly of shorter sequences ranging from 1 to 100 kbp. This distribution is consistent with our reference-based scaffolding approach, where some contigs were either unplaced or placed ambiguously due to limited or conflicting homology information.

Assessment of contamination via two complementary approaches indicated negligible levels of sequences predicted to derive from sources other than the genuine *C. maenas* genome (Fig. 1b, Supplementary Table 2). Using broad deep-learning-based classification, scaffolds larger than 1 Mbp were exclusively classified as eukaryotic, while three-fourths of smaller scaffolds were classified as eukaryotic and one-fourth as eukaryotic viruses. In the majority, the latter sequences represent retrotransposons that are expected to be abundant in this genome (Verbruggen 2016). Consistently, homology classification placed 89% of sequences with last common ancestors along the Brachyura lineage, with less than 2% of sequences potentially assigned to other kingdoms (1.8% fungi, <1% plants, bacteria and viruses). Completeness estimates based on lineage-specific universal eukaryotic marker genes indicate a highly complete genome assembly, with 98.4% of markers detected (Fig. 1c). Most marker genes were assembled at full length and located within large scaffolds. Additionally, the low level of duplicated markers supports the assembly’s high contiguity and suggests minimal structural contamination.

Finally, assessment of the sequence complexity revealed expected high levels of sequences with repetitive nature among the unscaffolded short contigs (Fig. 1d). The presence of repetitive regions in these contigs is consistent with both the classification of a significant fraction of these contigs as putative retrotransposons and the apparent difficulty of scaffolding these contigs unambiguously into a longer scaffold context, thereby contributing to the bimodal sequence length distribution observed.

While our initial draft assembly did not contain a detectable mitochondrial genome, we recovered a single

