The University of South Bohemia in České Budějovice
Faculty of Science

Optimizing soil DNA extraction protocols to fit 3rd generation, single-strand sequencing technologies

Bachelor thesis

Dmytro Shumlianskyi

Supervisor: Dr. rer. nat. Roey Angel

České Budějovice 2023

Shumlianskyi D.: Optimizing soil DNA extraction protocols to fit 3rd generation, single-strand sequencing technologies. Bc. Thesis, in English 31 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 extraction and create a fast, straightforward, and high-yield purification method for obtaining high-quality nucleic acids.

Acknowledgments

In this part, I want to express my gratitude to Dr. rer. nat. Roey Angel, the head of our laboratory, for affording me the privilege to join as a team member and for his scholarly mentorship and invaluable counsel. My appreciation extends to RNDr. Eva Petrová, for her contributions to the formulation of experimental methodologies and her indispensable aid throughout laboratory endeavors.

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, 17.08. 2023

Dmytro Shumlianskyi

Contents

1. Introduction

1.1 Basics of Nucleic Acid Extraction

Nucleic Acid (NA) isolation is one of the fundamental processes in molecular biology. The isolation of RNA and DNA simultaneously from a biological sample (hereafter, TNA extraction) opens the path to a wide range of downstream applications, 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."

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. Isolate was determined by Miescher as a new type of molecule equal in importance to proteins.

He 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 - the destruction of cells and tissues
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 - resuspending the nucleic acids in water or a compatible buffer solution, ensuring their preservation without interfering with subsequent procedures.

Initially, scientists would prepare their solutions for TNA 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, R 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, for decades, sequencing methods: chain termination sequencing [8].

This method, known as Sanger sequencing, involves using the enzyme DNA polymerase to elongate the DNA strand by incorporating fluorescently labeled nucleotides that stop the chain from elongating at different positions, forming a series of fragments of different lengths, each ending with a fluorescently labeled nucleotide. Then all the fragments are distributed in the order of their length via gel or capillary electrophoresis, and information about the last base is used to discover the original sequence. This method allows an average read length of 800 bases but may be extended to above 1000 bases [9–11]. 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 chief limitation: the small amounts of DNA that could be processed per one unit of time.

Then, in the mid-90s, new methods were developed. So-called next-generation sequencing, massively parallel sequencing, or 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.
Platforms differ in terms of engineering configurations and sequencing chemistry. However, they can be divided according to their underlying detection chemistries, including sequencing by ligation and sequencing by synthesis, further divided into proton detection, pyrosequencing, and reversible terminator.
However, all these methods share the same technical paradigm of massively parallel sequencing using spatially separated, clonally amplified DNA templates or individual DNA molecules in 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, is an emerging DNA sequencing technology that allows for reading much longer DNA sequences 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].

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 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 high DNA concentration since sequencing is done directly on each strand without amplification. Second, fresh material or

intact cells are needed since degraded DNA significantly impairs the process. Third-generation sequencing is 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, including Oxford Nanopore Technologies and Pacific Biosciences. These technologies are still evolving, and researchers continue to explore the potential applications and limitations of third-generation sequencing.

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

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

1.3 Current state of DNA/RNA extraction

Modern extraction methods can be divided into two groups: those that allow simultaneous extraction of DNA and RNA and those that extract DNA and RNA at different stages. While isolating nucleic acids separately is generally easier due to the wide variety of highly optimized kits and methods, joint DNA and RNA extraction allows for more comparable data, especially from highly heterogeneous 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 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 choice of extraction method affects the subsequent purity and yield of the nucleic acid, which in turn affects downstream processes. Therefore, fast and straightforward extraction and purification methods are required to produce high-quality nucleic acids in large quantities.

However, this is usually not possible due to the presence of inhibitory compounds. These well-known but still poorly understood compounds are widespread in most environments. They are in varying amounts in soils and are often classified as "humic and fulvic compounds or polyphenolic compounds" [17–20]. 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 [20–24] and large multinational companies, as well as numerous comparisons of such methods and kits [18, 25–31, 41]. Despite extensive testing of methods with and without kits, no single method has been found to work for all types of environments [29, 18, 25, 28], and the "best" method is often difficult to determine, as one kit or reagent may provide, for example, better replication or quantity but at the expense of quality [18, 27, 32, 28].

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, 33–37]. Because of this, some studies have recommended that many DNA extraction kits be

tested for environmental soil samples at the beginning of the study [39–40]. 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-prepared commercial kits. In the example of one such research, it can be shown that commercial sets have similar characteristics to non-kit, non-optimized methods [41]. Some may show high results regarding the amount of extracted NA but poor purity [41]. Studies also show that one type of kit can work with different efficiency depending on the nature of the sample [38, 41].

