The University of South Bohemia in České Budějovice

Faculty of Science

Optimizing soil nucleic acid extraction protocol to fit third-generation, single-strand sequencing technologies

Bachelor thesis

Dmytro Shumlianskyi

Advisor: Dr. Roey Angel

Co-supervisor: RNDr. Eva Petrová

České Budějovice 2023

Shumlianskyi, D., 2023: Optimizing soil nucleic acid extraction protocol to fit third-generation, single-strand sequencing technologies. Bc. Thesis, in English, 34 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

Nucleic acid extraction is one of the basic procedures in molecular biology, allowing isolating of RNA and DNA from biological samples. This process, known as total nucleic acid extraction, enables various downstream applications encompassing gene expression analysis, genotyping, sequencing, and pathogen detection. However, current methods for generating high-quality and high-molecular-weight nucleic acids have yet to keep pace with the requirements of modern third-generation sequencing methods. Therefore, we aim to improve the extraction protocol and create a fast, straightforward, and high-yield purification method for obtaining high-quality nucleic acids.

Acknowledgments

In this part, I wish to thank the head of our laboratory Roey Angel for the opportunity to become a member of his team and for his scientific guidance and valuable advice. I thank Eva Petrová for her help in designing my experiments and for her assistance during work in the laboratory.

Declaration

I declare that I am the author of this qualification thesis and that in writing it, I have used the sources and literature displayed in the list of used sources only.

České Budějovice, 14.12. 2023

Dmytro Shumlianskyi

Contents

1. Introduction

1.1 Basics of Nucleic Acid Extraction

Nucleic Acid (NA) isolation is one of the most basic procedures in molecular biology. The isolation of RNA and DNA simultaneously from a biological sample (hereafter referred to as total nucleic acids or TNA extraction) opens access to various subsequent applications, encompassing, including gene expression analysis, genotyping, sequencing, and pathogen detection, all carried out using a single sample. In 1869, the process of isolating nucleic acids was accomplished for the first time by Friedrich Miescher when he studied the chemical nature of white blood cell nuclei. He discovered a novel type of molecule from the nuclei of white blood cells, which he named "nuclein [1]."

Miescher initially extracted this substance from pus obtained from discarded surgical bandages. However, he later found that white blood cells from other sources, such as lymph nodes and the spleen, also contained nuclein. The isolate was determined by Miescher as a new type of molecule, with equal importance to proteins. Miescher also realized that it is an acid with a high molecular weight and phosphorus content [1] and that nuclein was resistant to most chemical treatments but could be dissolved in alkaline solutions. Later, in 1889, the term "nucleic acid" was coined by Richard Altmann, another German biochemist, to describe the same substance [2]. Miescher's discovery of nucleic acids was a crucial milestone in the development of modern genetics and biochemistry, and it paved the way for numerous breakthroughs in our understanding of the genetic code and its role in the development of living organisms.

Nucleic acid extraction can be generally divided into three steps, each of which can be optimized depending on the type of sample and the subsequent applications for which the nucleic acids will be used:

1) Cell lysis - mechanical or chemical destruction of tissues and cells
2) Separation - selectively precipitating, binding, and washing proteins, lipids, and other contaminants from the nucleic acids to remove them.
3) Recovery of DNA or RNA - purification the nucleic acids in water or a compatible buffer solution, ensuring their preservation without interfering with subsequent procedures.

Initially, scientists would prepare the solutions needed for NA extraction, but over time, commercial kits were developed to speed up and simplify the process.

1.2 DNA sequencing. Third generation sequencing

DNA sequencing is precisely determining the nucleotide sequence within a DNA molecule. In 1953, Watson and Crick elucidated the structure of DNA, drawing upon the essential insights from Rosalind Franklin's DNA crystallography and X-ray diffraction studies [3, 4]. However, the first molecule to be sequenced was tRNA – in 1965 by Robert Holley, and the RNA of bacteriophage MS2 later on [5, 6]. In 1970–1973, Wu, Padmanabhan, and their colleagues demonstrated that their method could be used to determine any DNA sequence by using synthetic site-specific primers [7]. Later, in 1977, Frederick Sanger developed this sequencing method and created one of the most common, sequencing methods, for decades to come: chain termination sequencing [8]. This method, known as Sanger sequencing, involves using the enzyme DNA polymerase to elongate the DNA strand, just like in a PCR reaction. The difference here, however is that fluorescently-labeled nucleotides are included in the mixture. When a fluorescently-labeled nucleotide is incorporated into the 3' end of a strand, the DNA polymerase cannot continue synthesizing beyond this point and the reaction is terminated for this strand. The result is a series of fragments of different lengths, each ending with a fluorescently labeled nucleotide. Then fragments are distributed accordingly to their length by acrylamide-gel or capillary electrophoresis, and the information about the 3' base is used to reconstruct the original sequence. This approach enables reading, on average, fragments with a length of 800 bases, with the potential for extension beyond 1000 bases [9, 10, 11]. However, while fully automated implementations of this method were the main instrument for the original sequencing of the human genome, it took a neat ten years and three billion dollars because of their main limitation: the small amounts of DNA that could be processed per one unit of time[15].

This method allows an average read length of 800 bases but may be extended to above 1000 bases [9,10,11]. Nevertheless, even though the primary tool for initially sequencing the human genome was fully automated implementations of this method, it required a substantial ten years and three billion dollars. This extended timeline and cost were primarily attributed to a significant limitation: it can process only small amounts of DNA at one time.

Then, in the mid-90s, new methods were developed. The so-called next-generation sequencing, second-generation sequencing, massively parallel sequencing and high-throughput sequencing.

These methods use miniaturized and parallelized platforms to sequence from millions to billions of short reads (50 to 400 bases each) simultaneously in a single run. The platforms exist in numerous variations in engineering configurations and sequencing chemistry. Nonetheless, they can be categorized based on their foundational detection chemistries into sequencing by ligation and sequencing by synthesis and then subdivided into proton detection, pyrosequencing, and reversible terminator methodologies..

However, all these approaches adhere to a common technical paradigm of conducting massively parallel sequencing through the utilization of spatially isolated, clonally amplified DNA templates or individual DNA molecules within a flow cell. These technologies have made it possible to perform sequencing on a larger scale.

In contrast, third-generation sequencing, also known as long-read sequencing, or single-strand sequencing is an emerging DNA sequencing technology that allows for reading much longer DNA strands than previous generations of sequencing technologies. Sequencing technologies with a different approach than the second generation were described as "third generation" by Erica Chek Hayden in 2009 [12]. In contrast to previous DNA-sequencing technologies, these technologies use single-molecule sequencing, which means they can directly read DNA strands without amplification or fragmentation.

Third-generation sequencing has several advantages over previous generations of sequencing technology. First, it can read much longer DNA sequences, which can be particularly useful for assembling complex genomes or identifying structural variations in a genome. Second, it can potentially provide more accurate information about DNA modifications, such as methylation, which can significantly impact gene expression and regulation. Finally, third-generation sequencing is much faster than other methods [13] and potentially cheaper. However, they have several limitations. First, DNA libraries require a large amounts of DNA since sequencing is done directly on each strand without amplification. Second, high-quality, non-degraded or sheered DNA is needed since degraded DNA significantly impairs the process. Third-generation sequencing is also fraught with problems associated with higher sequencing error rates and systematic errors [14]. Lastly, to sequence long reads of DNA, one first needs to be able to extract ample amounts of high molecular weight DNA.

Several companies offer third-generation sequencing technologies. However, currently only the sequencing platforms from Oxford Nanopore Technologies and Pacific Biosciences are commercially available. These technologies are still evolving, and researchers continue to explore

the potential applications and limitations of third-generation sequencing. A summary comparing the different sequencing generations can be found in Table 1.

Table. 1. Comparison of first, second, and third-generation sequencing [15]

	First generation	Second generation[a]	Third generation[a]
Fundamental technology	Size-separation of specifically end-labeled DNA fragments, produced by SBS or degradation	Wash-and-scan SBS	SBS, by degradation, or direct physical inspection of the DNA molecule
Resolution	Averaged across many copies of the DNA molecule being sequenced	Averaged across many copies of the DNA molecule being sequenced	Single-molecule resolution
Current raw read accuracy	High	High	Moderate
Current read length	Moderate (800–1000 bp)	Short, generally much shorter than Sanger sequencing	Long, 1000 bp and longer in commercial systems
Current throughput	Low	High	Moderate
Current cost	High cost per base Low cost per run	Low cost per base High cost per run	Low-to-moderate cost per base Low cost per run
RNA-sequencing method	cDNA sequencing	cDNA sequencing	Direct RNA sequencing and cDNA sequencing
Time from start of sequencing reaction to result	Hours	Days	Hours
Sample preparation	Moderately complex, PCR amplification not required	Complex, PCR amplification required	Ranges from complex to very simple depending on technology
Data analysis	Routine	Complex because of large data volumes and because short reads complicate assembly and alignment algorithms	Complex because of large data volumes and because technologies yield new types of information and new signal processing challenges
Primary results	Base calls with quality values	Base calls with quality values	Base calls with quality values, potentially other base information such as kinetics

1.3 Current state of DNA/RNA extraction methods

Modern extraction methods can be classified as methods that combine chemical and physical methods and those that only use chemistry; those that allow simultaneous extraction of DNA and RNA and those that extract DNA and RNA at different stages. The chemistry-pure methods usually are more relevant because they generate longer DNA fragments but they are not aggressive enough to lyse cells collected from the environment, especially from soils, where many of the cells are not actively growing and hence have a tough cell wall. Although the isolation of nucleic acids individually is generally easier due to a diverse array of well-optimized kits and methods, the concurrent extraction of DNA and RNA permits to obtain of more commensurate data, particularly when dealing with complex samples such as soils

As mentioned above, the development of third-generation sequencers led to the requirement to produce high-quality and high molecular weight NA. However, lab methods for generating such NA have fallen behind in development since the development of bead-beating based protocols and kits and do not fully meet modern needs. After all, current DNA extraction procedures and cell lysis also lead to DNA shearing, thereby limiting the sequencing read length [16]. The selected extraction technique has an impact on the eventual purity and quantity of nucleic acid, influencing subsequent processes. Consequently, it is crucial to employ efficient and

uncomplicated extraction and purification methods to generate substantial amounts of high-quality nucleic acids.

