Degree project

UNIVERSITY OF SKÖVDE

DETECTION OF CHINESE MITTEN CRAB IN SWEDISH WATER WITH MOLECULAR METHODS

Master Degree Project (60 credits) in Molecular Ecology

Second Cycle 30 credits

Spring semester 2025

Student: Felix Östervald

Supervisor: Sonja Leidenberger

John Baxter

Examiner: Magnus Fagerlind

Abstract

The invasive Chinese Mitten Crab, *Eriocheir sinensis*, poses an increasing ecological risk in Swedish waters, prompting the need for effective detection methods. This study aimed to detect *E. sinensis* using environmental DNA extracted from controlled aquarium setups and field sediment samples from Lake Vänern, employing both qPCR and Nanopore sequencing. Effectiveness of water and sediment as matrices for recovery and detection of *E. sinensis* eDNA was evaluated. Aquarium experiments revealed unpredictable variability in DNA concentrations due to complex DNA dynamics likely influenced by both biotic and abiotic factors. Water samples consistently showed higher DNA concentrations, potentially influenced by the sediment type and the agitation performed before sampling, and better qPCR detection rates compared to sediment samples. Although sediment yielded less DNA here, it remains promising as a complementary matrix because DNA can adsorb to particles and persist, potentially integrating signals over longer periods; with optimized extraction and inhibitor mitigation, sediment may improve site-level detectability. Although Nanopore sequencing demonstrated substantial potential for eDNA analysis due to its portability and PCR-free approach, limitations were observed, including low data output, classification errors, and dependency on high-quality, comprehensive reference databases. The qPCR approach emerged as a more reliable, practical, and cost-effective method for targeted *E. sinensis* detection. Field sampling, performed in known hotspots during the low-activity spring season did not yield any definitive detections with either method. Future research should focus on optimizing sample processing, improving reference databases, and exploring more sensitive and specific molecular detection technologies.

Popular Scientific Summary

The Chinese mitten crab, *Eriocheir sinensis*, is an invasive species that poses a growing threat to Swedish freshwater ecosystems. Originally from East Asia, it has spread to many parts of Europe via ship ballast water. The species is catadromous, growing in fresh water but reproducing in brackish or marine habitats, and can move long distances, even overland. In March 2025, Swedish authorities raised its invasion risk from "high" to "very high" due to its adaptability and potential to harm native species. Once established, mitten crabs can destabilise riverbanks through burrowing, compete with native fauna, and spread diseases such as crayfish plague to vulnerable species like the European crayfish, *Astacus astacus*.

In Sweden, mitten crabs have so far only been found as single adults, mainly in Lake Vänern and its connected waterways. No signs of successful reproduction have been recorded, but climate change and potential hybridisation with the Japanese mitten crab, *Eriocheir japonica*, may change this. Such hybridisation can cause introgression (gene flow), potentially spreading traits like cold tolerance into Swedish *E. sinensis* populations. Early detection is therefore essential to prevent a permanent population.

One modern approach to detecting invasive species is environmental DNA (eDNA) analysis. All organisms shed DNA into their surroundings through skin cells, waste, or decomposition. By collecting and analysing samples from water, sediment, soil, or even air, scientists can detect the presence of species without ever seeing or catching them. This study tested two molecular methods, quantitative PCR (qPCR) and Oxford Nanopore sequencing, to detect mitten crab DNA from water and sediment samples collected both in the lab and in the field. In simple terms, qPCR is a targeted "DNA detector" that copies a short, species-specific region and reports a fluorescent signal if the target is present. Nanopore sequencing reads much longer DNA fragments directly and can identify many species at once using a handheld device, potentially even outside the lab.

In the laboratory, two aquarium experiments were set up with mitten crab tissue placed in tanks containing either natural pond sediment or clean commercial sand. Water and sediment samples were collected at different time points to compare which sample type gave higher DNA yields and better detection rates. In the field, sediment samples were taken from two locations in Kinneviken, an area of Lake Vänern where mitten crabs have been spotted for decades.

The results showed that water samples generally contained more total DNA than sediment samples, and qPCR detected mitten crab DNA more consistently in water. However, sediment can still be a valuable source of eDNA because it may preserve DNA for longer periods, acting as a "local archive," though co-extracted inhibitors can reduce PCR sensitivity. In both sample types, qPCR proved to be more sensitive and reliable than Nanopore sequencing for detecting mitten crab DNA under the conditions of this study.

Nanopore sequencing has exciting potential: it can analyse DNA directly in the field, without laboratory PCR amplification, and can survey many species simultaneously. In this study, however, performance was limited by low data yield, classification errors, and gaps in genetic reference databases for mitten crab and related species. Adding high-quality *Eriocheir* references and refining workflows could improve results.

Field sampling did not detect mitten crab DNA at either location, possibly because the crabs are less active during early spring and occupy deeper water than where samples were taken. Future monitoring should consider season and sampling depth and coordinate with reports from fishers and public observations to focus effort on likely hotspots.

Overall, this study found that qPCR remains the more practical and cost-effective method for targeted detection of mitten crab DNA in Swedish waters. Nanopore sequencing offers greater versatility and potential for broader biodiversity surveys, but further development is needed before it can match the sensitivity of qPCR for detecting low-abundance invasive species.

Early detection remains the key to preventing mitten crab establishment in Sweden. As molecular technologies improve, combining sensitive, targeted methods like qPCR with broad-range sequencing approaches could offer the most powerful strategy for protecting our freshwater ecosystems from invasive threats. Members of the public can contribute by reporting suspected mitten crab sightings through established Swedish biodiversity platforms, helping researchers and managers act quickly when and where it matters most.

Abbreviations

Annealing temperature – T_a

Chinese Mitten Crab – CMC

Cytochrome B Subunit – cyt b

Deoxyribonucleic acid – DNA

Environmental Deoxyribonucleic acid – eDNA

Genus-Wide – GW

NADH Dehydrogenase Subunit 4 – ND4

NADH Dehydrogenase Subunit 5 – ND5

Non-Template Control – NTC

Oxford Nanopore Technology – ONT

Polymerase Chain Reaction - PCR

Quantitative PCR – qPCR

Quantification Cycle – Cq

Species-Specific – SS

Uracil-DNA Glycosylase – UDG

Urokinase-type Plasminogen – UP

Whole-Genome – WG

Table of Contents

Introduction

The Chinese Mitten Crab (CMC) *Eriocheir sinensis* H. Milne-Edwards, 1853, is native to the coast and freshwater bodies of China and Korea (Bentley, 2011). It is a crab with a catadromous life cycle, meaning that it spends most of its life in freshwater, but reproduces in saline waters such as estuaries. When the larvae have developed to the juvenile stages, they start their journey back to the freshwaters their parents came from. This way of life has provided CMC with a high adaptability to salinity levels and the ability to migrate long distances overland (Bentley, 2011; Jung et al., 2022). The species arrived in Europe through the ballast water of ships, travelling between Europe and Asia and established an invasive population in the German Aller river in 1912 (Peters, 1933). From there the CMC spread to neighbouring countries, e.g., the Netherlands and Belgium, with a coastline towards the North Sea and English Channel during the late 1920s - 1930s (Bentley, 2011; Herborg et al., 2003; McDermott, 2011).

The species has so far not been able to establish stable populations further north into Sweden or Norway, which is likely due to challenges acclimatising to colder weather and the lower salinity levels (Anger, 1991). However, in March 2025, Sveriges lantbruksuniversitet (SLU) Artdatabanken raised the species' invasiveness risk category from "high" to "very high." This decision was based on its strong capacity for survival and spread, as well as its potential to negatively impact threatened and keystone species (SLU Artdatabanken, 2025a). The urgency of preventing CMC's establishment has increased further due to confirmed occurrences of the closely related Japanese mitten crab, *Eriocheir japonica* (De Haan, 1835) in Europe, which is capable of hybridizing with CMC (Homberger et al., 2022; Palero et al., 2022). *E. japonica* was also listed as a species with a very high risk of becoming invasive in Sweden, reinforcing the need for immediate preventative measures (SLU Artdatabanken, 2025b).

At the moment only adult individuals are caught in Swedish waters, no juveniles or larvae stages have been detected yet indicating that CMC have so far not managed to establish itself in Sweden (Leidenberger et al. 2021). The risk of CMC becoming invasive here is however increasing, due to both i) increasing global temperatures and ii) introgression of cold adaptive genes from the related *E. japonica* in populations in Belgium and Germany (Homberger et al., 2022). These two factors could allow CMC to overcome the colder Scandinavia climate. The area with the highest risk of getting an established invasive population of CMC is the inlet in Gothenburg, which offers a breeding ground with the appropriate marine salinity levels and offers the crab a suitable migration path upstream through the river Göta älv to Lake Vänern (Leidenberger et al., 2015). Lake Vänern, which is one of Europe's largest freshwater lakes, offers a huge habitat for the mature CMC, which also provides access to many rivers for the CMC to further migrate into. The habitat that lake Vänern offers is markedly different from both the native habitat of CMC as well as the areas where it has established invasive populations previously seeing as it is a lake of exceptional size compared to the rivers where it usually found. This contrast in habitat may complicate detection and removal efforts of the species from this area. The coastal areas toward the Baltic Sea do not offer the right salinity levels for reproduction and are therefore less likely as invasive hotspots (Anger, 1991; Homberger et al., 2022). The risk for CMC to adapt to suboptimal salinity conditions is however still present.

The first recorded finding of CMC in lake Vänern was in 1954 (Havs- och vattenmyndigheten, 2016). More findings were observed in the early 2000s and continued until 2010, when they seemingly decreased again. Specimens captured between 2003 and 2008 were sequenced and determined to originate from a population in the Elbe River in Germany (Czerniejewski et al., 2012). A study has since been performed by the University of Skövde to investigate the occurrence of CMC in lake Vänern (Leidenberger et al., 2021). This study discovered that another increase had occurred between 2018 and 2020 with a total of 16 observations, with half of them being caught in 2020 alone. The observations mainly took place during the summer and autumn, with observations being rare during spring and no-catches during winter. Majority of the observed specimens were caught in soft sediment bottom, consisting of sand, clay and mixed material. The specimens were mainly caught with Tangle net (bottensatta nät), at a depth between 4.5 down to 40 meters, and Fyke net (storryssja), at a depth between 0 and 10 meters. A mapping of the observations showed that most of the specimens were found on the east coast of lake Vänern, between Kristinehamn and Vänersborg, with a lot of observations being observed in Kinneviken (Bay of Kinne, Figure 1) (Leidenberger et al., 2021).

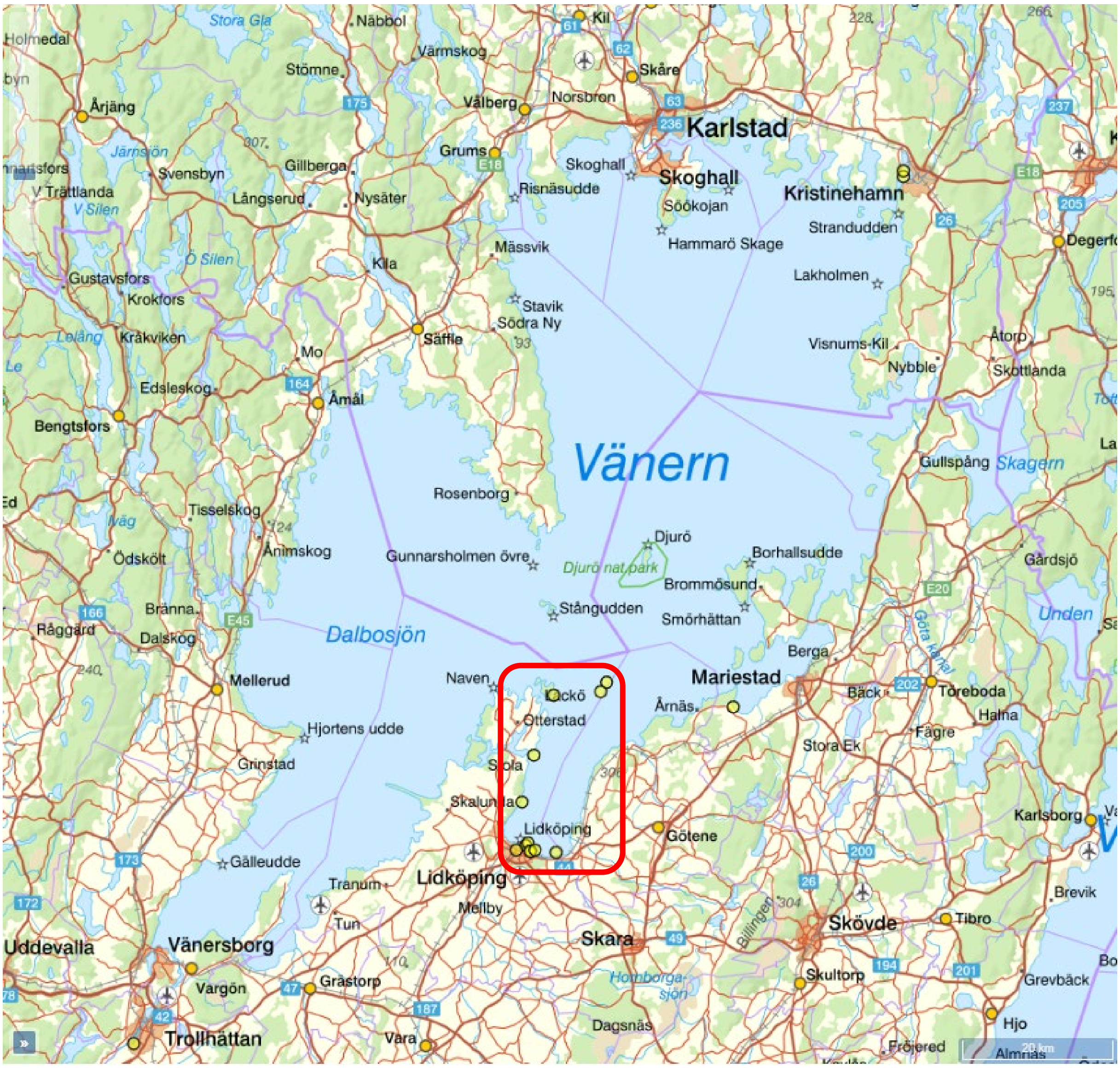

Figure 1. Dot map of reported findings of CMC in lake Vänern from 2010 to 2025. Area within the red square show the region of Kinneviken. Picture retrieved 21/05/2025 from https://www.artportalen.se/ViewSighting/ViewSightingAsMap.

The establishment of a stable CMC population in Swedish waters present several dangers, such as the fact that the species is omnivorous which gives it a favoured chance to find a niche in the ecosystem and outcompete native species (Panning, 1938). Another important behaviour of CMC that is worth considering is its habit of burrowing which could cause it to be a disruptive ecosystem engineer who can damage stream banks through erosion, further increasing its danger as an invasive species (Bentley, 2011; McDermott, 2011; Rato et al., 2024; Wang et al., 2024). As with most other species, CMC is a carrier of parasites, such as the lung fluke *Paragonimus westermani* (Kerbert 1878) Braun, 1899 which can infect humans. However, infection requires an intermediate host in the form of a snail of the genus *Semisulcospira* O. Boettger, 1886, but neither *P. westermani* nor members of the *Semisulcospira* genus are yet to be observed in Europe (Bentley, 2011; McDermott, 2011). If this will continue to be the case is still unsure, but what is known is that CMC can be affected with native parasites. For example, specimens caught in lake Vänern, have been shown to be infected with the fungi that causes the crayfish plague, *Aphanomyces astaci* Schikora, 1906. Since its introduction in Sweden, the plague has severely decimated the population of the native European crayfish, *Astacus astacus* (Linnaeus, 1758). The fungus is already a great threat to the vulnerable *A. astacus* populations, but CMC is a further risk factor that could spread the disease to isolated waterbodies through their highly migratory behaviour (Svoboda et al., 2014). Developing an early detection system for invasive species in general, but especially for CMC, is of outmost importance for the preservation of the ecosystems and the biodiversity of our large freshwater bodies.

The sampling and analysis of DNA (deoxyribonucleic acid) found in environmental samples, such as soil, sediment, water and air, is a developing technique, known as environmental DNA (eDNA), which has the potential to revolutionize biomonitoring and bioassessment (Pawlowski et al., 2020). Detecting crustaceans through eDNA has been shown to be less effective than for other animal groups, possibly due to the exoskeleton reducing the amount of eDNA shed into the environment (Pawlowski et al., 2020). Previous studies have however, shown that analysis of eDNA is a viable method for the detection of CMC (Chevrinais et al., 2023). Through the collection of eDNA out in field and analysing it through methods such as quantitative PCR (qPCR) or sequencing (e.g., Nanopore), the presence or absence of a species can be detected (Pawlowski et al., 2020). This can help to identify which bodies of water host populations of CMC and hence where actions plans are urgently needed to eradicate this aggressive invader. Previous generations of sequencing techniques can be used for this purpose as well, but the advantages of Oxford Nanopore Techniques (ONT) sequencing are the long-read lengths and the portability of the device (at least the MinION) allows on-site sequencing and immediate basecalling/demultiplexing on a laptop, and the avoidance of external sequencing services which decreases time until results are retrieved (Egeter et al., 2022; Pomerantz et al., 2018). ONT do currently have a higher error rate than the older and more commonly used sequencing technique Illumina (Stevens et al., 2023), but the method is continuing to be developed and becoming more reliable, through advancements with the flow cells, the sequencing kits and the basecalling algorithms as well as the development of error correcting tools (Chiou et al., 2023; Liu et al., 2024). This coupled with the previously mentioned advantages of ONT makes it a promising tool in the field of molecular ecology.

Genetic barcoding is a molecular technique that can be used to identify species through DNA samples, by utilizing specific regions of DNA that varies between species but remains relatively conserved within a species. Mitochondrial genes are often used as genetic barcodes since they occur in higher abundance than nuclear genes, they also don't contain any introns and have a high mutation rate, producing greater interspecific than intraspecific variation (the 'barcoding gap'), which aids species-level discrimination (Kress & Erickson, 2008; Luo et al., 2011). The protein coding gene Cytochrome Oxidase Subunit 1 (COI) is of the most commonly used genetic barcode for animals, but other mitochondrial protein coding genes, such as Cytochrome B Subunit (cyt b) and NADH Dehydrogenase Subunit 4 and 5 (ND4 and ND5), as well as the mitochondrial ribosomal RNA coding genes 12S and 16S have been shown to be suitable genetic barcodes for animals (Kress & Erickson, 2008; Li et al., 2024; Luo et al., 2011; Shen et al., 2013)

Previous studies have performed similar experiments, such as a study from Canada, which aimed to detect CMC as well as the European green crab, *Carcinus maenas* (Linnaeus, 1758), using eDNA from water samples collected in the wild and qPCR (Chevrinais et al., 2023). Here, the study was focused on CMC detection in Swedish waters, and used both, sediment and water samples from the hot-spot area Lake Vänern and samples from controlled environments, in the form of aquarium set-ups. Similarly, to Chevrinais et al (2023) qPCR was used to try to detect the presence of CMC under these different conditions, but it also used ONT to try to detect the presence of CMC from eDNA samples which has not previously been performed.

Aim and Objectives

Aim

The aim of this study is to attempt develop a reliable method to detect CMC in controlled environments, in the form of aquarium set-ups, and in the field using eDNA extracted from both water and sediment samples analysed with qPCR and ONT. This study will attempt to compare the efficiency of extracting CMC eDNA from water and sediment samples and to compare the efficiency of detection of CMC eDNA using qPCR and ONT.

Objectives

Study design and sampling

Set up two aquarium experiments with CMC tissue and collect time-series water and sediment samples; collect field sediment at two Kinneviken sites (Sannorna, Stenhammar).

eDNA sample types

Collect and extract eDNA from water samples (filters) and sediment samples from the aquarium set-ups as well as sediment samples from the field. Compare the two sample types for their ability to detect CMC.

Primer verification

Verify the species-specific (SS) and genus-wide (GW) primer sets, using tissue extracted DNA, and document specificity before application on eDNA samples.

qPCR workflow

Run SYBR-based qPCR on aquarium and field extracts with appropriate controls; record detection frequency/consistency (Cq/melt-curve criteria).

ONT workflow

Sequence water and sediment libraries on MinION; perform basecalling/demultiplexing and downstream classification. (Two ONT runs, one per sample type.)

Reference libraries & classification

Build and apply custom Kraken2 databases (Whole-genome, COI-only and multi-marker/mitogenome "MitoGun") and evaluate parameter settings (confidence thresholds) for read classification.

Performance comparison

Compare water vs. sediment sample types and qPCR vs. ONT on detection outcomes and practical considerations (e.g., data yield, classification behaviour), using statistical analyses.

Materials and methods

The study consisted of aquarium experiments, field sediment sampling, eDNA extraction, and analyses with qPCR and Nanopore sequencing. An overview of the experimental workflow, from sampling to downstream analyses, is shown in Figure 2.