Method/Kit	Nicolaisen's method[a]			MP[b]		PM[b]		PS + AK[b]			Optimized method		
TNA purification prior to digest[c]	−			+		−		+			+		
Soils tested	FH	FL	Å	FH	FL	FH	FL	FH	FL	Å	FH	FL	Å
Amplifiable DNA[d]	+	+	+	+	+	+	+	+	±[f]	+	+	+	+
Complete removal of DNA after 1st digestion[d,e]	+	−	−	+	−	+	−	+	−	−	+	+	−
Complete removal of DNA after 2nd digestion[d,e]	+	−	−	+	±[f]	+	−	+	−	+	+	+	+
cDNA synthesis	+	NT	NT	+	±[f]	+	NT	+	NT	+	+	+	+

[a] *Method from Nicolaisen et al. (2008).*
[b] *See **Table 1** for list of kit abbreviations.*
[c] *TNA purification with the OPIR kit.*
[d] *See text for details on DNA amplification and removal assessment.*
[e] *DNA was digested with TURBO DNase, and RCC kit was used for purification after each digestion.*
[f] *Results from replicates varied, likely due to the presence of inhibitory compounds.*
gDNA, genomic DNA; NT, not tested because of residual gDNA.

Fig. 2. Comparison of DNA and RNA co-extraction methods and kits tested on soils FH [high pH peat, pH 7.39], FL [low pH peat, pH 3.65], and Å [low pH clay soil, pH 5.5]. [41]

1.5 Goal of the thesis

Therefore, we aim to improve the extraction process and create a fast, straightforward, and high-yield extraction method for obtaining high-quality nucleic acids.

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.

2.1. Soils

Two soil samples were chosen because of their representation of very poor and very rich for humic acids soil and were used to determine the quality of DNA and RNA from co-extraction reactions. One sample was collected in Avdat, Negev Plateau, Israel, on a natural field, and the second was collected near Certovo Lake, Pilsen Region in the Czech Republic. For extraction was used 0.25g of the soil samples.

2.1.1. Soil Treatment

Several successive experiments were performed where different modifications of the original method were tested. Figure 3 illustrates a schematic diagram of the main steps investigated to develop an optimized protocol for co-extracting DNA and RNA from the soil. Our criteria for successful modification was the ability to obtain high-quality DNA and RNA from our samples; however, we mainly focused on DNA. Quality was assessed as follows:

- Number of base pairs in extracted DNA fragments.
- DNA Integrity Number (DIN) parameter.
- The purity of the sample is defined by the A260/A280 ratio.
- Concentrations of DNA in the extract.

The original protocol, a phenol-chloroform co-extraction protocol, was designed based on two protocols published by Henckel et al. (1999) [43] and Griffiths et al. (2000) [22], with several critical modifications [42]. Table 1 presents modifications that were tested to improve this protocol.

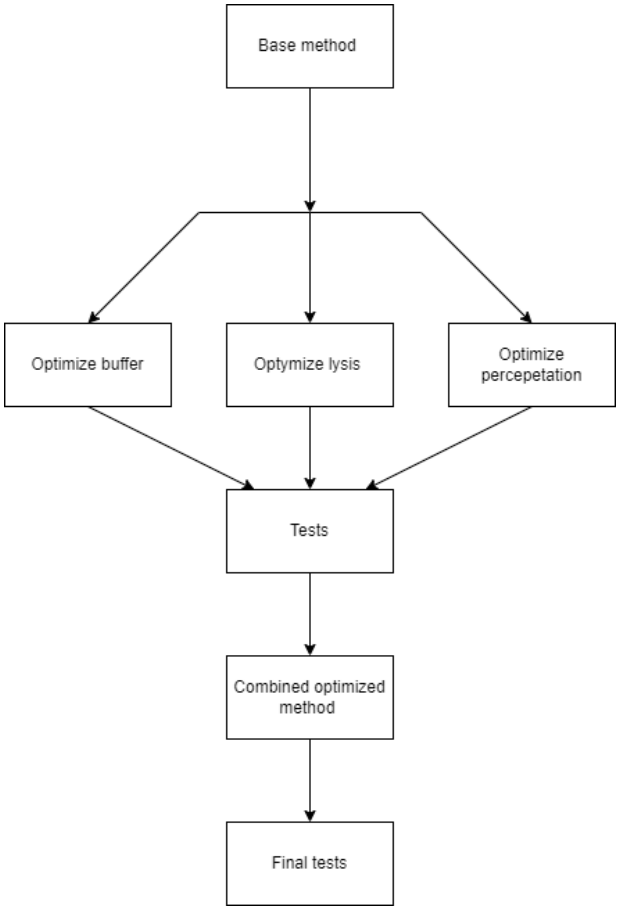

Fig. 3

Table 1

№ sample	Sample	Weight [g]	Changes in the protocol
1	Avdat soil	0.25	Standard protocol
2	Avdat soil	0.25	
3	Avdat soil	0.25	
4	Certovo soil	0.25	
5	Certovo soil	0.25	
6	Certovo soil	0.25	
7	Avdat soil	0.25	2h in ice and 30-minute centrifugation instead of standard 1h centrifugation for precipitation
8	Avdat soil	0.25	
9	Avdat soil	0.25	