Unfortunately, achieving this goal is often hindered by the presence of inhibitory compounds. These widely prevalent yet inadequately comprehended substances are commonly found in various environments in different quantities within soils and are frequently categorized as either "humic and fulvic compounds" or "polyphenolic compounds." [17, 18, 19, 20]. However, this is not the only problem when developing extraction methods. Among others, pH, inorganic salt composition and concentration, and the number of cells in the sample, also significantly impact. This has led to the emergence of many new methods and kits from independent researchers [20, 21, 22, 23, 24] and large multinational companies, as well as numerous comparisons of such methods and kits [18, 25, 26, 27, 28, 29, 30, 31, 40]. In turn, this has led to the creation of numerous new protocols and kits by both independent researchers [20, 21, 22, 23, 24] and major multinational corporations. Additionally, there have been numerous comparisons of these methods and kits [18, 25, 26, 27, 28, 29, 30, 31, 40]. Despite thorough evaluations of methods with and without kits, no single approach has proven universally effective across all environmental conditions [29, 18, 25, 28]. Therefore, it is impossible to objectively determine the "best" method or kit for extraction, since one or another modification may offer better sample quality, but worse quantitative yield. [18, 27, 32, 28]. In summary, the problem of evaluation depends on the goal that the user sets before himself.

1.4 Advantages and disadvantages of using kits

Selecting a DNA extraction kit or protocol is crucial to achieving consistent results. Many previous studies have examined the composition of microbial taxonomic groups in soils and shown that unbiased DNA extraction kits and methods are necessary to obtain accurate results [28, 33, 34, 35, 36]. Because of this, some studies have recommended that many DNA extraction kits be tested for environmental soil samples at the beginning of the study [38,39]. However, it can be troublesome for some laboratories because of funding, time limitations, and other factors. Consequently, numerous studies have been conducted to compare the effectiveness of various extraction techniques as well as pre-fabricated commercial kits. In the example of one such research, it has been shown that commercial sets have similar characteristics to non-kit, non-optimized methods [40]. Some may show high results regarding the amount of extracted NA but

poor purity [40]. Studies also show that one type of kit can work with different efficiency depending on the nature of the sample [37, 40].

Three more advantages of labs protocols over kits:

1. Kits contain reagents that are patent-protected and do not disclose what they contain. Lab protocols are much more transparent
2. Because of the above, lab protocols are much easier to adapt and optimize for different samples
3. Kits are often contaminated with bacterial biomass. This is difficult to detect or control. [45]

1.5 Goals of the project

Considering the above, the aim of this project was to improve the current TNA extraction protocol for soil samples used in the Anaerobic and Molecular Microbiology lab at the Biology Centre CAS and create a fast, straightforward, and high-yield extraction method for obtaining high-quality samples and with the minimum amount of impurities.

Our criteria for successful modification was the ability to obtain high-quality, long fragments of DNA and RNA from our samples; however, we mainly focused on DNA. Quality was assessed as follows:

- Average length of the DNA fragments.
- DNA Integrity Number (DIN) parameter.
- The purity of the DNA as defined by the A260/A280 ratio.
- Concentrations of DNA in the extract.

2. Materials and Methods

A detailed description of the original protocol, the final version of the modified protocol, and a full list of the used chemicals and equipment can be found in the supplemental materials. Original protocol also described in Appendix 1: Original protocol for TNA extraction p. 32

2.1. Soils

Two soil samples were chosen because of their representation of very poor and very rich in soil organic matter (humic acids) and with significant difference in biomass amount per g of soil and were used to determine the quality and quantity of DNA and RNA obtained by co-extraction protocol. One sample was collected in Avdat, Negev Plateau, Israel, on a natural field (humics-poor soil), and the second was collected near Certovo Lake, Pilsen Region in the Czech Republic (humics-rich soil). For extraction, we used three samples of 0.25 g each of each soil for each treatment.

2.1.1. Soil Treatment

Seven consecutive trials were conducted, assessing various adaptations of the initial methodology. Figure 3 shows a scheme of the main steps investigated to develop an optimized protocol for co-extracting DNA and RNA from the soil.

The original protocol, a phenol-chloroform co-extraction protocol, was designed by Angel et al. [47] , based on two protocols published by Henckel et al. [42] and Griffiths et al. [22], with several critical modifications [41]. Table 2 presents modifications that were tested to improve this protocol.

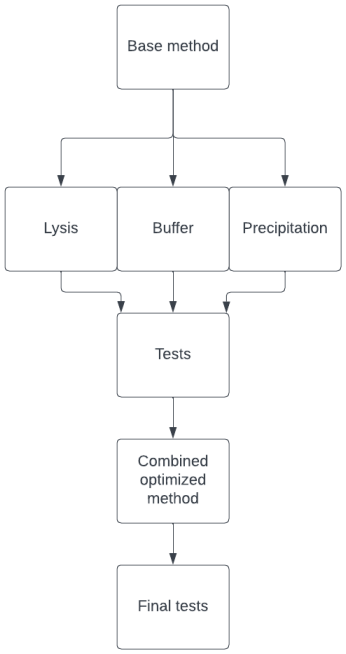

Fig. 2 Principal scheme of the work

Table 2 List of the modifications that were tested.

Test №	Sample	Weight [g]	Changes in the protocol	Name of the modified protocol
1-3	Avdat	0.25	Standard protocol	Standard
4-6	Certovo	0.25		
7-9	Avdat	0.25	2 h in ice and 30 min centrifugation instead of standard 1 h centrifugation for precipitation	Ice
10-12	Certovo	0.25		
13-15	Avdat	0.25	Sodium buffer (Na2HPO4/NaH2PO4 120 mM, pH 7.9) instead of standard potassium buffer (K2HPO4/KH2PO4 120 mM, pH 8.0)	Na
16-18	Certovo	0.25		
19-21	Avdat	0.25	Sodium buffer instead of standard buffer and TNS solution (TRIZMA 0.65 M, NaCl 0.1 M, SDS 0.347 M) instead TNC solution	Na TNS
22-24	Certovo	0.25		

			(TRIZMA 0.65 M, NaCl 0.1 M, CTAB 0.274 M)	
25-27	Avdat	0.25	PEG 5% instead of 30% PEG	Precipitation
28-30	Certovo	0.25		
31-33	Avdat	0.25	Vortex adapter for microcentrifuge instead of bead beater homogenizer and sodium buffer + TNS instead of standard buffer + TNC; 1-time extraction as in the prototype protocol; 33% lower volume of extraction and precipitation solutions	Adapter +TNS
34-36	Certovo	0.25		
37-39	Avdat	0.25	Vortex adapter instead of bead beater homogenizer + standard buffer + 2 h in ice and 30-minute centrifugation instead of standard 1 h centrifugation for precipitation	Adapter + Ice
40-42	Certovo	0.25		

2.1.2. Cell lysis

Cell lysis was performed in a 2 ml Lysing Matrix E tube MP Biomedicals. In the standard protocol, a FastPrep-24TM 5G bead beating grinder and lysis system were used, along with 375 µl of potassium phosphate buffer (120 mM pH 8.0), 125 µl of TNC, and 400 µl of TE-saturated phenol solution. Some modified versions used sodium phosphate buffer (120 mM, pH 7.9), which should prevent precipitation of some chemical or 125 µl of TNS instead of TNC, which should result in higher yields, but potentially lower DNA quailty. Lastly, we also tested the effect of lysing the cells with a lower mechanical force over an extended period of time using a or a tabletop vortex and an adapter for the microcentrifuge (15 min) instead of the more powerful bead-beating for 30 sec.

2.1.3. Separation

For separating the nucleic acids from the cell debris and proteins, we used a two-step extraction procedure, selective precipitation, and washing. The extraction and washing steps were not modified from the original protocol.

For the precipitation step in the original protocol, to each tube 2 µl RNA-grade glycogen and 1 ml PEG Precipitation Solution was added and the tube was then centrifugated at 14000 rpm (20817 RCF) at 4 °C, for 1 h. The PEG precipitation solution was prepared by dissolving 30 g of PEG (MW 7000-9000) and 9.35 g NaCl in 100 ml RNase-free water. For some of the modified versions, was prepared precipitation solution by dissolving 5 g of PEG (MW.7000–9000) 9.35 g NaCl in 100 ml RNase-free water, and a combination of 2 h of chilling in an ice bath and 30 minutes of centrifugation was used instead of the original 1 h centrifugation step.

