UNIVERSITA' DEGLI STUDI DI VERONA

DEPARTMENT OF

Biotechnology

GRADUATE SCHOOL OF

Natural Sciences and Engineering

DOCTORAL PROGRAM IN

Biotechnology

XXXII cycle

TITLE OF THE DOCTORAL THESIS

Optimization of High and Ultra High Molecular Weight DNA purification for Third Generation Sequencing and Optical Mapping in algae

S.S.D. BIO/18

Coordinator: Prof. Matteo Ballottari

Tutor: Prof. Massimo Delledonne

Doctoral Student: Emanuela Cosentino

Optimization of High and Ultra High Molecular Weight DNA purification for Third Generation Sequencing and Optical Mapping in algae

Emanuela Cosentino

Tesi di Dottorato

Verona, 21 Maggio 2020

A chi crede in me più di me stessa!

A mia madre, mio padre, Giacomo, la mia famiglia ed i miei veri amici…

sapere di non deludervi mai mi permette di sbagliare.

ABSTRACT

The analysis of long DNA molecules by novel genomic technologies, such as Bionano optical mapping and Third Generation Sequencing, including PacBio Single Molecule Real Time Sequencing and Oxford Nanopore sequencing, provide the opportunity for complete genome characterization and reconstruction, allowing to identify large (balanced) structural variants, to determine the variant phasing and haplotype, to sequence full-length repeated regions and to assemble and scaffold genomes de-novo. Implementation of these technologies requires a combination of highly pure and High Molecular Weight (HMW) DNA, >10^5bp (Bionano Optical Mapping) or >10^4bp (Third Generation Sequencing) in length. However, standardized and suitable extraction methods to obtain highly pure HMW DNA are still missing for many organisms and tissues. In particular, plants and algae store a large amount of phenolic compounds, polysaccharides and a high copy number of chloroplast and mitochondrial DNA, making the extraction of both pure and HMW genomic DNA challenging.

The aim of this work was the optimization of methods for the purification of highly pure and (Ultra)HMW DNA from a microalgae selected as case study, *Haematococcus pluvialis* (*H.pluvialis*), suitable for Third Generation sequencing and Bionano optical mapping. Despite *H.pluvialis* is unicellular green microalgae extensively studied for industrial applications, a high quality genome for its biotechnological application is still missing. Therefore, an extensive benchmarking of DNA and nuclei isolation methods was conducted to produce high-quality HMW DNA suitable to generate Third Generation sequencing and Bionano optical mapping data for the reconstruction of its genome *de-novo*.

4 (U)HMW DNA extraction methods and 8 nuclei isolation methods and 4 post-extraction DNA purification methods were evaluated independently or in combination. To further improve DNA purity and optimize the production of high-quality sequencing data, 4 post-extraction DNA purification methods were also tested. The methods were compared in terms of yield, length and purity of extracted DNA and its analysis by Third Generation sequencing and optical mapping. Only 3 specific combinations of these protocols yielded suitable DNA

to generate successful results with PacBio (CTAB buffer+AMPureXP beads purification), Oxford Nanopore (MEB buffer+G-tip- DNA based extraction) and Bionano (MEB buffer+plug- DNA based extraction). The data produced herein can be used to obtain a highly contiguous genome for *H.pluvialis* with the efficient reconstruction of repetitive genomic portions (highly present in *H.pluvialis* genome), by eliminating ambiguity in the positions or size of genomic elements.

INTRODUCTION

Progress in genomics has been moving steadily forward due to a revolution in DNA sequencing technologies [1]. This fast progress resulted in a substantial reduction in costs and increase in throughput and accuracy. These advancements were mainly due to the advent of second generation technology, better known as next generation sequencing (NGS) technology, which employed massively parallel reactions for high throughput. NGS is currently the most widely used DNA sequencing technology to characterize genomes. However, despite NGS is able to perform very efficiently and accurately, it is characterized by some limitations. The largest hurdle is related to sequencing low-complexity or highly homologous regions, that are longer than typical NGS reads. In fact, NGS relies upon short-read sequences that lack contextual information; this limits their utility in genome assembly and in resolving complex and repetitive regions of the genome [2]. Hence, the limitations of short-read sequencing have left a substantial fraction of most genomes inaccessible and much of their true complexity hidden [3][4].

Long reads sequencing approaches

To overcome these limitations a third generation of sequencing technologies (3GS) producing long-reads (>10Kb) was developed. In comparison with NGS, 3GS produce similar throughput but longer read lengths. The analysis of long DNA fragments provides the opportunity to acquire a complete genome characterization through the capability to identify large balanced structural variants (SVs), to determine the variant phasing (haplotypes), to characterize highly homologous regions (paralogous genes and pseudogenes), to sequence full-length repeated regions and to perform assembly and scaffolding to reconstruct genomes de-novo (**Table 1**). These long reads can help closing the gaps in genomes by spanning the low complexity regions that would otherwise require many costly YAC, BAC and fosmid clones to be created and sequenced. In addition, the preparation of DNA libraries for 3GS does not require any clonal

amplification step by PCR, thus reducing biases and generating a more homogeneous genome coverage [5].

		Short reads	Long reads
Highlights	Cheap	+	-
	PCR-free	-	++
	Accuracy	++	+
Critical genomic regions/features	High GC content	-	+
	Repetive regions	-	+
	Balanced structural variations	-	+
	Paralogous regions	-	+
	Haplotype phasing	-	+

Table 1: Advantages of long-read based analysis as compared to short-reads.

The companies currently at the heart of 3GS technology are two: Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT).

i) PacBio SMRT sequencing

PacBio sequencing is based on single-molecule real-time (SMRT) sequencing, which takes advantage of the natural process of DNA replication. A single DNA polymerase is fixed on the bottom of a very tiny well called zero-mode waveguide (ZMW) inside the SMRT cell. During the reaction, the enzyme incorporates each of four nucleotide labeled with a different colored fluorophore into the complementary strand cleaving off the fluorescent dye; the camera inside the machine can capture the light signal in real-time [6]. PacBio sequencer are: RSII, Sequel I and Sequel II that respectively produce approximately 1Gb, 10Gb, and 100Gb per flowcell (**Figure 1**).

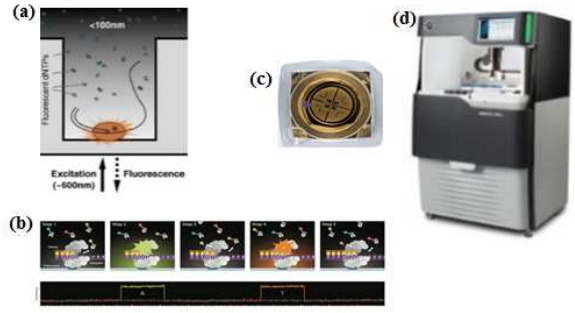


Figure 1: (a) ZMW with immobilized polymerase; (b) Nucleotide base is incorporated by the polymerase and the fluorescent dye is cleaved-off from the nucleotide; (c) A single SMRT flowcell (d) PacBio Sequel

With PacBio sequencing, the sample preparation consists of fragmenting the DNA into desired lengths (usually 10, 20 or 30Kb), blunting the ends, and ligating hairpin adaptors that create a circularized template [7] (**Figure 2**). Different DNA amount are required according to the library size, with an average of 5µg for the 20Kb library, DNA (**Table 2**) and the read length obtained is on average correspondent to the length of DNA fragment size.

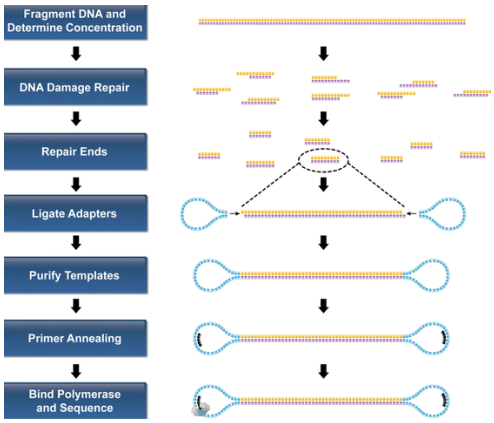


Figure 2: PacBio library preparation (SMRTbell Template Preparation Workflow).

ii) Oxford Nanopore sequencing

Oxford Nanopore technology (ONT) exploits nanopore sequencing and provides the possibility to analyze reads up to 2Mb. In nanopore sequencing, a biological nanopore (e.g. α-haemolysin) is inserted into an artificial membrane (**Figure 3**) and a voltage is applied across the membrane, resulting in ion flow through the pore. A protein motor, attached to nucleic acid, controls the translocation of the

DNA strand through the nanopore; translocation of DNA alters ion flux and thus DNA sequence is inferred from the alterations of current registered through the pore [8]. The MinION was the first device provided by ONT— it is a portable and low-cost device — and it was followed by larger devices: GridION, PromethION P24 and P48. These sequencing platforms can utilize simultaneously 1, 5, 24 or 48 flowcells respectively, producing up to 5,2 Tb.

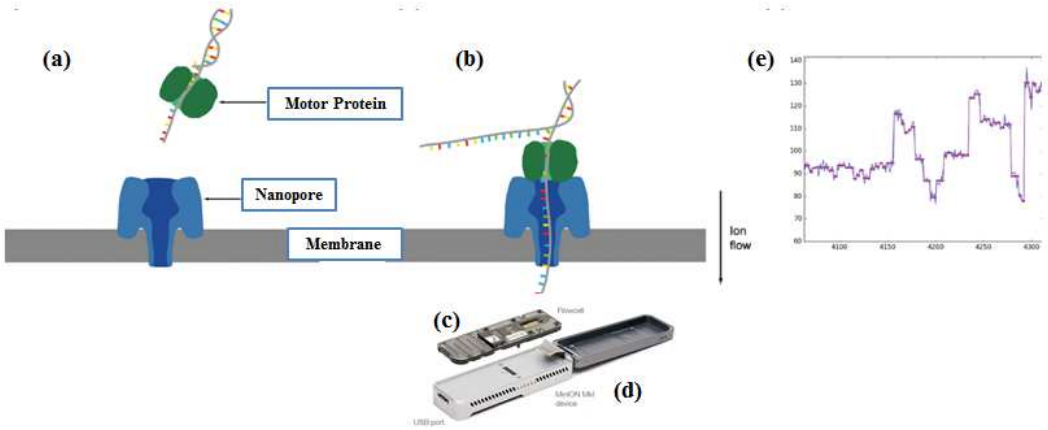


Figure 3: (a-b) Biological nanopore inserted into the electrically-polarized membrane. Library DNA molecules with motor protein, adaptors and aliphatic tethers (not shown) translocating though the nanopore; (c) Flowcell; (d) MinION device; (e) electric signal detection.

Two types of approaches can be used to generate long reads using nanopore sequencing: (i) mechanical DNA fragmentation followed by adapter ligation (1D Ligation kit), or (ii) tagmentation-based library construction using transposase (Rapid kit) (**Figure 4**). The second approach allows to produce the longest reads (so far the longest reported was 2,2Mb [9]). Increase in read length is achieved by unbalancing the ratio between DNA quantity and transposase in favor of DNA. Generating such ultra-long reads, however, greatly reduces the output of the flow cell [10]. In contrast, higher output can be obtained with the ligation approach, but at the expense of DNA length; in fact the longest read reported with this library type was 882Kb [11]. To maximize the output using the ligation approach, it is preferable to have fragments of homogeneous size, obtained through mechanical shearing [12].

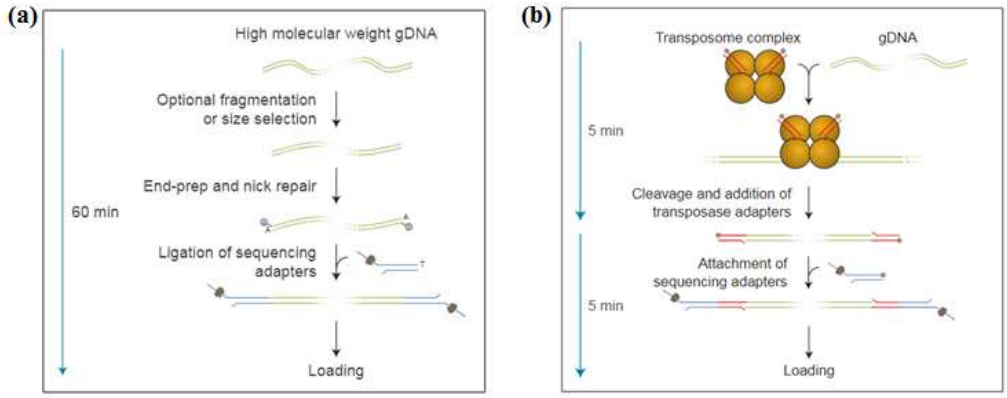


Figure 4: Oxford Nanopore Library Workflow by (a) Ligation sequencing kit (b) Rapid sequencing kit

Genome mapping approaches

Next to 3GS, also physical mapping has improved its high-throughput solution with the birth of Next Generation Mapping. Next Generation Mapping technologies allow the large-scale analysis of DNA structure by physical mapping, in a high-throughput fashion. The current leader in physical genome mapping is BioNano Genomics, generating optical maps, and other companies are emerging such as Nabsys, that provides the possibility to generate electronic maps.

Bionano optical mapping

Optical mapping (OM) was already invented at the end of the last century, but recent automation of this process has led to the development of commercial high-throughput platforms, such as the Irys or Saphyr system released by BioNano Genomics [13], generating respectively <10-50<Gb and 1300Gb per run. Bionano OM generates physical maps by imaging extremely long genomic molecules typically from 150 Kb and up to Megabases in length.

Bionano OM is built by locating the recognition site of a chosen restriction enzyme along the unknown DNA molecule and providing absolute distance information between different sites. Preparation of DNA for OM with Bionano system can be done either with the Nick-Label-Repair-Staining (NLRS) or with the Direct-Label-Staining (DLS). In the NLRS, the DNA is initially nicked on

single strand using modified restriction enzymes and subsequently labeled at these sites with green-fluorescent nucleotide using a polymerase reaction. The DNA backbone is stained with a blue intercalating dye. Labelled DNA molecules are then linearized inside a chip featured with micro- and nanochannels and imaged thanks to a fluorescence microscope (**Figure 5**). The spectrum of resulting DNA marks serves as a unique "fingerprint" or "barcode" for that sequence.

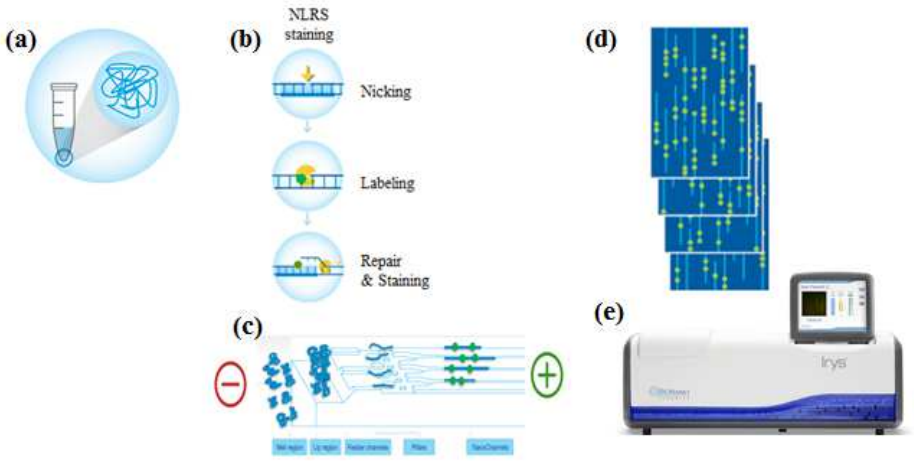


Figure 5: (a) label specific sequences across the entire genome; (b) NLRS staining workflow; (c) analysis of HMW DNA on an Irys Chip; (d) DNA molecules image acquisition (e) Bionano Irys Platform

Bionano Genomics provides different kit for the extraction of Ultra-High Molecular Weight (HMW) DNA from animal tissues/cells or plants, but none specific for algae. The protocols for plants are multiple, with a major distinction when considering tissue (cell wall) disruption, that can be either performed by blending on previously fixed tissue or by liquid nitrogen without prior fixation.

Long-read sequencing and genome mapping require pure and High Molecular Weight (HMW) DNA

The current yield and high cost per base of these long range technologies remains a barrier for most genomic projects, especially those targeting large genomes. Furthermore, any irreversible DNA damage present in the input material leads to poor quality DNA sequencing. Thus, high integrity, purity, and concentration of genomic DNA are imperative to obtaining reads with long lengths. The possibility

to analyze long molecules, without short fragments, allows to utilize the flow cell to the maximum, without wasting sequencing resources.

Each technology has different requirements of DNA input, in terms of quantity, DNA purity, and DNA length as shown in **Table 2**.

In particular, the *purity* of total DNA must be evaluated not only based on the absorbance ratio (260/230 and 260/280) but also based on the consistency between spectrophotometer- (Nanodrop) and fluorescence-based (Qubit) quantification, that should be very close to 1, meaning a good purity of the sample[14]. Spectrophotometer absorbance can be indeed influenced by the presence of contaminating compounds that can absorb light at 260nm (e.g. phenols, proteins) and may affect the quality of DNA. In contrast, Qubit uses a fluorescent dye that binds specifically dsDNA providing a more accurate measurement of the molecule of interest. Therefore consistency between spectrophotometer- and fluorescence-based quantification indicates a good purity of DNA samples [14].

In terms of DNA *length*, 3GS and Bionano OM require genomic DNA higher than $10^{\wedge 4}$bp and $10^{\wedge 5}$bp respectively, for whole genome sequencing.

	LIBRARY PREPARATION	DNA AMOUNT ng/ul	Total ng	PURITY 260/280	260/230	NANODROP/ QUBIT ratio	LENGTH (bp)
PacBio	≥ 10Kb	10	>1000	~1,8	<2-2,2<	Closest to 1	>10^4
	>15Kb	<200-300<	>2000				
	>30Kb	<25-50<	>3000				
Oxford Nanopore	Rapid kit	50	>400				
	1D Ligation kit	20	>1000				
Bionano	NLRS	<35-200<	300	-	-	-	10^5

Table 2: Optimal DNA features to perform Third Generation Sequencing (PacBio; Oxford Nanopore) and Bionano mapping

Methods for HMW DNA extraction

A wide range of methods is available for DNA extraction from different tissues and species, however only a subset of them is suitable for HMW DNA extraction. Isolation of HMW DNA is relatively routine for species such as bacteria and mammalian cell, while, for plants or algae, the extraction of high-quality and HMW DNA presents a substantial obstacle for long-molecule analysis [15] [16].