10	Certovo soil	0.25	
11	Certovo soil	0.25	
12	Certovo soil	0.25	
13	Avdat soil	0.25	Sodium buffer (Na2HPO4/NaH2PO4 120 mM, pH 7.9) instead of standard potassium buffer (K2HPO4/KH2PO4 120 mM, pH 8.0)
14	Avdat soil	0.25	
15	Avdat soil	0.25	
16	Certovo soil	0.25	
17	Certovo soil	0.25	
18	Certovo soil	0.25	
19	Avdat soil	0.25	Sodium buffer instead of standard buffer and TNC solution (TRIZMA, NaCl, CTAB) instead of TNS solution (TRIZMA, NaCl, SDS)
20	Avdat soil	0.25	
21	Avdat soil	0.25	
22	Certovo soil	0.25	
23	Certovo soil	0.25	
24	Certovo soil	0.25	
25	Avdat soil	0.25	PEG 5% instead of 30% PEG
26	Avdat soil	0.25	
27	Avdat soil	0.25	
28	Certovo soil	0.25	
29	Certovo soil	0.25	
30	Certovo soil	0.25	
31	Avdat soil	0.25	Vortex adapter for microcentrifuge instead of Fast-Prep 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
32	Avdat soil	0.25	
33	Avdat soil	0.25	
34	Certovo soil	0.25	
35	Certovo soil	0.25	
36	Certovo soil	0.25	
37	Avdat soil	0.25	Vortex adapter instead of Fast-Prep + standard buffer + 2h in ice and 30-minute centrifugation instead of standard 1h centrifugation for
38	Avdat soil	0.25	
39	Avdat soil	0.25	
40	Certovo soil	0.25	

41	Certovo soil	0.25	precipitation
42	Certovo soil	0.25	

2.1.2. Cell lysis

Cell lysis was provided in a 2 ml Lysing Matrix E tube. 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 TNS, and 400 µL of TE-saturated phenol solution. Some modified versions used sodium phosphate buffer (120 mM, pH 7.9), 125 µL of TNC, and a vortex adapter for the microcentrifuge were used.

2.1.3. Separation

For separation were used two-step extraction, selective precipitation, and washing. The extraction and washing steps were not modified and did not differ from the same steps in the original protocol.

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

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 100µL of Low-EDTA TE buffer, which contains 500 µL Tris-HCl 1 M, pH 8.0, and 10 µL EDTA 0.5 M, pH 8.0, in 50 mL RNase-free water.

2.2. Sample analysis

2.2.1 Purity of the sample (A260/A280 ratio)

To measure the absorption ratio, A260/280 was used spectrophotometer NanoDrop One/One from Thermo Fisher Scientific was calibrated with Low-EDTA TE buffer before measurements.

After these measurements, samples 1–36 were cleaned using the OneStep PCR Inhibitor Removal KitZymo Research purification kit.

2.2.2. Concentration of DNA

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 times diluted TNA sample. Fluorescence was measured using Synergy™ 2 Multi-Mode Microplate Reader from BioTek® Instruments, Incorporated. Measures were provided by the following protocol: Qant-iT™ PicoGreen® dsDNA Quantification [44].

The DNA concentration in samples 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. Measurements were provided by following instructions from the manufacturer [45].

For parallel determination of DNA concentration in samples 1–42, samples were sent for analysis via the 4150 TapeStation System from Agilent Technologies, Inc.

2.2.3. DIN

To determine the DIN parameter, samples were sent for analysis via the 4150 TapeStation System from Agilent Technologies, Inc.

2.2.4. Length of DNA fragments.

To ascertain the length of DNA fragments acquired during extraction, samples were sent for analysis via the 4150 TapeStation System from Agilent Technologies, Inc.

For measuring TNA fragments length, gel electrophoresis was used. To prepare the gel, we used 100 ml of a 0.5% solution of Agarose Broad Range, ROTI®Garose BioScience Grade in ROTIPHORESE®, both from ROTH. Gel was prepared according to manufacturer instructions [46]. GeneRuler DNA Ladder Mix Thermo Scientific has been used as a ladder from 100 to 10000 bp. 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.

2.2.5. Statistical data analysis

Obtained 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 result of the obtained data is presented in Table 2. Concentration for samples 37–42 was measured not with the help of PicoGreen™ but Qubit®. 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 incredibly poor results; therefore, data from these samples were not considered during statistical analysis.

Also, data obtained from samples 31–36 were not considered during statistical analysis for NA concentration because of significant losses caused by the need to modify the protocol at the lysis stage and the errors in the working technique caused by this, which were eliminated in the next version of the modification. However, these samples were considered for quality, DIN, and DNA fragment length during analysis.

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

Substitution of TNS for TNC as a detergent agent in lysis did not show statistically significant results in either case and therefore was not considered separately in subsections.

