Degree project

UNIVERSITY OF SKÖVDE

# Sequencing the genome of *Unio tumidus* for future pollution impact research

Master's Degree Project in Bioscience
Second Cycle, 30 credits

Academy Year 2026

Adedamola Johnson

Supervisor: Magnus Fagerlind

Examiner: Sanja Jurcevic

# Abstract

Freshwater mussels play vital ecological roles by improving water clarity, cycling nutrients, and serving as bioindicators of environmental health. Among them, the swollen river mussel (*Unio tumidus*) is one of the least impacted species, while many other unionids continue to decline due to pollution, habitat degradation, and climate change. To establish a genomic foundation for ecological and conservation research, this study presents the first draft genome of *U. tumidus*, providing valuable resources for future genomic and environmental studies. Mussels were collected from Lake Vindsjön, Sweden, and high-molecular-weight DNA was extracted from foot tissue using the Monarch® HMW DNA Extraction Kit. DNA integrity was verified by pulsed-field gel electrophoresis, and sequencing libraries were prepared using the Oxford Nanopore Ligation Sequencing Kit before sequencing on a MinION Mk1B platform. A total of seven million reads were generated, with read lengths reaching up to 188 kb across both Flongle and MinION runs. De novo assembly using Flye produced a 461 Mb genome with a GC content of 34.3%, consisting of 71,708 contigs and an N50 of 9.8 kb. Although the assembly was fragmented, it effectively demonstrated successful sequencing and assembly of *U. tumidus* genomic DNA. Further improvements in contiguity could be achieved through deeper sequencing coverage or hybrid scaffolding approaches, establishing a foundation for future comparative and conservation genomics of *Unionidae* mussels.

# Popular Scientific Summary

River and lake mussels are keystone species and important biofilters in freshwater ecosystems. Unionid mussels and their relatives make up a large portion of aquatic invertebrate biomass and support many ecosystem services. By grazing, mussels remove zooplankton and phytoplankton, algae, bacteria, and fine suspended particles improving water clarity and cycling nutrients. Because they filter large volumes of water, mussels also accumulate pollutants such as heavy metals and pesticides, which makes them useful bioindicators of aquatic health.

A common European species, the swollen river mussel (*Unio tumidus*), is currently listed as Least Concern by the IUCN. Unlike many freshwater mussels that have suffered sharp declines from habitat change, invasive species, and climate effects, *U. tumidus* has remained relatively stable. Still, it faces ongoing risks from environmental stressors. Sequencing its genome can reveal genetic traits that help explain resilience and vulnerability, which in turn improves our understanding of how stressors affect populations.

To provide a template genome for *U. tumidus*, tissue from a mussel was used to extract high-quality DNA and sequence it using Oxford Nanopore technology. One portable device often used with this technology, the MinION, produces very long DNA reads in this study some reads were as long as 188 kilobases and that length helps assemble difficult parts of the genome. (A “read” is a single continuous stretch of DNA sequence produced by the machine.)

The initial assembly produced about 461 million base pairs with a GC content of 34.3% (GC content is the proportion of the bases guanine and cytosine, which is a basic property often compared across species). The draft assembly is still fragmented into more than 71,000 pieces called contigs short contiguous sequences that will need to be joined together to form complete chromosomes. With deeper sequencing and methods that link pieces across larger distances (for example Hi-C or hybrid approaches that combine different types of data), many of those pieces can be ordered and connected into longer scaffolds to produce a more complete genome.

This draft is an important first step. Ten years ago, a project like this would have been financially and technically unrealistic; now draft genomes can be produced quickly and at much lower cost. That matters for conservation and environmental monitoring: a reference genome for *U. tumidus* makes it possible to identify genes involved in stress tolerance or pollutant processing, develop genetic markers for tracking populations, and detect early signs of decline. In short, genome data can improve targeted conservation actions and give managers better tools for monitoring freshwater health.

# Table of Contents


1 Introduction ..... 1

2 Aim ..... 6

3 Materials and Methods ..... 7
Sample Preparation ..... 7
Pulsed-Field Gel Electrophoresis ..... 7
Library Preparation ..... 8
Nanopore Sequencing ..... 9
Genome Assembly and Quality Assessment ..... 9

4 Results ..... 11

5 Discussion ..... 17
DNA Quality and Sequencing Readiness ..... 17
Effects of Mechanical Shearing and DNA Handling ..... 17
Library Preparation ..... 17
Sequencing Output ..... 18
Assembly Characteristics and Implications ..... 18
Comparison with Other Unionid Genome Projects ..... 19
Biological and Conservation Relevance ..... 19
Limitations and Future Directions ..... 20
Conclusions ..... 20

6 Ethical Aspects and Impact on Society ..... 21

7 Future Perspective ..... 22

8 References ..... 23

# Abbreviations

**bp** – Base pair

**BUSCO**- Benchmarking Universal Single-Copy Orthologs

**DNA** – Deoxyribonucleic Acid

**HMW** – High Molecular Weight

**IUCN** – International Union for Conservation of Nature

**kb** – Kilobase

**Mb** – Megabase

**NGS** – Next-Generation Sequencing

**ONT** – Oxford Nanopore Technologies

**PCR** – Polymerase Chain Reaction

**PFGE** – Pulsed-Field Gel Electrophoresis

**TBE** – Tris-Borate-EDTA

**TGS** – Third-Generation Sequencing

**WGS** – Whole Genome Sequencing

# INTRODUCTION

Freshwater mussels are essential filter feeders that play a critical role in aquatic ecosystems by removing suspended particles from the water column. Through their filtration process, mussels enhance water clarity, regulate nutrient cycles, and influence both primary and secondary production, thereby maintaining biogeochemical cycles and sedimentation rates (Aldridge et al., 2023). In addition, they serve as ecosystem engineers that modify habitat structures and water quality. However, rising levels of pollution, habitat destruction, and climate change have placed mussel populations under severe stress worldwide, leading to significant declines in many species (Guo et al., 2023). *Unio tumidus* is widely distributed throughout European freshwater systems and is notable for its relative resilience compared to other unionid species that have experienced marked population declines (Ollard & Aldridge, 2023). Classified as "Least Concern" by the International Union for Conservation of Nature (IUCN), *U. tumidus* is found in both pristine and contaminated habitats, such as the contrasting river basins in Ukraine (Martyniuk et al., 2022). However, localized pressures from pollution, habitat alteration, and exposure to contaminants like pesticides could affect its density and size over time, underscoring the importance of continuous monitoring to ensure its long-term viability (Chmist et al., 2019). Freshwater mussels display remarkable species diversity and high levels of endemism (Bolotov et al., 2019), setting them apart from the large marine habitats of most other bivalves. Understanding how these species have adapted to diverse freshwater environments, such as rivers, streams, and lakes, is therefore essential to their conservation, and genomic data offer a way to investigate this.

Advances in genomic technologies have opened new avenues for investigating the genetic basis of adaptation in freshwater mussels. These advances build on a long history of DNA sequencing (Figure 1). The foundation for DNA sequencing was established in 1952 when Chase and Hershey confirmed DNA as the hereditary material (Burian et al., 2021), and the first sequencing methods were introduced in 1977 by Sanger and by Maxam and Gilbert. Second-generation sequencing, or next-generation sequencing (NGS), emerged in 2005 using a sequencing-by-synthesis approach that generates millions of short, high-accuracy reads (Pareek et al., 2011). Although NGS is excellent for detecting small-scale variants, most methods rely on polymerase chain reaction (PCR) amplification, which can introduce biases and remove native DNA modifications, and its short reads can complicate the assembly of repetitive regions and the detection of large structural variants (Scarano et al., 2024). Third-generation sequencing (TGS) was developed to overcome these limitations.