HMW DNA extraction methods can be grouped as follow: organic, adsorption and plug-based extractions. Even if in different fashion, all methods include these fundamental steps to extract DNA: (1) disruption of cell wall and membranes to liberate cellular components, (2) inactivation of DNA -degrading enzymes (DNases), (3) separation of nucleic acids from other cellular components, (4) DNA extraction/precipitation. These steps must be carefully adjusted according to the tissues and species of interest as their modulation can determine quite profound differences in the resulting DNA prep.

i) Organic extraction method

The purification of DNA by organic extraction is one of the basic procedures in molecular biology. It is based on cell lysis buffer using detergent (CTAB, SDS, Triton-X, Tween-20), designed to lyse the outer cell membrane and the nuclear membrane. Detergents break cell membranes by attaching to the lipids and proteins present in the membranes. In addition, in combination with chelating agents, such as EDTA or proteolytic enzymes such as proteinase K, detergents protect DNA from degradation by inactivating nucleases. After cell lysis, in the organic extraction method, a mixture of phenol:chloroform:isoamyl alcohol is added to lysate to promote the partitioning of lipids and cellular debris into the organic phase, leaving isolated DNA in the aqueous phase. The different phases are separated by centrifugation obtaining: an aqueous phase on top with DNA, a denser organic phase rich of proteins at the bottom, and a white layer of precipitate of proteins in the interphase. This process can be repeated a few times to ensure the complete removal of proteins, before precipitating the DNA with ethanol. The organic extraction method generates DNA of 100-150Kb in size without damages due to physical forces and it is also very effective at removing nucleases. Despite this method remains one of the most reliable and efficient, it is

also quite time-consuming, it uses hazardous chemicals, and involves a lot of hands-on effort and multiple tube transfers which can introduces increased opportunities for contamination, sample mishandling and losing of material.

ii) Adsorption extraction method

Adsorption method is based on chemistry developed in 1990s [17] using either silica or anion exchange resins to reversibly bind DNA, that allow to separate it from cellular proteins, lipids and polysaccharides. The most common types of DNA extraction kits that anyone will encounter in a laboratory are based on spin columns or gravity flow columns. Spin columns are so called because reagents are added to the top of the tube and then forced through the binding matrix by spinning in a centrifuge. This approach is very fast and easy-to-performed; however the DNA is sheared during binding and elution due to the large physical forces experienced during centrifugation. The average size of DNA fragments obtained is around 60Kb. On the other hand, the gravity flow columns employ the same binding technology as spin columns. The difference is that they are not placed in the centrifuge but left in a rack allowing the lysate/wash solutions to drip through them by gravity. These can be used to recover DNA with an average size of 100-200 Kb due to the gentle handling of the sample, but are more time consuming. DNA is eluted from the column in a large volume then precipitated with isopropanol to concentrate it, producing a DNA higher in size and yield compared to the spin columns. Gravity flow columns are especially useful to isolate large quantities of DNA with a good ratio of purity.

iii) Agarose plug extraction method

To date, nuclei embedding method in low melting-point agarose in the form of plugs or microbeads is the only extraction method which provides chromosome-sized DNA or DNA fragments up to Mega-base pairs (Mb) in length [18]. In this method, nuclei are usually first isolated by physically grinding the source tissues or by blending it. The cytoplast organelle genomes and metabolites are removed by centrifugation and sequential washing. The isolated nuclei are then embedded in low-melting-point agarose in the form of small blocks, preserving the integrity

of DNA[18]. The cons of this Mega-base extraction method is that it is time consuming (the entire protocol takes ~3 days) and costly. In addition the DNA extracted from agarose matrix is usually not enough pure for sequencing technologies such as 3GS.

Challenges in HMW DNA extraction from plants and algae

Plant and algae are subjected to intense research because of their importance for the agro-alimentary sector, as well as for their exploitation in the industrial sector, as source of important metabolites or for biofuel production. One of the major bottlenecks in the study of plants and algae is the need to break cell wall to get access to their cellular content. This issue implies difficulties also for the extraction of DNA from these organisms, especially when considering that the cell wall composition greatly vary between -but also within- species, as well as between different developmental stages[19]. Different approaches to disrupt the plants and algae cell walls with the aim to extract DNA are listed in the table below, along with their advantages (Pros) and disadvantages (Cons).

CELL WALL DISRUPTION			
Classification	**Method**	**Pros**	**Cons**
Enzymatic Lysis (only algae)	• Cellulase • Chitinase • Pectinase..	no mechanical DNA damage	• Knowledge about cell-wall composition is needed • Not always as efficient as mechanical brute force
Chemical Disruption (only algae)	• Acetone • NaOH • HCl…	no mechanical DNA damage	• Not always as efficient as mechanical brute force
Mechanical Disruption	• Solid (Liquid Nitrogen grinding, Bead Beating, High speed homogenization) • Liquid (high pressure, ultrasonication) • Other (microwave, Pulse elctrical field)	good yield	• Might shear DNA • Cell wall residuals must be thoroughly removed

Table 3: Solution adopted in literature to disrupt plants and algae cell wall. Pro e Cons for HMW DNA extraction are indicated.

As the cell wall, also the composition and concentration of polysaccharides and polyphenolic compounds can vary greatly between species, developmental stages and distributions, making difficult to develop a ubiquitously effective extraction technique [20] [16] [21]. These compounds are released during DNA extraction procedures, often persist despite purification steps [15] and may strongly affect the final DNA purity. The incomplete removal of polysaccharides and polyphenols can inhibit downstream laboratory techniques reducing the activity of enzymes, including restriction enzymes, polymerases, and ligases [16][22]. Moreover, during DNA extractions, a substantial amount of proteins and RNA may get precipitated with DNA. **Table 4** lists the principal issues [15] [16] for HMW DNA extraction from plants/algae and their impact on DNA quality.

ISSUE	IMPACT ON DNA
Tough cell wall requires mechanical disruption	DNA damage (integrity, size)
High polyphenols (Flavonoids, terpenoids, tannis…)	Low purity (oxidation)
High polysaccharides (starch grains, cellulose, glycan…)	Low purity
Organic compounds (lignin, terpens, alkaloids,…)	Low purity and DNA damage (ROS, RNS..)
Pigments (Proteinaceous, Carotenoids, Flavonoids,..)	Low purity and DNA damage (ROS, RNS..)
DNAse rich	DNA degradation (integrity, size) and low purity

Table 4: Issues in DNA extraction from plants and algae and their impact on DNA.

Methods to extract high-quality and HMW DNA from plants and algae

In the preceding two to three decades, researchers have developed numerous protocols and procedures to isolate high quality DNA from a variety of plant species, which can be used depending on the downstream applications [23][24][11][25]. **Table 5** lists some strategies to rise above different issues related to plants and algae DNA extraction.

ISSUE	APPROACH (USE OF…)
Tough cell wall	Grinding tissue on ice with LN_2
Lipid membrane	Surfactant: • CTAB • SDS
High polyphenols	• PVP • BME • DTT • sodium sulfite, • ascorbic acid
High polysaccharides	• High salt concentration+cationic detergent
DNAse	• EDTA • EGTA • Phenol:Chloroform:Isoamylalcohol (24:25:1)
Proteins and RNA	• SDS • proteinase K • DTT • BME
Non-nuclear DNA	• Osmoprotectant (glucose,sucrose) • Membrane stabilizer(spermine,spermidine) • Non ionic detergent (Triton X100) • Density gradient

Table 5: Solutions adopted in literature to overcome the issues of high-quality DNA extraction from plants and algae

Cell wall disruption: the most widely-used approach to break the cell wall is the grinding of starting tissue (or biomass for algae) with pestle and mortal in Liquid Nitrogen (LN_2), making sure to keep the working area cold. The preservation of the tissue on ice, adding repetitively LN_2 during cell wall disruption, prevents the warming up of the sample. This blocks the binding of DNA to polysaccharides, and the generation of short DNA fragments and it is therefore preferred for the extraction of HMW DNA. This approach was the one selected to conduct the work in this thesis.

Cell lysis and cell components solubilization. Once the cell wall is disrupted, plant or alga cells are like any other cell whose membrane can be lysed using a detergent. At this aim, one of the most utilized protocols for plant and algae DNA extraction was developed by Murray & Thompson in 1980 and it employs the detergent **CTAB**. CTAB became the most-commonly utilized detergent for plants

and algae [26] thanks to its capability to dissolve both polar as well as non-polar compounds present in the cell wall; in addition it can strongly bind the polysaccharides and protein complexing with them, thus facilitating the separation of carbohydrates from DNA [27]. The CTAB-based method is widely used for the extraction of HMW DNA. The specific CTAB-based protocol utilized in this work has been used for example for the generation of 3GS from many plants tissue, such as *Secale*, *Festuca*, *Malus*, *Oryza*, and *Arabidopsis*[28].

In case of protein rich tissues [29], **SDS buffer** with proteinase K may be required for plants and algae DNA isolation. The SDS-based protocol utilized in this study was selected from Mayjonade et al. (2016)[30] and Schalamun et al. (2019)[12]. It was demonstrated to be very efficient in the extraction of HMW genomic DNA from *Sunflowers* and *Eucalyptus pauciflora* for Oxford Nanopore long-read sequencing that contain high polysaccharides and tannins. SDS is also very efficient in neutralizing and inactivating proteins, especially endonucleases, that degrade DNA [31]. Of note, for *Eucalyptus pauciflora*, Schalamun et al. (2019) have shown that this approach produces cleaner DNA than CTAB-based extraction [12].

Recent publications frequently report the use of **gravity columns** for the extraction of HMW from plants. Genomic tip-100 (G-tip) from Qiagen is one of the most widely use kit as the procedure is very gentle and results in negligible DNA shearing; moreover anion-exchange resins allow a more efficient purification of DNA as compared to organic extraction methods. Genomic Tip-100 method allowed to obtain high quality HMW DNA from plants and algae (i.e. *Vitis riparia* grape [32], *Caulerpa lentillifera* alga [33]) sequenced by PacBio platform.

Cationic detergents (such as CTAB), in combination with **high concentration of salts (KCl, NaCl),** maintain ionic strength of the solution, that has been proven to be beneficial in DNA isolation from polysaccharide-rich tissue such as plants and algae [23]. Otherwise, the use of divalent cations (such as Mg^{2+}, Ca^{2+}) with concentration higher than 10mM must be avoided, because it could lead to DNA degradation. To prevent DNA degradation and also to block DNase activity, it is

important the addition of **chelating agent (EDTA, EGTA)** during plant or alga DNA extraction [29]. The polyphenols and alkaloids can be removed by using compounds as **polyvinylpyrrolidone (PVP) or ascorbic acid**, which inhibit any modification of proteins or hindrance in spectrophotometric determinations of protein content caused by those phytochemicals[27][34]. It is good practice for plants and algae DNA extraction, to combine the use of the above mentioned compounds with **antioxidants like β-mercaptoethanol (BME) or 1,4dithiothreitol (DTT)** to maintain cysteine residues in a reduced form and protect DNA from oxidative damage too. The denaturation of proteins is performed by all these mentioned redox reagents in combination with organic solvents like phenol and chloroform.

Methods for nuclei isolation

When performing a *de-novo* genome assembly project, DNA is frequently extracted from nuclei, obtained after removal of cytosolic content along with organelles (mitochondria and chloroplasts, if plant/algae) and their genomes. Even if mitochondria and chloroplasts have usually a genome smaller in size than the nuclear genome, they are indeed present in multiple copies, and thus the fraction of plastidial DNA in total DNA prep can be consistent. Extraction of DNA from nuclei can therefore concentrate all sequencing on the nuclear genome, thus allowing to obtain more focused data and better genome-assembly. This is of utmost importance when considering the relatively high costs of long reads technologies: using DNA extracted from nuclei is therefore also economically convenient for these approaches. A second advantage of isolating nuclei is the possibility to exclude contaminants from the DNA prep. By removing all the cytosolic content prior to nuclei lysis, one can avoid that a large fraction of metabolites and cell wall/cytosol components get in direct contact with nuclear DNA and affect its final purity.

In general, for nuclei isolation the presence of **osmo-protectants** (like hexylene glycol, glucose, sucrose) and polyamines (like spermine, spermidine) in the

homogenization buffer stabilize the nuclear membrane, facilitating the isolation of intact nuclei by centrifugation. Addition of **Triton X-100**, specifically, lyses chloroplasts and mitochondria, leaving the nuclei intact [18]. In addition, centrifugation of cell lysates over a **density gradient** (performed using solutions at pre-defined density like: percoll, cesium chloride, sucrose, and zinc sulfate) is considered one of the most efficient method of separating pure nuclei according to their density[36]. Last but not least, the extraction buffers with a **low pH** have been reported to improve nuclear yield with minimal cytoplasmic contamination and nuclei clumping formation from plant protoplasts [18].
Despite the fact that organisms as plants and algae have, in addition to nuclear genome, two plastidial genomes (chloroplast and mitochondrion) with a high variability on copy-number and size, in literature only a limited amount of methods to isolate nuclei from this kingdom is reported. The methods mostly used to isolate nuclear DNA from plants and algae are based on the following lysis buffers: NIBTM, MEB, HONDA, BOLGER, and SEBM.

The NIBTM-based method has been published by Zhang et al. (2012) and it represents the antecedent of Bionano method for extraction of nuclear DNA with mega-bases in length. NIBTM buffer contains very high concentration of PVP that allows efficient removal polysaccharides from plant [18]. In its original form, this method includes a gradient step to improve the nuclei purity, at cost of DNA yield though.

MEB buffer was identified to be very useful in nuclei isolation for plants rich in secondary metabolites[36](distinguished roughly into three classes of chemical compounds, namely: alkaloids, phenolic compounds and terpenes) [29]. Relevant components are Hexylene Glycol and Pipes KOH that allow respectively to stabilize and purify the nuclei [36]. In addition, the acid pH=5 helps to inhibit the aggregation of nuclei to protoplasts [35], thus allowing a more efficient nuclei isolation.

HONDA buffer [37]was cited in the literature to isolate nuclei from an algae genus, Chlamydomonas. The high concentration of sugars, as 5% Dextran-40 and 2,5% Ficoll, helps to break the cell wall by creating an osmotic pressure; in this

manner, the cell bursts and cell contents are released by the action of proteolytic enzymes present with the sugars in the buffer.

BOLGER buffer [38] was designed for recalcitrant plant species (containing elevated concentrations of polysaccharide and polyphenolic compounds) for Oxford Nanopore long-read sequencing. The composition is similar to MEB buffer, but presents pH=6 and the double percentage of PVP, so it is expected to remove more efficiently secondary metabolites.

SEBM buffer [39] was used to extract nuclear DNA from recalcitrant plant extraction in our lab and it worked very well in the isolation of nuclei from eggplant [40] as well as to yield suitable DNA for Oxford Nanopore sequencing (**Appendix Table A1**). It contains high concentration of Triton X-100 (10%), a non-ionic detergent that is useful to disrupt chloroplasts, to release nuclei and most importantly to prevent that cytoplasmic debris will aggregate and bind the nuclear surface [41].

Still, whatever is the buffer employed, it is important to prevent high concentrations of detergents or prolonged exposure to it, because this may also disrupt the nuclear membrane [21]. HMW DNA extraction must be performed immediately after the isolation of nuclei. The immediate DNA extraction prevents damage to DNA in terms of size, while longer incubation times can lead to nuclear membrane damage and aggregation nuclei [42]. Considering these points, HMW DNA extraction from nuclei implies the use of **very long protocols**. While total HMW DNA extraction is usually performed in approximately 5 hours (except for agarose plug method), nuclear HMW DNA extractions need ~12 hours of continuous work.

Methods for the purification of extracted DNA

Sometimes, after DNA extraction, DNA still carries contaminants such as polysaccharide, phenols, proteins and RNA. Post-extraction purification steps can

thus improve the quality of DNA and they can be grouped in: (i) alcohol-based precipitation, (ii) beads-based absorption and (iii) gel-based extraction methods.

Alcohol based purification is based on "old school" protocols [40][44]. In this procedure Phenol:Chloroform:Isoamyl alcohol precipitation is used to completely remove the phenol compounds derived by the DNA extraction, subsequently a second step using a low-ethanol mixture (0,3X v/v of ethanol 99,99%) precipitates the excess of polysaccharides and a third steps uses high ethanol volume precipitation to recover gDNA. The number of precipitations steps to purify the DNA depends on its purity.

Gel-based extraction can be performed for example using BluePippin, an automated preparative gel electrophoresis system, capable of performing a very tight-selection of DNA fragments of a pre-defined size. In addition to being able to cut-off short DNA fragments, BluePippin can also very efficient clean the treated DNA as result of gel-based extraction[45].

Beads-based purification is a fast purification procedure, which utilizes an optimized buffer to selectively bind only DNA fragments to paramagnetic beads, while allows removing contaminants or other components that remain in the supernatant (e.g. metabolites). This clean-up is based on balancing the content between PEG and salt (NaCl) to selectively bind nucleic acids by type and size that allows also a size selection, based on the controlled beads to sample volume ratio.

A second reason why DNA can be purified after extraction is to remove short fragments that will produce sequencing reads with sub-optimal length. At this aim, three are the major approaches utilized in literature: (i) *Blue-pippin size-selection* described above [46], (ii) *Beads-based purification* as described above [47] [48], (iii) *Short Reads Eliminator (SRE) kit*, a commercial solution that allows the size-selective precipitation and recovery of large DNA fragments (complete elimination up to 10Kb, and progressive elimination up to 25Kb) [49].

The case study of *Haematococcus pluvialis*

Haematococcus pluvialis (*H.pluvialis*) is a unicellular green alga intensively studied and commercialized being the best natural source of astaxanthin, a "super-antioxidant"[43]. Despite its importance in different fields, the available genome of *H.pluvialis* is highly repeated and still very fragmented (Transposable elements content=32.2% and contig N50=8,2Kb) [50]. Although being necessary for biotechnological applications, a high quality genome is therefore still missing and long-molecules technologies can provide suitable data at this aim. However, no publication reported protocols to obtain HMW DNA from *H.pluvialis* to date and a quite limited number of studies investigated its genomic content.

AIM OF THE THESIS

The aim of this work was the optimization of methods for the extraction of highly pure and (Ultra)HMW DNA from the microalgae *Haematococcus pluvialis* (*H.pluvialis*), suitable for 3GS (PacBio and ONT) and Next Generation Mapping (Bionano OM), namely the technologies that have been selected to reconstruct the genome of *H.pluvialis de-novo.*

To achieve this aim, an extensive benchmarking of DNA and nuclei isolation methods was conducted, also in combination with post-extraction DNA purification methods. DNA preps extracted with the different protocols, or their combinations were assessed in terms of yield, length and purity and the best samples were also analyzed by Third Generation sequencing and Optical Mapping.

