Scientific Reports

https://doi.org/10.1038/s41598-026-38534-3

Article in Press

Evaluation of commercial kits and purification approaches for DNA extraction from atmospheric samples for 3rd generation sequencing without amplification

Received: 1 June 2024

Accepted: 29 January 2026

Published online: 11 February 2026

Cite this article as: Salokas J., Sofieva-Rios S., Paatero J. *et al.* Evaluation of commercial kits and purification approaches for DNA extraction from atmospheric samples for 3rd generation sequencing without amplification. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-38534-3

Julija Salokas, Svetlana Sofieva-Rios, Jussi Paatero, Eija Asmi, Ari Karppinen & Mikhail Sofiev

Evaluation of commercial kits and purification approaches for DNA extraction from atmospheric samples for 3rd generation sequencing without amplification

Salokas, J.[1]; Sofieva-Rios, S.[1,2]; Paatero, J.[3]; Asmi, E.[1]; Karppinen, A.[1]; Sofiev, M[1*].

[1] Finnish meteorological institute, Atmospheric composition, P.O. Box 503, FI-00101, Helsinki, Finland
[2] University of Helsinki, Department of Molecular and Integrative Biosciences, 00010 University of Helsinki
[3]Finnish Meteorological Institute (FMI), Observation Services, P.O. Box 503, FI-00101, Helsinki, Finland

Corresponding author: Mikhail Sofiev (mikhail.sofiev@fmi.fi)
The first two authors contributed equally.

Abstract

We present a DNA extraction protocol for atmospheric bioaerosol samples collected on glass-fiber filters widely used in air quality monitoring. The protocol produces high-quality molecules suitable for third-generation sequencing and other applications. The initial protocol was developed and applied in a Bioaerosol campaign performed in Finland and Lithuania in 2021 using low-volume air samplers, which posed stringent requirements to the method sensitivity. The protocol included a phenol–chloroform step for DNA purification, thus involving aggressive reagents; it was also quite time consuming and laborious. The present study advances this protocol to exclude the use of hazardous chemicals by using the SPRI paramagnetic bead technology for DNA purification and compares it to several commercial extraction methods. Despite trailing in efficiency to the initial method, the new development proved to be more efficient than several column-based commercial kits. The updated protocol was effective for a relatively high mass ratio of biological material to filter material: 70 nanograms of potential DNA on the filter to one milligram of filter fiber, as detected with the initial phenol–chloroform-based method. However, the new approach was not effective for a mass ratio lower than 15 nanograms of potential DNA per milligram of the filter material. The applicability of the new protocol for preparation of samples for the 3rd generation sequencing was confirmed by subsequent processing of the samples with the Oxford Nanopore (ONT) GridION sequencer.

Keywords

bioaerosols, atmospheric metagenomic, DNA extraction kits

1. Introduction

Primary biological aerosols in the Earth atmosphere, including pollen, fungal spores, bacteria, and viruses, constitute a substantial fraction of aerosols and are actively dispersed by winds over large distances[1,2]. This expansive journey, while transcending geographical confinement, introduces health risks by spreading allergenic particles and pathogenic microbes that contribute to respiratory illnesses, infections, allergies, and even cancer[1,3–5]. The composition of bioaerosols is determined by vegetation

composition at the surface and influenced by seasonal shifts and weather patterns. It can also be modulated by air pollution levels [6,7].

Aerial microbes, which make up less than 1% of airborne entities, have often been overlooked due to challenges in traditional monitoring methods, such as culturing[8]. Metagenomics has addressed this gap, allowing for exploration of species diversity through DNA extraction and culture-independent analyses[9–12]. This approach is particularly pertinent for understanding enigmatic aerosols, such as pollen, where accurate DNA extraction is essential for precise metagenomic studies, microbial profiling, and pathogen detection [13,14].