Table 2

				NanoDrop	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)**
1	Avdat soil	0.25	14.558	1.90	145.58	145	5.5	5612
2	Avdat soil	0.25	10.599	1.56	105.99	105	5.6	5715
3	Avdat soil	0.25	11.418	1.90	114.18	113	5.6	5794
4	Certovo soil	0.25	11.327	1.34	113.27	117	5	6282
5	Certovo soil	0.25	35.263	1.29	352.63	353	5.2	6150
6	Certovo soil	0.25	26.026	1.32	260.26	259	5	5518
7	Avdat soil	0.25	12.783	1.92	127.83	127	5.7	6030

8	Avdat soil	0.25	8.278	1.90	82.78	105	5.8	5838
9	Avdat soil	0.25	11.373	1.92	113.73	113	5.7	5980
10	Certovo soil	0.25	30.94	1.34	309.4	298	5	5280
11	Certovo soil	0.25	16.424	1.33	164.24	161	4.5	5174
12	Certovo soil	0.25	34.99	1.34	349.9	349	5.1	5628
13	Avdat soil	0.25	14.33	1.83	143.3	142	5.5	5528
14	Avdat soil	0.25	9.734	1.80	97.34	96	5.4	5280
15	Avdat soil	0.25	11.691	1.82	116.91	116	5.5	5114
16	Certovo soil	0.25	19.928	1.33	199.28	202	4.4	4626
17	Certovo soil	0.25	20.747	1.33	207.47	210	4.6	5299
18	Certovo soil	0.25	23.614	1.33	236.14	238	4.3	4629
19	Avdat soil	0.25	10.189	1.70	101.89	101	5.6	5910
20	Avdat soil	0.25	4.092	1.73	40.92	96	5.4	5632
21	Avdat soil	0.25	13.375	1.71	133.75	133	5.6	5916
22	Certovo soil	0.25	31.168	1.29	311.68	32	4.9	5562
23	Certovo soil	0.25	42.544	1.34	425.44	43	5.1	5979
24	Certovo soil	0.25	-	1.32	-	-	-	-
25	Avdat soil	0.25	0.5	1.40	5	1	3	2003
26	Avdat soil	0.25	0.145	1.39	1.45	0	2.3	1428
27	Avdat soil	0.25	0.132	1.45	1.32	0	2.3	1787
28	Certovo soil	0.25	1.634	1.29	16.34	1	4.4	4928

29	Certovo soil	0.25	1.998	1.30	19.98	2	4.8	5893
30	Certovo soil	0.25	1.179	1.30	11.79	1	4.3	4621
31	Avdat soil	0.25	3.877	1.56	38.77	39	6.4	10083
32	Avdat soil	0.25	2.807	1.59	28.07	28	6.5	9739
33	Avdat soil	0.25	2.507	1.52	25.07	25	6.5	9603
34	Certovo soil	0.25	14.831	1.35	148.31	15	5.8	9670
35	Certovo soil	0.25	11.373	1.35	113.73	11	5.7	9761
36	Certovo soil	0.25	12.283	1.35	122.83	12	5.9	9693
37	Avdat soil	0.25	9.91	1.729	99.1	99.1	6.6	12744
38	Avdat soil	0.25	6.71	1.722	67.1	67.1	6.5	12646
39	Avdat soil	0.25	2.68	1.794	26.8	26.8	6.6	12114
40	Certovo soil	0.25	13.6	1.335	136	136	2	9113
41	Certovo soil	0.25	20.7	1.350	207	207	2.1	8813
42	Certovo soil	0.25	25.8	1.352	258	258	2.7	10247

3.2. Concentration

Based on a statistical analysis of the acquired data, it is evident that protocol modification did not show any statistical difference ($p >> 0.05$) in the concentration of NA in the obtained extracts in the case of samples extracted from Certovo soil. Nevertheless, in the case of Avdat soil ($p = 0.0459$), there is some decrease when using an Adapter instead of FastPrep for cell lysing. The results of the statistical analysis are listed in Table 3 and Table 4 and visualized in Figure 4.

Table 3

Avdat soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	0.2030	0.6612	-
Lysis method	1.0000	5060	0.0459	42.2347
Precipitation method	1.0000	0.2820	0.6063	-

Table 4

Certovo soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	0.482	0.503	-
Lysis method	1.0000	0.734	0.412	-
Precipitation method	1.0000	0.173	0.686	-

