Comparing DNA extraction methods for successful PacBio HiFi sequencing: a case study of the freshwater mussel *Anodonta anatina* (Bivalvia: Unionidae)

Authors:
Marco Giulio[1+], Urs Lergster[1+], Philine G. D. Feulner[2,3], Alexandra A.-T. Weber[1]*

[1] Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland
[2] Department of Fish Ecology & Evolution, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Kastanienbaum, Switzerland
[3] University of Bern, Bern, Switzerland

* Corresponding author
[+] Equal contribution

Abstract

High-quality reference genomes are increasingly recognized as essential resources in biodiversity genomics and conservation. However, successful DNA extraction and long-read sequencing remain highly organism-dependent. Molluscs, a diverse phylum of invertebrates, pose particular challenges due to inhibitory compounds and degraded or fragmented DNA, often necessitating careful optimization of extraction protocols. Here, we present a case study on the freshwater mussel *Anodonta anatina* (Bivalvia: Unionidae), evaluating two preservation methods, six DNA extraction protocols, and two post-extraction clean-up steps for their effects on DNA quality and PacBio HiFi sequencing yield from a single individual. The PacBio Nanobind and CTAB protocols produced high-quality DNA from fresh foot tissue but performed poorly on flash-frozen tissue, which is the most common preservation method. Post-extraction clean-up generally degraded DNA and did not improve sequencing yield. Unexpectedly, the column-based Omega Mollusc Kit—although not designed for high-molecular-weight DNA—outperformed the PacBio-recommended Nanobind kit and the manual CTAB method on flash-frozen tissue. It generated high DNA quantity and purity, sufficient integrity for HiFi sequencing, and effective contaminant removal. While the resulting DNA may be too fragmented for Oxford Nanopore sequencing, the Omega Mollusc Kit offers a practical, cost-effective first approach for testing DNA extraction and PacBio sequencing in new mollusc genome projects of relatively large species when tissue is flash-frozen. When fresh tissue is available, Nanobind or CTAB remain the recommended options. This strategy could reduce the need for extensive protocol optimization and facilitate future mollusc genomics efforts.

Keywords

Mollusc genomics; polysaccharides; inhibitors; reference genome

Introduction

The generation of reference genomes plays a crucial role in the conservation and management of biodiversity, providing a foundation for understanding species' evolutionary histories, adaptive potentials, and responses to environmental pressures (Formenti et al., 2022; Paez et al., 2022; Theissinger et al., 2023). Large-scale initiatives such as the Earth BioGenome Project (Lewin et al., 2018, 2022) with dedicated local centers, for instance the European Reference Genome Atlas (Mc Cartney et al., 2024), have emerged to coordinate the global effort to sequence and catalog genomes from all eukaryotic species. These projects are especially important for species-rich groups such as molluscs, one of the most diverse animal phyla with around 87,000 described extant species, with gastropods and bivalves being the two most species-rich classes (MolluscaBase eds. 2025).

Despite the high number of mollusc species, genomic resources for this phylum lag behind other species-rich taxa (Gomes-dos-Santos et al., 2020). The reasons for this gap are multifaceted. While some challenges are attributed to genome complexity (e.g. heterozygosity and repeat content) (Halstead-Nussloch et al., 2024; Chen et al., 2025), a more pervasive issue involves the co-extraction of compounds that inhibit DNA polymerase activity during sequencing. Molluscs are known to produce polyphenolic proteins, mucopolysaccharides, and other biochemical contaminants that complicate the extraction of high-quality DNA, hampering efforts to obtain accurate and complete genomic data (Adema, 2021).

Although numerous DNA extraction protocols have been developed for molluscs (Sokolov, 2000; Arseneau et al., 2017), ensuring sufficient DNA quantity, integrity, and purity for successful sequencing remains a challenge. This is often due to the presence of the above-mentioned contaminants which are not always adequately removed during extraction. Even when high yields of DNA are obtained, the presence of inhibitors can compromise sequencing. Commercial kits specifically designed for molluscs, such as the E.Z.N.A. Mollusc & Insect DNA Kit (Omega Bio-tek), have been developed to address these challenges. These kits typically produce DNA with good yield and purity, making them a popular choice for routine DNA extractions. However, their efficacy is supposedly limited when long-read sequencing technologies are required (e.g. to assemble a reference genome). The frequent use of centrifugation and column-based purification steps in these kits typically results in DNA fragmentation, which is not optimal for techniques such as Oxford Nanopore or PacBio where long DNA fragments (typically >50 kb) are usually required for *de novo* genome sequencing and assembly (Arseneau et al., 2017; Angthong et al., 2020). Nevertheless, this kit was recently applied in the *Anodonta cygnea* genome project, where it enabled successful PacBio HiFi sequencing (Faust et al., 2025), suggesting that its utility for mollusc genome projects may be greater than previously assumed.