MATERIALS & METHODS

H.pluvialis samples

H.pluvialis samples (frequently referred in the thesis also as "biomass") used for DNA extractions were kindly supplied by Sole Lab directed by Prof. Ballottari at the Department of Biotechnology, University of Verona. HMW DNA extractions were performed in duplicate or triplicate starting from 43*10^7 cells. Algal cells, suspended in TAPS buffer, were pelleted at 3000g in a swing-bucket rotor centrifuge at 4°C and snap-frozen in liquid nitrogen before storing at -80°C.
When required, the biomass was fixed by adding 1% of fresh formaldehyde in PBS and the solution was mixed at 140 rpm for 20 minutes. The quenching was performed adding 1,25mM Glycine and mixing 15 minutes at Room Temperature. The supernatant was removed and the pellet was washed with PBS centrifuging at 1000g for 2 minutes twice. The last centrifuge was performed without any buffer at 3500g for 2 minutes to remove the excess of supernatant.

H.pluvialis cell wall disruption

Cell wall disruption was performed by grinding frozen biomass in the presence of liquid nitrogen as described by Jagielski et al. (2017) [28]. Briefly, frozen pellet was ground for 30 minutes to a fine powder by using an autoclaved, pre-chilled mortar and pestle kept on ice all the time. The powdered sample was suspended in different cell lysis buffers according to the downstream protocol applied.

Nuclei isolation methods

Nuclei isolation based on MEB buffer. Method based on MEB buffer was performed according to Lutz et al. (2011) [36] with minor revision. Powder of ground cells was transferred into 45 ml freshly prepared MEB buffer (1M 2-methyl-2,4-pentanediol (MPD), 10mM PIPES-KOH, 10mM $MgCl_2$, 2%PVP-10, 10mM sodium metabisulfite, 0.5%Sodium diethyldithiocarbamate, 6mM EGTA, 200mML-lysine-HCl, and 5mM β-mercaptoethanol) just before use; pH 5.0). The homogenate was filtered through first a 100µm cell strainer and then through a

40µm cell strainer. Triton X-100 was added to a final concentration of 0.5%, placed on ice for 30minutes and centrifuged at 800g for 20 minutes at 4°C. The pellet was re-suspended in 45ml of MPDB (0.5M 2-methyl-2,4-pentanediol, 10mM PIPES-KOH, 10mM $MgCl_2$, 0.5%Triton X-100, 10mM Sodium metabisulfite, 5mM β-mercaptoethanol added just before use, pH 7.0) and centrifuged at 800g for 20 minutes at 4°C. This last step was repeated up to 4 times carrying on with only the whiter layer of the pellet, representing the nuclei fraction. The green pellet layer was discarded. The white pellet was layered on top of a 37.5% Percoll bed (20ml of 37.5% Percoll (7.5ml percoll + 12.5ml MPDB) in a 15ml glass centrifuge tube. The gradient was centrifuged at 650g for 1 hour after which the nuclear pellet was on the bottom of the tube. The white part of the pellet was recovered and re-suspended in 10ml MPDB buffer, centrifuged at 2500g for 10minutes at 4°C twice. The nuclei pellet was re-suspended in TE buffer and DNA was isolated using one of the HMW DNA extraction protocols described below.

Nuclei isolation based on BOLGER buffer. Method based on Bolger buffer was performed according to according to Bolger et al. (2014)[38] with minor revisions to adapt this mothod to algae. Powder of ground cells were transferred into 45ml freshly prepared nuclei isolation buffer (1M Hexylen Glycol, 10mM PIPES pH6, 10mM MgCl2, 10mM Sodium metabisulfite, 6mM EGTA, 0,5% Sodium diethyldithiocarbamate, 4%PVP10, 200mM L-Lysine, 1mM DTT, 0,2%b-mercaptoethanol added just before use) adjusted to pH 6 with HCl. The homogenate was filtered through 2 cell strainer (100µm and 40µm) and it was incubated on ice for 30 minutes adding 0,5% of Triton X-100. Then the homogenate was centrifuge at 600g for 20 minutes at 4°C. The whiter part of the pellet was re-suspended in 45ml of nuclei wash buffer that is similar to nuclei isolation with some difference (0,5M Hexylen Glycol and the pH=7) [20] and centrifuged at 600g for 20 minutes at 4°C for 4 times to enrich for the whiter part of the pellet with no trace of green particulates. After that, the white layer of the pellet was re-suspended in 10ml Bolger buffer, centrifuged at 2500g for 10

minutes at 4°C twice. DNA was isolated from the nuclei preparation using one of the HMW DNA extraction protocols described below.

Nuclei isolation based on HONDA buffer. Method based on Honda buffer was performed according to (Luthe and Quatrano 1980) [37] with minor revision. Powder of ground cells was transferred into 45ml fresh Honda extraction buffer (25mM Tris, 10mM MgCl2, 0,44M sucrose, 10mM BME, 5%Dextran 40, 1%Triton X-100, 2,50% Ficoll) and filtered through 2 cell strainers (100µm and 40µm), as described before. The filtrate was kept 30minutes on ice and then it was centrifuged at 4,500g for 5 minutes at 4°C and the supernatant was discarded. The pellet was re-suspended with a soaked paint brush and bring to 45ml of volume with Honda buffer to perform 4 spin at 600g for 20 minutes each at 4°C. The pellet was re-suspended in 10ml Honda buffer, centrifuged at 2500g for 10 minutes at 4°C twice. DNA was isolated using one of the HMW DNA extraction protocols described below.

Nuclei isolation based on NIBTM buffer. Method based on NIBTM buffer was performed with and without gradient according to an internal Bionano protocol with minor revisions. Powder of ground cells were transferred into 45ml fresh nuclei isolation buffer (10mM Tris pH8, 10mM EDTA pH8, 500mM Sucrose, 80mM KCl, 8% (w/v) PVP-10, 100mM Spermine, 100mM Spermidine) adjusted to pH9. Add 0,5% of Triton-X and 0,2% BME just before starting the procedure. The homogenate was filtered through 2 cell strainers (100µm and 40µm) and centrifuged at 3000g for 20 minutes at 4°C. The pellet was re-suspended with a soaked paint brush and bring to 45ml of volume with NIBTM, centrifuged at 60g for 2min to remove debris. The supernatant was filtered again through a 40µm cell strainer and centrifuge at 3000g for 20 minutes to pellet nuclei. These steps were performed up to 4 times trying to go ahead only with the white layer of the pellet. The pellet was re-suspended in 10ml NIBTM buffer, centrifuged at 2500g for 10minutes at 4°C twice. DNA was isolated using one of the HMW DNA extraction protocols described below.

Nuclei isolation based on SEBM buffer. Method based on SEBM buffer was performed according to Carrier et al. (2011) [39]. The manufacture instructions were strictly followed a part for the grinding step that is performed as previously described. Powder of ground cells was transferred into 45ml of SEBM buffer (500mM Sucrose, 1mM Spermina, 4mM Spermine Tetrahydrochloride, 0,13% Carbamic Acid, 0,25% PVP40, 0,2% BME), incubated on ice for 30min and filtered through 2 cell strainers (100µm and 40µm) it was incubated on ice for 30 mins adding 10% of Triton X-100. The homogenate was centrifuged at 600g for 9minutes at 4°C. The supernatant was discarded and the whiter layer of the pellet was transferred into a new tube with 20ml of SEBM buffer and it was homogenate mixing gently. The content of the new tube was filtered through a 40µm cell strainer. The centrifugation, the homogenization and the filtration steps were repeated up to 4times. DNA was isolated using one of the HMW DNA extraction protocols described below.

Nuclei isolation based on BIONANO buffer and Bionano protocol. The IrysPrep® Plant Tissue DNA Isolation kit provided by Bionano (BioNano Genomics, San Diego, CA) was used following Bionano Liquid Nitrogen Grinding protocol.
Powder of ground cells were transferred into 45ml fresh nuclei Homogenization buffer (Bionano) that was supplemented with 0,2%BME, 100mM spermine, 100mM spermidine, 7.5mL IrysPrep Triton X-100. Subsequently the nuclei were cleaned via filtration of the homogenate through 100µm and 40µm cell strainers and centrifuged. The nuclei collection spinning was performed at 3500g for 20 minutes at 4°C. The supernatant was discarded and the pellet was re-suspended with a small paintbrush pre-soaked in ice cold Bionano Buffer. Once re-suspended, 3mL ice-cold Bionano Buffer was added and swirled gently by hand for 30 seconds while the tube is touching ice. The suspended solution was brought to 40mL of volume and a centrifugation at 60g for 2 minutes at 4°C was performed to remove intact cells, cell debris, unbroken tissue, and other residues. The supernatant was collected through a 40µm cell strainer on top of a new

chilled 50 mL tube. The centrifugation steps at 3500g and 60g were performed twice. A density gradient purification based on two solutions at different density (provided by Bionano Genomics) was performed to separate nuclei from cell debris. Only the white pellet was transferred on top of the gradient solution. The tube was the centrifuged using a swing-bucket rotor at 4500g for 45 minutes at 4°C. Cytoplasmic proteins remain on top of the lysis layer. Nuclei continue to pass through the nuclei wash layer and get banded on top of the floating layer. It is important to stratify the gradient immediately before the sample loading to avoid the breaking of the interface. After Bionano nuclei isolation workflow, DNA was isolated using plug-based extraction described below.

HMW DNA extraction

Organic-based DNA extractions

CTAB-based DNA extraction. Organic HMW extraction based on CTAB (Cetyl TrimethylAmmonium Bromide) was performed using DNA extraction methodology of Doyle, J. J. & Doyle, J. L. et al. (1987) [24] with minor adjustments. Powder of grinded cells was transferred into 500µl of pre-warmed CTAB isolation buffer (2% CTAB (Sigma, Saint Louis, USA), 1.4M NaCl, 100mM Tris pH 8.0, 20mM EDTA supplemented with 1%PVP40 and 0,2% of fresh fresh β-mercaptoethanol, and incubated at 65°C for 30minutes. DNA was extracted from total cell lysate with chloroform-isoamyl alcohol (Chl/IAA, 24:1) (Sigma, Saint Louis, USA) and precipitated with 0,6 volumes of isopropanol. The obtained pellet was washed with 70% EtOH, dried, and dissolved in 50µL TE buffer. Extracted DNA was treated with 80ug/ml RNAase A at 37°C for 20 minutes and subsequently purified using 1.8X v/v AMPureXP beads.

SDS-based DNA extraction. Powder of grinded cells was transferred into 600µL of pre-warmed SDS-lysis buffer (1%PVP40, 1% $Na_2S_2O_5$, 500mM NaCl, 100mM Tris pH 8.0, 50mM EDTA, 1,25% SDS) supplied with 4µL RNase A (100 mg/ml) (Thermo Fisher) and 0,2% of BME. The mixture was incubated at 50°C for 30

minutes in a water bath (Julabo). 1/3 of the lysis buffer volume of 5M Potassium Acetate was added to the sample, mixed by inverting the tube 20 times and incubated at 4°C for 10minutes. The precipitates, containing debris and proteins, were removed by centrifugation at 5000g for 10 minutes at 4°C. The supernatant was transferred to new tubes and 1:18 (v:v) of Serapure beads solution was added.

The Serapure were prepared using an adapted AMPureXP beads (Beckman Coulter, Brea, CA, USA) using the following protocol. The AMPureXP beads were washed 4 times with water and re-suspended in their initial buffer (20% PEG, 2,5M NaCl) to obtain the right composition of Serapure beads required by this method. The solution with 1:18 of Serapure beads was mixed by inverting the tube 20 times and incubating on a rotator for 10 minutes at room temperature. The tube was placed in a magnetic rack for 3 minutes (until the solution becomes clear) and washed twice with 70% of ethanol. The beads were dried for no longer than 1min before elution of the DNA in 50µl TE buffer.

Adsorption-based DNA extraction by gravity columns

Gravity flow columns were provided by QIAGEN Genomic-tips 100/G kit (Cat No./ID:9110243), and the extraction of DNA was performed according to manufacturer instruction following the protocol for DNA isolation from tissue.
Powder of grinded cells was transferred into 9,5ml of Buffer G2 (with 190µl of RNase A (10mg/ml) and incubated 3h at 50°C as suggested by Girollet N., et al (2019) [32]. The digested sample was centrifuged at 4500g for 15minutes to prevent clogging of the Qiagen Genomic tip 100 (G-tip). Then the debris-free sample was transferred into the column, previously equilibrated with 4ml of QBT buffer based on isopropanol. Genomic-tip designed to operate by gravity flow was left in a rack allowing the lysate/wash solutions to drip though by gravity. After several washing steps, based on QC buffer, the DNA was eluted from the column, then desalted and concentrated by precipitation adding 0,7 volume of isopropanol. The pellet was re-suspended 50 in TE buffer.

Plug-based DNA extraction

H.pluvialis nuclei pellet resuspended in in 60µL of ice-cold Density Gradient Buffer were embedded in low agarose (2%) in disposable plug molds according to Staňková et al. (2016) [51]. To obtain the optimal DNA concentration (35-200ng/µl) with the least amount of contaminants, the extracted nuclei were embedded at different dilutions. After agarose solidification, the nuclei embedded in plugs were treated with 200µl of Qiagen Proteinase K enzyme in 2,5ml of Lysis Buffer (BioNano Genomics) for 2hours at 50°C, and subsequently they were transferred in a fresh solution of Proteinase K and Lysis Buffer for an overnight incubation at 50°C. After that, 50µl of Qiagen RNaseA enzyme was added to the tube and incubated in Thermomixer for 1 hour at 37°C with intermittent mixing. Following these digestion steps, the plugs were washed 7 times in Tris-EDTA buffer for 15 minutes on a horizontal platform mixer with continuous mixing at 180 rpm. After that the plugs were scooped with a metal spatula and the excess liquid was drained from plugs before to melt them at 70°C for 2 minutes. Immediately later, the plugs were transferred at 43°C and the agarose was digested with 2µL of 0.5U/µL Agarase enzyme letting the samples in a Thermomixer for 45 minutes with intermittent mixing. Agarose was removed from DNA by dialysis: DNA drop were positioned on a 0,1µm dialysis membrane floating on the surface of the 15ml of TE Buffer. Extracted DNA was kept overnight on the bench at RT to allow DNA relaxation and homogenization.

Post-extraction DNA size selection and cleanup

Alcohol-based precipitation clean up

Alcohol precipitations were performed according to a combination of two methods Japelaghi, R et al. (2011) [52] and Healey, A. et al. (2014) [31] with minor revisions. The extracted DNA eluted in Tris 10mM was precipitated adding 5M NaCl, 0,5M EDTA and 400µl of Phenol:Chloroform:Isoamyl Alcohol (25:24:1), and pelleted by centrifuging at 15000rpm for 10 minutes. The aqueous layer was saved aside, and the organic phase was precipitated again to recover more DNA. An equal volume of Chloroform:Isoamyl Alcohol (24:1) was added to

recovered DNA and spin 10 minutes at 15000 rpm. The aqueous layer was removed, combined with 0,3X v/v ethanol (99,99%) and centrifuged for 15 min at 15000rpm. Polysaccharide pellet was discarded and the supernatant was transferred in a falcon tube where a 1,7X v/v ethanol was added. The tube was centrifuge 45 minutes at 5000 rpm. The gDNA pellet was washed twice with 70% of ethanol and re-suspended in 50µl of Tris 10mM pH8.

Beads-based size selection and clean up

Solid-phase reversible immobilization beads (SPRI) were used to purify genomic DNA. SPRI beads are uniform polystyrene and magnetite microspheres with a carboxyl coating. AMPureXP SPRI beads (Beckman Coulter, Brea, CA, USA) were used to clean up DNA extracted from *H.pluvialis*. 0,45X v/v of AMPureXP beads was added to the sample (**Appendix Figure A1**); let it incubate for 6 minutes on a rotator platform to mix it gently. After a brief (pulse) centrifugation step in a mini-centrifuge, the tube was placed in a magnetic field and the supernatant was removed. The beads were washed twice with 1 ml 70% ethanol. All the extra solution was removed making sure the beads were never completely dried. Finally the DNA was eluted from beads with 50µl Tris 10mM pH8.

Bluepippin-based size-selection and clean up

BluePippin system (Sage science) was used with 0.75% dye-free agarose cassettes with high-pass threshold range <15-20< Kb program. These high-pass cassette definitions are able to collect up to 165Kb fragment.
The broad range mode was selected and the DNA external marker S1 (3-10Kb Sage Science) was loaded into reference lane field. To select DNA over 20 Kb size-selection threshold was set at 16Kb (as suggested by Sage Science recommendation). The 'End base pair' values were not modified (50Kb). 4µg *H.pluvialis* DNA was loaded for each lane. After 6 hours run, for best recovery the samples were let in the elution module for 45minutes. The DNA recovered was purify from salt with 1X v/v AMPure XP (Beckman) and eluted in 25µl of TE buffer.

Short Reads Eliminator-based size selection

Short Reads Eliminator (SRE-SS100-101-01) kit was provided by Circulomics Inc. It is based on a size selective precipitation that depletes completely short fragments up to 10Kb and reduce progressively read up to 25Kb. 1X vol/vol SRE precipitation buffer was added to 9µg of HMW DNA sample, centrifuged at 10000g for 30 min at RT. The DNA pellet was washed twice with 70% of ethanol and re-suspended in 50µl of Tris10mM pH8 incubating at 50°C for 2 hours and 30minutes. This kit has been thoroughly tested on Oxford Nanopore MinION/GridION/PromethION.

G-TUBE-based DNA shearing

In selected experiment, HMW DNA was sheared prior to nanopore library preparation using g-TUBE™ (Covaris). G-TUBE is a single-use device that uses centrifugal force to push the DNA sample through a precisely manufactured orifice that determines DNA shearing of a specific size. To obtain fragment of 20Kb in size the DNA sample was centrifuged at 4200g for 1'30''. The tube was inverted and the same speed centrifugation was performed before recover the sheared sample. The size of the sheared DNA samples was verified by 2200 TapeStation (Agilent Technologies) using Genomic DNA ScreenTape analysis.

HMW DNA quality control

DNA quantification by Nanodrop spectrophotometer

DNA was quantified based on the Beer–Lambert law by measuring the absorbance at 260nm using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, Germany). The purity of DNA was assessed based on the absorbance readings at 230, 260, and 280 nm, and by calculating the 260:280 and 260:230 ratios, indicating proteins, phenols, glycogen contamination and salt, polyphenols, guanidine contamination respectively.

DNA quantification by Qubit fluorimetric assay

The concentration of genomic DNA was measured fluorometrically using dsDNA Broad Range Assay Kit in combination with Qubit® 4.0 (Thermo Fisher Scientific, Waltham, USA). For ultra-HMW DNA three quantifications from top-medium-bottom of the samples were performed after 10 minutes of sonication. In these cases, for each quantification 2ul DNA were pipetted using a positive displacement pipet.

Pulsed field gel electrophoresis (PFGE)

The size of DNA was evaluated using the CHEF Mapper electrophoresis system (Bio-Rad Laboratories, California). PFGE allows a more precise analysis of HMW DNA than traditional electrophoretic system as it uses the electrical field between spatially distinct pairs of electrodes. This technique results in the separation of DNA fragments up to Mb by their reorientation and movement at different speeds through the pores of an agarose gel.