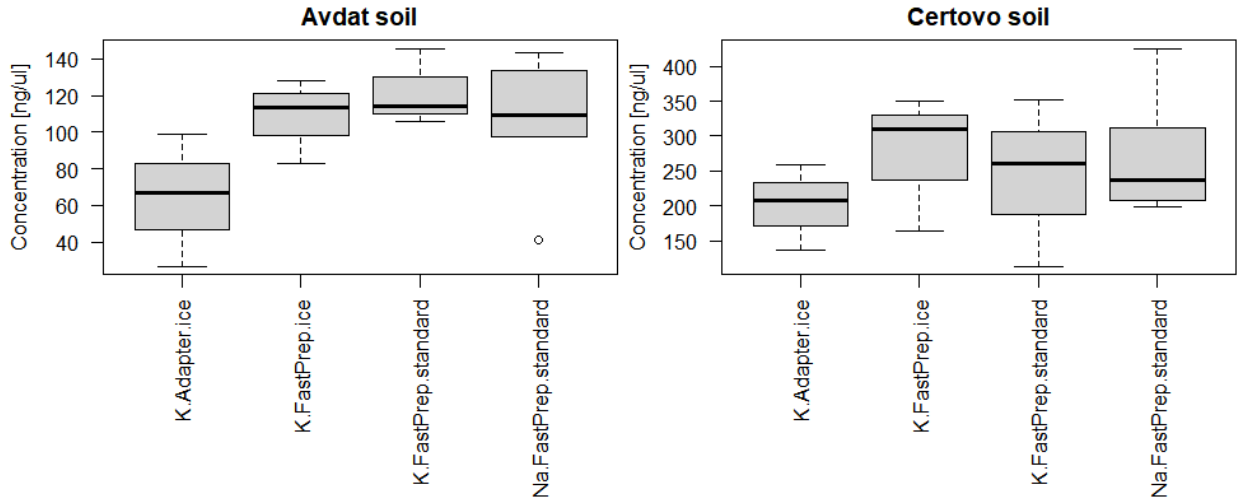

Fig. 4

3.3. Purity

Statistical analysis of the data shows that protocol modification did show a significant 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 sample purity, while shorter centrifugal times combined with precipitation in an ice bath improve sample purity to some extent. In the case of Certovo soil, only applying an Adapter instead of Fast Prep for cell lysing demonstrated a statistical difference in sample purity; its magnitude was non-significant. The results of the statistical analysis are listed in Table 5 and Table 6 and visualized in Figure 5.

Table 5

Avdat soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	9.0200	0.0095	0.1206
Lysis method	1.0000	13.2540	0.0027	0.1550
Precipitation method	1.0000	4.6660	0.04859	-0.0608

Table 6

Certovo soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	0.009	0.9239	-
Lysis method	1.0000	7.7280	0.0156	-0.0223
Precipitation method	1.0000	1.6180	0.2257	-

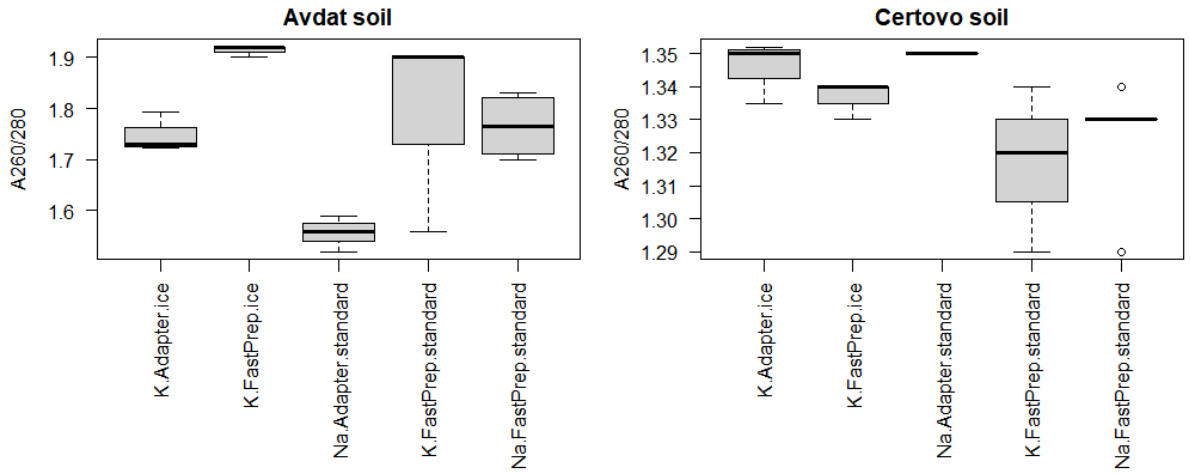

Fig. 5

3.4. DIN

Statistical analysis shows that protocol modification did not show any statistical difference in DIN parameters in obtained samples extracted from Certovo soil. However, in the case of Avdat soil modification, significant statistical differences in DNA integrity numbers occur. Modifying the lysis method shows some significant increase in DNA quality obtained from Avdat soil, while modification of the precipitation method shows some non-significant improvement. Application of sodium buffer shows 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

Avdat soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	13.6740	0.00239	0.1333333
Lysis method	1.0000	606.2800	$6.31 * 10^{-13}$	-0.9417
Precipitation method	1.0000	4.9450	0.04314	-0.0563

Table 8

Certovo soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	4.5360	0.0529	-
Lysis method	1.0000	2.8230	0.1168	-
Precipitation method	1.0000	2.7480	0.1213	-