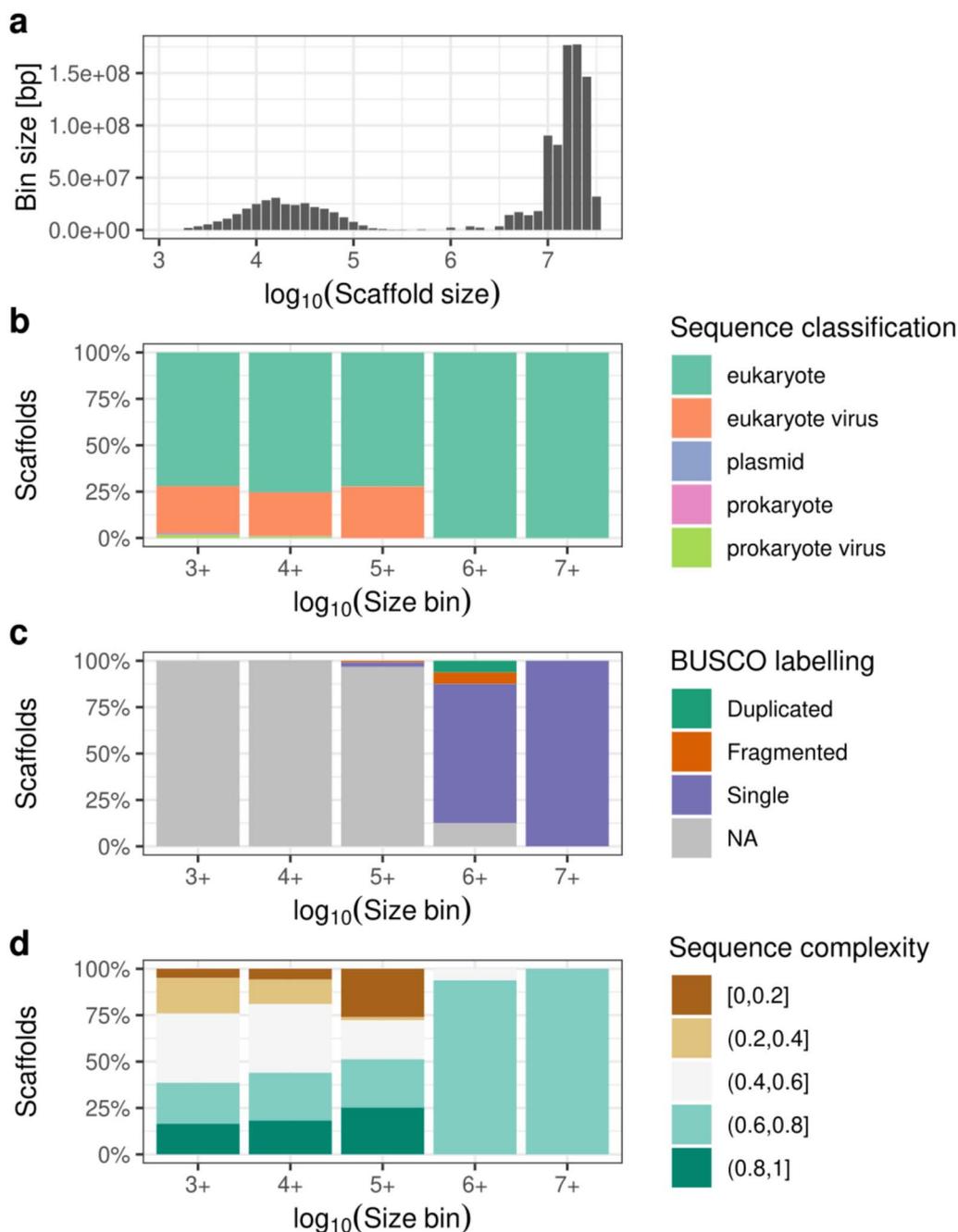


Fig. 1. Overview of assembly composition and quality metrics. (a) Size distribution histogram of assembled contigs and scaffolds. (b) Broad classification profiles across contigs/scaffolds of different length bins (minimum sizes: 1 kb, 10 kb, 100 kb, 1 Mb, 10 Mb), with colors indicating the predicted source. (c) Distribution of universal eukaryotic marker genes (BUSCOs) across the same contigs/scaffolds length bins with colors indicating the fraction of sequences identified as low complexity by the tantan algorithm.

mitochondrial contig of 15,474 bp using a targeted assembly approach. The obtained mitochondrial genome matches those of three brachyuran crabs (*Scylla paramamosain*—JAHFWG010003966.1; *Callinectes sapidus* Rathbun, 1896—NC_012572.1, *Portunus trituberculatus* (Miers, 1876)—NC_005037.1) in size and divergence and appears syntenic in its overall genomic organization (Supplementary Fig. 3).

Discussion

Here we present a high-quality genome assembly of the green shore crab, *C. maenas*, with a total size of approximately

1.09 Gbp. The assembly is highly complete, containing 98.4% of universal eukaryotic markers, and achieves high contiguity through homology-based scaffolding. This genome fills an important gap within the Brachyura, bringing the total number of publicly available true crab genomes to 11.

Despite its ecological significance as an invasive species and its widespread use as a model organism across multiple research fields, particularly ecotoxicology (Leignel et al. 2014; Rodrigues and Pardal 2014), a complete genome for *C. maenas* was not previously available. Earlier sequencing efforts in 2016 produced a fragmented draft assembly covering only 36% of the estimated genome size, hindered primarily by

the genome's high repetitive content (Verbruggen 2016). Our assembly thus provides a valuable foundation for future studies on the genomic basis of invasion success and supports the species' utility as a model organism in other lines of research.

