Degree project

UNIVERSITY OF SKÖVDE

NANOPORE SEQUENCING & DE NOVO GENOME ASSEMBLY OF THE DUCK MUSSEL *ANODONTA ANATINA*

Master Degree Project in Bioscience

Second Cycle 30 credits

Spring semester 2025

Student: Filiph Carlsson

Supervisor: Magnus Fagerlind

Examiner: Helene Lindholm

Abstract

Climate change, pollution and human interference have led to a decrease in biodiversity and species being threatened with extinction. Freshwater mussels are more susceptible to environmental changes than most due to them being filter feeders. Freshwater mussels are widespread and important to numerous and diverse ecosystems around the world. Despite this importance, there is a concerning lack of knowledge and, research on the genomes of different freshwater mussels. The aim of this research is to increase the knowledge of the genome for the duck mussel species *Anodonta anatina* through performing genome assembly. The investigation into the genome of the *A. anatina* started with sample preparation, continued with DNA extraction, Nanopore sequencing before genome assembly was performed and evaluated. The results of the experiment were the generation of several genome assemblies through the usage of Flye, which were then evaluated by QUAST and compared to other published genome assemblies. The evaluation of the assemblies showed smaller assembly sizes than what was desired based on an estimated genome size of 1.5-2.9 Gb. The estimated genome size range was based on recent published studies on other freshwater mussel species. With the most generous assumption of a genome size at 1.5 Gb, the largest assembly created in this experiment covers less than 33% and therefore did not yield a satisfactory assembly. However, it can potentially serve as a foundation for future research to improve upon with regards to the techniques used and through learning from problems encountered during the experiment.

Popular Scientific Summary

A topic that is very important and extremely relevant in today's society, all around the world, is the environment. Two of the most common topics in day-to-day life connected to the environment are climate change and pollution. Regardless of whether someone is rich or poor, the place of their birth or their age, the effects of these two are impossible to escape or ignore. Both of these factors are harming animals all around the world, many of which are important for human' survival.

One group of animals that are showcasing the effect of these problems is freshwater mussels. Freshwater mussels are due to environmental changes and other human-related factors such as pollution or infringement onto their habitats facing the possibility of extinction all around the globe. A decrease in their numbers has the potential to be detrimental due to how important they are for ecosystems all around the world. Freshwater mussels fulfil many different important roles including, for example, water filtration in the ecosystems they inhabit.

Despite the acknowledged importance of freshwater mussels for the environment, there is a lack of research when it comes to knowledge about the composition of DNA in many of the different freshwater mussel species. In the start of the century, a famous project to sequence the DNA composition of a human was performed. It took a large number of experts, a lot of work, many years and a substantial amount of money to complete. Technology has progressed a lot since then and similar projects can now be performed in smaller teams or even individually with a lot less resources, opening it up for use in smaller projects.

The lack of knowledge of the DNA composition of freshwater mussel extends to the duck mussel species denoted as *Anodonta anatina*, found commonly around Europe. The aim of this research is to increase the knowledge about the DNA composition of the *A. anatina* and through this help with environmental studies and guiding conservation efforts. This is achieved through, in very simplified terms, extraction of DNA from tissue that is excised from the *A. anatina*, the DNA is drawn out from the tissue to separate the DNA from other undesired stuff that are present in the tissue as, for example proteins. The DNA is then run through a machine that outputs and identifies the building blocks of DNA known as nucleotides. This output is in turn used with software and algorithms to put the building blocks back together to give an estimation of how the DNA composition looks for the *A. anatina.* There are a few reasons that it is important to improve the existing knowledge about the DNA composition of the *A. anatina*. Knowledge about the DNA composition allows for the design of studies that can be used to measure and indicate the presence of pollution in their habitat and potentially even identify, the specific pollution.

Unfortunately, while a composition, that is representative of a smaller part of the *A. anatina* was created, a representation of the whole DNA composition of the *A. anatina* was not achieved and therefore the conclusion of the experiment is that it was not completely successful. The generated DNA composition does however have the possibility of being useful for future projects. This is also the case with other knowledge obtained in the experiment with regards to the different processes and optimizations which could be used to inform and help future experiments. In other words, despite not fulfilling the original desire it can serve as a foundation for further research to improve upon.

Table of Contents

Introduction 1
- Environmental Challenges 1
- Freshwater Mussels 1
- Environmental Research 1
- Genome Knowledge 2
- Genome Size 2
- Genome Assembly Experimental Outline 3
- High Molecular Weight DNA Extraction 3
- Whole Genome Sequencing 3
- Nanopore Sequencing (TGS) 4
- Genome Assembly & Evaluation 4
- Research Aim 5

Materials and methods 6
- Collection of *A. anatina* 6
- Dissection 6
- High Molecular Weight DNA extraction 6
- Pulsed-Field Gel Electrophoresis Sample Preparation 7
- Pulsed-Field Gel Electrophoresis (PFGE) 7
- Nanopore sequencing (Library preparation and sequencing) 7
- Sequencing data processing 8
- Genome Assembly 8

Results 9
- DNA extraction 9
- Pulsed-Field Gel Electrophoresis 9
- Nanopore Sequencing 10
- Genome Assembly 12

Discussion 13
- Preparation 13
- Species Identification & Collection 13
- Extraction & Pulsed-Field Gel Electrophoresis 13
- Nanopore Sequencing 14
- Genome Assembly 15
- Conclusion 16

Ethical Considerations & Societal Impact 17

Future Research 18

References 19

List of abbreviations

QUAST – Quality Assessment Tool

WGS - Whole Genome Sequencing

ONT - Oxford Nanopore Technologies

NEB - New England Biolabs

BUSCO - Benchmarking Universal Single-Copy Orthologs

HMW DNA – High Molecular Weight DNA

PFGE – Pulsed Field Gel Electrophoresis

Kb - Kilobase

Mb - Megabase

Gb – Gigabase

Introduction

Environmental Challenges

Numerous different environmental challenges have appeared due to the actions of humankind, one of the major environmental challenges for society is pollution. Pollution can be separated into different categories dependent on the environment in which they are present such as land, air or aquatic pollution (Chen et al., 2022; Goodman et al., 2025). Many different species are threatened by this increase in pollution and other environmental changes, often due to actions related to humans. This is not isolated to any singular type of environment or location, instead it is a global issue that affects aquatic animals, terrestrial animals and amphibians alike. (Di Marco et al., 2018; Hirt et al., 2021).

Freshwater Mussels

Freshwater mussels, also known under the term freshwater bivalves, are present worldwide and are in most of their habitats around the world facing the possibility of extinction due to both environmental factors and other factors such as habitat fragmentation (Bolotov et al., 2019; Saavedra et al., 2022; Sousa et al., 2023). A common freshwater mussel species in Sweden is the *Anodonta anatina* more commonly known under the term duck mussel. The *A. anatina* is a part of the larger family of *Unionidae* mussels. Unionidae mussels have a highly varied lifespan depending on different factors; however, it is mainly dependent on how clean the water is in their habitats due to pollutants having the potential to adversely affect them (Jonsson et al., 2013).

The lifespans for species of freshwater mussels range from as short as ten years to as long as 100 years. These lifespan estimates are based on freshwater mussels that are living in a clean unpolluted environment (Jonsson et al., 2013). Since the *A. anatina* are filter feeders, the importance of clean water and environment are of the utmost importance. Filter feeder is a term used for animals or plants that consume nutrients through the intake of water filtered through an organ. This consumption method means that in addition to nutrients it also intakes other harmful substances present in the water which in turn negatively impacts the *A. anatina* (Jonsson et al., 2013). The existence and presence of freshwater mussels are of key importance in aquatic ecosystems when their important environmental functions such as water filtration are taken into consideration. In addition to water filtration, freshwater mussels have shown themselves to be a useful indicator for environmental changes in their habitats since the mussels generally stay immobile or move small distances meaning that the mussels generally stay in their local habitats for large periods of time which makes monitoring of local environmental changes in specific habitats or locations easier. These characteristics make freshwater mussels of interest in environmental research, more specifically with regard to research concerning mutagenic pollution since mussel's intake pollutants present in the water when feeding. (Ferreira-Rodríguez et al., 2023; Zieritz et al., 2022).

Environmental Research

Research using freshwater mussel species as environmental research species is nothing new, however, research using *A. anatina* as an indicator species is still in its infancy. However, studies do exist, which indicates the possibility of further usage of *A. anatina* as an indicator species (Lance et al., 2025). There has in addition been at least two studies conducted using *A. anatina* and associated biomarkers (Ekelund Ugge et al., 2020, 2021). Neither study could be considered completely successful, however promise was shown and further studies with regards to suitable biomarkers were recommended (Ekelund Ugge et al., 2020, 2021).

Freshwater mussels are highly important to numerous and diverse aquatic ecosystems worldwide and display potential usage in environmental research, despite this there is a concerning lack of knowledge with regard to the genome of a large majority of freshwater mussel species (Smith, 2021). The cost and complexity of performing whole genome sequencing and genome assembly have historically been a significant barrier of entry when applied to complex organisms such as eukaryotes. Improvements in sequencing technology and assemblers in conjunction with the decline in associated costs has made this area of research more accessible (Bagger et al., 2024).