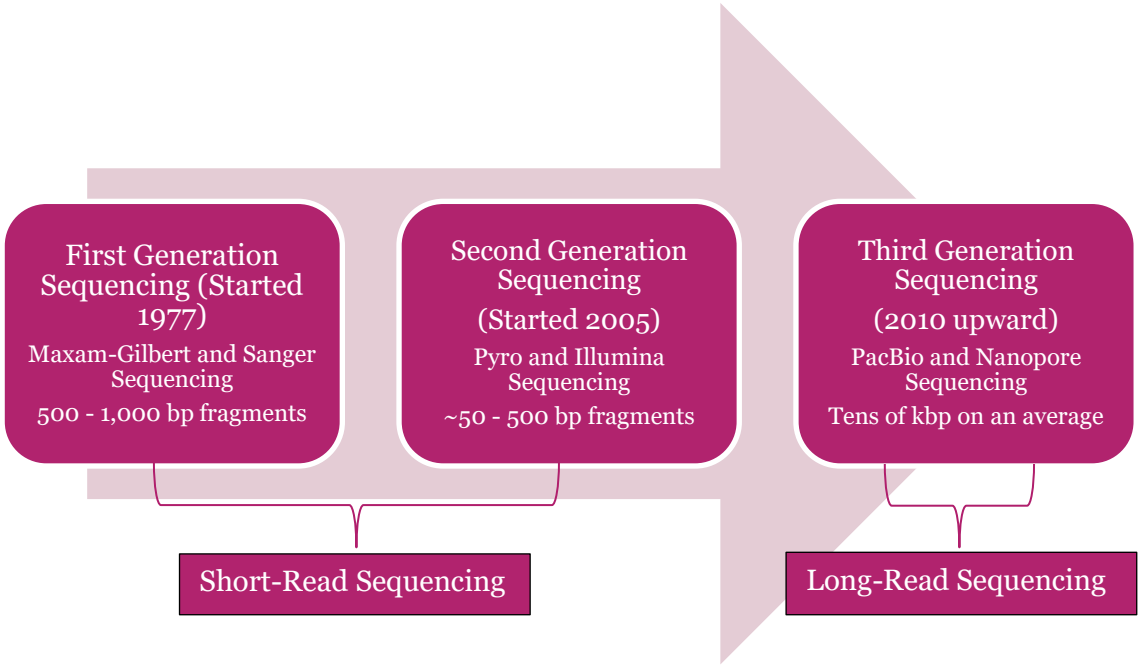


**Figure 1:** Evolution of DNA sequencing technologies from first-generation to third-generation platforms.

Unlike earlier methods, TGS platforms sequence single native DNA or RNA molecules in real time, without requiring PCR amplification, thereby reducing the associated biases (Scarano et al., 2024). The two primary TGS platforms are Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). PacBio HiFi mode generates long reads with extremely high per-base accuracy, while ONT enables the production of ultra-long reads that can reach tens to hundreds of kilobases using portable devices capable of real-time data generation (Scarano et al., 2024; Lin et al., 2021a). A key advantage of TGS is therefore its ability to produce long reads, often averaging more than 10 kilobases, which greatly facilitates *de novo* genome assembly and the detailed analysis of complex genomic regions, including large insertions, deletions, and repetitive sequences. TGS can also resolve structural variants and directly detect epigenetic modifications by sequencing native DNA or RNA molecules without prior conversion or amplification (Scarano et al., 2024; Burian et al., 2021; Lin et al., 2021a). These capabilities facilitate the study of genome architecture, particularly in non-model organisms such as freshwater mussels (Burian et al., 2021; Scarano et al., 2024).

Although early versions of TGS technologies had higher raw error rates than NGS platforms, improvements in chemistry, pore design, and consensus algorithms, including PacBio HiFi and ONT duplex/polished workflows, have substantially increased accuracy, making TGS suitable for use either alone or as the long-read backbone in hybrid assemblies (Lin et al., 2021a; Wenger et al., 2019). Among these platforms, ONT devices such as MinION provide additional advantages because of their portability, low cost, and real-time sequencing capability, and they can generate long reads in field settings, making them suitable for rapid genomic applications beyond the laboratory (Lamb et al., 2021). Studies have shown that ONT assemblies can achieve comparable accuracy to Illumina sequencing for single-nucleotide variants while also detecting

structural variants that are often missed by short-read technologies (Razook et al., 2020). This makes ONT particularly powerful for ecological genomics, conservation, and real-time environmental monitoring, and it is the platform used in this study.

Whichever platform is used, the first step of a successful whole-genome sequencing (WGS) project is the extraction of high-molecular-weight (HMW) DNA. This procedure usually involves physical disruption of the tissue, nuclear enrichment to concentrate nuclear DNA, and chemical treatments to remove contaminants, resulting in long, purified DNA fragments (Kang et al., 2023). The quality and integrity of these fragments are typically assessed using techniques such as pulsed-field gel electrophoresis (PFGE) (Lin et al., 2021b). PFGE is an advanced technique that separates large DNA fragments by periodically changing the direction of the electric field. In this process, the electric field forces the DNA molecules to reorient and migrate through the gel matrix based on size, with smaller molecules moving faster than larger ones. This high-resolution separation technique is valuable for structural analysis and DNA fingerprinting (Sharma Kuinkel et al., 2016). Once HMW DNA of sufficient quality has been obtained, library preparation transforms it into sequencing-ready templates through end-repair and adapter ligation steps, allowing the DNA to be efficiently read by high-throughput sequencing platforms (Goodwin et al., 2016).

Following nanopore sequencing, genome assembly and bioinformatics processing are required to reconstruct the genome from the generated sequencing reads. This process involves multiple stages, including preassembly, consensus sequence generation, and consensus polishing. An overview of this workflow, from sample collection and DNA extraction through PFGE, sequencing, and assembly to downstream quality assessment, is illustrated in Figure 2. The choice of assembler matters: in a study by Mochizuki et al. (2023), long-read-only assemblers including Canu (v2.1.1) and Flye (v2.8.3) were benchmarked across six diploid genomes spanning a range of heterozygosity levels (0.055-3.00%), using PacBio long-read data. Canu showed reduced contig continuity in genomes with heterozygosity below 1%, while Flye achieved higher continuity in that same lower-heterozygosity range (Mochizuki et al., 2023). While no freshwater mussel genome was included in this benchmark, these findings provide a general basis for assembler selection in diploid, non-model organisms with comparatively low heterozygosity. These assessments were complemented by evaluations of computational resource usage and completeness metrics, guiding the choice of assemblers based on genome characteristics (Cosma et al., 2023). These advances now make chromosome-scale assemblies accessible even for large, heterozygous genomes.

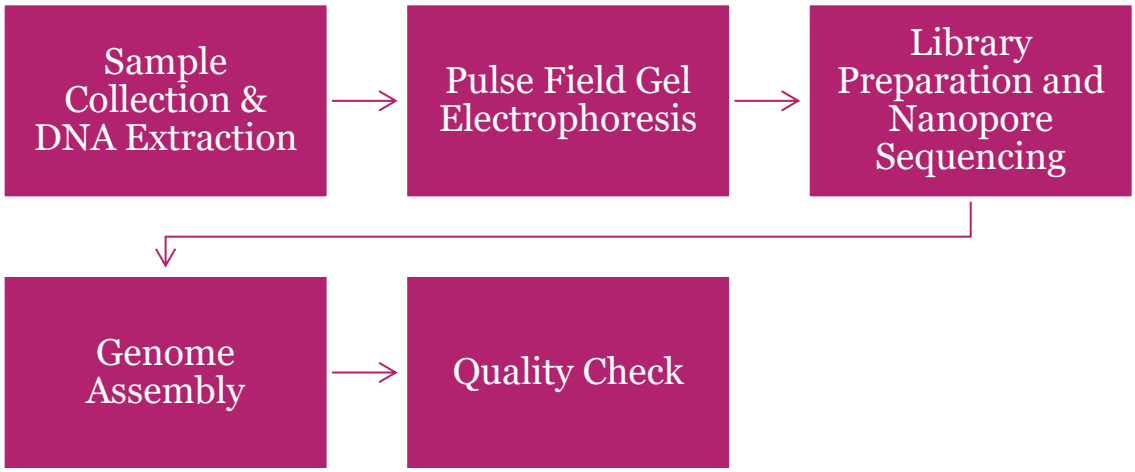


**Figure 2.** Experimental workflow for *Unio tumidus* genome sequencing: from sample collection and HMW DNA extraction to PFGE verification, Nanopore sequencing, genome assembly, and quality assessment

Once an assembly has been produced, its quality must be evaluated. QUAST (Quality Assessment Tool for Genome Assemblies) is a widely used bioinformatics tool for evaluating and comparing genome assemblies (Gurevich et al., 2013). It generates a range of summary statistics that describe the structural quality and completeness of an assembly, including the number of contigs, total assembly length, N50 and L50 values, GC content, and the largest contig size. When a reference genome is provided, QUAST can additionally assess accuracy by identifying misassemblies, mismatches, indels, and genome fraction covered, allowing direct comparison between the assembly and the expected sequence. In reference-free mode, it instead relies on assembly-intrinsic metrics to indicate contiguity and completeness. QUAST is commonly applied after genome assembly (for example, using tools such as SPAdes, Unicycler, or Flye) to benchmark assembler performance, compare different assembly parameters, or validate assembly quality before downstream analyses such as annotation or comparative genomics. In this study, QUAST was used to evaluate the assembly produced after sequencing.