2.1.4 Recovery

The recovery step was not modified and did not differ from the same step in the original protocol. For dissolving precipitated nucleic acid, was used 100 µl of Low-EDTA-TE RNase-free buffer (Tris-HCl 0.01 M and EDTA 0.0001 M).

2.2. Sample analysis

2.2.1 Purity of the sample (A260/A280 ratio)

DNA has a characteristic peak of absorbance at 260 nm, but to measure purity, measuring the absorbance at another wavelength also required. Typically ratio A260/280 is used. A ratio of 1.8 is generally accepted as “pure” for DNA. If the ratio is appreciably lower (≤1.6), it may indicate the presence of proteins, phenol, or other contaminants that absorb strongly at or near 280 nm [49]. To measure the absorption ratio at 260 and 280 nm, the NanoDrop One spectrophotometer from Thermo Fisher Scientific was used, with Low-EDTA TE buffer serving as the blank measurement.

After these measurements, samples 1–36 were purified using the OneStep PCR Inhibitor Removal KitZymo Research purification kit (Sigma Aldrich).

2.2.2. DNA concentration measurements

The DNA concentration in samples 1-36 was determined by measuring the fluorescence of the solution prepared from Quant-iT™ PicoGreen™ dsDNA Assay Kit Invitrogen - Thermo Fisher and 1 µl of ×10 diluted TNA sample. Fluorescence was measured using Synergy™ 2 Multi-Mode

Microplate Reader from BioTek® Instruments, Incorporated, using the following protocol: Quant-iT™ PicoGreen® dsDNA Quantification [43].
The DNA concentration in samples 37-42 was determined by measuring the fluorescence of solution prepared from the Qubit® assay kit Thermo Fisher and 2 µl of TNA sample. Fluorescence was measured using The Invitrogen Qubit 4 Fluorometer from Thermo Fisher.

2.2.3. DNA Integrity Number (DIN) and length of DNA fragments

DNA Integrity Number (DIN) determines the level of sample degradation as opposed to the classical gel electrophoresis method that cannot adequately determine the sample integrity [48]. To determine the DIN parameter and ascertain the length of DNA fragments, samples were sent for analysis via the 4150 TapeStation System from Agilent Technologies, Inc at the University of Vienna, Austria. The device also measures the DNA concentration in parallel.
For preliminary measuring of TNA fragments length, gel electrophoresis was used. To prepare the gel, 100 ml of a 0.5% solution of Agarose Broad Range was used (ROTI®Garose BioScience Grade in ROTIPHORESE®, ROTH). The gel was prepared according to manufacturer's instructions [45]. GeneRuler DNA Ladder Mix Thermo Scientific was used as a ladder from 100 to 10000 bp. The gel was loaded with 5 µl of each sample mixed with DNA Gel Loading Dye, Thermo Scientific. Electrophoresis was performed for 45 minutes at 110 volts.
Both methods were used because gel electrophoresis is a fast and easy method that can be used in our lab, but it is not very precise, and is limitated with measuring the length of the DNA fragments in terms of precision and length. Otherwise, analyzing samples by tape station allows obtaining precise data but requires sending samples to a third-party laboratory in Vienna, which is time and money consuming.

2.2.4. Droplet digital polymerase chain reaction

The third generation of polymerase chain reaction, droplet digital polymerase chain reaction (ddPCR), is a biotechnological refinement of conventional polymerase chain reaction methods that can be used to directly quantify and clonally amplify DNA [44]. ddPCR can provide accurate quantification of the total copy number of the 16S rRNA gene, the gene usually exploited for assessing total bacterial abundance in metagenomic DNA samples [45].
This method was used in this project to count the number of 16S gene copies in the extract to evaluate the efficiency of bacterial cell lysis. PCR solutions were prepared from 11 µl of Master

mix EvaGreen from Bio-Rad, 0.2 µl primer BAC338F and primer BAC805R, 8.6 µl PCR-grade water, and 2 µl of samples diluted to the order of magnitude 1 ng μl^{-1}. PCR program was: 95 ºC - 5 min, 95 ºC - 30 sec, 55 ºC - 30 sec, 60 ºC - 3 min, repeat 40 times, then 4 ºC - 5 min, 90 ºC - 5 min, 10 ºC - hold. Ramp rate 2ºC sec^{-1}. Dropplets were generated by Automated Droplet Generator from Bio-Rad, PCR cycles were provided by T100TM Thermo Cycler from Bio-Rad, and, measurements were performed in QX200TM Droplet Reader from Bio-Rad. Result number of copy were normalized according to dilution factor and recalculated for 1 µl of extract and represent in millions.

2.2.5. Statistical data analysis

The data were divided by soil type and analyzed by multiple two-way ANOVA tests with the following Tukey test, with consideration of the type of lysis method, type of buffer, and separation methods, with the aim of finding differences in concentration, purity of samples, and length of DNA fragments. Analyses were provided in RStudio 2023.06.0 Build 421. As the null hypothesis was set that the average values of all groups are equal to each other.

3. Results

3.1. General description

The summary results are presented in Table 3. Concentration for samples 37–42 was measured not with the help of PicoGreen™ but Qubit®, which is based on the same physical principle and uses comparable chemistry. Unfortunately, sample 24 was lost during kit purification.
Protocol modification of replacing 30% PEG with 5% PEG in samples 25–30 was considered a failure due to the poor DNA yields; therefore, data from these samples were not considered for statistical analysis.
Data obtained from samples 31–36 were not considered for the statistical analysis of DNA concentration because of significant losses caused by mistakes in the technique. However, these samples were considered for purity, DIN, and DNA fragment length during analysis.
Data obtained from samples 10 and 23 were excluded from analysis in case of ddPCR because of abnormally low numbers and considered as mistakes that occured during the ddPCR procedure.

NA concentrations measured with the TapeStation and PicoGreen™ and Qubit® were generally consistent but differed significantly in some low-purity samples, so it was decided to use the PicoGreen™ and Qubit® data for analysis as it is more convenient.

Table 3 The summary of the results

				Nano Drop	PicoGreen*	Tape Station			
№	Sample origin	Soil weight [g]	Extracted NA [ug]	A 260/280	The concentration of NA [ng/ul]	The concentration of NA [ng/ul]	DIN	Upper peak [bp]	ddPCR [10^6]
1	Avdat	0.25	14.558	1.90	145.58	145	5.5	5612	323.4
2	Avdat	0.25	10.599	1.56	105.99	105	5.6	5715	170.06
3	Avdat	0.25	11.418	1.90	114.18	113	5.6	5794	974.6
4	Certovo	0.25	11.327	1.34	113.27	117	5	6282	576.4
5	Certovo	0.25	35.263	1.29	352.63	353	5.2	6150	635.8
6	Certovo	0.25	26.026	1.32	260.26	259	5	5518	385
7	Avdat	0.25	12.783	1.92	127.83	127	5.7	6030	338.8
8	Avdat	0.25	8.278	1.90	82.78	105	5.8	5838	238.48
9	Avdat	0.25	11.373	1.92	113.73	113	5.7	5980	1542.2

10	Certovo	0.25	30.94	1.34	309.4	298	5	5280	4.18
11	Certovo	0.25	16.424	1.33	164.24	161	4.5	5174	875.6
12	Certovo	0.25	34.99	1.34	349.9	349	5.1	5628	1881
13	Avdat	0.25	14.33	1.83	143.3	142	5.5	5528	407
14	Avdat	0.25	9.734	1.80	97.34	96	5.4	5280	429
15	Avdat	0.25	11.691	1.82	116.91	116	5.5	5114	466.4
16	Certovo	0.25	19.928	1.33	199.28	202	4.4	4626	888.8
17	Certovo	0.25	20.747	1.33	207.47	210	4.6	5299	915.2
18	Certovo	0.25	23.614	1.33	236.14	238	4.3	4629	459.8
19	Avdat	0.25	10.189	1.70	101.89	101	5.6	5910	179.3
20	Avdat	0.25	4.092	1.73	40.92	96	5.4	5632	1053.8
21	Avdat	0.25	13.375	1.71	133.75	133	5.6	5916	1243
22	Certovo	0.25	31.168	1.29	311.68	32	4.9	5562	1181.4
23	Certovo	0.25	42.544	1.34	425.44	43	5.1	5979	5.72
24	Certovo	0.25	-	1.32	-	-	-	-	0
25	Avdat	0.25	0.5	1.40	5	1	3	2003	0

26	Avdat	0.25	0.145	1.39	1.45	0	2.3	1428	0
27	Avdat	0.25	0.132	1.45	1.32	0	2.3	1787	0
28	Certovo	0.25	1.634	1.29	16.34	1	4.4	4928	0
29	Certovo	0.25	1.998	1.30	19.98	2	4.8	5893	0
30	Certovo	0.25	1.179	1.30	11.79	1	4.3	4621	0
31	Avdat	0.25	3.877	1.56	38.77	39	6.4	10083	223.3
32	Avdat	0.25	2.807	1.59	28.07	28	6.5	9739	150.92
33	Avdat	0.25	2.507	1.52	25.07	25	6.5	9603	124.3
34	Certovo	0.25	14.831	1.35	148.31	15	5.8	9670	213.18
35	Certovo	0.25	11.373	1.35	113.73	11	5.7	9761	299.2
36	Certovo	0.25	12.283	1.35	122.83	12	5.9	9693	239.8
37	Avdat	0.25	9.91	1.729	99.1	99.1	6.6	12744	1018.6
38	Avdat	0.25	6.71	1.722	67.1	67.1	6.5	12646	241.78
39	Avdat	0.25	2.68	1.794	26.8	26.8	6.6	12114	315.04
40	Certovo	0.25	13.6	1.335	136	136	2	9113	327.36
41	Certovo	0.25	20.7	1.350	207	207	2.1	8813	358.6

4 2	Certov o	0.25	25.8	1.352	258	258	2.7	10247	415.8

3.2. DNA yields

The result of the statistical analysis are listed in Table 3 and Table 4 and visualized in Figure 4. Based on a statistical analysis of the acquired data, it is evident that no protocol modification resulted in a statistically significant difference in the amount of the extracted DNA from both types of soil. However, using the vortex instead of bead beating yielded lower DNA amounts.