Our *C. maenas* specimens were sequenced using an Oxford Nanopore R10.4.1 platform, which offers high-accuracy long reads suitable for de novo genome assembly (Guiglielmoni and Schiffer 2024). We faced recurring pore blockage after approximately 2 h of sequencing. Initially, we suspected that pore blockage was caused by contaminants. To address this, we tested two preservation fluids, optimized DNA extraction protocols, incorporated an InhibitEX step, applied SFR kits and WGA. Despite these efforts, pore blockage persisted, leading us to conclude that secondary structures in the DNA were likely responsible. To maximize data yield and minimize pore blockage, periodic pore scans were performed at 30-min intervals during sequencing. After each run (generally 2 h), a flow cell wash was applied before running the next sample, which helped reopen some of the blocked pores. Similar challenges have been reported in other decapods, such as the tiger prawn (*Penaeus monodon*; Fabricius 1798). The authors also attributed these blockages to the formation of complex secondary structures during sequencing and considered them irreversible (Van Quyen et al. 2020).

To facilitate genomic research in brachyuran crabs and other non-model marine invertebrates, we provide a detailed protocol for long-read sequencing using Oxford Nanopore's MinION platform, from DNA extraction to strategies bypassing pore blockage. This resource highlights the importance of methodological optimization in generating high-quality genomic data for ecologically and evolutionarily important marine invertebrate taxa.

Acknowledgments

We thank Britas Klemens Eriksson, Jorn Claassen, Yichen Liu, Lea Simon (all University of Groningen), Bastian Reijnen, and two little girls for providing us with *C. maenas* specimens for (preliminary) genetic analyses and testing of sequencing protocols. We furthermore thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high-performance computing cluster. We thank an anonymous reviewer for their constructive comments and suggestions on an earlier version of our manuscript.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

None declared.

Data availability

Raw sequence data and the assembled, curated genome presented here have been deposited at the European Nucleotide Archive under the project accession PRJEB86588.

Author contributions

Jolanda K Brons (Data curation, Investigation, Methodology, Project administration, Validation, Visualization, Writing—original draft), Thomas Hackl (Conceptualization, Data curation, Formal analysis, Validation, Visualization, Writing—original draft), Riccardo Iacovelli

(Data curation, Investigation, Methodology, Writing—review & editing), Kristina Haslinger (Resources, Writing—review & editing), Sebastian Lequime (Conceptualization, Resources, Writing—review & editing), and Sancia ET van der Meij (Conceptualization, Project administration, Resources, Visualization, Writing—original draft)