Sequencing the genome of *Unio tumidus* will generate a comprehensive genomic blueprint that forms the foundation for conservation and pollution-impact research. A reference genome provides an ordered, annotated catalog of genes, allowing gene expression and genetic variation data to be mapped onto specific molecular pathways involved in detoxification, oxidative stress, and immune defense, which enables the development of biomarkers for pollution response (Rogers et al., 2021). It also facilitates the creation of population-genetic markers for assessing gene flow, resilience, and vulnerability across habitats, supporting conservation planning for freshwater mussels (Gomes-dos-Santos et al., 2021). Furthermore, the genome allows integration with transcriptomic studies, making it possible to link environmental stressors to gene

expression changes, while environmental DNA (eDNA) approaches enable monitoring of species presence and health in aquatic ecosystems (Razook et al., 2020). Because freshwater mussels are sensitive bioindicators of aquatic health, such high-resolution genomic data can reveal genetic responses to pollution, such as shifts in stress-response genes or changes in gene expression profiles under environmental stress. Recent studies have demonstrated the effectiveness of such approaches; for instance, Roznere et al. (2023) used integrated transcriptomic and metabolomic profiling in freshwater mussels to identify biomarkers responsive to contaminants, highlighting the potential of genomic resources in environmental monitoring. Together, these capabilities (biomarker discovery, population-genetic assessment, and integration with transcriptomic and eDNA monitoring) make the *U. tumidus* genome a molecular resource for both investigating how pollution affects mussel biology and guiding management strategies to protect declining unionid populations.

The genomic resources generated through long-read sequencing and draft genome assembly therefore provide a foundation for exploring genetic diversity, adaptation mechanisms, and evolutionary resilience in *U. tumidus*. These resources ultimately support the development of targeted conservation strategies aimed at mitigating the impacts of pollution on this species' populations (Harvey et al., 2023; Dorado et al., 2021). Recent work has also demonstrated the ecological potential of long-read genomics across a wide range of organisms. For example, large-scale sequencing of lepidopteran pest genomes has revealed the genetic basis of host-plant adaptation, invasion biology, and resistance traits, offering valuable insights for both ecological research and species management (Zhang et al., 2022). Collectively, these studies highlight how long-read sequencing enables the creation of robust genomic resources for previously understudied taxa, thereby advancing both conservation genomics and broader ecological understanding.

# Aim and Objectives

This study aims to construct a reference genome for *Unio tumidus*, a freshwater mussel species collected from Vindsjön in Sweden. By sequencing and assembling the genome of this species, the study aims to generate a genomic resource that supports conservation and environmental monitoring efforts

The objectives of this study are as follows:

- Obtain high-quality, high-molecular-weight DNA from *U. tumidus* tissue for long-read sequencing.
- Prepare sequencing libraries for Oxford Nanopore sequencing.
- Assemble the reads into a draft *U. tumidus* genome.
- Evaluate the assembled genome's quality and completeness.

# Materials and Methods

## Sample Preparation

Fresh specimens of *Unio tumidus* were collected from Lake Vindsjön, Skövde, Sweden. Two mussels were transported in a water-filled jar for each extraction to maintain their viability. Tissue sampling was performed by carefully opening the shells with a scalpel, exposing the soft tissues, and excising the foot, which was identified by its distinct muscular appearance compared to other tissues. Excised foot tissues were transferred into six 1.5 ml Eppendorf tubes, 3 stored on ice for immediate processing and 3 frozen at -20 °C for 48 hours to compare both recommended techniques from the Monarch HMW DNA Extraction kit protocol. These two storage conditions were compared to assess their effect on DNA quality after extraction.

## DNA Extraction

Three mussels were used for the test extraction step, with two tissue samples taken from each mussel. For each sample, approximately 25 mg of foot tissue was weighed and processed using the Monarch HMW DNA Extraction Kit for Tissue (New England Biolabs), following the manufacturer's tissue protocol. As part of the initial optimisation, two tissue storage methods were compared: three samples were stored on ice and three were frozen, to determine which method performed better.

Following the test extraction, a new set of six extractions was performed on three new mussels (M1, M2 and M3), using fresh tissue stored on ice, the method selected based on observations from the test extraction. Two tissue samples were taken from each mussel and labelled by mussel and sample number: M1-1 and M1-2 (mussel 1), M2-1 and M2-2 (mussel 2), and M3-1 and M3-2 (mussel 3). The protocol was followed exactly, with a

single modification: the lysis incubation was extended by 10 minutes beyond the manufacturer's recommended 5 minutes to improve HMW DNA yield and integrity.

After extraction, and prior to any clean-up or library preparation, DNA concentration and purity were measured for each extract. Concentration was quantified on a Qubit 4.0 fluorometer using the dsDNA High-Sensitivity Assay Kit (Thermo Fisher Scientific); three technical replicate readings were taken per sample and the mean value was reported. Purity was assessed on a NanoDrop 2000 spectrophotometer (Thermo Scientific) and recorded as A260/280 and A260/230 ratios.
The three extracts with the best DNA quality, M2-1, M2-2 and M3-1, were selected for fragment-size and DNA integrity assessment. Each selected extract was divided into two aliquots: one was left untreated (suffix U) and the other was mechanically sheared (suffix S). Because sheared DNA is expected to show smaller fragments, a clear size shift between the two aliquots would confirm that the untreated DNA was HMW.
For mechanical shearing, two 0.5-mm metal beads were added to a 1.5-mL Eppendorf tube containing the DNA aliquot. The tubes were vortexed for 3 min 00 s (M2-1-S), 3 min 30 s (M2-2-S) and 4 min 00 s (M3-1-S). The corresponding untreated aliquots were labelled M2-1-U, M2-2-U and M3-1-U. All six samples were then analysed by pulsed-field gel electrophoresis (PFGE) to compare their DNA fragment-size distributions and assess the integrity of the HMW DNA.

## Pulsed-Field Gel Electrophoresis (PFGE)

DNA fragment length was assessed using a CHEF-DR III Pulsed-Field Gel Electrophoresis System (Bio-Rad Laboratories) with a MidRange PFG Marker ranging from 50 kb to 6 Mb (Bio-Rad Laboratories). For each sample, 12 µg of DNA was mixed with 6× Purple Loading Dye (Thermo Fisher Scientific) to obtain a final loading-dye concentration of 1×.
A 1% agarose gel was prepared in 0.5× Tris-Borate-EDTA (TBE) buffer. A total of 2.2 L of 0.5× TBE was used for preparation of the gel and to fill the electrophoresis chamber. Electrophoresis was performed at 14 °C with a field strength of 6 V/cm for 24 hours using the CHEF-DR III system.
The MidRange PFG Marker was loaded in the lanes labelled L and used as the molecular-size reference. Samples A–F were loaded in the remaining lanes. Following electrophoresis, the gel was stained with GelRed nucleic acid stain (Biotium) according to the manufacturer's instructions and imaged using a ChemiDoc MP Imaging System (Bio-Rad Laboratories).
Following electrophoresis, gels were stained with GelRed nucleic acid stain (Biotium) according to the manufacturer's protocol. Gel images were captured on a ChemiDoc MP Imaging System (Bio-Rad), which provides the required resolution for PFGE gels.

## Library Preparation

Nanopore sequencing libraries were prepared from the selected high-molecular-weight (HMW) DNA sample using the Ligation Sequencing Kit V14 (SQK-LSK114, Oxford Nanopore Technologies). A preliminary Flongle sequencing run was first performed to evaluate library quality and overall DNA suitability for long-read sequencing before committing to a full MinION run. Both preparations followed the manufacturer's ligation sequencing protocol, including the use of Long Fragment Buffer (LFB) for long DNA fragments (Oxford Nanopore Technologies).