Table 3 The result of statistical analysis of mass measurements of Avdat soil

Avdat soil	Df	F value	p	Mean value standard method **[µg]**	Mean value modified method **[µg]**	difference of means (standard - modified) **[µg]**
Type of buffer	1.000	0.204	0.661	9.81	10.57	-0.76
Lysis method	1.000	5.097	0.0646	11.04	6.43	4.60
Precipitation method	1.000	0.285	0.738	11.11	8.62	2.49
Detergent	1.000	1.082	0.430	10.34	9.22	1.12

Table 4 The result of statistical analysis of mass measurements of Certovo soil

Certovo soil	Df	F value	p	Mean value standard method **[µg]**	Mean value modified method **[µg]**	difference of means (standard - modified) **[µg]**
Type of buffer	1.000	0.631	0.448	23.90	27.60	-3.70
Lysis method	1.000	0.961	0.390	26.63	20.03	6.60
Precipitation method	1.000	0.226	0.761	26.33	23.74	2.58
Detergent	1.000	4.084	0.248	24.36	26.76	-3.4

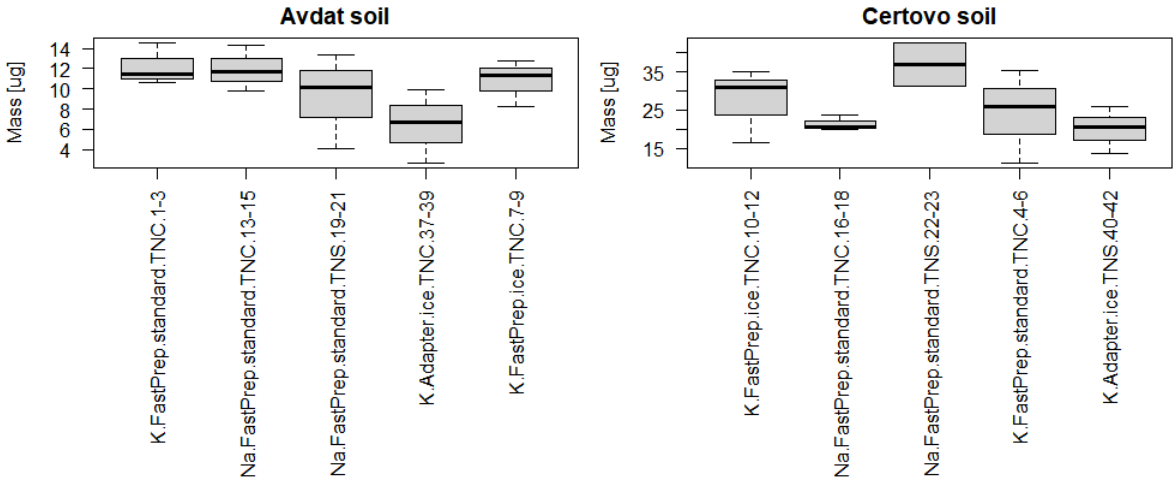

Fig. 4 Graph of the concentration of the sample depend on protocol version

3.3. Purity

Statistical analysis of the data shows that protocol modification did show statistical difference in the A260/280 ratio in extracts obtained from Avdat soil. Modifications of the lysis method and extraction with sodium buffer show some decrease in the sample purity. In the case of Certovo soil, only applying an adapter instead of bead beater homogenizer for cell lysing demonstrated a statistically significant but minor improvement but it remain bellow criteria of pure sample and need additional purification. Other modifacations did not show any statistically significant effect. The results of the statistical analysis are listed in Table 5 and Table 6 and visualized in Figure 5.

Table 5 The result of the statistical analysis of purity measurements of Avdat soil

Avdat soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	10.068	0.00734	1.82	1.70	0.12
Lysis method	1.000	15.091	0.00188	1.81	1.65	0.16
Precipitation method	1.000	5.221	0.155	1.72	1.83	-0.11
Detergent	1.000	2.708	0.301	1.82	1.64	0.18

Table 6 The result of the statistical analysis of purity measurements of Certovo soil

Certovo soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	0.003	0.955	1.33	1.33	0
Lysis method	1.000	8.951	0.0113	1.33	1.35	-0.02
Precipitation method	1.000	1.856	0.379	1.33	1.34	-0.01
Detergent	1.000	1.798	0.466	1.33	1.34	-0.01

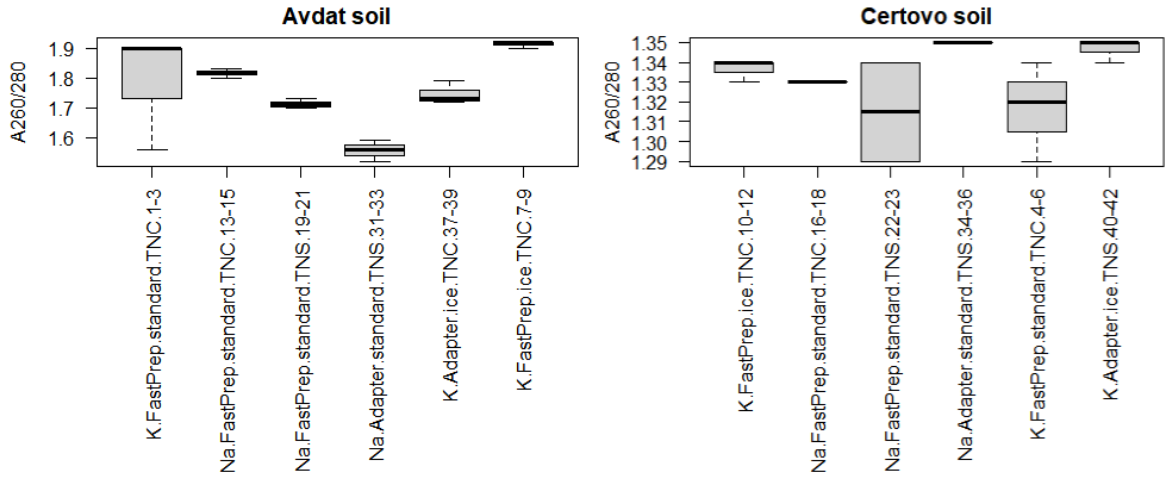

Fig. 5 Graph of the A260/280 of the sample depend on protocol version

3.4. DIN

The statistical analysis of the DNA integrity showed that none of the protocol modifications had any statistically significant difference in DIN parameters in the samples extracted from Certovo soil. However, in the case of Avdat soils, significant statistical differences in DNA integrity numbers occured. Modifying the lysis method to vortexing showed a significant increase of DNA quality obtained from Avdat, while application of sodium buffer showed some decrease in the DIN parameter. The results of the statistical analysis are listed in Table 7 and Table 8 and visualized in Figure 8.

