Genomic DNA extraction optimization and validation for genome sequencing using the marine gastropod Kellet's whelk

Benjamin N Daniels[1,2], Jenna Nurge[2], Olivia Sleeper[2], Andy Lee[3], Cataixa López[5], Mark Christie[3,4], Robert J Toonen[5], Jean Davidson[2], and Crow White[1]

[1]Center for Coastal Marine Sciences, California Polytechnic State University, San Luis Obispo, CA 93407, USA
[2]Department of Biological Sciences, California Polytechnic State University, San Luis Obispo, CA 93407, USA
[3]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-2054, USA
[4]Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-2054, USA
[5]Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kāne'ohe, HI 96744, USA

Corresponding author:

Benjamin N. Daniels[1]

Email address: ben.daniels255@gmail.com

ABSTRACT

Next-generation sequencing technologies, such as Nanopore MinION, Illumina Hiseq and Novaseq, and PacBio Sequel II, hold immense potential for advancing genomic research on non-model organisms, including the vast majority of marine species. However, application of these technologies to marine invertebrate species is often impeded by challenges in extracting and purifying their genomic DNA due to high polysaccharide content and other secondary metabolites. In this study, we help resolve this issue by developing and testing DNA extraction protocols for Kellet's whelk (*Kelletia kelletii*), a subtidal gastropod with ecological and commercial importance, by comparing four DNA extraction methods commonly used in marine invertebrate studies. We identified an optimized protocol that balances efficiency, cost, and DNA extraction quality; the Salting out protocol. We validated the protocol using an independent set of tissue samples, then applied it to extract high-molecular-weight (HMW) DNA from over three thousand Kellet's whelk tissue samples. The protocol demonstrated scalability and, with added clean-up, suitability for RAD-seq, GT-seq, as well as whole genome sequencing using both long read (ONT MinION) and short read (Illumina Novaseq) sequencing platforms. Our findings offer a robust and versatile DNA extraction and clean-up protocol for supporting genomic research efforts on non-model marine organisms, to help mediate the under-representation of invertebrates in genomic studies.

INTRODUCTION

Rapidly advancing next generation sequencing technologies methods, such as whole genome sequencing, genotyping-in-thousands by sequencing (GT-seq), and restriction-site-associated DNA sequencing

(RAD-seq), are becoming more available and affordable for non-model organisms (Park and Kim (2016); Ellegren (2014); Van Wyngaarden et al. (2017); Bootsma et al. (2020)). However, applying these technologies to marine invertebrate species, which are nearly all non-model organisms, is hindered by challenges in extracting and purifying their genomic DNA (Boughattas et al. (2021)). DNA extraction of marine invertebrate species can be riddled with contaminants that result in poor sequencing results (Adema (2021); Chakraborty et al. (2020); Panova et al. (2016); Angthong et al. (2020)). The contaminants are commonly attributed to mucopolysaccharides that co-precipitate with genomic DNA, increasing the viscosity of the sample during extractions and preventing the activity of enzymes as well as proper homogenization of mixtures (Boughattas et al. (2021)). As a result, researchers studying marine invertebrates species have had to dedicate extensive time and resources to troubleshooting and optimizing their genomic DNA extraction methods (Ferrara et al. (2006); Panova et al. (2016); Boughattas et al. (2021); Adema (2021)).

This study aims to develop and test DNA extraction protocols and generate an optimal protocol for the marine gastropod, Kellet's whelk (*Kelletia kelletii*), for supporting genomic research on the species and marine invertebrates in general. Kellet's whelk is a subtidal gastropod distributed along the North American west coast (Zacherl et al. (2003)), a significant predator in kelp forest ecosystems (Halpern et al. (2006)), and a species of interest to commercial fisheries (Aseltine-Neilson et al. (2006)). Furthermore, Kellet's whelk recently exhibited a >400 km northward range expansion (Herrlinger (1981)), potentially driven by intensifying El Niño Southern Oscillation (ENSO) oceanographic conditions due to climate change (Zacherl et al. (2003); Harley et al. (2006)). Current research on the species seeks to use genomic-based assignment testing (*sensu* Christie et al. (2017)) to investigate this hypothesis and, more broadly, elucidate spatio-temporal patterns of gene flow and population dynamics of Kellet's whelk (López et al. (2023)). Such information holds significant promise in addressing pivotal marine ecological and evolutionary inquiries and providing valuable support for the sustainable management of the Kellet's whelk fishery, as well as that of other ecologically and economically important coastal species with analogous life history traits (e.g., large populations, slow growth, sedentary demersal adults, and pelagic larvae) such as spiny spiny lobster (*Panulirus interruptus*), kelp bass (*Paralabrax clathratus*), sheephead (*Semicossyphus pulcher*), and some rockfish (*Sebastes atrovirens*; Allen et al. (2006); Froese and Pauly (2011)).

An experimental design for revealing genomic features of Kellet's whelk could be supported by applying next generation sequencing technologies to a large sample size of individuals in parent and recruit (1-2 year old Kellet's whelk individuals) populations across the species' biogeographic range and over multiple years (Christie et al. (2017)). For instance, the application of GT-seq and RAD-seq techniques targeting neutral loci, along with GT-seq focusing on geographically-associated loci harboring putatively adaptive alleles, can offer insights into the underlying adaptive genetic mechanisms and historical population dynamics within the species (Reitzel et al. (2013)). Moreover, these approaches can delineate spatially distinct population units characterized by unique genotypic patterns, facilitating the assignment testing of recruits to adult populations, thereby unraveling population connectivity in the species (Malde (2014); Christie et al. (2017)). To support such genomics research, a reliable method for generating high-molecular weight (HMW) genomic DNA is required from each sample. Furthermore, analysis of a large sample size (e.g., 1000s of individuals, necessary for generating a complete population connectivity matrix; Christie et al. (2017)) requires an optimized high-throughput HMW DNA extraction methodology. To help meet this need for Kellet's whelk and potentially other non-model marine invertebrate species, we tested and compared the efficacy of DNA extraction protocols on Kellet's whelk tissue. We evaluated four DNA extraction methods from marine invertebrate studies: 1) Qiagen Blood and Tissue kit (Qiagen, Germany) (White and Toonen (2008); Song et al. (2019)), 2) Zymo Quick-DNA HMW Magbead kit (Zymo, USA) (Akinde et al. (2022)), 3) Phenol Chloroform protocol (PCI) (Thermo Fisher Scientific, USA) (Nagel et al. (2015)), and, 4) an optimized version of a Salting out protocol (Li et al. (2011); Ferrara et al. (2006)). For each, we conducted DNA extraction on 4-5 tissue samples, then compared the quality of DNA extractions produced, as well as the difficulty, toxicity and cost of the extraction method. We then used our test results and lessons learned from other marine genomic research projects to develop a modified protocol intended to optimally balance efficiency, cost, and quality of extraction of Kellet's whelk DNA. We validated our protocol by evaluating DNA yield, DNA quality, and DNA purity generated by DNA extractions on an independent set of 16 tissue samples. We then applied our protocol to a large set of over three thousand tissue samples. We conducted the DNA extractions in batches using a group of trained student

researchers, and assessed the quality of the DNA extractions produced on a random subset of the samples. We also tested a subset of the extractions for both RAD-seq and GT-seq methods. Finally, we tested our protocol by applying DNA extractions generated from it to whole genome sequencing using both long read sequencing (ONT MinION) and short read sequencing (Illumina Novaseq).