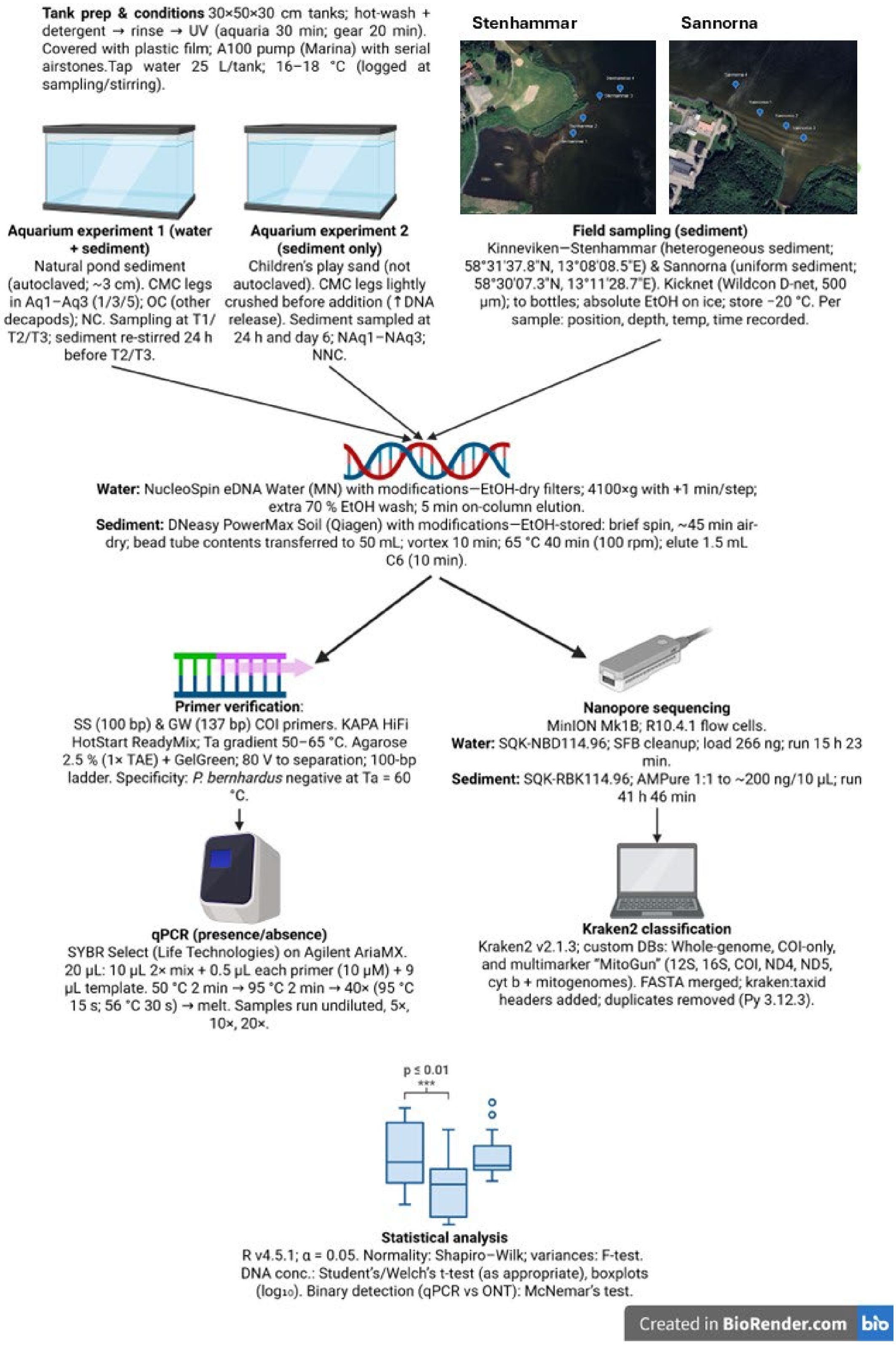

Figure 2. Workflow of the experiment and analyses.

1. Aquarium Experiment

Two aquarium experiments were conducted to evaluate detection of CMC from water and/or sediment (Figure 2). All tanks (30 × 50 × 30 cm) and equipment (airline, airstones, spoons) were washed with detergent, rinsed, and UV-irradiated (aquariums: 30 min; equipment: 20 min). Each aquarium was covered with plastic film to reduce airborne contamination, and an A100 air pump (Marina) with serially coupled airstones circulated water in all tanks. Water temperature during sampling and sediment-stirring ranged from 16–18 °C (Tables S1, S3).

Experiment 1 (10-19 March 2025, natural pond sediment)

Sediment (sand with some detritus) was collected from the nearby man-made pond Boulognersjön, autoclaved for 20 min at 121 °C, and added as a ~3 cm layer; tanks were then filled with 25 L tap water. Legs from frozen CMC were thawed, segmented at the joints, and added to Aq1–Aq3 in increasing numbers; the OC aquarium received legs from other decapods (OC) and NC served as a negative control (Table 1). Sediment and legs were mixed with sterilised metal spoons. Sediment was allowed to settle for ~2 h, before water and sediment samples were taken (10 Mar). Sediment was resuspended 24 h before the second and third sampling dates (14 and 19 Mar), after which sampling was repeated.

Experiment 2 (13-19 April 2025, commercial sand)

Tanks were prepared as above but filled with non-autoclaved children's play sand (Sandlådesand, Södra Århults Torv AB). CMC legs were crushed with a mortar and pestle before to increase DNA release. The set-up is shown in Table 2 (NAq1–NAq3 and a new negative control, NNC). Sediment was sampled at two time-points: 24 h and 6 days after legs were added.

Table 1. Set-up for aquarium experiments 1 and 2.

Aquarium 1	Species	Number of legs
Aq1	*E. sinensis*	1
Aq2	*E. sinensis*	3
Aq3	*E. sinensis*	5
OC	*Cancer pagurus*	2
	Carcinus maenas	1
	Polybius depurator	2
	Pandalus borealis	1
	Pagurus bernhardus	1
NC	Negative control	0
Aquarium 2		
NAq1	*E. sinensis*	1
Naq2	*E. sinensis*	3
Naq3	*E. sinensis*	5
NNC	Negative control	0

2. Field samples

Two sites, Stenhammar and Sannorna (Fig. 2), in Kinneviken (Lake Vänern) were sampled based on previous records and local fishermen. Four sediment samples each (03/04/2025) were collected with a kicknet (Wildcon D-net, 500 μm), transferred to bottles, topped up with absolute ethanol, and kept on ice until storage at −20 °C in the laboratory. For each sample position, sediment type, water depth, and temperature were noticed (Table S4). For habitat details see Figures 2 and S1-4.

3. eDNA extraction

3.1. Water samples

Water from the first aquarium experiment was collected with separate 1 L Erlenmeyer flasks. Samples were vacuum-filtered through 45 mm EO-treated glass-fiber filters mounted in a Nalgene Analytical Test Filter Funnel (0.45 µm, 250 mL, cat. 145-0020). Tweezers were soaked in 5% bleach for 10 min and rinsed with Milli-Q water before handling filters. Filtration used either a hand-driven Mityvac or a Welch vacuum pump. Due to highly particulate rich water, the filters were saturated between 250 ml and 350 ml. Filters were stored in absolute ethanol at −20 °C until extraction.

eDNA was extracted with the NucleoSpin eDNA Water kit (Macherey-Nagel, 740402.50) following the manufacturer's protocol. Work surfaces were cleaned with 5% bleach, then water, then 70% ethanol; filters were removed from ethanol and briefly air-dried on sterile Petri dishes; centrifugation was performed at 4,100 × g with one extra minute added to each step; an additional wash of 500 µL 70% ethanol was inserted between the kit's wash steps; and the elution incubation on the NucleoSpin eDNA XS column was extended from 1 to 5 min.

3.2. Sediment samples

Sediment was collected from the aquariums with metal spoons and transferred to 50 mL Falcon tubes. Samples from the first aquarium experiment were topped with absolute ethanol and stored at −20 °C; samples from the second experiment were stored directly at −20 °C.

eDNA was extracted with the DNeasy PowerMax Soil kit (Qiagen, 12988-10) following the manufacturer's protocol. For ethanol-stored samples, tubes were centrifuged at 4100 rpm for 6 min, the supernatant was discarded, and pellets were air-dried ~45 min to remove residual ethanol. The contents of the PowerBead tubes were then emptied into the 50 mL sample tubes (to avoid sediment loss during transfer). Tubes were sealed with Parafilm, vortexed 10 min at full speed, and incubated at 65 °C on a shaking platform (100 rpm, 40 min) to promote homogenisation and lysis, followed by 1 min vortex prior to centrifugation per kit instructions. Elution used 1.5 mL Solution C6 with a 10 min room-temperature incubation before the final centrifugation, to increase DNA yield.

The DNA quantity and purity of all samples were measured using Qubit 4 fluorometer, using the 1x HS dsDNA assay (Qiagen), and NanoDrop spectrophotometer (Thermo Fisher) respectively.

4. Primer verification

Two COI primer pairs were evaluated: a species-specific (SS) pair for CMC; forward: `CAGATATAGCTTTCCCACGAATAAACA-3´` and reverse: `CCAACTCCTCTTTCTACTATTCTTCTTGT-3´` (100 bp; J. E. Ironside, pers. comm.) and a genus-wide (GW) pair for Eriocheir; forward: `CATCAGTTGATCTTGGTATCTTYTCTCTACA-3´` and reverse: `TCAAACAAAAAGAGGTATTTGATCCATT-3´` (137 bp; Chevrinais et al., 2023).

To confirm amplification and establish annealing conditions, DNA was extracted from CMC tissue (Specimen 1; Table S2) with the DNeasy Blood & Tissue Kit (Qiagen, 69504) following the manufacturer's protocol, including overnight lysis on a thermomixer (100 rpm). ~12.5 mg leg tissue was homogenised; tools were decontaminated as described above. PCRs used KAPA HiFi HotStart ReadyMix (Roche, 07958927001). A gradient PCR was run to identify the optimal Ta (50 – 65 °C) (table 2-3). Products were quantified on a Qubit 4 fluorometer (1x HS dsDNA assay) and purity ratios read on a NanoDrop spectrophotometer; fragment size and the selected Ta were verified by agarose gel electrophoresis (2.5% agarose in 1x TAE, GelGreen (Biotium)) with a Quick-Load 100 bp DNA ladder (NEB); 5 µl PCR product was mixed with 1 µl 6x Gel Loading Dye, Purple (NEB) and run at 80 V until sufficient separation.

Table 2. Reaction set-up for gradient and specificity PCR, using the KAPA HiFi Hotstart Ready Mix Kit (Roche)

Reagents	Volume
2X KAPA HiFi Hotstart ReadyMix	10 µl
10 µm Forward primer	0.6 µl
10 µm Reverse primer	0.6 µl
Template	Variable (10 ng)
H_2O	Up to 20 µl
Total	20 µl

Table 3. Cycling protocol for PCR with KAPA HiFi Hotstart Ready Mix Kit (Roche)

Cycle Step	Temperature	Duration	Cycles
Initial Denaturation	95 °C	3 min	1
Denaturation	98 °C	20 sec	
Annealing	50 – 65 °C; 60 °C	15 sec	25
Extension	72 °C	15 sec	
Final Extension	72 °C	1 min	1
Hold	4 °C	∞	1

Primer specificity was assessed for both primer pairs by extracting DNA from *P. bernhardus* using the same protocol and running PCRs, with CMC and *P. bernhardus* DNA, and gel electrophoresis following the same procedure as for the gradient PCR, with T_a of 60 °C (Table 2-3).

5. qPCR, Optimization and Standard Curve generation

qPCR used SYBR Select Master Mix (Life Technologies, 4472908) on an Agilent AriaMX Real-Time PCR System. For the standard curve, PCR products from each primer pair were purified with the QIAquick PCR Purification Kit (Qiagen, 28104) and serially diluted. An initial 7-point 10-fold series starting at 10 ng was replaced by an 8-point series starting at 0.1 ng. Samples were run using three replicates.

Reactions (20 µL) contained 10 µL 2x SYBR Select Master Mix, 0.5 µL of each primer (10 µM), and 9 µL template. Samples were run undiluted and at 5, 10, and 20-fold dilutions. Cycling conditions: 50 °C 2 min (UDG activation), 95 °C 2 min (polymerase activation), then 40 cycles of 95 °C 15 s and 56 °C 30 s. A melt curve followed from 60 °C with 0.5 °C increments (5 s/step).

6. Nanopore Sequencing

All sequencing was performed on a MinION Mk1B (ONT) with R10.4.1 flow cells.

6.1. Sequencing Water samples

Water-sample library was prepared with the Ligation Sequencing gDNA Native Barcoding Kit 96 V14 (SQK-NBD114.96; ONT). The run included the nine CMC samples from (Aq1–Aq3, T1–T3), the negative control, the OC sample, and an additional Aq3 T3 extracted with the standard protocol (Aq3 T3 UW).

Up to 400 ng DNA in 11 µL per sample (Table S5) entered DNA repair/end-prep, followed by barcoding. Because concentrations varied widely, median concentration (17.95 ng/µL) was applied: samples below the median used the full 7.5 µL in the barcode ligation; samples above targeted a constant input mass of 134.625 ng and were topped up with nuclease-free water. Barcoded samples were pooled in full, bead-cleaned with Short Fragment Buffer, and 266 ng in 1.7 µL was loaded. Sequencing lasted 15 h 23 min.

6.2. Sequencing Sediment samples

Sediment library was prepared with the Rapid Sequencing DNA V14 barcoding kit (SQK-RBK114.96; ONT) following the manufacturer's protocol. When inputs were low, samples were concentrated by magnetic-bead cleanup in 2 mL tubes using an equal volume of AMPure XP beads (to retain short fragments) and eluted in 15 µL nuclease-free water. For each sample, the volume required an input of 300 ng (assuming 40% loss), with a 1 mL maximum per tube (Table S6). Cleaned DNA was quantified and up to 200 ng per sample was barcoded.

Barcoded libraries were pooled and eluted in 16.5 µL elution buffer; the pooled concentration was 118 ng/µL (Qubit, 1x HS). Rapid Adapter was added to 11 µL of the prepared library and stored at −70 °C, then thawed on ice and 12 µL was combined with Library Beads and Sequencing Buffer to finalise the library. Sequencing lasted 41 h 46 min.

7. Kraken2 Custom Database

Sequencing reads were classified with Kraken2 (v2.1.3). Three custom databases were built: i) Whole-Genome (WG) database using all NCBI genomes for the study species (Table 4), with default settings (k-mer 35, minimizer 31, minimizer-spaces 7). ii) COI-only reference libraries assembled from COI sequences on NCBI and BOLD for the same species (Table 4), compiled into multiple databases spanning a grid of k-mer/minimizer lengths with the maximum allowed minimizer-spaces for each setup (Table S7). iii) a multi-marker database ("MitoGun") comprising 12S, 16S, COI, ND4, ND5, cyt b, and available mitochondrial genomes retrieved via Entrez Direct (Script 1), where species-specific mitogenomes were absent, congeneric mitogenomes were included (Table 6). Parameter variants matching those in Table 5 were also built for MitoGun.

For each database, sequences were merged into a single FASTA, kraken:taxid headers were added, and duplicates removed using Python 3.12.3 (Scripts 2–4).

Table 4. Content of the reference library, number of whole genomes and COI sequences respectively, of the two custom databases.

Species	Whole Genome available	COI sequences available
E. sinensis	5	289
E. japonica	0	355
C. pagurus	0	82
C. maenas	0	796
P. bernhardus	0	175
P. borealis	1	61
P. depurator	0	1105

Table 6. Content of the reference library used for the custom multimarker database "MitoGun" showing 12S, 16S, COI, ND4, ND5, cytb sequences and mitochondrial genomes (MG) available for each species.

Species	12S	16S	COI	MG	ND4	ND5	cyt b
C. pagurus	1	9	58	0	0	0	1
C. maenas	5	20	834	1	4	5	48
E. japonica	6	8	353	2	0	0	175
E. sinensis	25	27	258	7	0	0	170
P. depurator	0	6	1106	0	0	0	0
P. bernhardus	4	131	176	0	0	0	0
P. ochotensis	0	0	0	1	0	0	0
P. borealis	1	20	46	1	0	1	1

Statistical Analysis

Statistical analyses were run in R (v4.5.1). Normality and variance homogeneity were tested with Shapiro–Wilk and F-test, respectively. DNA concentrations were compared using Welch's or Student´s two-sample t-test, depending on assumptions and log10-transformed for boxplots. Binary detection (qPCR vs. ONT) was compared with McNemar's test at $\alpha = 0.05$.

Results

1. Aquarium Experiment

A comparison of the change in DNA concentration over time of the different aquariums was performed through line plots (Figure S5), which showed no discernible pattern across the sample groups (water samples, sediment samples from first experiment and from the second experiment).

3. eDNA extraction

3.1. Water samples

Extractions performed strictly according to the NucleoSpin eDNA Water protocol yielded high DNA concentrations (13.6–57.0 ng/µl) and acceptable 260/280 ratios (1.96–2.00; Table S8), but low 260/230 ratios (0.56–1.19) (Thermo Fisher Scientific, 2015). Using the optimized protocol, DNA concentrations were similar (8.36–58.0 ng/µl) with 260/280 ratios of 1.88–2.06 and improved 260/230 ratios (0.94–2.33; Table S9). Thus, protein contamination was unlikely (260/280 ~2.0), while some samples still showed residual phenol and carbohydrate impurities based on 260/230 values.

3.2. Sediment samples

The DNA concentrations of the sediment samples extracted, using the DNeasy PowerMax Soil Kit (Qiagen), from the first aquarium experiment ranged from 0.265 ng/µl to 0.941 ng/µl, the 260/280 values ranged from -4.55 to 40.35 and 260/230 values ranged from 0.06 to 1.74 (Table S12). The extractions from the second aquarium experiment had a concentration varying between 0.0690 and 1.25, with 260/280 values from 2.19 to 4.38 and 260/230 values from 0.11 to 0.21 (Table S12). While the extractions from the field samples varied in concentration 0.196 ng/µl to 11.6 ng/µl, with 260/280 values from 0.63 to 3.91 and 260/230 values from 0.09 to 1.39 (Table S12). Since all concentrations were <20 ng/µl, Nanodrop purity ratios should not be trusted, as accuracy declines at low DNA concentrations (Koetsier & Cantor, 2019).

3.3. Comparing sample types

DNA concentrations of the water and sediment samples from the first aquarium experiment were compared. Normality was assessed with the Shapiro–Wilk's test (water $W=0.94567$, $p=0.5747$; sediment $W=0.89216$, $p=0.148$); equality of variances was rejected ($F=0.00020198$, $df1=10$, $df2=11$, $p=2.20E-16$). Consequently, Welch's two-sample t-test was applied, showing higher DNA concentrations in water than sediment ($t=-7.0133$, $df=11.005$, $p=2.23e-05$). The lower DNA concentrations in the sediment samples are shown in Figure 3A (data $\log_{10}$-transformed).

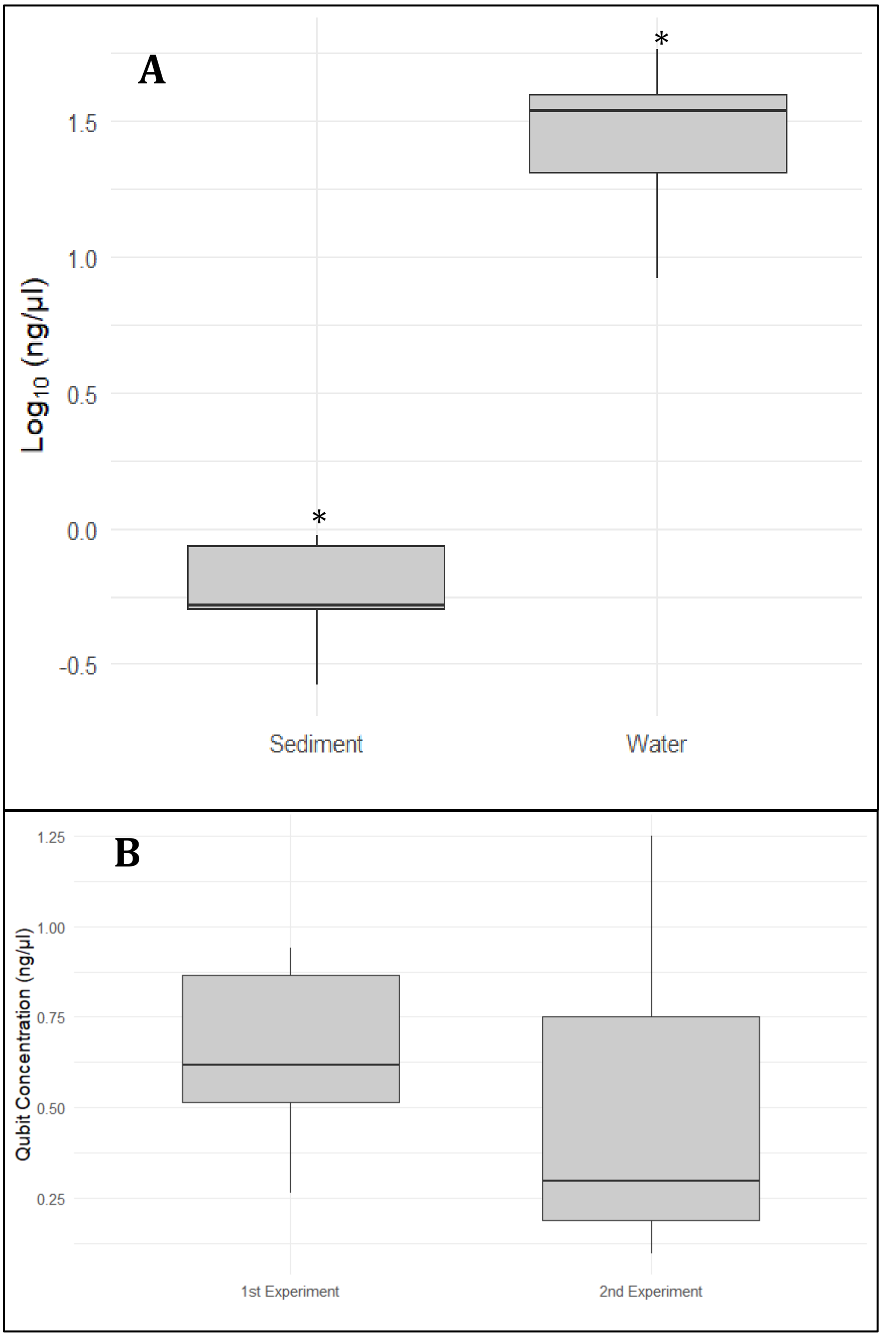

Figure 3. Boxplots of DNA concentrations from aquarium experiments. (A) Log_{10}-transformed DNA concentrations of water samples extracted with the NucleoSpin® eDNA Water kit (Macherey-Nagel) and sediment samples from the first aquarium experiment extracted with the DNeasy PowerMax Soil kit (Qiagen). Asterisks (*) indicate significant differences between sample types ($p < 0.05$). (B) DNA concentrations of sediment samples from aquarium experiment 1 and experiment 2, both extracted with the DNeasy PowerMax Soil kit (Qiagen).