Long-read DNA sequencing technologies, such as PacBio and Oxford Nanopore, have transformed biodiversity studies by providing much more comprehensive and accurate genetic information[15]. These technologies produce reads that can cover entire genes or genomes, making them invaluable for studying complex ecosystems[16]. In the field of bioaerosols, long-read sequencing can help to identify new species, detect genetic diversity, and enhance our understanding of microbial community functions, also simplifying the task of genome assembly [17].

Metagenomic studies of the atmospheric bioaerosols face challenges due to low concentration of biological material in the air in comparison with water and soil. To overcome this roadblock, one has to use high-volume samplers (expensive and difficult as well) and a highly sensitive and precise procedure of the sample treatment and sequencing. This paper addresses the second challenge by presenting a DNA extraction protocol applicable to atmospheric samples with moderate-to-low amount of biological material.

The initial version of the protocol presented in this article was developed for the Bioaerosol campaign performed in 2021 in Helsinki (Finland) and Šiauliai (Lithuania) using several types of sampling devices, with the air flow rate varying from 2 to 50 l min^{-1} and collection time ranging from 2 to 24 hours [12]. Briefly, the extraction process consisted of three steps: an enzymatic pretreatment of samples in a modified Longmire’s buffer, a chemical degradation with detergents, and a treatment with proteinase K. The DNA purification was carried out using phenol–chloroform-isoamyl alcohol, followed by sodium acetate-isopropanol precipitation, and finished with ethanol washing.[12]

The original protocol performed well for DNA extraction from bioaerosols, producing high-quality DNA from various sample types: polycarbonate air filters, glass slides or plastic tape coated with an organic adhesive surface, and liquid samples. This approach was effective even for comparatively low-yield samples. However, it was time-consuming, laborious, and involved hazardous chemicals, such as phenol and chloroform. To upscale the DNA extraction and analysis process and to make it suitable for operational tasks and large-scale bioaerosol analysis, modifications to the DNA extraction process were necessary.

DNA extraction methods have evolved significantly in recent years, with a variety of commercial kits available alongside traditional methods, such as phenol–chloroform–isoamyl alcohol– or chloroform–based techniques. These methods commonly utilize resin columns to bind DNA molecules. While the legacy methods often involved lengthy procedures with multiple steps, modern methods proved to be more efficient and able to extract high-quality DNA suitable for downstream applications, such as PCR and

sequencing.[18] Commercial kits offer standardized protocols, and are widely used in research owing to their reported efficiency and ease of use[19].

Magnetic beads with affinity for binding DNA are another method that is considered to be the gold-standard solution for DNA purification for many sequencing platforms, including Oxford Nanopore Technologies[20,21]. Additionally, studies have shown that magnetic beads enable efficient extraction of microbial DNA directly from plant roots[22,23]. Therefore, the update of the initial DNA extraction protocol incorporated this technology as one of possible options, aiming at higher efficiency and high yields of DNA with simultaneous elimination of the time-consuming steps and hazardous materials.

The current study evaluated efficiency of five commercial DNA extraction kits: DNA and DNA/RNA miniprep kits (Zymobiomics cat. D4300 and R2002, hereinafter referred to as M3 and M4, respectively), a PeqGOLD Plant DNA Mini Kit (Peqlab VWR Chemicals cat. 13-3486, M5), a Soil DNA Isolation Plus Kit (Norgen Biotek Corp. cat. 64000, M6), and a Qiagen DNeasy Plant Pro Kit (Qiagen cat. 69204, M7). The efficiency and quality of these commercial kits were compared with the upgraded in-house protocol coupled with the AMpure magnetic beads (Beckman Coulter cat. A63881, M1). Finally, the initial version of the protocol involving phenol–chloroform DNA extraction and purification set the reference for the comparison (method M2).