Table 7 The result of the statistical analysis of DIN measurements of Avdat soil

Avdat soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	15.407	0.00174	5.96	5.82	0.14
Lysis method	1.000	683.115	0	5.58	6.52	-0.94
Precipitation method	1.000	5.571	0.142	5.76	6.15	-0.39
Detergent	1.000	2.774	0.295	5.83	6.00	-0.17

Table 8 The result of the statistical analysis of DIN measurements of Certovo soil

Certovo soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	4.269	0.061	4.07	5.09	-1.02
Lysis method	1.000	2.657	0.129	4.83	4.03	0.8
Precipitation method	1.000	2.586	0.303	5.08	3.57	1.51
Detergent	1.000	0.233	0.791	4.79	4.28	0.51

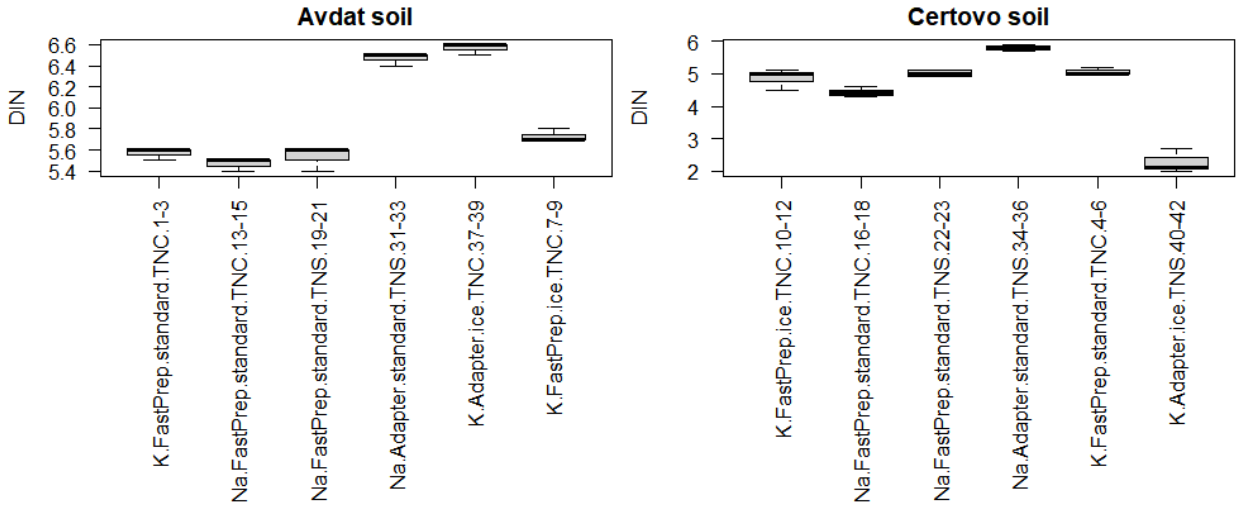

Fig. 6 Graph of the DIN parameter of the sample depend on protocol version

3.5. ddPCR Number of copy

Statistical analysis of the data showed that protocol modification did not have a statistical difference in the copy numbers of 16S gene in extracts obtained from Avdat soil in any protocol variation. In the case of Certovo soil, only applying an adapter instead of the bead beater homogenizer for cell lysing demonstrated a statistically significant and large magnitude decrease of numbers copy. Other modifications didn't show any statistically significant effect. The results of the statistical analysis are listed in Table 9 and Table 10 and visualized in Figure 7.

Table 9 The result of the statistical analysis of ddPRC measurements of Avdat soil

Table 10 The result of the statistical analysis of ddPRC measurements of Certovo soil

Avdat soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	0.212	0.649	573.66	475.22	98.44
Lysis method	1.000	1.434	0.252	613.84	345.66	268.18
Precipitation method	1.000	0.774	0.571	478.76	262.57	216.19
Detergent	1.000	0.631	0.631	538.78	495.77	43.01

Certovo soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	0.298	0.597	681.95	599.63	82.32
Lysis method	1.000	12.992	0.00485	866.56	308.99	557.57
Precipitation method	1.000	8.892	0.0715	579.46	771.67	-192.21
Detergent	1.000	0.922	0.667	827.20	433.62	393.58

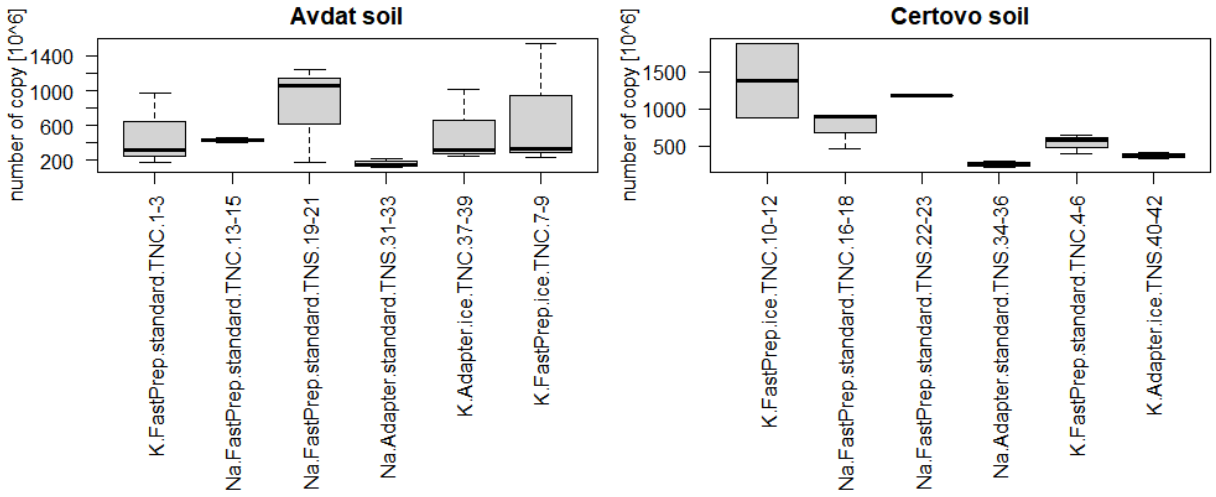

Fig. 7 Graph of the A260/280 of the sample depend on protocol version

3.6. Length of DNA fragments

The implication of the vortex adapter instead of bead beater allowed us to obtain DNA fragments almost twice as long, 9–12 thousand instead of 5–6 thousand base pairs, as in the original protocol in both Avdat and Certovo soil samples. In comparison, the application of sodium buffer showed a negative effect in Avdat soil samples and no statistically significant effect in Certovo soil samples. Modifying the precipitation step and type of detergent showed no statistically significant effect in both soils. The statistical analysis results are listed in Table 11 and Table 12 and visualized in Figure 8.

Figure 9 represents the result of gel electrophoresis of samples 1–12 and 37–42, which clearly illustrates the increase in the length of the DNA fragments obtained due to the modification of the

lysis method compared with the original protocol. Also, this image confirms that we extracted not only DNA fragments but RNA with is a double band characteristic for 16S ribosomal RNA at sizes of 2000 bp and 1500 bp.

Plots of the size distribution of DNA fragments obtained from Tape Station demonstrate that long-length DNA fragments are a significant fraction of the obtained extract rather than a narrow peak. The illustration of it can be seen in Figures 10 and 11 with graphs of DNA fragments length distribution for samples 1 and 37, respectively.

Table 11 The result of the statistical analysis of length of the DNA fragments measurements of Avdat soil

Avdat soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	12.375	0.00378	8052.56	6978.33	1074.22
Lysis method	1.000	284.087	0	5695.75	11154.83	-5459.08
Precipitation method	1.000	3.389	0.245	6660.50	9225.33	-2564.83
Detergent	1.000	0.068	0.867	7366.25	7813.83	-447.58

Table 12 The result of the statistical analysis of length of the DNA fragments measurements of Certovo soil

Certovo soil	Df	F value	p	Mean value standard method	Mean value modified method	difference of means (standard - modified)
Type of buffer	1.000	0.002	0.967	6911.67	6902.38	9.29
Lysis method	1.000	312.058	0	5466.09	9549.50	-4083.41
Precipitation method	1.000	4.714	0.171	6651.73	7375.83	-724.11
Detergent	1.000	3.257	0.331	5398,44	8604,75	-3206,31