Genome Knowledge

Knowledge of the genome is important to be able to evaluate the health of individuals of the associated species and changes in their living environment such as pollution. With an established reference genome, mutations in the DNA of individual *A. anatina* could be discovered and potentially indicate the presence of mutagenic pollution (Rosendahl Huber et al., 2021). A complete reference genome would additionally make it easier to detect structural variations or more numerous mutations that are present throughout the genome. A reference genome gives a more coherent view of the organism and the information that can gleaned from this can help with development of different diagnostic tests which could help with ongoing and future research (Ekelund Ugge et al., 2021). Mutations are not inherently negative and can be due to numerous different factors and be both spontaneous or as with the case of mutagenic pollution, occur due to external factors and therefore further investigation into suspected mutations and the reason behind the mutations would be necessary before drawing conclusions. Another factor that should be taken into consideration is that is possible that differences can exist in the genome of different populations of the same species due to their specific habitats or inheritable mutations that have occurred in that specific population (Gallardo-Escárate et al., 2023; Harney et al., 2022; Lemiere et al., 2005).

Research with regards to the genome of a minority of freshwater mussel species do exist and has been performed successfully. Examples of this are two North American species *Venustaconcha ellipsiformis* (Ellipse Mussel) and *Potamilus streckersoni* (Brazos Heelsplitter) which hade genomes assembled in 2018 and 2021 respectively (Renaut et al., 2018; Smith, 2021). Successful genome assemblies of freshwater mussel species are sparse and the genome of the vast majority, remains unknown. Existing estimates of the number of freshwater species are at approximately a thousand different species and only six species have whole genome assemblies produced at the time of writing (Gomes-Dos-Santos et al., 2024). *A. anatina* is unfortunately not one of the exceptions.

Genome Size

A rough estimation of the genome size of *A. anatina* was desired for purpose of genome assembly evaluation. At the time of writing, a published complete genome assembly did not exist, instead an estimation based on published freshwater mussel genome assemblies was used. The sizes of genome assemblies of five different freshwater mussel species ranged from 1.5 Gb - 2.9 Gb and the genome of the *A. anatina* was assumed to be around this range (Bai et al., 2024; Gomes-Dos-Santos et al., 2023a, 2023b, 2024; Renaut et al., 2018).

Increasing the knowledge of different freshwater mussel species through sequencing and genome assembly has the potential be used as a foundation for continued studies both on the freshwater mussels and for usage in environmental studies as an indicator species. It should be noted however that not all pollutants or environmental changes can be discovered through sequencing since not all pollutants or environmental changes lead to structural change in the DNA. The aforementioned factors could instead lead to epigenetic changes in the freshwater mussels which would not be observable through the commonly used DNA sequencing techniques.

Genome Assembly Experimental Outline

A generalized outline consisting of the main parts of this *A. anatina* genome assembly experiment broken down into five main parts is presented in Figure 1.

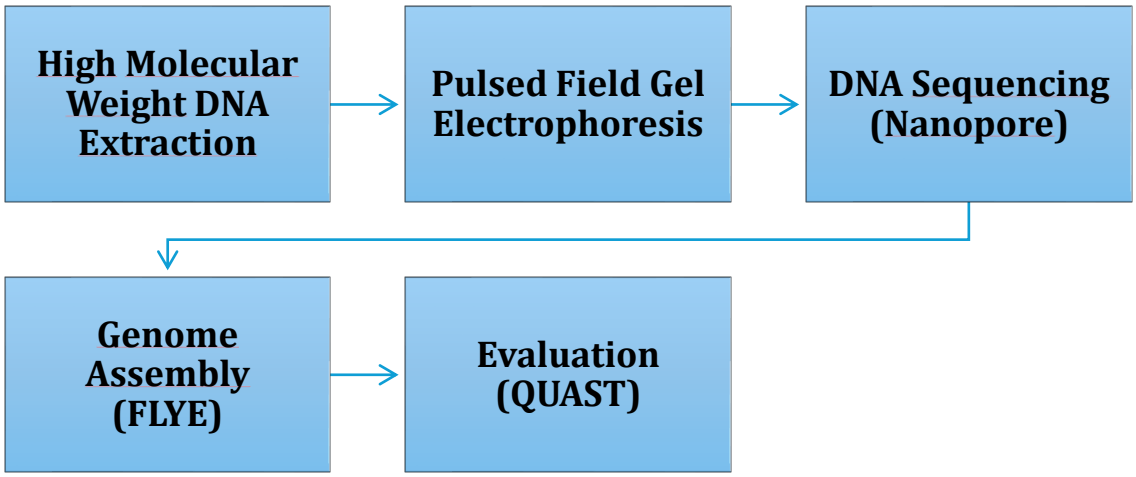

Figure 1. Breakdown of sequencing and genome assembly experiment performed on *Anodonta anatina* into five main steps.

It is important however, to note that the outline of the necessary steps for a successful genome assembly differs between different researchers. This is due to the fact that depending on how and how thoroughly the different processes are divided or broken down the number of steps will change. This means that the exact number of steps can be considered somewhat arbitrary. (Jung et al., 2020; Raghavan et al., 2022). Figure 1 is a breakdown specifically created for this *A. anodonta* experiment

High Molecular Weight DNA Extraction

The starting point of any research performed on a specific species is to collect individuals of the specified species. It is of the highest importance that the specimens collected, belong to the desired species. This verification can be performed in different ways. A commonly used method is DNA barcoding (Antil et al., 2023). When it comes to experiments where DNA extraction is necessary, it is of key importance that pure DNA can be extracted to increase the success rate of downstream applications. Numerous different techniques, protocols, kit exist with their own specializations making the choice of what to use of major importance. The success or failure of DNA extractions is generally evaluated through purity measurements and validation of the size of extracted DNA fragments through methods such as pulsed field gel electrophoresis (PFGE) to determine if they are of suitable size for downstream applications. What is a suitable DNA fragment size varies depending on the end goal of the experiment. In a typical whole genome sequencing assembly, high molecular weight DNA (HMW DNA) is typically desired. (Maghini et al., 2021; Ye & Lei, 2023).

Whole Genome Sequencing

Whole genome sequencing (WGS) is the process of sequencing the entirety of the DNA of an organism. With the most famous case of sequencing the whole genome of an organism being the human genome project, which was a massive undertaking (Lander et al., 2001). However, with the decline in cost and the increased ease of use due to technological advancements in recent years, the use of whole genome sequencing has also started to be used for different clinical purposes and smaller research projects. (Bagger et al., 2024). Sequencing techniques are divided into three different generations. Sanger sequencing is considered to be of the first generation and

was used in the sequencing of the human genome and now sequencing techniques have reached the third generation. In contrast to many other technologies, newer generations are not inherently superior to the older generations. The generations have different advantages and drawbacks, meaning that depending on the aim of the experiment the suitable generation technique differs (Heather & Chain, 2016).

First generation sequencing is characterized by high costs, complexity and low throughput; however the output is considered to be of high quality. Sanger sequencing is the most famous first-generation technique and is still commonly used due to high accuracy when performed on short sequences (Liu et al., 2012; Pareek et al., 2011). Second-generation techniques improved on the throughput output and lowered costs, however read lengths are short and the output needs additional bioinformatic processing when compared to Sanger sequencing, one of the most famous second-generation techniques is Illumina sequencing (Pareek et al., 2011). Third-generation sequencing differ from earlier generations significantly. While the first and second-generation underlying techniques can be considered quite similar, third-generation techniques function completely different and work on the level of singular nucleotides. Third generation techniques (TGS) have a significant increase in read length, substantial decrease in the time to perform sequencing and high throughput while demanding significant bioinformatic processing. Two famous third generation techniques at the forefront of research today are Nanopore sequencing and PacBio sequencing (Lu et al., 2016).

Nanopore Sequencing (TGS)

The concept of Nanopore sequencing is relatively old, however it is in recent years that the usage of Nanopore sequencing has started on a grander scale due to improvements both in the technique itself and in the surrounding field of omics (Head et al., 2024). Nanopore sequencing can be used both for simpler experiments using short DNA fragments and for experiments that need long and ultra-long read capability (Wang et al., 2021). Nanopore sequencing functions with the help of numerous Nanopores located in what is denoted as flow cells. Single stranded DNA are then identified based on electrical currents that are generated when DNA is propelled through the Nanopores with the help of motor proteins attached during the preparation process (Scarano et al., 2024). The nucleotide identification process occurs in real time and simultaneously, the nucleotide identification process is usually denoted as basecalling. Basecalling is the process of an algorithm converting the electrical currents into a string of nucleotide bases which in turn allow for identification on the nucleotide level. This information is then stored and transferred to fastq or BAM files and ready for further analysis. Different algorithms can be used for basecallling; however, Dorado is integrated into Nanopore sequencing systems and is recommended (B. Lin et al., 2021)

Nanopore sequencing demand specific preparation to function. Oxford Nanopore Technologies offers several protocols and kits to tailor to different experiments and therefore determining which kit and protocol to use are highly important. Additionally, the choice of flow cells and associated systems are also of significant importance. TGS technologies such as Nanopore sequencing have reached the point that average read lengths of >10 kb are now possible (Scarano et al., 2024). This and the continued improvement with regards to accuracy, and the streamlined process that is associated with Nanopore sequencing, all contributed to the decision to use Nanopore sequencing for this experiment.

Genome Assembly & Evaluation

The aforementioned technique of Nanopore sequencing can be combined with sequence assembly when studies are performed related to whole genome sequencing. Genome assembly is a process of combining different small or large fragments to obtain larger representative fragments or even

a complete representation. These larger representative fragment combinations are commonly known as contigs or scaffolds. Sequence or genome assembly can be performed with short or long fragments and can go from combining fragments to reach sizes in the low hundreds bp or be performed to create a representation of a full genome (Jiao & Schneeberger, 2017; Simpson & Pop, 2015).