Despite adherence to established DNA quality control measures, such as achieving appropriate Nanodrop ratios, good DNA integrity, and quantity, as well as successful library preparation for long-read sequencing platforms like PacBio, the expected sequencing outputs are not always met. For example, cases have been reported where sequencing results yielded

shorter reads and lower output than anticipated, despite following correct protocols (Mollusc Genomics Conference 2024, pers. comm.). This issue has become particularly evident with the latest PacBio Revio chemistry and enzyme, which appear to be more sensitive than the previous version used in the Sequel II system. Specifically, some anectodal observations of sequencing centers have noted a reduction in sequencing output on the Revio platform compared to Sequel II, indicating that despite meeting the standard metrics for DNA quality—such as integrity and purity—there may still be unknown contaminants co-purified with the DNA. These contaminants could potentially inhibit the performance of the high-fidelity polymerase, ultimately compromising the sequencing output. Unfortunately, such issues remain largely undetectable prior to sequencing, as both the DNA and libraries may appear high-quality based on standard QC metrics. As a result, incorporating additional post-extraction DNA clean-up steps may help eliminate residual contaminants and improve sequencing performance. Finally, it has been shown that tissue preservation method (e.g. flash freezing in liquid nitrogen and preservation at -80°C; preservation in ethanol 100%; or using fresh tissue directly) can also influence the outcome of DNA extractions (Alberola-Mora et al., 2025), with flash freezing in liquid nitrogen being the current gold standard for *de novo* reference genome projects (Reichel et al., 2025).

To date, systematic studies that compare long-read sequencing output from different preservation methods and DNA extraction protocols remain uncommon, with some recent exceptions (Angthong et al., 2020; Alberola-Mora et al., 2025). This is a significant gap in the context of large-scale genome initiatives where optimizing genomic resources for non-model organisms such molluscs is critical. In this study, we address this gap by conducting a comparison of various factors that can influence DNA extraction and sequencing success in the freshwater mussel *Anodonta anatina* (Bivalvia: Unionidae), a species notorioulsy difficult to extract high-molecular weigth DNA from (Faust et al., 2025). Specifically, we: 1) Compare two different preservation methods (flash freezing in liquid nitrogen; fresh tissue); 2) Compare six DNA extraction protocols; 3) Evaluate the impact of two additional DNA cleanup steps applied before library preparation; and 4) Assess the sequencing output from five libraries sequenced on one full PacBio Revio cell. Our goal was to provide a comprehensive evaluation of the factors influencing genome sequencing in a freshwater mussel, thereby offering practical guidance for improving genomic resource generation in molluscs and other non-model organisms.

Methods

Sampling and molecular species identification

A collection permit (Nr. 24091) was obtained from Canton Zürich (Switzerland) and five adult individuals of the duck mussel *Anodonta anatina* were collected on the 5th of August 2024 by snorkeling in Türlersee, Switzerland (GPS coordinates: 47.265602, 8.510178). They were brought back to the laboratory and kept alive in the Eawag experimental facility Aquatikum until processing. They were maintained in an open-water circuit at 19°C and fed twice a week using a mixture of lab-grown algae. Given the morphological plasticity of *Anodonta* mussels,

species identification was initially based on shell morphology and subsequently confirmed by sequencing a COI barcode. Briefly, a small tissue biopsy was taken from the foot of each of the five individual mussels and DNA was extracted using the E.Z.N.A Tissue DNA kit (Omega Bio-tek) according to the manufacturer's instructions. DNA integrity, quantity, and purity were assessed using NanoDrop Eight. A fragment of the COI mitochondrial gene was amplified with forward primer LCO22me2: 5'- GGTCAACAAAYCATAARGATATTGG-3' and reverse primer HCO700dy2: 5'- TCAGGGTGACCAAAAAAYCA-3' (Walker et al., 2006). PCR cycling conditions were: an initial step of 95°C for 15 min, followed by 40 cycles of 94°C for 30 s, 48°C for 40 s, 72°C for 60 s and then a final extension of 72°C for 10 min. Amplified PCR products were purified and Sanger sequenced by Microsynth AG, Balgach, Switzerland. The COI sequences of the five *Anodonta* individuals were compared to the NCBI GenBank database using the Basic Local Alignment Search Tool (BLAST) to confirm morphological species identification.

Comparison of DNA extraction methods and quality control

All DNA extractions were performed using ~30 mg of foot tissue from a single *Anodonta anatina* individual to eliminate inter-individual variability and isolate the effects of methodological factors. We tested the impact of two variables on DNA extraction success (Fig. 1): 1) Preservation method: fresh tissue vs. flash-frozen in liquid nitrogen and subsequently preserved at -80°C; 2) DNA extraction method: six protocols, five of which being specifically designed to extract high-molecular weight (HMW) DNA (Table 1). Each extraction condition was performed in triplicate, except for the GenomicTip and MegaLong protocols with frozen tissue, which were conducted in duplicate. All DNA extractions were conducted following the manufacturer's instructions or using a dedicated protocol (Table 1). For the Nanobind and Omega kits, tissue homogeneisation was performed with a TissueRuptor II (QIAGEN). For the Omega kit, a single final elution step was conducted. DNA extractions using the six protocols were first performed on fresh foot tissue, as this allowed non-lethal sampling (small biopsy) and kept the individual alive. Then, the specimen was killed and remaining tissues (foot, gills, and whole body) were flash-frozen in liquid nitrogen and stored at –80°C. Subsequent extractions from flash-frozen foot tissue were conducted using the six DNA extraction methods. Finally, to evaluate the effect of clean-up on DNA quality, we applied two commercial clean-up kits (Zymo and PowerClean) to extracts from the two top-performing column-free methods (CTAB and Nanobind).