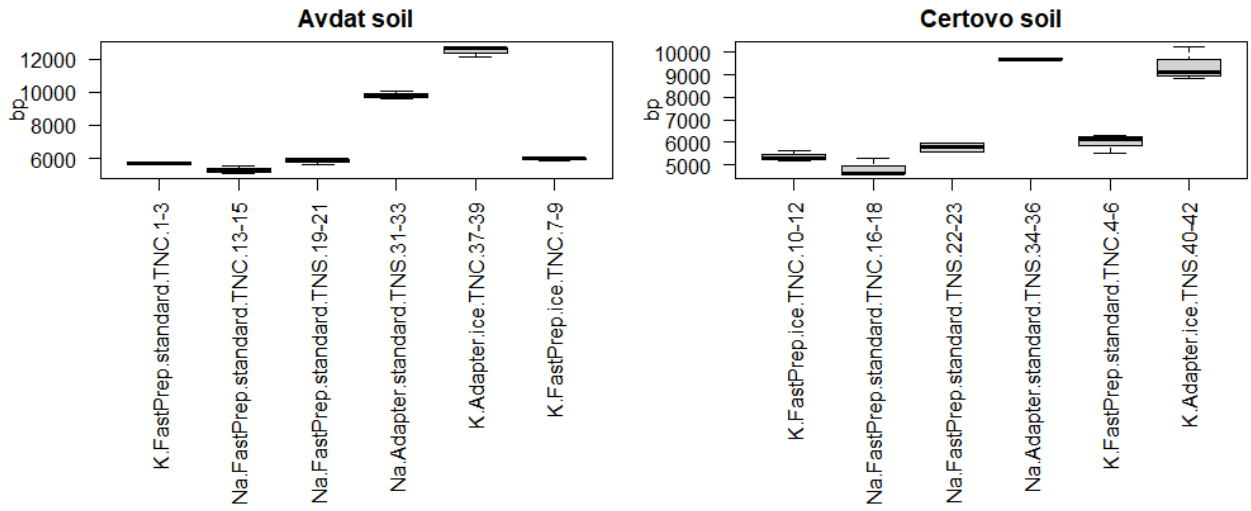

Fig. 8 Graph of the bp of the sample depend on protocol version

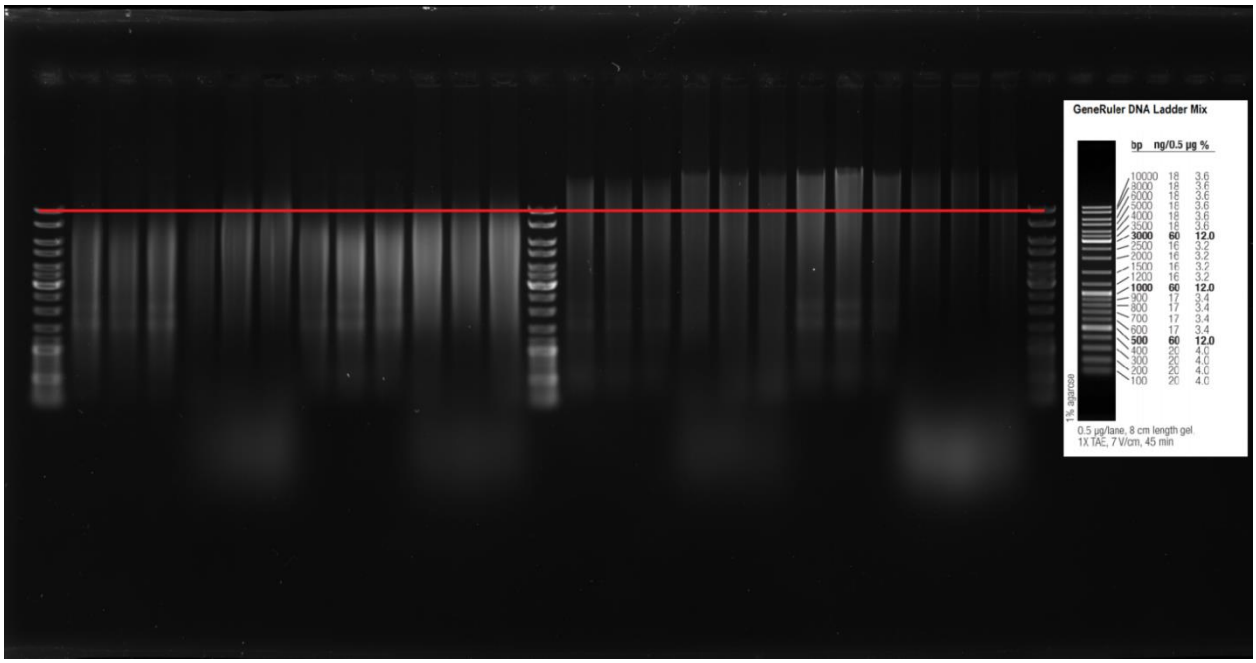

Fig. 9 Result of gel electrophoresis of samples 1–12 and 37–42

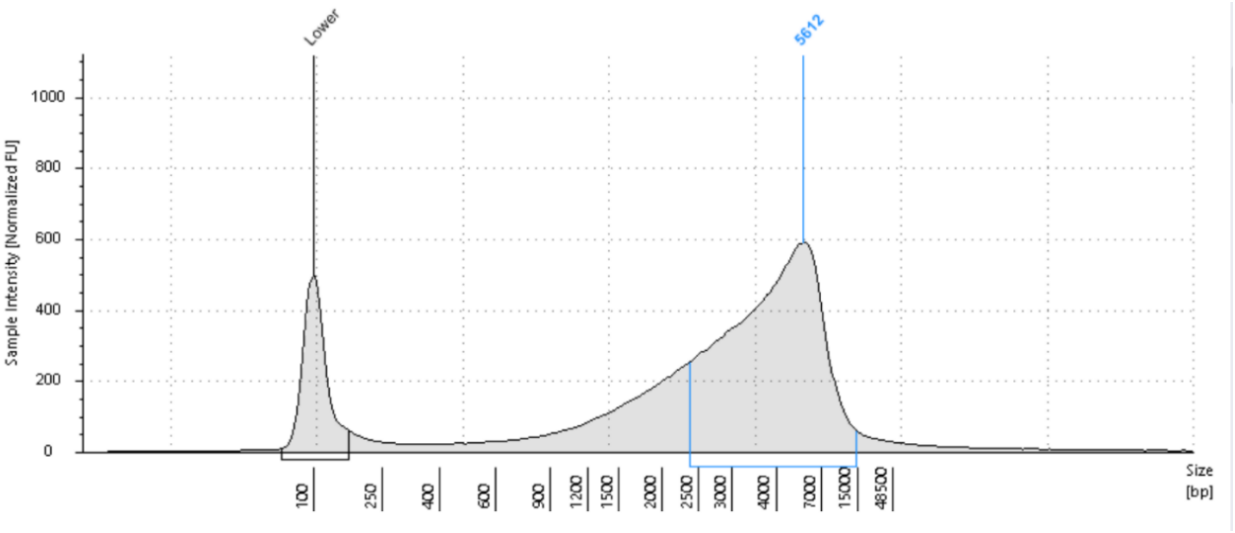

Fig. 10 Graph of DNA fragment length distribution for sample 1

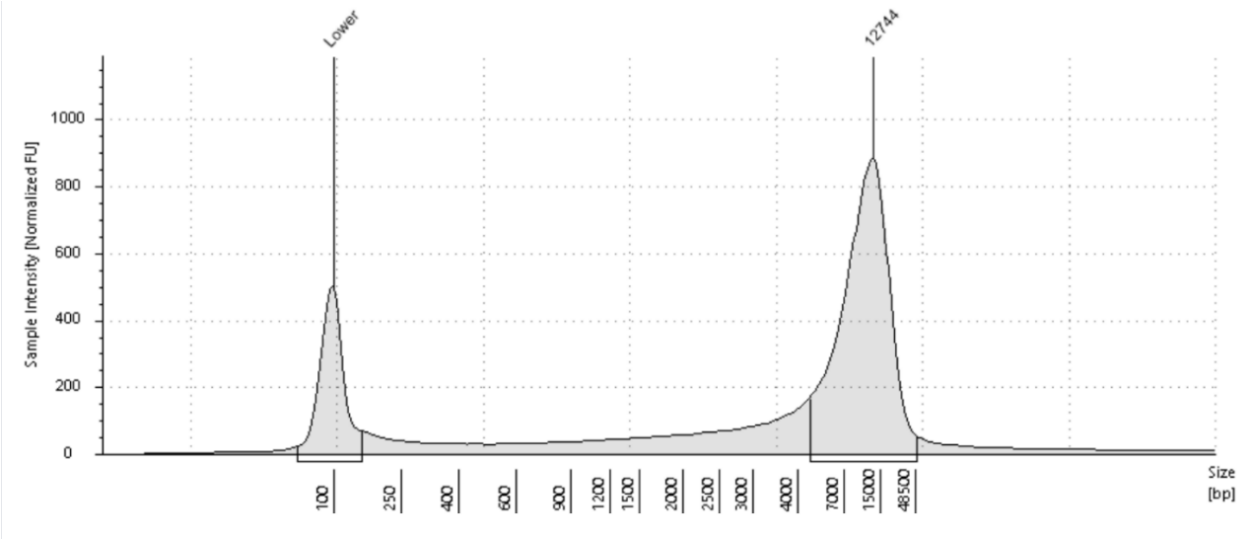

Fig. 11 Graph of DNA fragment length distribution for sample 37

4. Discussion

This work focused on improving our nucleic acids extraction protocol. We reviewed almost all the steps and tested the changes to create a more efficient method. From the results we obtained it appears that most of the modifications had an insignificant effect on the results. However, replacing the bead-beating step with vortexing resulted in a significant increase in the length of the extracted DNA fragments and, accordingly, the DIN parameter, with some decrease in the

purity. Kits using vortexing to lyse the cells are common. However, in all known cases the yields are typically much lower compared to bead-beating (data not shown). This is an important point since third generation sequencing methods also require high yields, in addition to high integrity and quality.

It can be seen that, as was known [17, 18, 19, 20], the purity and quality of the obtained samples largely depend on the amount of humic compounds in the soil samples. While extracts from soil samples poor in humic compounds from the Avdat in Negev desert mostly demonstrate adequate purity (parameter A 260/280 in the range of 1.7-2 [40 supplemental materials]) and parameter DIN at least 5.4, extracts obtained from soil samples from a neighborhood of Chertovo lake show significantly lower results, especially in the case of samples that did not undergo additional cleaning - samples 40 - 42. Even after cleaning, the obtained samples still show significantly lower DIN indicators than samples extracted similarly from Avdat soil. This indicates that one of the essential areas of further improvement of extraction methods is purifying the obtained extracts from humic compounds and other impurities. As one of the way for improved purity in samples obtained from humic reach source can be further optimization of the lysis method by using a adapter instead of bead beating, which improved the purity in case of Čertovo soil in contrast to Avdat where it acted in an opposite way.