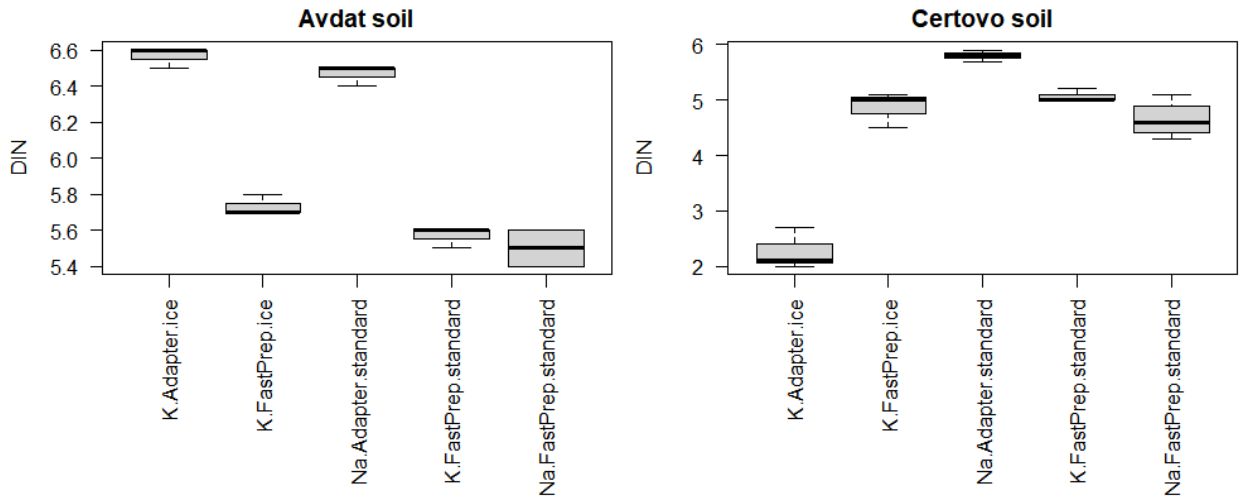

Fig. 8

3.5. Length of DNA fragments

The implication of the vortex adapter instead of FastPrep allows 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 showed no statistically significant effect in both soils. The statistical analysis results are listed in Table 9 and Table 10 and visualized in Figure 10.

Figure 11 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 since, in the region of 1500–2000 bp, there is a double band characteristic for 16S ribosomal RNA.

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. For example, Figures 12 and 13 represent graphs of DNA length distribution for samples 1 and 37, respectively.

Table 9

Avdat soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	13.2580	0.00267	1074.2220
Lysis method	1.0000	304.3540	6.79e-11	-5459.0830
Precipitation method	1.0000	3.6310	0.07746	-

Table 10

Certovo soil	Df	F value	Pr(>F)	difference of means (standard - modified)
Type of buffer	1.0000	0.001	0.9697	-
Lysis method	1.0000	265.8980	6.79e-114.91e-10	-4083.8310
Precipitation method	1.0000	4.0170	0.0663	-