DNA quality control included three metrics: 1) Yield (Qubit dsDNA High Sensitivity kit); 2) Purity (Nanodrop 260/280 ratio); 3) Integrity (Agilent TapeStation gDNA ScreenTape). A DNA extraction was considered successful if it met all of the following thresholds: 1) Yield: >1 µg total DNA; 2) Purity: 260/280 ratio between 1.8 and 2.0; 3) Integrity: largest TapeStation peak >15 kb.

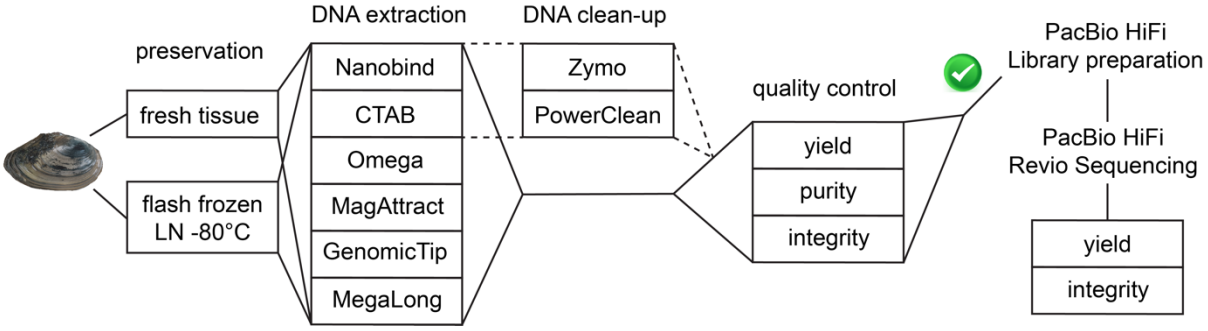

Figure 1: Experimental design for comparing DNA extraction methods and sequencing output in the freshwater mussel *Anodonta anatina*. LN: liquid nitrogen. See Table 1 for details regarding DNA extraction methods.

164 **Table 1:** List of the DNA extraction methods and DNA clean-up kits used in this study. HMW: high-molecular weight.
165

Kit or protocol name	Purpose	Short name	Kit price (CHF)	Number of preps	Price per prep (CHF)	Link
PacBio Nanobind panDNA kit	HMW DNA extraction for animals	Nanobind	315	24	13.1	https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-extraction-manual-moll-14egn36nyl5d/v1
CTAB for molluscs	HMW DNA extraction for molluscs	CTAB	237	~3000	0.08	https://www.protocols.io/view/ctab-dna-extraction-protocol-for-mollusks-e6nvwjpr2lmk/v1
Omega Bio-tek E.Z.N.A. Mollusc & Insect DNA Kit	DNA extraction for mollusc	Omega	480	200	2.4	https://omegabiotek.com/product/e-z-n-a-mollusc-dna-kit/
Qiagen MagAttract	HMW DNA extraction	MagAttract	350	48	7.3	https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/magattract-hmw-dna-kit-48
Qiagen Genomic-tip 100G	HMW DNA extraction	GenomicTip	422	25	16.9	https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiagen-genomic-tips?catno=10223
G-Biosciences MegaLong	HMW DNA extraction	MegaLong	768	25	30.7	https://www.gbiosciences.com/MegaLong/MegaLong-359
Qiagen DNeasy PowerClean Pro Cleanup Kit	DNA clean-up	Powerclean	362	50	7.2	https://www.qiagen.com/us/products/discovery-translational-research/dna-rna-purification/dna-purification/dna-clean-up/dneasy-powerclean-pro-cleanup-kit
ZYMO Genomic DNA Clean & Concentrator-10	DNA clean-up	Zymo	133	25	5.3	https://zymoresearch.eu/collections/genomic-dna-clean-concentrator-gdcc/products/genomic-dna-clean-concentrator-25