METHODS

We tested four DNA extraction methods used in previous marine invertebrate studies. Each method was applied to tissue samples dissected from adult Kellet's whelk collected in the wild and maintained prior to dissection in flow-through filtered seawater aquaria at the California Polytechnic State University research pier, located in Avila Beach, California, USA. The adult whelks were collected from sub-tidal reefs located near Monterey (36.6181670 N, 121.897 W), Naples (34.4219670 N, 119.952283 W), Diablo Canyon (35.2244500 N, 120.877483 W), and Point Loma (32.665333 N, 117.261517 W) California in 2019, transported live to the aquaria, and maintained under ambient conditions with food in the form of frozen seafood provided ad libitum (CDFW Scientific Collection Permit 8018 to C.W.). Tissue samples were dissected in March 2021. Approximately 200 mg tissue was dissected non-invasively from the foot of the individual, immediately frozen in liquid nitrogen, then transported to the Center for Applied Biotechnology at California Polytechnic State University and stored at -80°C. A portion of each tissue sample was used for DNA extractions as specified in each method.

Five Kellet's whelk individuals were dissected for the Qiagen and Zymo kit extractions with each sample having a duplicate (10 total extractions conducted using the Qiagen kit and 10 total extractions conducted using the Zymo kit) in March 2021. Four Kellet's whelk individuals were dissected for the Salting out and PCI protocols with duplicates for the Salting out protocol but not for the PCI protocol (8 total extractions conducted using the Salting out protocol and 4 total extractions conducted using the PCI protocol) in May 2021.

Each method was evaluated based on protocol difficulty, cost per extraction, DNA yield, DNA quality, and toxicity. Difficulty was measured on a three-level qualitative scale. Easy: requires very little wet bench expertise - simple pipetting volumes and reagents as well as basic techniques. Medium: requires some wet bench experience - simple pipetting volumes and reagents with more advanced techniques (e.g., bead binding, DNA washing and drying). Hard: requires more extensive wet bench experience - precise pipetting and handling of toxic reagents (e.g., pipetting layered solutions, working in fume hoods and more extensive use of personal protective equipment (PPE)). Cost per extraction was calculated by the sum of costs per method (e.g., price of Qiagen kit, cost of all reagents for PCI method) divided by the number of possible DNA extractions. DNA yield was measured by a Qubit fluorometer using the DNA Broad Range assay (Invitrogen, USA). DNA quality was evaluated based on gel electrophoresis using $\sim$250 ng DNA per sample (or 10 µL sample if concentration was under 25 µg/µL) on a 1% agarose gel at 80V for 45 minutes with a 1kb DNA ladder. Quality was measured as good: high molecular weight (HMW) band; medium: HMW band with degradation; or, bad: digested. Examples of each quality are shown in Figure 5. Toxicity of each method was measured by the reagent classification in accordance with Occupational Safety and Health Administration (OSHA), International Agency for Research on Cancer (IARC), or National Toxicology Program (NTP). Low toxicity: no ingredients are registered as carcinogenic, aspirationally or reproductively toxic. Medium toxicity: no ingredients are registered as carcinogenic, aspirationally or reproductively toxic, but some ingredients are registered as causing skin irritation, serious irritation, or damage to eyes. High toxicity: ingredients are registered as carcinogenic, aspirationally or reproductively toxic.

Qiagen DNeasy Blood and Tissue kit

DNA extraction procedures followed the manufacturer's protocol. For the Qiagen DNeasy Blood and Tissue kit (Qiagen, Germany), 25 mg tissue per sample was cut into small pieces and placed in a 1.5 mL microcentrifuge tube. 180 µL Buffer ATL and 20 µL Proteinase K were added. The sample was mixed by vortexing, then incubated at 56°C overnight. The sample was vortexed for 15 seconds, then 200 µL Buffer

AL was added. The sample was again vortexed, then 200 µL 100% ethanol was added. The sample was vortexed one last time, then pipetted into a DNeasy Mini spin column placed in a 2 mL collection tube. The sample was centrifuged at 6,000 x g for 1 minute and the flow-through was discarded. The DNeasy Mini spin column was placed in a new 2 mL collection tube, 500 µL Buffer AW1 was added, and then the sample was centrifuged at 6,000 x g for 1 minute. The flow-through was discarded, 500 µL Buffer AW2 was added, and then the sample was centrifuged at 20,000 x g for 3 minutes. The flow-through and collection tube were discarded and the DNeasy Mini spin column was placed in a clean 1.5 mL microcentrifuge tube. 100 µL Buffer AE was added to the DNeasy membrane and incubated at room temperature for 1 minute. The sample was centrifuged at 6,000 x g for 1 minute to elute the DNA. The previous step was repeated with 100 µL Buffer AE to maximize DNA yield. Eluted DNA was stored at -20°C. Protocol can be found in the Supplementary Material section (SI 7).

Zymo Quick-DNA HMW Magbead kit

DNA extraction procedures followed the manufacturer's protocol. For the Zymo Quick-DNA HMW Magbead kit (Zymo, USA), 25 mg tissue per sample was added to a 1.5 mL microcentrifuge tube. 95 µL DNA Elution Buffer, 95 µL Biofluid and Solid Tissue Buffer, and 10 µL Proteinase K were added to the sample. The sample mixture was pipette mixed 5 times and incubated at 55°C for 1-3 hours until tissue became soluble. The sample was centrifuged at 10,000 x g for 1 minute. The supernatant was removed and transferred to a new clean 1.5 mL microcentrifuge tube. 400 µL Quick-DNA MagBinding Buffer was added to the sample and mixed by pipetting. 33 µL MagBinding Beads were added to the sample, pipette mixed 5 times, and placed on a shaker for 10 minutes. The sample was placed on a magnetic stand until the beads separated from the solution. The supernatant was removed and discarded and the sample was taken off of the magnetic stand. 500 µL Quick-DNA MagBinding Buffer was added to the sample, pipette mixed 5 times, and placed on a shaker for 5 minutes. The sample was placed on a magnetic stand until the beads separated from the solution. The supernatant was removed and discarded and the sample was taken off of the magnetic stand. 500 µL DNA Pre-Wash Buffer was added to the sample and pipette mixed 10 times. The sample was placed on a magnetic stand until the beads separated from the solution. The supernatant was removed and discarded and the sample was taken off of the magnetic stand. 900 µL g-DNA Wash Buffer was added to the sample and pipette mixed 10 times. All liquid was transferred into a new microcentrifuge tube and placed on a magnetic stand until the beads separated from the solution.The supernatant was removed and discarded and the sample was taken off of the magnetic stand. The DNA wash step was repeated. The beads were dried for 20 minutes at room temperature and 50 µL DNA Elution Buffer was added. The sample mixture was pipette mixed 20 times and incubated at room temperature for 5 minutes. The sample was placed on a magnetic stand until the beads separated from the solution. The supernatant (eluted DNA) was removed and transferred to a new clean 1.5 mL microcentrifuge tube. Eluted DNA was stored at -20°C. Protocol can be found in the Supplementary Material section (SI 8).

PCI Protocol

The PCI DNA extraction procedure followed an adopted method from Thermo Fisher Scientific, USA. For each sample, 30 mg tissue was homogenized mechanically in 1% SDS cell lysis buffer, 0.5M EDTA, and Proteinase K. The sample was then placed in a 65°C water bath for 1 hour, and vortexed every 20 minutes. 1x volume Phenol Chloroform (PCI) was added to the sample and vortexed. The sample was centrifuged for 5 minutes at 16,000 x g and the upper aqueous phase was transferred to a new tube. 20 µg Glycogen, .5x volume Ammonium Acetate 7.5M, and 2.5x volume 100% ethanol were added to the sample. The sample was pipette mixed and kept at -20°C overnight, then centrifuged at 4°C for 30 minutes at 16,000 x g and the supernatant was removed. 150 µL 70% ethanol was added to the sample, it was centrifuged at 4°C for 2 minutes at 16,000 x g and the supernatant was removed. The previous step was repeated. DNA was dried in a speedvac for 2 minutes and resuspended in 200 µL Elution Buffer. Eluted DNA was stored at -20°C. Protocol can be found in the Supplementary Material section (SI 9).