There exist different software's and algorithms for sequence assembly with inherent down and upsides, therefore depending on the kind of experiment and the size of the assembly, the optimal software will vary. With regard to Nanopore sequencing reads, Flye has in recent years become the preferred method for genome assembly according to Oxford Nanopore Technologies (ONT), the company behind the Nanopore sequencing technology and this is backed up by third party researcher's genome assembly comparisons (Chen et al., 2024; Khrenova et al., 2022). There are, however, often still difficulties with choosing which software are optimal and often multiple different software are used and the results compared to each other. Evaluation and comparison of results from different assemblers are unfortunately not always easy either, however, there exists tools such as QUAST (Quality Assessment Tool) which is used in this experiment, that can help with performing comparisons and evaluate genome assemblies if desired. A basic explanation of some of the important metrics of QUAST will be provided. N50 are the cutoff length of where 50% of the genome assembly size are covered by contigs larger than the presented value. For N90, this is the same with the exception of the fact that the cutoff is instead at 90%. L50 is related to N50 and instead of looking at sizes, it denotes the number of contigs that comprise 50% of the assembly (Gurevich et al., 2013)

Research Aim

The aim of this research is to create a complete genome assembly of the species *A. anatina* through a combination of whole genome sequencing and genome assembly. The creation of a complete genome assembly of the *A. anatina* will be performed though the fulfilment of three objectives.

1. DNA extraction: The first objective is to extract High Molecular Weight DNA (HMW) from the *A. anatina*. PFGE is used to verify the desired lengths following DNA extraction (50-250 kb).
2. Nanopore sequencing: Sequencing of the extracted HMW DNA using, Nanopore sequencing in conjunction with MinION. Sequencing is followed by evaluation of quality score of the sequencing data using MinKNOW and PycoQC before proceeding to genome assembly.
3. Genome Assembly: Processing of the Nanopore sequencing data is followed by performing genome assembly of the sequencing data with Flye. Followed by evaluation of genome assembly quality through the usage of QUAST. Completeness of the genome assembly was then evaluated through comparison to the expected genome size of the *A. anatina*.

Materials and methods

Collection of *A. anatina*

The *A. anatina* used for this experiment was collected locally from the lake Vingsjön located in Axvall Sweden during the the time period March – May 2025, with the help of Annie Jonsson at Skövde University who have the necessary knowledge for collection and species identification of the *A. anatina*.

Following collection of the *A. anatina*, at multiple occasions during the previously stated time period, the specimens were stored in individual glass jars filled with tap water and stored in a fridge at a temperature of 4°C until dissection could be performed. Due to potential detrimental effects on the specimens if stored in the fridge for too long, a time limit of 48 hours was used. New specimens were only collected if necessary, however, the decision was taken to only use fresh tissue due to recommendations stated in the Monarch® HMW DNA Extraction Kit for Tissue (NEB) which was used for DNA extraction. The use of only fresh tissue necessitated the use of more specimens than if frozen tissue would have been used instead, however, to achieve the optimal results this was deemed to be necessary. Two specimens were used during the final experiment and a further eight were used during the optimization process.

Dissection

Dissection was carried out within 48 hours of collection to avoid potential health deterioration from the *A. anatina* being transferred out of their natural habitat. The collected *A. anatina* specimens were taken out of the fridge right before dissection and were rinsed gently with water to remove dirt and debris from the shell. Following the gentle rinse, the *A. anatina* was dissected through the use of pincers, scalpels and scissors. A scalpel with blade attached were inserted in the small opening for the "foot" of the mussel and thereafter the scalpel was moved alongside both the shell in both directions until the scalpel reached the anterior and posterior abductor mussels which were then excised using the scalpel after which the *A. anatina* was opened up. Following the opening of the *A. anatina*, a tweezer was used to hold the foot steady while the foot was simultaneously excised using scissors and then immediately transferred into a 1.5 ml Eppendorf tube and kept on ice.

High Molecular Weight DNA extraction

Following dissection, the tissue was in accordance with instructions present in the Monarch® HMW DNA Extraction Kit for Tissue (New England Biolabs) separated into smaller pieces and weighed to create 25 mg of tissue which was then further separated through the use of scalpels into smaller pieces and then transferred into new 1.5 ml Eppendorf tubes and put on ice in preparation of homogenization.

Homogenization in accordance with instructions present in Monarch® HMW DNA Extraction Kit for Tissue (NEB) for pestle homogenizations for fresh tissue were followed. For the first part of the protocol denoted as "Tissue lysis" modifications to the protocol were made. Lysis was performed following the pestle homogenization, and the samples were run on a thermal mixer for 15 minutes at 1400 rpm and then kept stationary on the same thermal mixer at a temperature of 56°C for 45 minutes until fully dissolved as instructed for muscular tissue. For the rest of the lysis process, the standard input for the protocol was followed, and no further changes were made.

For the second part of the protocol denoted as "HMW gDNA BINDING AND ELUTION", no changes were made, and standard input instructions were followed. Following the completion of the elution of the extractions, several different methods were available to use to increase the trustworthiness of concentrations and purity measurements. The different methods were tested during the optimization phase and based on the results of the tests; one method was chosen to be

used throughout the experiment. The method chosen was to store the samples for more than 24 hours at 4°C after elution in 100 µl of elution buffer. Following elution, a step consisting of pipette mixing of the elution was added to ensure dispersion of the DNA throughout the elution buffer. After 24 hours of storage the samples were measured using Nanodrop 2000 (ThermoFischer) for concentration and purity values.

Purity of the samples were measured through the use of a spectrophotometer that detects the absorbance of samples at specific wavelengths to detect contaminants and produce ratios depicting this, two commonly used ratios for sample purity are A260/A280 and A260/A230 with the desirable ratios being 1.8-2.0 for A260/A280 and 2.0-2.2 for A260/A230.

Pulsed-Field Gel Electrophoresis Sample Preparation

Two samples from two different individual mussels, A and B were chosen to proceed with following completion of the extraction based on purity and concentration values. Each individual sample were split into three separate 1.5 ml Eppendorf tubes, each containing 30 µl from the original extraction. The first series of samples are denoted as A1-A3 and the second B1-B3 to clarify which sample they originate from. The first tube from each sample series were left unmodified. For the other four tubes an individual 4mm stainless steel bead was added to each tube and then the samples were pulse vortexed at maximum speed starting at 2 minutes for sample A2 with an increase of 30 seconds for each subsequent sample, ending with 3.5 minutes for the last sample B3 meaning each sample spent different amount of time being vortexed.

Pulsed-Field Gel Electrophoresis (PFGE)

After sample preparation was completed, PFGE was performed to determine DNA fragment size. DNA fragment size was important for the project due to the desire to perform sequencing and assembly on HMW DNA and this step was added for verification of this. The settings and gel concentration used for the PFGE was chosen to reflect instructions present in the "CHEF-DR III Pulsed Field Electrophoresis Systems" (BioRad) for the 48.5 kb to 1000 kb which was the size marker Lambda ladder (BioRad) used for purposes of size determination of the extracted DNA fragments. The specific ladder was chosen due to the expected DNA fragment sizes of 50-250 kb. With regards to sample loading, the total amount loaded for each sample was 20 µl using a wide bore pipette. The 20 µl was taken from a mixture containing 6 µl of sample, 4 µl of 6X purple dye (NEB) and 14 µl of milli-Q ultra-pure water (Millipore).

The specific settings used for the CHEF-DR III was a ramping switch time of 50–90 seconds, The field angle was set to 120°, voltage was set to 5 V/cm, and the run time was set to 22 hours. The cooling module was used to keep the temperature steady at 14°C throughout the run. A gel concentration of 1.0% was used and 150 ml gel was created. The PFGE was imaged using ChemiDoc MP Imaging System (BioRad). The chosen buffer for the gel and circulation throughout the run was 0.5% TBE-Buffer. The alternative TAE-buffer was tested; however, results were unusable and therefore TBE-Buffer had to be used despite its toxicity.

Nanopore sequencing (Library preparation and sequencing)

Following completion of size verification, library preparation and sequencing was performed in accordance with the protocol Ligation sequencing DNA V14 (SQK-LSK114, ONT) for the sample B3. B3 was the sample chosen to proceed with due to the smaller sized fragments in comparison to the other samples. This was based on the results of the PFGE, which was deemed to be desirable due to earlier test runs of the experiment and communication with Oxford Nanopore Technologies Support.

There were some key modifications to the protocol for sample B3, all reagents and DNA amounts were increased to 1.5x, this change of the protocol was chosen due to poor DNA recovery in test

runs of the experiment. In addition, the end elution step was performed twice, this was done to create two separate elution's that could be loaded onto the MinION flow cell separately. This was done to increase the potential sequencing output. For the library preparation, long fragment buffer was used due to DNA sizes being >3 kb based on earlier size verification. Measurements of DNA concentrations were performed both during and at the end of the library preparation and were performed using the Qubit™ 1X dsDNA High Sensitivity (HS) kit in conjunction with Qubit Fluorometer 4 (Thermo Fisher)

After completion of the library preparation, the prepared libraries were loaded onto individual MinION Flowcells (R10.4.1) (ONT) following Ligation sequencing DNA V14 (SQK-LSK114) (ONT). For sequencing, the MinKNOW software (24.11.10) was used. The total length of sequencing was 46 hours, split in two parts. After 24 hours, the sequencing was stopped, the flow cell was washed following the Flow Cell Wash Kit (EXP-WSH004 or EXP-WSH004-XL) (ONT) and reloaded using the second elution from the sample and left to run for an additional 22 hours before termination. Both elutions of sample B3 were basecalled in conjunction with sequencing, with the settings set to high accuracy basecalling which sets the cutoff for the Q-value at 9.