2. Results and discussion

To determine the optimal DNA extraction from air samples, we conducted a comparative analysis of the seven extraction methodologies. The comparison criteria were as follows: (i) DNA yield, (ii) purity, (iii) reproducibility of the outcome, and (iv) quality of the extracted DNA as revealed by a subsequent sequencing with Nanopore GridION: sequence length, accuracy, and completeness.

Two types of samples were analyzed, with low and high amount of DNA, as determined by the reference method M2. All methods were successfully tested against the control samples: DNA was not observed in any of the controls with un-exposed glass-fiber filters.

2.1. DNA extraction efficiency for high-yield filters

The results (**Error! Reference source not found.**, blue-colored bars) show that more than 20 ng DNA per mg of filter material can be extracted from the high-yield samples using three methods: the phenol–chloroform-based initial method M2, the M1 protocol with AMpure XP beads, and the Norgen Biotek Soil DNA Isolation Plus Kit (M6). The relative amounts of DNA extracted from the same samples were 3:7:5 for the M1, M2, and M6 methods, respectively. This extraction efficiency was sufficient for sequencing without the multiplication step. Similar experiment with low-yield filters, less than 15 ng DNA per mg of filter material, as extracted by the M2 method, (**Error! Reference source not found.**, yellow-colored bars) were unsuccessful for all methods but the phenol–chloroform M2 itself. Most of DNA was lost in all repetitions of the analysis, and the extracted DNA amount was comparable to that in the control extraction from the unused filter. For such samples, only the initial M2 protocol version proved usable.

The M2 method also demonstrated a weakness: a high variability of the yield (**Figure 1**). This variability can be attributed to many factors, such as an uneven distribution of aerosols over the filter, which should be particularly significant for M2 due to its high sensitivity. Reagent preparation and slight alterations in extraction duration, temperature and pH can also affect the results. Finally, complexity of the protocol also

contributed to the uncertainty: it includes multiple critical points prone to human error. Both M1 and M6 methods produced more stable results than M2.

The statistical significance of the differences between the methods was evaluated with Kruskal-Wallis ANOVA test coupled with Dunn's test (see Supplementary material). The results show that for the high-yield samples the only significant difference is between the M3 and M6 methods, with M6 outperforming M3. Differences between other methods turned out to be statistically insignificant (see Supplementary information S7). For low-yield filter samples, the DNA quantity extracted with M2 method was significantly higher than that from methods M1 and M6.