References

- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022;23:258. <https://doi.org/10.1186/s13059-022-02823-7>
- Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, Pinto BL, Orr A, Paysan-Lafosse T, Ponamareva I, et al. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* 2025;53:D444–D456. <https://doi.org/10.1093/nar/gkae1082>
- Boughattas S, Albatesh D, Al-Khater A, Giraldez BW, Althani AA, Benslimane FM. Whole genome sequencing of marine organisms by Oxford nanopore technologies: assessment and optimization of HMW-DNA extraction protocols. *Ecol Evol.* 2021;11:18505–18513. <https://doi.org/10.1002/ece3.8447>
- Bourne, SD, Hudson, J, Holman, LE, Rius, M (2018). Marine invasion genomics: Revealing ecological and evolutionary consequences of biological invasions. In M. Oleksiak, O. Rajora (Eds.), *Population genomics: marine organisms* (pp. 363–398). Springer, Cham. https://doi.org/10.1007/13836_2018_21
- Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods.* 2021;18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Carlton JT, Cohen AN. Episodic global dispersal in shallow water marine organisms: the case history of the European shore crabs *Carcinus maenas* and *Aestuarii*. *J Biogeogr.* 2003;30:1809–1820. <https://doi.org/10.1111/j.1365-2699.2003.00962.x>
- Colautti RI, Ricciardi A, Grigorovich IA, MacIsaac HJ. Is invasion success explained by the enemy release hypothesis? *Ecol Lett.* 2004;7:721–733. <https://doi.org/10.1111/j.1461-0248.2004.00616.x>
- Cui Z, Liu Y, Yuan J, Zhang X, Ventura T, Ma KY, Sun S, Song C, Zhan D, Yang Y, et al. The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. *Nat Commun.* 2021;12:2395. <https://doi.org/10.1038/s41467-021-22604-3>
- Daniels BN, Nurge J, Sleeper O, Lee A, López C, Christie MR, Toonen RJ, White C, Davidson JM. Genomic DNA extraction optimization and validation for genome sequencing using the marine gastropod *Kelletia kelletii*. *PeerJ.* 2023;11:e16510. <https://doi.org/10.7717/peerj.16510>
- Davidson AM, Jennions M, Nicotra AB. Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecol Lett.* 2011;14:419–431. <https://doi.org/10.1111/j.1461-0248.2011.01596.x>
- De Grave S, Decock W, Dekeyser S, Davie PJF, Franssen CHJM, Boyko CB, Poore GCB, Macpherson E, Ah Yong ST, Crandall KA, et al. Benchmarking global biodiversity of decapod crustaceans (crustacea: Decapoda). *J Crustac Biol.* 2023;43:uad042. <https://doi.org/10.1093/jcbiol/ruad042>
- Ens NJ, Harvey B, Davies MM, Thomson HM, Meyers KJ, Yakimishyn J, Lee LC, McCord ME, Gerwing TG. The green wave: reviewing the environmental impacts of the invasive European green crab (*Carcinus maenas*) and potential management approaches. *Environ Rev.* 2022;30:306–322. <https://doi.org/10.1139/er-2021-0059>
- Frederich, M, Lancaster, ER (2024). The European green crab, *Carcinus maenas*: Where did they come from and why are they here? In D Wehrauch, IJ McGaw (Eds.), *Ecophysiology of the European green crab (Carcinus maenas) and related species* (pp. 1–20). London (UK): Academic Press. <https://doi.org/10.1016/B978-0-323-99694-5.00002-7>
- Frith MC. A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res.* 2011;39:e23. <https://doi.org/10.1093/nar/gkq1212>