Sequencing data processing

The sequencing data was investigated for both individual runs to evaluate the quality of the sequencing data. MinKNOW presents data such as Q-values, read length, passed/failed reads and more. However, PycoQC (2.5.2) was used, in addition, to evaluate the sequencing data quality and was used with default settings with the exception of increasing the Q-value threshold to 9 to match the cutoff used when basecalling was performed and the results from MinKNOW and PycoQC were compared, MinKNOW is a part of the Nanopore sequencing system and PycoQC was added for purposes of comparison to other published results not using Nanopore sequencing (Leger & Leonardi, 2019).

Genome Assembly

The data from the two individual runs contained in fastq.gz files were merged into one combined file to create the largest possible data for use in genome assembly. Genome assembly was performed using the genome assembly program Flye (2.9.6). One change was made from the default settings, and this was "Mode" being changed to either Nanopore Raw, Nanopore High Quality or Nanopore Corrected for three different analyses (Lin et al., 2016).

Following completion of Flye, QUAST Genome Assembly Quality (QUAST) (5.3.0) was run on the different Flye assemblies to get an easier overview of data with metrics such as number of contigs and total length of the assembly. Options changed from the default options for QUAST were type of organism to Eukaryote and activating the setting for large genome (>100 Mbp). Additionally, the type of reads was set to Nanopore since Nanopore sequencing is what was used for sequencing of the DNA. The bioinformatics processing with regards to the programs PycoQC, Flye and QUAST were performed using the Galaxy EU server (Abueg et al., 2024).

Results

DNA extraction

The extracted DNA samples were measured for concentration and sample purity with nanodrop 2000 (ThermoFischer). In Table 1. purity and concentrations values are presented for these samples, which were from two different individuals denoted as A and B.

Table 1. Samples from tissue excised from the "foot" of the *A. anatina* from two different individuals denoted as A and B. Concentrations and purity measurements of the DNA samples was measured with nanodrop 2000 (ThermoFischer) after High Molecular Weight DNA extraction had been performed.

Sample	DNA Conc (ng/ul)	A260/A230	A260/A280
A	105	2.15	1.85
B	116	2.05	1.85

Pulsed-Field Gel Electrophoresis

Due to issues occurring when following the original experimental setup an extra step was added before proceeding to PFGE for size verification. An additional shearing step of the DNA samples were added. This step was performed with 4mm stainless steel beads in conjunction with a vortex after A and B had been separated into three new equal parts, A1-A3 and B1-B3. The new samples were sheared for different amounts of time with the effects of the shearing process presented in Table 2. A1 and B1 were left untouched and used as baselines for the PFGE.

Table 2. Concentrations and purity measurements using nanodrop 2000 (ThermoFischer), following the separation of sample A1 into A2 and A3. Similarly sample B1 were split into B2, B3 and the amount of time each sample were vortexed.

Sample	Time (Seconds)	DNA conc (ng/ul)	A260/A230	A260/A280
A1	0	105	2.15	1.85
A2	120	190	1.29	1.88
A3	150	121	1.14	1.87
B1	0	116	2.05	1.85
B2	180	197	1.20	1.83
B3	210	107	1.20	1.87

Following completed PFGE run, the fragment sizes of the different samples were evaluated and compared to the baseline, meaning: A2, A3 to A1 and B2, B3 to B1 to allow for selection of a suitable sample (50-100 kb) to proceed with to Nanopore sequencing. The results of the PFGE are presented in Figure 2.

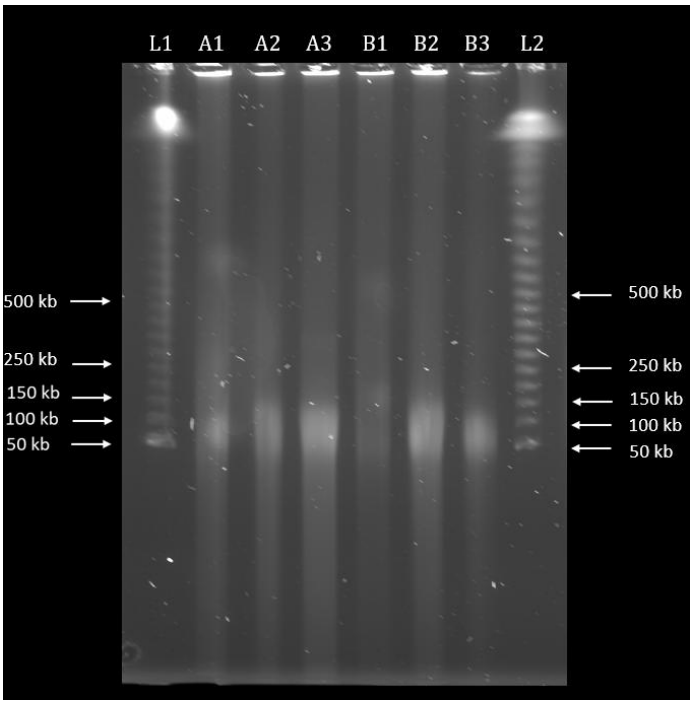

Figure 2. Pulsed Field Gel Electrophoresis imaged using ChemiDoc MP Imaging System (BioRad) following 22-hour PFGE run of samples extracted from *A. anatina* denoted as A1-A3 and B1-B3, L1, L2, denotes loaded CHEF DNA size standard - Lambda Ladder (BioRad) and covers 48.5 kb to 1000 kb.

The picture taken following the PFGE presented as Figure 2 run shows differences between A1 and its associated samples that it served as baseline for, the samples A2 and A3. The results of the PFGE show that the majority of the DNA sizes of the samples A2 and A3 are in the range of 50 - 150 kb. A clear difference in DNA size between baseline sample B1 and its associated samples B2 and B3 can be observed as well. Sample B2 resembles the same 50-150 kb size range as samples A2 and A3. For sample B3, size instead ranges mainly between 50-100 kb with some fainter fragments above and below.

Nanopore Sequencing

Following PFGE, the sample denoted as B3 was chosen to proceed with sequencing due to the desire of smaller fragments based on earlier sequencing runs during optimization of the experiment that were unsuccessful with larger sizes. Measurements of DNA concentration were performed with Qubit™ dsDNA High Sensitivity (HS) kit in conjunction with Qubit Fluorometer 4 (Thermo Fisher) to determine the absolute DNA mass present during the preparation and loading process of sample B3. These measurements could be used to estimate how well the DNA recovery present in library preparation functioned.

Starting amount of DNA for library preparation was chosen to be 1500 ng from the sample B3. Following the step of "DNA repair and end-prep" in the Ligation sequencing DNA V14 kit (SQK-LSK114, ONT), 810 ng of DNA was measured to be present in the sample. Measurement was additionally performed following completion of the library preparation. The total amount of DNA mass available for loading of the MinION flow cell were 237.2 ng of DNA from two subsequent elution's of the sample B3. 128.4 ng from the first elution and 108.8 ng from the second elution. Due to the load limit of 15 ul for one run of the MinION, two separate runs, one for each elution had to be performed.

Following completion of library preparation, flow cell check was performed, and the check was successful. Following completed flow cell and hardware check loading of the first prepared library was performed. The maximum amount of DNA that could be loaded from the first elution consisted of 128.4 ng which is below the 300 ng that is desired for loading and therefore all available elution was loaded. The available pores in the flow cell at the start of the sequencing run were 1180 pores and at the time of manual termination after 24 hours, available pores had decreased to 312.

The first sequencing run used the setting for high accuracy for basecalling, which set the cutoff for passed reads at the q-value of 9. The first sequencing run generated 4.19 M reads and basecalled 4.9 Gb. Of the 4.19 M reads, 3.7 M of them passed the quality score of 9 while 0.5M did not pass the quality score cutoff. Of the 4.9 Gb basecalled bases, 4.56 Gb passed, and 0.351 Gb did not pass the quality score cutoff. The ratio of pass and fail of the basecalled bases is over 90%, while with regards to the numbers of reads, the pass rate falls below 90% down to 88%.

Following the manual termination, cleaning of the flow cell was performed and the pores available for sequencing was increased from 312 to 452, indicating that the flow cell wash was successful. The second elution containing 108.9 ng, which is less than the desired amount of 300 ng and therefore all was loaded, the settings from the first sequencing run was kept. The second sequencing run was left to run for 22 hours before it was terminated. The second sequencing generated lower reads and basecalled bases than the first run. The amount of reads in the second run ended at 1.84M with 1.57M of the reads passing the quality score and 0.27M of the reads failed the Q-score. The basecalled bases ended at 2.49 Gb with 2.29 Gb bases passing the quality score and 0.2 Gb failing. The pass rate for the reads is barely lower than for the first sequencing at 85% while the passing rate for basecalling was relatively consistent at over 90%.

The sequencing data from the two runs were further analysed separately using PycoQC (2.5.2) to visualize potential differences in quality and read length between the two runs. Relevant statistics from the PycoQC are median read lengths, N50 values and quality scores for the two different runs. For the first run the median read length of all reads is 464 bp while the N50 is 2860 bp. For the second run the median read length of all reads is 481 bp while the N50 is higher at 4130 bp. With regard to quality scores these are graphically represented in In Figure 3. The density of the reads is presented on the Y-axis and the quality scores of the reads on the X-axis giving a quick overview of the quality of the sequencing runs. Figure 3a is for the first run while Figure 3b is for the second run. Median values for both runs are close to each other with the median varying between 15-16.