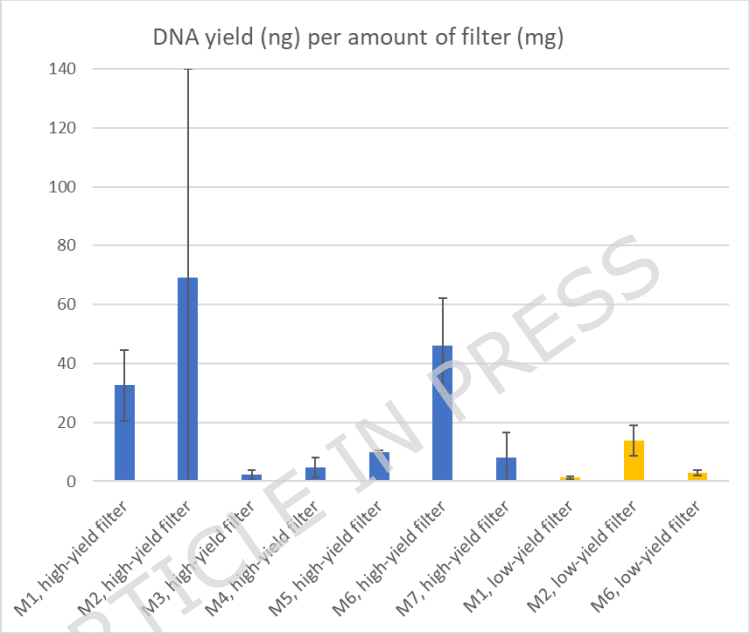

***Figure 1**. The DNA yield from each sample presented in nanograms of DNA per milligram of the filter; only nonzero yields are shown. M1 is the paramagnetic bead-based extraction method presented in this manuscript, M2 is the phenol–chloroform-based reference method tested in the 2021 Bioaerosol campaign[11], M3 is the Zymobiomics DNA miniprep kit, M4 is the Zymobiomics DNA/RNA miniprep kit, M5 is the Peqlab PegGOLD Plant DNA Mini Kit, M6 is the Norgen Biotek Soil DNA Isolation Plus Kit, and M7 is the QIAGEN DNeasy Plant Pro Kit. The bars represent standard deviation for each method. The significance of the differences in yields between methods was assessed in Supplementary material S7 according to the method described in section 2.3.*

2.2. Quality of the extracted DNA

To assess the quality of the extracted DNA, the ratios of sample absorbances at wavelengths of 260 nm/230 nm and 260 nm/280 nm were calculated (**Table 1**). Samples with a ratio of ~1.8 for 260 nm / 230 nm and ~2 for 260 nm / 280 nm were considered free from contaminants. Both substantially lower and substantially higher ratios indicate contamination. In extreme cases, the values can be even negative because the optical thickness is referred to that of the elution buffer with no DNA (in case of near-zero DNA amount in the sample the ratio is practically a zero-divided-by-zero). For environmental samples, it is exceedingly difficult to obtain near-optimal

values. Therefore, the general guideline is that ratios of between 1.5 and 2.5 are already suitable for further applications.

The difference between minimum and maximum ratios is the smallest for the traditional phenol–chloroform method M2, which also reached both target ranges in at least some repetitions and stayed close to these ranges in all test. In average, the quality of the extracted DNA reached the target quality range (ratio A260/A280 > 1,5) with M1 – but it failed some replicas (the minimum of the A260/A230 ratio was 0.556). However, the risk of such a failure of some repetitions can be mitigated with a larger number of DNA extraction repetitions, albeit at a cost of extra work and reagents.

Table 1. *Ratios of absorbances at wavelengths 260 nm / 230 nm and 260 nm / 280 nm of the DNA samples. M1 is the paramagnetic bead-based extraction method presented in this manuscript, M2 is the initial phenol–chloroform-based version of the method tested in the 2021 Bioaerosol campaign[11], M3 is the Zymobiomics DNA miniprep kit, M4 is the Zymobiomics DNA/RNA miniprep kit, M5 is the Peqlab PegGOLD Plant DNA Mini Kit, M6 is the Norgen Biotek Soil DNA Isolation Plus Kit and M7 is the QIAGEN DNeasy Plant Pro Kit. The target range for A(260 nm) / A(230 nm) is 1.5-2, and that for A(260 nm) / A(280 nm) is 1.6-1.9.*

	A(260 nm) / A(230 nm)			A(260 nm) / A(280 nm)		
	Maximum	**Minimum**	**Mean**	**Maximum**	**Minimum**	**Mean**
M1	2.500	0.556	1.278	2.300	1.667	1.946
M2	1.667	1.390	1.542	1.763	1.295	1.465
M3	0.500	0.429	0.465	4.000	3.000	3.500
M4	0.304	-0.136	0.039	1.400	1.167	1.322
M5	1.200	0.727	0.315	2.000	0.875	1.564
M6	7.231	0.188	1.695	2.040	1.500	1.728
M7	0.929	-0.093	0.406	5.000	1.125	2.309

Due to high variability in the sample quality between repetitions, each absorbance spectrum was examined for presence of a slope typical for DNA sample (see example in Supplementary Figure S1). After examination of the absorbance spectra, we had to conclude that the tested kits were not successful in extracting enough DNA from atmospheric filters for spectrophotometric observation: the curves typical for DNA samples with a peak of absorbance at 260nm wavelength were not observed. None of the commercial kits managed to reliably reach the desired range of sample absorbance ratios and demonstrate absorbance spectrum shape typical for DNA samples, including M6, which showed promising DNA concentrations in the efficiency tests. An example of the spectra for four repetitions of the M6 extraction is presented in Supplementary material S6.