- GBIF. *Carcinus maenas* (Linnaeus, 1758) global occurrence in GBIF. Global biodiversity information facility. 2025: <https://www.gbif.org/species/5178595> [Accessed 15 May 2025].
- Gregory TR. Animal genome size database. 2025; <http://www.genome-size.com>
- Guiglielmoni N, Schiffer PH. Phasing or purging: tackling the genome assembly of a highly heterozygous animal species in the era of high-accuracy long reads. 2024; [bioRxiv:2024.06.16.599187](https://doi.org/10.1101/2024.06.16.599187). <https://doi.org/10.1101/2024.06.16.599187>
- Han GD, Ma DD, Du LN, Zhao ZJ. Chromosomal-scale genome assembly of the Mediterranean mussel *Mytilus galloprovincialis*. *Scientific Data*. 2024;11:644. <https://doi.org/10.1038/s41597-024-03497-5>
- Hou S, Tang T, Cheng S, Liu Y, Xia T, Chen T, Fuhrman JA, Sun F. DeepMicroClass sorts metagenomic contigs into prokaryotes, eukaryotes and viruses. *NAR Genomics and Bioinformatics*. 2024;6:lqae044. <https://doi.org/10.1093/nargab/lqae044>
- Huang N, Li H. Compleasm: a faster and more accurate reimplement of BUSCO. *Bioinformatics*. 2023;39:brad595. <https://doi.org/10.1093/bioinformatics/btad595>
- Jaspers C, Ehrlich M, Pujolar JM, Künzel S, Bayer T, Limborg MT, Lombard F, Browne WE, Stefanova K, Reusch TB. Invasion genomics uncover contrasting scenarios of genetic diversity in a widespread marine invader. *Proc Natl Acad Sci*. 2021;118:e2116211118. <https://doi.org/10.1073/pnas.2116211118>
- Jeffery NW, DiBacco C, Wringe BF, Stanley RRE, Hamilton LC, Ravindran PN, Bradbury IR. Genomic evidence of hybridization between two independent invasions of European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Heredity*. 2017;119:154–165. <https://doi.org/10.1038/hdy.2017.22>
- Jones A, Torkel C, Stanley D, Nasim J, Borevitz J, Schwesinger B. High-molecular-weight DNA extraction, clean-up, and size selection for long-read sequencing. *PLoS One*. 2021;16:e0253830. <https://doi.org/10.1371/journal.pone.0253830>
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Kolodziejczyk J, Fijarczyk A, Porth I, Robakowski P, Vella N, Vella A, Kloch A, Biedrzycka A. Genomic investigations of successful invasions: the picture emerging from recent studies. *Biological Reviews, Advance online publication*. 2025;100:1396–1418. <https://doi.org/10.1111/brv.70005>
- de Lannoy C, de Ridder D, Risse J. The long reads ahead: *De novo* genome assembly using the MinION. *F1000Research*. 2017;6:1083. <https://doi.org/10.12688/f1000research.12012.2>
- Lee CE. Evolutionary genetics of invasive species. *Trends Ecol Evol*. 2002;17:386–391. [https://doi.org/10.1016/S0169-5347\(02\)02554-5](https://doi.org/10.1016/S0169-5347(02)02554-5)
- Leignel VSJH, Stillman JH, Baringou S, Thabet R, Metais I. Overview on the European green crab *Carcinus* spp. (Portunidae, Decapoda), one of the most famous marine invaders and ecotoxicological models. *Environ Sci Pollut Res*. 2014;21:9129–9144. <https://doi.org/10.1007/s11356-014-2979-4>
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Lowe S, Browne M, Boudjelas S, De Poorter M. *100 of the world's worst invasive alien species: a selection from the global invasive species database*. Auckland (New Zealand): Hollands Printing Ltd.; 2000.
- Lu Y, Luo F, Zhou A, Yi C, Chen H, Li J, Guo Y, Xie Y, Zhang W, Lin D, et al. Whole-genome sequencing of the invasive golden apple snail *Pomacea canaliculata* from Asia reveals rapid expansion and adaptive evolution. *GigaScience*. 2024;13:giae064. <https://doi.org/10.1093/gigascience/giae064>
- McCartney MA, Auch B, Kono T, Mallez S, Zhang Y, Obille A, Becker A, Abraham JE, Garbe J, Badalamenti JP, et al. The genome of the zebra mussel, *Dreissena polymorpha*: a resource for comparative genomics, invasion genetics, and biocontrol. *G3*. 2022;12:jkab423. <https://doi.org/10.1093/g3journal/jkab423>
- Mirdita M, Steinegger M, Breitwieser F, Söding J, Levy Karin E. Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics*. 2021;37:3029–3031. <https://doi.org/10.1093/bioinformatics/btab184>
- Molnar JL, Gamboa RL, Revenga C, Spalding MD. Assessing the global threat of invasive species to marine biodiversity. *Front Ecol Environ*. 2008;6:485–492. <https://doi.org/10.1890/070064>