Sediment DNA concentrations were also compared between aquarium experiments after excluding OC, NC and NNC (low concentrations; OC absent in experiment 2). Normality was supported (experiment 1 W=0.90333, p=0.2719; experiment 2 W=0.8468, p=0.1483), variances were equal (F=3.5687, df1=5, df2=8, p=0.1083), and Student's two-sample t-test indicated no difference (t=−0.87283, df=13, p=0.3986; Figure 3B).

4. Primer verification

Gradient PCR produced single bands at the expected sizes for both primer pairs (SS: 100 bp; GW: 137 bp) across all annealing temperatures tested (Ta = 50–65 °C), with no amplification in non-template controls and no visible secondary products (Figure S6A). The DNA concentrations for the PCR products amplified using the SS primers ranged from 13.2 to 14.4, with 260/280 values ranging from 1.30 to 1.84 and 260/230 values from 1.26 to 1.80 (Table S13). While for the PCR products amplified using the GW primer the DNA concentration ranged from 22.5 to 26.1, with 260/280 values ranging from 1.80 to 1.86 and 260/230 values from 1.72 to 1.82 (Table S13).

Specificity testing showed amplification from Eriocheir sinensis DNA and no amplification from *P. bernhardus* for both primer pairs; non-template controls were negative (Figure S6B). The observed pattern indicates target specificity under the tested conditions and absence of reagent contamination. Amplicon yields and purity metrics are summarized in Table S14.

5. qPCR, Optimization and Standard Curve generation

The first standard-curve attempt produced early, high-signal amplification at the two most concentrated points (10 ng and 1 ng) for both primer pairs, resulting in no recorded Cq values despite visible amplification (Figures S7–S8). The threshold was set to 0.7 (mid-exponential phase of the first above-baseline curve; 1:100 GW). Remaining points for both primer sets showed baseline-originating amplification with Cq increasing by dilution; NTCs were negative (Table S14). Melt curves indicated a single product for SS (Tm 74 °C) and a near-single product for GW (mostly 75 °C, one 75.5 °C).

Excluding the two saturated points enabled five-point standard curves (Figure 4A-4B). SS met quality criteria (R^2=0.996; slope −3.564; efficiency 90.79%), consistent with NEN-EN-ISO-15216-1 acceptance ranges (R^2≥0.98; slope −3.6 to −3.1; efficiency 90–110%). GW had acceptable R^2 (0.998) but unacceptable slope and efficiency (−4.351; 69.75%). Based on these metrics and the minor Tm variability, subsequent work proceeded with the SS primer.

A second SS standard-curve attempt (eight 10-fold dilutions, 0.1 ng to 1×10^{-8} ng) again showed a single melt-curve peak at 74 °C. With threshold 0.7, all but the lowest point produced Cq values that increased by dilution; the lowest amplified too late to cross the threshold (Figure S9). This could indicate an LOD between 10^{-7} and 10^{-8} ng, but further testing with more replicates would be needed to be certain. Despite R^2=0.999, slope and efficiency remained outside acceptable ranges (−4.339; 70.02%) (Figure 4C). Because acceptable efficiency could not be achieved within the timeframe, SS was used for presence/absence only, not for quantification.

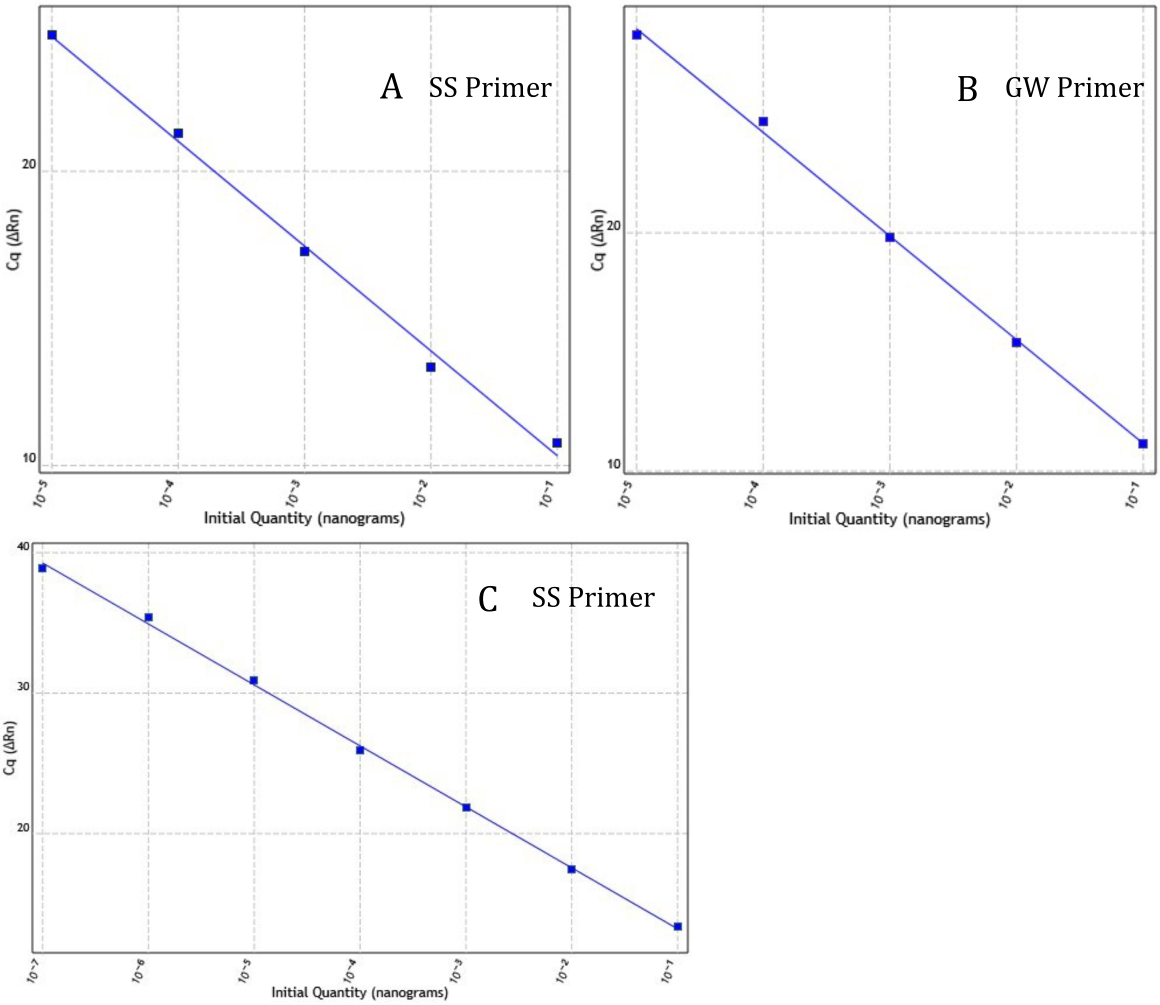

Figure 4. Standard curves of the primer pairs based on 10-fold dilution series, using three replicates. (A) SS primer pair and (B) GW primer pair, each generated from a 5-point dilution series starting at 0.1 ng and ending at 1.0×10^{-5} ng. (C) Second standard curve generated for the SS primer pair. The x-axis shows the template amount (ng) and the y-axis shows the Cq value.

qPCR detection using the SS primer pair was evaluated on experimental and field sediment samples and on experimental water samples (threshold 0.15). In sediment samples from the first aquarium experiment, undiluted extracts showed no amplification except the positive control; the positive-control Tm (73.5 °C) matched the expected product. After 10-fold dilution, amplification was detected in three sediment samples (Aq1 T1, Aq3 T1, Aq3 T2); OC and NC were negative. Tm values for positives were within ~1.5 °C of the expected product at ~74 °C (Table 7; Figure S10).

In the water samples (10-fold diluted), amplification was detected in all samples expected to contain CMC DNA except Aq3 T3; OC and NC were negative. Product Tm values were ~74 °C (approximately half at 73.5 °C). In sediment samples from aquarium experiment 2 (10-fold diluted), amplification was detected for all NAq samples except NAq1 T2; NNC and NTC were negative. Tm varied by ≤1 °C from the expected value. Field sediment samples showed no amplification and observed Tm values were ≥3.5 °C from the expected product (Table 7).

Table 7. Cq values and T_m of the 10-fold diluted sediment samples from the first aquarium experiment, water samples, sediment samples from the second experiment, and the fields samples. Threshold set to 0.15.

Well Name	Cq (ΔRn)	Tm Product 1 (-Rn'(T)) (˚C)
Aq1 T1 Sediment 1	37.52	73

Aq2 T1 Sediment 1	No Cq	72.5
Aq3 T1 Sediment 1	34.63	72.5
Aq1 T2 Sediment 1	No Cq	73
Aq2 T2 Sediment 1	No Cq	72.5
Aq3 T2 Sediment 1	39.38	73
Aq1 T3 Sediment 1	No Cq	73
Aq2 T3 Sediment 1	No Cq	73
Aq3 T3 Sediment 1	No Cq	70
NC Sediment 1	No Cq	70.5
OC Sediment 1	No Cq	74
Positive Control	19.87	74
Aq1 T1 Water	32.82	74
Aq2 T1 Water	36.25	73.5
Aq3 T1 Water	29.19	73.5
Aq1 T2 Water	39.94	73.5
Aq2 T2 Water	39.44	73.5
Aq3 T2 Water	35.69	73.5
Aq1 T3 Water	39.77	74
Aq2 T3 Water	38.36	74
Aq3 T3 Water	No Cq	74
OC Water	No Cq	63
NC Water	No Cq	70.5
Positive Control	20.4	74
NAq1 T1 Sediment 2	36.18	73.5
NAq2 T1 Sediment 2	30.71	73
NAq3 T1 Sediment 2	32.16	73
NAq1 T2 Sediment 2	No Cq	73
NAq2 T2 Sediment 2	35.75	73
NAq3 T2 Sediment 2	35.44	73
NNC Sediment 2	No Cq	72.5
NTC Sediment 2	No Cq	63
Positive Control	20.4	74
Sannorna 1	No Cq	70.5
Sannorna 2	No Cq	63
Sannorna 3	No Cq	63
Sannorna 4	No Cq	63
Stenhammar 1	No Cq	70
Stenhammar 2	No Cq	70.5
Stenhammar 3	No Cq	65.5
Stenhammar 4	No Cq	63

Dilution testing at 5-fold and 20-fold confirmed these patterns: NAq samples remained positive except NAq1 T2; NNC, NTC, and all field sediment samples remained negative, with NAq Tm deviations up to ~2.5 °C (Table S17)

6. Nanopore Sequencing

Both runs used FLO-MIN114 flow cells. The flow cell for the water samples started with ~1,500 available pores and remained comparatively stable during sequencing (Figure S18A). The flow cell for the sediment samples started with ~500 available pores and declined markedly over time; sequencing was therefore extended to increase yield (Figure S18B).

Yield metrics differed substantially between runs. The water run produced ~15.5-fold more reads and ~5.7-fold more bases, whereas the sediment run produced ~5-fold longer reads (N50). The proportion of reads passing quality filters was also higher in the water run (96.6%) compared to the sediment run (80.4%). Both runs had a minimum q-score of 9 (Table 12).

Table 12. Performance of the ONT sequencing of both the run with the water samples and the run with the sediment samples.

Run	Kit	Estimated bases	Reads generated	Estimated N50	Passed reads	Failed reads	Run duration	Min Q Score
Water	SQK-NBD114-96	8.22 Gb	10.14 M	1.06 kb	9.8 M	643.79 k	15 hrs 23 min	9
Sediment	SQK-RBK114-96	1.43 Gb	653.11 k	5.13 kb	525.15 k	152.03 k	41 hrs 46 min	9

7. Kraken2 Classification with Custom Databases

For the classification performed with the WG database classification was performed with confidence levels ranging from 0.0 to 1.0 in 0.1 increments. Increasing confidence reduced the number of classified reads. At low thresholds, non-target assignments occurred: reads classified as *P. borealis* in aquarium (Aq) and field samples, CMC reads in OC, and both *P. borealis* and CMC in NC (Tables S16–S17). NC reads disappeared by confidence 0.1 in the sediment run, but not until 0.5 in the water run. Wrongly assigned *P. borealis* reads in aquarium samples fully disappeared from the water samples at confidence 0.5 and from the sediment samples at confidence 0.4. For OC, the sediment run showed no CMC at confidence ≥0.1 and no *P. borealis* even at 0.0. In the water OC sample, P. *borealis* disappeared at 0.3, whereas one CMC read remained even at 1.0. For field samples, both CMC and *P. borealis* were detected in all Sannorna samples up to 0.7 and 0.8 respectively. At Stenhammar, one sample retained CMC to 0.3 and *P. borealis* to 0.6.

The COI-only database was not able to generate any classified reads, even at 0.0 confidence. Few reads were classified when the k-mer and minimizer settings were lowered, several of them were however assigned wrongly.

Using the MitoGun database, classified read counts were drastically higher than with the COI-only database (Figure 5; Tables S20–S24). At confidence 0.1, CMC reads were detected in multiple aquarium (Aq) water samples; CMC reads were also present in the NC and OC water samples. The OC water sample contained no reads from the other decapod species, and two Aq water samples contained three reads classified as *P. bernhardus*. In sediment samples, numerous CMC reads were observed from the second aquarium experiment, whereas the first experiment only yielded CMC reads at the first sampling time; NC and OC sediment samples had no classified reads. In field samples, three CMC reads were detected (Sannorna 1: 1 read; Sannorna 2: 2 reads).

Increasing confidence reduced non-target assignments but also removed some target detections. In water, CMC reads in the NC disappeared at 0.2; the OC retained 10 CMC reads at 0.2 and none at 0.6. At 0.6, five of ten Aq water samples, all Aq sediment samples from experiment 1, and two of six NAq sediment samples from experiment 2 had no remaining CMC reads (Figure 5; Tables S22–S24).

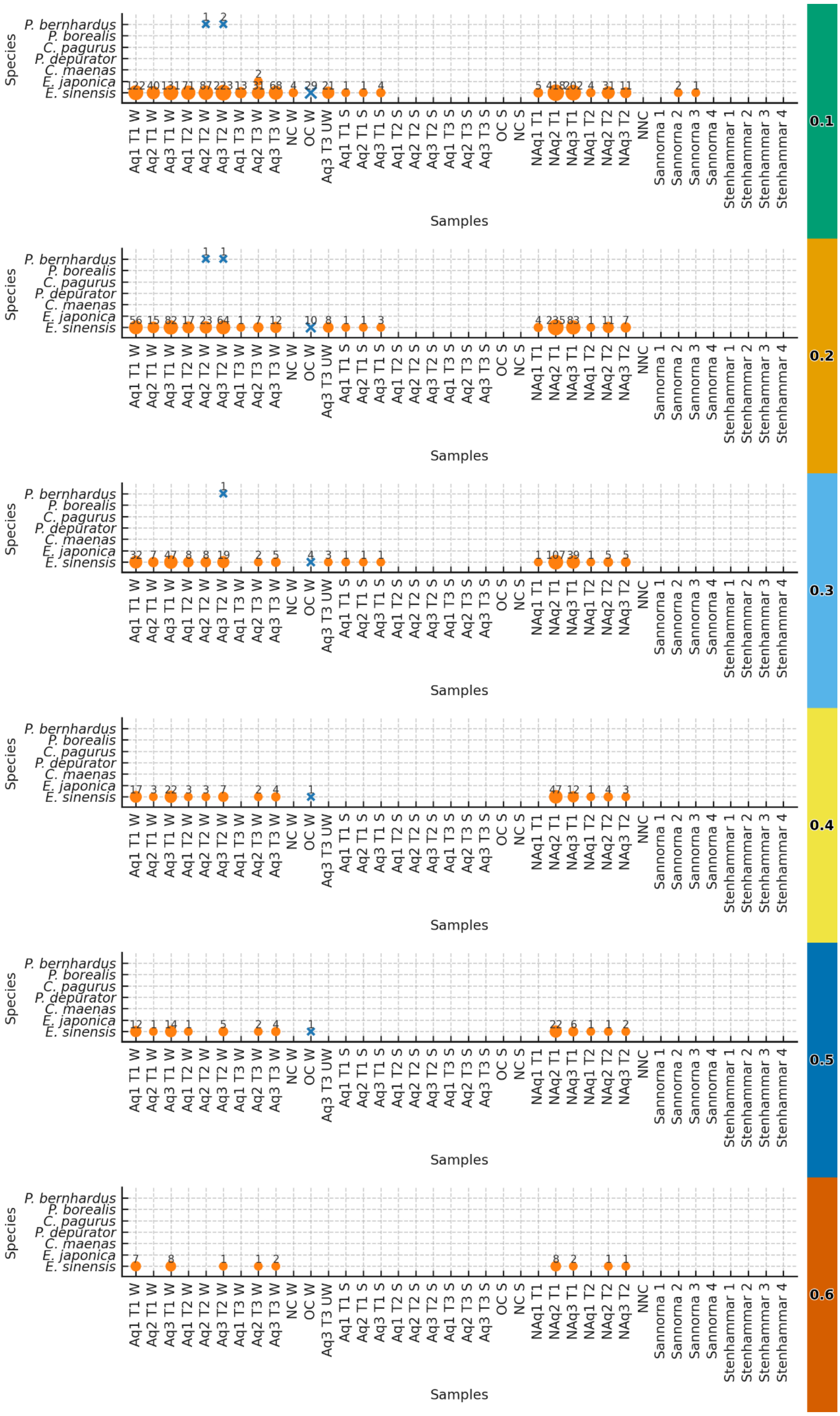

Figure 5. Reads classified with the MitoGun Database using the default parameters and confidence ranging from 0.1 to 0.6. Each panel combines the water samples from the first and second aquarium experiment as well as the field samples. The y-axis lists all species in the reference set. Symbols indicate outcome (○

expected, × non-target); the marker area is proportional to Log10(reads + 1), and the numeric label gives the read count. "Expected" was defined by sample type (*Eriocheir* spp. expected in Aq/NAq and field samples; non-*Eriocheir* expected in OC; none expected in NC); *Eriocheir* spp). Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); W = water samples; S = sediment samples; OC = other-crustaceans control; NC/NNC = negative controls; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

8. Comparing qPCR and Nanopore Sequencing

Detection outcomes for all aquarium (Aq/NAq) samples from both experiments were converted to a binary scale (1 = CMC detected; 0 = not detected; Table S29). qPCR positives were defined as any amplification at threshold 0.150. ONT positives were defined as ≥1 read classified as CMC using the MitoGun database (default parameters) at confidence 0.6, selected as the lowest threshold with no apparent false positives.

The binary outcomes were compared with McNemar's test (χ^2 = 4.0, df = 1, p = 0.045), indicating a significant difference between methods. McNemar's test showed that among the samples, 7 were negative by both qPCR and ONT, 1 was ONT-positive but qPCR-negative, 8 were qPCR-positive but ONT-negative, and 8 were positive by both methods. Thus, under these criteria, qPCR detected CMC more often than ONT.

Discussion

1. Aquarium Experiment

eDNA levels reflect three processes: release (shedding/decay), transport (currents/disturbance), and degradation (enzymes/UV/chemistry) (Mauvisseau et al., 2022; Pawlowski et al., 2020; Qin et al., 2019).

eDNA samples inevitably contain abundant non-target DNA, primarily from microbial and planktonic taxa that dominate biomass and cell counts (Djurhuus et al., 2018; Pawlowski et al., 2020). The aquariums therefore likely included background DNA from the sediment and from untreated tap water (Lührig et al., 2015; Mauvisseau et al., 2022). DNA concentrations in water and sediment were expected to scale with leg number and the time between deployment and sampling; this pattern was not observed (Figure S5). However, since the DNA concentration was not measured before adding the tissue no true comparison could be performed.

2. Field samples

Kinneviken (Figure 1) is a consistent hotspot for CMC observations (Leidenberger et al., 2021; SLU Artdatabanken, 2025c). The two sites sampled, Stenhammar and Sannorna, were selected based on Artportalen records and advice from a local commercial fisher, and their substrates (sand, clay, mixed) matched previously reported capture habitats in the bay (Leidenberger et al., 2021). Despite this, non-detection is plausibly explained by timing and depth. Sampling occurred in early April, when crabs are less active and observations decline relative to summer–autumn; in spring they are often situated in deeper water (Leidenberger et al., 2021; Veilleux & de Lafontaine, 2007). Sediment was collected from ~0.2–1 m water depth, whereas prior catches span 0–45 m in this region (Leidenberger et al., 2021). Future surveys should prioritise late summer–autumn and include deeper sampling, which will likely require alternative field methods.

3. eDNA extraction

3.1. Comparing water and sediment samples

Both water and sediment were used as eDNA sources, and comparing their efficacy is important for invasive-species detection (Pawlowski et al., 2020). Because different ONT library kits and flow-cell starting conditions were used, ONT data are not directly comparable between sample types. However, extraction yield can be compared: Welch's two-sample t-test and the boxplot (Figure 3A) showed significantly higher total DNA concentrations in water than in sediment, suggesting more efficient DNA recovery from water. The fraction of CMC DNA within total DNA remains unknown. Higher totals in water could reflect more non-target DNA (e.g., bacteria, plants, algae).

The sediment samples were expected to yield higher concentrations than observed seeing as the amount sediment was high but still within that prescribed by the manufacturer. The cause of this could be that the sediment type, sand, provides limited surface area and few reactive binding sites, resulting in poor adsorption of extracellular DNA (Buxton et al., 2017; Ogata et al., 2021). DNA associated with finer suspended particles may instead have remained in the water column, especially after agitation prior to sampling, leading to higher concentrations in the water samples as sediment-bound eDNA can be resuspended into the water column (Harrison et al., 2019).

qPCR results also favoured water: CMC was detected in 8/9 water samples (90%) versus 3/9 (33%) sediment samples from Aquarium Experiment 1; sediment from Experiment 2 performed better (5/6; 83%), giving 53% overall for sediment. This pattern suggests water may be more reliable for qPCR, but results should be interpreted cautiously because the first sediment run could be affected by inexperience. When comparing water to Experiment 2 sediment only, detection was similar (water 8/9; sediment 5/6).

Sediment might be expected to perform better given CMC's benthic habits and the longer persistence of DNA in sediments (Ogata et al., 2021; Veilleux & de Lafontaine, 2007). Robinson et al. (2019) reported more positive detections within sediment samples, yet water yielded positives at more locations. Further work is therefore needed to resolve which matrix offers the most reliable detection for CMC under field-relevant conditions.