For the full MinION library preparation, the manufacturer's SQK-LSK114 ligation sequencing protocol was modified. The modification consisted of doubling the manufacturer's recommended reagent volumes used in the DNA repair and ligation reactions, while maintaining the same DNA input and final elution volume. The Long Fragment Buffer (LFB) cleanup procedure was retained as specified in the manufacturer's protocol.

The DNA concentration was measured at three stages of the library preparation: immediately before library preparation, following the bead-cleanup step, and after final elution. DNA concentration was measured using a Qubit 4.0 fluorometer with the dsDNA High-Sensitivity Assay Kit (Thermo Fisher Scientific). The final library was kept on ice and loaded onto the flow cell immediately after preparation.

Quantification was performed using a Qubit 4.0 fluorometer and the dsDNA High-Sensitivity Assay kit(Invitrogen, Thermo Fisher Scientific, USA), which provides accurate measurements of double-stranded DNA and avoids the overestimation commonly associated with absorbance-based quantification (Thermo Fisher Scientific). Tracking concentrations through these steps provided a direct assessment of whether doubling reagent volumes improved overall library retention.

Modifying reagent volumes in ligation-based protocols can influence enzymatic reaction kinetics, buffer composition, and ligation efficiency, potentially affecting downstream sequencing performance.

Final sequencing libraries were stored on ice and loaded immediately onto the MinION instrument following standard Oxford Nanopore loading procedures (Oxford Nanopore Technologies).

## Nanopore Sequencing

For the Flongle test, a Flongle adapter was inserted into the MinION Mk1B device and controlled through MinKNOW software (Oxford Nanopore Technologies). The prepared library was loaded onto a Flongle flow cell, which contained 53 active pores seen from the flow cell check. Sequencing was performed for 24 hours with live basecalling enabled in MinKNOW using a minimum Q score threshold of 10, which represents Oxford Nanopore Technologies' standard quality cutoff for distinguishing high-confidence reads from lower-quality signals.

For the main MinION sequencing run, a standard R10.4.1 MinION flow cell was inserted into the MinION Mk1B device, and a flow cell check was performed, confirming 1,437 active pores prior to sequencing. A total of 150 fmol of the prepared library was loaded

according to the SQK-LSK114 protocol. Sequencing was run for 48 hours with live basecalling enabled at a minimum Q score of 10. All sequencing data generated were transferred to an external hard drive for downstream analysis.

## Genome Assembly and Quality Assessment

Basecalled FASTQ files generated from MinKNOW were merged into a single dataset using the Linux cat command. The merged FASTQ file was uploaded to the usegalaxy.eu (The Galaxy Community, 2026) platform for assembly and downstream analysis, allowing computationally intensive processing to be performed without requiring a local high-performance computing environment.

De novo genome assembly was performed using the Flye assembler (Kolmogorov et al., 2019) with default parameters consisting of the genome size which was set to auto, minimum overlap which also auto determined, threads and polishing iterations which were set to 1. The assembly process completed in 7 hours and 42 minutes, generating an assembled genome in FASTA format. Assembly quality was then assessed using QUAST in the reference free mode, which calculated key metrics including total assembly length, number of contigs, largest contig size, N50, L50, and GC content. The QUAST analysis was conducted on the usegalaxy.eu platform to ensure standardized, reproducible evaluation of the assembly.

# Results

## Test extraction: fresh versus frozen tissue

To establish which tissue storage method gave the better DNA, a test extraction was performed on foot tissue from three mussels, with six extractions in total: three from fresh tissue stored on ice (Fresh 1-3) and three from frozen tissue (Frozen 1-3). DNA concentration was quantified by fluorometry (Qubit) and purity was assessed by spectrophotometry (NanoDrop). Table 1 summarizes the Qubit concentrations together with the A260/280 and A260/230 purity ratios for all six extractions.

Fresh tissue gave higher DNA concentrations than frozen tissue in every comparison. The Qubit concentrations of the fresh extracts were 23.2, 35.6 and 85.2 ng/µL (mean 48.0 ng/µL), whereas those of the frozen extracts were 15.2, 23.8 and 12.5 ng/µL (mean 17.2 ng/µL). The mean concentration of the fresh extracts was therefore about 2.8 times that of the frozen extracts. The highest concentration overall, 85.2 ng/µL, came from a fresh extract and the lowest, 12.5 ng/µL, from a frozen extract, a difference of nearly seven-fold. The fresh extracts also varied more between samples (23.2 to 85.2 ng/µL) than the frozen extracts (12.5 to 23.8 ng/µL), which indicates that yield differed between individual animals even when the same protocol and tissue amount were used. With only three extractions per storage condition, the comparison is descriptive and was not tested statistically.

Fresh tissue stored on ice gave the higher yields and was therefore chosen as the storage method for all later extractions.

**Table 1.** DNA concentration and purity measurements of *Unio tumidus* samples extracted from frozen and fresh tissue. Concentration was measured with a Qubit 4.0 fluorometer (mean of three replicate readings), while purity was assessed using a NanoDrop spectrophotometer.

| **Extraction Method** | **Frozen Sample** | | | **Fresh Sample** | | |
|---|---|---|---|---|---|---|
| | **Frozen 1** | **Frozen 2** | **Frozen 3** | **Fresh 1** | **Fresh 2** | **Fresh 3** |
| **A260/230** | 0.39 | 0.89 | 0.39 | 2.24 | 1.92 | 1.87 |
| **A260/280** | 0.86 | 1.45 | 0.86 | 2.07 | 2.20 | 1.75 |
| **DNA Concentration (Qubit) ng/µL** | 15.2 | 23.8 | 12.5 | 23.2 | 35.6 | 85.2 |

## DNA extraction from three new mussels

Following the test extraction, six new DNA extractions were performed on fresh foot tissue, stored on ice, from three new mussels (M1, M2 and M3). Two tissue samples were taken from each mussel, giving six extracts labelled M1-1, M1-2, M2-1, M2-2, M3-1 and M3-2, where the first number identifies the mussel and the second the tissue sample. Table 2 reports the DNA concentration (Qubit, mean of three technical readings) and the purity ratios (NanoDrop A260/280 and A260/230) for each extract.

The two methods describe different aspects of extract quality. Qubit measures double-stranded DNA specifically and therefore indicates how much usable DNA is present, whereas the NanoDrop ratios indicate how much protein, salt or organic carry-over accompanies it. Extract quality was judged from these measurements together. On this basis, three extracts were selected as having the best DNA quality: M2-1 and M2-2, both from mussel 2, and M3-1 from mussel 3. The remaining extracts (M1-1, M1-2 and M3-2) were not analysed further. The three selected extracts were carried forward to fragment-size assessment by PFGE.

**Table 2.** DNA concentration and purity measurements of *Unio tumidus* samples extracted from fresh tissue. Concentration was measured with a Qubit 4.0 fluorometer, while purity was assessed using a NanoDrop spectrophotometer.

| | **Samples** | | | | | |
|---|---|---|---|---|---|---|
| | **M1-1** | **M1-2** | **M2-1** | **M2-2** | **M3-1** | **M3-2** |
| **A260/230** | 2.24 | 1.92 | 2.26 | 2.06 | 2.11 | 2.17 |
| **A260/280** | 2.07 | 2.20 | 1.98 | 2.18 | 2.26 | 2.09 |
| **DNA Concentration (Qubit) ng/µL** | 23.2 | 32.7 | 48.9 | 60.2 | 57.9 | 65.4 |

## Fragment size assessment by PFGE

The three selected extracts were analysed by PFGE, each in two states, giving six sample lanes on a single gel (Figure 3): the untreated aliquots M2-1-U, M2-2-U and M3-1-U, and the sheared aliquots M2-1-S, M2-2-S and M3-1-S, which had been vortexed with beads for 3 min 00 s, 3 min 30 s and 4 min 00 s, respectively. A size ladder (L) spanning 50 kb to 6 Mb was run in the marker lane, which allowed fragment sizes to be judged far beyond the range of conventional agarose gels. Each sheared aliquot was compared with the untreated aliquot of the same extract.