166

167 *PacBio library preparation and sequencing*
168 A total of 15 DNA extractions from nine treatments were brought to the Lausanne Genomic
169 Technologies Facility (GTF) sequencing platform in Lausanne (Switzerland) for quality control,
170 PacBio library preparation and sequencing. Integrity of DNA was first assessed using a 5200
171 Fragment Analyzer (Agilent Technologies). Based on quality control, a total of six DNA
172 extractions were selected for PacBio library preparation. DNA fragments below <10 kb were
173 partially removed with a short-read eliminator (SRE) kit (PacBio). DNA of Omega_frozen1
174 sample was not sheared because it was already fragmented enough. For the five remaining
175 DNA extractions, DNA was sheared in a Megaruptor 3 DNA shearing system (Diagenode) with
176 a target mean fragment length of 15-20 kb. Fragments profile was assessed by capillary
177 electrophoresis on a Fragment Analyzer using an HS large fragment 50 kb kit (Agilent
178 Technologies). Repair and A-tailing of DNA fragments, SMRTbell adapters ligation, nuclease
179 treatment and removal of fragments <3kb were performed according to the SMRTbell prep
180 kit 3.0 protocol (Pacific Biosciences). Primer annealing and polymerase binding were
181 performed using the Pacific Biosciences Revio SPRQ polymerase kit including a DNA internal
182 control according to the manufacturer's protocol. Sequencing run was set up using SMRT Link
183 v25.01. Samples were sequenced in HiFi mode in a Pacific Biosciences Revio platform using
184 Revio SPRQ Sequencing plate and one 25M ZMW SMRT cell with a 30-hour movie time and 2
185 hours of pre-extension time.
186
187 **Results**
188 *Molecular species identification, comparison of DNA extraction performance and DNA clean-*
189 *up*
190 Using COI barcodes, all five analyzed individuals were assigned to *Anodonta anatina*,
191 confirming the morphological species identification. A single representative individual (AE01)
192 was subsequently used for all DNA extraction tests. Three high-molecular weight (HMW) DNA
193 extraction protocols—MagAttract, GenomicTip, and MegaLong—failed to produce DNA of
194 sufficient quality, regardless of whether fresh or flash-frozen tissue was used. These protocols
195 did not meet the minimum quality criteria for yield, purity, or integrity (Fig. 2; Table S1). In
196 particular, the GenomicTip and MegaLong kits yielded almost no detectable DNA, and
197 therefore none of these extracts were selected for sequencing.
198 In contrast, the remaining three protocols—Nanobind, CTAB, and Omega—produced
199 satisfactory DNA yield, purity, and integrity from fresh tissue, with CTAB yielding the highest
200 integrity (Fig. 2; Table S1). As expected, the high-molecular-weight extraction methods
201 (Nanobind and CTAB) generated DNA with greater integrity than the column-based Omega
202 kit. Nonetheless, the Omega kit still produced relatively long fragments (approximately 18–
203 31 kb; Table S1), sufficient for several long-read sequencing applications. Interestingly, all
204 three methods yielded substantially more DNA when using flash-frozen tissue, but this
205 increase came at the expense of integrity, which was markedly reduced compared to
206 extractions from fresh tissue (Fig. S1). Actually, the Omega kit did not meet our integrity
207 threshold on flash-frozen material (largest TapeStation peak >15 kb). However, because the

DNA profiles showed a clean and promising fragment distribution, we proceeded to submit these extracts to the sequencing center for further quality assessment.

To further assess potential improvements in DNA quality, we subjected extracts obtained with the two most promising column-free protocols (Nanobind and CTAB) from fresh tissue to additional clean-up steps using two commercial kits (Zymo and PowerClean; Fig. 3). Unexpectedly, the Zymo clean-up reduced the purity of Nanobind DNA, whereas CTAB-derived DNA remained stable after the same treatment. In both cases, the PowerClean protocol markedly degraded the DNA, rendering it unsuitable for sequencing. However, DNA integrity after the Zymo clean-up remained sufficient for downstream long-read sequencing.

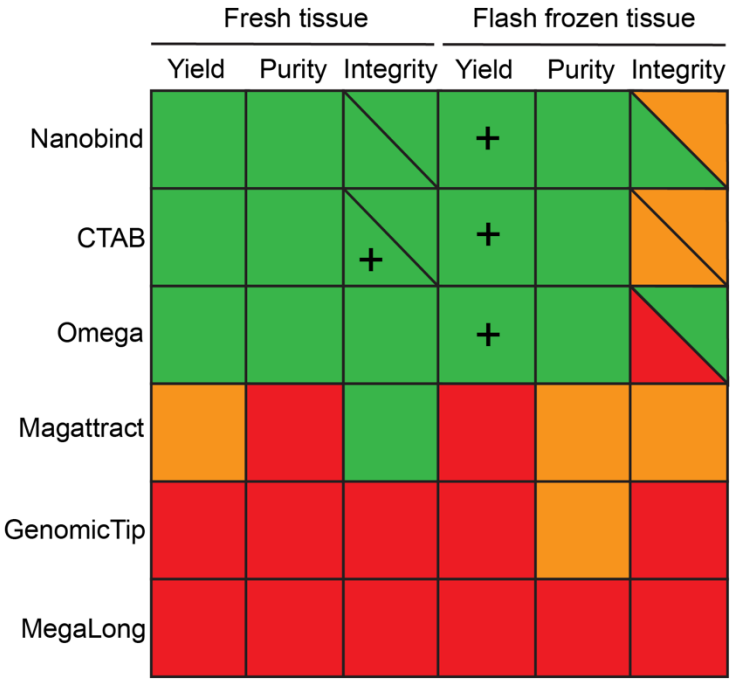

Figure 2. DNA extraction performance across methods.

Thresholds for successful extraction were defined as: yield >1 µg total DNA ("+" indicates >10 µg across three replicates), purity with an A260/A280 ratio of 1.8–2.0, and integrity with the largest TapeStation peak >15 kb ("+" indicates largest TapeStation peak >60 kb across three replicates). Color coding: green = success in all replicates; orange = success in one to two replicates; red = no success. Extractions from five treatments (Nanobind fresh and frozen, CTAB fresh and frozen, Omega frozen) were submitted to the sequencing facility, where DNA integrity was confirmed on a Fragment Analyzer. Integrity values are shown as diagonals: lower left = TapeStation results from the extraction laboratory; upper right = Fragment Analyzer results from the sequencing center.