3.2. Comparing the sediment samples from the two aquarium experiments

The sediment sets differed in substrate source and handling, leg processing, and preservation (ethanol present only in Experiment 1). The Student's t-test and the boxplot (Figure 3B) showed no significant difference in total DNA concentration between experiments, suggesting that these factors did not measurably affect yield. Direct evidence for sediments is scarce, but for water samples, ethanol preservation for ≤4 days performed comparably to freezing at −20 °C (Hinlo et al., 2017), indicating ethanol can be an acceptable eDNA preservative.

Leg handling may have mattered, repeated freeze–thaw cycles fragment high-molecular-weight DNA and can reduce recoverable yield (Shao et al., 2012). Due to the EU-(Regulation (EU) No 1143/2014) living CMC cannot be kept or transported; consequently, only dead/frozen material was used. Follow-up work using fresh or living material would however be of great interest.

4. Primer verification

Both primer pairs performed reliably for CMC detection. Their short amplicons better tolerate eDNA degradation (Jo et al., 2022). Each amplified CMC tissue across a range of Ta. Given Sweden lacking freshwater brachyuran crabs and freshwater Decapoda are limited to crayfish (*A. astacus*, *Pacifastacus leniusculus*) only, false positives from other freshwater decapods are unlikely (Sundelöf et al., 2021). No amplification was observed from *Pagurus bernhardus* tissue (Figure 6B). GW specificity to *Eriocheir* was independently supported by Chevrinais et al. (2023), who reported no amplification in sympatric taxa. SS specificity was further supported by the absence of amplification in all OC water/sediment extracts (Table 7).

5. qPCR, Optimization and Standard Curve generation

The SS primer produced an acceptable standard curve (R^2 = 0.996, slope −3.56, efficiency 90.8%), while the GW primer failed to meet efficiency requirements. A second SS curve again showed high linearity (R^2 = 0.999) but poor efficiency (70.0%). This suggests that with further optimisation and additional dilution points, reliable curves could likely be generated for both primers, but time constraints limited attempts to improve them. As a result, quantification was abandoned, and qPCR was used solely for presence/absence detection. The estimated detection limit of the SS primer may be as low as 10^{-7}–10^{-8} ng, but this requires further confirmation.

Undiluted sediment samples failed to amplify, indicating strong PCR inhibition, most likely from humic substances binding polymerase or chelating ions (Matheson et al., 2010; Stark et al., 2000). Dilution markedly improved performance, consistent with previous studies on eDNA inhibition (Buxton et al., 2017; McKee et al., 2015). Testing at 5-, 10- and 20-fold dilutions in the second aquarium experiment showed nearly identical detection success, suggesting that a 5-fold dilution is sufficient to reduce inhibitors while minimising target DNA loss. Ideally, all samples should have been run at these dilutions in parallel to systematically identify the optimal balance.

All field samples, tested undiluted and diluted, remained negative, supporting the absence of CMC DNA at the time of sampling.

Because many positive detections occurred at late cycles, increasing cycle numbers (e.g., to 50) could potentially improve sensitivity, as shown in other eDNA studies (Veldhoen et al., 2016), though at the cost of higher risk of non-specific products. The use of the cost-effective SYBR Green, also added uncertainty, since it fluoresces with any dsDNA. Melt-curve verification was therefore critical, but probe-based assays would offer greater specificity by producing signal only from target binding (Goldberg et al., 2016; Pawlowski et al., 2020). Droplet digital PCR (ddPCR) represents a further improvement, providing absolute quantification and enhanced sensitivity at low concentrations (Guri et al., 2024).

6. Nanopore Sequencing

Direct comparison of the two ONT runs is confounded by three factors that differed simultaneously, matrix (water vs. sediment), library kit (Ligation vs. Rapid), and initial pore count, so differences in yield and read length cannot be attributed to a single cause.

The water run produced ~8.22 Gb and 10.14 M reads vs. 1.43 Gb and 0.65 M for sediment, despite a shorter runtime. Kit chemistry can affect output: Rapid libraries often yield less than Ligation due to transposase-driven fragmentation and the streamlined workflow (Oxford Nanopore Technologies, 2024), and empirical comparisons have reported markedly higher yields with Ligation (≈3× in bacterial genomes) (Sauvage et al., 2023). Although a bead-cleaning step was added for the sediment DNA to reach input, chemistry differences likely still mattered (Oxford Nanopore Technologies, 2024). However, the dominant driver of the yield gap was almost certainly flow-cell quality: the water run began with ~1,500 active pores and the sediment run with ~500. Total data output scales with starting pore availability (De La Cerda et al., 2023), and low-pore cells benefit from reduced sample loading to avoid oversaturation and reduced reads (Meyer et al., 2025), a consideration not applied here. This together with the flow cell health most likely explains its low yield.

Paradoxically, the sediment run had a higher N50 (5.13 kb) than the water run (1.06 kb). N50 shows little association with initial pore count (De La Cerda et al., 2023) and is primarily shaped by input DNA integrity and handling. Sediment-bound eDNA can adsorb to particles and be protected from degradation relative to water-column DNA (Sakata et al., 2020). Workflow can further modulate length: multi-step Ligation with repeated pipetting/clean-ups risks mechanical shearing, whereas Rapid uses single-step enzymatic tagmentation that is physically gentler (Oxford Nanopore Technologies, 2024; Wick et al., 2021). Literature on kit effects is mixed: some report longer reads with Ligation than Rapid (Chen et al., 2017), others report no consistent difference (Sauvage et al., 2023; Wick et al., 2021), and some show Rapid producing longer reads, potentially due to less shear (Jain et al., 2018). Also, datasets with more reads can trend toward shorter N50s (negative association), which fits the water vs. sediment pattern (De La Cerda et al., 2023).

Environmental ONT datasets span a wide N50 range. With optimised extraction, soil eDNA can reach ~8–14 kb (Cousson et al., 2024), placing the sediment N50 as plausible yet improvable. For complex water metagenomes, N50s <1–3 kb has been reported (Reddington et al., 2020; Serite et al., 2023), so the water run's ~1.06 kb is low but not unprecedented. Overall, while kit and matrix likely contributed, the overwhelmingly low pore count on the sediment flow cell best explains its poor yield.

7. Kraken2 Classification

Sequencing reads were classified with Kraken2 using custom databases because no pre-built set covered the study taxa. WG classification was attempted first, which in principle best matches Kraken2's design and would provide more reference material for shotgun reads (Odom et al., 2023). Coverage gaps limited its usefulness: only one of five OC species had a reference genome (*P. borealis*), CMC was represented by five genomes, and *E. japonica* had none, meaning introgression testing was not feasible. WG classification produced many apparent false hits that decreased as confidence increased, but at the cost of losing assumed true positives, particularly in sediment samples. Lower confidence thresholds (0.1–0.3) removed most assumed false negatives, yet NC and OC water still contained reads classified as CMC at 0.4–1.0, including one read at 1.0 in OC water (Tables S18-S19). A plausible explanation is misassignment from other OC species that are more similar to CMC than to *P. borealis*, which is not brachyuran. Reads assigned to *P. borealis* in field samples, especially Sannorna, persisted up to confidence 0.7 despite the species being marine and absent from Lake Vänern; given the lack of freshwater shrimps in Vänern, misclassification or low-level contamination is a more likely explanation than true presence (Sundelöf et al., 2021). These outcomes highlight the value of broader reference libraries and additional genomes.

Kraken2 can perform well with marker-gene and mitogenome libraries, sometimes outperforming traditional 16S pipelines (Bayer et al., 2025; Odom et al., 2023). A COI-only database yielded no classified reads with default parameters even at confidence 0.0. Guided by reports that shorter k-mers/minimizers can help with error-prone ONT reads (Lu et al., 2022), multiple COI databases with reduced parameters were tested. Total classifications increased as parameters decreased, but most assignments vanished already at confidence 0.1, consistent with spurious matches at very short k-mers (Tables S20-S21). The k21_m17 build was selected as a compromise, yet at confidence 0.1 only one CMC read was detected while OC species produced more hits in Aq water and field samples, reflecting the low probability of capturing COI fragments in shotgun eDNA without targeted enrichment (Table S22).

The MitoGun database performed markedly better, even at default settings, as expected from broader target coverage. However, OC water still contained reads classified as CMC at higher confidences (up to 0.5), indicating residual false positives even with comprehensive mitochondrial coverage. Possible contributors include reference bias toward CMC mitogenomes, contamination, sequencing errors, or barcode misassignment (Xu et al., 2018). Only at confidence 0.6 did these CMC calls disappear, but by then more than half of Aq and NAq samples lost all CMC reads (Figure 7; Table S24-S28), illustrating the sensitivity–specificity trade-off for ONT shotgun eDNA paired with custom Kraken2 databases.

Classification performance could likely be improved with expanded taxonomic coverage, especially additional whole genomes and mitogenomes, or by adding a PCR enrichment step for the chosen marker prior to sequencing. Reported optimal Kraken2 confidence thresholds can be lower (approximately 0.2–0.4) when using comprehensive pre-built libraries (Liu et al., 2024), but higher thresholds were required here, consistent with the less complete, custom libraries used. KrakenUniq might allow higher specificity through unique k-mer tracking, albeit with reduced speed (Breitwieser et al., 2018; Marcelino et al., 2020).

8. Comparing qPCR and Nanopore sequencing

Both methods detected CMC from eDNA but require optimisation for reliable use. For qPCR, moving from dye-based assays to probe-based assays, or preferably ddPCR, would improve specificity and sensitivity. For ONT, detection would benefit from a more comprehensive reference library and, potentially, classifiers with higher specificity. Using the MitoGun database at confidence 0.6, McNemar's test showed a significantly higher detection rate for qPCR, with approximately twice as many positives as ONT. Under the conditions used here, qPCR was therefore the more effective targeted approach. Its lower per-sample cost is an additional advantage, enabling larger sample sizes and increased monitoring effort.

Conclusion

This study evaluated detection of Chinese mitten crab from eDNA using qPCR and Oxford Nanopore sequencing across water and sediment, in aquariums and in the field. Results underscored the stochastic nature of eDNA dynamics and yielded no confirmed field detections from the two sites, likely affected by season and depth. Water produced higher DNA concentrations and slightly better qPCR detection, whereas sediment may preserve DNA longer but was limited here by extraction and inhibition. Primer verification supported both SS and GW assays, with SS used for presence/absence. ONT performance was constrained by variable library preparation, flow cell quality, and incomplete references, emphasizing the need for broader mitochondrial and genomic coverage and careful classifier settings. A direct comparison showed qPCR detecting CMC more often than ONT under the thresholds used, and its lower per-sample cost favours larger monitoring effort. Overall, qPCR is presently the more reliable and cost-effective option for targeted CMC surveillance, while ONT remains promising for broader biodiversity surveys pending improved extraction, pore health, and reference databases.

Future perspectives

This study has identified several key areas for future research to enhance molecular methods for detecting CMC in Swedish waters.

Further investigation on the optimal sample type, water or sediment, for extracting eDNA is needed. Although water samples showed higher DNA concentrations and qPCR consistency, sediment samples may still offer advantages due to their potential long-term stability and DNA concentration capacity (Sakata et al., 2020). Addressing PCR inhibition through systematic testing of dilution strategies is essential. Future research should attempt to clearly define optimal dilution practices to maximize DNA recovery while effectively mitigating inhibitors. Additionally, further studies should isolate factors affecting ONT sequencing performance, including sample type, DNA integrity, sequencing kits, and flow cell quality. Controlled comparative experiments varying these factors individually would help clarify their impacts on sequencing outcomes. Improving Kraken2 classification by expanding genomic reference databases, especially comprehensive whole genomes, and mitochondrial genomes, remains critical. Evaluating alternative bioinformatics tools, such as KrakenUniq, could also enhance classification specificity and sensitivity (Breitwieser et al., 2018). The performance of dye-based qPCR could benefit from validation against more specific methods like probe-based qPCR or ddPCR, establishing standardized methodologies for low-abundance eDNA detection (Guri et al., 2024). The presence of *E. japonica* and/or *Eriocheir* hybrids in Sweden, which has been noted in other European countries by Homberger et al. (2022) and Palero et al. (2022), need to be investigated further in order to have sufficient monitoring of the genus. Finally, field sampling conducted across varying seasons and depths is recommended to better understand CMC distribution in Lake Vänern, Sweden. Addressing these research areas will significantly refine eDNA detection methods and therewith help to improve invasive species management strategies, independently of the target organism.

Ethical aspects and Impact on society

Other species of crabs have previously been shown to experience pain (Kasiouras et al., 2024), it would therefore be logical to assume that CMC does so as well. The specimens used in this project were humanely euthanized, by freezing them to death, in order to avoid unnecessary suffering. The species plays an important role in its native habitat, the specimens for this project did however originate from captured specimens from invasive areas where it is important to remove them from the ecosystem. Since they need to be captured and killed anyway, it is important that they are used to gain scientific knowledge about the species and its detection.

Care was taken while taking samples in the field to avoid contaminating and harming the environment with any foreign material or chemicals. Care was also taken to avoid contaminating the samples, all material were therefore either single-use or cleaned to be DNA-free.

Artificial intelligence was used in a limited, assistive role. ChatGPT (OpenAI, v.4.0, v.4.5 and v5.0) was employed to draft and debug scripts for figure generation and data handling, to generate Figure 11 and 12, to accelerate literature searches and preliminary citation formatting, and to refine phrasing for clarity. Experimental design, data acquisition, statistical analyses, and final interpretation were performed by the author. AI outputs were treated as suggestions only; all text was independently reviewed and revised, bibliographic details were verified against the original sources, and no AI system was used to fabricate, manipulate, or filter any results.

The early detection of invasive species such as the Chinese mitten crab has a clear societal relevance. If established, the species could negatively impact local biodiversity, ecosystem services, and economically important species such as the native European crayfish, *A. astacus*. By evaluating and comparing molecular detection methods, this study contributes to developing tools that can improve monitoring programs and inform management decisions. Reliable detection methods make it possible to act before populations become established and costly eradication or control measures are required. Thus, the results of this project can support conservation efforts, reduce long-term economic costs, and help safeguard both ecological and cultural values connected to Swedish freshwater systems.

Acknowledgements

I would like to thank my supervisor Sonja Leidenberger for providing me with such an interesting project and her guidance with seeing it through to the end. I would like to thank John Baxter for helping me structure the practical work and for his help in the lab and teaching me how to use Kraken2, as well as providing moral support. I want to thank Matthew Herring for providing his expertise in qPCR. I would like to thank the eDNA group (Adriana Lopez Leon, Vasilijs Nekrasevics and Teodor Bäckström) for sharing their experience with the eDNA extraction kits, for being so inviting and for their companionship. I would like to thank Filiph Carlsson for always providing an ear when problems and frustrations arouse. I would like to thank Joel Hildingsson for his help with Python scripts. Lastly, I would like to thank my family and friends who have helped me through this challenging time.

References

Anger, K. (1991). Effects of temperature and salinity on the larval development of the Chinese mitten crab Eriocheir sinensis (Decapoda: Grapsidae). *Marine Ecology Progress*, *72*(72), 103–110.

Bayer, P. E., Bennett, A., Nester, G., Corrigan, S., Raes, E. J., Cooper, M., Ayad, M. E., McVey, P., Kardailsky, A., Pearce, J., Fraser, M. W., Goncalves, P., Burnell, S., & Rauschert, S. (2025). A Comprehensive Evaluation of Taxonomic Classifiers in Marine Vertebrate eDNA Studies. *Molecular Ecology Resources*, e14107. https://doi.org/10.1111/1755-0998.14107;PAGEGROUP:STRING:PUBLICATION

Bentley, M. G. (2011). The Global Spread of the Chinese Mitten Crab Eriocheir sinensis. *In the Wrong Place - Alien Marine Crustaceans: Distribution, Biology and Impacts*, 107–127. https://doi.org/10.1007/978-94-007-0591-3_3

Breitwieser, F. P., Baker, D. N., & Salzberg, S. L. (2018). KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biology*, *19*(1), 1–10. https://doi.org/10.1186/S13059-018-1568-0

Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2017). Is the detection of aquatic environmental DNA influenced by substrate type? *PLoS ONE*, *12*(8), e0183371. https://doi.org/10.1371/journal.pone.0183371

Chen, X., Wang, J., Huang, L., Yue, W., Zou, J., Yuan, C., Lu, G., & Wang, C. (2017). Evolutionary relationship of three mitten crabs (Eriocheir sp) revealed by mitogenome and 5S ribosomal DNA analysis. *Aquaculture and Fisheries*, *2*(6), 256–261. https://doi.org/10.1016/J.AAF.2017.10.004

Chevrinais, Marion., Demers, A., Dumoulin, L.-A., Gagné, N., Larivière, J., LeBlanc, F., Simard, N., Therien, A., Weise, A. M., & Parent, G. J. (2023). *Targeted detection of European green crab (Carcinus maenas) and Chinese mitten crab (Eriocheir sinensis) using environmental DNA [Canadian Technical Report of Fisheries and Aquatic Sciences No. 3539]*. Regional Sciences Branch, Quebec region, Benthic and Demersal Sciences Division, Fisheries and Oceans Canada, Maurice Lamontagne Institute. https://publications.gc.ca/collections/collection_2023/mpo-dfo/Fs97-6-3539-eng.pdf

Chiou, C. S., Chen, B. H., Wang, Y. W., Kuo, N. T., Chang, C. H., & Huang, Y. T. (2023). Correcting modification-mediated errors in nanopore sequencing by nucleotide demodification and reference-based correction. *Communications Biology*, *6*(1), 1215. https://doi.org/10.1038/S42003-023-05605-4

Cousson, A., Pablo, A. L., Cournac, L., Piton, G., Dezette, D., Robin, A., Taschen, E., & Bernard, L. (2024). Ultra pure high molecular weight DNA from soil for Nanopore shotgun metagenomics and metabarcoding sequencing. *MethodsX*, *14*. https://doi.org/10.1016/j.mex.2024.103134

Czerniejewski, P., Skuza, L., Drotz, M. K., & Berggren, M. (2012). Molecular connectedness between self and none self-sustainable populations of Chinese mitten crab (Eriocheir sinensis, H. Milne Edwards, 1853) with focus to the Swedish Lake Vänern and the Oder and Vistula River in Poland. *Hereditas*, *149*(2), 55–61. https://doi.org/10.1111/J.1601-5223.2012.02246.X

De La Cerda, G. Y., Landis, J. B., Eifler, E., Hernandez, A. I., Li, F. W., Zhang, J., Tribble, C. M., Karimi, N., Chan, P., Givnish, T., Strickler, S. R., & Specht, C. D. (2023). Balancing read length and sequencing depth: Optimizing Nanopore long-read sequencing for monocots with an emphasis on the Liliales. *Applications in Plant Sciences*, *11*(3), e11524. https://doi.org/10.1002/APS3.11524

Djurhuus, A., Pitz, K., Sawaya, N., Rojas-Márquez, J., Michaud, B., Montes, E., Muller-Karger, F. E., & Breitbart, M. (2018). Evaluation of Marine Zooplankton Community Structure Through Environmental DNA Metabarcoding. *Limnology and Oceanography: Methods*, *16*(4), 209–221. https://doi.org/https://doi.org/10.1002/lom3.10237

Egeter, B., Veríssimo, J., Lopes-Lima, M., Chaves, C., Pinto, J., Riccardi, N., Beja, P., & Fonseca, N. A. (2022). Speeding up the detection of invasive bivalve species using environmental DNA: A Nanopore and Illumina sequencing comparison. *Molecular Ecology Resources*, *22*(6), 2232–2247. https://doi.org/10.1111/1755-0998.13610

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307. https://doi.org/10.1111/2041-210X.12595

Guri, G., Ray, J. L., Shelton, A. O., Kelly, R. P., Præbel, K., Andruszkiewicz Allan, E., Yoccoz, N., Johansen, T., Wangensteen, O. S., Hanebrekke, T., & Westgaard, J. I. (2024). Quantifying the Detection Sensitivity and Precision of qPCR and ddPCR Mechanisms for eDNA Samples. *Ecology and Evolution*, *14*(12). https://doi.org/10.1002/ECE3.70678,

Harrison, J. B., Sunday, J. M., & Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1915), 20191409. https://doi.org/10.1098/RSPB.2019.1409

Havs- och vattenmyndigheten. (2016). *Eriocheir sinensis: Kinesisk ullhandskrabba [Faktablad]*. https://www.havochvatten.se/download/18.1690613b166605675de4bf35/1599212248193/faktablad-eriocheir-sinensis.pdf

Herborg, L. M., Rushton, S. P., Clare, A. S., & Bentley, M. G. (2003). Spread of the Chinese mitten crab (Eriocheir sinensis H. Milne Edwards) in Continental Europe: Analysis of a historical data set. *Hydrobiologia*, *503*(1), 21–28. https://doi.org/10.1023/B:HYDR.0000008483.63314.3C/METRICS

Hinlo, R., Gleeson, D., Lintermans, M., & Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PLOS ONE*, *12*(6), e0179251. https://doi.org/10.1371/JOURNAL.PONE.0179251

Homberger, L., Xu, J., Brandis, D., Chan, T. Y., Keirsebelik, H., Normant-Saremba, M., Schoelynck, J., Chu, K. H., & Ewers-Saucedo, C. (2022). Genetic and morphological evidence indicates the persistence of Japanese mitten crab mitochondrial DNA in Europe for over 20 years and its introgression into Chinese mitten crabs. *NeoBiota*, *73*, 137–152. https://doi.org/10.3897/NEOBIOTA.73.72566

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., … Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345. https://doi.org/10.1038/NBT.4060

Jo, T., Takao, K., & Minamoto, T. (2022). Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. *Environmental DNA*, *4*(2), 271–283. https://doi.org/10.1002/EDN3.253

Jung, J., Wu, X., Park, T., Lee, S. hyun, Hwang, H. seung, Jung, J., & Jung, J. (2022). Population genetic analysis reveals secondary contact between Eriocheir sinensis and E. japonica in South Korea. *Genes & Genomics*, *44*(5), 593–602. https://doi.org/10.1007/S13258-021-01198-6