Salting Out Protocol

The Salting out protocol was adopted from Li et al. (2011). For each sample, 30 mg tissue was homogenized mechanically in 1% SDS cell lysis buffer, 0.5M EDTA, and Proteinase K. The sample was placed in a 65°C water bath for 1 hour, and vortexed for 5 seconds every 20 minutes. RNase A was added, the sample was lightly vortexed, then incubated at 37°C for 10 minutes. 150 µL 7.5M Ammonium Acetate was added to the sample and mixed by vortexing. The sample was incubated at 4°C for 5 minutes and centrifuged at 12,000 x g for 10 minutes at 4°C. The supernatant was transferred to a new tube and 60 µL 7.5M Ammonium Acetate was added. The sample was mixed by vortexing, incubated at 4°C for 5 minutes, and centrifuged at 12,000 x g for 10 minutes at 4°C. The supernatant was transferred to a new tube and 1x volume 100% isopropanol was added. The sample was inverted 50 times and centrifuged at 8,000 x g for 5 minutes at 4°C. The supernatant was discarded and 400 µL 70% ethanol was added. The sample was inverted 3 times and centrifuged at 4°C at 16,000 x g for 5 minutes. The previous two steps were repeated. The sample was dried in a speedvac at 30°C for 4 minutes. DNA was resuspended in 100 µL TE buffer and incubated at 4°C overnight. Eluted DNA was stored at -20°C. Protocol can be found in the Supplementary Material section (SI 10).

Validation

The protocol selected as the most optimal among the four methods (based on the four selection criteria; see above and Results), was then validated on an independent set of 16 tissue samples from 9 adult and 7 recruit Kellet's whelks collected in the wild at 15 sub-tidal reef locations across the species' range in 2015, 2016, and 2017 (CDFW Scientific Collection Permit 8018 to C.W.)(Fig. 1). Tissue samples were collected from the foot of each individual Kellet's whelk using a modified scalpel. Tissue samples from adults were collected non-lethally in the field, while recruits were collected whole in the field (because of their small size) and tissue samples were collected from them in the lab. Adult tissues and whole recruits collected in the field were frozen on dry ice or in liquid nitrogen, then transported to Cal Poly and stored at -80°C until used for this study in 2021 and 2022. DNA extractions were evaluated based on DNA yield, DNA quality, and DNA purity. DNA yield and DNA quality were assessed using the same methods described above for comparing the four DNA extraction methods. DNA purity was measured by absorbance at 260/280 (contamination of proteins absorb at $\sim$280 nm, ideal ratio = $\sim$1.8) and 260/230 (contamination of salts, carbohydrates, and/or phenols absorb at $\sim$230 nm, ideal ratio = $\sim$1.8-2.2) (Scientific (2009)) using a Nanodrop-1000 spectrophotometer (Invitrogen, USA).

Scaling-up

The selected optimal method was applied to 3,325 Kellet's whelk tissue samples and assessed by DNA yield and DNA quality using the same methods above. The tissue samples represent 2026 adults and 1299 recruits from the same set of wild-collected samples used in the validation step described above, but covering more sites (Table S1; Fig. 1). The extractions procedure was conducted by 13 undergraduate students at California Polytechnic State University in the Center for Applied Biotechnology from January 2022 to May 2022. DNA was extracted from samples in groups of 20 each week by each student. Three extractions were randomly selected from each set of 20 extractions bi-weekly for quality check assurance throughout the extraction procedure. Quality checks were conducted based on DNA yield and DNA quality, using the same methods described above for comparing the four DNA extraction methods. Sample quality and DNA concentration correlations were tested in R and visualized using ggplot2.

RAD-seq and GT-seq

A subset of the scaled-up DNA extractions were tested for sequencing quality on both GT-seq and RAD-seq platforms. A range of varying locations, years, and tissue types were used for each subset (Table S3; Table S4; Table S5).

For GT-seq 96 DNA extractions were shipped frozen to Twin Falls, ID, USA and processed at the

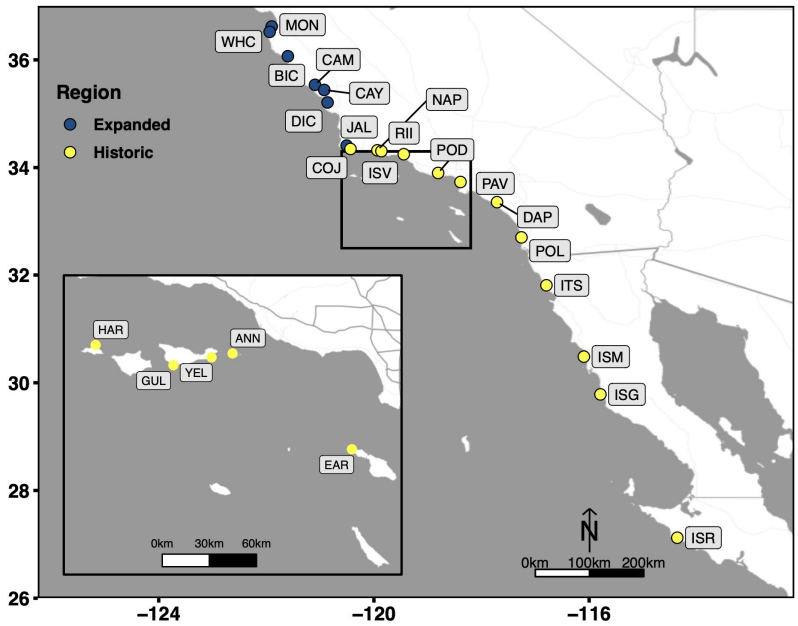

Figure 1. Collection sites which span the entire biogeographic range of Kellet's whelk. Kellet's whelk populations north of Pt. Conception represents a recent range expansion into colder-water habitat (Herrlinger (1981)). See Supplementary Material section for location coordinates (Table S1).

GTseek LLC laboratory. Briefly, exonuclease 1 was added to each gDNA extraction. gDNA was amplified using the set GT-seq primers. PCR products were indexed for Illumina sequencing with i7 and i5 index. Samples were normalized using Nate's Plates Tagging and Normalization kit (SI 18). Size selection was conducted using SPRI beads. In depth methods can be found in the Supplementary Material section. Further optimization was conducted for GT-seq using the Zymo Clean and concentrator kit on aliquots of the same samples initially tested. The same methods were used for sequencing these samples. The genotyping pipeline used can be found at https://github.com/GTseq/GTseek_utils.

For RAD-seq, 172 DNA extractions were shipped frozen to the Hawai'i Institute of Marine Biology, Hawai'i, USA and processed at the ToBo genomics lab (https://tobolab.org/). The gDNA digestion was performed with isoschizomer restriction enzymes (New England Bio-Labs, Ipswich, MA). Libraries were generated using the KAPA Hyper Prep DNA kit (Roche Sequencing and Life Science) and then shipped frozen to the University of California, Davis, USA (UCD) and sequenced at UCD Genome Center through HiSeq2500. RAD-seq reads were filtered by their quality scores and the presence of the RADcutsite using process_radtags in Stacks (Rivera-Colón and Catchen (2022)). Protocols can be found in the Supplementary Material section (SI 11-12).