The untreated aliquots retained prominent high-molecular-weight bands. M3-1-U showed a pronounced band above the ladder bands, whereas M2-1-U and M2-2-U

showed broader signal distributed above and below the 50 kb marker, so that their DNA was spread over a wider range of sizes. Several of the lanes also contained material migrating in the region of hundreds of kilobases.

The sheared aliquots showed more pronounced smearing than their untreated counterparts, meaning that their DNA was spread continuously over a range of sizes and not concentrated in distinct bands. Even so, no lane showed a smear made up only of low-molecular-weight DNA, and in every lane, including the sheared ones, the main signal corresponded to fragments larger than 50 kb.

All three extracts therefore showed the same pattern: a clear difference between the untreated and the sheared aliquot, with the untreated DNA retaining larger fragments. This confirms that the extracts contained HMW DNA. The persistence of the main signal above 50 kb after shearing also suggests that vortexing with beads under these conditions reduced fragment size only partly.

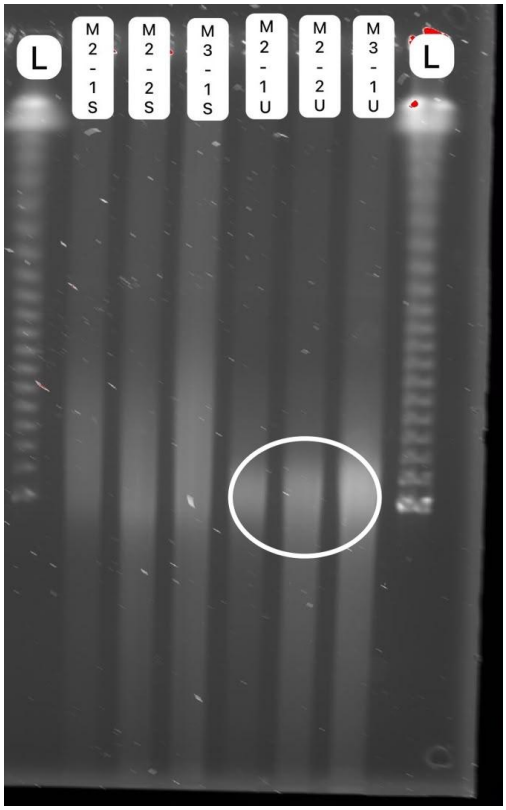


**Figure 3**. **Pulsed-field gel electrophoresis (PFGE) of genomic DNA extracted from *Unio tumidus*.** Lanes labelled L contains the MidRange RFD Marker ladder with fragment sizes ranging from 50 kb to 6 Mb.

## DNA concentration during library preparation

To follow the DNA through library preparation, its concentration was measured at three points in the workflow (Table 3): immediately before library preparation, after the end-prep and ligation clean-up steps, and in the final eluted library.

The starting sample had a concentration of 23 ng/μL. After the end-prep and ligation clean-up steps, the concentration was 15.2 ng/μL, a decrease of 7.8 ng/μL, which corresponds to about 66% of the starting concentration. The final eluted library had a concentration of 52.8 ng/μL, which was about 3.5 times the post-clean-up concentration and 2.3 times the starting concentration. Concentration therefore fell after the clean-up and rose after the final elution, and the final library was then taken forward to sequencing.

These values are concentrations, not total amounts of DNA. Because the sample volume can differ at each step, a higher or lower concentration does not by itself mean that more or less DNA was present, and the changes cannot be read as DNA recovery without taking the volumes into account. Total DNA is obtained by multiplying each concentration by the volume at the same step.

**Table 3.** Progression of DNA concentration for library preparation during cleanup to final elution.

| | **Starting Concentration** | **After Cleanup** | **Final Concentration** |
|---|---|---|---|
| **DNA Concentration (Qubit) ng/μL** | 23 | 15.2 | 52.8 |

## Nanopore sequencing output

The final library was sequenced on a MinION flow cell, which generated a broad distribution of read lengths (Figure 4, with total bases on the y-axis). The distribution was strongly skewed towards short reads: most reads were shorter than 10 kb, and a tail of much longer reads extended up to 188 kb. The read length N50 was 2.61 kb, meaning that half of all sequenced bases were in reads of 2.61 kb or longer. The longest read was therefore about 70 times the N50, which shows that very long molecules were sequenced but made up only a small part of the data.

Sequencing yield over time is shown in Figure 5. The cumulative number of reads increased steadily throughout the 48-hour run and reached approximately 7 million by the end. Reads are classified as “passed” (above the quality threshold, green), “failed” (below the threshold, red) or “skipped” (grey, typically short fragments or pore problems), and each category is plotted separately alongside the total. The number of passed reads kept rising over the whole run and did not plateau, which indicates that the pores remained productive and that sequencing output was sustained. A substantial

fraction of the reads passed the quality (Q-score) filter, while the rest failed or were excluded. Overall, the run produced a high read count with a balance between passed and failed reads that is typical of nanopore sequencing.

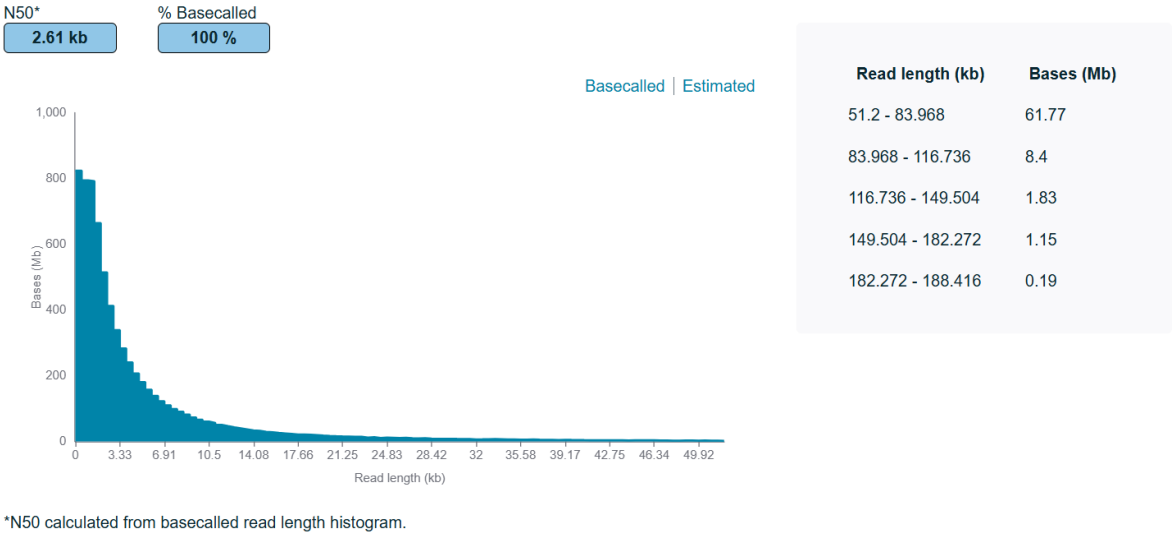


**Figure 4. Read length distribution of Nanopore sequencing data.** Histogram shows the total bases generated across different read lengths, with an N50 of 2.61 kb. The longest 1% of reads are classified as outliers and summarized in the accompanying table, with the longest read reaching 188 kb.

**Figure 5. Sequencing yield over time.** The histogram shows sequencing yield over time from Nanopore MinION run. Total, passed, failed, and skipped reads are shown

## De novo genome assembly

The raw reads were assembled de novo with Flye, and the assembly was evaluated with QUAST (Table 4). The assembly had a total length of 461.2 Mb and a GC content of 34.31%, which indicates an AT-rich sequence composition. It was made up of 71,708 contigs, of which 66,301 (92.5%) were 1 kb or longer and 5,407 were shorter than 1 kb. The mean contig length was about 6.4 kb, and the largest contig was 103.3 kb, which is only a very small part of the total assembly length.

Contiguity is described by the N50 and N90 values and their corresponding L values. The N50 is the contig length at which half of the assembly is contained in contigs of that length or longer; here it was 9,793 bp (about 9.8 kb). The corresponding L50 was 13,529, meaning that half of the assembly lies in the 13,529 largest contigs, or about 19% of all contigs. The N90 is the same measure for 90% of the assembly; it was 2,982 bp, with an L90 of 46,586, so that 90% of the assembly lies in the 46,586 largest contigs, or about 65% of all contigs. The large number of contigs needed to reach each threshold, together with an N50 below 10 kb, shows that the assembly is highly fragmented.