Kasiouras, E., Hubbard, P. C., Gräns, A., & Sneddon, L. U. (2024). Putative Nociceptive Responses in a Decapod Crustacean: The Shore Crab (Carcinus maenas). *Biology*, *13*(11), 851. https://doi.org/10.3390/BIOLOGY13110851

Koetsier, G., & Cantor, E. (2019). *A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers*. https://www.neb.com/en/-/media/nebus/files/application-notes/technote_mvs_analysis_of_nucleic_acid_concentration_and_purity.pdf?rev=c24cea043416420d84fb6bf7b554dbbb&hash=09004EB6A516A960025243A057F9B7FB&srsltid=AfmBOoptHnhiEbC7lt6O0JJWZ4_YIi76lDSm6a0tXghRj1k0SHV2IM1d

Kress, W. J., & Erickson, D. L. (2008). DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(8), 2761. https://doi.org/10.1073/PNAS.0800476105

Leidenberger, S., Kilströmer, A., Herring, M., & Bergwall, E. (2021). *Inventering av invasiva främmande arter i Vänern – Kinesisk ullhandskrabban [Rapport 126].* https://www.vanern.se/wp-content/uploads/Rapport-nr-126.pdf?utm_

Leidenberger, S., Obst, M., Kulawik, R., Stelzer, K., Heyer, K., Hardisty, A., & Bourlat, S. J. (2015). Evaluating the potential of ecological niche modelling as a component in marine non-indigenous species risk assessments. *Marine Pollution Bulletin*, *97*(1–2), 470–487. https://doi.org/10.1016/J.MARPOLBUL.2015.04.033

Li, C., Wang, B., Ji, Y., Huang, L., Wang, X., Zhao, W., Wang, Y., Wang, H., & Yao, Y. (2024). Mitochondrial genome provides species-specific targets for the rapid detection of early invasive populations of Hylurgus ligniperda in China. *BMC Genomics*, *25*(1), 1–11. https://doi.org/10.1186/S12864-024-10011-Z/TABLES/2

Liu, Y., Ghaffari, M. H., Ma, T., & Tu, Y. (2024). Impact of database choice and confidence score on the performance of taxonomic classification using Kraken2. *ABIOTECH*, *5*(4), 465–475. https://doi.org/10.1007/S42994-024-00178-0

Liu, Y., Li, Y., Chen, E., Xu, J., Zhang, W., Zeng, X., & Luo, X. (2024). Repeat and haplotype aware error correction in nanopore sequencing reads with DeChat. *Communications Biology 2024 7:1*, *7*(1), 1–10. https://doi.org/10.1038/s42003-024-07376-y

Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., Salzberg, S. L., & Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, *17*(12), 2815–2839. https://doi.org/10.1038/S41596-022-00738-Y, Lührig, K., Canbäck, B., Paul, C. J., Johansson, T., Persson, K. M., & Rådström, P. (2015). Bacterial Community Analysis of Drinking Water Biofilms in Southern Sweden. *Microbes and Environments*, *30*(1), 99–107. https://doi.org/10.1264/JSME2.ME14123

Luo, A., Zhang, A., Ho, S. Y. W., Xu, W., Zhang, Y., Shi, W., Cameron, S. L., & Zhu, C. (2011). Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC Genomics*, *12*, 84. https://doi.org/10.1186/1471-2164-12-84

Marcelino, V. R., Clausen, P. T. L. C., Buchmann, J. P., Wille, M., Iredell, J. R., Meyer, W., Lund, O., Sorrell, T. C., & Holmes, E. C. (2020). CCMetagen: Comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. *Genome Biology*, *21*(1), 1–15. https://doi.org/10.1186/S13059-020-02014-2

Matheson, C. D., Gurney, C., Esau, N., & Lehto, R. (2010). Assessing PCR Inhibition from Humic Substances. In *The Open Enzyme Inhibition Journal* (Vol. 3).

Mauvisseau, Q., Harper, L. R., Sander, M., Hanner, R. H., Kleyer, H., & Deiner, K. (2022). The Multiple States of Environmental DNA and What Is Known about Their Persistence in Aquatic Environments. *Environmental Science and Technology*, *56*(9), 5322–5333. https://doi.org/10.1021/ACS.EST.1C07638/SUPPL_FILE/ES1C07638_SI_006.XLSX

McDermott, J. J. (2011). Parasites of shore crabs in the genus Hemigrapsus (Decapoda: Brachyura: Varunidae) and their status in crabs geographically displaced: a review. *Journal of Natural History*, *45*(39/40), 2419–2441. https://doi.org/10.1080/00222933.2011.596636

McKee, A. M., Spear, S. F., & Pierson, T. W. (2015). The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation*, *183*, 70–76. https://doi.org/10.1016/J.BIOCON.2014.11.031

Meyer, D., Goettsch, W., Spannenberg, J., Stieber, B., Krautwurst, S., Olzer, M. H. ¨, Brandt, C., Org Linde, J. ¨, Oner Zu Siederdissen, C. H. ¨, Srivastava, A., Zarkovic, M., Wollny, D., & Marz, M. (2025). Unlocking the Full Potential of Nanopore Sequencing: Tips, Tricks, and Advanced Data Analysis Techniques. *BioRxiv*, *xx*, 2023.12.06.570356. https://doi.org/10.1101/2023.12.06.570356

Odom, A. R., Faits, T., Castro-Nallar, E., Crandall, K. A., & Johnson, W. E. (2023). Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. *Scientific Reports*, *13*(1). https://doi.org/10.1038/S41598-023-40799-X,

Ogata, M., Masuda, R., Harino, H., Sakata, M. K., Hatakeyama, M., Yokoyama, K., Yamashita, Y., & Minamoto, T. (2021). Environmental DNA preserved in marine sediment for detecting jellyfish blooms after a tsunami. *Scientific Reports*, *11*(1), 16830. https://doi.org/10.1038/S41598-021-94286-2

Oxford Nanopore Technologies. (2024). *Chemistry Technical Document (CHTD_500_v1_revAR_25Nov2024)*. https://nanoporetech.com/document/chemistry-technical-document

Palero, F., Ferrer-Mateu, I., Wray, B., Hughes, R., Morritt, D., Lepage, M., Kotterman, M., van der Meer, M., Tate, M., Kamanli, S. A., Smith, L., Llewellyn-Hughes, J., & Clark, P. F. (2022). Presence of a second Eriocheir species in Europe as confirmed by molecular and morphological data. *Aquatic Invasions*, *17*(3), 374–392. https://doi.org/10.3391/ai.2022.17.3.03

Panning, A. (1938). *THE CHINESE MITTEN CRAB*.

Pawlowski, J., Apothéloz-Perret-Gentil, L., Mächler, E., & Altermatt, F. (2020). *Environmental DNA applications in biomonitoring and bioassessment of aquatic ecosystems: Guidelines. [Environmental Studies No. 2010]*. Federal Office for the Environment (FOEN). https://doi.org/10.5167/uzh-187800

Peters, N. (1933). *Die chinesische Wollhandkrabbe (Eriocheir sinensis H. Milne-Edwards) in Deutschland (Zoologischer Anzeiger, Bd. 104, Ergänzungsband)* (N. Peters, A. Panning, & W. Schnakenbeck, Eds.). Akademische Verlagsgesellschaft m.b.H.

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., Barrio-Amorós, C. L., Salazar-Valenzuela, D., & Prost, S. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*, *7*(4), giy033. https://doi.org/10.1093/GIGASCIENCE/GIY033

Qin, C., Yang, B., Zhang, W., Ling, W., Liu, C., Liu, J., Li, X., & Gao, Y. (2019). Organochlorinated pesticides expedite the enzymatic degradation of DNA. *Communications Biology*, *2*(1), 1–9. https://doi.org/10.1038/S42003-019-0326-5

Rato, L. D., Simões, T., Novais, S. C., Damasceno, J. M., Van der Meer, J., Thieltges, D. W., Marques, J. C., & Lemos, M. F. L. (2024). Thermal performance of native and invasive crab species: investigating the invasion potential of Hemigrapsus takanoi in southern European Carcinus maenas' habitats. *Biological Invasions*, *26*(11), 3587–3601. https://doi.org/10.1007/S10530-024-03396-1/FIGURES/5

Reddington, K., Eccles, D., O'Grady, J., Drown, D. M., Hansen, L. H., Nielsen, T. K., Ducluzeau, A. L., Leggett, R. M., Heavens, D., Peel, N., Snutch, T. P., Bayega, A., Oikonomopoulos, S., Ragoussis, I., Barry, T., Van Der Helm, E., Jolic, D., Richardson, H., Jansen, H., … Brown, B. L. (2020). Metagenomic analysis of planktonic riverine microbial consortia using nanopore sequencing reveals insight into river microbe taxonomy and function. *GigaScience*, *9*(6), giaa053. https://doi.org/10.1093/gigascience/giaa053

Regulation (EU) No 1143/2014 of the European Parliament and of the Council of 22 October 2014 on the Prevention and Management of the Introduction and Spread of Invasive Alien Species, Pub. L. No. Regulation (EU) No 1143/2014, Official Journal of the European Union 35 (2014). https://eur-lex.europa.eu/legal-content/ENG/TXT/PDF/?uri=CELEX%3A32014R1143&utm

Robinson, C. V., de Leaniz, C. G., & Consuegra, S. (2019). Effect of artificial barriers on the distribution of the invasive signal crayfish and Chinese mitten crab. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/S41598-019-43570-3;TECHMETA=22,45,77;SUBJMETA=158,2178,2452,631;KWRD=INVASIVE+SPECIES,MOLECULAR+ECOLOGY

Sakata, M. K., Yamamoto, S., Gotoh, R. O., Miya, M., Yamanaka, H., & Minamoto, T. (2020). Sedimentary eDNA provides different information on timescale and fish species composition compared with aqueous eDNA. *Environmental DNA*, *2*(4), 505–518. https://doi.org/10.1002/EDN3.75;JOURNAL:JOURNAL:26374943

Sauvage, T., Cormier, A., & Delphine, P. (2023). A comparison of Oxford nanopore library strategies for bacterial genomics. *BMC Genomics*, *24*(1), 1–16. https://doi.org/10.1186/S12864-023-09729-Z/TABLES/6

Serite, C. P., Emami-Khoyi, A., Ntshudisane, O. K., James, N. C., Jansen van Vuuren, B., Bodill, T., Cowley, P. D., Whitfield, A. K., & Teske, P. R. (2023). eDNA metabarcoding vs metagenomics: an assessment of dietary competition in two estuarine pipefishes. *Frontiers in Marine Science*, *10*, 1116741. https://doi.org/10.3389/FMARS.2023.1116741

Shao, W., Khin, S., & Kopp, W. C. (2012). Characterization of Effect of Repeated Freeze and Thaw Cycles on Stability of Genomic DNA Using Pulsed Field Gel Electrophoresis. *Biopreservation and Biobanking*, *10*(1), 4. https://doi.org/10.1089/BIO.2011.0016

Shen, Y. Y., Chen, X., & Murphy, R. W. (2013). Assessing DNA Barcoding as a Tool for Species Identification and Data Quality Control. *PLOS ONE*, *8*(2), e57125. https://doi.org/10.1371/JOURNAL.PONE.0057125

SLU Artdatabanken. (2025a). *Artfakta: Eriocheir japonica*. Artfakta (SLU Artdatabanken). https://artfakta.se/taxa/6330418/information

SLU Artdatabanken. (2025b). *Artfakta: Eriocheir sinensis*. Artfakta (SLU Artdatabanken). https://artfakta.se/taxa/217771/information

SLU Artdatabanken. (2025c). *Artportalen: Fyndkarta Eriocheir sinensis,2010–2025*. LU Artdatabanken. https://www.artportalen.se/ViewSighting/ViewSightingAsMap

Stark, P. C., Mullen, K. I., Banton, K., Russotti, R., Soran, D., & Kuske, C. R. (2000). Pre-PCR DNA quantitation of soil and sediment samples: method development and instrument design. *Soil Biology and Biochemistry*, *32*(8–9), 1101–1110. https://doi.org/10.1016/S0038-0717(00)00020-1

Stevens, B. M., Creed, T. B., Reardon, C. L., & Manter, D. K. (2023). Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. *Scientific Reports*, *13*(1), 1–11. https://doi.org/10.1038/S41598-023-36101-8;SUBJMETA=208,2254,2565,326,41,514,61,631;KWRD=BACTERIA,MICROBIAL+COMMUNITIES,NEXT-GENERATION+SEQUENCING

Sundelöf, A., Florin, A.-B., Rogell, B., Bolund, E., Vitale, F. S., Sundblad, G., Strömberg, H., Ahlbeck Bergendahl, I., Sundin, J., Lundström, K., Wikström, K., Magnusson, K., Fetterplace, L., Wennerström, L., Ogonowski, M., Bergenius Nord, M., Holmgren, N., Kaljuste, O., Bohman, P., … Östman, Ö. (2021). *Fisk- och skaldjursbestånd i hav och sötvatten 2021 - Resursöversikt*. Havs- och vattenmyndigheten/SLU. https://www.diva-portal.org/smash/get/diva2%3A1733825/FULLTEXT01.pdf?utm_

Svoboda, J., Strand, D. A., Vrålstad, T., Grandjean, F., Edsman, L., Kozák, P., Kouba, A., Fristad, R. F., Bahadir Koca, S., & Petrusek, A. (2014). The crayfish plague pathogen can infect freshwater-inhabiting crabs. *Freshwater Biology*, *59*(5), 918–929. https://doi.org/10.1111/FWB.12315

Thermo Fisher Scientific. (2015). *Assessment of Nucleic Acid Purity (Tech. Bull. T042)*. www.thermoscientific.com

Veilleux, É., & de Lafontaine, Y. (2007). Biological synopsis of the Chinese mitten crab (Eriocheir sinensis). In *Fisheries and Oceans Canada*. https://publications.gc.ca/site/eng/9.618621/publication.html

Veldhoen, N., Hobbs, J., Ikonomou, G., Hii, M., Lesperance, M., & Helbing, C. C. (2016). Implementation of Novel Design Features for qPCR-Based eDNA Assessment. *PLoS ONE*, *11*(11), e0164907. https://doi.org/10.1371/JOURNAL.PONE.0164907

Wang, Y., Hu, Y., Liu, S., Hu, X., Chen, W., & Xiao, M. (2024). Research on Undisturbed Monitoring of Eriocheir Sinensis based on Biological Habits. *2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 1547–1551. https://doi.org/10.1109/ICSP62122.2024.10743245

Wick, R. R., Judd, L. M., Wyres, K. L., & Holt, K. E. (2021). Recovery of small plasmid sequences via Oxford Nanopore sequencing. *Microbial Genomics*, *7*(8), 000631. https://doi.org/10.1099/MGEN.0.000631

Xu, Y., Lewandowski, K., Lumley, S., Pullan, S., Vipond, R., Carroll, M., Foster, D., Matthews, P. C., Peto, T., & Crook, D. (2018). Detection of viral pathogens with multiplex nanopore MinION sequencing: Be careful with cross-Talk. *Frontiers in Microbiology*, *9*(SEP), 390055. https://doi.org/10.3389/FMICB.2018.02225/BIBTEX

Appendix – Supplementary material

Table S1. Temperatures measured of the first aquarium experiment.

Date	Aq1 (°C)	Aq2 (°C)	Aq3 (°C)	OC (°C)	NC (°C)
10/03/2025	16	17	17	16	16
13/03/2025	17	18	18	17	16
14/03/2025	16	17	18	17	16
18/03/2025	17	18	18	17	17
19/03/2025	16	17	17	16	16

Table S2. Specimens of CMC used in this experiment. Each specimen was stored at -20˚C until legs were broken off to be used in the experiment.

Specimen	Date gathered	Location gathered	Gender	Aquarium experiment
Specimen 1	08/11/2024	Kinneviken, Vänern	Male	1
Specimen 2	16/09/2024	Kristinehamn, Vänern	Female	1 & 2
Specimen 3	16/10/2024	N.d.	Male	1 & 2
Specimen 4	23/09/2024	Kinneviken, Vänern	Male	2

Table S3. Temperatures measured of the second aquarium experiment.

Date	NAq1 (°C)	NAq2 (°C)	NAq3 (°C)	NNC (°C)
13/04/2025	17	18	18	17
14/04/2025	16	17	17	16
18/04/2025	17	17	18	16
19/04/2025	16	16	17	16

Figure S1. Picture taken from the small peninsula looking out towards the bay at Stenhammar. Photo taken by Sonja Leidenberger.

Figure S2. Picture taken from the water looking in towards the peninsula at Stenhammar. Photo taken by Sonja Leidenberger.

Figure S3. Picture taken from shore showing the bottom of the water on the southern side of the peninsula at Stenhammar. Photo taken by Sonja Leidenberger.

Figure S4. Picture taken from land, looking east, showing how the Sannorna location looked like. Photo taken by Felix Östervald.

Table S4. Field samples taken along the shore of the area of Kinneviken in lake Vänern, Sweden.

Sample	Sediment	Depth	Temperature	Location
Stenhammar 1	Sand	30-40 cm	8.5˚C	58.52667, 13.13511

Stenhammar 2	Sand	30-40 cm	8.5°C	58.52691, 13.13543
Stenhammar 3	Clay	Ca 1 m	8°C	58.52728, 13.13594
Stenhammar 4	Sand/Gravel	60-80 cm	8°C	58.52740, 13.13658
Sannorna 1	Sand	60-80 cm	9.5°C	58.50205, 13.19246
Sannorna 2	Sand	Ca 20 cm	9.5°C	58.50164, 13.19392
Sannorna 3	Sand	40-50 cm	9.5°C	58.50127, 13.19488
Sannorna 4	Sand	Ca 1 m	9.5°C	58.50289, 13.19111

Table S5. Amount used of each water extracted sample to reach the 400 ng in 11 µl for the sequencing with the Ligation sequencing gDNA - Native Barcoding Kit 96 V14 (SQK-NBD114.96) kit (Oxford Nanopore Technologies). Numbers marked with an * in the third row indicate samples that could not reach the required mass in the required volume, hence the full volume of 11 µl was used. Codes: Aq1 = 1 CMC leg; Aq2 = 3 CMC legs; Aq3 = 5 CMC legs; T1–T3 denote sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); OC = other-crabs control; NC = Negative Control; UW = sample performed without additional wash step.

Sample	Qubit concentration (ng/µl)	Volume to reach ≤400 ng (µl)	Volume of water to reach 11 µl	Mass of DNA (ng)
Aq1 T1	35.2	11*	0	387.2
Aq2 T1	58.0	6.9	4.1	400
Aq3 T1	19.0	11*	0	209
Aq1 T2	21.0	11*	0	231
Aq2 T2	34.4	11*	0	378.4
Aq3 T2	33	11*	0	363
Aq1 T3	16.4	11*	0	180.4
Aq2 T3	49.3	8.11	2.89	400
Aq3 T3	8.36	11*	0	91.96
OC	36.3	11*	0	399.3
NC	23.7	11*	0	260.7
Aq3 T3 (UW)	57.0	7.02	3.98	400

Table S6. Concentration, measured with Qubit 4 fluorometer using 1X HS DNA Assay, of the sediment samples. Calculated volume required to reach 20 ng/µl after magnetic bead cleaning, the possibility of reaching that volume in ≤1000 µl and the concentration, measured with Qubit 4 fluorometer using 1X HS DNA Assay, after magnetic bead cleaning. Codes: NAq1 = 1 CMC leg; NAq2 = 3 CMC legs; NAq3 = 5 CMC legs; T1–T2 denote sequential sampling time points within an experiment (T1 = first, T2 = second); Sannorna 1-4 = field samples from Sannorna; Stenhammar 1-4 = field samples from Stenhammar.

Sample	Qubit before Bead cleaning (ng/µl)	Required Volume (µl)	Possible	Qubit after Bead cleaning (ng/µl)
NAq1 T1	0.096	5208.33	No	5.42
NAq2 T1	0.17	2941.18	No	8.44
NAq3 T1	0.248	2016.13	No	14.9
NNC T1	0.069	7246.38	No	3.16
NAq1 T2	0.349	1432.66	No	12.5
NAq2 T2	0.887	563.7	Yes	34.4

NAq3 T2	1.25	400	Yes	27.6
Aq1 T1	0.43	1162.79	No	20.6
Aq2 T1	0.949	526.87	Yes	23.0
Aq3 T1	0.961	520.29	Yes	24.6
Aq1 T2	0.519	963.39	Yes	16.5
Aq2 T2	0.865	578.03	Yes	17.6
Aq3 T2	0.866	577.37	Yes	17.4
Aq1 T3	0.265	1886.79	No	9.00
Aq2 T3	0.515	970.87	Yes	15.0
Aq3 T3	0.789	633.71	Yes	17.7
OC	0.513	974.66	Yes	14.9
NC	0.505	990.1	Yes	11.2
Sannorna 1	8.42	59.38	Yes	31.8
Sannorna 2	4.42	113.12	Yes	57.0
Sannorna 3	6.31	79.24	Yes	37.4
Sannorna 4	8.17	61.2	Yes	26.6
Stenhammar 1	9.69	51.6	Yes	26.2
Stenhammar 2	11.6	43.1	Yes	26.2
Stenhammar 3	0.196	2551.02	No	3.08
Stenhammar 4	0.645	775.19	Yes	21.8

Table S7. K-mer, minimizer lengths, and number of minimizer-spaces for the different custom COI and MitoGun databases.