Library Preparation

To optimize and validate the selected DNA extraction method for next generation sequencing, a library from one of the DNA extractions generated using the Salting out protocol was prepared for Nanopore MinION sequencing using the rapid sequencing kit (SQK-RAD004, ONT, Oxford, UK) without any further cleaning or repair. Three DNA libraries from the same DNA extraction were prepared with further purification and concentration using the Genomic DNA Clean and Concentrator kit (Zymo Research, Irvine, CA, USA). The DNA was then selected for high molecular weight DNA using the PacBio Short Read Eliminator kit (PacBio, Menlo Park, CA, USA). Concentration was assessed using the dsDNA BR assay on a Qubit fluorometer (Thermo Fisher Scientific, Singapore). Two libraries were prepared for Nanopore MinION sequencing using the Ligation Sequencing kit (SQK-LSK108, ONT, Oxford, UK) and NEBnext DNA Repair kit reagents (NEB, Ipswich, MA, USA) according to the manufacturer's instructions. The other library was shipped frozen to the University of Oregon (Eugene, OR) and prepared

for Illumina Novaseq sequencing using the NEB Ultra II kit (NEB, Ipswich, MA, USA) according to the manufacturer's instructions with a few modifications by the Genomics and Cell Characterization Facility (GC3F) at the University of Oregon. DNA was mixed with the fragmentation reagents as described in the kit instructions and fragmented at 37°C for 8 minutes, and end repaired as described in the manual. The end repaired sample was mixed with the ligation buffer and enhancer from the NEB kit, and 2.5 µL of 15 µm pre-annealed Tru-Seq style Y-adapter and ligated for 15 minutes at 20°C. The sample was cleaned with 2, 0.75x bead cleans to remove adapter dimer. Samples were quantified by qPCR and loaded on the NovaSeq. Protocol methods can be found in the Supplementary Material section (SI 13-17).

Whole Genome Sequencing

For Nanopore MinION sequencing, each library was loaded onto an R9 flow cell (FLO-MIN106, ONT). Priming and loading of the SpotON Flow Cells were performed using the standard protocol (sequencing gDNA (SQK-RAD004 or SQK-LSK109) Protocol) (Lu et al. (2016)). MinION sequencing was operated with MinKNOW v5.2.13 without basecalling (SI 17). Each flow cell was sequenced until <2 pores were sequencing. Basecalling was conducted on the raw Nanopore MinION reads using Guppy v6.2.1 and high accuracy mode (<5% error rate). Illumina Novaseq sequencing was conducted by GC3F at the University of Oregon (Eugene, OR) using the NEBNext Ultra 2 DNA Library Prep kit for Illumina with slight modifications as stated above. All methods and protocols described above can be found in the Supplementary Material section.

RESULTS

Extraction Performance

Each extraction method was evaluated based on its difficulty, cost, yield, quality, and toxicity. Regarding DNA extraction difficulty (Table 1), the Qiagen kit was considered "easy" due to its lack of complex pipetting steps, while the Zymo kit and Salting out protocol were considered "medium" due to steps such as DNA washing and drying that required proficient lab skills. The PCI protocol was considered "hard" because it required steps with precise pipetting skills such as layered toxic solutions and under a fume hood.

Comparing the cost per extraction of each method revealed the PCI and Salting out protocols, which required separate purchasing and assembly of reagents, to be approximately half as expensive as the Qiagen and Zymo kit methods, which contain pre-assembled reagents with the kit purchase (Table 1).

DNA yield was broadly low in all samples prepared by the Qiagen Blood and Tissue kit (Table 1). DNA extractions produced by the Zymo kit produced optimal DNA yields (>1 µg of DNA) (Table 1). The Salting out protocol produced high and more variable DNA yields with a standard deviation of 9.45 µg (Table 1). The PCI method produced DNA yields lower than either the Zymo kit or salting out protocol, but higher than that by the Qiagen Blood and Tissue kit (Table 1).

The Qiagen and Zymo kits produced poor quality extractions. Samples prepared by the Qiagen Blood and Tissue kit had DNA degradation across most samples as observed by "digested" smears and low molecular weight bands on the agarose gel electrophoresis (Fig. 2, see Fig. 5 for example of different levels of quality). DNA extractions produced by the Zymo kit resulted in inconsistent HMW bands with DNA degradation (Fig. 2). The PCI method produced HMW bands with little DNA degradation (Fig. 2). The Salting out protocol produced consistent HMW bands without DNA degradation across all 2).

All methods had relatively low toxicity, except the PCI method which required the use of phenol and chloroform, which are dangerous chemicals that can be absorbed through the skin and are suspected mutagens.

The Salting out protocol was selected as optimal among the four methods, due to its lack of difficulty, lowest cost per extraction,low toxicity, high DNA yield, and best quality HMW DNA.

Protocol	Difficulty (easy, medium, hard)	Cost per extraction	DNA Yield (µg)	DNA Quality (Good=HMW Band, Medium=HMW band with smear, Bad=fragmented smear)	Toxicity (high, medium, low)
Qiagen Blood and Tissue Kit	Easy	$3.49	0.14 ± 0.084	Bad	low
Zymo Magbead HMW Kit	Medium	$3.15	13.38 ± 2.06	Medium - inconsistent	low
PCI Protocol	Hard	$1.83	2.84 ± 1.27	Good	high
Salting Out Protocol	Medium	$1.50	27.78 ± 9.45	Good	low

Table 1. DNA extraction protocol evaluation on commonly used DNA extraction methods in other marine invertebrate studies. Evaluation is based on difficulty, cost, yield, quality, and toxicity.

Validation

To test the reliability of the Salting out protocol, it was validated on 16 Kellet's whelk adult and recruit foot tissues from frozen samples collected across the species' entire biogeographic range over multiple years (Table 2; Fig. 1). DNA yield of these extractions was broadly high, with an average of 35.7±27.48 µg (Table 2). HMW bands were evident in all extractions, but signs of digestion/degradation were also evident (Fig. 3). The purity of these DNA extractions were evaluated using the samples absorbance ratios at A260/280 and A260/230. The average A260/280 ratio of these extractions was 1.85±0.046 and A260/230 was 1.81±0.27 indicating pure DNA with, potentially, some ∼230 nm absorbing contaminants.

Scaling

When the salting out protocol was tested on 3,325 samples, the randomly sampled DNA extractions (252 DNA extractions from all different locations, tissue type, and years) resulted in a range of quality, with most falling into high enough quality for downstream GT-seq and RAD-seq applications. The overall quality of each extraction fell into three groups; Digested (8.7%), HMW + Degradation (29.8%), and HMW (61.5%) (Fig. 4; Fig. 5). The majority (81.3%) fell into the groups optimal for RAD-seq and GT-seq; high quality HMW group or HMW + Degradation. The remaining Digested samples (8.7%), would be the least likely to produce sufficient RAD-seq and GT-seq libraries. DNA concentration was found to be correlated with DNA quality (Fig. 4). This result is most likely due to smaller sample sizes that may have undergone DNA degradation more rapidly during the storing and thawing phase of tissue.

RAD-seq and GT-seq

A subset of 172 of the scaled-up DNA extractions were analyzed via RAD-seq, producing a total of 793,871,366 reads that passed quality assessment. Among the sequenced samples, 9 (5.2%) exhibited a total read count lower than 410, with the remainder displaying a notably higher count of over 250,000 reads. All reads were further filtered by the presence of the anticipated RAD-seq cut site, resulting in an average of 78.9% of reads retained after all filtering from the raw reads (Fig. 6).

A subset of 96 of the scaled-up DNA extractions were also analyzed with GT-seq. Twenty five (26%) generated >90% GT, 15 (15.6%) generated 30-90% GT, and 56 (58.3%) generated <30% GT (Fig. 6). Approximately half of the samples (49%) had raw read counts below 100,000 reads indicating unsuccessful sequencing. In general, the number of raw reads per sample decreased with lower % GT (Fig. 6). No correlation was found between successful genotyping and sample location or DNA yield (Table S6).