Contrary to the previous studies, which showed that there is no difference between the use of vortex adapter and bead-beating ($p > 0.1$) [40 supplemental materials], we managed to achieve a significant increase in the length of the extracted fragments in the final version of the protocol, but with some decrease of DNA concentration, which however was not statistically significant but can be supported by significant decrease of 16S copy number in case of the Certovo soil. This can be explained by the fact that a more "gentle" method of mechanical leasing was used, which, although it did not shred the DNA to the same extent as the original method, could not open all the rather "tough" microbial cells. However, it is noteworthy that modification of the lysis and increases in the length of DNA fragments resulted in a corresponding elevation of the DIN index in samples derived from Avdat soil. In contrast, samples originating from Chertovo soil showed no statistically significant difference in the DIN parameter under these conditions, thereby warranting additional research.

Testing of the use of different types of buffers confirmed the data of previous studies[18] that changing the ionic strength of the buffer affects the quality of the obtained extracts. Nevertheless, it was unexpected that even such a minor change in the form of alteration of potassium cations with

sodium cations significantly affects the length of DNA fragments and the DIN parameter in case of Avdat soil. This indicates the need for further tests to select the optimal lysis buffer

Modifying the precipitation method by reducing the centrifugation time and adding a precipitation step during chilling in an ice bath showed no difference in yield or purity. Moreover, it made the procedure more time-consuming and challenging to perform, because the resulting NA pellets were sensitive to touch and did precipitate well on the inner surface of the low-binding tubes. Furthermore replacing TNC with TNS as a detergent agent in lysis did not show a statistically significant difference.

5. Conclusion

Even though we achieved the goals we set before starting the study only partially, we managed to obtain a significant increase in the length of extracted DNA fragments, without compromising quantity, quality or integrity of the DNA. All these are critical parameters in the context of nucleic acid extraction for third-generation sequencing. Future research should focus on improving purification methods, selecting a more effective buffer solution, and comparing the results obtained using third-generation sequencing.

6. List of literature

1. Thess, Andreas, Hoerr, Ingmar, Panah, Benyamin Yazdan, Jung, Günther and Dahm, Ralf. "Historic nucleic acids isolated by Friedrich Miescher contain RNA besides DNA" Biological Chemistry, vol. 402, no. 10, 2021, pp. 1179-1185. https://doi.org/10.1515/hsz-2021-0226
2. Ralf Dahm, Friedrich Miescher and the discovery of DNA, Developmental Biology, Volume 278, Issue 2,2005,Pages 274-288, ISSN 0012-1606, https://doi.org/10.1016/j.ydbio.2004.11.028.
3. WATSON, J., CRICK, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738 (1953). https://doi.org/10.1038/171737a0
4. FRANKLIN RE, GOSLING RG. Molecular configuration in sodium thymonucleate. Nature. 1953 Apr 25;171(4356):740-1. doi: 10.1038/171740a0.

5. HOLLEY RW, APGAR J, EVERETT GA, MADISON JT, MARQUISEE M, MERRILL SH, PENSWICK JR, ZAMIR A. STRUCTURE OF A RIBONUCLEIC ACID. SCIENCE. 1965 MAR 19;147(3664):1462-5. DOI: 10.1126/SCIENCE.147.3664.1462. PMID: 14263761.
6. Fiers, W., Contreras, R., Duerinck, F. *et al.* Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 260, 500–507 (1976). https://doi.org/10.1038/260500a0
7. R. Padmanabhan, Raji Padmanabhan, Ray Wu, Nucleotide sequence analysis of DNA: IX. Use of oligonucleotides of defined sequence as primers in DNA sequence analysis, Biochemical and Biophysical Research Communications, Volume 48, Issue 5, 1972, Pages 1295-1302, ISSN 0006-291X, https://doi.org/10.1016/0006-291X(72)90852-2.
8. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7. doi: 10.1073/pnas.74.12.5463
9. Hert DG, Fredlake CP, Barron AE. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. Electrophoresis. 2008 Dec;29(23):4618-26. doi: 10.1002/elps.200800456.
10. Schloss JA. How to get genomes at one ten-thousandth the cost. Nat Biotechnol. 2008 Oct;26(10):1113-5. doi: 10.1038/nbt1008-1113.
11. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. (2001) The sequence of the human genome. Science. 2001 Feb 16;291(5507):1304-51. doi: 10.1126/science.1058040. Erratum in: Science 2001 Jun 5;292(5523):1838.
12. Check Hayden, E. Genome sequencing: the third generation. *Nature* (2009). https://doi.org/10.1038/news.2009.86
13. Erwin L. van Dijk, Yan Jaszczyszyn, Delphine Naquin, Claude Thermes, The Third Revolution in Sequencing Technology, Trends in Genetics, Volume 34, Issue 9,2018, Pages 666-681, ISSN 0168-9525, https://doi.org/10.1016/j.tig.2018.05.008.
14. Xiao T, Zhou W. The third generation sequencing: the advanced approach to genetic diseases. Transl Pediatr. 2020 Apr;9(2):163-173. doi: 10.21037/tp.2020.03.06.
15. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet. 2010 Oct 15;19(R2):R227-40. doi: 10.1093/hmg/ddq416. Epub 2010 Sep 21. Erratum in: Hum Mol Genet. 2011 Feb 15;20(4):853.

16. Olson ND, Morrow JB. DNA extract characterization process for microbial detection methods development and validation. BMC Res Notes. 2012 Dec 3;5:668. doi: 10.1186/1756-0500-5-668. .
17. Tebbe CC, Vahjen W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl Environ Microbiol. 1993 Aug;59(8):2657-65. doi: 10.1128/aem.59.8.2657-2665.1993.
18. Krsek M, Wellington EM. Comparison of different methods for the isolation and purification of total community DNA from soil. J Microbiol Methods. 1999 Dec;39(1):1-16. doi: 10.1016/s0167-7012(99)00093-7.
19. Hirsch, P. R., Mauchline, T. H., and Clark, I. M. (2010). Culture-independent molecular techniques for soil microbial ecology. Soil Biol. Biochem. 42, 878–887. doi: 10.1016/j.soilbio.2010.02.019
20. Mettel, C., Kim, Y., Shrestha, P. M., and Liesack, W. (2010). Extraction of mRNA from Soil. Appl. Environ. Microbiol. 76, 5995–6000. doi: 10.1128/aem.03047-09
21. Purdy, K. J., Embley, T. M., Takii, S., and Nedwell, D. B. (1996). Rapid extraction of DNA and rRNA from sediments by a novel hydroxyapatite spin-column method. Appl. Environ. Microbiol. 62, 3905–3907.
22. Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., and Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl. Environ. Microbiol. 66, 5488–5491. doi: 10.1128/aem.66.12.5488-5491.2000
23. Peršoh, D., Theuerl, S., Buscot, F., and Rambold, G. (2008). Towards a universally adaptable method for quantitative extraction of high-purity nucleic acids from soil. J. Microbiol. Methods 75, 19–24. doi: 10.1016/j.mimet.2008.04.009
24. Lever, M. A., Torti, A., Eickenbusch, P., Michaud, A. B., Šantl-Temkiv, T., and Jørgensen, B. B. (2015). A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. Front. Microbiol. 6:476. doi: 10.3389/fmicb.2015.00476
25. LaMontagne, M. G., Michel, F. C., Holden, P. A., and Reddy, C. A. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. J. Microbiol. Methods 49, 255–264. doi: 10.1016/S0167-7012(01)00377-3Dineen et al., 2010;

26. Mahmoudi, N., Slater, G. F., and Fulthorpe, R. R. (2011). Comparison of commercial DNA extraction kits for isolation and purification of bacterial and eukaryotic DNA from PAH-contaminated soils. Can. J. Microbiol. 57, 623–628. doi: 10.1139/w11-049
27. Vishnivetskaya, T. A., Layton, A. C., Lau, M. C. Y., Chauhan, A., Cheng, K. R., Meyers, A. J., et al. (2014). Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. FEMS Microbiol. Ecol. 87, 217–230. doi: 10.1111/1574-6941.12219
28. Frostegård,Å., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., et al. (1999). Quantification of bias related to the extraction of DNA directly from soils. Appl. Environ. Microbiol. 65, 5409–5420.
29. LaMontagne, M. G., Michel, F. C., Holden, P. A., and Reddy, C. A. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. J. Microbiol. Methods 49, 255–264. doi: 10.1016/S0167-7012(01)00377-3
30. Mahmoudi, N., Slater, G. F., and Fulthorpe, R. R. (2011). Comparison of commercial DNA extraction kits for isolation and purification of bacterial and eukaryotic DNA from PAH-contaminated soils. Can. J. Microbiol. 57, 623–628. doi: 10.1139/w11-049Cruaud et al., 2014;
31. Shantelle Claassen, Elloise du Toit, Mamadou Kaba, Clinton Moodley, Heather J. Zar, Mark P. Nicol, A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples, Journal of Microbiological Methods, Volume 94, Issue 2, 2013, Pages 103-110, ISSN 0167-7012, https://doi.org/10.1016/j.mimet.2013.05.008.
32. Cruaud, P., Vigneron, A., Lucchetti-Miganeh, C., Ciron, P. E., Godfroy, A., and Cambon-Bonavita, M.-A. (2014). Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. Appl. Environ. Microbiol. 80, 4626–4639. doi: 10.1128/AEM.00592-14
33. Deiner, Kristy & Walser, Jean-Claude & Mächler, Elvira & Altermatt, Florian. (2014). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation. 10.1016/j.biocon.2014.11.018.