These statistics therefore describe a fragmented draft assembly, which provides a first sequence resource for *Unio tumidus*. QUAST was used here only to assess assembly size and continuity, so no contamination or misassembly metrics are reported.

**Table 4.** Assembly statistics of the *Unio tumidus* genome generated with Flye, evaluated using QUAST.

| Metric | Value |
|---|---|
| Total number of contigs (all sizes) | 71,708 |
| Number of contigs ≥ 1 kb | 66,301 |
| Total assembly length | 461.2 Mb |
| Largest contig | 103.3 kb |
| GC content (%) | 34.31 |
| N50 | 9,793 bp |
| L50 | 13,529 |
| N90 | 2,982 bp |
| L90 | 46,586 |

# Discussion

## DNA extraction and tissue storage

HMW DNA is a prerequisite for long-read sequencing, because fragment length and integrity directly determine read length, yield and assembly contiguity (Logsdon et al., 2021; Trigodet et al., 2022). In the test extraction, fresh foot tissue gave higher DNA concentrations than frozen tissue in all three comparisons (Table 1). Purity followed the same pattern: the fresh extracts had A260/280 ratios close to the accepted range for pure DNA [state values] and higher A260/230 ratios, whereas the frozen extracts had markedly depressed A260/230 ratios [state values]. Low A260/230 ratios are typically associated with carry-over of polysaccharides, phenolic compounds or other organic contaminants, which can be released by freeze-thaw damage and which suppress end-repair and adapter ligation in ligation-based nanopore workflows (Amarasinghe et al., 2020; Wang et al., 2021). Fresh tissue stored on ice was therefore used for all subsequent extractions (Table 2). This comparison rests on only three extractions per storage condition, and the test extracts were not compared by PFGE, so the difference is established for yield and purity but not for DNA integrity.

## HMW DNA integrity and the shearing comparison

PFGE provided information on DNA integrity that concentration and purity cannot. The sheared aliquots (M2-1-S, M2-2-S and M3-1-S) showed more smearing than their untreated counterparts (M2-1-U, M2-2-U and M3-1-U), which retained prominent HMW bands (Figure 3). This supports the presence of HMW DNA in extracts M2-1, M2-2 and M3-1. It also shows that the gel could detect fragmentation, so the HMW bands in the untreated lanes are unlikely to be a gel artefact.

The comparison has limits. PFGE is qualitative, so these results support DNA integrity but do not prove that every molecule was intact. Each vortexing time was applied to a different extract, so the effect of shearing duration cannot be separated from differences between extracts, and no dose-response can be concluded. The main signal in the sheared lanes also stayed above 50 kb, which suggests that vortexing with beads reduced fragment size only partly. A more controlled shearing method, with several treatment times applied to the same extract, would give a more informative comparison.

## Library preparation

Library preparation is a major bottleneck for non-model organisms with limited input DNA. Reagent volumes in the end-prep and ligation steps were doubled while the elution volume was kept constant, to improve enzymatic efficiency and reduce proportional losses during bead clean-up. Concentration after the end-prep and ligation clean-up was about 66% of the starting concentration, and the final library reached 52.8 ng/µL (Table 3). Measuring concentration at each checkpoint allowed losses to be identified before sequencing, not afterwards in the output data (Wang et al., 2021). Two cautions apply. Concentration is not total mass, so these values only approximate. Without an

unmodified control library, the benefit of doubling reagent volumes cannot be separated from the effect of the elution volume.

## Sequencing output

The MinION run produced about seven million reads with lengths from below 1 kb up to 188 kb (Figure 4), and passed reads accumulated steadily throughout the run (Figure 5). The read N50 of 2.61 kb is consistent with reports that nanopore datasets contain many short reads and a minority of very long ones, particularly when the input fragment size is moderate (Rang et al., 2022). Reads of up to 188 kb show that long molecules survived to sequencing, and such ultra-long reads contribute disproportionately to contiguity by spanning repeats (Logsdon et al., 2021; Nurk et al., 2022).

The read lengths were nevertheless much shorter than the PFGE profile predicts. The untreated aliquot of M3-1, which was used for library preparation [confirm], had its main signal above 50 kb and a prominent band above the ladder bands (Figure 3), yet most reads were shorter than 10 kb. Because the sequenced aliquot was unsheared, deliberate shearing cannot explain this. Several other causes are possible, and this study did not test them:

- **Breakage during handling and library preparation.** Pipetting, mixing, end-prep, adapter ligation and bead clean-up can all break long DNA.
- **Single-strand nicks.** Nanopores read one strand, so a nick can end a read even when the double-stranded molecule looks intact on a native gel.
- **Contaminants.** Residual contaminants can block pores and terminate reads early [compare with the purity of M3-1 in Table 2].
- **Capture bias.** Shorter molecules are often captured by pores more readily than very long ones, so the sequenced population can be shorter than the DNA present in the tube.

Tracking fragment size by PFGE at more points in the workflow, and not only after extraction, would show where length is lost.

## Assembly quality and genome size

The Flye assembly contained 71,708 contigs (66,301 of them ≥1 kb) with a total length of 461.2 Mb (Table 4). The N50 of 9.8 kb is only about four times the read N50, which is consistent with contiguity being limited by read length. The largest contig (103.3 kb) is also shorter than the longest read (188 kb). This suggests that the longest reads had too little overlapping support to assemble intact, possibly because the long-read fraction was thin. Such fragmentation is expected for a nanopore-only assembly at modest depth without polishing or scaffolding (Nurk et al., 2022; Rhie et al., 2021), and it limits the resolution of gene order, regulatory regions and repeat families, which make up a large part of bivalve genomes (Rogers et al., 2021; Gomes-dos-Santos et al., 2021).

The total length of 461.2 Mb should not be read as the genome size of *Unio tumidus*. Sequenced unionid genomes are typically around 1.6 to 2.4 Gb, for example about 2.4 Gb in *Margaritifera margaritifera* (Gomes-dos-Santos et al., 2021, 2023). This assembly therefore probably represents only a fraction, roughly a fifth to a quarter, of the genome. The most likely causes are low sequencing depth and short reads, so that much of the genome was not covered well enough to assemble. Reporting the total sequenced bases and the expected coverage would make this clearer, and the genome size should be estimated independently, for example by k-mer analysis or flow cytometry.

## Comparison with other unionid genome projects

Higher-quality unionid assemblies, such as *M. margaritifera* (about 2.4 Gb, contig N50 3.4 Mb) and *Unio pictorum*, used PacBio long reads, generally combined with Illumina polishing and, for chromosome-scale work, Hi-C scaffolding. These address the three main constraints on assembly quality: read accuracy, read length and long-range connectivity (Rhie et al., 2021; Gomes-dos-Santos et al., 2023). The contrast with the present draft reflects the sequencing strategy and depth, not any property of the *U. tumidus* genome.

## Biological and conservation relevance

Unionid mussels are filter feeders that process large volumes of water and contribute to nutrient cycling, phytoplankton regulation and sediment stabilisation. Their accumulation of contaminants makes them useful bioindicators, but it also exposes them to cumulative toxic burdens that can affect survival, reproduction and immune function over time (Aldridge et al., 2023; Guo et al., 2023; Ollard & Aldridge, 2023). Freshwater systems in Europe face growing pressure from agricultural runoff, urbanisation and climate change (Theissinger et al., 2023; Tickner et al., 2020), which makes molecular resources for freshwater mussels increasingly relevant to conservation.

In principle, a draft genome allows the identification of candidate genes and gene families involved in xenobiotic metabolism, oxidative stress and immune defence, such as cytochrome P450s and ABC transporters, which have been reported to differ between unionid species under different pollution pressures (Rogers et al., 2021; Gomes-dos-Santos et al., 2021). These candidates could then be used as targets in transcriptomic studies of gene expression under controlled pollutant exposure. The sequence data can also be used to develop SNP and microsatellite markers for population studies, enabling assessment of gene flow, effective population size and adaptive variation between sites that differ in pollution history and habitat quality, which is increasingly regarded as essential for evidence-based conservation planning (Theissinger et al., 2023). Given the fragmentation and incompleteness described above, such work should follow a completeness assessment such as Benchmarking Universal Single-Copy Orthologs (BUSCO), and conclusions drawn from absent genes should be avoided.