After the same samples were cleaned using the Zymo Clean and Concentrator there was a 49% increase in successful genotype calling (>90%). Of the samples, 72 (75%) were above 90% GT, 13 (13.5%) were between 30 and 90% GT, and only 11 (11.5%) under 30% GT (Fig. 6). The number of raw reads across each sample was far more consistent with only one sample having less than 100,000 reads, a 50% increase

Sample number	Tissue Type	Site	Year	DNA Yield (µg)	DNA Quality (Good, Medium, Bad)	DNA Purity	
						260/280	260/230
1	Adult	POD	2015	27.71	Good	1.89	2.13
2	Adult	CAK	2015	72.87	Medium	1.74	1.54
3	Adult	PUB	2015	26.33	Good	1.86	1.89
4	Adult	ISM	2015	28.7	Good	1.89	2.24
5	Adult	PAV	2015	107.71	Medium	1.91	2.23
6	Adult	JAL	2016	13.43	Bad	1.84	1.81
7	Adult	MON	2016	13.67	Medium	1.85	1.81
8	Adult	RHR	2016	2.97	Good	1.82	1.83
9	Adult	COJ	2016	67.96	Bad	1.9	2.15
10	Recruit	POD	2015	31.32	Medium	1.86	1.79
11	Recruit	ISG	2015	20.16	Medium	1.84	1.53
12	Recruit	GUI	2016	30.2	Good	1.85	1.6
13	Recruit	BIC	2016	7.4	Medium	1.83	1.33
14	Recruit	MON	2016	40.35	Medium	1.87	1.87
15	Recruit	RII	2016	27.64	Bad	1.76	1.53
16	Recruit	WHC	2017	52.73	Bad	1.87	1.71

Table 2. Salting out protocol validation with 16 samples from different tissue types, location (see Fig. 1), and year of collection. Resulting DNA yield, DNA quality, and DNA purity based on A260/280 and A260/230 ratios are reported. Average DNA yield = 35.7±27.48µg, A260/280 =1.85±0.046 , A260/230 = 1.81±0.27. See DNA quality descriptions in table 1.

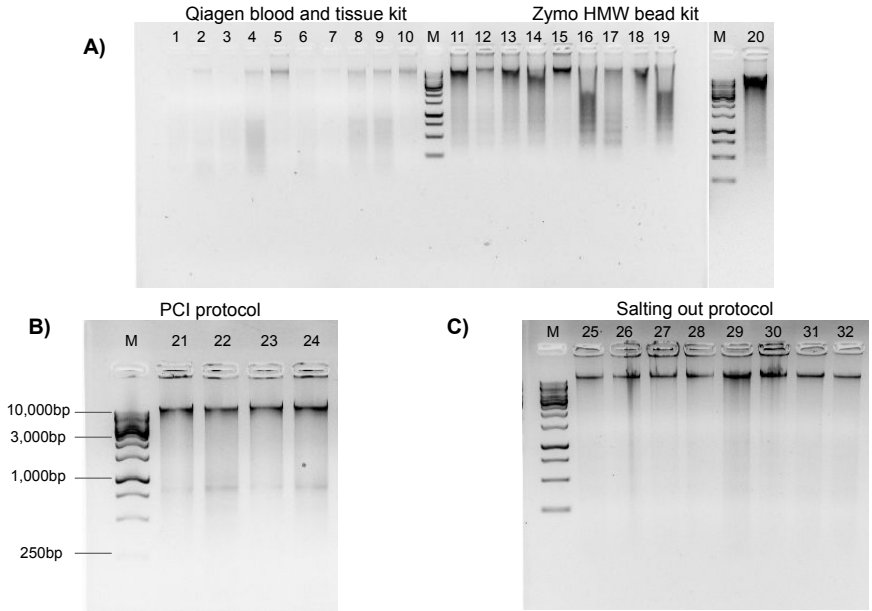

Figure 2. Genomic DNA extractions conducted on flash frozen Kellet's whelk tissue on 1% gel with ∼250 ng gDNA (M = 1kb DNA ladder). **A)** Wells 1-10 display Qiagen Blood and Tissue kit, wells 11-20 display Zymo Quick-DNA HMW Magbead kit. **B)** Wells 21-24 display PCI protocol (∼400bp bands are most likely RNA contaminants due to lack of RNase use in protocol) with 1kb DNA markers **C)** Wells 25-32 display modified Salting out protocol.

in samples with optimal raw read counts after cleaning (Fig. 6).

Genome Sequencing

The DNA from the Salting out protocol was sequenced on the Nanopore MinION and the Illumina Novaseq. The first library prepared using the rapid sequencing kit resulted in a failed sequencing run (Table 3). The sequencing run produced an estimated 14.01k reads and 8.94Mb bases, yet when finalized using the MinKONOW software, no data was produced (Table 3). Both Nanopore MinION libraries prepared using the Genomic DNA Clean Concentrator kit, the PacBio Short Read Eliminator kit, and the ligation sequencing protocol produced successful whole genome sequencing (Table 3). Both sequencing runs produced more than 80 times the estimated number of bases from the rapid sequencing kit run (Table 3). The KK_Ligationseq library produced a total of 2.6 Gb from 446.26 k reads and the KK_Ligationseq2 library produced a total of 1.4 Gb from 98.57 k reads of data that passed high quality base calling using Guppy. Each run had an estimated retention rate of 80.02% and 87.53% for reads that passed high quality filtering, respectively (Table 3).

The DNA from the salting out protocol extraction was also subjected to sequencing on the Illumina Novaseq. The same DNA cleaning and selection method used to produce successful libraries for nanopore sequencing was used. The Illumina Novaseq run produced 715.94 million reads and 113.83 Gb of bases. After filtering, there were 682.18 million reads and 104.85 Gb of bases. The filtered data consisted of 88.78% Q30 bases and 95.27% Q20 bases.

DISCUSSION

Investigating the spatial and temporal scales of genetic variation among populations of non-model organisms, such as marine invertebrates, can reveal emergent, novel patterns, and drivers of population connectivity in a marine system, with significant implications for science and management (Cowen

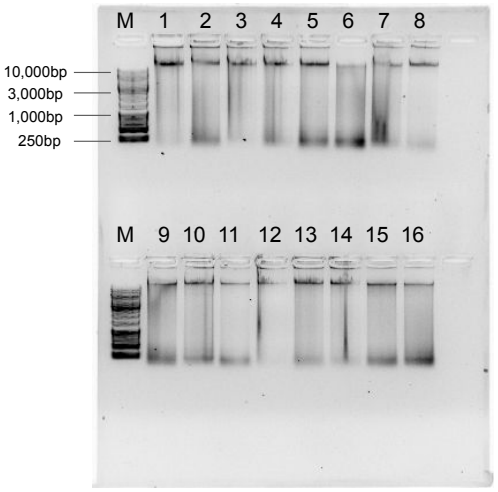

Figure 3. Genomic DNA extractions using the Salting out protocol on 16 adult foot tissue or recruit samples from frozen samples collected in different years (2015, 2016, and 2017) and locations (POD, CAK, PUB, ISM, PAV, JAL, MON, RHR, COJ, POD, ISG, GUI, BIC, RII, WHC; See sample locations and year in Table 2 and location of site code indicated in Fig. 2) on a 1% gel with ∼250 ng gDNA (M = 1kb DNA ladder).

et al. (2007); Fogarty and Botsford (2007)). Recent advances in genome sequencing technologies hold tremendous potential for genomic research across the tree of life, particularly in under-represented organisms. While these new technologies hold the capacity for sequencing numerous non-model species, However, it is essential to recognize that sequencing many organisms, especially marine invertebrates, remains inherently challenging (Chakraborty et al. (2020)). Marine organisms retain chemical composites within their tissue that interfere with both common DNA extraction methods as well as many downstream sequencing techniques (Boughattas et al. (2021)). Moreover, they often exhibit notably high heterozygosity and large numbers of repetitive regions within their genomes (e.g., transposable elements), underscoring the need for DNA extraction methods capable of producing high molecular weight DNA for long-read sequencing to overcome these genome assembly limitations (Angthong et al. (2020)).