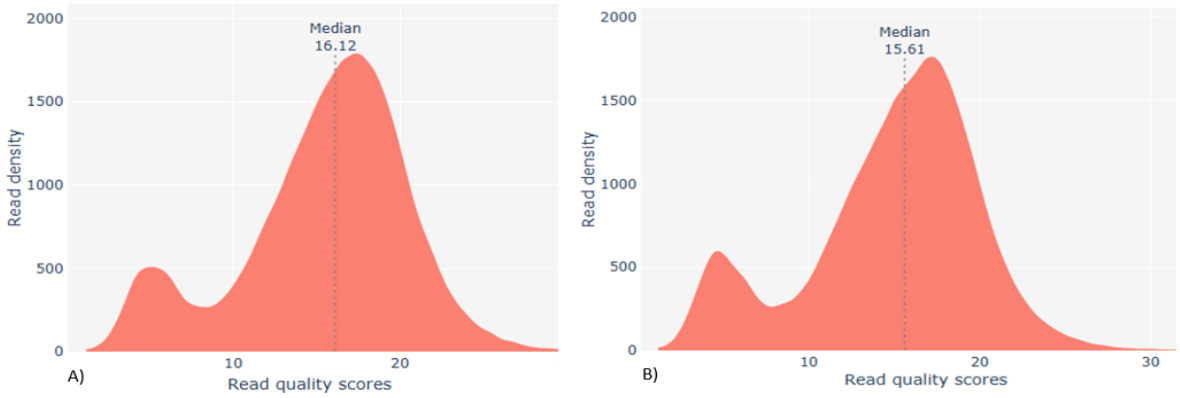

Figure 3. A graphical representation of the quality of the sequencing data created using PycoQC (2.5.2) with read density on the Y-axis (The amount of reads with specific quality scores) and quality scores of reads on the X-axis giving a quick overview of the quality of the two sequencing runs. A) is for the first run while B) is for the second run. First run has a median value of 16.1 while the second run has a median value of 15.6.

Genome Assembly

The data from the Nanopore sequencing stored in POD5 files were merged for the individual runs and then subsequently merged for both runs to create the largest possible dataset available for genome assembly to achieve the best possible result. The combined amount of basecalled bases following the mergers ended at 6.85 Gb.

The genome assembly was performed using Flye for the merged data for the two runs and was performed thrice with one change which was that the first run used the setting for Nanopore raw the second one used Nanopore High Quality and lastly the third run used Nanopore Corrected instead. QUAST software was used on all three assemblies, some important metrics for the QUAST output are N50, N90, L50 and total length. The key parameters for the different Flye runs are shown in Table 3 to give an easier overview of the differences between the Flye assemblies.

Table 3. Important parameters from QUAST analysis of three different genome assemblies: Nanopore Raw, Nanopore High Quality and Nanopore Corrected. All were performed using Flye and QUAST settings were kept identical for all assemblies.

Read settings	**Nanopore (Raw)**	**Nanopore (Hq)**	**Nanopore (Corrected)**
Contigs (>0 bp)	55 537	44 045	45 450
Contigs (>1000 bp)	48 422	40 090	40 296
Total Length (>0 bp)	393 149 kb	438 532 kb	446 258 kb
Total Length (>1000 bp)	387 736 kb	435 816 kb	443 377 kb
Largest Contig	179 kb	176 kb	180 kb
N50	12 749 bp	18 544 bp	19 708 bp
N90	3201 bp	4693 bp	4695 bp
L50	8582	6676	6492

Discussion

The process of performing a genome assembly has historically been considered expensive, time consuming and difficult. Rapid advancements in recent years have made it cheaper and simpler than before. (Cosma et al., 2023). Genome assembly is however not trivial. Especially when performing *de novo* genome assembly. *De novo* genome assembly is what has been performed in this study and refers to assembly performed without a reference genome. (Michael et al., 2018; Sohn & Nam, 2018).

Preparation

Preparation and planning of the entire the experiment was of the utmost importance. Two key parameters were decided for this experiment based on the available resources and the aim of performing *de novo* genome assembly. Nanopore sequencing with MinION would be used and HMW DNA was desired. HMW DNA was desired since it tends to produce more coherent assemblies. (Maghini et al., 2021; Petersen et al., 2022). HMW DNA is however more fragile due to its length and therefore demands specialized techniques, kits and equipment. There are difficulties with regards to measurements of concentration and purity of HMW DNA and optimization of commercially available kits are generally needed for success, regardless, using HMW DNA has proven successful in several freshwater mussel assemblies in recent years. (Gomes-Dos-Santos et al., 2024; Volarić et al., 2021)

Species Identification & Collection

Help with the collection and identification of the *A. anatina* were available and morphological identification, specifically visual inspection was used. Ideally both morphological and molecular identification would have been performed; however, it was determined to not be possible with the allocated time and resources. The downside of visual inspection is the possibility of misidentification due to it being a subjective evaluation instead of objective (Velasquez-Restrepo et al., 2024; Yang et al., 2022). However, visual identification has been used successfully in published genome assemblies of freshwater mussels (Bai et al., 2024). An objective method that could be used in future experiments would be DNA barcoding which compares sequenced DNA from an organism to an existing database to allow for species identification. (Hauck et al., 2023).

Extraction & Pulsed-Field Gel Electrophoresis

Extraction is of one of the most important processes for downstream analysis, a low-quality DNA sample have massive influence on every step following which compounds on itself when multiple steps are necessary before analysis (Schalamun et al., 2019). What is considered good concentrations of DNA is dependent on the type of tissue and what is needed for the downstream applications and therefore cannot be generalized.

With regards to purity, there are generalized accepted guidelines for what is considered desirable. Desirable purity values of A260/A230 and A260/A280 are 2.0-2.2 and 1.8-2.0 respectively (Cruz et al., 2025). These ratios functions on detection of contaminants present in the wavelengths of 230 nm and 280 nm respectively. These guidelines are therefore used when evaluating the results of this experiment. The ratios of 2.05 and 2.15 for A260/A230 and 1.85 with regards to A260/280 presented in Table 1. fall within the desired range of both ratios (Cruz et al., 2025; Schalamun et al., 2019). Therefore, based on purity values, both extractions can be considered successful. When comparing the purity ratios to a published study using the same extraction kit, similar values were observed which additionally indicates a successful extraction (Volarić et al., 2021). The amount of DNA following elution is presented in Table 1. and is substantially more than what is needed to perform both PFGE and Nanopore sequencing. Therefore, the extraction could be considered successful with regards to both purity and concentration.

Size verification was performed using PFGE. The PFGE run additionally worked as a verification of the added shearing step which was inspired by the research article "A Simple Approach for Effective Shearing and Reliable Concentration Measurement of Ultra-High-Molecular-Weight DNA" with some modifications (Koetsier & Cantor, 2021). The shearing step was added due to undesirable sequencing results in the test run of the experiment despite good concentrations and purity values. This decision came after communication with Oxford Nanopore Technologies support, and it was decided to attempt the same process with smaller sized fragments necessitating the addition of a shearing step. (Personal Communication, Oxford Nanopore Technologies Support, April 2025). It should however be noted that this decrease in DNA fragment size contradicts the stated guideline for Monarch® HMW DNA Extraction Kit for Tissue (NEB) used in this experiment, where the stated desired size for downstream applications was 50-250 kb and the fragment size used in this experiment instead was around 50-100 kb. The downside of the smaller DNA fragments is that the point of using HMW DNA was its improved ability to generate more coherent genome assemblies, which could potentially be compromised by reduction in size (Petersen et al., 2022). There is however no set definition of what constitutes HMW DNA and depending on the source, 50-100 kb could be considered HMW DNA.

The sample B3 was chosen due it best fulfilling the desired objective of the added shearing step of lowering the DNA fragment size and was estimated to be around 50-100 kb based on the PFGE. However, purity measurements performed after the shearing step showed a decrease in the purity values of the DNA samples presented in Table 1 and Table 2. The A260/A280 values were relatively consistent, and the small differences could be considered natural variations between different measurements. For A260/A230 values however, a drop to around 1.2 was observed which could have significant negative effects on the Nanopore sequencing. A lowered A260/A230 ratio indicates the presence of contaminants that absorbs at 230 nm, shearing of DNA using beads is nothing that inherently should lower the A260/A230 values, however the measurements for all four samples have this decrease. The beads used in this experiment was from an opened container and it is possible that contaminants were present on the beads or otherwise introduced during the shearing (Arseneau et al., 2017; Schalamun et al., 2019). Studies using HMW DNA extraction with shearing steps as a part of the extraction and reaching the desired purity values does exist indicating that shearing itself is not likely to be the reason (Volarić et al., 2021). It was decided to proceed with the experiment despite the lowered purity values. Optimally, additional purifications steps would have been performed. However, due to the limited amount of DNA available for downstream applications following the sample split this was decided against due to the risk of potential DNA loss during the purification process.

Nanopore Sequencing

During the library preparation for the Nanopore sequencing, large amounts of DNA was lost leading to substantially less DNA being available for loading on the MinION flow cell than the 300-ng desired. A possible reason for this is the lower than desired A260/A230 values, it is possible that the library preparation did not function as intended due to contaminants (Schalamun et al., 2019). Another possibility is wrongly estimated DNA size, according to the kit manufacturer, an inaccurate estimation of fragment sizes has the potential to lead to lower recovery (Oxford Nanopore Technologies, 2025). With no comparative runs performed, where the 300-ng requirement was fulfilled, it is not possible to state with certainty the effect, this lower amount of DNA had. It could, however, be inferred that lower amount of DNA would lead to less sequencing output. With the earlier stated concerns about the DNA sample quality following the shearing process, an unfortunate expectation was that the sequence output would not be optimal (Arseneau et al., 2017; Schalamun et al., 2019).

More stringent demands of sequencing are used in some recent published works performed with Nanopore sequencing with cutoffs of Q-score far surpassing the 9 used in this experiment with values >20 being used instead, however for a smaller experiment performed by an individual, those cutoffs are not realistic with the available time and resources (de Souza et al., 2023; Zubach et al., 2025). The combined output of the sequencing runs ended at 6.85 Gb basecalled bases, which is far below the desired minimum output for the genome assembly and far lower than the claimed potential output from Oxford Nanopore of 50 Gb from one singular MinION flowcell.