700ng DNA was resolved by PFGE using a 1% agarose gel that was let solidify overnight at RT for 3hours. The electrophoresis chamber was filled by 2,2 Liters of Tris-Borate-EDTA (TBE) buffer 0,5X. According to the expected size of extracted DNA the run was set:

- From 250-2200 kb with a Two State Mode (24 hours) that consist of two field vectors, with each vector having the same voltage and duration but separated in direction by a 120°definable included angle, with an Initial Switch Time of 60seconds and a final switch time of 90 seconds.
- From 50-1000 kb with a Two State Mode (20 hours) that consist of two field vectors, with each vector having the same voltage and duration but separated in direction by a 120°definable included angle, with an Initial Switch Time of 35seconds and a final switch time of 90 seconds.
- From 5-450 kb with a Two State Mode (20 hours) that consist of two field vectors, with each vector having the same voltage and duration but

separated in direction by a 120°definable included angle, with an Initial Switch Time of 5seconds and a final switch time of 35 seconds.

Three λ DNA markers (either liquid or embedded in agarose) were chosen for PFGE run differing in their size range:

i) CHEF DNA Size Standard, Size range <8,3-48,5<kb (Biorad#1703707)
ii) MidRange PFG Marker- Size range <15-291<kb (NEB#N0342S)

iii) Lambda PFG Ladder- Size range <48.5-1,018<kb (NEB#N0341S)

iv) CHEF DNA Size Marker- Size range <225-2200<kb (Biorad#1703605)

After the run, the gel was stained for 30 minutes on Hula mixer at 80rpm in 400ml of 0.5X TBE buffer supplemented with 40µl Syber Gold (Thermo Fisher Scientific, USA). Subsequently the gel was washed for 30 minutes with fresh 0.5X TBE buffer. ChemiDoc Touch Imaging System with Image Lab Touch Software (Bio-Rad) was used for gel imaging.

DNA capillary electrophoresis by Tape Station

DNA integrity was verified on the 2200 TapeStation (Agilent Technologies). According to the manufacture's specifications, 11µl of reaction (10µl buffer + 1µl DNA) for both the ladder and the sample was set and the capillary gel electrophoresis was run on a Genomic ScreenTape (Agilent Technologies).

Handling and preservation of HMW DNA

For long-read sequencing applications, shearing of DNA is a significant issue as it limits the read lengths generated by sequencing. During extraction and storage of DNA the DNA polymers are susceptible to physical and chemical shearing [53]. HMW DNA was pipet exclusively with wide bore tips to avoid double strand breaks due to the hydrodynamic forces in moving fluids.

HMW DNA is very viscous and the preservation at high concentration helps to reduce shearing. After HMW DNA extraction, a little extra care was taken to ensure that DNA quality was maintained during storage. HMW DNA was re-suspended in elution buffer (EB; 10mM Tris-HCl pH 8.0) or Tris-EDTA buffer (TE; 10mM Tris-HCl pH 8.0, 1mM EDTA). While TE protects DNA against nuclease activity by chelating any Mg^{2+} ions, the pH8 helps by keeping it to a pH where the nucleases work less efficiently. HMW DNA was always stored in the fridge at 4°C as freezing will result in physical shearing [53]. DNA was stable for a year or more at this temperature when kept in a nucleases free environment.

Library preparation and DNA sequencing/mapping

Illumina library preparation and sequencing

WGS DNAseq libraries for Illumina sequencing were prepared using KAPA Hyper Prep Kit (Kapa Biosystems) using a PCR-free protocol. The HMW DNA was sheared using a Covaris M220 ultra-sonicator (Thermo Fisher Scientific, Waltham, USA), adjusting the treatment time to obtain ~350bp of DNA fragment length setting 60seconds of fragmentation time. End repair and A-tailing reactions were carried out in 50µL reaction volumes containing 3µL of End Repair & A-Tailing Enzyme Mix, 7µL of End Repair & A-Tailing Buffer and 20-200 ng of input DNA. Reaction mixtures were incubated at 20 °C for 30 minutes and then 65 °C for 30 minutes in a thermal cycler without using a heated lid. Next, 50µL aliquots of reaction products were added to adapter ligation reaction mixtures consisting of 10µL DNA Ligase, 30µL Ligation Buffer, 5µL Agilent SureSelect Adapter Oligo Mix and 5µL nuclease-free water. The 110µL reaction mix was incubated at 20 °C for 30 minutes. Then 0.8X SPRI cleanup was performed by adding 88µL of AMPureXP (Beckman Coulter, Danvers, MA) reagent to the adapter-ligated DNA according to the manufacturer's instructions. DNA was eluted in 100µL of water. This product was size selected adding 0,85X of AMPureXP (Beckman) and it was eluted in 25µl of water. The size of generated libraries was assessed by capillary electrophoresis on a Bioanalyzer High

Sensitivity DNA chip. Library quantification was performed by qPCR using a standard curve. Libraries were sequenced on a Illumina NovaSeq™ 6000 (at the Department of Biology, University of Firenze) aiming to generate on average 1,7 Million fragments using 150nt reads in paired end. Illumina sequencing data were mapped to the *H.pluvialis* chloroplast genome using BWA v0.7.17-r1188 at the Functional Genomic Laboratory of the University of Verona.

Oxford Nanopore library preparation and sequencing

Oxford Nanopore library preparation was conducted according to ONT's specifications using the Ligation Sequencing kit (LSK-SQK 108 for total DNA extraction and LSK-SQK109 for nuclear extractions) with modifications specifically required for long-read sequencing.

5000ng of DNA was usually end-repaired and dA-tailed using the NEB Next End Repair/dA-tailing module. The incubation time was extended at 20°C for 10 minutes and 65°C for 10 minutes according to Josh Quick advice [50]. AMPureXP beads clean-up 1x v/v was performed immediately later and 0,2pmol of the end-prepped DNA were ligated to 5µl of Adapter Mix (provided by ONT) with 10µl of NEBNext Quick T4 DNA ligase during the adapter ligation step, incubating the reaction at RT for 10minutes. Of note, 0,2pmol DNA was the ideal concentration for DNA fragments in order to improve the adapters' ligation to both DNA ends, since library preparations are generated about molarity and not mass. The subsequent clean-up was performed adding 0,4X v/v of AMPure XP beads and washing the beads with 250µl of Long Fragment Buffer (LFB) designed to enrich for DNA fragments of >3Kb. Library yield (usually ~2000 ng) was evaluated by Qubit fluorimetric analysis, while library size by capillary electrophoresis using Tape station 2200.

Sequencing was performed using the MinKNOW software, on a MinION device. Prior to sequencing, the quality control of a SpotON flow cell (FLO-MIN106 R9.4.1) was launched. The flow cell priming was carried out according to the manufacturer's specifications.

The library loading was prepared following manufacturer instructions. 75µl of library were loaded on the SpotON port in a dropwise fashion and the sequencing protocol (NC_48Hr_sequencing_Run_FLO_MIN106_SQK-LSK109) was used. No more than ~15 fmol DNA library were loaded into the flowcell for each run that was usually conducted for ~3h for each loading. After 3h, pores activity was usually reduced to 5% and therefore the run was stopped. In order to use the whole sequencing capacity and the whole amount of prepared library, a "Nuclease flush treatment" was performed between each library loading. The flow cell was incubated with 40units/ml of DNaseI (NEB) to revert the pores to the "single pore" state removing residual DNA and restores 50-80% of pores activity. After the treatment a second library aliquot was loaded and the procedure was repeated until the whole library amount was used. Nanopore sequencing data were mapped to the *H.pluvialis* genome using BWA v0.7.17-r1188 using -x ont2d parameters at the Functional Genomic Laboratory of the University of Verona.

PacBio (SMRT) library preparation and sequencing

PacBio library preparation and sequencing was performed by the company Macrogen in Seoul (South Korea). For preparing >15Kb libraries gDNA was selected by BluePippin (Sage Science). The evaluation of the distribution of the resulting sheared gDNA was performed by running the sheared samples on Agilent Bioanalyzer. PacBio Template Prep Kit was used to repair the ends of fragmented DNA (or non-phosphorylated 5' ends of PCR products) with 2,5µl of End Repair Mix 20X, incubating at 37°C for 20minutes, After purification with 0.45x volume of AMPure PB beads to the End-Repaired DNA, 0,5µM of blunt hairpin adapters were ligated at 65ºC for 10minutes. To remove failed ligation products, 100U/µl of ExoIII and 10U/µl of ExoVII were added incubating the solution at 37ºC for 1 hour. Three step of size selection with 0,45X v/v of AMPure PB beads was performed. The size of PacBio-SMRT bell library was measured by running 30ng of the sample using an Agilent Bioanalyzer 12000 chip. PacBio libraries were sequenced on a Sequel sequencer. PacBio sequencing data were assembled using Flye v2.5 at Functional genomic laboratory of the University of Verona.

BioNano Genome Mapping: DNA labeling and optical map acquisition

In order to acquire Bionano optical maps, the DNA was stained using the Nick Label Repair Stain (NLRS) protocol by Bionano Genomics.

The protocol consists of four sequential steps (Nick, Label, Repair and Stain), and was performed strictly following manufacturer's guidelines (Bionano Prep™ Labeling - NLRS Protocol, Bionano Genomics, #30024). Briefly, 300 ng of purified HMW DNA was nicked by two nicking endonucleases Nt.BspQI 10U/µl and Nb.BssSI (20U/µl) (New England BioLabs) in their respectively buffer provided by Bionano Genomics, and incubating the reaction at 37 °C for 2 hours. Using Taq polymerase (5U/µl) (New England BioLabs), with 1,5µl of 10X Labeling Buffer (Bionano Genomics), the nicked DNA was labeled at 72°C for 1 hour by fluorophore-labeled nucleotides mixed in 1,5µl of 10X Labeling mix (Bionano Genomics). In the third step, labeled DNA was repaired with 40U/µL of Taq ligase (NEB) at 37°C for 30 minutes to restore integrated double strands DNA. In the last step, the DNA backbone was stained overnight in a dark environment at 4°C with 3µl of Bionano DNA Stain reagent mixed with 1 µl of Bionano Stop Solution, 4X Flow Buffer and 5X DTT for visualization and size identification. DNA quantification was carried out using Qubit 4.0 Fluorometer (ThermoFisher Scientific) with Qubit dsDNA assay HS kits after 10minutes of DNA sonication of Top, Middle and Bottom layer of the sample. Only DNA samples with concentration between 4-10ng/µl with a coefficient of variation <25% were chosen to be loaded in the next step. The nicked and labeled DNA was then loaded onto on a Syphyr chip at the ETH functional Genomic Center of the University of Zurich for imaging on the Saphyr system (BioNano Genomics). Imaging was conducted for at least 30cycles and until a mapped coverage of 100X (of molecules >150Kb) was reached. Data analysis was conducted using the Bionano Access software using the PacBio-based draft assembly of *H.pluvialis* genome.

RESULTS

Benchmarking of methods for the extraction of total HMW DNA from H.pluvialis, suitable for Third Generation sequencing

Selection of HMW DNA extraction methods for 3GS

To identify a suitable method that could yield HMW DNA from *Haematococcus pluvialis* for 3GS, three different methods for the extraction of total DNA (namely genomic DNA + plastidial DNA) were initially tested. Two methods exploited organic DNA extraction: CTAB-based and SDS-based methods, that were selected for their capability to efficiently remove carbohydrates and polyphenols from DNA [52], as these factors can heavily affect the purity of DNA extracted from algae [15]. The specific CTAB-based protocol utilized was successfully utilized for the generation of PacBio data from another alga (*Chlorella vulgaris*) [56] (**Appendix Figure A2**). The SDS-based extraction was chosen as a potentially valid alternative to CTAB-based extraction as it can produce cleaner DNA than CTAB-based extraction for some plant species [12] as outlined in the introduction. The third method was gravity column-based, specifically the Genomic tip-100 (G-tip) from Qiagen, proven successful to extract HMW DNA also from plants and algae as described above.

The three HMW DNA extraction approaches were applied starting from *H.pluvialis* biomass ground by liquid nitrogen. The total HMW DNA extractions were evaluated according to three parameters:

1. DNA yield and concentration
2. DNA purity
3. DNA length

Evaluation of results was performed on the basis of requirements for 3GS (**Table 6**), namely >$10^{\wedge 4}$bp DNA molecules that are also highly pure (absorbance ratio of

1,8 for the 260/280nm and <2-2,2< for 260/230nm) and consistency between Qubit and Nanodrop quantification as extensively described in the introduction.

	DNA AMOUNT		PURITY		NANODROP/ QUBIT ratio	LENGTH (bp)
	ng/ul	Total ng	260/280	260/230		
PacBio	10	>1000	~1,8	<2-2,2<	Closest to 1	>10^4
Oxford Nanopore	20	>1000				

Table 6: 3GS HMW DNA requirements

DNA yield and concentration– The highest yield was obtained for SDS- and CTAB-based extractions compared to G-tip extraction (**Table 7**; **Figure 6**). All methods produced sufficient yield and DNA concentration for PacBio and Oxford Nanopore sequencing.

DNA purity– CTAB-based DNA extraction generated DNA with the lowest purity, as demonstrated by a very high Nanodrop/Qubit ratio, which suggests a high presence of contaminants. Instead, SDS yielded DNA with better purity and G-tip yielded optimal purity for 3GS (**Table 7**).

	Yield		Purity			N
	Total ng	ng/ul	$A_{260/280}$	$A_{260/230}$	Nanodrop/ Qubit ratio	
CTAB	4406±1468	88±29	2,11±0,1	2,22±0,06	20,74	3
SDS	5068±772	101±15	1,95±0,1	1,85±0,1	2,02	3
G-tip	3162±1920	63±38	1,8±0,1	2,05±0,04	0,95	3

Table 7: Results of total HMW DNA extraction methods for: DNA yield and concentration (assessed by fluorimetric assay) and DNA purity assessed by spectrophotometer analysis. The mean±sd is reported. N= number of replicates.

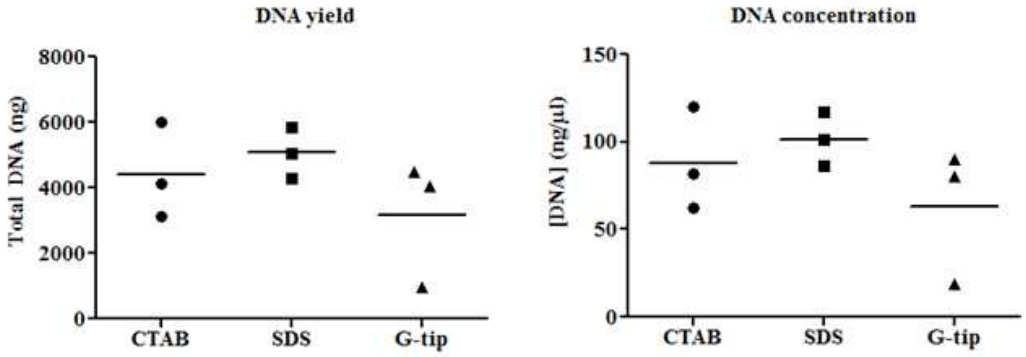
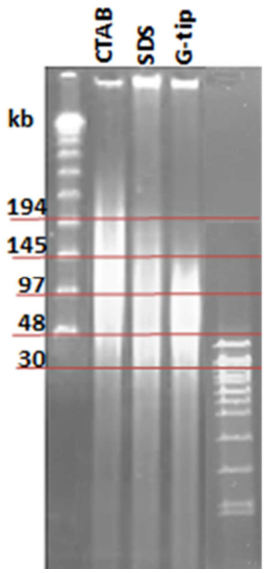


Figure 6: Plotted graphs for total HMW DNA extraction methods: DNA yield and DNA concentration (assessed by fluorimetric assay)

DNA length– In terms of DNA integrity, CTAB-based extraction produced the longest DNA fragments with a size enrichment of <48-194< Kb, while SDS- and G-tip-based extraction produced lower enrichment <30-110/145< (**Figure 7**).



	Length (kb)	
	up to	enrichment zone
CTAB	194	<48-194<
SDS	194	<30-145<
G-tip	145	<30-110<

Figure 7: Results of total HMW DNA extraction methods for DNA length (assessed by PFGE)

The results obtained for total HMW DNA extractions are summarized in a qualitative manner in **Table 8**.

HMW DNA extraction Method	Yield-Concentration (Qubit)	Purity		Length
		Absorbance	Nanodrop/ Qubit ratio	
CTAB	+ +	-	-	+ + +
SDS	+ +	+	+	+ +
G-tip	+	+ +	+ +	++

Table 8: Qualitative results comparison between 3 HMW DNA extraction methods evaluated by: i) DNA yield and DNA concentration ii) DNA purity and iii) DNA length parameters

Benchmarking of post-extraction purification methods

Since CTAB-based extraction from *H.pluvialis* provided the longest DNA fragments, but the DNA had the lowest purity (**Table 7**), a further benchmarking of post-extraction purification methods was performed on DNA isolated using the CTAB-based method. The three purification methods selected were: size-selection (BluePippin-based cleanup), SPRI beads (AMPureXP beads-based cleanup) and alcohol precipitation (ethanol-based cleanup).

The resulting DNA was analyzed as described above and considering the requirement for 3GS (**Table 6**).