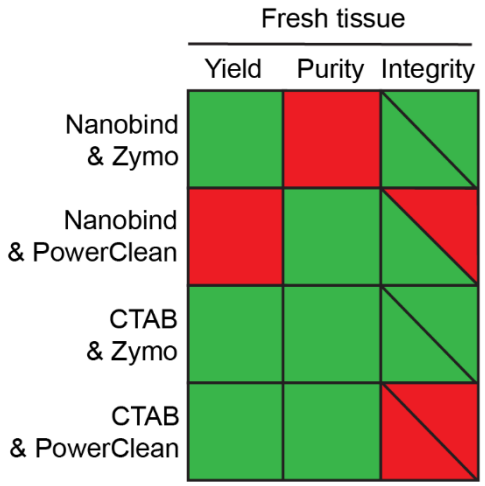

Figure 3: DNA quality metrics after additional clean-up step. Color coding: green = success. Red = no success. All four extractions were submitted to the sequencing facility, where DNA integrity was confirmed on a Fragment Analyzer. Integrity values are shown as diagonals: lower left = TapeStation results from the extraction laboratory; upper right = Fragment Analyzer results from the sequencing center.

DNA extraction quality control, library preparation and PacBio sequencing at the sequencing center

Based on the quality control conducted at the DNA extraction laboratory in Dübendorf, 15 DNA extractions representing eight independent treatment conditions were selected for further quality control, library preparation, and sequencing at the GTF platform in Lausanne (Table 2). During the second round of quality control, substantial discrepancies in DNA integrity values were observed between instruments: the TapeStation (Eawag Dübendorf) and the Fragment Analyzer (GTF Lausanne) produced several inconsistent results (Fig. S2). These differences likely reflect both known platform-specific biases—since the TapeStation is known to slightly overestimate DNA fragment lengths—and most likely post-extraction DNA degradation (Fig. 2; Fig. 3).

For instance, Nanobind extractions from flash-frozen tissue displayed relatively high integrity when analyzed at the extraction laboratory but were found to be completely degraded upon re-evaluation at the sequencing facility (Fig. 2; Table 2; Fig. S2). Interestingly, this degradation was not observed for Nanobind extractions from fresh tissue, which retained their integrity across both analyses. A comparable pattern was found for CTAB extractions, where DNA from frozen tissue appeared substantially degraded at the sequencing center, while DNA from fresh tissue remained stable (Fig. S2). In contrast, no evidence of post-extraction degradation was detected for DNA obtained with the Omega kit; in fact, integrity values were relatively higher when assessed with the Fragment Analyzer (~22-24 kb) compared to the TapeStation (~13-14 kb), which was unexpected.

258 **Table 2:** Quality control summary of the 15 DNA extractions brought to the sequencing facility GTF. Extraction protocol, tissue preservation and potential clean-up steps are
259 included in the sample name. All DNA extractions were performed using foot tissue from the same *A. anatina* individual. Technical replicates are numbered 1-3. The size
260 detection limit of the genomic DNA Tapestation is 60,000 bp. Tapestation QC was performed in the extraction laboratory in Dübendorf. Fragment Analyzer QC was performed
261 at the sequencing facility GTF in Lausanne. *Samples used for library preparation. [1] yield calculated from Qubit concentration.

Sample name	Treatment	Yield[1] [ug]	Qubit [ng/µl]	Nanodrop [ng/µl]	A260/A280	A260/A230	Largest peak Tapestation [bp]	Largest peak Fragment Analyzer [bp]
*CTAB_fresh1	1	4.8	48.4	122.5	1.94	1.62	>60,000	48,313
*CTAB_fresh_Zymo1	2	1.9	41.6	53.7	1.88	3.07	51,900	47,003
CTAB_fresh_Powerclean 1	3	1.5	33	57.9	1.84	1.92	12,973	147
*Nanobind_fresh1	4	7.5	74.6	159.8	1.86	2.24	>60,000	47,751
*Nanobind_fresh_Zymo 1	5	1.6	36.2	40.4	1.76	3.16	>60,000	40,640
Nanobind_fresh_Power clean1	6	0.6	13.4	16.3	1.92	1.04	17,352	8,557
*Omega_frozen1	6	18.8	125	209.7	1.84	2.17	13,080	24,733
Omega_frozen2	6	93.8	625	902.8	1.87	2.34	13,519	22,861
Omega_frozen3	6	75.8	505	1111.6	1.86	2.37	14,075	22,300
Nanobind_frozen1	7	90	600	920.8	1.93	2.34	27,603	3,810
Nanobind_frozen2	7	38.1	254	493.5	1.96	2.33	>60,000	3,953
Nanobind_frozen3	7	124.5	830	974.9	1.91	2.39	22,985	>60,000
*CTAB_frozen1	8	42.6	284	461.3	1.92	2.71	18,259	21,177
CTAB_frozen2	8	41.6	277	457.5	1.93	2.72	10,900	84
CTAB_frozen3	8	39.9	266	426.7	1.92	2.76	14,120	12,813