Genome Assembly

The fact that sequencing output fell below the desired output for genome assembly was based on the estimated genome size of the *A. anatina*. Genome size was as stated in the introduction based on recent successful *de novo* genome assemblies of other freshwater mussel species. The genome size estimation range of 1.5Gb to 2.9Gb was then used in conjunction with the sequencing data output for purposes of evaluation. Flye demands a minimum coverage of 30x of the genome size for good quality assemblies meaning that at minimum the necessary sequencing data output would be 45Gb. (Y. Lin et al., 2016). It should however be noted that these different freshwater mussels differ in the degree of taxonomical relationship to the *A. anatina* and is used only as a rough estimation.

When evaluating the amount of sequencing data produced using the lower end of the estimated size range (1.5 Gb) the coverage is approximately 4.5x, which is far lower than desired. Despite this, three assemblies were successfully generated with varying read settings input which were then analysed using QUAST of which the results can be observed in Table 3.

There are differences in the results of all three of the different assemblies in the parameters of assembly length, contigs, largest contig, N50 and L50. The results of the Nanopore high quality and Nanopore corrected could be considered relatively close, while the results of the Nanopore raw assembly differ throughout. When evaluating the genome assemblies against each other using the QUAST result, a smaller number of contigs is desired since this indicates a more coherent assembly. When it comes to the parameter of total length, the largest value is what is desired which is similarly the case with the parameters N50 and N90. The parameter of L50 in similarity with the number of contigs, desires instead lower values. Based on how these different parameters compares in the different assemblies, the best one can then be chosen. (Mikheenko et al., 2023). The stated desired parameters for the different metrics in the output from QUAST in Table 3 give a clear indication that the worst assembly is Nanopore raw. Reaching a conclusion when it comes to determining the best out of Nanopore high quality and Nanopore corrected is much more difficult. This is because the values of these assemblies are much more similar. Another factor that significantly contributes to difficulty when it comes to comparisons of similar Flye assemblies is that Flye is by nature not deterministic, meaning that small variations occur despite the same input data (Kolmogorov et al., 2019). The sequence input data for this experiment is recommended to be used with Nanopore raw or high-quality since no correction of the sequencing data was performed. Therefore, the Nanopore high-quality assembly will be evaluated. With even the most generous assumptions about the estimated genome size of the *A. anatina* at 1.5 Gb, the assembly does not come close to full coverage and ends at below 33% coverage at a total assembly size of 0.4 Gb.

The program QUAST, was used to evaluate the genome assembly. The results were not indicative of a desirable result when compared to other freshwater mussel studies. The assemblies used for size estimation are generally of higher quality which shows in a substantial amount of fewer contigs, larger N50 values and the trend of detrimental differences are unfortunately present with regards to largest contig, L50 and N90 values as well (Bai et al., 2024; Gomes-Dos-Santos et al.,

2023a, 2023b, 2024; Renaut et al., 2018). QUAST was used to evaluate the assemblies in all of the aforementioned studies which makes comparisons to the genome assemblies produced in this study relevant. However, analysis of the genome assemblies was additionally performed using the tool BUSCO (Simão et al., 2015). BUSCO uses other metrics than QUAST and the two tools are therefore often used in conjunction (Bai et al., 2024; Gomes-Dos-Santos et al., 2023a, 2023b, 2024; Renaut et al., 2018).

For the referenced works of freshwater mussel genome assemblies, different sequencing technologies, assemblers and other bioinformatics processing software have been used (Bai et al., 2024; Gomes-Dos-Santos et al., 2023a, 2023b, 2024; Renaut et al., 2018). The main sequencing technology used in these articles are HiFi-sequencing which according to some studies display an increased accuracy when compared with Nanopore sequencing, however a trade-off is that it generally has a lower read length (Yu et al., 2024). With regards to assemblers, Flye was used in this experiment while Hifiasm was used in the other works. Comparisons between different assemblers are difficult, the sequencing technology used is often what determines which assembler is suitable and studies using both ONT reads and Hifi reads for comparison are sparse (Yu et al., 2024). Studies do even contradict each other with regards to what assemblers is suitable for a specific sequencing technology. For example, two different studies on Hifi-reads assembled with Flye and Hifiasm showed different results. In one of the studies Flye was evaluated to be superior on Hifi-reads while in the other Hifiasm was evaluated to be superior (Yu et al., 2024; Zhang et al., 2022). Additionally, depending on the processing of the sequencing data with regards to polishing, using reference genome and other processing options, studies have shown different results with regards to the optimal assembler (Burger et al., 2025). Therefore, it is unwise to state a superior sequencing technology or assembler, this is especially the case with assemblers since Flye, Hifiasm and others are still under active development which means that the most optimal or suitable might change. Two programs that were used in other genome assembly studies and for whom no suitable replacement was found were Jellyfish and Genomescope. These two programs work in conjunction to create an estimated genome size based on the sequencing reads of the experiment and functions on both Hifi-reads and Nanopore. This makes the programs useful when optimising the methods used in this experiment or for evaluation when using the methods present in the published freshwater mussel studies (Marçais & Kingsford, 2011; Ranallo-Benavidez et al., 2020).

Conclusion

To conclude, a genome assembly of the *A. anatina* was created and therefore the individual objectives of sample preparation, Nanopore sequencing and genome assembly can be considered fulfilled. The same can however not be said with regards to the aim of the experiment. The assembled genome covers less than 33% of the smallest (1.5Gb) of the recent published freshwater mussel assemblies and therefore the experiment did not yield a satisfactory assembly. In addition, the evaluation metrics when it comes to the quality and coherence of the genome assembly falters when compared to recent published works of freshwater mussel genome assemblies. The fulfilment of individual objectives and that an actual genome assembly of the *A. anatina* was produced, makes it possible for this experiment to serve as a base for improvement of the associated experimental techniques and help guide the direction of future experiments.

Ethical Considerations & Societal Impact

When performing research that requires the usage of living organisms, ethical considerations with regards to whether it is necessary to use these specific living organisms are needed. The desired data, meaning a published complete genome of the *A. anatina* did at the time of the experiment not exist. No other alternatives than using *A. anatina* to attempt to create a complete genome of the *A. anatina* was deemed possible.

The hope is that this and other similar studies can serve as a foundation for continued studies on the *A. anatina* itself with regards to its potential usage in environmental studies. However, despite it being necessary to use living organisms for this experiment, it is still important to treat the samples created from these organisms with care so that no more than the necessary number of organisms are used. The primary consideration when it comes to usage of the *A. anatina* is to use as few individuals as possible. No permission from Länsstyrelsen, Jordbruksverket or equivalent are needed to perform this experiment.

One of the major challenges that society face is pollution and how it effects climate change and biodiversity. Pollution has the ability to have massive detrimental effects on the whole of mankind and change the structure of society itself. When considering the added effect, it has on climate change and biodiversity, the challenge these connected factors, create for today's societal structure and mankind are even more pronounced. This research project has the potential to function as a foundation for further continued research with relevant implications on studies of pollution, biodiversity and climate change. With regards to pollution, the *A. anatina* could potentially be used as an indicator for the presence of mutagenic pollution in their habitats or as an indicator of general water quality. The usage of *A. anatina* as an indicator species would allow for a quick and easy way to indicate the general health of an aquatic environment. This would in turn allow for more informed decisions regarding if further investigation is necessary or if resources should be allocated to other areas. Resource prioritization is of the utmost importance since limited resources are available for environmental efforts. Therefore, it is highly important to use the limited resources efficiently and where the resources have the largest possible impact.

The increased knowledge obtained from continued studies also has the possibility to aid in conservation efforts. Determining what efforts are most impactful for conservation of a species demand a significant amount of information both with regards to the organism itself but also about its habitats and numerous other factors. This means that knowledge about the genome of the *A. anatina* is in of itself not enough however it is one important factor.

Future Research

The conclusion of this report is that the genome assembly unfortunately was not of the desired completeness. The starting point for future research should therefore be to achieve a complete genome. As can be seen in published works on other *de novo* assemblies in the last five years, it is not impossible to create high quality genome assemblies for other freshwater mussels (Bai et al., 2024; Gomes-Dos-Santos et al., 2023a, 2023b, 2024). Therefore, there is nothing that indicates that the associated technologies used in this study cannot achieve this. Worth noting however, is that for the referenced published works of other recent freshwater mussel genome assemblies, other techniques and technologies have been used (Bai et al., 2024; Gomes-Dos-Santos et al., 2023a, 2023b, 2024; Renaut et al., 2018). Therefore, a natural progression would be to start and see what could be adapted to this project to determine if a higher quality genome assembly could be created using those methods.

A commonality for the assemblies is the use of multiple sequencing techniques, further quality control and deeper bioinformatics processing than what was used in this study. The pros of looking into these already published articles is that thanks to the methods and material sections, there are a lot of information given that makes it possible to use their tested method and apply these on the *A. anatina* (Bai et al., 2024; Gomes-Dos-Santos et al., 2023a, 2023b, 2024; Renaut et al., 2018). It should, however, be noted that these individual studies were all performed jointly by multiple people, and this, combined with the associated techniques would mandate the use of significantly more resources than what was used in this study.