The method that most efficiently clean-up the extracted DNA was the gel-based selection using BluePippin (**Table 9**; **Figure 8**). However, this method generated too low yield to proceed further with 3GS and to assess the DNA size on PFGE. Alcohol-based precipitation yielded good amount of DNA but with the lowest purity among the methods tested. AMPureXP bead-based clean-up provided the best combination between yield, purity and length (with fragments up to 145Kb) (**Table 9**; **Figure 8**; **Figure 9**). Alcohol-based and AMPure XP beads-based purification methods reduced also the length of DNA fragments (**Figure 9**) as compared to the starting DNA (**Figure 8**)

	Yield		Purity			N
	Total ng	ng/ul	$A_{260/280}$	$A_{260/230}$	Nanodrop/ Qubit ratio	
BluePippin	508±179	35±5	1,88±0,11	2,28±0,59	1,05	3
AMPureXP beads	1187±706	59,3±35	1,89±0,07	1,94±0,03	1,29	3
Alcohol Precipitation	1336±758	59,8±47	2,0±0,02	1,88±0,07	4,58	2

Table 9: Results of post extraction purification methods for: DNA yield and concentration (assessed by fluorimetric assay) and DNA purity assessed by spectrophotometer analysis. The mean±sd is reported N=number of replicates

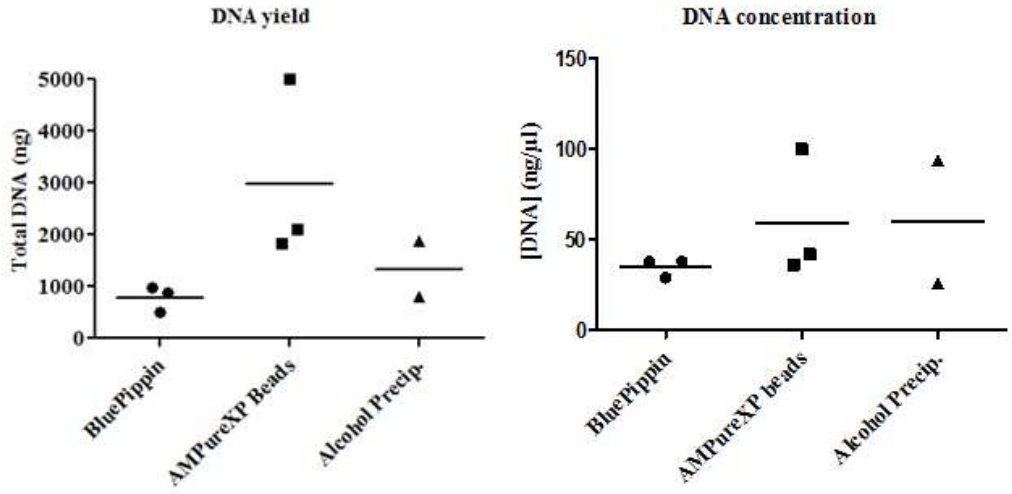
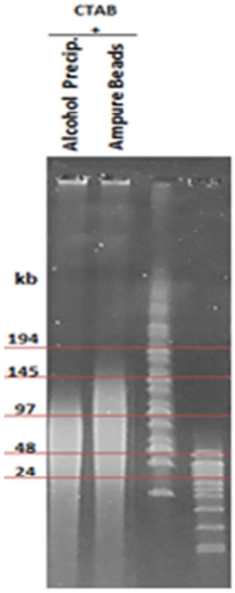


Figure 8: Plotted graphs for total HMW DNA extraction methods: DNA yield and DNA concentration (assessed by fluorimetric assay)



	Length (kb)	
	up to	enrichment zone
BluePippin	na	na
AMPure XP beads	194	<24-145<
Alcohol Precipitation	145	<24-97<

Figure 9: Results of post extraction purification methods for DNA length (assessed by PFGE).

In conclusion, the benchmarking of total DNA extraction methods and post-extraction clean-up for *H.pluvialis* allowed to determine that SDS-based and CTAB-based organic extraction in combination with AMPureXP beads-based

clean-up provided total HMW DNA from *H.pluvialis* with the best feature for 3GS, with the best purity results for the CTAB method (summarized in **Table 10**).

	Yield		Purity			Length (kb)	
	Total ng	ng/ul	$A_{260/280}$	$A_{260/230}$	Nanodrop/ Qubit ratio	up to	enrichment zone
SDS	5068±772	101±15	1,95±0,06	1,85±0,11	2,02	194	<24-145<
CTAB+AMPureXP beads	1187±706	59,3±35	1,89±0,07	1,94±0,03	1,29	194	<24-145<

Table 10: Results and comparison between SDS and CTAB+AMPure XP purification method on: i) DNA yield and concentration; ii) DNA purity, iii)DNA length. The mean±sd is reported.

Analysis of Total HMW DNA by PacBio and Nanopore sequencing

Due to a slightly better purity, CTAB+AMPureXP beads extraction method was selected for PacBio sequencing.

For preparing >15Kb gDNA libraries, 5µg of total HMW DNA was initially size-selected by BluePippin (Sage Science) to provide an enrichment of fragments >15Kb in length. Size selected DNA (~3µg) underwent end-repair treatment and SMRTbell hairpin adapters were ligated according to PacBio indications. Generated libraries were sequenced on 22 SMRT cells on a Sequel I instrument. PacBio sequencing produced good results with a yield over 20 Gb raw data and sequencing read N50 of 11,5Kb. All sequencing parameters were in line with expectation (**Table 11**).

	PacBio sequencing results	
	Expected	CTAB+ AMPureXP beads
#Reads (k)	--	2828,8
#Output (Gb)	18	21,8
Avg length (kp)	10	7,7
Largest (kb)	--	48,5
N50 (kb)	>10	11,5

Table 11: Results of PacBio sequencing for *H.pluvialis* DNA isolated with CTAB+AMPure XP beads method

At the light of results obtained with PacBio sequencing we analyzed the same DNA prep (CTAB+AMPureXP beads) with Oxford Nanopore sequencing.

Aiming to obtain an optimal balance between yield and long-reads, we selected the Ligation protocol (SQK-LSK108) to generate ONT libraries. 2µg of HMW DNA was used into end-repair and dA-tailing reaction and ligated to ONT adapters according to the manufacturer protocol. 372 ng library were loaded into a flowcell (FLO-MIN106-R9.4). Oxford Nanopore sequencing produced very low yield with only 0,08 Gb total data and sequencing read with N50 7,31 Kb (**Table 12**).

To verify whether this poor-quality result was specifically linked to CTAB-based extraction or it was a general issue of total DNA isolated from *H.pluvialis*, nanopore sequencing was performed, with the same library preparation workflow, also on the DNA isolated with the SDS-based method that was meeting the requirement for nanopore sequencing too. Oxford Nanopore sequencing obtained from DNA isolated with the SDS-method provided an improvement in sequencing read length (N50=14,6 Kb) while the sequencing output was still very sub-optimal (0,28 Gb) (**Table 12**).

	Oxford Nanopore sequencing results		
	Expected	**CTAB+ AMPureXP Beads**	**SDS**
#Reads (k)	--	30,3	65,4
#Output (Gb)	>1	0,08	0,28
Avg lenght (kb)	---	2,8	4,3
Largest (kb)	Up to 2Mb	73	140
N50 (kb)	>10	7,3	14,6

Table 12: Results of Nanopore sequencing for two total HMW DNA extraction methods: CTAB+AMPure XP beads extraction and SDS-based extraction.

Considering that the yield of nanopore sequencing was low for both CTAB and SDS-DNA, we concluded that metabolites remaining in the total DNA preparations from *H.pluvialis* could inhibit the pore activity, as reported in the nanopore community for many plants [57][58].

In addition, we noticed that a big fraction of sequencing reads was derived from chloroplast genome and 26,20% and 31,46% of reads respectively for CTAB and SDS-DNA mapped on chloroplast. Indeed, draft-assembly data on the basis of PacBio sequencing allowed to determine that *H.pluvialis* is characterized by a very big chloroplast genome (1,4 Mb).

In the light of these results, in the next step we performed an extensive benchmarking of nuclei extraction from *H.pluvialis*, in order to exclude both chloroplasts as well as metabolites, both of which could cause low sequencing performance. Nuclear DNA would also focus the sequencing data generated on genomic DNA thus allowing a potentially better genome assembly and to reduce sequencing costs that are rather high for Oxford Nanopore.

Benchmarking of methods for the extraction of HMW DNA from *H.pluvialis* nuclei, suitable for Oxford Nanopore sequencing

Selection of nuclear extraction buffers

To extract nuclear DNA from *H.pluvialis*, a benchmarking of 5 nuclei isolation buffers was performed. The different methods (NIBTM, MEB, HONDA, BOLGER and SEBM nuclear isolation buffers) were selected based on their chemical properties and based on literature (**Table 13**). The NIBTM was tested either in combination with a gradient to further purify nuclei (Bionano gradient) or without.

Method	Buffer Composition	Literature Information	Reference
NIBTM (+/- Bionano gradient)	10mM Tris, 80mM KCl, 10mM EDTA, 0,5M Sucrose, 1mM Spermidina, 1mM Spermina, 0,2%BME , **8% PVP10**, 0,5% Triton X-100 (pH9)	Used for Chlorella Vulgaris	Zhang et al. (2012)
MEB	**1M Hexylen Glycol,10mM Pipes-KOH,** 10mM Sodium Metabisulfite, 200mM L-Lysine,6mM EGTA, 10mM $MgCl_2$, 0,5% Sodium diethyldithiocarbamate, 2%PVP 10, 0,5%Triton, 5mM BME **(pH5)**	Used for plant highly rich of metabolities	Lutz, 2011
HONDA	25mM tris, 10mM MgCl2, 0,44 M sucrose, 2mM Spermina, 10mM BME, 5% **Dextran 40**, 1 %Triton, **2,50% Ficoll**	Used for Chlamydomonas	Winck F. et al,2011 (Luthe,Quatrano 1980)
BOLGER	**1M Hexylen Glycol, 10mM PIPES**, 10mM $MgCl_2$, 10mM Sodio metabisulfite, 6mM EGTA, 0,5% Sodium diethyldithiocarbamate, 4% PVP10, 200mM L-Lysine 1mM DTT **(pH6)**	Used for recalcitrant plant sequenced with nanopore	Bolger et al. (2014)
SEBM	10mM Tris, 80mM KCl, 10mM EDTA, 0,5M Sucrose, 1mM Spermidina, 1mM Spermina, 0,2% BME , **10% Triton X-100**, **0,13% Carbamic Acid**, 0,25% PVP40	Used for recalcitrant plant	Carrier et al. (2011)

Table 13: Nuclei isolation buffers selected for benchmarking. In bold are underlined the different components of the solutions.

After isolating nuclei using the 6 methods, nuclear DNA was extracted using the same SDS-based extraction method evaluated in the previous section.

Comparison of nuclei isolation methods

The nuclei isolation methods were evaluated according to two parameters:

1. DNA yield and concentration
2. Nuclei purity

The DNA length was not evaluated because the goal of this step was to identify methods allowing the efficient isolation of nuclei from *H.pluvialis,* as well as producing sufficient DNA for the identification of chloroplast contamination by Illumina sequencing.

DNA yield and concentration–MEB, HONDA and NIBTM (both with and w/o gradient) yielded DNA >1000ng. BOLGER buffer produced lower yield, but still sufficient for downstream DNA sequencing, and SEBM method was inefficient producing very low yield and it was therefore excluded from further evaluations (**Table 15**; **Figure 10**).

	Yield		N
	Total ng	ng/ul	
MEB	7220±2854	361±143	3
HONDA	3759±2549	188±127	3
NIBTM	2667±326	136±15	3
NIBTM+Grad	1257±357	43±35	3
BOLGER	719±141	36±7	3
SEBM	8,3±2,9	3±2	3

Table 15: Results of nuclear HMW DNA extraction methods for: DNA yield and concentration (assessed by fluorimetric assay) and concentration. The mean±sd is reported. N=number of replicates

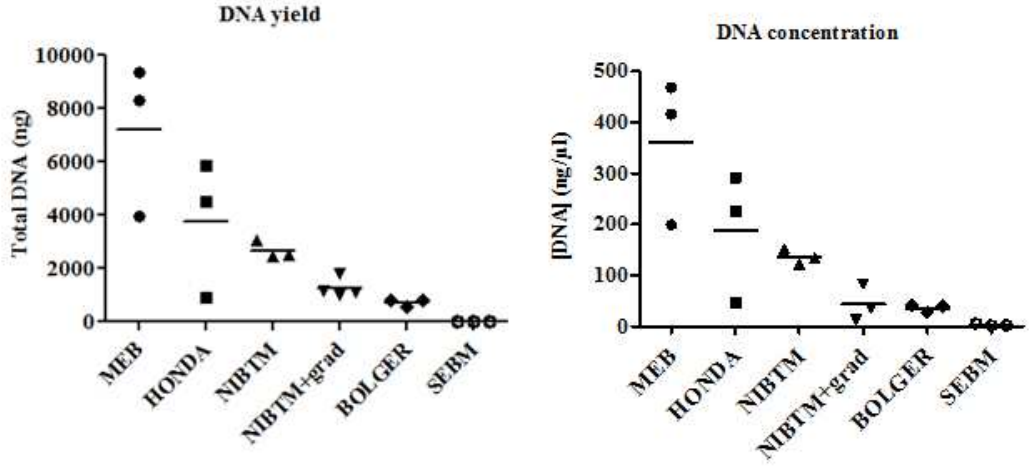


Figure 10: Plotted graphs for nuclear DNA extraction methods: DNA yield and DNA concentration (assessed by fluorimetric assay)

Nuclei purity– The efficiency of nuclear isolation methods to remove chloroplast DNA was assessed by sequencing DNA with short-reads (Illumina sequencing), that allowed such analysis at low cost. Contamination of chloroplast was estimated as equivalent to the percentage of reads mapping on draft *H.pluvialis* chloroplast genome.

According to this analysis, we could conclude that MEB and BOLGER buffers, presenting respectively the 13,1% and 13,3% of reads mapping on chloroplast genome, were the methods providing the purest nuclear preparations from *H.pluvialis*. They were followed by HONDA (22,4%), NIBTM with gradient (24,1%) and NIBTM w/o gradient (25,6%) (**Table 16**; **Figure 11**).

% reads mapping on chloroplast

	Mean±sd	N
TOTAL DNA	25,8 ± 4,7	3
MEB	13,1 ± 4,5	3
HONDA	22,5 ± 7	3
NIBTM	25,6 ± 3,9	3
NIBTM+Grad	24,1 ± 2,7	3
BOLGER	13,3 ± 4,5	3

Table 16: Results of DNA purity -represented as percentage of reads mapping on chloroplast genome- for 6 nuclear isolation methods evaluated by Illumina sequencing. The mean±sd is reported. N=number of replicates.

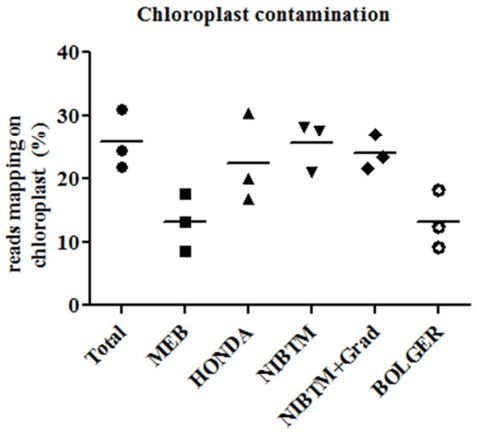


Figure 11: Plotted graphs results for nuclear DNA extraction methods: nuclear DNA purity represented as percentage of reads mapping on chloroplast genome- for 6 nuclear isolation methods evaluated by Illumina sequencing.

Considering that a total (nuclear+plastidial) DNA extraction provided 25,8% of reads mapping on chloroplast genome, it was evident that NIBTM buffer either with or without gradient was not efficient in the removal of chloroplast DNA. Nevertheless, the same buffer worked properly to isolate nuclei in a plant (*V.vinifera*) and in another alga (*C.vulgaris*), as shown in **Appendix Table A2; Figure A3**. Considering the inefficiency of NIBTM in isolating nuclei from *H.pluvialis*, this method was excluded from further analysis.

Benchmarking of HMW DNA extraction methods starting from H.pluvialis nuclei

In the light of nuclear purity evaluation, HMW DNA was subsequently extracted from nuclei isolated with MEB, BOLGER and HONDA methods by using two HMW extraction methods that were benchmarked in parallel, namely SDS- and G-tip-based extraction. The CTAB based-method+AMPureXP beads was not tested in combination with nuclei isolation because it did not yield significantly better results than SDS-method for the extraction of total DNA. Moreover, the method is very time-consuming (10 vs 5 hours of the SDS-based methods) and quicker DNA extraction methods are recommended to prevent DNA damage [15].

DNA extractions were compared based on the same parameters assessed for total DNA extraction (yield, purity and length) on the basis of DNA requirements for Oxford Nanopore (**Table 6**).

DNA yield and concentration– All nuclear isolation buffers, in combination with SDS-based extraction yielded higher DNA amount in comparison to G-tip-based extraction. Furthermore, BOLGER buffer+G-tip did not produce sufficient DNA yield for Oxford Nanopore. SDS-based extraction in combination with MEB yielded the highest DNA amount (**Table 17**; **Figure 12**).

SDS-based extraction

	Yield		N
	Total ng	ng/ul	
MEB	7220±2854	361±143	3
HONDA	3759±2549	188±127	3
BOLGER	719±141	36±7	3

G-tip-based extraction

	Yield		N
	Total ng	ng/ul	
MEB	1987±887	168±98	3
HONDA	2022±1323	237±49	3
BOLGER	229±193	37±43	3

Table 17: Results of nuclear SDS- and G-tip-based HMW DNA extractions for DNA yield and concentration (assessed by fluorimetric assay). The mean±sd is reported. N=number of replicates

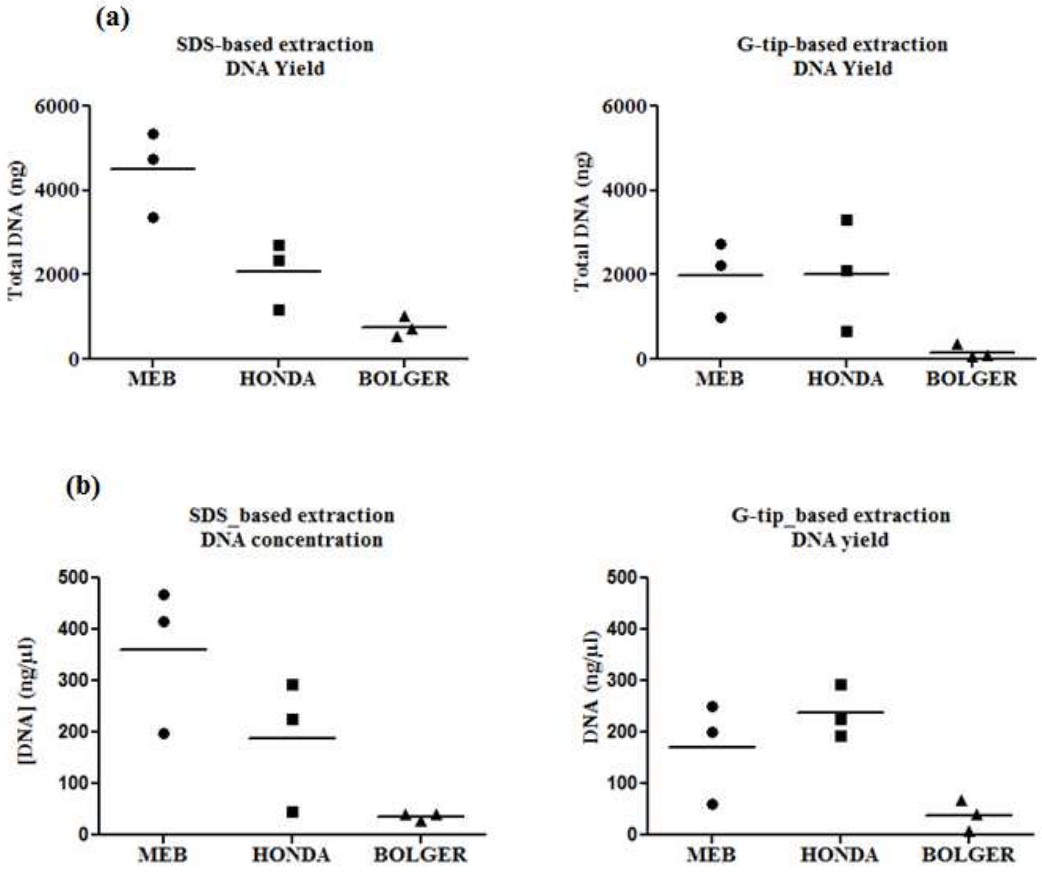


Figure 12: Plotted graphs for nuclear SDS- and G-tip-based HMW DNA extraction methods for (a) DNA yield (b) DNA concentration (assessed by fluorimetric assay)