Name	K-mer length	Minimizer length	Minimizer-spaces
COI_k35_m31_ms7	35	31	7
COI_k31_m25_ms6	31	25	6
COI_k27_m21_ms5	27	21	5
COI_k25_m19_ms4	25	19	4
COI_k21_m17_ms4	21	17	4
COI_k19_m15_ms2	19	15	2
COI_k17_m13_ms1	17	13	1

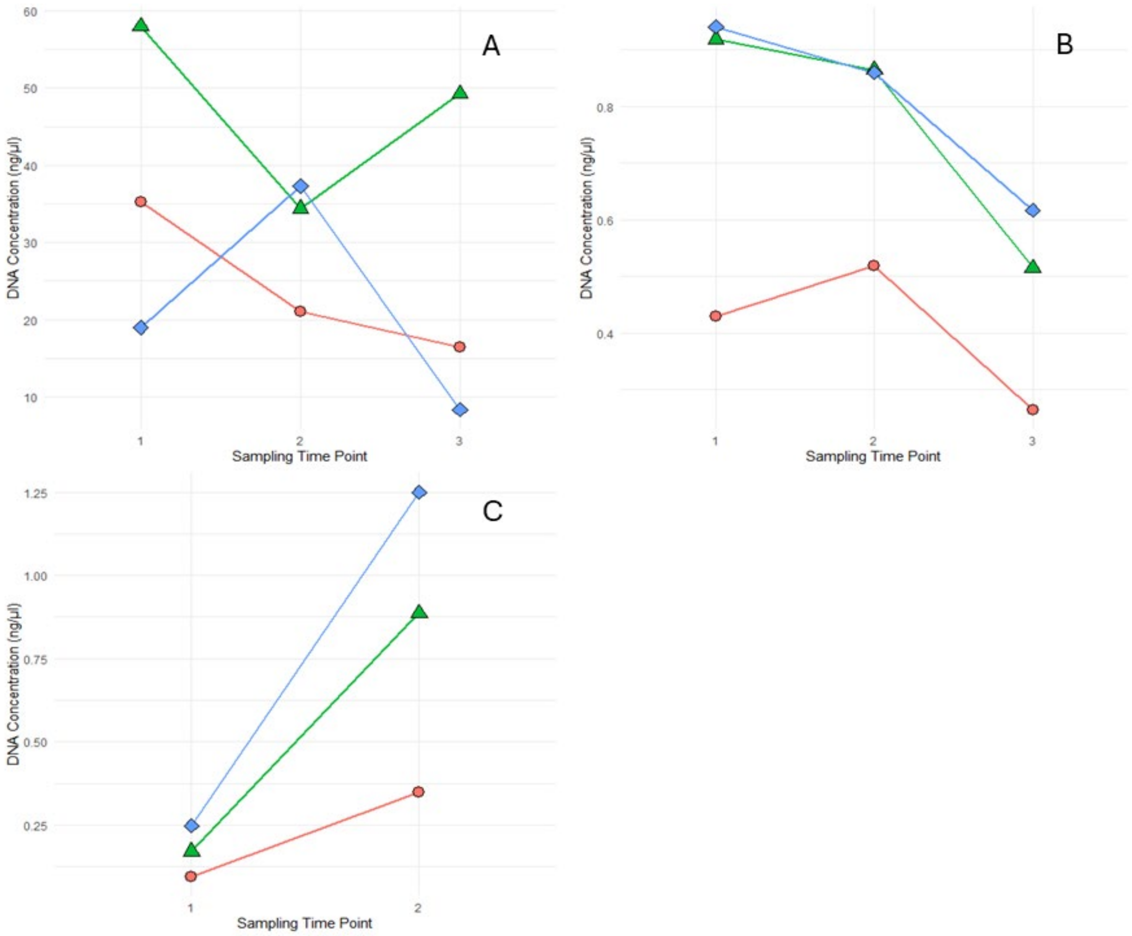

Figure S5. Line plots representing the DNA concentrations of the eDNA samples from the aquarium experiments. The x-axis shows the sampling time points, while the y-axis shows the DNA concentration in ng/µl. The images represent, A; the water samples from the first aquarium experiment, B; the sediment samples from the first aquarium experiment, C; the sediment samples from the second aquarium experiment. Within each plot the lines with the following symbols represent, ●; the Aq1/NAq1 aquariums, ▲; the Aq2/NAq2 aquariums, ◆; the Aq3/NAq3 aquariums.

Table S8. DNA concentration and purity, measured using Qubit fluorometer 4 with the 1x HS dsDNA assay and Nanodrop spectrophotometer respectively, of the water samples extracted using NucleoSpin® eDNA Water kit (Macherey-Nagel, 740402.50) following the standard protocol. Codes: Aq1 = 1 CMC leg; Aq2 = 3 CMC legs; Aq3 = 5 CMC legs; T1–T3 denote sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); OC = other-crabs control.

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
Aq3 T3	57.0	186.2	2.0	1.18
Aq2 T3	50.0	143.8	1.97	1.19
Aq1 T3	13.6	47.4	1.96	0.56
OC	47.4	137.9	1.97	0.95

Table S9. DNA concentration and purity, measured using Qubit fluorometer 4 with the 1x HS dsDNA assay and Nanodrop spectrophotometer respectively, of the water samples extracted using NucleoSpin® eDNA Water kit (Macherey-Nagel, 740402.50) following the optimized protocol. Codes: Aq1 = 1 CMC leg; Aq2 = 3 CMC legs; Aq3 = 5 CMC legs; T1–T3 denote sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); OC = other-crabs control; NC = Negative Control; UW = sample performed without additional wash step.

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
Aq1 T1	35.2	70.0	1.92	1.53
Aq2 T1	58.0	208.4	1.99	1.83
Aq3 T1	19.0	63.1	1.94	1.47
Aq1 T2	21.0	55.3	1.92	1.25
Aq2 T2	34.4	83.4	1.88	1.45
Aq3 T2	37.3	89.2	2.00	1.49
Aq1 T3	16.4	44.2	1.92	1.11
Aq2 T3	49.3	124.4	2.06	2.07
Aq3 T3	8.36	20.2	1.95	1.21
OC	36.5	109.5	2.03	2.15
NC	23.7	50.4	1.91	0.94
Aq3 T3 UW	57.0	186.2	2.00	1.18

Table S10. Concentration and purity, measured using Qubit fluorometer 4 with the 1x HS dsDNA assay and Nanodrop spectrophotometer respectively, of the Sediment samples extracted using DNeasy PowerSoil Pro kit using the standard protocol. Codes: Aq1 = 1 CMC leg; Aq2 = 3 CMC legs; Aq3 = 5 CMC legs; T1–T3 denote sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); OC = other-crabs control.

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
Aq1 T3	Too low	-1.5	0.83	-0.03
Aq2 T3	Too low	-0.9	0.63	-0.02
Aq3 T3	Too low	-3.3	1.11	-0.39
OC	Too low	-2.8	1.16	0.64

Table S11. Concentration and purity, measured using Qubit fluorometer 4 with the 1x HS dsDNA assay and Nanodrop spectrophotometer respectively, of the sediment samples extracted using DNeasy PowerSoil Pro kit the optimized protocol. Codes: Aq1 = 1 CMC leg; Aq2 = 3 CMC legs; Aq3 = 5 CMC legs; T1–T3 denote sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); OC = other-crabs control; NC = Negative Control.

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
Aq1 T1	0.127	0.9	-25.08	0.00
Aq2 T1	0.176	1.2	-23.48	0.01
Aq3 T1	0.100	2.1	20.16	0.00
Aq1 T2	Too low	59.4	1.08	0.57
Aq2 T2	Too low	1.5	1.75	0.00
Aq3 T2	0,0980	0.1	-0.89	0.00
Aq1 T3	Too low	1.015	0.66	0.05
Aq2 T3	Too low	1.698	0.95	0.01
Aq3 T3	Too low	1.296	0.71	0.19
OC	Too low	1.120	0.55	0.02
NC	0.0580	-0.6	2.95	-0.04
Aq3 T3	0.144	0.4	2.19	0.01

Table S12. Concentration and purity measurements, measured using Qubit fluorometer 4 with the 1x HS dsDNA assay and Nanodrop spectrophotometer respectively, of sediment samples extracted using DNeasy PowerMax Soil kit with the optimized protocol. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
Aq1 T1	0.430	2.2	40.35	0.17
Aq2 T1	0.919	1.9	-4.11	0.16
Aq3 T1	0.941	1.6	-4.55	0.13
Aq1 T2	0.519	3.1	2.51	0.12
Aq2 T2	0.865	14.1	1.13	1.74
Aq3 T2	0.860	4.2	2.39	0.15
Aq1 T3	0.265	24.3	1.29	1.41
Aq2 T3	0.515	6.1	3.22	0.21
Aq3 T3	0.617	5.4	3.10	0.17
OC	0.513	4.0	1.85	0.07
NC	0.505	2.0	-1.47	0.12
NAq1 T1	0.0960	2.0	3.45	0.11
NAq2 T1	0.171	2.5	2.19	0.12
NAq3 T1	0.248	2.2	4.21	0.10

NAq1 T2	0.349	3.9	2.36	0.20
NAq2 T2	0.887	4.4	2.72	0.20
NAq3 T2	1.25	5.4	2.28	0.21
NNC	0.0690	2.5	3.03	0.11
Sannorna 1	8.42	10.5	1.82	0.50
Sannorna 2	9.42	13.9	2.23	0.56
Sannorna3	6.31	24.5	2.39	0.58
Sannorna 4	8.17	60.1	1.29	1.39
Stenhammar 1	9.69	14.5	2.06	0.48
Stenhammar 2	11.6	20.1	0.63	0.79
Stenhammar 3	0.196	1.7	10.08	0.09
Stenhammar 4	0.645	2.7	3.91	0.14

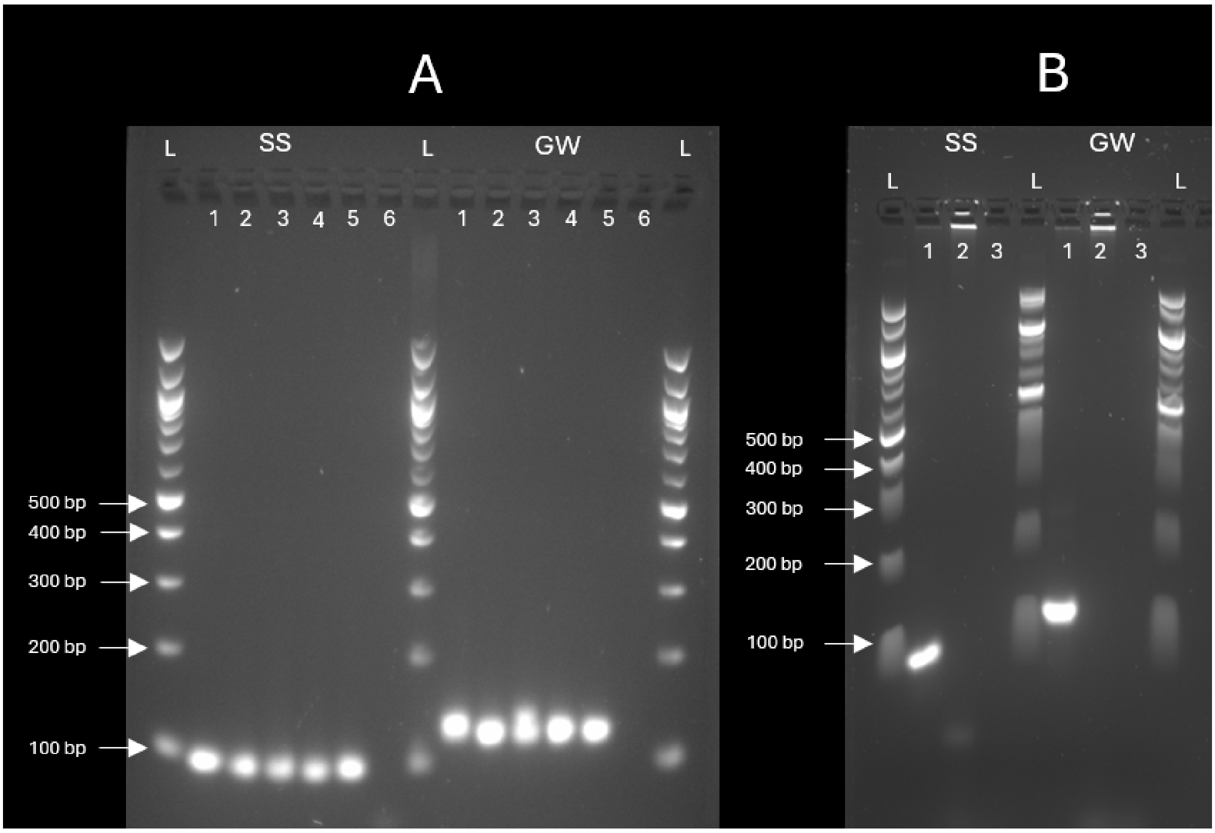

Figure S6. Gel electrophoresis images of PCR products generated with two primer pairs: the species-specific (SS) and genus-wide (GW) primers. (A) Gradient PCR at five annealing temperatures (T_a): 1 = 50 °C, 2 = 53 °C, 3 = 56 °C, 4 = 60 °C, and 5 = 65 °C. L = Quick-Load® Purple 100 bp DNA Ladder. Wells numbered 6 contain non-template controls (NTCs). (B) PCR amplification of DNA from two crab species. L = Quick-Load® Purple 100 bp DNA Ladder (NEB). 1 = *Eriocheir sinensis* (CMC) DNA, 2 = *Pagurus bernhardus* DNA, 3 = non-template control (NTC).

Table S13. Concentration and purity measurements, measured with Qubit 4 fluorometer, with 1x HS DNA assay, and Nanodrop spectrophotometer respectively, for the gradient PCR with KAPA HiFi HotStart Ready mix of both the Species-Specific (SS) and Genus-Wide (GW) primer pairs.

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
SS 50˚C	14.2	644.2	1.84	1.79
SS 53˚C	14.4	239.0	1.30	1.27
SS 55˚C	14.0	662.0	1.84	1.79

SS 60˚C	13.6	699.4	1.83	1.80
SS 65˚C	13.2	234.8	1.54	1.26
SS NTC	0.949	653.0	1.86	1.76
GW 50˚C	25.7	668.3	1.84	1.75
GW 53˚C	25.7	663.2	1.83	1.78
GW 56˚C	25.8	621.8	1.84	1.79
GW 60˚C	26.1	699.2	1.80	1.82
GW 65˚C	22.5	644.6	1.82	1.72
GW NTC	0.696	633.2	1.86	1.82

Table S14. Concentration and purity measurements, measured with Qubit 4 fluorometer, with 1x HS DNA assay, and Nanodrop spectrophotometer respectively, for the PCR with KAPA HiFi HotStart Ready mix of both the Species-Specific (SS) and Genus-Wide (GW) primer pairs checking the specificity of the primer pairs. CMC = Chinese Mitten Crab (*E. sinensis*).

Sample	Qubit concentration (ng/µl)	Nanodrop concentration (ng/µl)	260/280	260/230
SS CMC	22.1	538.9	1.50	1.41
SS *P. bernhardus*	10.4	705.4	1.78	1.83
SS NTC	0.218	819.8	1.79	1.92
GW CMC	36.0	557.6	1.83	1.87
GW *P. bernhardus*	10.7	646.3	1.78	1.91
GW NTC	0.595	676.0	1.77	1.97

Table S15. Cq values and T_m of the standard curves for the Species-Specific (SS) and Genus-Wide (GW) primer, with the threshold set to 0.7.

Well Name	Cq (ΔRn)	Tm Product 1 (-Rn'(T)) (˚C)
1:1 SS	No Cq	74
1:10 SS	No Cq	74
1:100 SS	13.06	74
1:1000 SS	15.06	74
1:10000 SS	19.25	74
1:100000 SS	23.26	74
1:1000000 SS	27.15	74
NTC SS	No Cq	74
1:1 GW	No Cq	75
1:10 GW	No Cq	75
1:100 GW	15.68	75
1:1000 GW	18.84	75
1:10000 GW	23.37	75
1:100000 GW	28.54	75.5
1:1000000 GW	32.07	75
NTC GW	No Cq	75.5

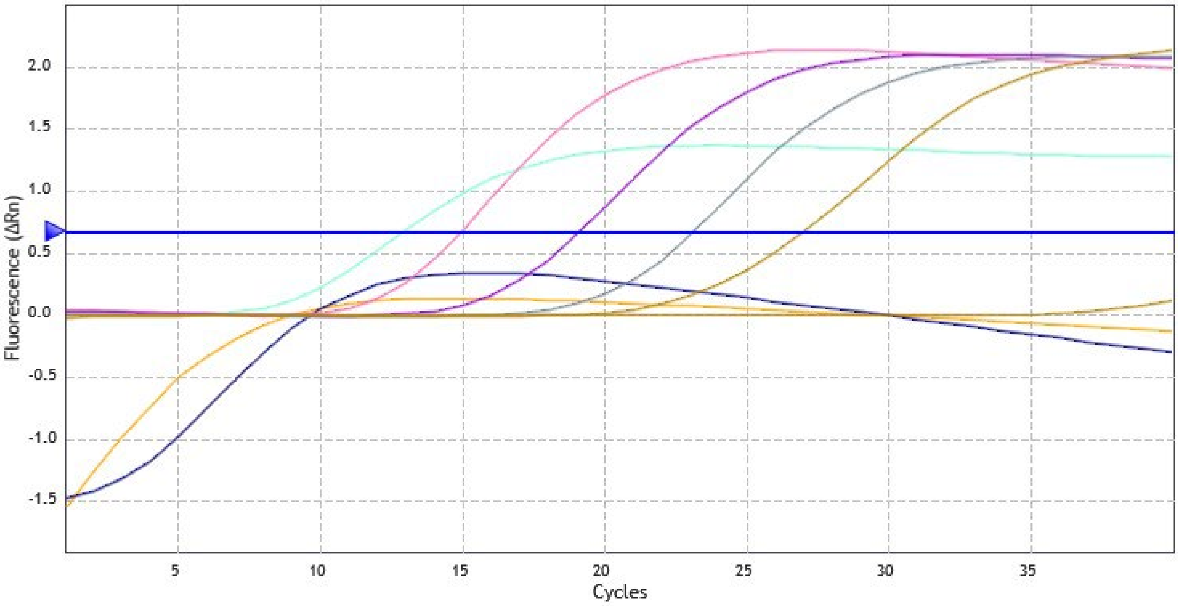

Figure S7. Amplification plots of the first attempt to create a standard curve with the Species-Specific (SS) primer pair. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn).Three technical replicates were used for each sample.

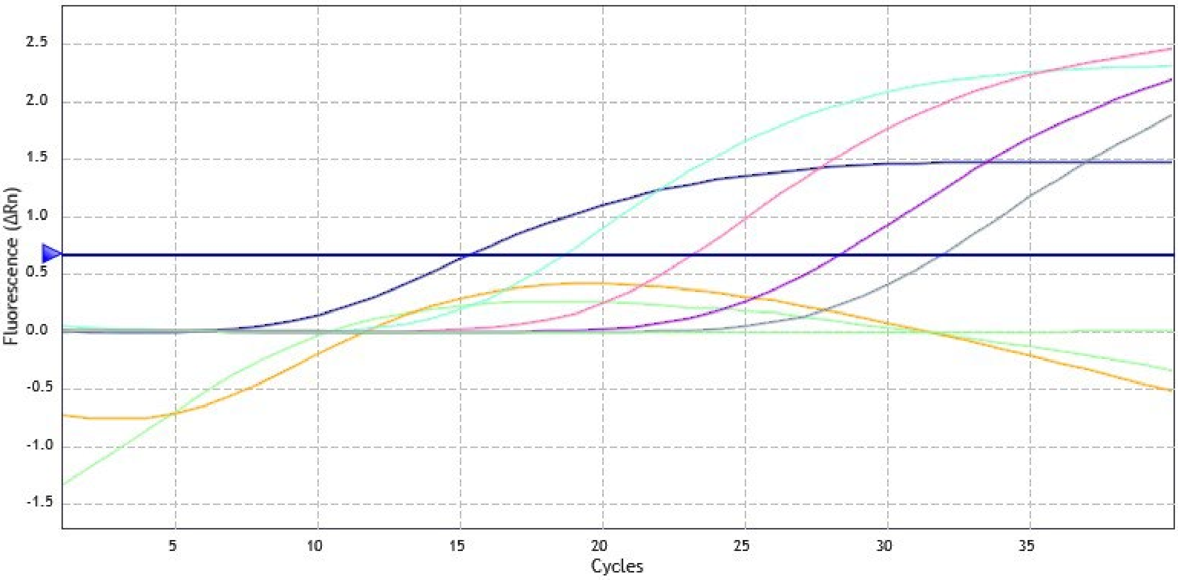

Figure S8. Amplification plots of the attempt to create a standard curve with the Genus-Wide (GW) primer pair. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

Table S16. Cq values and T_m of the second standard curve for the Species-Specific (SS) primer pair, with the threshold set to 0.7. Three technical replicates were used for each sample.

Well name	Cq (ΔRn)	Tm Product 1 (-Rn'(T)) (˚C)
1e^{-1} SS	13.36	74
1e^{-2} SS	17.41	74
1e^{-3} SS	21.81	74
1e^{-4} SS	25.92	74
1e^{-5} SS	30.87	74
1e^{-6} SS	35.41	74
1e^{-7} SS	38.84	74

1e-8 SS	No Cq	74
NTC	No Cq	74

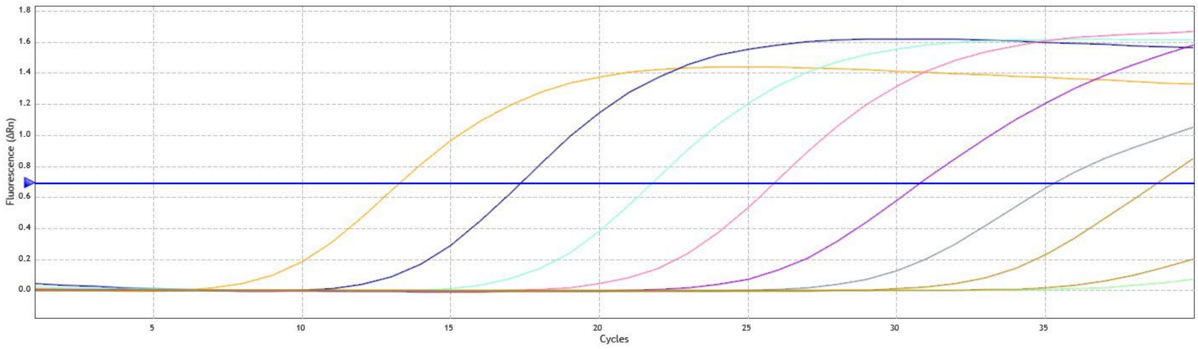

Figure S9. Amplification plots of the second attempt to create a standard curve with the Species-Specific (SS) primer pair. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

Table S17. Cq values and T_m of the 5-fold and 10-fold diluted sediment samples from the second aquarium experiment and 5-fold and 10-fold diluted field samples, with the threshold set to 0.15. Three technical replicates were used for each sample. Codes: Naq = Aquarium (Experiment 2); Naq1 – Naq3 denotes sequential increase of number of crab legs (Naq1 = 1 CMC leg, Naq2 = 3 CMC legs, Naq3 = 5 CMC legs); NNC = negative controls; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; (5) = 5-fold diluted samples; (20) = 20-fold diluted samples.