2.3. Sequence quality and species composition

The above tests suggested two methods for further analysis – M1 and M2. The method M6 could also be a candidate unless the problems with reliability of the spectrum shape (Figure S1) and inability to extract reasonable amount of DNA from the low-yield filters.

The Nanopore sequencing was performed on three samples extracted with M1 and two samples extracted with M2 methods. The read ends were processed to meet the quality standards required for the following classification analysis and assembled into contigs. The general statistics for the assembly are presented in **Table 2**. The second repetition of the DNA sample extracted with M2 resulted in an extremely low yield sequence with a median contig length less than 200 base pairs. This was most likely due to DNA disruption during one of the critical steps in the DNA extraction process, to which M2 is prone, as mentioned above. It might be the reason for the much lower fraction of bacterial sequences in the repetition 2 than in the repetition 1.

The GC content varied greatly between the samples, revealing a high tolerance of ONT sequencing chemistry and platform for such variation. Illumina sequencing is known to be less tolerant.[24,25] This feature of the ONT platform makes it more flexible and amenable to use for analysis of environmental samples, and atmospheric samples in particular.

The M1 and M2 methods produced similar results for the fraction of unclassifiable sequences, which varied from 22.1% to 33.3%. The fraction of sequences classified as “other than plants or bacteria” varied from 6.9% to 22.2%, also quite synchronously for both methods. The “other” fraction mainly consisted of insects, but also included fungal (average abundance of ~0.1%) and human-originated (~0.2%) DNA.

Table 2. *General statistics on the assembly of DNA samples extracted from a glass fiber filter collected on 31.5.2021 in Helsinki, methods M1 and M2.*

Assembly statistics									
Sample Name	**% GC**	**Length range (bp)**	**Average Contig Length (bp)**	**Median Contig Length (bp)**	**Total sequence length (kbp)**	**Fraction of plant sequences (%)**	**Fraction of bacterial sequences (%)**	**Fraction of other sequences (%)**	**Fraction of unclassified sequences (%)**
M2, repetition 1	78%	150-79407	2021	999	93.2	13.4	50.7	10.4	25.4
M2, repetition 2	40%	150-747	210	169	5.6	33.3	11.1	22.2	33.3
M1, repetition 1	42%	151-25735	1357	499	32.9	65.5	3.4	6.9	24.1
M1, repetition 2	56%	150-18927	578	249	42.9	37.8	18.3	11.0	32.9
M1, repetition 3	45%	152-39551	1746	499	101.5	48.5	19.1	10.3	22.1

One of the greatest concerns in the DNA extraction for metagenomic analysis is a bias due to the varying extraction efficiencies of the methods for different bioaerosol types. According to **Table 2**, the relative abundance of bacterial sequences was low in M1 (3.4%-19.1%), whereas the relative abundance of sequences classified as plants was much greater (37.8%-65.5%). The results for M2, especially the repetition 1, indicated the opposite, with a lower abundance of plant sequences and a greater fraction of bacterial sequences. A likely explanation lies in the extraction method: M2 did not include physical disruption, whereas M1 did. Physical extraction may disrupt DNA from softer or partially damaged cells more extensively, thus favoring more rigid particles, such as pollen. Since the filter was collected 2 years before the treatment, soft cells or

spores with shorter lifetimes could be damaged already before the beginning of the extraction process.

2.4. Other comparative studies of DNA extraction protocols

Different performance of the extraction methods and kits has been noticed by previous studies comparing the extraction methods. They emphasized importance of selecting a suitable method for a particular sample type and downstream applications[26]. For example, the presence of interfering compounds, such as tannins and essential oils in plant samples, can influence the extraction process[27]. Especially in the column-based approach, the efficiency of the column in binding DNA molecules is strongly dependent on composition of the sample. Column-based methods using commercial kits were fast but provided low yields, while more efficient methods were more time-consuming[28].