DNA purity– All nuclear isolation buffers, in combination with SDS-based extraction produced lower purity than ONT requirement for 260/230 ratio. On the contrary, all three nuclear isolation buffers in combination with G-tip extraction generated good DNA purity, especially for MEB and BOLGER buffer (**Table 18**)

Purity_SDS extraction

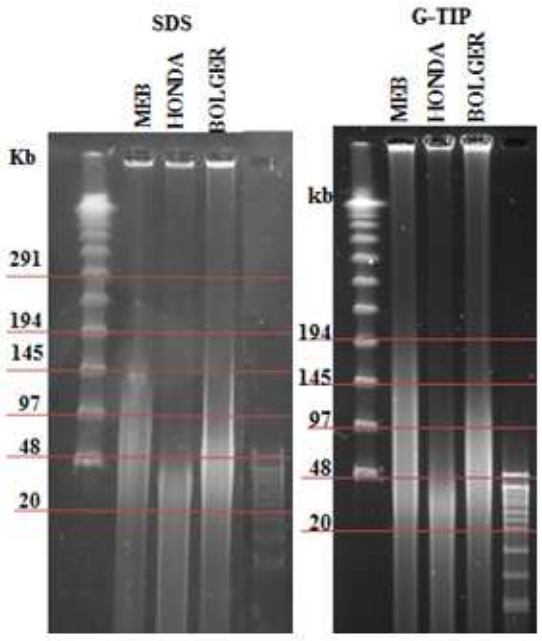
	$A_{260/280}$ Mean±sd	$A_{260/230}$ Mean±sd	Nanodrop/ Qubit ratio	N
MEB	1,82±0,01	1,83±0,01	1,18	3
HONDA	1,84±0,1	1,82±0,1	0,88	3
BOLGER	1,95±0,2	1,94±0,1	0,94	3

Purity_G-tip extraction

	$A_{260/280}$ Mean±sd	$A_{260/230}$ Mean±sd	Nanodrop/ Qubit ratio	N
MEB	1,86±0,05	2,07±0,04	1,18	3
HONDA	1,83±0,01	2±0,04	1,08	2
BOLGER	1,84±0,01	2,1±0,1	0,9	3

Table 18: Results of nuclear SDS- and G-tip-based HMW DNA extraction methods for DNA purity (assessed by fluorimetric assay). The mean±sd is reported. N=number of replicates

DNA length– MEB and BOLGER produced the highest enrichment zone <20-145<Kb, while HONDA buffer produced DNA fragments of <48Kb in size (**Figure 13**). The DNA extraction method (SDS vs G-tip) did not influence the size of DNA while the nuclei isolation methods affected the DNA length.



Nuclei Isolation Buffer	HMW DNA extraction	up to (kb)	Enrichment zone (kb)
MEB	SDS	194	≤20-145<
HONDA		48	<20-48<
BOLGER		194	<20-97<
MEB	G-TIP	194	<20-145<
HONDA		48	<20-48<
BOLGER		194	<20-97<

Figure 13: Results of nuclei isolation+ HMW DNA extraction methods for DNA length (assessed by PFGE: range 5-450Kb)

In conclusion, after this extensive benchmarking, as it is shown in a qualitative manner in **Table 19**, we concluded that MEB nuclei isolation buffer in combination with G-tip produced more than 1µg of DNA yield, with optimal absorbance ratio and DNA fragments up to 194 Kb in length with an enrichment zone between <24-145< Kb. For these reasons, this approach was selected for the subsequent analysis by Oxford Nanopore sequencing.

Nuclei Isolation Buffer	HMW DNA extraction	Yield	Purity		Length
			Absorbance	Nanodrop/ Qubit ratio	
MEB	SDS	++	+	+	+++
HONDA		+	+	+	+
BOLGER		-	+	+	++
MEB	G-TIP	+	++	+	+++
HONDA		+	++	+	+
BOLGER		-	++	+	++

Table19: Qualitative results comparison between 3nuclear isolation buffer matched with 2 HMW DNA extraction methods evaluated by: i) DNA yield ii) DNA purity and iii) DNA length parameters

Analysis of nuclear HMW DNA by Oxford Nanopore sequencing

Oxford Nanopore sequencing was performed on nuclear DNA isolated with G-tip columns in combination with MEB method for nuclei isolation.

Despite the nuclear DNA showed an enrichment of fragments with ideal length for 3GS, it showed also a high amount of short fragments, visible as a smear in the gel at molecular weight lower than ~20Kbp (**Figure 13**). For this reason, the nuclear DNA was sequenced as such or after a treatment with Short Read Eliminator (SRE-Circulomics), capable of eliminating short DNA fragments (up to 25Kb) that are favored during Oxford Nanopore sequencing as compared to longer ones [10] [60]. As it is shown in **Figure 14**, SRE treatment indeed efficiently removed DNA fragments below 4Kb and progressively up to 15Kb from *H.pluvialis* DNA.

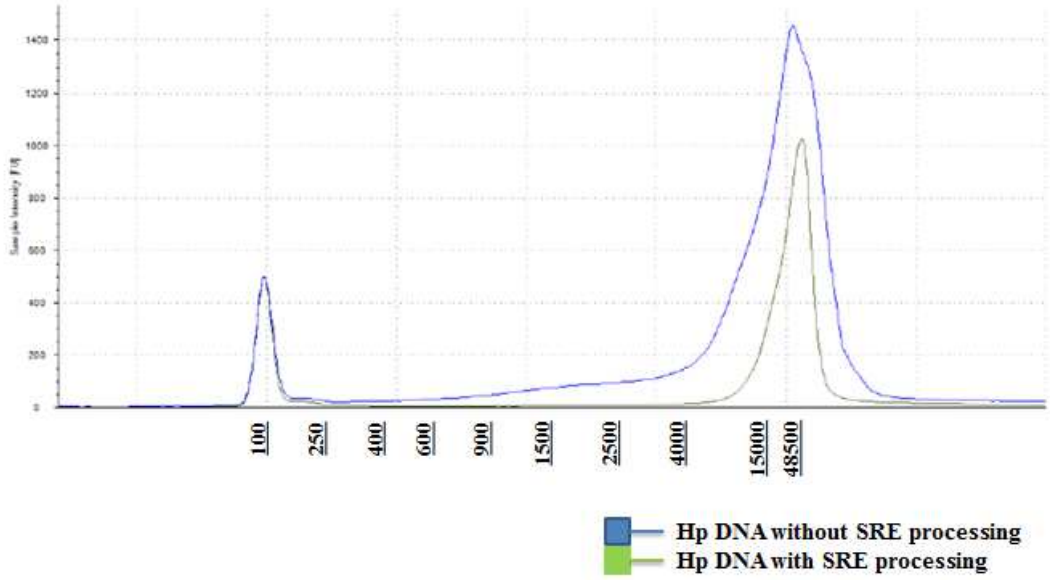


Figure 14: Result of SRE processing evaluated by Tape station 2200. The blu line shows the DNA without SRE processing. The green line shows the DNA with SRE processing.

Oxford Nanopore library was performed as described above from both DNA samples (+/-SRE). 5µg of DNA was end-repaired and dA-tailed using the NEB Next End Repair/dA-tailing module and libraries were prepared using the 1D ligation sequencing protocol (SQK-LSK109) according to the manufacturer instructions, with the exception of incubation times that were elongated in line with Josh Quick advices [54] for long reads sequencing. ~2,5µg of DNA library was obtained but no more than ~15 fmol library were loaded into the flowcell (FLO-MIN106_R9.4.1).

Oxford Nanopore sequencing of nuclear HMW DNA (**Table 20**) without any further treatment produced higher yield (0,53Gb) as compared to total extraction (0,08-0,28Gb) (**Table 12**). In contrast, the length of Oxford Nanopore reads was lower, with N50=3Kb versus 7-14Kb of total HMW DNA. This was consistent with the fact that the enrichment zone for nuclear DNA started from 20Kb and the DNA showed a high level of partial degradation (smear at lower MW) (**Figure 13**), while for total extraction the enrichment started from ≥30 Kb and the smear was less evident (**Figure 7**).

Sequencing of nuclear HMW DNA with SRE-treatment, removing fragments below 15Kb, generated an improved N50 that reached 22,5Kb with 146,6Kb for

the longest read. However, even in this analysis the output was still lower than the minimum for Oxford Nanopore (1Gb) (**Table 20**).

	Oxford Nanopore sequencing results		
	Expected	**No treatment**	**SRE**
#Reads (k)	--	269,36	76,35
#Output (Gb)	>1	0,53	0,9
Avg length (kp)	---	1,99	11,8
Largest (kb)	Up to 2Mb	156,9	146,6
N50 (kb)	>10	3,0	22,5

Table 20: Results Oxford Nanopore sequencing of MEB+G-tip-based extraction with i)NO treatment post extraction ii) with SRE

To increase the output of nanopore sequencing by disrupting potential secondary structures present in HMW DNA that can block pores, we fragmented DNA at ~20Kb using g-TUBE™ (Covaris) prior to sequencing. Nanopore sequencing of DNA fragmented either by sonication with Megaruptor or by physical shearing using g-TUBE is indeed reported to improve sequencing yield[60][61]. To obtain homogeneous fragments of the same length and have a good balance between length and shearing, it is important to start from HMW DNA; in this way the procedure allows to produce very narrow-size distribution of fragments with an enrichment of 20Kb size (**Figure 15**).

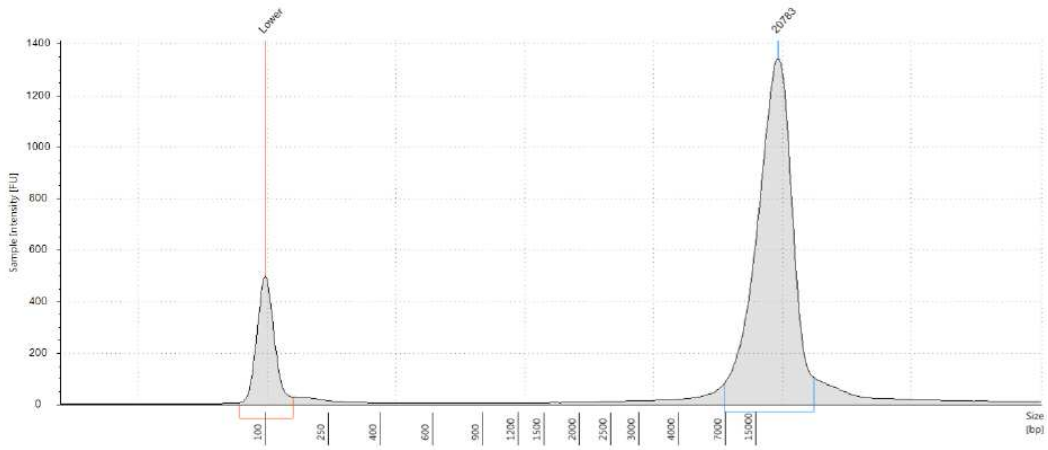


Figure 15: Result of g-TUBE fragmentation at 20Kb evaluated by capillary electrophoresis (Tape station 2200).

The removal of short fragments with SRE-treatment in combination with g-TUBE allowed to get both higher yield (2,3 Gb) as well as to maintain a good read length with N50 of 19,2 Kb (**Table 21**).

	Oxford Nanopore sequencing results	
	Expected	**SRE+g-TUBE**
#Reads (k)	--	187,2
#Output (Gb)	>1	2,3
Avg length (kp)	---	12,2
Largest (kb)	Up to 2Mb	133,9
N50 (kb)	>10	19,2

Table 21: Results of Oxford Nanopore sequencing for MEB+G-TIP-based extraction with SRE+g-TUBE treatment.

Overall, the benchmarking of nuclear HMW DNA extraction methods allowed to purify DNA with lower chloroplast contamination and suitable features for Oxford Nanopore sequencing. In conclusion, extra post-processing adjustments (SRE+g-TUBE) allowed to further increase read length and sequencing yield.

Benchmarking of methods for the extraction of Ultra-HMW DNA from *H.pluvialis* nuclei, suitable for Bionano optical mapping

To identify a suitable method that could yield Ultra-HMW DNA from *H.pluvialis* for Bionano optical map generation, the previous 3 methods validated for the isolation of nuclei from *H.pluvialis* (namely MEB, BOLGER and HONDA) we coupled to plug-based extraction method to obtain Ultra-HMW in the order of $10^{\wedge 5}$bp in length.

In addition, we tested two additional nuclei isolation protocols based on the commercial kit of Bionano Genomics for plants, also coupled to plug-based extraction method. Indeed, one of these Bionano protocols worked properly in our hands to isolate UHMW from another algae, Chlorella vulgaris and to generate optical maps from this species (**Appendix Figure A4**) [56]. The difference between the two Bionano protocols tested was the use of frozen biomass, or biomass that was fixed with formaldehyde before freezing. Fixation should reduce the shearing of DNA due to processing and thus yield longer DNA.

The nuclei isolation methods were evaluated according to three parameters:

1. DNA yield and concentration
2. DNA length
3. Nuclei purity

Evaluation of results was performed on the basis of DNA requirements for Bionano optical map generation (**Table 22**).

	DNA AMOUNT		LENGTH
	ng/ul	Total ng	(bp)
Bionano	<35-200<	300	$10^{\wedge 5}$

Table 22: Bionano OM Ultra-HMW DNA requirements

DNA yield and concentration– The highest yield was obtained for MEB isolation buffer, followed by Bionano buffer (both with and without fixing) and HONDA

buffer. BOLGER confirmed to produce the lowest DNA yield that was below than Bionano optical mapping requirement (**Table 23**; **Figure 16**).

	Yield		N
	Total ng	**ng/ul**	
MEB	1766±482	88,3±24	3
HONDA	365	18,3	1
BOLGER	218±103	10,9±5	5
BIONANO	853±243	43±12	3
BIONANO+FIX	1203±447	60±22	3

Table 23: Results of nuclear isolation buffer for UHMW DNA extraction for DNA yield and concentration (assessed by fluorimetric assay). The mean±sd is reported. N=number of replicates.

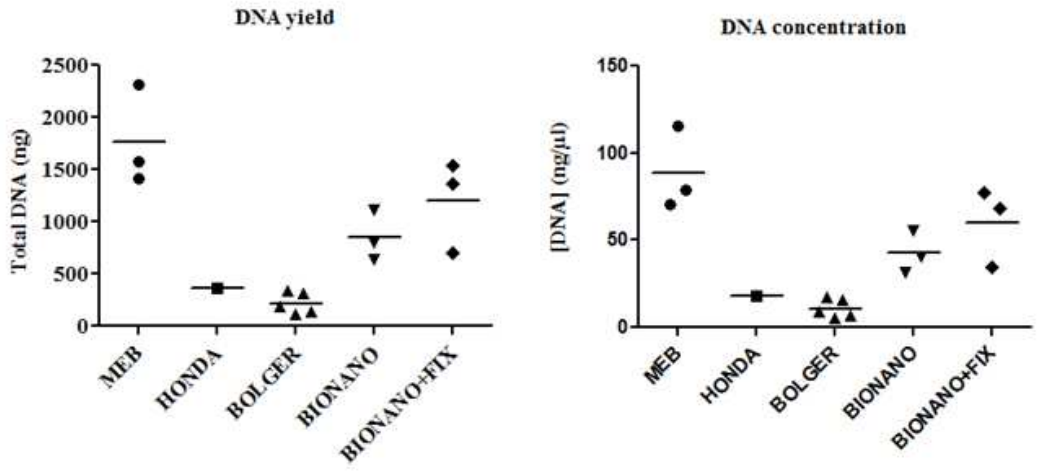
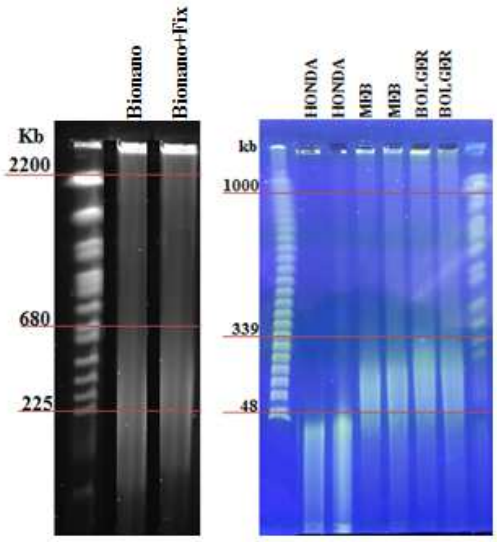


Figure 16: Plotted graphs for nuclear UHMW DNA extraction methods: DNA yield and DNA concentration (assessed by fluorimetric assay).

DNA length– Extraction based on agarose plug effectively allowed to recover DNA up to Megabases in length, namely higher than the previous extraction methods tested except for HONDA buffer that confirmed to produce DNA fragments up to 48Kb. MEB and BOLGER yielded DNA with similar enrichment zone <48-339< Kb. For MEB and BOLGER a small fraction of fragments above 339 and up to 1Mb was also evident. Bionano buffers used with and without fixing produced the highest DNA fragments up to 2,2Mb in length and an enrichment of fragment size of <225-680< Kb (**Figure 17**).



Nuclei Isolation Buffer	HMW DNA extraction	up to (kb)	Enrichment zone (kb)
MEB	PLUG	1000	<48-339<
HONDA		48	<48
BOLGER		1000	<48-339<
BIONANO		2200	<225-680<
BIONANO+FIX		2200	<225-680<

Figure 17: Results of UHMW DNA extraction for DNA length (assessed by PFGE)

Nuclei purity**-** Nuclei purity was evaluated in parallel for all isolation methods as described. MEB buffer confirmed to provide the lowest % of reads mapping on chloroplast genome followed by BOLGER buffer. Bionano buffer could not efficiently remove chloroplast from *H.pluvialis* nuclei. In addition, we observed that tissue fixation strongly increased (+12%) the percentage of reads mapping on chloroplast genome (**Table 24**; **Figure 18**). HONDA buffer was not included in this analysis because it did not satisfy the Bionano length parameter.

% reads mapping on chloroplast

	mean±sd	N
TOTAL DNA	25,8±4,7	3
MEB	7,6±4,1	3
BOLGER	12,8±1,3	4
BIONANO	24,6±10	3
BIONANO+FIX	37,9±7,6	4

Table24: Results of Ultra-HMW DNA purity -represented as percentage of reads mapping on chloroplast genome- for 4 nuclear isolation methods, in comporasinon with Total DNA, evaluated by Illumina sequencing. The mean±sd is reported. N=number of replicates.