Well Name	Cq (ΔRn)	Tm Product 1 (-Rn'(T)) (°C)
NAq1 T1 (5)	38.72	73
NAq2 T1 (5)	32.14	73
NAq3 T1 (5)	33.72	72.5
NAq1 T2 (5)	No Cq	72.5
NAq2 T2 (5)	35.89	72.5
NAq3 T2 (5)	36.19	71.5
NNC (5)	No Cq	68.5
NAq1 T1 (20)	37.6	73.5
NAq2 T1 (20)	33.55	73.5
NAq3 T1 (20)	33.23	73.5
NAq1 T2 (20)	No Cq	73.5
NAq2 T2 (20)	36.26	73.5
NAq3 T2 (20)	36.23	73.5
NNC (20)	No Cq	73
Sa 1 (5)	No Cq	63
Sa 2 (5)	No Cq	63
Sa 3 (5)	No Cq	63.5
Sa 4 (5)	No Cq	72.5
SH 1 (5)	No Cq	63
SH 2 (5)	No Cq	69.5
SH 3 (5)	No Cq	72.5
SH 4 (5)	No Cq	73
Sa 1 (20)	No Cq	64

Sa 2 (20)	No Cq	64
Sa 3 (20)	No Cq	70.5
Sa 4 (20)	No Cq	63.5
SH 1 (20)	No Cq	63.5
SH 2 (20)	No Cq	63
SH 3 (20)	No Cq	65
SH 4 (20)	No Cq	73.5
NTC	No Cq	63.5

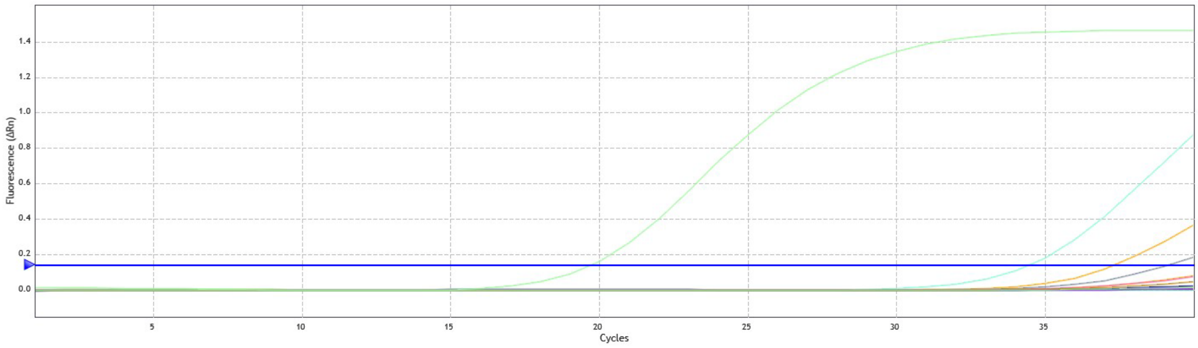

Figure S10. Amplification plots of the sediment samples from the first aquarium experiment diluted 10-fold, using the Species-Specific (SS) primer pair. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

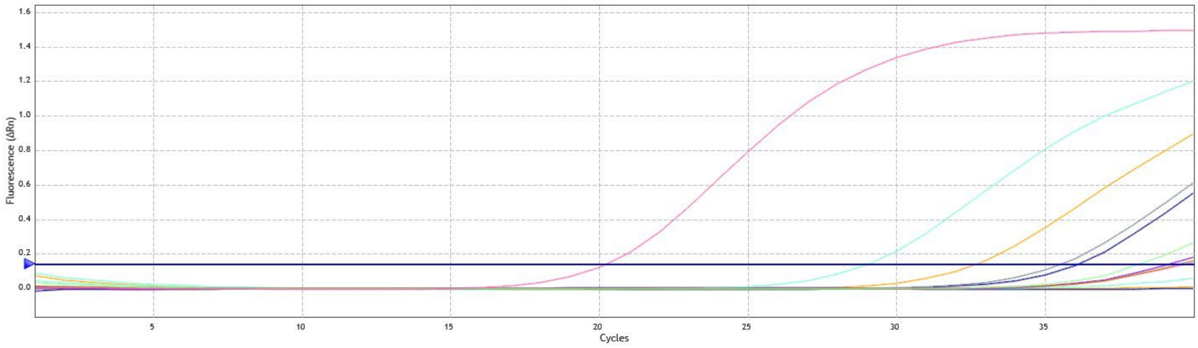

Figure S11. Amplification plots of the water samples diluted 10-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

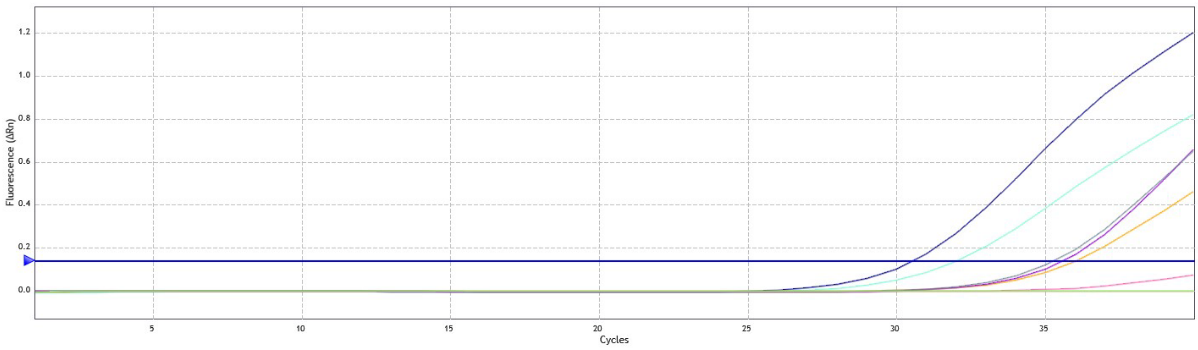

Figure S12. Amplification plots of the sediment samples from the second aquarium experiment diluted 10-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

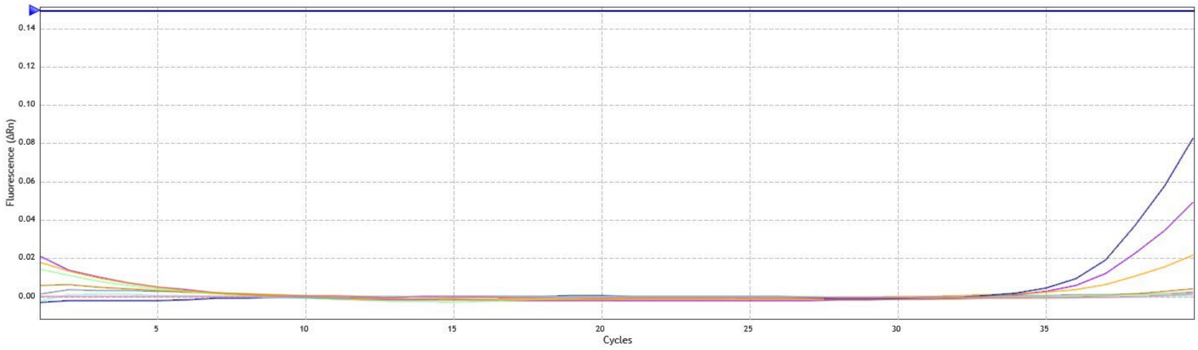

Figure S13. Amplification plots of the field samples diluted 10-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

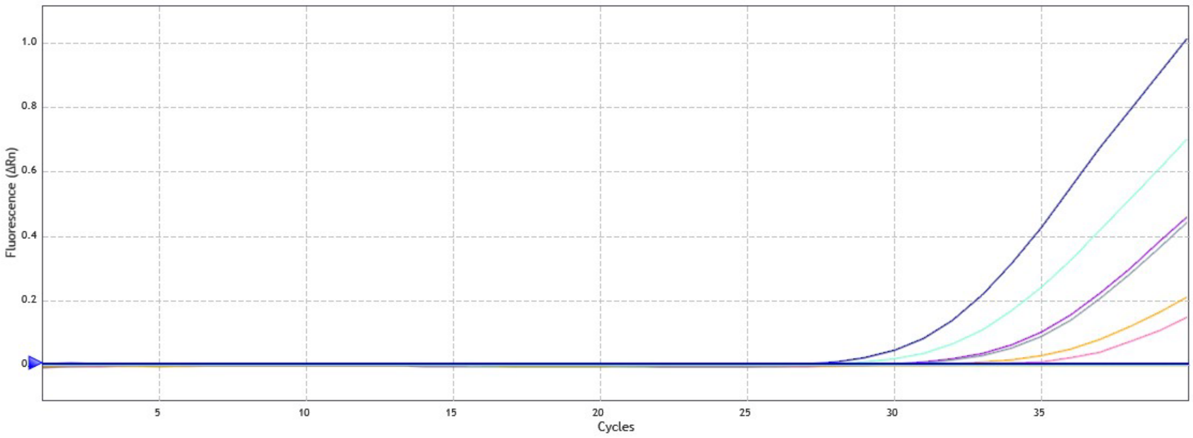

Figure S14. Amplification plots of the sediment samples from the second aquarium experiment diluted 5-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

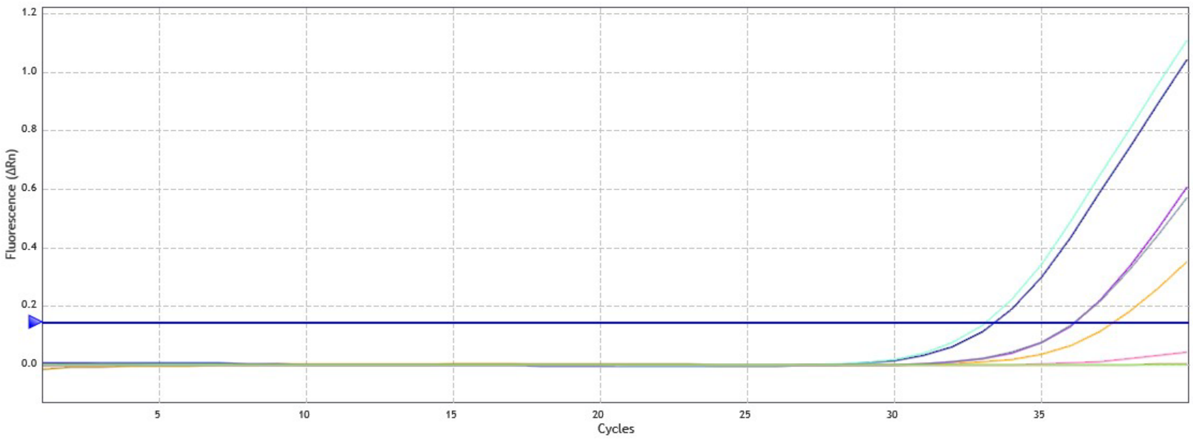

Figure S15. Amplification plots of the sediment samples from the second aquarium experiment diluted 20-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

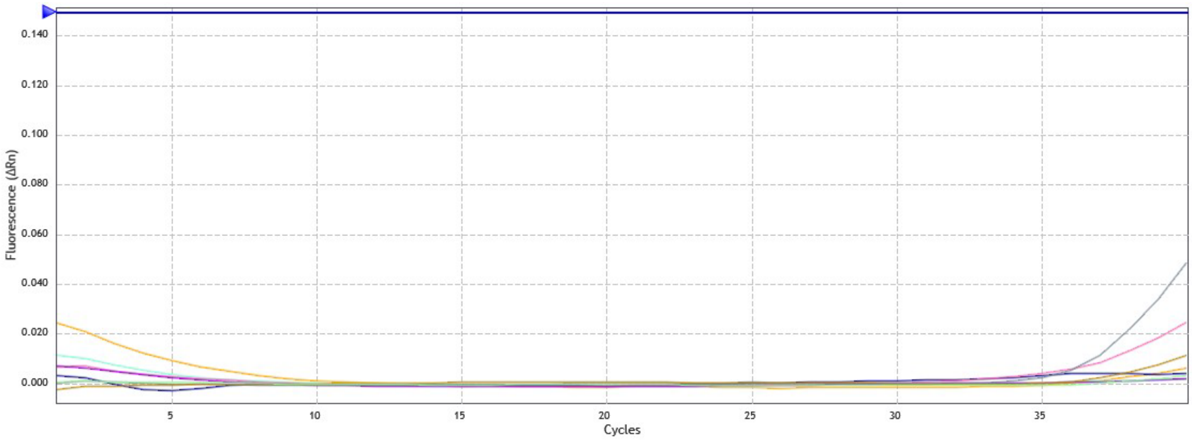

Figure S16. Amplification plots of the field samples diluted 5-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

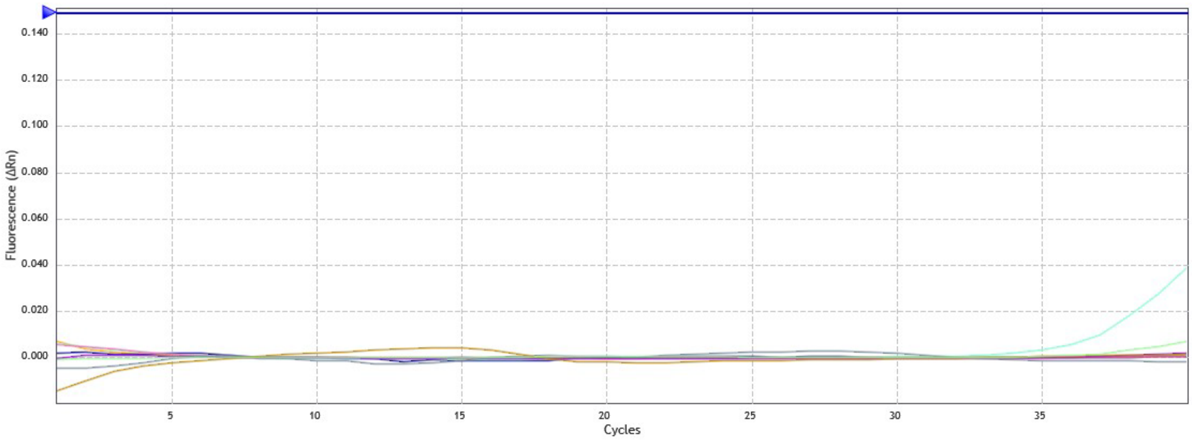

Figure S17. Amplification plots of the field samples diluted 20-fold. The x-axis shows the number of cycles, while the y-axis shows the fluorescence (ΔRn). Three technical replicates were used for each sample.

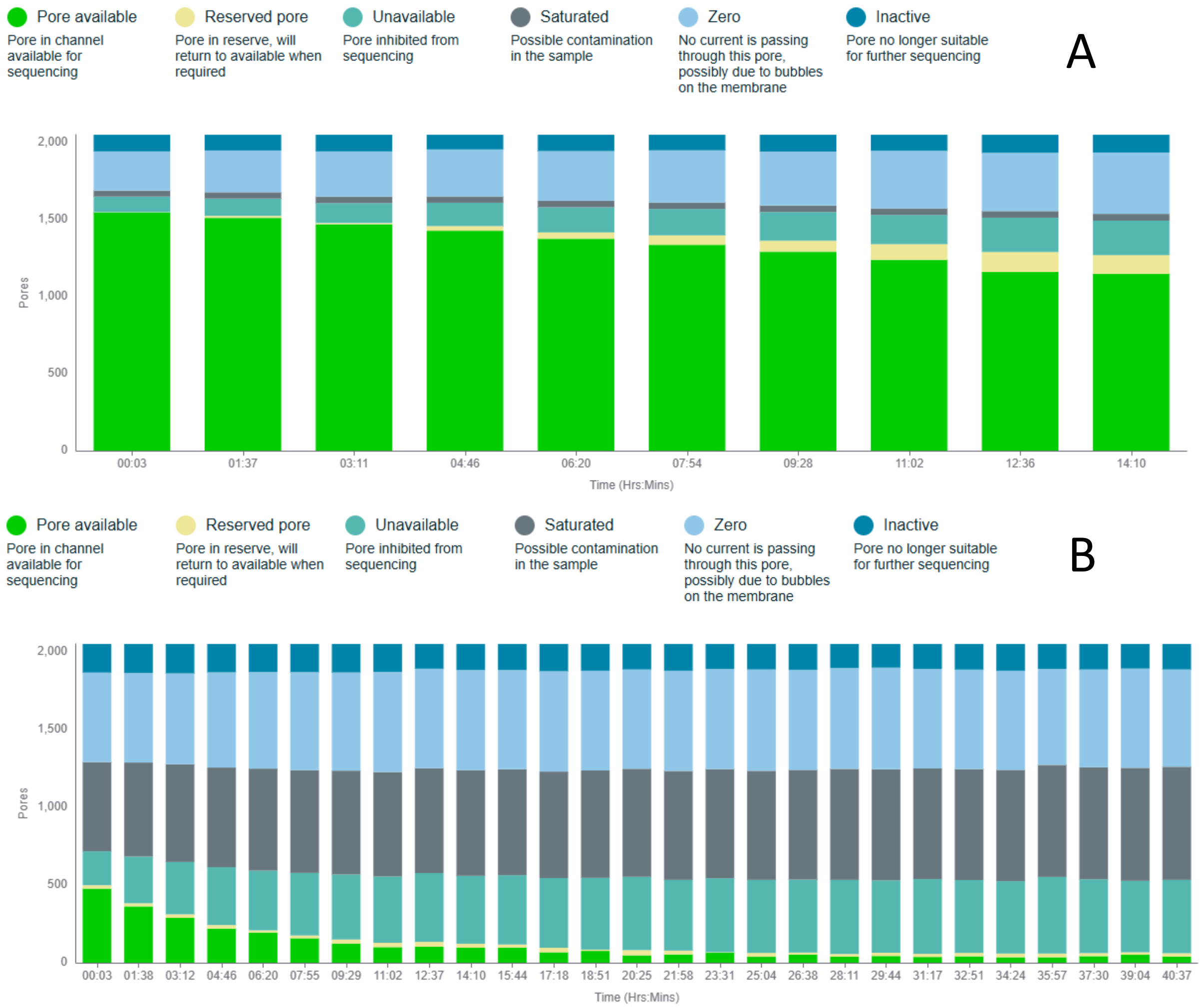

Figure 18. Pore activity scans of the flow cell during nanopore sequencing. (A) Sequencing of water samples. (B) Sequencing of sediment samples. In both panels, the x-axis represents the number of active pores, and the y-axis represents the sequencing runtime in hours and minutes.

Table S18. Classified reads of the water samples using the custom whole genome database. The first row shows the confidence level, which ranged from 0.0 to 1.0. The second row shows which species the reads belong to, CMC; *Eriocheir sinensis*, *P. bo*; *Pandalus borealis*. The cells filled with green indicate reads expected to be present/absent, while the cells filled with red indicate reads not expected to be present/absent. Codes: Aq = Aquarium (Experiment 1); Aq1 – Aq3 denotes sequential increase of number of crab legs (Aq1 = 1 CMC leg, Aq2 = 3 CMC legs, Aq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

	0		**0.1**		**0.2**		**0.3**		**0.4**		**0.5**		**0.6**		**0.7**		**0.8**		**0.9**		**1**	
Sample	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***
Aq1 T1	2235	38	1348	0	702	0	401	0	305	0	238	0	180	0	124	0	69	0	31	0	11	0
Aq2 T1	939	62	427	0	143	0	34	0	8	0	3	0	3	0	0	0	0	0	0	0	0	0
Aq3 T1	1086	32	744	0	496	0	375	0	327	0	271	0	207	0	136	0	55	0	17	0	7	0
Aq1 T2	2206	263	136	2	47	0	12	0	8	0	6	0	4	0	3	0	2	0	1	0	0	0
Aq2 T2	1956	166	311	0	105	0	41	0	17	0	6	0	5	0	1	0	1	0	1	0	1	0
Aq3 T2	3829	422	301	7	95	3	48	1	33	0	24	0	18	0	14	0	6	0	3	0	1	0
Aq1 T3	741	130	63	0	21	0	12	0	9	0	7	0	5	0	5	0	1	0	0	0	0	0
Aq2 T3	1283	215	127	2	49	0	18	0	10	0	7	0	3	0	2	0	1	0	1	0	0	0
Aq3 T3	1462	107	337	7	94	3	41	2	25	1	12	0	9	0	4	0	0	0	0	0	0	0
NC	697	21	360	0	96	0	30	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
OC	507	81	156	3	69	2	40	0	28	0	17	0	8	0	6	0	3	0	1	0	1	0
Aq3 T3	1523	177	210	4	68	0	39	0	22	0	12	0	8	0	1	0	0	0	0	0	0	0

Table S19. Classified reads of the sediment samples, from both the first and second aquarium experiment, using the custom whole genome database. The first row shows the confidence level, which ranged from 0.0 to 1.0. The second row shows which species the reads belong to, CMC; *Eriocheir sinensis*, *P. bo*; *Pandalus borealis*. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

	0		**0.1**		**0.2**		**0.3**		**0.4**		**0.5**		**0.6**		**0.7**		**0.8**		**0.9**		**1**	
	CMC	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***	**CMC**	***P. bo***
Aq1 T1	36	3	9	0	6	0	5	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0
Aq2 T1	5	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aq3 T1	61	4	25	0	12	0	8	0	4	0	2	0	1	0	0	0	0	0	0	0	0	0
Aq1 T2	40	3	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aq2 T2	10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aq3 T2	30	5	9	0	4	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Aq1 T3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aq2 T3	17	4	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aq3 T3	164	32	25	0	7	0	3	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
NAq1 T1	96	1	85	0	61	0	39	0	25	0	12	0	11	0	3	0	1	0	0	0	0	0
NAq2 T1	2527	12	2175	1	1732	1	1268	1	842	0	454	0	204	0	50	0	4	0	0	0	0	0

NAq3 T1	1212	11	992	0	781	0	536	0	341	0	193	0	67	0	20	0	2	0	0	0	0	0
NAq1 T2	233	8	69	0	27	0	9	0	6	0	5	0	3	0	2	0	0	0	0	0	0	0
NAq2 T2	432	13	196	0	114	0	68	0	38	0	24	0	11	0	2	0	0	0	0	0	0	0
NAq3 T2	466	39	234	0	131	0	76	0	52	0	17	0	4	0	1	0	0	0	0	0	0	0
OC	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NC	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NNC	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sa 1	544	149	37	12	18	7	11	5	9	5	7	3	3	2	1	1	0	0	0	0	0	0
Sa 2	584	230	52	13	21	7	15	5	11	3	8	3	5	1	2	1	1	0	0	0	0	0
Sa 3	334	84	22	4	13	3	6	2	5	2	5	2	5	1	3	1	3	0	0	0	0	0
Sa 4	295	68	21	2	12	2	6	1	6	1	5	1	4	1	4	1	3	0	0	0	0	0
SH 1	55	19	8	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SH 2	24	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SH 3	88	14	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SH 4	253	59	20	2	5	2	2	2	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Table 20. Percentage of classified reads of all samples for the databases built using only COI sequences as the reference library, with different k-mer and minimizer lengths and number of minimizer spaces, at confidence level 0.0. The name of the databases shows the length of the k-mer, k, and minimizer, m, and the number of minimizer spaces, ms.