34. McOrist AL, Jackson M, Bird AR. A comparison of five methods for extraction of bacterial DNA from human faecal samples. J Microbiol Methods. 2002 Jul;50(2):131-9. doi: 10.1016/s0167-7012(02)00018-0. PMID: 11997164.;
35. Jun-ni Tang et al., An effective method for isolation of DNA from pig faeces and comparison of five different methods, Journal of Microbiological Methods, Volume 75, Issue 3,2008, Pages 432-436, ISSN 0167-7012,https://doi.org/10.1016/j.mimet.2008.07.014.
36. Trigodet F, Lolans K, Fogarty E, Shaiber A, Morrison HG, Barreiro L, Jabri B, Eren AM. High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. Mol Ecol Resour. 2022 Jul;22(5):1786-1802. doi: 10.1111/1755-0998.13588.
37. Trigodet, Florian & Lolans, Karen & Fogarty, Emily & Shaiber, Alon & Morrison, Hilary & Barreiro, Luis & Jabri, Bana & Eren, A.. (2022). High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. Molecular Ecology Resources. 22. 10.1111/1755-0998.13588.
38. Lang, Jenna & Darling, Aaron & Eisen, Jonathan. (2010). Morgan JL, Darling AE, Eisen JA. Metagenomic sequencing of an in vitro-simulated microbial community. PLoS One 5: e10209. PloS one. 5. e10209. 10.1371/journal.pone.0010209.
39. Zielińska, S, Radkowski, P, Blendowska, A, Ludwig-Gałęzowska, A, Łoś, JM, Łoś, M. The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. MicrobiologyOpen. 2017; 6:e453. https://doi.org/10.1002/mbo3.453
40. Natalie Y. N. Lim, Constance A. Roco1 and Åsa Frostegård Transparent DNA/RNA Co-extraction Workflow Protocol Suitable for Inhibitor-Rich Environmental Samples That Focuses on Complete DNA Removal for Transcriptomic Analyses Front. Microbiol., 18 October 2016 Sec. Terrestrial Microbiology Volume 7 - 2016 https://doi.org/10.3389/fmicb.2016.01588
41. Roey Angel, Eva Petrova, Ana Lara-Rodriguez Total Nucleic Acids Extraction from Soil V.6 dx.doi.org/10.17504/protocols.io.bwxcpfiw
42. Henckel T, Friedrich M, Conrad R. Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. Appl Environ Microbiol. 1999 May;65(5):1980-90. doi: 10.1128/AEM.65.5.1980-1990.1999.

43. Roey Angel, Eva Petrova, Ana Lara-Rodriguez Qant-iT™ PicoGreen® dsDNA Quantification V.1 https://dx.doi.org/10.17504/protocols.io.qdrds56
44. Li H, Bai R, Zhao Z, Tao L, Ma M, Ji Z, Jian M, Ding Z, Dai X, Bao F, Liu A. Application of droplet digital PCR to detect the pathogens of infectious diseases. Biosci Rep. 2018 Nov 15;38(6):BSR20181170. doi: 10.1042/BSR20181170.
45. Manzari C, Oranger A, Fosso B, Piancone E, Pesole G, D'Erchia AM. Accurate quantification of bacterial abundance in metagenomic DNAs accounting for variable DNA integrity levels. Microb Genom. 2020 Oct;6(10):mgen000417. doi: 10.1099/mgen.0.000417.
46. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014 Nov 12;12:87. doi: 10.1186/s12915-014-0087-z.
47. Angel R, Claus P, Conrad R. Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. ISME J. 2012;6:847–62.
48. Kong, Nguyet & Ng, Whitney & Cai, Lucy & Leonardo, Alvin & Weimer, Bart & Kelly, Lenore. (2014). Integrating the DNA Integrity Number (DIN) to Assess Genomic DNA (gDNA) Quality Control Using the Agilent 2200 TapeStation System. doi: 10.13140/RG.2.1.3616.8409.
49. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R. DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreserv Biobank. 2016 Aug;14(4):264-70. doi: 10.1089/bio.2015.0064.

Appendix 1: Original protocol for TNA extraction

1 Prepare the following solutions for TNA extraction. Use clean and preferably baked glassware (make sure all non-glass components can withstand the high temperatures).

1.1 Phosphate buffer

Phosphate buffer ([M] 120 millimolar (mM) pH 8; 12.43 g K2HP04.3 H20, 0.751 g KH2P04, 500 mL RNase-free water. Dissolve the salts in RNase-free water and fill up to 500 ml. Autoclave. Store at Room temperature

1.2 TNC

I 15.76 g TRIZMA, 1.17 g NaCI, 20 g CTAB, 200 mL RNase-free water. Dissolve the salts in RNase-free water and fill up to 200 ml. Autoclave. Store at Room temperature.

1.3 PEG precipitation solution (30%)

60 g PEG (M.W 7000-9000), 18.7 g NaCI, 200 mL RNase-free water. Add ingredients to a graduated Duran bottle. Add water to fill up to 200 ml, shake vigorously by hand, autoclave and mix well while hot (solution turns milky when hot, but then turns clear when cooled to room temperature). Store at Room temperature

1.4 Molecular-grade ethanol solution (75%)

75 mL Absolute ethanol, 25 mL RNase-free water. Store at -20°C

1.5 Low-EDTA TE buffer

500 uL Tris-HCI 1 M, pH 8.0 (Trizma), 10 uL EDTA 0.5 M, pH 8.0, 50 mL RNase-free water Prepare in a laminar-flow hood (to protect stocks), filter sterilise (0.2 um) and autoclave. Store at Room temperature.

2 Weigh 0.25g of soil (0.2-0.7 g) into a Lysing Matrix E tube

3 Add 375 µl of PB pH 8.0, 125 µl of TNC, and 400 µl of TE-saturated phenol.

4 Immediately place the tube in a sample homogeniser and process for 00:00:30 at 6.5 ms-1. We recommend using the FastPrep-24™ sample homogenizer with the CoolPrep™ adapter for 24 x 2 mlOn ice (dry ice).

5 Centrifuge at 14000 rpm, 15°C, 00:03:00 (centrifugation at RT is also possible).

6 Transfer the entire liquid (aqueous and organic phases) to a fresh 2 ml tube by decanting or pipetting.

7 Repeat Steps 3-7 two more times using the same Lysing Matrix E tube. Be careful not to overfill the tube as this might cause phenol leakage during the sample homogenising process. If there is not enough space in the tube for all the reagents, decrease the phenol volume (down to

200 µl). For the second homogenising repetition use a fresh 2 ml tube to collect the supernatant in Step 7, while for the third repetition divide the supernatant between the two tubes from the previous repetitions to achieve equal volumes.

8 Add 800 uL phenol/chloroform/isoamyl alcohol 25:24:1 (or 1 volume) to each of the tubes containing the supernatant, to a maximum of 2 ml total volume in each tube.

9 Mix the two phases , by hand or using a vortex. Centrifuge at 14000 rpm, 15°C, 00:03:00

10 Using a 1-ml pipette tip, carefully transfer the aqueous phase (the upper phase) from each tube to two fresh 2 ml tubes.

Be careful not to touch or pipette the interphase or the organic phase (the lower phase)

11 Add 800 uL chloroform/isoamyl alcohol 24:1 (or 1 volume) to each tube.

12 Mix the phases vigorously by hand or using a vortex. Centrifuge 14000 rpm, 15°C, 00:03:00

13 Carefully transfer the supernatant from each tube to a fresh 2 ml low-binding microcentrifuge tubes. At this point, you should have two low-binding tubes per sample.

14 Add to each tube I 2 uL RNA-grade glycogen and 1 mL PEG Precipitation Solution or twice th extract's volume.

15 Centrifuge at 14000 rpm, 4C, 01:00:00

A pellet should be visible at the bottom of the LoBind tube after centrifugation. The pellet should be white/opaque in colour. The size of the pellet will depend on the TNA content in a sample, but also on the amount of co-extracted contaminants.

16 Decant the supernatant, briefly centrifuge or spin-down again to collect the drops and using a pipette, remove as much as possible from the remaining precipitation solution. Be careful not to disturb the pellet.

17 Add 1 mL ice-cold 75% EtOH, Centrifuge at 14000 rpm, 4°C, 00:20:00

18 Remove the supernatant using a pipette, shortly centrifuge again to collect the drops and using a pipette remove as much as possible from the remaining ethanol. Be careful not to disturb the pellet.

19 Resuspend each pellet in 50 uL to 100 uL Low-EDTA TE Buffer and combine both subsamples into one of the non-stick tubes.

20 Remove co-extracted humic substances using OneStep™ PCR Inhibitor Removal Kit

20.1 Resuspend the resin in the column by brief vortexing (if the column is dry, add 100 µl of RNAse-free water prior to vortexing.)

20. 2 Loosen the cap by a quarter twist, place the column in a collection tube and centrifuge 8000 x g, Room temperature, 00:03:00

20.3 Place the column in a fresh 1.5 ml tube, pipette the entire TNA extract (up to 200 µl) on top of the resin

20.4 Centrifuge 8000 x g, Room temperature, 00:01:00

20.5 Discard the column and retain the extract in the tube.