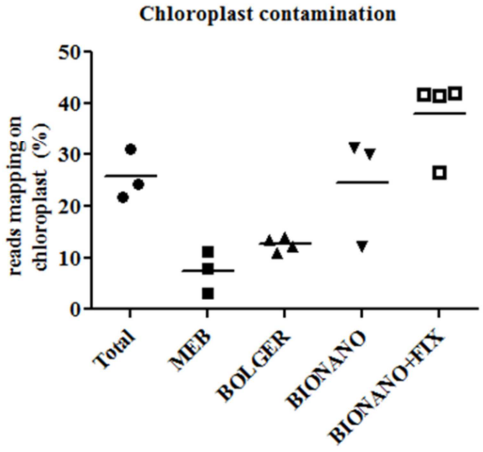


Figure18: Plotted graph for Ultra-HMW DNA purity -represented as percentage of reads mapping on chloroplast genome- for 4 nuclear isolation methods, in comporasinon with Total DNA, evaluated by Illumina sequencing.

In conclusion, as it is shown in a qualitative manner in **Table 25**, MEB, BOLGER and Bionano buffer both with and without fixation in combination with plug-based Ultra HMW DNA extraction method produced DNA fragments up to megabases in length. MEB buffer generated also enough yield and the highest nuclei purity to perform optical mapping.

Nuclei Isolation Buffer	HMW DNA extraction	Yield	Nuclear DNA purity	Lenght
MEB	PLUG	+	++	+
HONDA		-	na	-
BOLGER		-	+	+
BIONANO		+	-	++
BIONANO+FIX		+	- -	++

Table25: Qualitative results comparison between 3nuclear isolation buffer matched with plug-based extraction method evaluated by: i) DNA yield ii) Nuclear DNA purity; iii) DNA length parameters.

Analysis of UHMW DNA by optical mapping

Optical Maps were acquired after Nick-Label-Repair (NLRS) staining of the Ultra-HMW DNA extracted by MEB nuclei isolation buffer+plug. The nicking enzyme selected for Bionano optical map generation were NtBspQI and NbBssSI, which provided an expected label/100Kb of 7 and 20, respectively, for the *H.pluvialis* genome (optimal labeling density is <6-20<). Respectively 300 ng of Ultra-HMW DNA was nicked by two nicking endonucleases Nt.BspQI (10U/µl) and Nb.BssSI (20U/µl) (New England BioLabs), previously selected. NLRS was performed strictly following manufacturer's guidelines (Bionano Prep™ Labeling - NLRS Protocol, Bionano Genomics, #30024) and Optical Maps were acquired using the Saphyr instrument.

Data acquired from NtBspQI- and NbBssSI-labelled DNA showed that DNA molecules had an average length of 176 and 279Kb after filtering for >150Kb (filter required optical map assembly) respectively. The label density detected was 9/100Kb and 19,9/100Kb in line with the expected ones for both nickases. The output was more than sufficient for optical map assembly, namely 100X the genome size (30 Gb for *H.pluvialis*).

Overall the results obtained demonstrated that the selected protocol for Ultra-HMW DNA extraction from *H.pluvialis* (MEB+plug) yielded suitable DNA for optical maps generation (**Table 26**).

	Bionano optical Mapping generation results		
	Expected	**Nt.BspQI**	**Nb.BssSI**
Expected Labels/100kb	NtBspQ1=7,8 Nt.BsSSI= 20,7	9	19,9
#Output (Gb)	>30*	198,6	163,3
N50 (>150kb)	>150	196	210
Largest molecule (kb)	___	2812,8	2814,4
Positive label variance (%)	<3-10<	5,6%	5,1%

Table26: Results of optical mapping acquisition by Saphyr instrument (Bionano Genomics) starting from MEB+plug DNA stained with Nt.BspQI and Nt.BsSSI. *100X genome coverage required for OM assembly

DISCUSSION

Extracting high quality, High and Ultra MW DNA from plants and algae can represent the bottleneck in the era of whole genome sequencing by long reads technologies and optical mapping. These organisms can indeed carry a though cell wall and different metabolites that can strongly affect DNA purity and length that influence the quality of the down-stream -long-read sequencing or mapping-analysis [62][63]. Standardized extraction methods to obtain highly pure (U)HMW DNA from these organisms are still missing and only a limited set of methods is available. The work presented in this study was devoted to the benchmarking of methods for the extraction of total/nuclear (Ultra)HMW DNA from a microalgae selected as case-study, *Haematococcus pluvialis* (*H.pluvialis*), for which no protocol for (U)HMW DNA extraction has been reported so far. Overall, 4 (U)HMW DNA extraction methods, 8 nuclei isolation methods and 4 post-extraction DNA purification methods were tested, in order to produce good quality DNA to be analyzed by Third generation Sequencing and Optical Mapping. Only 3 specific combinations of these protocols yielded suitable DNA to generate successful results with PacBio (CTAB buffer+AMPureXP beads purification), Oxford Nanopore (MEB buffer+G-tip- DNA based extraction) and Bionano (MEB buffer+plug- DNA based extraction).

Overall the work showed that the optimization of HMW DNA extraction protocols can be very challenging and time consuming. HMW DNA extraction protocols must be carefully tested and adjusted according to the tissues and species of interest as their modulation can determine quite profound differences in the resulting DNA prep. This was perfectly evident on the fact that at least three methods, that did not produce suitable DNA from *H.pluvialis*, were perfectly effective on another algae *C.vulgaris* (namely: CTAB based extraction, NIBTM nuclei isolation, and Bionano nuclei isolation buffer+plug-based DNA extraction). Another striking example was the SEBM method that didn’t yield almost any DNA from *H.pluvialis*, while it was very effective in eggplant [40].

In addition, profound differences exist in the tolerance of each sequencing/mapping technology to the impurities carried by the DNA to be analyzed. For example, the CTAB protocol was effective to yield DNA of sufficient purity for PacBio sequencing in our hands, and it was also used by others for Illumina sequencing from algae [64]. In contrast this DNA prep was not suitable for Oxford Nanopore sequencing in *H.pluvialis*, where it produced very poor results. Consistently with other reports[20][21], these results demonstrated that ubiquitous extraction methods effective for different organisms do not exist and that the extraction procedure must be carefully adjusted also according to the downstream sequencing technology to be used.

Noteworthy, Oxford Nanopore sequencing is very sensitive to contaminants co-purified with DNA that can lead to pores clogging[57][58], directly or indirectly. Consistently, also with the best DNA prep produced in this work (MEB nuclei isolation+G-tip-based DNA extraction), we observed a constant pore inhibition. Indeed, activity of the pores was usually strongly decreased down to 5% after only 3h of sequencing. This suggests that *H.pluvialis* DNA carries (unknown) metabolites that can affect pore activity either directly or by inducing/favoring the creation of DNA secondary structures that indirectly lead to pore clogging. This may be exacerbated by the fact that *H.pluvialis* genome is characterized by a large fraction of (long) repeated sequences (59.32%, *H.pluvialis* draft genome, unpublished results) that can favor this process. Consistently with the last hypothesis, treatment with DNase is capable to remove the issue and restore pore activity, with a minor pore loss due to the treatment itself. This phenomenon is widely observed also for other plants and algae [Oxford Nanopore and circulomics personal communication].

Evaluation of post-extraction purification methods allowed to identify approaches that can effectively ameliorate DNA purity, however this came at cost of yield and/or DNA size. BluePippin could perform a very effective DNA size-selection and purification; however a big loss of DNA was experienced in our hand using this method (only 15% DNA was recovered after bluepippin selection). In contrast, alcohol-based and AMPureXP-based purifications allowed to obtain

higher DNA recovery but the first procedure left behind some contaminants (most likely due to phenols employed in the procedure) while the second consistently shortened the DNA size (-50Kb). Therefore, even if these treatments may be effective in improving DNA purity, attention must be paid to their impact on the final DNA. As it happened for *H.pluvialis*, rather than attempting the purification of already extracted DNA, it is probably more effective to identify alternative extraction methods that are more suitable for the species of interest. It must be noted also that the time required to perform these post-processing procedures is not negligible, especially for Bluepippin (7h) and Alcohol-bases precipitation (8h).

In contrast, post-extraction processing procedure aimed at changing the size of DNA was successful in our hands to improve Oxford Nanopore sequencing results. In fact, removal of short DNA fragments using the selective precipitation with the SRE kit allowed to consistently increase the read length (+19,5Kb) as compared to the untreated DNA.

Short reads are a real issue for oxford nanopore technology when one is aiming at long read sequencing [59][49]. Indeed, while ONT flow cells can sequence extremely long DNA fragments (up to 2Mb), the generated read length directly reflects the size distribution of fragments present in the input sequencing library[59]. The sequencing of short-fragments generates significantly more individual reads within a given time than a long-fragment sequencing run[65][10]. DNA molecules of different length behave differently in solution, which can affect the efficiency of adapter ligation and influence preferential sequencing; short-fragments outcompete longer DNA molecules in both cases [12]. Therefore, sequencing of very-long fragments could overall reduce the output of nanopore flowcell. In addition, with very HMW DNA we basically introduce in the flowcell a big intertangled ball of DNA that can block the pores, as described above. Consistently, the short-fragment removal in combination with DNA fragmentation capable of generating homogeneous DNA fragments and to disrupt DNA secondary structure (g-TUBE) further improved our results in terms of yield (+1,77Gb as compared to untreated sample). According to these evidences, the

post-extraction treatments on DNA should account for a delicate balance between DNA size, appropriate fragmentation and short-read removal to obtain efficient Oxford Nanopore runs.

DNA purity can be ameliorated also by extracting DNA from pre-isolated nuclei. This procedure can avoid that a large fraction of metabolites and cell wall/cytosol components get in direct contact with nuclear DNA and bind to it, thus affect its final purity. This procedure allows also to exclude plastidial genomes (present frequently in multiple copies) and focus sequencing data on nuclear genome, thus saving costs and improving final assembly data. This was of utmost importance for *H.pluvialis* that is also characterized by an extremely large chloroplast genome, 1.4Mb versus a genome of 300Mb. Indeed, sequencing data mapping on chloroplast (starting from total DNA prep) reached up to 30% in *H.pluvialis*, while it was only 7,1% for *C.vulgaris* and 8,4% for *V.vinifera*, that have a chloroplast of 0,17Mb and 0,16Mb, respectively. Also, this high percentage is due to the fact that the chloroplast genome is present in multiple copies inside each cell that carry only a single genomic DNA copy instead.

Literature reports a limited amount of protocols for nuclei isolation from plants and algae and specifically none for *H.pluvialis.* Therefore a benchmarking of 8 nuclear isolation methods was performed. Out of these, only 2 methods effectively decreased the percentage of chloroplast-mapping reads lower than 15%, while the others did not effectively reduce the contamination. In particular, the NIBTM nuclear isolation buffer, that worked properly in algae and plants in our hands, was inefficient for nuclear isolation from *H.pluvialis*. The SEBM nuclei isolation buffer almost did not yield any DNA from *H.pluvialis*, despite providing high quality DNA for eggplant[40]. This buffer uses a high concentration of Triton X-100 (10%) that could destroy also *H.pluvialis* nuclei in addition to cell membrane.

The inefficiency/difficulty of methods in removing chloroplasts from *H.pluvialis,* may be explained also by the physical conformation of this alga. According to the literature [66], the outer envelope of algae chloroplast is continuous with the outer membrane of the nuclear envelope as a double-membraned as in other unicellular

algae. In addition, specifically in *H.pluvialis*, chrloroplast occupies more than 41,7% of the total cell volume [67]. This physical conformation determines that residual of chloroplasts could remain bound to the nuclear membrane after the isolation of nuclei; this, in combination with large chloroplast genome size, may lead to high chloroplast contamination in the final genomic DNA prep. These results demonstrated once more that a careful species-specific evaluation of the most appropriate nuclear isolation buffer is required to obtain optimal nuclear DNA preps.

Surprisingly, another factor strongly affecting nuclear purity was the fixation of biomass with formaldehyde prior to DNA extraction. Again, this may be due to the physical conformation of *H.pluvialis*, placing in close proximity chloroplast and nuclear genomes. Fixation is a procedure utilized by different protocols to preserve integrity when isolating HWM DNA. In addition, tissue fixation is required for multiple methods that allow to investigate chromatin structure (e.g. 3C or HiC) or the interaction between protein and DNA (e.g. ChIPseq). Even if additional studies should be performed to determine whether and how much fixation can affect these approaches, aspecific chrloroplast linkage to nuclear genome may lead to false positive results or altered derived conformations.

The size of nuclear DNA preps was in general shorter than that of total DNA extractions. This is most likely due to the fact that the isolation of nuclei is a quite time-consuming procedure using a long set of sample manipulations that can affect DNA size. In addition, in contrast to our expectations, the benchmarking of HMW DNA extraction from pre-isolated nuclei highlighted that the final DNA length is much more influenced by the buffer utilized for nuclei isolation rather than by the DNA extraction method. The most evident result came from the HONDA buffer that never generated DNA fragments higher than 48Kb with any DNA isolation method. In general, the ranking in term of DNA size for the other buffers was always maintained despite the changes in the DNA extraction method, with BOLGER and MEB being always the best. The fact that the nuclear buffer is the parameter mostly influencing the DNA size was also demonstrated by fact that

extraction with agarose-plug method (that should preserve fully intact DNA) did not profoundly improve DNA size.

In conclusion, this work has identified three efficient methods to isolate nuclei from *H.pluvialis,* three efficient methods to extract High and Ultra-High MW DNA from *H.pluvialis* giving also the possibility to generate 3GS data for the assembly of *H.pluvialis* genome and to generate optical mapping data for the scaffolding of *H.pluvialis* genome. Moreover, this work underlined the importance to set up a nuclear DNA isolation method in combination with ideal HMW DNA extraction suitable for long molecules technologies. It also demonstrated that specific post-extraction DNA treatment can further improve the quality of generated data, in particular with Nanopore Sequencing. Last but not least, this work has repeatedly emphasized that a global efficient and excellent technique for the extraction of HMW DNA does not exist. Benchmarking studies as the one presented here are therefore very important, especially for plants and algae, to exploit the full potential of long read technologies.

APPENDIX

	Oxford Nanopore sequencing results	
	Expected	**SEBM buffer**
#Reads (k)	--	277,15
#Output (Gb)	>1	3,27
Avg length (kp)	---	10,35
Largest (kb)	Up to 2Mb	138,5
N50 (kb)	>10	17,2

Table A1: Eggplant Nanopore sequencing results for SEBM nuclei isolation method

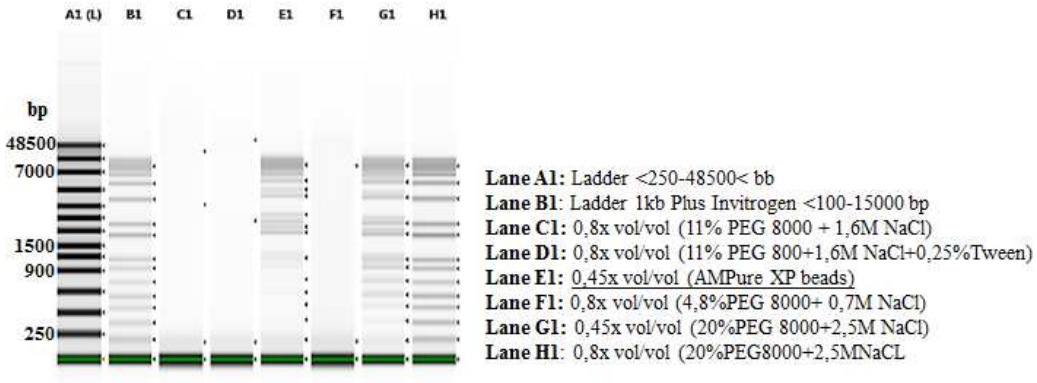
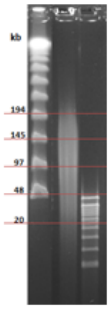


Figure A1: Internal benchmarking performed according to literature [30] [48] [47] to select the right PEG+NaCl combination to obtain longest fragments.

(a)

	Yield		Purity			N
	Total ng	ng/ul	$A_{260/280}$	$A_{260/230}$	Nanodrop/ Qubit ratio	
CTAB	2043±290	86±13	1,9±0,03	2,0±0,01	0,9	4
PacBio	>1000	>10	1,8	<2-2,2<	closest to 1	
Oxford Nanopore	>1000	>20				

(b)



	Length (kb)	
	up to	enrichment zone
CTAB	194	<48-194<
3GS	$>10^{4}$	

(c)

Yield (Qubit)	Purity		Lenght
	Absorbance	Nanodrop/ Qubit ratio	
+++	+++	+++	++

Figure A2: ***Chlorella vulgaris*** **control quality HMW DNA isolated with CTAB method. Evaluation of DNA yield, DNA purity, DNA length meeting with Pacbio and nanopore requirements. N=number of replicates.**

% reads mapping on chloroplast

	mean ± sd	N
V.vinifera **Total DNA**	8,4	1
V.vinifera **NIBTM nuclei isolation**	2,1±0,2	3
C.vulgaris **Total DNA**	7,1±5,5	2
C.vulgaris **NIBTM nuclei isolation**	2,79 ± 0,4	2

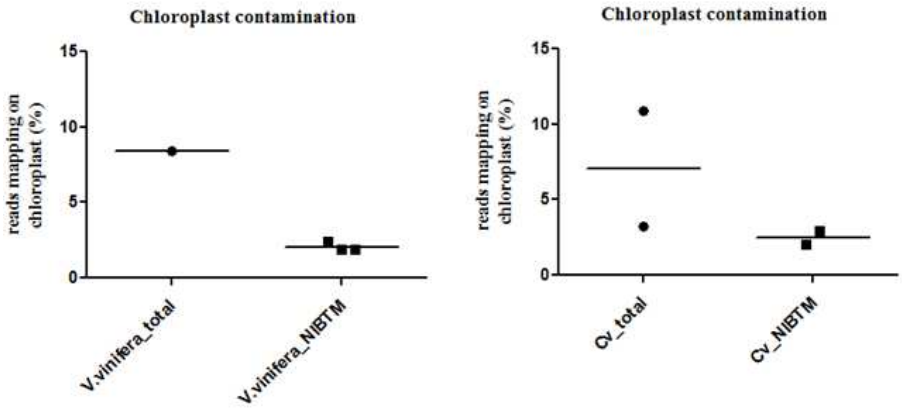
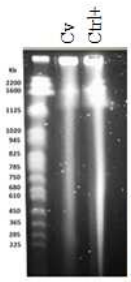


Table A2 and Figure A3: Percentage of reads mapping on chloroplast genome for ***V.vinifera*** **and** ***C.vulgaris*****. Evaluated by Illumina sequencing.**

(a)

	Yield		N
	Total ng	ng/ul	
BIONANO w/o FIX	2043±290	75±18	3
Bionano OM	>100	>35ng/ul	

(b)



	Length (Kp)	
	up to	enrichment zone
BIONANO w/o FIX	2200	<1600-2200<
Bionano OM	10^{-5}	

(c)

	Expected	Chlorella vulgaris
Expected Labels/100kbp (NtBspQ1)	>7,2	8,4
N50 (>150kb)	>150	349
Mol >150Kb	___	27%
False Positive (%)	<15%	0,97

Figure A4: Chlorella vulgaris control quality UHMW DNA isolated with Bionano buffer w/o fixation+Liquid Nitrogen protocol. (a) (b) Evaluation of DNA yield, DNA length meeting with OM requirements. (c): OM data generation on Irys platform after 30cycles. N=number of replicates.