Species	k35_m31 _ms7	k31_m25 _ms6	k27_m21 _ms5	k25_m19 _ms4	k21_m17 _ms4	k19_m15 _ms2	k17_m13 _ms1
E. sinensis	0	0	0	0.001	0.044	0.029	0.091
E. japonica	0	0	0.001	0.009	0.456	0.303	1.153
C. pagurus	0	0.001	0.005	0.031	1.461	1.049	4.245
C. maenas	0	0	0.004	0.028	1.227	1.073	4.432
P. bernhardus	0	0	0.001	0.004	0.073	0.07	0.275
P. borealis	0	0	0.005	0.021	1.267	0.971	4.01
P. depurator	0	0	0.001	0.005	0.141	0.117	0.411
Total classified	0	0.002	0.015	0.099	4.669	3.612	14.616

Table 21. Percentage of classified reads of all samples for the databases built using only COI sequences as the reference library, with different k-mer and minimizer lengths and number of minimizer spaces, at confidence level 0.1. The name of the databases shows the length of the k-mer, k, and minimizer, m, and the number of minimizer spaces, ms.

Species	k35_m31 _ms7	k31_m25 _ms6	k27_m21 _ms5	k25_m19 _ms4	k21_m17 _ms4	k19_m15 _ms2	k17_m13 _ms1
E. sinensis	0	0	0	0	0	0	0
E. japonica	0	0	0	0	0	0	0
C. pagurus	0	0	0	0	0.001	0	0.001
C. maenas	0	0	0	0	0	0.004	0.001
P. bernhardus	0	0	0	0	0	0	0
P. borealis	0	0	0	0	0	0	0.001
P. depurator	0	0	0	0	0	0	0
Total classified	0	0	0	0	0.002	0.005	0.004

Table S22. Classified reads of the water samples using the custom k21_m17_ms4 COI database at confidence level 0.1. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Aq1 – Aq3 denotes sequential increase of number of crab legs (Aq1 = 1 CMC leg, Aq2 = 3 CMC legs, Aq3 = 5 CMC legs); OC = other-crustaceans control; NC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); UW = sample extracted without additional wash step.

Species	***E. sinensis***	***E. japonica***	***C. maenas***	***P. depurator***	***C. pagurus***	***P. borealis***	***P. bernhardus***
Aq1 T1	0	0	0	0	0	1	0
Aq2 T1	0	0	0	0	1	0	0
Aq3 T1	0	0	0	0	0	0	0

Aq1 T2	0	2	0	0	0	0	0
Aq2 T2	0	0	0	0	0	1	0
Aq3 T2	0	0	1	1	1	0	0
Aq1 T3	1	0	0	0	0	0	0
Aq2 T3	0	0	2	0	1	1	0
Aq3 T3	0	1	2	1	0	2	0
NC	0	0	0	0	0	0	0
OC	0	0	3	0	1	1	2
Aq3 T3 UW	0	0	0	0	0	0	0

Table S23. Classified reads of the sediment samples using the custom k21_m17_ms4 COI database at confidence level 0.1. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

Species	***E. sinensis***	***E. japonica***	***C. maenas***	***P. depurator***	***C. pagurus***	***P. borealis***	***P. bernhardus***
Aq1 T1 S	0	0	0	0	0	0	0
Aq2 T1 S	0	0	0	0	0	0	0
Aq3 T1 S	0	0	0	0	0	0	0
Aq1 T2 S	0	0	0	0	0	0	0
Aq2 T2 S	0	0	0	0	0	0	0
Aq3 T2 S	0	0	0	0	0	0	0
Aq1 T3 S	0	0	0	0	0	0	0
Aq2 T3 S	0	0	0	0	0	0	0
Aq3 T3 S	0	0	0	0	0	0	0
OC S	0	0	0	0	0	0	0
NC S	0	0	0	0	0	0	0
NAq1 T1	0	0	0	0	0	0	0
NAq2 T1	0	0	0	0	0	0	0
NAq3 T1	0	0	0	0	0	0	0
NAq1 T2	0	0	0	0	0	0	0
NAq2 T2	0	0	0	0	0	0	0
NAq3 T2	0	0	0	0	0	0	0
NNC	0	0	0	0	0	0	0
Sannorna 1	0	0	0	0	4	4	1
Sannorna 2	0	0	1	0	12	0	2
Sannorna 3	0	0	0	0	2	1	0
Sannorna 4	0	0	0	0	1	1	0
Stenhammar 1	0	0	0	0	0	0	0
Stenhammar 2	0	0	0	0	0	0	0

Stenhammar 3	0	0	0	0	0	0	0
Stenhammar 4	0	0	0	0	0	0	0

Table S24. Classified reads, of the water samples, with the default parameters MitoGun database at confidence level 0.1. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); UW = sample extracted without additional wash step.

Species	***E. sinensis***	***E. japonica***	***C. maenas***	***P. depurator***	***C. pagurus***	***P. borealis***	***P. bernhardus***
Aq1 T1	122	0	0	0	0	0	0
Aq2 T1	40	0	0	0	0	0	0
Aq3 T1	131	0	0	0	0	0	0
Aq1 T2	71	0	0	0	0	0	0
Aq2 T2	87	0	0	0	0	0	1
Aq3 T2	223	0	0	0	0	0	2
Aq1 T3	13	0	0	0	0	0	0
Aq2 T3	31	2	0	0	0	0	0
Aq3 T3	68	0	0	0	0	0	0
NC	4	0	0	0	0	0	0
OC	29	0	0	0	0	0	0
Aq3 T3 UW	21	0	0	0	0	0	0

Table S25. Classified reads, of the sediment samples, with the default parameters MitoGun database at confidence level 0.1. Cells marked with green shows reads, or lack of reads, that are expected in that sample. Cells marked with red show reads, or lack of reads, that are not expected in that sample. Cells marked with yellow show reads of special interest. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

Species	***E. sinensis***	***E. japonica***	***C. maenas***	***P. depurator***	***C. pagurus***	***P. borealis***	***P. bernhardus***
Aq1 T1 S	1	0	0	0	0	0	0
Aq2 T1 S	1	0	0	0	0	0	0
Aq3 T1 S	4	0	0	0	0	0	0
Aq1 T2 S	0	0	0	0	0	0	0
Aq2 T2 S	0	0	0	0	0	0	0
Aq3 T2 S	0	0	0	0	0	0	0

Aq1 T3 S	0	0	0	0	0	0	0
Aq2 T3 S	0	0	0	0	0	0	0
Aq3 T3 S	0	0	0	0	0	0	0
OC S	0	0	0	0	0	0	0
NC S	0	0	0	0	0	0	0
NAq1 T1	5	0	0	0	0	0	0
NAq2 T1	418	0	0	0	0	0	0
NAq3 T1	202	0	0	0	0	0	0
NAq1 T2	4	0	0	0	0	0	0
NAq2 T2	31	0	0	0	0	0	0
NAq3 T2	11	0	0	0	0	0	0
NNC	0	0	0	0	0	0	0
Sannorna 1	0	0	0	0	0	0	0
Sannorna 2	2	0	0	0	0	0	0
Sannorna 3	1	0	0	0	0	0	0
Sannorna 4	0	0	0	0	0	0	0
Stenhammar 1	0	0	0	0	0	0	0
Stenhammar 2	0	0	0	0	0	0	0
Stenhammar 3	0	0	0	0	0	0	0
Stenhammar 4	0	0	0	0	0	0	0

Table S26. Classified reads, of the water samples, with the default parameters MitoGun database at confidence level 0.2. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); UW = sample extracted without additional wash step.

Species	***E. sinensis***	***E. japonica***	***C. maenas***	***P. depurator***	***C. pagurus***	***P. borealis***	***P. bernhardus***
Aq1 T1	56	0	0	0	0	0	0
Aq2 T1	15	0	0	0	0	0	0
Aq3 T1	82	0	0	0	0	0	0
Aq1 T2	17	0	0	0	0	0	0
Aq2 T2	23	0	0	0	0	0	1
Aq3 T2	64	0	0	0	0	0	1
Aq1 T3	1	0	0	0	0	0	0
Aq2 T3	7	0	0	0	0	0	0
Aq3 T3	12	0	0	0	0	0	0
NC	0	0	0	0	0	0	0
OC	10	0	0	0	0	0	0
Aq3 T3 UW	8	0	0	0	0	0	0

Table S27. Classified reads, of the water samples, with the default parameters MitoGun database at confidence level 0.6. The cells filled with green indicate reads expected to be present/absent, the cells filled with red indicate reads not expected to be present/absent, and the cells filled with yellow indicate reads that are uncertain if they were present in the sample or not. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); UW = sample extracted without additional wash step.

Species	*E. sinensis*	*E. japonica*	*C. maenas*	*P. depurator*	*C. pagurus*	*P. borealis*	*P. bernhardus*
Aq1 T1	7	0	0	0	0	0	0
Aq2 T1	0	0	0	0	0	0	0
Aq3 T1	8	0	0	0	0	0	0
Aq1 T2	0	0	0	0	0	0	0
Aq2 T2	0	0	0	0	0	0	0
Aq3 T2	1	0	0	0	0	0	0
Aq1 T3	0	0	0	0	0	0	0
Aq2 T3	1	0	0	0	0	0	0
Aq3 T3	2	0	0	0	0	0	0
NC	0	0	0	0	0	0	0
OC	0	0	0	0	0	0	0
Aq3 T3 UW	0	0	0	0	0	0	0

Table S28. Classified reads, of the sediment samples, with the default parameters MitoGun database at confidence level 0.6. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third).

Species	*E. sinensis*	*E. japonica*	*C. maenas*	*P. depurator*	*C. pagurus*	*P. borealis*	*P. bernhardus*
Aq1 T1 S	0	0	0	0	0	0	0
Aq2 T1 S	0	0	0	0	0	0	0
Aq3 T1 S	0	0	0	0	0	0	0
Aq1 T2 S	0	0	0	0	0	0	0
Aq2 T2 S	0	0	0	0	0	0	0
Aq3 T2 S	0	0	0	0	0	0	0
Aq1 T3 S	0	0	0	0	0	0	0
Aq2 T3 S	0	0	0	0	0	0	0
Aq3 T3 S	0	0	0	0	0	0	0
OC S	0	0	0	0	0	0	0
NC S	0	0	0	0	0	0	0
NAq1 T1	0	0	0	0	0	0	0
NAq2 T1	8	0	0	0	0	0	0
NAq3 T1	2	0	0	0	0	0	0
NAq1 T2	0	0	0	0	0	0	0

NAq2 T2	1	0	0	0	0	0	0
NAq3 T2	1	0	0	0	0	0	0
NNC	0	0	0	0	0	0	0
Sannorna 1	0	0	0	0	0	0	0
Sannorna 2	0	0	0	0	0	0	0
Sannorna 3	0	0	0	0	0	0	0
Sannorna 4	0	0	0	0	0	0	0
Stenhammar 1	0	0	0	0	0	0	0
Stenhammar 2	0	0	0	0	0	0	0
Stenhammar 3	0	0	0	0	0	0	0
Stenhammar 4	0	0	0	0	0	0	0

Table S29. Binary detection matrix of the qPCR and ONT results. Codes: Aq = Aquarium (Experiment 1); Naq = Aquarium (Experiment 2); Aq/Naq1 – Aq/Naq3 denotes sequential increase of number of crab legs (Aq/Naq1 = 1 CMC leg, Aq/Naq2 = 3 CMC legs, Aq/Naq3 = 5 CMC legs); OC = other-crustaceans control; NC/NNC = negative controls; Sa 1-4 = field samples from Sannorna; SH 1-4 = field samples from Stenhammar; T1-T3 denotes sequential sampling time points within an experiment (T1 = first, T2 = second, T3 = third); UW = sample extracted without additional wash step.

Sample	qPCR	ONT
Aq1 T1 W	1	1
Aq2 T1 W	1	0
Aq3 T1 W	1	1
Aq1 T2 W	1	0
Aq2 T2 W	1	0
Aq3 T2 W	1	1
Aq1 T3 W	1	0
Aq2 T3 W	1	1
Aq3 T3 W	0	1
Aq1 T1 S	1	0
Aq2 T1 S	0	0
Aq3 T1 S	1	0
Aq1 T2 S	0	0
Aq2 T2 S	0	0
Aq3 T2 S	1	0
Aq1 T3 S	0	0
Aq2 T3 S	0	0
Aq3 T3 S	0	0
NAq1 T1	1	0
NAq2 T1	1	1
NAq3 T1	1	1
NAq1 T2	0	0
NAq2 T2	1	1
NAq3 T2	1	1

Scripts

1. Entrez Direct

Entrez Direct was used to gather the multiple mitochondrial barcodes used for the MitoGun reference library. The program was run in a bash script through the command line using the following nano script:

```
OUTDIR=~/crab_db/sequences

mkdir -p "$OUTDIR"


declare -A species_taxids=(
 ["Eriocheir_sinensis"]="95602"
 ["Eriocheir_japonica"]="95603"
 ["Cancer_pagurus"]="6755"
 ["Carcinus_maenas"]="6759"
 ["Pagurus_bernhardus"]="174397"
 ["Pandalus_borealis"]="6703"
 ["Liocarcinus_depurator"]="313354"
)


declare -A marker_synonyms=(
 ["COI"]='COI OR CO1 OR "cytochrome oxidase subunit I" OR "cox1"'
 ["cytb"]='cytb OR cob OR "cytochrome b" OR "cytochrome oxidase b"'
 ["12S"]='12S OR "12S ribosomal RNA" OR "small subunit ribosomal RNA"'
 ["16S"]='16S OR "16S ribosomal RNA" OR "large subunit ribosomal RNA"'
 ["ND4"]='ND4 OR "NADH dehydrogenase subunit 4" OR "nad4"'
 ["ND5"]='ND5 OR "NADH dehydrogenase subunit 5" OR "nad5"'
)


for species in "${!species_taxids[@]}"; do
 taxid="${species_taxids[$species]}"
 for marker in "${!marker_synonyms[@]}"; do
  synonyms=${marker_synonyms[$marker]}
  echo "Downloading $marker (synonyms: $synonyms) for $species (taxid $taxid)..."
```

```
  esearch -db nucleotide -query "txid${taxid}[Organism:exp] AND (${synonyms}) NOT predicted[Title]" \
    | efetch -format fasta \
    | awk -v marker="$marker" -v species="$species" '
      BEGIN { RS = ">" ; ORS = "" }
      NR > 1 {
        header = $0
        gsub(/\n/, "", header)
        seq = substr(header, index(header, "\n")+1)
        header = substr(header, 1, index(header, "\n")-1)
        n_count = gsub(/N/, "", seq)
        if (length(seq) >= 200 && n_count <= 5) {
          print ">" marker "_" species "_" NR "\n"
          for (i = 1; i <= length(seq); i += 60)
            print substr(seq, i, 60) "\n"
        }
      }' >> "$OUTDIR/${species}_${marker}.fasta"

  echo "Saved: $OUTDIR/${species}_${marker}.fasta"
 done
done

echo "✅ All downloads complete (with synonyms)."
```

1. Combine sequence files and add Kraken:taxid.

The sequence files were compiled into one file and the Kraken:taxid was added to the header of each sequence using Python (v.3.12.3) in the command line with the following nano script:

```
import os
from Bio import SeqIO

input_dir = os.path.expanduser("~/crab_db/sequences")
output_fasta = os.path.expanduser("~/crab_db/output/combined_kraken_db.fasta")

# Taxonomy ID map
```

```
taxid_map = {
    "Eriocheir_sinensis": "95602",
    "Eriocheir_japonica": "95603",
    "Cancer_pagurus": "6755",
    "Carcinus_maenas": "6759",
    "Pagurus_bernhardus": "174397",
    "Pandalus_borealis": "6703",
    "Liocarcinus_depurator": "313354"
}

# Delete existing output file if present
if os.path.exists(output_fasta):
    os.remove(output_fasta)

with open(output_fasta, "w") as out_f:
    for filename in os.listdir(input_dir):
        if filename.endswith(".fasta"):
            parts = filename.replace(".fasta", "").split("_")
            if len(parts) >= 2:
                species = "_".join(parts[:-1])
                marker = parts[-1]
            else:
                continue

            taxid = taxid_map.get(species)
            if not taxid:
                print(f"Warning: No taxid for {species}, skipping...")
                continue

            fasta_path = os.path.join(input_dir, filename)
            for record in SeqIO.parse(fasta_path, "fasta"):
                accession = record.id.split()[0]
```

```
        new_id = f"kraken:taxid|{taxid}|{species}|{marker}|{accession}"
        record.id = new_id
        record.description = ""
        SeqIO.write(record, out_f, "fasta")

print(f"☑ New Kraken2-formatted FASTA written to {output_fasta}")
```

3. Kraken2 Header Check

The headers were checked for correct formatting, correct taxonomic ID and duplicates using Python (v.3.12.3) in the command line with the following nano script:

```
import os
import pandas as pd
from Bio import SeqIO

# File paths
fasta_path = os.path.expanduser("~/crab_db/output/combined_kraken_db.fasta")
excel_path = "/mnt/d/Crab_Taxid.xlsx"  # updated to your new file
report_path = os.path.expanduser("~/crab_db/output/header_check_report.csv")

# Load and clean taxid map
taxid_df = pd.read_excel(excel_path, dtype=str)
taxid_df.columns = [c.strip() for c in taxid_df.columns]
taxid_map = {
  row["Species"].strip(): row["Taxid"].strip()
  for _, row in taxid_df.iterrows()
  if pd.notnull(row["Species"]) and pd.notnull(row["Taxid"])
}

# Init
malformed_headers = []
taxid_mismatches = []
seen_sequences = set()
duplicates = []
```

```python
# Read FASTA and validate
records = list(SeqIO.parse(fasta_path, "fasta"))
for i, record in enumerate(records, 1):
    header = record.description
    sequence = str(record.seq).upper()

    # Check Kraken2 prefix
    if not header.startswith("kraken:taxid|"):
        malformed_headers.append((i, header))
        continue

    try:
        parts = header.split("|")
        taxid = parts[1].strip()
        species = parts[2].replace("_", " ").strip()
    except IndexError:
        malformed_headers.append((i, header))
        continue

    # Match taxid
    expected_taxid = taxid_map.get(species)
    if not expected_taxid:
        taxid_mismatches.append((i, species, taxid, "Species not in Excel"))
    elif taxid != expected_taxid:
        taxid_mismatches.append((i, species, taxid, f"Expected: {expected_taxid}"))

    # Check for duplicates
    if sequence in seen_sequences:
        duplicates.append((i, header))
    else:
        seen_sequences.add(sequence)
```

```
# Write .csv report
report_rows = []

for i, header in malformed_headers:
    report_rows.append({"Seq#": i, "Issue": "Malformed header", "Details": header})

for i, species, taxid, note in taxid_mismatches:
    report_rows.append({
        "Seq#": i,
        "Issue": "Taxid mismatch",
        "Details": f"{species} → {taxid} ({note})"
    })

for i, header in duplicates:
    report_rows.append({"Seq#": i, "Issue": "Duplicate sequence", "Details": header})

# Save report
os.makedirs(os.path.dirname(report_path), exist_ok=True)
pd.DataFrame(report_rows).to_csv(report_path, index=False)

# Summary output
print("\n📋 Validation Summary:")
print(f"  Total sequences checked: {len(records)}")
print(f"  Malformed headers:      {len(malformed_headers)}")
print(f"  Taxid mismatches:       {len(taxid_mismatches)}")
print(f"  Duplicate sequences:    {len(duplicates)}")
print(f"\n☑ Full report saved to: {report_path}")
```

4. Deduplication

Exact duplicate sequences were removed from the sequence file using Python (3.12.3) in the command line using the following nano script:

```
import os
```

```python
from Bio import SeqIO

# Input and output paths
input_fasta = os.path.expanduser("~/crab_db/output/combined_kraken_db.fasta")
output_fasta =
os.path.expanduser("~/crab_db/output/combined_kraken_db_deduplicated.fasta")

# Track unique sequences
seen_seqs = set()
unique_records = []

for record in SeqIO.parse(input_fasta, "fasta"):
    sequence = str(record.seq).upper()
    if sequence not in seen_seqs:
        seen_seqs.add(sequence)
        unique_records.append(record)

# Write deduplicated FASTA
SeqIO.write(unique_records, output_fasta, "fasta")

print(f"✅ Deduplicated FASTA written: {output_fasta}")
print(f"📉 Removed {len(seen_seqs) - len(unique_records)} duplicates")
print(f"📈 Final count: {len(unique_records)} unique sequences")
```