Our objective was to test four common DNA extraction methods using the non-model organism Kellet's whelk, generate an optimal protocol for producing HMW DNA for next generation sequencing, apply this protocol to a large collection of samples, and test its suitability for whole genome sequencing as well as RAD-seq and GT-seq platforms across large sample sizes. Among the four tested DNA extraction

Sequence run	Estimated reads	Estimated bases	Data produced post-filtering	Reads passed filtering
KK_Rapidseq	14.01k	8.94Mb	0B	0
KK_Ligationseq	557.70k	1347Mb	2.6Gb	446.26k
KK_Ligationseq2	112.61k	724Mb	1.4Gb	98.57k

Table 3. Each sequencing run's output from MinKNOW sequencing software (Estimated reads and Estimated bases) and Fastp (Data produced post-filtering and Reads passed filtering). Reads produced by the KK_Ligationseq and KK_Ligationseq2 libraries that passed quality filtering were 80.02% and 87.53% respectively indicating similar quality of sequencing for both libraries. Lower yields for the second round of ONT ligation sequencing library was most likely due to an observed air bubble within the sequencing pore.

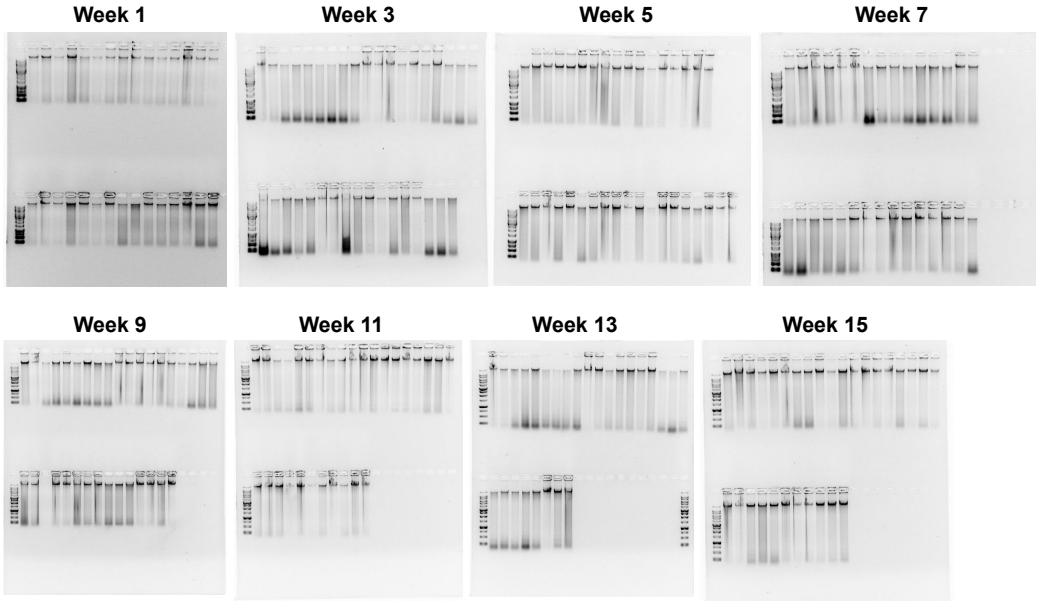

Figure 4. Quality check gel electrophoresis on randomly selected genomic DNA extractions bi-weekly. 1% gel with ∼250 ng of DNA and 1kb or XL DNA ladder in the first well of each row. See Supplementary Material section for sample location, concentration, and quality.

methods, the Salting out protocol stood out by producing the highest quality genomic DNA, while also being cost-effective and high yielding (Table 1). This method was improved by the use of simple low agitation mechanical tissue homogenization as well as a set duration for chemical tissue lysis with proteinase K (SI 10). The necessity of these improvements became evident during initial testing, as they played a vital role in producing HMW DNA, given that different methods for homogenization and tissue lysis yielded varying outcomes. Importantly, we validated this method using samples collected 3-5 years prior to extractions and stored at -80°C, which resulted in genomic DNA with high purity, as evident by A260/280 and A260/230 ratios suitable for population genetic platforms (Table 2). Notably, any signs of degradation observed in these samples collected years ago from disparate locations (Fig. 3) were not evident in samples produced in the lab from fresh tissue (Fig. 2). This finding strongly suggests that compromised DNA quality stems not from the extraction method but rather from DNA degradation during tissue collection or storage.

Conducting extractions on all 3,325 tissue samples using the Salting out protocol produced DNA with sufficient quality for RAD-seq and GT-seq, despite some variability among samples (Fig. 4). We attribute this variability to potential factors such as human error, low tissue input, and suboptimal storage conditions (frozen on dry ice and subjected to multiple transfers before being stored at -80°C). Nevertheless, the Salting out method yielded a majority (81.3%) of DNA extractions of optimal quality for RAD-seq and GT-seq platforms. We anticipate that further improvements can be achieved by collecting samples under more consistent conditions and employing robotics rather than manual operators (trained students, see Methods), which may elevate the overall DNA quality across samples.

While the Salting out protocol produced HMW DNA extraction, it likely did not remove all contaminants, as indicated by the failed ONT rapid sequencing library and initial suboptimal GT-seq libraries (Table 3, Fig. 6). These contaminants may have potentially interfered with the polymerase activity of the MinION Nanopore technology and enzyme activity during library preparation for the Illumina technologies (Healey et al. (2014)). With the introduction of additional cleanup steps, the ONT ligation sequencing libraries and Illumina Novaseq library successfully achieved high-quality genome sequencing. Moreover, following the same cleanup process, 80% of the failed GT-seq samples from the initial run produced successful genotyping (>90% call rate) (Fig. 6). The bead cleaning steps in combination with the PCR step employed in the library preparation of the RAD-seq libraries appeared to exhibit resistance

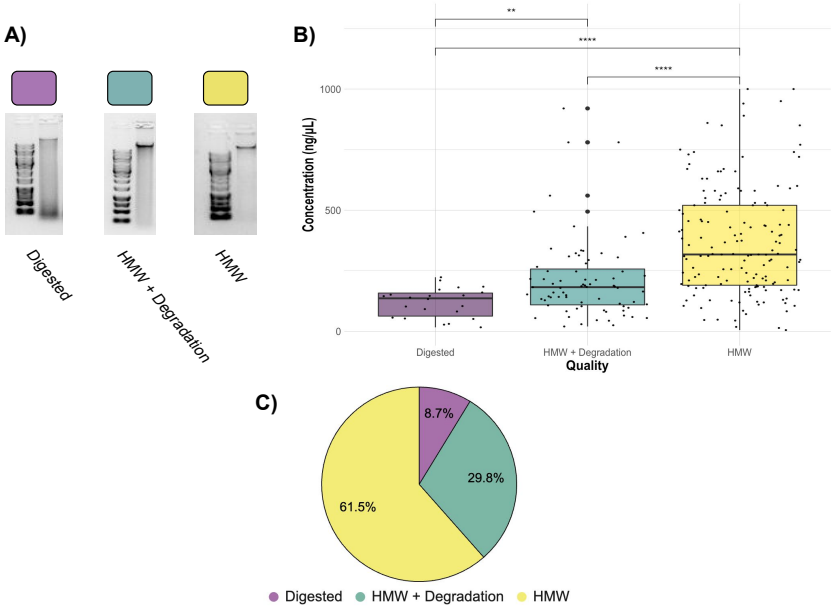

Figure 5. Assessment of 252 randomly selected DNA extractions **A)** Examples of Digested, HMW + Degradation, and HMW quality genomic DNA extractions determined by gel electrophoresis **B)** Quality assessment of genomic DNA extractions. Average concentration of Digested extractions (22 total samples), HMW + Degradation extractions (75 total samples), and HMW extractions (155 total samples). The top of the box represents the 75th percentile and the bottom represents the 25th percentile. Whiskers represent the highest and lowest values of the sample group. Kruskal Wallis tests are indicated by horizontal lines and asterisks (****: $p \leq 0.0001$, **: $p \leq 0.01$). **C)** Pie chart displaying percentage of the 252 randomly-selected samples of the 3,325 DNA extractions within each quality assessment group: Digested, HMW + Degradation or HMW (Table S2).

to inhibition by co-precipitants during sequencing. This observation implies a cleanup step is necessary before conducting PCR library amplification in order to facilitate successful sequencing.