## Limitations and future directions

The main limitations are the fragmented and apparently incomplete assembly, the lack of BUSCO-based completeness assessment (Manni et al., 2021), and the lack of polishing, so residual nanopore base errors probably remain (Hu et al., 2020). The genome comes from a single individual, and the PFGE assessment and fresh-versus-frozen comparison were qualitative and based on few samples.

Future work should:

- run BUSCO and polish the assembly (for example with Medaka or NextPolish);
- increase depth, including runs optimised for ultra-long reads, and add PacBio HiFi data;
- add Hi-C or optical mapping data for scaffolding;
- keep DNA longer through the workflow (gentle handling, extra clean-up, size selection);
- sample several individuals from different populations, and combine genomic data with transcriptomics under controlled pollutant exposure (Rogers et al., 2021).

## Conclusions

This study produced a draft genome assembly of *Unio tumidus* from Oxford Nanopore data. Fresh foot tissue gave higher DNA yields and better purity than frozen tissue, and PFGE analysis of sheared and untreated aliquots showed that the extracts contained HMW DNA. The MinION run generated about seven million reads, with a read N50 of 2.61 kb and a longest read of 188 kb, and the Flye assembly comprised 71,708 contigs with an N50 of 9.8 kb and a total length of 461.2 Mb.

The assembly is fragmented and, at 461.2 Mb, probably represents only part of the genome. Related unionids have much larger genomes, for example about 2.4 Gb in *Margaritifera margaritifera* (Gomes-dos-Santos et al., 2023). The sequencing reads were also much shorter than the long fragments seen on PFGE, which shows that long fragments were not fully preserved through library preparation and sequencing. Overall, the workflow showed that fresh tissue, careful handling and measuring DNA at each stage are important for long-read sequencing.

The draft is a starting point for further work. Completeness assessment with BUSCO, polishing, deeper sequencing and scaffolding would improve its quality and make it suitable for gene annotation and population studies aimed at understanding pollution resilience and supporting the conservation of freshwater mussels.

# Ethical Aspects and Impact on The Society

The ethical aspects of this research were considered throughout sample collection, laboratory handling, data processing, and reporting. Protecting environmental and biological integrity was central to the study, particularly because the freshwater mussel *Unio tumidus* is a native species that contributes to ecosystem stability. Mussels were collected from Lake Vindsjön, Sweden, with minimal disturbance to the habitat and in line with sustainable sampling practices. Only six individuals were collected, and care was taken to avoid harming the local population, since removing animals from a small or sensitive population could affect its long-term stability.

Because DNA extraction required fresh tissue, each mussel was killed individually and rapidly to minimize stress and suffering. The shell was opened and the animal was cut open immediately with a sterile scalpel, and tissue was collected straight afterwards, so that no animal was kept alive in the laboratory longer than necessary. Individuals were processed one at a time. No ethical approval was required, as freshwater mussels are invertebrates and fall outside the scope of animal experimentation legislation. Nevertheless, the work followed the principles of the 3Rs (replacement, reduction, refinement) by using the smallest number of individuals sufficient for the study and by ensuring a quick death. The same principle guided the laboratory work: the extraction protocol was optimized on a small test set before the final extractions, so that tissue was not wasted on repeated attempts. This approach is consistent with the principle of "do no harm" in ecological research, which promotes biodiversity conservation while obtaining the biological material needed for scientific purposes (European Commission, 2018).

Ethical responsibility also extends to how the data are processed and reported. The results are presented as they were obtained, including the limitations of the assembly, which is fragmented and probably covers only part of the genome. Presenting these shortcomings openly avoids overstating the value of the draft genome and gives future users an accurate basis for deciding how it can be used.

The project can also benefit society. Freshwater mussels filter large volumes of water, help cycle nutrients and stabilize sediment, and are widely used as indicators of water quality, yet they are among the most threatened animal groups worldwide. The draft genome produced here is one of the early resources for *Unio tumidus* and provides a starting point for conservation genetics, such as assessing genetic diversity, population structure and responses to pollution. Because related unionid species are severely threatened, a genome from a more widespread species can also serve as a reference for comparative studies that help to understand why some mussels cope with environmental pressure better than others. In the longer term, this knowledge can support evidence-based management of freshwater ecosystems and water-quality monitoring, which are of direct relevance to public health, fisheries and recreation, as healthy lakes and rivers depend on the services that mussel populations provide.

The study also has value beyond this species. The DNA extraction and library preparation workflow was optimized for a non-model organism with limited input material, and the observations on tissue storage, gentle handling and quality control at each step may help other researchers who work with similar species. Nanopore sequencing is relatively affordable and accessible compared with many other long-read platforms, so showing

that it can produce a draft genome in a small project may encourage similar work in laboratories with limited resources. Finally, sharing the sequence data through public repositories, in line with open-science practice, would make the resource available for future research, teaching and conservation planning, so that the benefit of sampling six animals is passed on to as many users as possible.

# Future Perspectives

The draft genome generated for *Unio tumidus* provides a foundation for several targeted future research directions. A primary next step is improving assembly contiguity and base accuracy by increasing sequencing depth and integrating complementary sequencing technologies. The current results show that high molecular weight DNA can be obtained from fresh tissue and that Nanopore sequencing yields both short and ultra long reads. Building on this, additional Nanopore runs optimized for ultra long reads, combined with high accuracy PacBio HiFi sequencing, would substantially improve contig continuity and consensus accuracy, as demonstrated in other molluscan genome projects (Wenger et al., 2019; Gomes dos Santos et al., 2023). The addition of Hi C or optical mapping data would further enable chromosome scale scaffolding and resolution of repetitive regions.

A more contiguous reference genome would support population genomic analyses aimed at assessing genetic diversity, population structure, and connectivity across environmental gradients. Such analyses are particularly relevant for freshwater mussels, which are sensitive to habitat fragmentation and pollution pressures (Rogers et al., 2021). Re-sequencing individuals from multiple populations using the improved reference would allow identification of loci associated with stress tolerance and local adaptation.

In parallel, integration of transcriptomic data under controlled pollutant exposure experiments would enable functional annotation of genes involved in detoxification, stress response, and immune function. Similar approaches have successfully linked gene expression patterns to contaminant exposure in other bivalves (Gomes dos Santos et al., 2021). Together, these future efforts would transform the current draft genome into a comprehensive genomic resource with direct applications in conservation genomics, biomonitoring, and freshwater ecosystem management.

# References


Aldridge, D. C., Ollard, I. S., Bespalaya, Y. V., Bolotov, I. N., Douda, K., Geist, J., Haag, W. R., Klunzinger, M. W., Lopes-Lima, M., Mlambo, M. C., Riccardi, N., Sousa, R., Strayer, D. L., Torres, S. H., Vaughn, C. C., Zając, T., & Zieritz, A. (2023). Freshwater mussel conservation: A global horizon scan of emerging threats and opportunities. Global Change Biology, 29(3), 575–589. https://doi.org/10.1111/gcb.16510

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long read sequencing data analysis. Genome Biology, 21, 30. https://doi.org/10.1186/s13059-020-1935-5

Bolotov, I. N., Konopleva, E. S., Vikhrev, I. V., Lopes-Lima, M., Bogan, A. E., Lunn, Z., Chan, N., Win, T., Aksenova, O. V., Gofarov, M. Y., Tomilova, A. A., & Kondakov, A. V. (2019). Eight new freshwater mussels (Unionidae) from tropical Asia. Scientific Reports, 9(1), 12053. https://doi.org/10.1038/s41598-019-48528-z

Burian, A. N., Zhao, W., Lo, T.-W., & Thurtle-Schmidt, D. M. (2021). Genome sequencing guide: An introductory toolbox to whole-genome analysis methods. Biochemistry and Molecular Biology Education, 49, 815–825. https://doi.org/10.1002/bmb.21561

Chmist, J., Szoszkiewicz, K., & Drożdżyński, D. (2019). Behavioural responses of *Unio tumidus* freshwater mussels to pesticide contamination. Archives of Environmental Contamination and Toxicology, 77, 432–442. https://doi.org/10.1007/s00244-019-00649-2

Cosma, B.-M., Zade, R. S. H., Jordan, E. N., van Lent, P., Peng, C., Pillay, S., & Abeel, T. (2023). Evaluating long-read de novo assembly tools for eukaryotic genomes: Insights and considerations. GigaScience, 12, 1–12. https://doi.org/10.1093/gigascience/giad100

Dorado, G., Gálvez, S., Rosales, T. E., Vásquez, V. F., & Hernández, P. (2021). Analyzing modern biomolecules: The revolution of nucleic-acid sequencing – Review. Biomolecules, 11(8), 1111. https://doi.org/10.3390/biom11081111

European Commission. (2018). Guidelines on ethics in research involving biodiversity and ecosystems. European Union Publications Office.