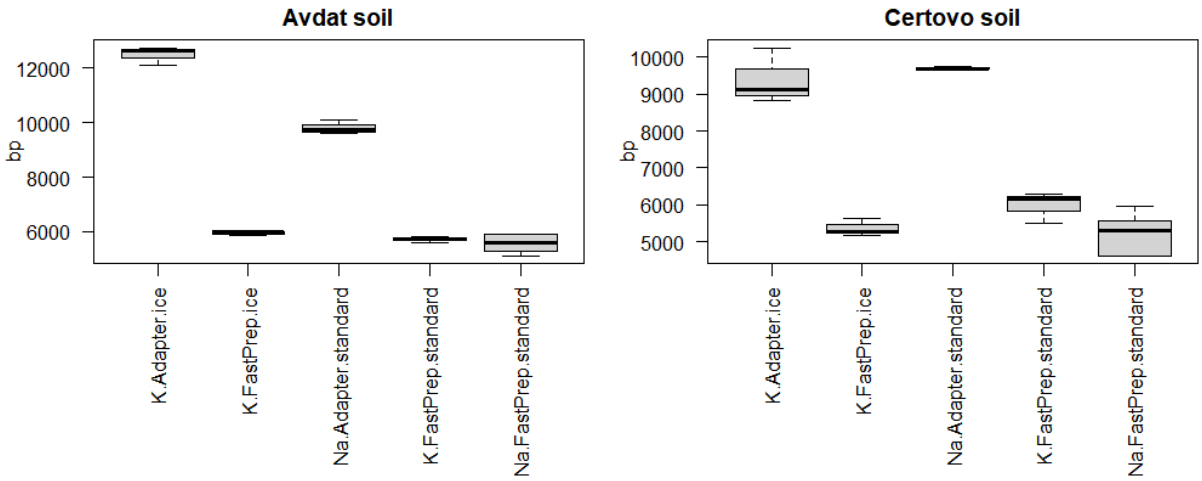

Fig. 10

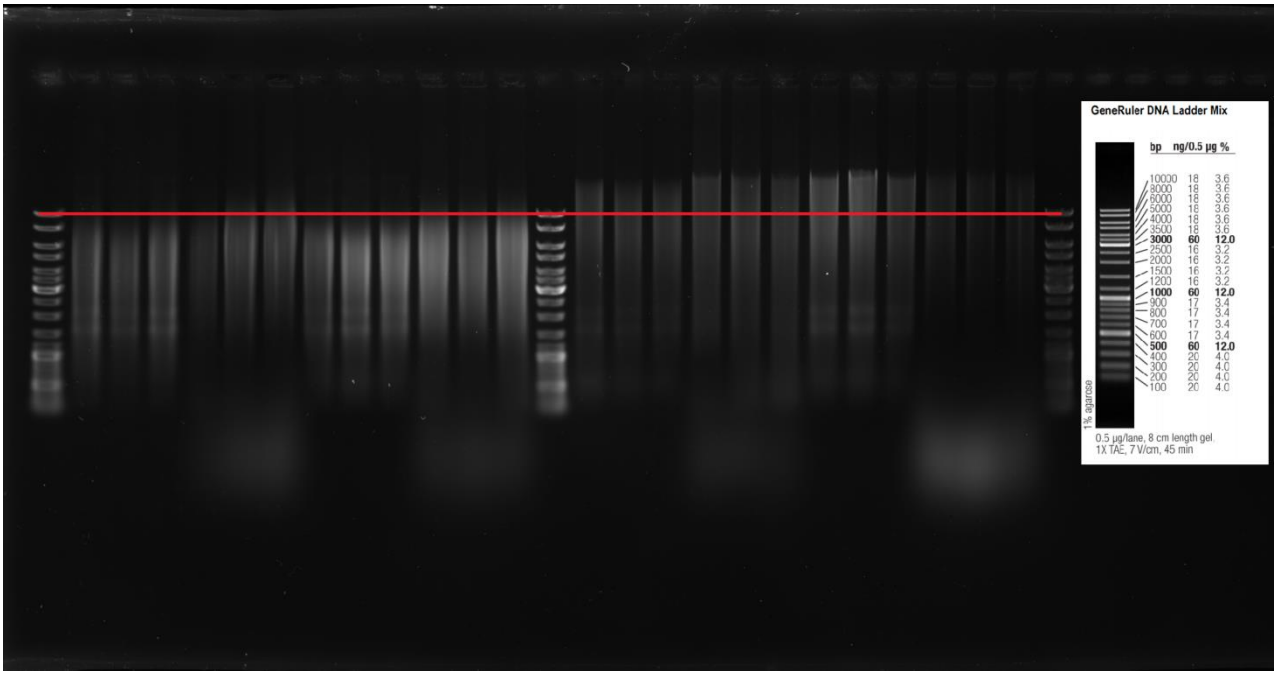
GeneRuler DNA Ladder Mix
bp ng/0.5 µg %
10000 18 3.6
8000 18 3.6
6000 18 3.6
5000 18 3.6
4000 18 3.6
3500 18 3.6
3000 60 12.0
2500 16 3.2
2000 16 3.2
1500 16 3.2
1200 16 3.2
1000 60 12.0
900 17 3.4
800 17 3.4
700 17 3.4
600 17 3.4
500 60 12.0
400 20 4.0
300 20 4.0
200 20 4.0
100 20 4.0
1% agarose
0.5 µg/lane, 8 cm length gel.
1X TAE, 7 V/cm, 45 min

Fig. 11

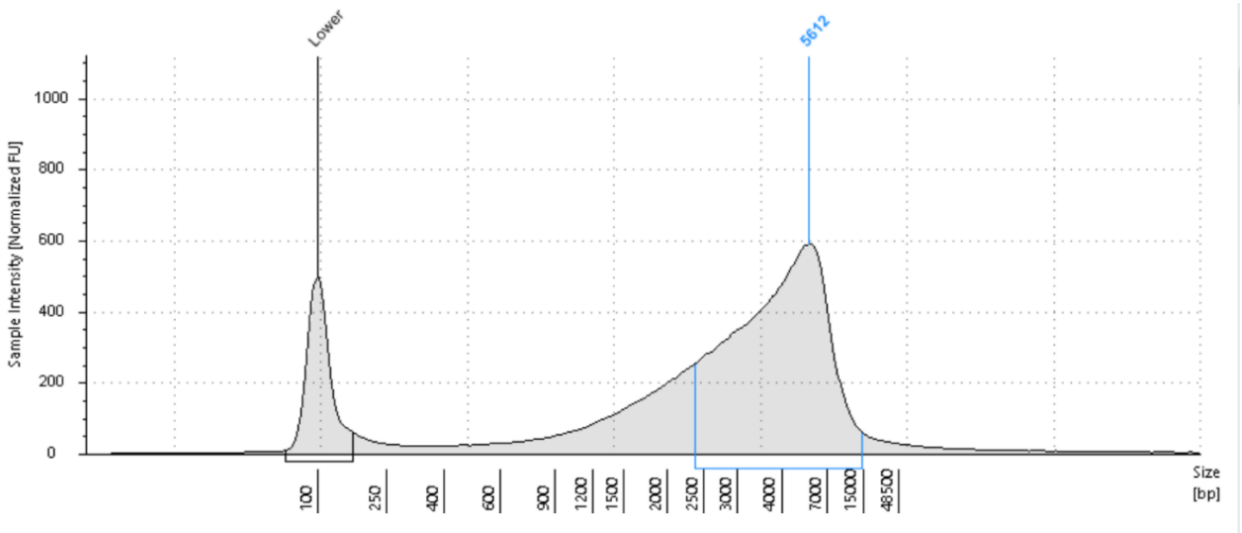
Lower
5612
Sample Intensity [Normalized FU]
1000
800
600
400
200
0
100
250
400
600
900
1200
1500
2000
2500
3000
4000
7000
15000
48500
Size [bp]