The issue of inhibition or blockage of Nanopore pores was also observed in the black tiger prawn, Penaeus monodon, believed to be caused by repetitive regions in the genome (Van Quyen et al. (2020)). Successful Nanopore sequencing occurred after PCR-based library preparation, suggesting that there may be unknown contaminants associated with mollusc DNA that are removed or destroyed during PCR (Adema (2021)). Similarly, we encountered comparable challenges in producing successful genome sequencing using the Nanopore MinION without added cleanup and purification. Our cleanup, which did not include PCR steps and maintained the integrity of our HMW DNA, allowing secondary structures to form prior to sequencing, is believed to remove the unidentified contaminant responsible for the failed Nanopore MinION sequencing runs (SI 14-15).

Our comparison of DNA extraction methods could have benefited from a method for isolating or quantifying co-precipitants. These co-precipitants, most likely a mucopolysaccharide (Sokolov (2000)), were not significantly implicated by the A260/280 ratio (A260/280 =1.85±0.046) or the A260/230 ratio (A260/230 = 1.81±0.27) and were only clearly evident after sequencing, not being implicated by PCR failure (Popa et al. (2007)). To prevent future failed sequencing runs, we strongly recommend implementing a method for detecting these co-precipitants (2D gel-electrophoresis) or ensuring their removal (using the added clean-up, HMW selection kits, or PCR). Other methods known for removing polysaccharides, such as the CTAB method (Chakraborty et al. (2020); Huelsken et al. (2011)), should also be explored, as well as other modified methods being published regularly, given the increasing number of non-model marine invertebrates under genetic analysis (Ardura et al. (2017); Panova et al. (2016); Angthong et al. (2020)).

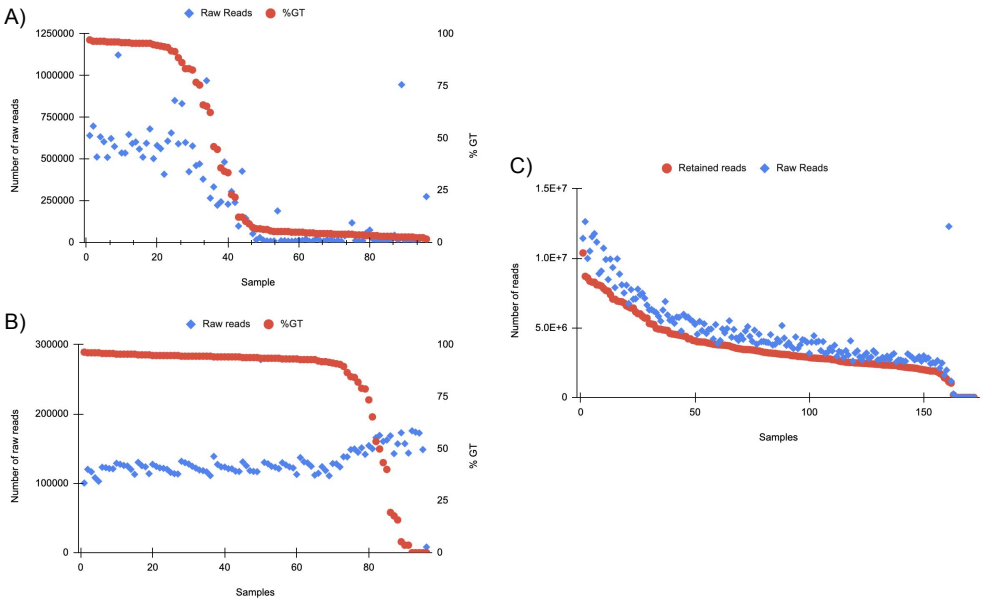

Figure 6. **A)** GT-seq tested samples in order of percent genotyped (%GT) from scaled DNA extractions. 96 total samples were tested (see sup. data). Varied read distribution indicates unsuccessful sequencing especially in samples 40+. **B)** GT-seq on the same 96 samples after Zymo clean and concentrator. Samples are ordered by %GT. Distribution of raw reads across samples indicates much more consistent and successful sequencing results. **C)** Samples tested on RAD-seq platform. Reads that passed the quality filter and contained the RADcutsite are indicated by Retained reads. Reads are ordered by retained reads. Overall, an average of 78.84% of reads were retained after filtering for quality and RADcutsite. 94.77% of samples produced >180,000 raw reads.

By applying whole genome sequencing using GT-seq and RAD-seq to the extensive collection of Kellet's whelk adult and recruit samples from populations spanning the U.S. and Mexico coasts, southern and central California, and the species' historical and expanded range, over multiple years covering ENSO and non-ENSO oceanographic conditions, researchers can hope to increase there understanding of population connectivity in this and other coastal marine species with extensive larval dispersal (Christie et al. (2017)). However, accomplishing such large-scale sequencing data on a non-model marine invertebrate species of this scale is challenging. This study serves as an important contribution towards enhancing our understanding of the performance and potential use of DNA extraction methods for future large-scale genomics.

SUPPLEMENTARY MATERIAL

Supplementary Data

Data

Table S1 - Metadata on Samples Extracted

Table S2 - Scaled Extractions QC

Table S3 - GT-seq Raw DNA

Table S4 - GT-seq Cleaned DNA

Table S5 - RAD-seq Raw DNA

Table S6- GT% vs Conc. & Location

Methods and Protocols

Files

SI 7 - Qiagen DNeasy Blood and Tissue kit

SI 8 - Zymo Quick-DNA HMW Magbead kit

SI 9 - PCI Protocol

SI 10 - Salting Out Protocol

SI 11 - ToBo lab ezRAD Protocol_v3.2

SI 12 - Protocol ezRAD whelks

SI 13 - Nanopore MinION Rapid Sequencing Kit

SI 14 - Zymo Clean and Concentrator

SI 15 - PacBio Short Read Eliminator kit

SI 16 - Nanopore MinION Ligation Sequencing kit

SI 17 - Illumina NEBNext Ultra II DNA Library Prep kit

SI 18 - Nate's Plates Protocol

Code and Sequence Data

SI 17 - Github

Sequence Read Archive

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1924537.

The 13 undergraduate students that conducted the DNA extractions in the Center for Biotechnology at California Polytechnic State University: Jaden Hansen, Anabel Sanchez, Gabriella Richardson, Hanna Jaynes, Tyler Weipert, Alyssa Queen, Kathryn Hutchinson, Olivia Watt, Jordan Reichhardt, Chanel De Smet, Alli Clark, Jenna Nurge, and Olivia Sleeper.

The RAD-sequencing was carried out by the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

The GT-sequencing testing was carried out by GTseek with special thanks to Nathan Campbell.

The Illumina Novaseq Genome sequencing was carried out by the Genomics & Cell Characterization Facility (GC3F) at the University of Oregon with special thanks to Jeff Bishop.

REFERENCES

Adema, C. M. (2021). Sticky problems: extraction of nucleic acids from molluscs. *Philosophical Transactions of the Royal Society B*, 376(1825):20200162.

Akinde, S. B., Adeyemi, F. M., Oyedara, O. O., Ajani, T. F., Sule, W. F., Olaitan, J. O., Iwalokun, B. A., and Ojo, O. O. (2022). 16s rrna metagenomics data on the bacterial communities in integrated poultry-fish farm ponds. *Data in Brief*, 42:108269.

Allen, L. G., Pondella, D. J., and Horn, M. H. (2006). *The ecology of marine fishes: California and adjacent waters*. Univ of California Press.