- NCBI. *Genome portal*. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2025: <https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=6752> [retrieved 2025 May 15].
- North HL, McLaughran A, Jiggins CD. Insights into invasive species from whole-genome resequencing. *Mol Ecol*. 2021;30:6289–6308. <https://doi.org/10.1111/mec.15999>
- Oosting T, Hilario E, Wellenreuther M, Ritchie PA. DNA degradation in fish: practical solutions and guidelines to improve DNA preservation for genomic research. *Ecol Evol*. 2020;10:8643–8651. <https://doi.org/10.1002/ece3.6558>
- Rodrigues ET, Pardal MÁ. The crab *Carcinus maenas* as a suitable experimental model in ecotoxicology. *Environ Int*. 2014;70:158–182. <https://doi.org/10.1016/j.envint.2014.05.018>
- Rutz C, Bonassin L, Kress A, Francesconi C, Boštjančić LL, Merlat D, Theissinger K, Lecompte O. Abundance and diversification of repetitive elements in Decapoda genomes. *Gene*. 2023;14:1627. <https://doi.org/10.3390/genes14081627>
- Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Smith SA, et al. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*. 2013;342:1242592. <https://doi.org/10.1126/science.1242592>
- Schlaepfer MA. Do non-native species contribute to biodiversity? *PLoS Biol*. 2018;16:e2005568. <https://doi.org/10.1371/journal.pbio.2005568>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, et al. No saturation in the accumulation of alien species worldwide. *Nat Commun*. 2017;8:14435. <https://doi.org/10.1038/ncomms14435>
- Simberloff D, Martin JL, Genovesi P, Maris V, Wardle DA, Aronson J, Courchamp F, Galil B, García-Berthou E, Pascal M, et al. Impacts of biological invasions: what's what and the way forward. *Trends Ecol Evol*. 2013;28:58–66. <https://doi.org/10.1016/j.tree.2012.07.013>
- Shen W, Sipos B, Zhao L. SeqKit2: a Swiss army knife for sequence and alignment processing. *iMeta*. 2024;3:e191. <https://doi.org/10.1002/imt2.191>
- Technologies ON. *Medaka*. n.d.: <https://github.com/nanoporetech/medaka> [Accessed May 19, 2025].
- Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res*. 2025;53:D516–D522. <https://doi.org/10.1093/nar/gkae987>
- Tepolt CK. Adaptation in marine invasion: a genetic perspective. *Biol Invasions*. 2015;17:887–903. <https://doi.org/10.1007/s10530-014-0825-8>
- The UniProt Consortium. UniProt: the universal protein knowledgebase in 2025. *Nucleic Acids Res*. 2025;53:D609–D617. <https://doi.org/10.1093/nar/gkae1010>
- Van Quyen D, Gan HM, Lee YP, Nguyen DD, Nguyen TH, Tran XT, Khang DD, Austin CM. Improved genomic resources for the black tiger prawn (*Penaeus monodon*). *Mar Genomics*. 2020;52:100751. <https://doi.org/10.1016/j.margen.2020.100751>
- Verbruggen B. *Generating genomic resources for two crustacean species and their application to the study of white spot disease (doctoral dissertation, University of Exeter)*. 2016: <https://ore.exeter.ac.uk/repository/handle/10871/25535>.
- Wang Y, Wang Y, Yang Y, Ni G, Li Y, Chen M. Chromosome-level genome assembly of the northern Pacific seastar *Asterias amurensis*. *Scientific Data*. 2023;10:767. <https://doi.org/10.1038/s41597-023-02688-w>

- Wang Z, Tang B, Li Y, Wang G, Liu H, Liu Q, Sun Y, Lu X, Han G, Zhang H, et al. Phylogenomics of crabs provides insights into their origin and evolution (version 1) [preprint]. *Research Square*. 2025. <https://doi.org/10.21203/rs.3.rs-5612812/v1>
- Weir SM, Salice CJ. Managing the risk of invasive species: how well do functional traits determine invasion strategy and success? *Integr Environ Assess Manag*. 2011;7:299–300. <https://doi.org/10.1002/ieam.171>
- Xu Z, Gao T, Xu Y, Li X, Li J, Lin H, Yan W, Pan J, Tang J. A chromosome-level reference genome of red swamp crayfish *Procambarus clarkii* provides insights into the gene families regarding growth or development in crustaceans. *Genomics*. 2021;113:3274–3284. <https://doi.org/10.1016/j.ygeno.2021.07.017>
- Yuan J, Yu Y, Zhang X, Li S, Xiang J, Li F. Recent advances in crustacean genomics and their potential application in aquaculture. *Rev Aquac*. 2023;15:1501–1521. <https://doi.org/10.1111/raq.12791>