REFERENCES

[1] L. Liu *et al.*, “Comparison of next-generation sequencing systems”, *Role Bioinforma. Agric.*, vol 2012, bll 1–25, 2014.

[2] M. O. Pollard, D. Gurdasani, A. J. Mentzer, T. Porter, en M. S. Sandhu, “Long reads: their purpose and place”, *Hum. Mol. Genet.*, vol 27, no R2, bll R234–R241, 2018.

[3] S. E. Levy en R. M. Myers, “Advancements in Next-Generation Sequencing”, *Annu. Rev. Genomics Hum. Genet.*, vol 17, no 1, bll 95–115, 2016.

[4] J. M. Heather en B. Chain, “The sequence of sequencers: The history of sequencing DNA”, *Genomics*, vol 107, no 1, bll 1–8, 2016.

[5] E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, en C. Thermes, “The Third Revolution in Sequencing Technology”, *Trends Genet.*, vol 34, no 9, bll 666–681, 2018.

[6] Glowicz, “乳鼠心肌提取 HHS Public Access”, *Physiol. Behav.*, vol 176, no 5, bll 139–148, 2017.

[7] E. E. Schadt, S. Turner, en A. Kasarskis, “A window into third-generation sequencing”, *Hum. Mol. Genet.*, vol 19, no R2, bll 227–240, 2010.

[8] D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, en H. Bayley, “Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore”, *Proc. Natl. Acad. Sci. U. S. A.*, vol 106, no 19, bll 7702–7707, 2009.

[9] S. KIYONO, “第1回 置かれた場所で咲こうとしました（1）”, *J. Japan Soc. Precis. Eng.*, vol 84, no 3, bll 250–250, 2018.

[10] S. Wei en Z. Williams, “Rapid short-read sequencing and aneuploidy detection using minION nanopore technology”, *Genetics*, vol 202, no 1, bll 37–44, 2016.

[11] M. Jain *et al.*, “Nanopore sequencing and assembly of a human genome with ultra-long reads”, *Nat. Biotechnol.*, vol 36, no 4, bll 338–345, 2018.

[12] M. Schalamun *et al.*, “Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from Eucalyptus pauciflora”, *Mol. Ecol. Resour.*, vol 19, no 1, bll 77–89, 2019.

[13] H. Tang, E. Lyons, en C. D. Town, “Optical mapping in plant comparative genomics”, *Gigascience*, vol 4, no 1, bll 1–6, 2015.

[14] Y. Nakayama, H. Yamaguchi, N. Einaga, en M. Esumi, “Pitfalls of DNA quantification using dnabinding fluorescent dyes and suggested solutions”, *PLoS One*, vol 11, no 3, bll 1–12, 2016.

[15] S. O. Rogers en A. J. Bendich, “Extraction of total cellular DNA from plants, algae and fungi”, *Plant Mol. Biol. Man.*, bll 183–190, 1994.

[16] E. Varela-Álvarez *et al.*, “Dinucleotide microsatellite markers in the genus Caulerpa”, *J. Appl. Phycol.*, vol 23, no 4, bll 715–719, 2011.

[17] M. J. Carter en I. D. Milton, “An inexpensive and simple method for DNA purifications on silica particles”, *Nucleic Acids Res.*, vol 21, no 4, bl 1044, 1993.

[18] M. Zhang, Y. Zhang, C. F. Scheuring, C. C. Wu, J. J. Dong, en H. Bin Zhang, “Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research”, *Nat. Protoc.*, vol 7, no 3, bll 467–478, 2012.

[19] Y. Kong, M. O’Neill, en G. Zhou, “Plant cell walls: Isolation and monosaccharide composition analysis”, *Methods Mol. Biol.*, vol 1744, bll 313–319, 2018.

[20] P. D. N. Hebert, “Semi-automated , Membrane-Based Protocol for DNA Isolation from Plants”, bll 186–198, 2008.

[21] Z. Hu, X. Zeng, A. Wang, C. Shi, en D. Duan, “An efficient method for DNA isolation from red algae”, bll 161–166, 2004.

[22] T. Maeda, T. Kawai, M. Nakaoka, en N. Yotsukura, “Effective DNA extraction method for fragment analysis using capillary sequencer of the kelp, Saccharina”, *J. Appl. Phycol.*, vol 25, no 1, bll 337–347, 2013.

[23] M. G. Murray en W. F. Thompson, “Nucleic Acids Research”, vol 8, no 19, bll 4321–4325, 1980.

[24] J. J. Doyle en J. L. Doyle, “Doyle_plantDNAextractCTAB_1987.pdf”. bll 11–15, 1987.

[25] M. A. Lodhi, G. N. Ye, N. F. Weeden, en B. I. Reisch, “A simple and efficient method for DNA extraction from grapevine cultivars and Vitis species”, *Plant Mol. Biol. Report.*, vol 12, no 1, bll 6–13, 1994.

[26] N. Mostafa, H. Omar, S. G. Tan, en S. Napis, “Studies on the genetic variation of the green unicellular alga haematococcus pluvialis (chlorophyceae) obtained from different geographical locations using ISSR and RAPD molecular marker”, *Molecules*, vol 16, no 3, bll 2599–2608, 2011.

[27] S. L. Dellaporta, J. Wood, en J. B. Hicks, “A plant DNA minipreparation: Version II”, *Plant Mol. Biol. Report.*, vol 1, no 4, bll 19–21, 1983.

[28] T. Jagielski, J. Gawor, Z. Bakuła, K. Zuchniewicz, I. Zak, en R. Gromadka, “An optimized method for high quality DNA extraction from microalga Prototheca wickerhamii for genome sequencing”, *Plant Methods*, vol 13, no 1, bll 1–8, 2017.

[29] M. Chomel *et al.*, “Plant secondary metabolites: a key driver of litter decomposition and soil nutrient cycling”, *J. Ecol.*, vol 104, no 6, bll 1527–1541,

2016.

[30] B. Mayjonade *et al.*, “Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules”, *Biotechniques*, vol 61, no 4, bll 203–205, 2016.

[31] A. Healey, A. Furtado, T. Cooper, en R. J. Henry, “Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species”, *Plant Methods*, vol 10, no 1, bll 1–8, 2014.

[32] N. Girollet, B. Rubio, C. Lopez-Roques, S. Valière, N. Ollat, en P.-F. Bert, “De novo phased assembly of the Vitis riparia grape genome”, *Sci. Data*, vol 6, no 1, bll 1–8, 2019.

[33] A. Arimoto *et al.*, “A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants”, *DNA Res.*, vol 26, no 2, bll 183–192, 2019.

[34] M. M. R. Shah, Y. Liang, J. J. Cheng, en M. Daroch, “Astaxanthin-producing green microalga Haematococcus pluvialis: From single cell to high value commercial products”, *Front. Plant Sci.*, vol 7, no APR2016, 2016.

[35] P. K. Saxena, L. C. Fowke, en J. King, “An efficient procedure for isolation of nuclei from plant protoplasts”, *Protoplasma*, vol 128, no 2–3, bll 184–189, 1985.

[36] K. A. Lutz, W. Wang, A. Zdepski, en T. P. Michael, “Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing”, *BMC Biotechnol.*, vol 11, no 1, bl 54, 2011.

[37] D. S. Luthe en R. Quatrano, “Transcription in Isolated Wheat Nuclei: I. ISOLATION OF NUCLEI AND ELIMINATION OF ENDOGENOUS RIBONUCLEASE ACTIVITY”, *Plant Physiol.*, vol 65, no 2, bll 305–308, 1980.

[38] A. Bolger *et al.*, “The genome of the stress-tolerant wild tomato species Solanum pennellii”, *Nat. Genet.*, vol 46, no 9, bll 1034–1038, 2014.

[39] G. Carrier *et al.*, “An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods”, *Am. J. Bot.*, vol 98, no 1, bll 15–17, 2011.

[40] L. Barchi *et al.*, “A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution”, *Sci. Rep.*, vol 9, no 1, bll 1–13, 2019.

[41] J. Loureiro, E. Rodriguez, J. Doležel, en C. Santos, “Comparison of four nuclear isolation buffers for plant DNA flow cytometry”, *Ann. Bot.*, vol 98, no 3, bll 679–689, 2006.

[42] A. Nabbi en K. Riabowol, “Isolation of nuclei”, *Cold Spring Harb. Protoc.*, vol

2015, no 8, bll 731–734, 2015.

[43] D. M. Webb en S. J. Knapp, “DNA extraction from a previously recalcitrant plant genus”, *Plant Mol. Biol. Report.*, vol 8, no 3, bll 180–185, 1990.

[44] W. Wray, P. M. Conn, en V. P. Wray, “Isolation of Nuclei Using Hexylene Glycol”, *Methods Cell Biol.*, vol 16, no C, bll 69–86, 1977.

[45] E. Wilson, “Construction and Analysis of a Variovorax Genomic DNA Library for use in Identifying Genes Related to Auxin Degradation in Arabidopsis thaliana”, فرهنـــــگ و زن, vol 1, no 4, bl 53, 1389.

[46] S. Deschamps *et al.*, “A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping”, *Nat. Commun.*, vol 9, no 1, bl 4844, 2018.

[47] M. Schalamun en B. Schwessinger, “DNA size selection (> 1kb) and clean up using an optimized SPRI beads mixture”, *Protocols.io*, bll 1–6, 2017.

[48] B. Schwessinger, “High quality DNA from Fungi for long read sequencing e.g. Pacbio”, *Protocols.io*, bll 1–6, 2016.

[49] K. Etebari, I. Filipovi, G. Ra, G. J. Devine, H. Tsatsia, en M. J. Furlong, “[short communication] Complete genome sequence of”, no 8, 2019.

[50] Q. Luo *et al.*, “Genome and Transcriptome Sequencing of the Astaxanthin-Producing Green Microalga, Haematococcus pluvialis”, *Genome Biol. Evol.*, vol 11, no 1, bll 166–173, 2019.

[51] H. Staňková *et al.*, “BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes”, *Plant Biotechnol. J.*, vol 14, no 7, bll 1523–1531, 2016.

[52] R. H. Japelaghi, R. Haddad, en G. A. Garoosi, “Rapid and efficient isolation of high quality nucleic acids from plant tissues rich in polyphenols and polysaccharides”, *Mol. Biotechnol.*, vol 49, no 2, bll 129–137, 2011.

[53] T. Klingström, E. Bongcam-Rudloff, en O. V. Pettersson, “A comprehensive model of DNA fragmentation for the preservation of High Molecular Weight DNA”, *bioRxiv*, bl 254276, 2018.

[54] J. Quick, “Ultra-long read sequencing protocol for RAD004”, *Protoc. Io.*, bll 1–16, 2018.

[55] D. Khumallambam, P. Kshetrimayum, S. Nandeibam, en S. Huidrom, “An efficient protocol for total DNA extraction from the members of order Zingiberales- suitable for diverse PCR based downstream applications”, *Springerplus*, bll 2–20, 2013.

[56] M. Cecchin *et al.*, " Chlorella vulgaris genome assembly and annotation reveals the molecular basis for metabolic acclimation to high light conditions ", *Plant J.*, bll 1–17, 2019.

[57] B. Vaillancourt en C. R. Buell, "High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing", *bioRxiv*, bl 783159, 2019.

[58] R. B. Brown en R. B. Brown, "E cient NGS ready gDNA from microalga of Queensland", bll 5–7.

[59] T. Sauvage, W. E. Schmidt, H. S. Yoon, V. J. Paul, en S. Fredericq, "Promising prospects of nanopore sequencing for algal hologenomics and structural variation discovery", *BMC Genomics*, vol 20, no 1, bl 850, 2019.

[60] M. H. W. Schmidt *et al.*, "De novo assembly of a new Solanum pennellii accession using nanopore sequencing", *Plant Cell*, vol 29, no 10, bll 2336–2348, 2017.

[61] K. J. Liu *et al.*, "RAPID HIGH MW DNA EXTRACTION FROM PLANT , INSECT , CELL AND TISSUE SAMPLES FOR LONG-READ SEQUENCING USING NANOBIND MAGNETIC DISKS Cells / Blood Ultra-Long Reads", bl 80.

[62] E. Günerken, E. D'Hondt, M. H. M. Eppink, L. Garcia-Gonzalez, K. Elst, en R. H. Wijffels, "Cell disruption for microalgae biorefineries", *Biotechnol. Adv.*, vol 33, no 2, bll 243–260, 2015.

[63] M. H. Rezadoost, M. Kordrostami, en H. H. Kumleh, "An efficient protocol for isolation of inhibitor-free nucleic acids even from recalcitrant plants", *3 Biotech*, vol 6, no 1, bll 1–7, 2016.

[64] L. Guo *et al.*, "Genome assembly of Nannochloropsis oceanica provides evidence of host nucleus overthrow by the symbiont nucleus during speciation", *Commun. Biol.*, vol 2, no 1, bll 1–12, 2019.

[65] F. Kraft en I. Kurth, "Long-read sequencing in human genetics", *Medizinische Genet.*, vol 31, no 2, bll 198–204, 2019.

[66] S. P. GIBBS, "Nuclear envelope-chloroplast relationships in algae.", *J. Cell Biol.*, vol 14, no 5, bll 433–444, 1962.

[67] M. Wayama, S. Ota, H. Matsuura, N. Nango, A. Hirata, en S. Kawano, "Three-Dimensional Ultrastructural Study of Oil and Astaxanthin Accumulation during Encystment in the Green Alga Haematococcus pluvialis", *PLoS One*, vol 8, no 1, bll 1–9, 2013.

ACKNOWLEDGEMENTS

Il primo ringraziamento va al mio capo, il Prof. Massimo Delledonne. Appena mi assunse mi promise che avrei raggiunto questo titolo se mi fossi impegnata. Se promette qualcosa la mantiene sempre, in quanto la sua determinazione lo porta ad esser un uomo di parola. Vorrei poi ringraziare il gruppo del Prof. Matteo Ballottari, che mi fornito tutto il materiale di cui avevo bisogno per portare avanti questo lavoro; soprattutto vorrei ringraziare Michela che mi ha sempre incoraggiato e fatto ridere anche quando i risultati erano ben lontani. Trai gruppi esterni vorrei ringraziare quello della Prof. Torriani, ed in particolare Veronica Gatto ed Ilenia Campedelli che mi hanno dato sempre piena disponibilità e supporto nell'utilizzo della PFGE. Ed ora, i ringraziamenti passano al mio numeroso gruppo. Vorrei partire ringraziando i "former members" che mi hanno affiancato di più, ossia: Maria Giovanna (la mia MG), la Cantaloni, la Griggio, Carla, Elisa, Alessia, Marianna, Anna e Ciano. Li ringrazio perché prima di abbandonarmi, ognuno di loro mi ha insegnato qualcosa, tanto da farmi diventare l'anziana del gruppo. Grazie ai miei "nuovi" colleghi wetlab (Cristina, Chiara, Marta, Elisabetta, Massimiliano, Valentina e Marzia V.-permettimi di inserirti qui) e bioinfo (Simonuccio, Denise, Stefania, Giulia e Salviati-mi permetta di inserirla qui), che mi sopportano ogni giorno. Un grazie particolare va a Barbara Giovannone che mi ha affiancato nella maggior parte di queste estrazioni senza mai mollare. Vorrei concludere l'elenco facendo un ringraziamento speciale a due "colleghi" che mi stanno particolarmente a cuore e con i quali ho condiviso e vissuto interamente questo percorso: Barbara e Luca. È importante confrontarti con qualcuno che vive le tue stesse emozioni e sa ascoltarti. Mia sorella Babaa, che grazie al suo Q.I. superiore alla media, alla sua perseveranza ed alla sua dedizione per quel che fa, riesce a farmi appassionare di più al nostro lavoro ed in questi anni è stata sempre presente anche fuori dal lab. Luca, che ha una pazienza infinita nel ripetermi le cose mille volte e non mi giudica mai se poi faccio di testa mia.

Poi, come sempre, vorrei ringraziare i miei amici: Ornella, Marika, Rossella, Ivan, M.Chiara, Debby, Anna, Mary, Gianna, Julia, Rita e Mimì… perché anche se io sono assente ai loro matrimoni ed a volte nelle loro vite, loro mi conserveranno

sempre un posto nell'album dei ricordi e si schiereranno sempre dalla mia parte. Vorrei ringraziare anche coloro che mi hanno accolto appena arrivata a Verona: la Polly, Cesare Valerio, Elena, Capasso, CC e Pier. Tutti loro hanno accettato i miei rifiuti per le uscite settimanali capendo che sono una "donna di scienza" cit.

E come in ogni traguardo, in ogni passo, in ogni respiro che faccio, vorrei dedicare un pensiero alla mia famiglia. Mamma, Papà e Giacomo che sono il mio scudo e la mia arma, la mia valvola di sfogo ed il mio silenziatore, il mio legame e la mia indipendenza… vorrei ringraziarli per quella che sono e per quello che fanno e faranno sempre per me. E ringrazio la mia famiglia, tutta: zii, zie, cugini e chi non c'è più ma mi guarda da lassù.

Ed infine, ma solo per farla sospirare, e perché le dovevo o il primo o l'ultimo nome, un immenso ringraziamento va a Marzia Rossato, la persona che mi ha rimproverato di più in questi anni. Colei che mi ha seguito in ogni mio esperimento anche agli orari più assurdi; colei che mi ha insegnato quasi tutto quello che so in questo settore; colei che quando ha capito le mie lacune iniziali ha avuto fiducia in me, seguendomi in ogni mio passo fino a fidarsi delle mie mani come se fossero le sue. Colei che, quando cado, cerca di attutirmi il colpo in tutti i modi riuscendo a vedere sempre il lato buono delle cose. Colei che lavora più di tutti e sta sempre un passo avanti a tutti, ma non te lo fa mai pesare. In un mondo ostile come oggi, avere una persona con un'onestà morale quale la sua mi ha invogliato a fare qualsiasi tipo di progetto (anche non prettamente legato alla mia tesi). La fiducia che lei ripone in me mi fa acquisire sicurezza nelle mie capacità sul "campo" e non, quindi GRAZIE.