262

263 Following quality control at GTF, six samples (marked with asterisks in Table 2) were selected
264 for library preparation. Although all six libraries were successfully prepared and loaded onto
265 a single PacBio Revio SMRT Cell, the initial run failed due to a technical issue with the PacBio
266 instrument and had to be repeated. Due to the exhaustion of the CTAB_fresh_Zymo1 library,
267 only five libraries were ultimately sequenced.
268 The final sequencing run was successful, yielding a total of 91 Gb, which is in the range of
269 expected output of a PacBio Revio cell (Table 3). One library (CTAB_frozen1) produced shorter
270 reads, likely due to more extensive shearing compared to other samples, despite using the
271 same shearing parameters. This resulted in reduced HiFi read length and overall yield. In
272 contrast, the remaining libraries produced yields between 18 and 22.9 Gb, with
273 Omega_frozen1 achieving the highest output. A trade-off between HiFi read length and
274 quality was observed, with shorter reads generally having higher quality. Nevertheless, even
275 the longest reads (mean ~13 kb) maintained excellent quality (Q38). Overall, all five libraries
276 were successfully sequenced, and read length distributions closely mirrored the respective
277 library fragment sizes, indicating the absence of sequencing inhibitors.
278
279 **Table 3:** PacBio HiFi sequencing output on one Revio cell for each of the five libraries.

Sample Name	HiFi Reads	HiFi Read Length (mean, bp)	HiFi Read Quality (median, QV)	HiFi Yield (Gbp)
CTAB_fresh1	1,792,289	10,087	Q41	18.08
Nanobind_fresh1	1,493,698	12,371	Q38	18.48
Nanobind_fresh_Zymo1	1,469,733	13,051	Q38	19.18
Omega_frozen1	1,957,907	11,698	Q39	22.91
CTAB_frozen1	2,002,334	6,347	Q47	12.71
Not barcoded	51,897	11,226	Q30	0.58
Total	**8,767,858**	**10,796**	**Q40**	**91.94**

280
281 **Discussion**
282 *Tissue preservation and extraction protocol shape extraction outcomes*
283 Molluscs present well-known challenges for DNA extraction due to high levels of mucus and
284 endogenous inhibitors (Adema, 2021). In this study, we evaluated different strategies for
285 extracting high-molecular-weight (HMW) DNA from the freshwater mussel *Anodonta anatina*,
286 a species for which we previously encountered difficulties in generating sufficient quality data
287 for PacBio sequencing (Faust et al., 2025). While our results are based on a single species, we
288 believe they provide valuable insights likely applicable to other unionids and molluscs more
289 broadly.
290 We found that tissue preservation had a critical influence on the performance of the top three
291 HMW DNA extraction protocols. This effect was evident both in terms of yield and integrity.
292 First, yield differed markedly between preservation methods, with approximately ten times
293 more DNA recovered from flash-frozen tissue compared with fresh tissue. Second, we
294 observed clear differences in DNA integrity between fresh and flash-frozen tissue extracted
295 using the Nanobind and CTAB protocols. While both protocols produced high-quality and

296 stable DNA from fresh tissue, DNA obtained from flash-frozen tissue initially showed relatively
297 high integrity at the extraction laboratory in Dübendorf but exhibited strong post-extraction
298 degradation when re-assessed at the sequencing center in Lausanne. This occurred despite
299 identical handling of all samples—storage at −20 °C after extraction and quality control, and
300 transport on dry ice. The cause of this degradation remains unclear. One possibility is that
301 residual compounds, such as DNases, remained active in extracts from frozen tissue, though
302 this explanation remains speculative. Nevertheless, these findings demonstrate a clear
303 advantage of extracting DNA from fresh tissue when feasible. We acknowledge, however, that
304 this is often impractical—particularly in field-based or conservation genomic contexts—
305 where optimized methods for preserved material remain indispensable (Reichel et al., 2025).
306 Among the flash-frozen samples, the Omega kit consistently yielded DNA of moderate
307 fragment length (~13–14 kb on the TapeStation; ~22-24 kb on the FragmentAnalyzer) and
308 showed high reproducibility across replicates. Importantly, no post-extraction degradation
309 was detected, suggesting that any residual compounds potentially responsible for
310 degradation in the Nanobind and CTAB extractions were effectively removed by the Omega
311 protocol. When applied to fresh tissue, the Omega kit also produced consistent results,
312 yielding DNA of moderate fragment length (~18–31 kb) and generally high purity. Together,
313 these findings highlight the robustness and versatility of the Omega kit, which performed
314 reliably across both fresh and flash-frozen samples.
315
316 *High-molecular weight DNA kits do not guarantee better results*
317 Three commercial HMW DNA kits (MagAttract, GenomicTip, Megalong) failed to yield usable
318 DNA in our tests, from both fresh and flash-frozen tissues. In contrast, we successfully
319 extracted high-quality DNA using a manual CTAB protocol, the Nanobind kit, and the Omega
320 kit—though success depended strongly on tissue preservation. These results emphasize that
321 dedicated HMW kits are not necessarily superior to manual or general-purpose protocols in
322 molluscs. Factors such as cost, processing time, and consistency should be considered when
323 choosing a protocol.
324 Because mollusc DNA extractions are notoriously prone to inhibition by co-extracted
325 contaminants, we tested additional post-extraction clean-up steps using PowerClean and
326 Zymo kits to further remove residual inhibitors after standard DNA extraction. The
327 PowerClean kit was unsuitable for genomic DNA, leading to severe degradation. In contrast,
328 the Zymo kit preserved fragment length more effectively. However, clean-up did not provide
329 clear added value, as libraries prepared from unpurified extractions (including from relatively
330 dirty samples like CTAB_fresh1) sequenced successfully. We therefore do not recommend an
331 additional clean-up of DNA extracts prior to PacBio library preparation.
332 Ultimately, we generated high-quality PacBio data from five different DNA extraction and
333 preservation combinations. Despite pooling all libraries on the same SMRT cell, we observed
334 no evidence that inhibitors from one sample compromised sequencing success in others.
335 Overall read yield was consistent across treatments, suggesting that inhibitor removal was