Angthong, P., Uengwetwanit, T., Pootakham, W., Sittikankaew, K., Sonthirod, C., Sangsrakru, D., Yoocha, T., Nookaew, I., Wongsurawat, T., Jenjaroenpun, P., et al. (2020). Optimization of high molecular weight dna extraction methods in shrimp for a long-read sequencing platform. *PeerJ*, 8:e10340.

Ardura, A., Zaiko, A., Morán, P., Planes, S., and Garcia-Vazquez, E. (2017). Epigenetic signatures of invasive status in populations of marine invertebrates. *Scientific Reports*, 7(1):42193.

Aseltine-Neilson, D., Brady, B., Connell, M., Kalvass, P., Laughlin, L., Leos, B., Mello, J., Neillands, G., Roberts, E., Taniguchi, I., et al. (2006). Review of some california fisheries for 2005: Coastal pelagic finfish, market squid, dungeness crab, sea urchin, abalone, kellet's whelk, groundfish, highly migratory species, ocean salmon, nearshore live-fish, pacific herring, and white seabass. *California Cooperative Oceanic Fisheries Investigations Reports*, 47:9–29.

Bootsma, M. L., Gruenthal, K. M., McKinney, G. J., Simmons, L., Miller, L., Sass, G. G., and Larson, W. A. (2020). A gt-seq panel for walleye (sander vitreus) provides a generalized workflow for efficient development and implementation of amplicon panels in non-model organisms. *BioRxiv*, pages 2020–02.

Boughattas, S., Albatesh, D., Al-Khater, A., Giraldes, B. W., Althani, A. A., and Benslimane, F. M. (2021). Whole genome sequencing of marine organisms by oxford nanopore technologies: Assessment and optimization of hmw-dna extraction protocols. *Ecology and Evolution*, 11(24):18505–18513.

Chakraborty, S., Saha, A., and Neelavar Ananthram, A. (2020). Comparison of dna extraction methods for non-marine molluscs: is modified ctab dna extraction method more efficient than dna extraction kits? *3 Biotech*, 10:1–6.

Christie, M. R., Meirmans, P. G., Gaggiotti, O. E., Toonen, R. J., and White, C. (2017). Disentangling the relative merits and disadvantages of parentage analysis and assignment tests for inferring population connectivity. *ICES Journal of Marine Science*, 74(6):1749–1762.

Cowen, R. K., Gawarkiewicz, G., Pineda, J., Thorrold, S. R., and Werner, F. E. (2007). Population connectivity in marine systems an overview. *Oceanography*, 20(3):14–21.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in ecology & evolution*, 29(1):51–63.

Ferrara, G. B., Murgia, B., Parodi, A. M., Valisano, L., Cerrano, C., Palmisano, G., Bavestrello, G., and Sara, M. (2006). The assessment of dna from marine organisms via a modified salting-out protocol. *Cellular & molecular biology letters*, 11:155–160.

Fogarty, M. J. and Botsford, L. W. (2007). Population connectivity and spatial management of marine fisheries. *Oceanography*, 20(3):112–123.

Froese, R. and Pauly, D. (2011). Fishbase. 2011. *World wide web electronic publication*.

Halpern, B. S., Cottenie, K., and Broitman, B. R. (2006). Strong top-down control in southern california kelp forest ecosystems. *Science*, 312(5777):1230–1232.

Harley, C. D., Randall Hughes, A., Hultgren, K. M., Miner, B. G., Sorte, C. J., Thornber, C. S., Rodriguez, L. F., Tomanek, L., and Williams, S. L. (2006). The impacts of climate change in coastal marine systems. *Ecology letters*, 9(2):228–241.

Healey, A., Furtado, A., Cooper, T., and Henry, R. J. (2014). Protocol: a simple method for extracting next-generation sequencing quality genomic dna from recalcitrant plant species. *Plant methods*, 10(1):1–8.

Herrlinger, T. (1981). Range extension of kelletia kelletii. *Veliger*, 24(1):78.

Huelsken, T., Schreiber, S., and Hollmann, M. (2011). Coi amplification success from mucus-rich marine gastropods (gastropoda: Naticidae) depends on dna extraction method and preserving agent. *Mitteilungen der Deutschen Malakozoologischen Gesellschaft*, 85(July 2011):17–26.

Li, Y., Wang, W., Liu, X., Luo, W., Zhang, J., and Gul, Y. (2011). Dna extraction from crayfish exoskeleton.

Lu, H., Giordano, F., and Ning, Z. (2016). Oxford nanopore minion sequencing and genome assembly. *Genomics, proteomics & bioinformatics*, 14(5):265–279.

López, C., Daniels, B. N., Lee, A., White, C., Christie, M., Davidson, J., Anderson, P., and Toonen, R. (2023). Quantifying population connectivity during a climate-driven northern range expansion. Presented at Human Impacts on Marine Functional Connectivity, Sesimbra, Portugal.

Malde, K. (2014). Estimating the information value of polymorphic sites using pooled sequences. *BMC genomics*, 15:1–11.

Nagel, M. M., Sewell, M. A., and Lavery, S. D. (2015). Differences in population connectivity of a benthic marine invertebrate evechinus chloroticus (echinodermata: Echinoidea) across large and small spatial scales. *Conservation Genetics*, 16:965–978.

Panova, M., Aronsson, H., Cameron, R. A., Dahl, P., Godhe, A., Lind, U., Ortega-Martinez, O., Pereyra, R., Tesson, S. V., Wrange, A.-L., et al. (2016). Dna extraction protocols for whole-genome sequencing in marine organisms. *Marine genomics: Methods and protocols*, pages 13–44.

Park, S. T. and Kim, J. (2016). Trends in next-generation sequencing and a new era for whole genome sequencing. *International neurourology journal*, 20(Suppl 2):S76.

Popa, O. P., Murariu, D., Popa, L. O., et al. (2007). Comparison of four dna extraction methods from invasive freshwater bivalve species (mollusca: Bivalvia) in romanian fauna. *Travaux du Muséum National d'Histoire Naturelle "Grigore Antipa*, 6:527–36.

Reitzel, A., Herrera, S., Layden, M., Martindale, M., and Shank, T. (2013). Going where traditional markers have not gone before: utility of and promise for rad sequencing in marine invertebrate phylogeography and population genomics. *Molecular ecology*, 22(11):2953–2970.

Rivera-Colón, A. G. and Catchen, J. (2022). Population genomics analysis with rad, reprised: Stacks 2. In *Marine Genomics: Methods and Protocols*, pages 99–149. Springer.

Scientific, T. F. (2009). Nanodrop 2000/2000c spectrophotometer v1. 0 user manual. *Wilmington, DE*, 19810.

Sokolov, E. P. (2000). An improved method for dna isolation from mucopolysaccharide-rich molluscan tissues. *Journal of Molluscan Studies*, 66(4):573–575.

Song, W., Thomas, T., and Edwards, R. J. (2019). Complete genome sequences of pooled genomic dna from 10 marine bacteria using pacbio long-read sequencing. *Marine genomics*, 48:100687.

Van Quyen, D., Gan, H. M., Lee, Y. P., Nguyen, D. D., Nguyen, T. H., Tran, X. T., Khang, D. D., Austin, C. M., et al. (2020). Improved genomic resources for the black tiger prawn (penaeus monodon). *Marine genomics*, 52:100751.

Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R., and Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, placopecten magellanicus, using rad seq-derived snp s. *Evolutionary Applications*, 10(1):102–117.

White, C. and Toonen, R. J. (2008). Isolation and characterization of polymorphic microsatellite and coi loci from the whelk kelletia kelletii. *Molecular ecology resources*, 8(4):881–883.

Zacherl, D., Gaines, S. D., and Lonhart, S. I. (2003). The limits to biogeographical distributions: insights from the northward range extension of the marine snail, kelletia kelletii (forbes, 1852). *Journal of Biogeography*, 30(6):913–924.