References

Abueg, L. A. L., Afgan, E., Allart, O., Awan, A. H., Bacon, W. A., Baker, D., Bassetti, M., Batut, B., Bernt, M., Blankenberg, D., Bombarely, A., Bretaudeau, A., Bromhead, C. J., Burke, M. L., Capon, P. K., Čech, M., Chavero-Díez, M., Chilton, J. M., Collins, T. J., … Zoabi, R. (2024). The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*, *52*(W1), W83–W94. https://doi.org/10.1093/NAR/GKAE410

Antil, S., Abraham, J. S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., Bhagat, P., Gupta, R., Sood, U., Lal, R., & Toteja, R. (2023). DNA barcoding, an effective tool for species identification: a review. *Molecular Biology Reports*, *50*(1), 761–775. https://doi.org/10.1007/S11033-022-08015-7,

Arseneau, J. R., Steeves, R., & Laflamme, M. (2017). Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Molecular Ecology Resources*, *17*(4), 686–693. https://doi.org/10.1111/1755-0998.12616

Bagger, F. O., Borgwardt, L., Jespersen, A. S., Hansen, A. R., Bertelsen, B., Kodama, M., & Nielsen, F. C. (2024). Whole genome sequencing in clinical practice. *BMC Medical Genomics*, *17*(1). https://doi.org/10.1186/S12920-024-01795-W

Bai, Z., Lu, Y., Hu, H., Yuan, Y., Li, Y., Liu, X., Wang, G., Huang, D., Wang, Z., Mao, Y., Wang, H., Chen, L., & Li, J. (2024). The First High-Quality Genome Assembly of Freshwater Pearl Mussel Sinohyriopsis cumingii: New Insights into Pearl Biomineralization. *International Journal of Molecular Sciences*, *25*(6), 3146. https://doi.org/10.3390/IJMS25063146/S1

Bolotov, I. N., Klass, A. L., Kondakov, A. V., Vikhrev, I. V., Bespalaya, Y. V., Gofarov, M. Y., Filippov, B. Y., Bogan, A. E., Lopes-Lima, M., Lunn, Z., Chan, N., Aksenova, O. V., Dvoryankin, G. A., Chapurina, Y. E., Kim, S. K., Kolosova, Y. S., Konopleva, E. S., Lee, J. H., Makhrov, A. A., … Vinarski, M. V. (2019). Freshwater mussels house a diverse mussel-associated leech assemblage. *Scientific Reports*, *9*(1), 16449. https://doi.org/10.1038/S41598-019-52688-3

Burger, N. F. V., Nicolis, V. F., & Botha, A. M. (2025). Evaluating long-read assemblers to assemble several aphididae genomes. *Briefings in Bioinformatics*, *26*(2). https://doi.org/10.1093/BIB/BBAF105

Chen, X., Guo, X., Xiong, W., & Zhan, A. (2022). Pollution-driven morphological plasticity in a running water ecosystem. *Environmental Science and Pollution Research*, *29*(2), 2783–2791. https://doi.org/10.1007/S11356-021-15802-5,

Chen, Z., Grim, C. J., Ramachandran, P., & Meng, J. (2024). Advancing metagenome-assembled genome-based pathogen identification: unraveling the power of long-read assembly algorithms in Oxford Nanopore sequencing. *Microbiology Spectrum*, *12*(6), e00117-24. https://doi.org/10.1128/SPECTRUM.00117-24

Cosma, B. M., Shirali Hossein Zade, R., 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*, giad100. https://doi.org/10.1093/GIGASCIENCE/GIAD100

Cruz, E., Wittstock, W., Snyder, B. A., & Sengupta, A. (2025). Tissue-specific DNA isolation from dissected millipedes for nanopore sequencing. *Biology Methods & Protocols*, *10*(1), bpaf042. https://doi.org/10.1093/BIOMETHODS/BPAF042

de Souza, L. M., de Oliveira, I. D., Sales, F. C. S., da Costa, A. C., Campos, K. R., Abbud, A., Guerra, J. M., dos Santos Cirqueira Borges, C., Takahashi, C. P. F. J., & de Araújo, L. J. T. (2023).

Technical comparison of MinIon and Illumina technologies for genotyping Chikungunya virus in clinical samples. *Journal of Genetic Engineering & Biotechnology*, *21*(1), 88. https://doi.org/10.1186/S43141-023-00536-3

Di Marco, M., Venter, O., Possingham, H. P., & Watson, J. E. M. (2018). Changes in human footprint drive changes in species extinction risk. *Nature Communications*, *9*(1), 4621. https://doi.org/10.1038/S41467-018-07049-5

Ekelund Ugge, G., Jonsson, A., & Berglund, O. (2021). Molecular biomarker responses in the freshwater mussel Anodonta anatina exposed to an industrial wastewater effluent. *Environmental Science and Pollution Research International*, *29*(2), 2158. https://doi.org/10.1007/S11356-021-15633-4

Ekelund Ugge, G., Jonsson, A., Olsson, B., Sjöback, R., & Berglund, O. (2020). Transcriptional and biochemical biomarker responses in a freshwater mussel (Anodonta anatina) under environmentally relevant Cu exposure. *Environmental Science and Pollution Research*, *27*(9), 9999–10010. https://doi.org/10.1007/s11356-020-07660-4

Ferreira-Rodríguez, N., Beggel, S., Geist, J. P., Modesto, V., Österling, M., Riccardi, N., Sousa, R., & Urbańska, M. (2023). Freshwater Mussels as Sentinels for Safe Drinking Water Supply in Europe. *ACS Es&t Water*, *3*(12), 3730. https://doi.org/10.1021/ACSESTWATER.3C00012

Gallardo-Escárate, C., Valenzuela-Muñoz, V., Nuñez-Acuña, G., Valenzuela-Miranda, D., Tapia, F. J., Yévenes, M., Gajardo, G., Toro, J. E., Oyarzún, P. A., Arriagada, G., Novoa, B., Figueras, A., Roberts, S., & Gerdol, M. (2023). Chromosome-Level Genome Assembly of the Blue Mussel Mytilus chilensis Reveals Molecular Signatures Facing the Marine Environment. *Genes*, *14*(4), 876. https://doi.org/10.3390/GENES14040876/S1

Gomes-Dos-Santos, A., Fonseca, E., Riccardi, N., Hinzmann, M., Lopes-Lima, M., & Froufe, E. (2024). The transcriptome assembly of the European freshwater mussel Unio elongatulus C. Pfeiffer, 1825. *Scientific Data*, *11*(1), 377. https://doi.org/10.1038/S41597-024-03226-Y

Gomes-Dos-Santos, A., Lopes-Lima, M., Machado, A. M., Forest, T., Achaz, G., Teixeira, A., Prié, V., Filipe, L., & Froufe, E. (2023a). The Crown Pearl V2: an improved genome assembly of the European freshwater pearl mussel Margaritifera margaritifera (Linnaeus, 1758). *GigaByte*, *2023*, gigabyte81. https://doi.org/10.46471/GIGABYTE.81

Gomes-Dos-Santos, A., Lopes-Lima, M., Machado, M. A., Teixeira, A., C. Castro, L. F., & Froufe, E. (2023b). PacBio Hi-Fi genome assembly of the Iberian dolphin freshwater mussel Unio delphinus Spengler, 1793. *Scientific Data*, *10*(1), 340. https://doi.org/10.1038/S41597-023-02251-7

Goodman, N., Campbell, S., Tong, M., Cameron, D., Brain, M., Borchers Arriagada, N., Wheeler, A. J., Matthews, V., Saini, B., Laachir, K., Walsh, E., Johnston, F. H., & Vardoulakis, S. (2025). Interventions for reducing exposure to air pollution from landscape fires in a changing environment: A systematic review. *Science of the Total Environment*, *966*. https://doi.org/10.1016/j.scitotenv.2025.178621

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072. https://doi.org/10.1093/BIOINFORMATICS/BTT086

Harney, E., Paterson, S., Collin, H., Chan, B. H. K., Bennett, D., & Plaistow, S. J. (2022). Pollution induces epigenetic effects that are stably transmitted across multiple generations. *Evolution Letters*, *6*(2), 118. https://doi.org/10.1002/EVL3.273

Hauck, L. L., Atkinson, C. L., Homyack, J. A., Penaluna, B. E., Mangum, C., Coble, A. A., Nettles, J., Thornton-Frost, J. E., & Fix, M. J. (2023). Molecular identity crisis: environmental DNA metabarcoding meets traditional taxonomy—assessing biodiversity and freshwater mussel populations (Unionidae) in Alabama. *PeerJ*, *11*, e15127. https://doi.org/10.7717/PEERJ.15127/SUPP-3

Head, J. A., Ewald, J. D., & Basu, N. (2024). The DIKW of Transcriptomics in Ecotoxicology: Extracting Information, Knowledge, and Wisdom From Big Data. *Environmental Toxicology and Chemistry*. https://doi.org/10.1002/etc.5954

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/J.YGENO.2015.11.003

Hirt, M. R., Barnes, A. D., Gentile, A., Pollock, L. J., Rosenbaum, B., Thuiller, W., Tucker, M. A., & Brose, U. (2021). Environmental and anthropogenic constraints on animal space use drive extinction risk worldwide. *Ecology Letters*, *24*(12), 2576–2585. https://doi.org/10.1111/ELE.13872;PAGEGROUP:STRING:PUBLICATION

Jiao, W. B., & Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology*, *36*, 64–70. https://doi.org/10.1016/J.PBI.2017.02.002

Jonsson, A., Bertilsson, A., & Rydgård, M. (2013). Spatial distribution and age structure of the freshwater unionid mussels Anodonta anatina and Unio tumidus: Implications for environmental monitoring. *Hydrobiologia*, *711*(1), 61–70. https://doi.org/10.1007/s10750-013-1455-3

Jung, H., Ventura, T., Sook Chung, J., Kim, W. J., Nam, B. H., Kong, H. J., Kim, Y. O., Jeon, M. S., & Eyun, S. Il. (2020). Twelve quick steps for genome assembly and annotation in the classroom. *PLOS Computational Biology*, *16*(11), e1008325. https://doi.org/10.1371/JOURNAL.PCBI.1008325