336 generally successful, and that the selected protocols are robust enough for long-read
337 sequencing applications in *A. anatina.*
338
339 *Recommendations for a new mollusc genome project*
340 We showed that several HMW DNA extraction kits failed to produce DNA of sufficient quality
341 for PacBio sequencing. Some protocols (PacBio Nanobind and CTAB) were successful, but
342 their performance depended strongly on tissue preservation, leading to high-quality DNA only
343 using fresh tissue. We would therefore recommend using one of these protocols if fresh tissue
344 is available. It is also noteworthy that the manual CTAB protocol requires considerably longer
345 hands-on time, while the Nanobind kit is comparatively rapid and user-friendly. Because we
346 focused on *A. anatina*—a species notoriously difficult for DNA extraction (Faust et al., 2025)—
347 rather than testing multiple species, we reduced confounding factors while still generating
348 results likely relevant across a broad range of mollusc species, tissue types, and preservation
349 methods.
350 In the present study, the Omega kit consistently produced DNA of sufficient quality for PacBio
351 sequencing, especially for flash-frozen tissue. The same kit has been successfully applied in
352 the *Anodonta cygnea* genome project using flash-frozen tissue (Faust et al., 2025). In another
353 genome project conducted in our lab, we also obtained high-quality PacBio HiFi data from an
354 RNAlater-preserved *A. exulcerata* sample stored at –80 °C using this kit – we however
355 acknowledge that this is an anecdotal observation. Although ethanol-preserved tissues were
356 not tested here, we expect that this preservation method would also work with the Omega
357 kit, given its routine use for standard DNA extractions.
358 We are aware that the Omega kit was not initally designed for long-read sequencing, and
359 therefore fragment sizes are relatively modest. Hence, small adjustments to PacBio library
360 preparation can - or should - be performed to increase the length of the resulting sequencing
361 data. For example, in the present study, DNA was not sheared and was used directly in library
362 preparation after the standard SRE step. In the above-mentioned *A. exulcerata* project, the
363 sequencing platform further recommended a BluePippin purification step to remove small
364 fragments (<10 kb) after PacBio library preparation. Nevertheless, such adjustments are
365 straightforward, add little cost compared with months of DNA extraction optimization, and
366 ensure effective sequencing output. In all cases, we recommend close contact and open
367 communication with the experienced sequencing platform conducting library preparation
368 and sequencing.
369 For new mollusc genome projects, it is often most practical to begin with broadly reliable
370 strategies rather than investing heavily in expensive HMW extraction kits and extensive
371 protocol optimization. Our results show that the choice of DNA extraction method should be
372 guided primarily by tissue preservation. For fresh tissue, the PacBio Nanobind and manual
373 CTAB protocols consistently produced the highest DNA integrity and are therefore the
374 preferred options. For flash-frozen tissue, which is commonly used in field collections, the
375 column-based Omega kit performed unexpectedly well, providing sufficient DNA quantity,
376 purity, and fragment length for PacBio HiFi sequencing despite not being designed for HMW

extraction. Indeed, to our knoweldge, it has yielded high-quality PacBio data in three *Anodonta* species (Faust et al., 2025). The Omega kit does have limitations—adequate tissue input is needed to allow size selection, short fragments may require removal (e.g., via BluePippin), and it is not suitable when ultra-long reads are required. In such cases, Nanobind or CTAB remain better choices. Overall, our findings suggest a practical decision framework: Nanobind or CTAB should be used when high-quality fresh tissue is available, while the Omega kit should be considered as an efficient and cost-effective first attempt when working with flash-frozen material. This approach may reduce unnecessary optimization efforts and accelerate future mollusc genome projects.

References

Adema, C. M. (2021). Sticky problems: extraction of nucleic acids from molluscs. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 20200162. doi: 10.1098/rstb.2020.0162

Alberola-Mora, I., Guerra-Font, O., Espinoza-Calderón, O. D., Galià-Camps, C., and Džunková, M. (2025). Combination of Sample Preservation Approaches and DNA Extraction Methods for Long-Read Sequencing of Nudibranchs' Genomes. *Ecol. Evol.* 15, e71262. doi: 10.1002/ece3.71262

Angthong, P., Uengwetwanit, T., Pootakham, W., Sittikankaew, K., Sonthirod, C., Sangsrakru, D., et al. (2020). Optimization of high molecular weight DNA extraction methods in shrimp for a long-read sequencing platform. *PeerJ* 8, e10340. doi: 10.7717/peerj.10340

Arseneau, J.-R., Steeves, R., and Laflamme, M. (2017). Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Mol. Ecol. Resour.* 17, 686–693. doi: 10.1111/1755-0998.12616