Gomes-dos-Santos, A., Afonso, S., Megías, M., Lopes-Lima, M., & Machado, A. M. (2021). The Crown Pearl: A draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). DNA Research, 28(2), dsab002. https://doi.org/10.1093/dnares/dsab002

Gomes-dos-Santos, A., Fonseca, E., Riccardi, N., Hinzmann, M., Lopes-Lima, M., & Froufe, E. (2023). PacBio Hi-Fi genome assembly of the painter's mussel *Unio pictorum* (Linnaeus, 1758). Genome Biology and Evolution, 15(7), evad116. https://doi.org/10.1093/gbe/evad116

Gomes-Dos-Santos, A., Lopes-Lima, M., Machado, A. M., Forest, T., Achaz, G., Teixeira, A., Prié, V., C Castro, L. F., & Froufe, E. (2023). The Crown Pearl V2: an improved

genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). *GigaByte (Hong Kong, China)*, *2023*, gigabyte81. https://doi.org/10.46471/gigabyte.81

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Guo, F., Ye, Y., Zhu, K., Lin, S., Wang, Y., Dong, Z., Yao, R., Li, H., Wang, W., Liao, Z., & Yan, X. (2023). Genetic diversity, population structure, and environmental adaptation signatures of Chinese coastal hard-shell mussel *Mytilus coruscus* revealed by whole-genome sequencing. International Journal of Molecular Sciences, 24(17), 13641. https://doi.org/10.3390/ijms241713641

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Harvey, W. T., Ebert, P., Ebler, J., Audano, P. A., Munson, K. M., Hoekzema, K., Porubsky, D., Beck, C. R., Marschall, T., Garimella, K., & Eichler, E. E. (2023). Whole-genome long-read sequencing downsampling and its effect on variant-calling precision and recall. Genome Research, 33, 2029–2040.

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: A fast and efficient genome polishing tool for long-read assembly. Bioinformatics, 36(7), 2253–2255. https://doi.org/10.1093/bioinformatics/btz891

Kang, M., Chanderbali, A., Lee, S., Soltis, D. E., Soltis, P. S., & Kim, S. (2023). High-molecular-weight DNA extraction for long-read sequencing of plant genomes: An optimization of standard methods. Applications in Plant Sciences, 11(3), e11528. https://doi.org/10.1002/aps3.11528

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology, 37(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Lamb, H. J., Ross, E. M., Nguyen, L. T., Lyons, R. E., Hayes, B. J., & Chandler, H. D. (2021). Genomic prediction using low-coverage portable Nanopore sequencing. Frontiers in Genetics, 12, 710269. https://doi.org/10.3389/fgene.2021.710269

Lin, B., Hui, J., & Mao, H. (2021a). Nanopore technology and its applications in gene sequencing. Biosensors, 11(7), 214. https://doi.org/10.3390/bios11070214

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M., Chaisson, M., & Pevzner, P. A. (2021b). Assembly of long error-prone reads using repeat graphs. Genome Research, 31(3), 517–527. https://doi.org/10.1101/gr.260604.119

Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2021). Long-read human genome sequencing and its applications. Nature Reviews Genetics, 22(10), 597–614. https://doi.org/10.1038/s41576-021-00367-3

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution, 38(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Martyniuk, V., Khoma, V., Matskiv, T., Baranovsky, V., Orlova-Hudim, K., Gylyte, B., et al. (2022). Indication of the impact of environmental stress on the responses of the bivalve mollusk *Unio tumidus* to ibuprofen and microplastics based on biomarkers of reductive stress and apoptosis. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 261, 109425. https://doi.org/10.1016/j.cbpc.2022.109425

Mochizuki, T., Sakamoto, M., Tanizawa, Y., Nakayama, T., Tanifuji, G., Kamikawa, R., & Nakamura, Y. (2023). A practical assembly guideline for genomes with various levels of heterozygosity. Briefings in Bioinformatics, 24(6), 1–13. https://doi.org/10.1093/bib/bbad337

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., & Phillippy, A. M. (2022). The complete sequence of a human genome. Science, 376(6588), 44–53. https://doi.org/10.1126/science.abj6987

Ollard, I., & Aldridge, D. C. (2023). Declines in freshwater mussel density, size and productivity in the River Thames over the past half century. Journal of Animal Ecology, 92, 112–123. https://doi.org/10.1111/1365-2656.13835

Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. Journal of Applied Genetics, 52(4), 413–435. https://doi.org/10.1007/s13353-011-0057-x

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2022). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biology, 23(1), 228. https://doi.org/10.1186/s13059-022-02793-2

Razook, M., Smith, L., & Thomas, C. (2020). Benchmarking Nanopore sequencing assemblies in environmental genomics. Frontiers in Genetics, 11, 567. https://doi.org/10.3389/fgene.2020.00567

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., et al. (2021). Towards complete and error free genome assemblies of all vertebrate species. Nature, 592(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rogers, R. L., Grizzard, S. L., Titus-McQuillan, J. E., Bockrath, K., Patel, S., Wares, J. P., & Garner, J. T. (2021). Gene family amplification facilitates adaptation in the freshwater unionid bivalve *Megalonaias nervosa*. Molecular Ecology, 30(5), 1155–1173. https://doi.org/10.1111/mec.15817

Roznere, I., Watters, G. T., & Lee, K. L. (2023). Multi-omics characterization of stress responses in freshwater mussels exposed to environmental contaminants. Ecotoxicology and Environmental Safety, 253, 114699. https://doi.org/10.1016/j.ecoenv.2023.114699

Scarano, C., Veneruso, I., De Simone, R. R., Di Bonito, G., Secondino, A., & D'Argenio, V. (2024). The third-generation sequencing challenge: Novel insights for the genomic sciences. Biomolecules, 14(5), 568. https://doi.org/10.3390/biom14050568

Sharma-Kuinkel, B. K., Rude, T. H., & Fowler, V. G., Jr. (2016). Pulse field gel electrophoresis. Methods in Molecular Biology, 1373, 117–130. https://doi.org/10.1007/7651_2014_191

Swedish Work Environment Authority. (2023). Biosafety and laboratory safety guidelines in Sweden. SWEA Publications.

The Galaxy Community. Galaxy for accessible, reproducible, and collaborative data analyses: 2026 update. *Nucleic Acids Research*, 2026; gkag469, https://doi.org/10.1093/nar/gkag469

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bomfleur, B., Brewer, M. S., Browne, L., Carle, J., & Leese, F. (2023). How genomics can help biodiversity conservation. Trends in Genetics, 39(7), 545–559. https://doi.org/10.1016/j.tig.2023.02.005

Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., Colloff, M. J., Colvin, J., Davies, J., Fluet-Chouinard, E., Froend, R., Grodska, H., Horne, A., Hughes, J., Jensen, O., Johnson, W. C., King, A., Marchand, M., Meitzen, K. M., & Young, L. (2020). Bending the curve of global freshwater biodiversity loss: An emergency recovery plan. BioScience, 70(4), 330–342. https://doi.org/10.1093/biosci/biaa002

Trigodet, F., Lolans, K., Fogarty, E., Shaiber, A., Morrison, H. G., Barreiro, L., Jabri, B., & Eren, A. M. (2022). High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. Molecular Ecology Resources, 22(5), 1786–1802. https://doi.org/10.1111/1755-0998.13588

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology, 39(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly. Nature Biotechnology, 37(10), 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Zhang, T., Xing, W., Wang, A., Zhang, N., Jia, L., Ma, S., & Xia, Q. (2022). Comparison of long-read methods for sequencing and assembly of lepidopteran pest genomes. International Journal of Molecular Sciences, 24(1), 649. https://doi.org/10.3390/ijms24010649