Khrenova, M. G., Panova, T. V., Rodin, V. A., Kryakvin, M. A., Lukyanov, D. A., Osterman, I. A., & Zvereva, M. I. (2022). Nanopore Sequencing for De Novo Bacterial Genome Assembly and Search for Single-Nucleotide Polymorphism. *International Journal of Molecular Sciences*, *23*(15), 8569. https://doi.org/10.3390/IJMS23158569

Koetsier, P. A., & Cantor, E. J. (2021). A Simple Approach for Effective Shearing and Reliable Concentration Measurement of Ultra-High-Molecular-Weight DNA. *BioTechniques*, *71*(2), 439–444. https://doi.org/10.2144/BTN-2021-0051

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

Lance, E., Lepoutre, A., Brient, L., Maurin, N., Guillon, E., Geffard, A., & Amon-Moreau, D. (2025). Survey of Microcystin-Producing Cyanobacteria in French Lakes of Various Trophic Status Using Environmental and Cyanobacterial Parameters and an Active Mussel Biomonitoring. *Toxins* , *17*(5), 245. https://doi.org/10.3390/TOXINS17050245/S1

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature 2001 409:6822*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Leger, A., & Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source Software*, *4*(34), 1236. https://doi.org/10.21105/JOSS.01236

Lemiere, S., Cossu-Leguille, C., Bispo, A., Jourdain, M. J., Lanhers, M. C., Burnel, D., & Vasseur, P. (2005). DNA damage measured by the single-cell gel electrophoresis (Comet) assay in mammals fed with mussels contaminated by the 'Erika' oil-spill. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *581*(1–2), 11–21. https://doi.org/10.1016/J.MRGENTOX.2004.10.015

Lin, B., Hui, J., & Mao, H. (2021). 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. W., Chaisson, M., & Pevzner, P. A. (2016). Assembly of long error-prone reads using de Bruijn graphs. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(52), E8396–E8405. https://doi.org/10.1073/PNAS.1604560113/SUPPL_FILE/PNAS.1604560113.SAPP.PDF

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*, *2012*, 251364. https://doi.org/10.1155/2012/251364

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265. https://doi.org/10.1016/J.GPB.2016.05.004

Maghini, D. G., Moss, E. L., Vance, S. E., & Bhatt, A. S. (2021). Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome HHS Public Access. *Nat Protoc*, *16*(1), 458–471. https://doi.org/10.1038/s41596-020-00424-x

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/BIOINFORMATICS/BTR011

Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., Loudet, O., Weigel, D., & Ecker, J. R. (2018). High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. *Nature Communications*, *9*(1), 541. https://doi.org/10.1038/S41467-018-03016-2

Mikheenko, A., Saveliev, V., Hirsch, P., & Gurevich, A. (2023). WebQUAST: online evaluation of genome assemblies. *Nucleic Acids Research*, *51*(W1), W601–W606. https://doi.org/10.1093/NAR/GKAD406

Oxford Nanopore Technologies. (2025). *Ligation sequencing DNA V14 (SQK-LSK114)*. https://nanoporetech.com/document/genomic-dna-by-ligation-sqk-lsk114

Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, *52*(4), 413. https://doi.org/10.1007/S13353-011-0057-X

Petersen, C., Sørensen, T., Westphal, K. R., Fechete, L. I., Sondergaard, T. E., Sørensen, J. L., & Nielsen, K. L. (2022). High molecular weight DNA extraction methods lead to high quality filamentous ascomycete fungal genome assemblies using Oxford Nanopore sequencing. *Microbial Genomics*, *8*(4), 000816. https://doi.org/10.1099/MGEN.0.000816

Raghavan, V., Kraft, L., Mesny, F., & Rigerte, L. (2022). A simple guide to de novo transcriptome assembly and annotation. *Briefings in Bioinformatics*, *23*(2), bbab563. https://doi.org/10.1093/BIB/BBAB563

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/S41467-020-14998-3;SUBJMETA=114,1647,2217,2785,631,794;KWRD=COMPUTATIONAL+BIOLOGY+AND+BIOINFORMATICS,GENOME+INFORMATICS,GENOMIC+ANALYSIS,SOFTWARE

Renaut, S., Guerra, D., Hoeh, W. R., Stewart, D. T., Bogan, A. E., Ghiselli, F., Milani, L., Passamonti, M., & Breton, S. (2018). Genome Survey of the Freshwater Mussel Venustaconcha ellipsiformis (Bivalvia: Unionida) Using a Hybrid De Novo Assembly Approach. *Genome Biology and Evolution*, *10*(7), 1637. https://doi.org/10.1093/GBE/EVY117

Saavedra, M. J., Fernandes, C., Teixeira, A., Álvarez, X., & Varandas, S. (2022). Multiresistant bacteria: Invisible enemies of freshwater mussels. *Environmental Pollution*, *295*, 118671. https://doi.org/10.1016/J.ENVPOL.2021.118671

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 Omic Sciences. *Biomolecules*, *14*(5), 568. https://doi.org/10.3390/BIOM14050568

Schalamun, M., Nagar, R., Kainer, D., Beavan, E., Eccles, D., Rathjen, J. P., Lanfear, R., & Schwessinger, B. (2019). Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from Eucalyptus pauciflora. *Molecular Ecology Resources*, *19*(1), 77–89. https://doi.org/10.1111/1755-0998.12938;WEBSITE:WEBSITE:PERICLES;REQUESTEDJOURNAL:JOURNAL:17550998;JOURNAL:JOURNAL:14718286;WGROUP:STRING:PUBLICATION

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/BIOINFORMATICS/BTV351,

Simpson, J. T., & Pop, M. (2015). The Theory and Practice of Genome Sequence Assembly. *Annual Review of Genomics and Human Genetics*, *16*(Volume 16, 2015), 153–172. https://doi.org/10.1146/ANNUREV-GENOM-090314-050032/CITE/REFWORKS

Smith, C. H. (2021). A High-Quality Reference Genome for a Parasitic Bivalve with Doubly Uniparental Inheritance (Bivalvia: Unionida). *Genome Biology and Evolution*, *13*(3). https://doi.org/10.1093/gbe/evab029

Sohn, J. Il, & Nam, J. W. (2018). The present and future of de novo whole-genome assembly. *Briefings in Bioinformatics*, *19*(1), 23–40. https://doi.org/10.1093/BIB/BBW096

Sousa, R., Zając, T., Halabowski, D., Aksenova, O. V., Bespalaya, Y. V., Carvalho, F., Castro, P., Douda, K., da Silva, J. P., Ferreira-Rodríguez, N., Geist, J., Gumpinger, C., Labecka, A. M., Lajtner, J., Lewin, I., Lopes-Lima, M., Meira, A., Nakamura, K., Nogueira, J. G., … Aldridge, D. C. (2023). A roadmap for the conservation of freshwater mussels in Europe. *Conservation Biology*, *37*(2), e13994. https://doi.org/10.1111/COBI.13994

Velasquez-Restrepo, S., Orozco, M. C., Franco-Sierra, N. D., Martínez-Cerón, J. M., & Díaz-Nieto, J. F. (2024). Identification of non-model mammal species using the MinION DNA sequencer from Oxford Nanopore. *PeerJ*, *12*(9), e17887. https://doi.org/10.7717/PEERJ.17887/SUPP-9

Volarić, M., Veseljak, D., Mravinac, B., Meštrović, N., & Despot-Slade, E. (2021). Isolation of high molecular weight DNA from the model beetle Tribolium for nanopore sequencing. *Genes*, *12*(8), 1114. https://doi.org/10.3390/GENES12081114/S1

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

Yang, H., Yang, Z., & Dong, W. (2022). Morphological Identification and Phylogenetic Analysis of Laelapin Mite Species (Acari: Mesostigmata: Laelapidae) from China. *The Korean Journal of Parasitology*, *60*(4), 273. https://doi.org/10.3347/KJP.2022.60.4.273

Ye, X., & Lei, B. (2023). The current status and trends of DNA extraction. *BioEssays*, *45*(8), 2200242. https://doi.org/10.1002/BIES.202200242

Yu, W., Luo, H., Yang, J., Zhang, S., Jiang, H., Zhao, X., Hui, X., Sun, D., Li, L., Wei, X. Q., Lonardi, S., & Pan, W. (2024). Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes. *Genome Research*, *34*(2), 326. https://doi.org/10.1101/GR.278232.123

Zhang, X., Liu, C. G., Yang, S. H., Wang, X., Bai, F. W., & Wang, Z. (2022). Benchmarking of long-read sequencing, assemblers and polishers for yeast genome. *Briefings in Bioinformatics*, *23*(3). https://doi.org/10.1093/BIB/BBAC146

Zieritz, A., Sousa, R., Aldridge, D. C., Douda, K., Esteves, E., Ferreira-Rodríguez, N., Mageroy, J. H., Nizzoli, D., Osterling, M., Reis, J., Riccardi, N., Daill, D., Gumpinger, C., & Vaz, A. S. (2022). A global synthesis of ecosystem services provided and disrupted by freshwater bivalve molluscs. *Biological Reviews of the Cambridge Philosophical Society*, *97*(5), 1967. https://doi.org/10.1111/BRV.12878

Zubach, V., Osahan, G., Kolsun, K., Pfeifer, C., Severini, A., & Hiebert, J. (2025). Development and evaluation of a next-generation sequencing methodology for measles virus using Oxford Nanopore Technology. *Journal of Clinical Microbiology*, *63*(3), e01456-24. https://doi.org/10.1128/JCM.01456-24