Fig. 12

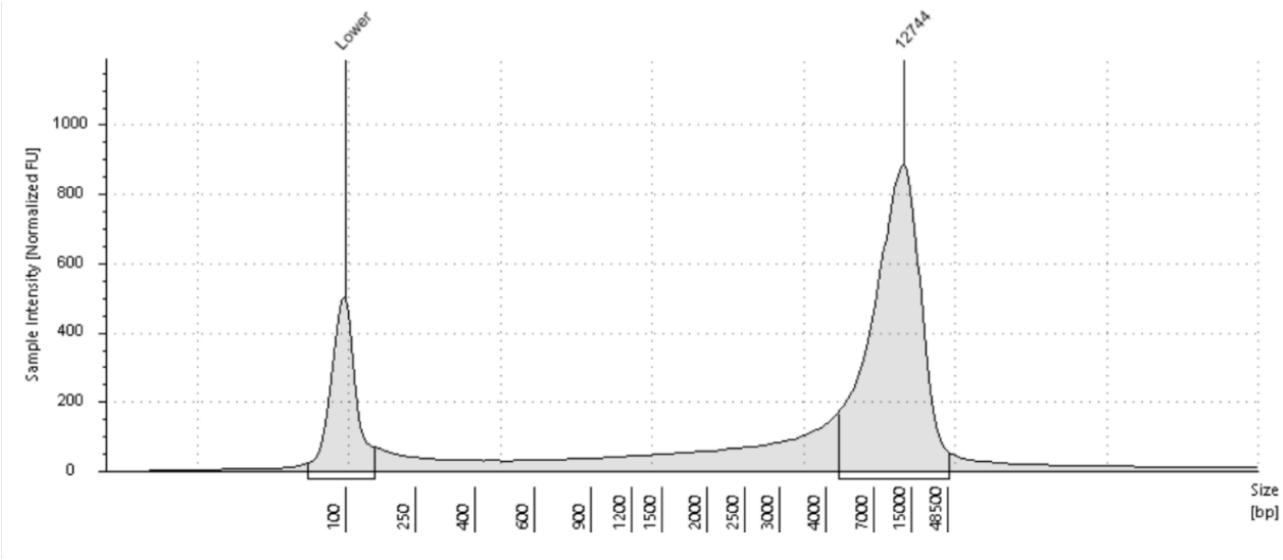

Fig. 13

4. Discussion

This work focused on improving our NA extraction protocol. We reviewed almost all the steps and tested the changes to create a more efficient method. From the obtained results, we failed to fully fulfill all the goals we set for ourselves before starting the research. As a final result, we obtained a significant increase in the length of the extracted DNA fragments and, accordingly, the DIN parameter, with some decrease in the quantitative yield and purity.

It can be seen that, as was known [17-20], 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 desert mostly demonstrate adequate purity (parameter A 260/280 in the range of 1.7-2 [41 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.

Contrary to the previous studies, which showed that there is no difference between the use of vortex adapter and FastPrep ($p > 0.1$)[41 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. 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.

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 and quantitative yield 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, the DIN parameter, and the purity of the obtained extracts. 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 minor improvements in the purity and DIN parameters of the resulting product. However, it made the procedure more time-consuming and challenging to perform since the resulting NA pellets were extremely sensitive, did not sit well on the inner surface of the low-binding tubes, and required very delicate handling.

Furthermore, as was mentioned, replacing TNS with TNC as a detergent agent in lysis did not show a statistically significant difference.

5. Conclusion

Even though we did not achieve all the goals we set before starting the study, we managed to obtain a significant increase in the length of extracted DNA fragments, which is one of the 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. PMID: 13054694.
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. PMID: 271968; PMCID: PMC431765.
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. PMID: 19053153..
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. PMID: 18846084.

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. PMID: 11181995.
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. PMID: 32477917; PMCID: PMC7237973.
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. PMID: 20858600.
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. PMID: 23206278; PMCID: PMC3599793.
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. PMID: 7690221; PMCID: PMC182335.
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. PMID: 10579502.
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,
36. 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.
37. 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. Epub 2022 Feb 7. PMID: 35068060; PMCID: PMC9177515.
38. 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.

39. 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.
40. 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
41. 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
42. Roey Angel, Eva Petrova, Ana Lara-Rodriguez Total Nucleic Acids Extraction from Soil V.6 dx.doi.org/10.17504/protocols.io.bwxcpfiw
43. 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. PMID: 10223989; PMCID: PMC91286.
44. Roey Angel, Eva Petrova, Ana Lara-Rodriguez Qant-iT™ PicoGreen® dsDNA Quantification V.1 https://dx.doi.org/10.17504/protocols.io.qdrds56
45. https://www.thermofisher.com/at/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-assays.html
46. https://www.carlroth.com/medias/BA-T846-EN.pdf?context=bWFzdGVyfGluc3RydWN0aW9uc3wzMDg5OTl8YXBwbGljYXRpb24vcGRmfGluc3RydWN0aW9ucy9oNjQvaDA0LzkwMTAzNzcxMzAwMTQucGRmfGJjNmQ4NWEyNmQ3YzFkMjNhZDkyOTQyMzAxYmFjYjgxZWE5MTkwNmUwNTUxYWE0YzRlZTY3MjUyYjViMmYxMDU