Chen, Z., Baeza, J. A., Chen, C., Gonzalez, M. T., González, V. L., Greve, C., et al. (2025). A genome-based phylogeny for Mollusca is concordant with fossils and morphology. *Science* 387, 1001–1007. doi: 10.1126/science.ads0215

Faust, E., Conrads, J., Giulio, M., Ciofi, C., Natali, C., Feulner, P. G. D., et al. (2025). Beyond genetic indicators: how reproductive mode and hybridisation challenge freshwater mussel conservation. *Mol. Ecol.* 34, e70066. doi: 10.1111/mec.70066

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., et al. (2022). The era of reference genomes in conservation genomics. *Trends Ecol. Evol.* 37, 197–202. doi: 10.1016/j.tree.2021.11.008

Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L. F. C., and Froufe, E. (2020). Molluscan genomics: the road so far and the way forward. *Hydrobiologia* 847, 1705–1726. doi: 10.1007/s10750-019-04111-1

Halstead-Nussloch, G., Signorini, S. G., Giulio, M., Crocetta, F., Munari, M., Della Torre, C., et al. (2024). The genome of the rayed Mediterranean limpet *Patella caerulea* (Linnaeus, 1758). *Genome Biol. Evol.* 16, evae070. doi: 10.1093/gbe/evae070

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., et al. (2022). The Earth BioGenome Project 2020: Starting the clock. *Proc. Natl. Acad. Sci.* 119, e2115635118. doi: 10.1073/pnas.2115635118

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., et al. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci.* 115, 4325–4333. doi: 10.1073/pnas.1720115115

Mc Cartney, A. M., Formenti, G., Mouton, A., De Panis, D., Marins, L. S., Leitão, H. G., et al. (2024). The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. *Npj Biodivers.* 3, 1–17. doi: 10.1038/s44185-024-00054-6

MolluscaBase eds. (2025). MolluscaBase. Accessed at: https://www.molluscabase.org on 26 November 2025.

Paez, S., Kraus, R. H. S., Shapiro, B., Gilbert, M. T. P., Jarvis, E. D., and VERTEBRATE GENOMES PROJECT CONSERVATION GROUP (2022). Reference genomes for conservation. *Science* 377, 364–366. doi: 10.1126/science.abm8127

Reichel, K., Pohjoismäki, J., Astrin, J., Böhne, A., Bortoluzzi, C., Buzan, E., et al. (2025). Addressing key challenges in sample handling for high-quality reference genome generation. Available at: https://www.authorea.com/users/930028/articles/1301342-addressing-key-challenges-in-sample-handling-for-high-quality-reference-genome-generation (Accessed November 26, 2025).

Sokolov, E. P. (2000). An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *J. Molluscan Stud.* 66, 573–575.

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., et al. (2023). How genomics can help biodiversity conservation. *Trends Genet.* 39, 545–559. doi: 10.1016/j.tig.2023.01.005

Walker, J. M., Curole, J. P., Wade, D. E., Chapman, E. G., Bogan, A. E., Watters, G. T., et al. (2006). Taxonomic distribution and phylogenetic utility of gender-associated mitochondrial genomes in the Unionoida (Bivalvia). *Malacologia* 48, 265.

Acknowledgments

We thank the Canton of Zürich for providing the *Anodonta anatina* sampling permit, and Julie Conrads and Joana L. Santos for their assistance with field sampling. We are also grateful to Julien Marquis and Mélanie Dupasquier from the GTF platform at the University of Lausanne (UNIL) for their support with library preparation and PacBio sequencing. This work was supported by an Eawag Discretionary Fund Grant awarded to PGDF and AATW.

Data availability

PacBio HiFi whole-genome sequencing data are available from the European Nucleotide Archive (ENA) under BioProject accession PRJEB105293. The dataset comprises five barcoded libraries and one set of unassigned reads generated from a single sequencing flow cell, all

457 derived from the same BioSample (SAMEA120807678), with sequencing runs deposited
458 as ERR16075080 and ERR16075215–ERR16075219.
459
460 **Author contributions**
461 MG: Investigation, Supervision, Writing – review and editing.
462 UL: Data curation, Investigation.
463 PGDF: Funding acquisition, Writing – review and editing.
464 AATW: Conceptualization, Funding acquisition, Project administration, Resources,
465 Supervision, Visualization, Writing – original draft, Writing – review and editing.
466
467 **Supplementary material**
468
469 **Figure S1:** Tapestation profiles of all DNA extractions, treatments and replicates. **A.** Nanobind
470 extractions. **B.** CTAB extractions. **C.** Omega extractions. **D.** MagAttract extractions. **E.**
471 GenomicTip extractions. **F.** MegaLong extractions. **G.** Nanobind and CTAB extractions after
472 DNA clean-up.
473
474 **Figure S2:** Tapestation vs. Fragment Analyzer profiles of the 15 DNA extractions brought to
475 the sequencing center in Lausanne, Switzerland. **A.** CTAB extractions using fresh tissue. **B.**
476 Nanobind extractions using fresh tissue. **C.** Omega extractions using frozen tissue. **D.**
477 Nanobind extractions using frozen tissue. **E.** CTAB extractions using frozen tissue.
478
479 **Table S1:** Quality control results of all DNA extractions, treatments and replicates.
480
481 **Table S2:** Quality control results of DNA extractions following DNA clean-up steps.
482