Another difficulty is that the available extraction kits with physical sample disruption generally assume a high abundance of the biomatter in the sample, and the chemistry of the kits is optimized for it[29]. However, concentration of bioaerosols on filters is often low, so that a large fraction of DNA in a sample needs to be extracted to collect a sufficient amount of DNA for sequencing[30].

These findings corroborate with our conclusions, whereas the highlighted peculiarities are particularly important for atmospheric samples, which are usually poor in biological material.

3. Methods

3.1. Experiment design

The filter samples for testing the DNA extraction kits were collected at the roof of the FMI main building in Kumpula university campus, Helsinki, Finland (60°12’N, 24°58’E), 18 m above street level. The high-volume (3500 m^3 d^{-1}) aerosol samples were collected onto 240 mm diameter round glass-fiber filters (Munktell MGA), daily.[31] For test purposes, two filters, with high and low DNA yield, were selected from the archive of the bioaerosol campaign of 2021 in Helsinki [12]. The DNA yield was determined with the phenol-chloroform extraction (method M2 in this paper). The high yield filter with 17.7 µg of DNA per 100mg of filter was obtained on 31.5.2021, and the low yield filter was from 23.4.2021, with 5.4 µg of DNA per 100mg of filter. Large size of the filters (240 mm in diameter) allowed cutting 70 mg filter slices for each experiment and supplying exactly the same samples to all tests. For each DNA extraction kit, the initial execution of the corresponding protocol was performed in accordance with the manufacturer's instructions. It was conducted immediately after the sample(s) were subjected to in-house enzymatic pretreatment (**Figure 2**).

Differences in DNA yields were statistically assessed using IBM SPSS Statistics V. 30 software, with a Kruskal-Wallis ANOVA test coupled with Dunn’s test for pairwise comparisons.[32] The p-value <0.05 was selected as a cutoff for significance.

3.2. Procedure of the improved protocol

The protocol M1 consisted of three steps: filtration, DNA extraction, and DNA purification (**Figure 2**). To prepare for the DNA extraction, glass fiber filters were cut into pieces of approximately 12 cm^2 (~0.07–0.08 g) in a sterile environment. After that, the protocol

proceeded to the DNA extraction with the enzymatic pretreatment. Each filter piece was submerged in a solution containing 0.1 M Tris-HCl, 0.05 M NaCl, 0.1 M EDTA buffer, 0.8 µl of lysostaphin, 1.2 µl of lyticase, and 8 µl of lysozyme and incubated for 1 hour at +37°C. After the incubation, the sample was gently broken down into smaller pieces using a clean pipette tip.

After enzymatic pretreatment, chemical degradation was performed with a detergent mixture supplemented with proteinase K. The mixture was incubated at +55°C for 30 minutes. Following chemical disruption and proteinase K treatment, the filter was homogenized with a tissue disruption tube from the Qiagen DNeasy Plant Pro Kit by vortexing. The filter pieces and cell debris were removed from the sample with centrifugation.

After the DNA purification steps, sodium acetate (3 M, pH 5.2) was added with an amount corresponding to 1/10 of the sample volume, and the mixture was gently mixed. Next, 7/10 of the sample volume of isopropanol was added at room temperature. The solution was centrifuged at maximum speed (17,000 × g) for 30 minutes at room temperature. The supernatant was discarded, and residual isopropanol was carefully removed with a pipette.

AMpure XP beads were prepared by mixing and transferring the required amount to a clean tube, after which the mixture was allowed to reach room temperature. Subsequently, 30 µl of AMpure beads and 110 µl of 80% EtOH were added and the sample was incubated on a slow mixer for 15 minutes at +4°C, followed by pelleting on a magnetic rack and removal of the supernatant.

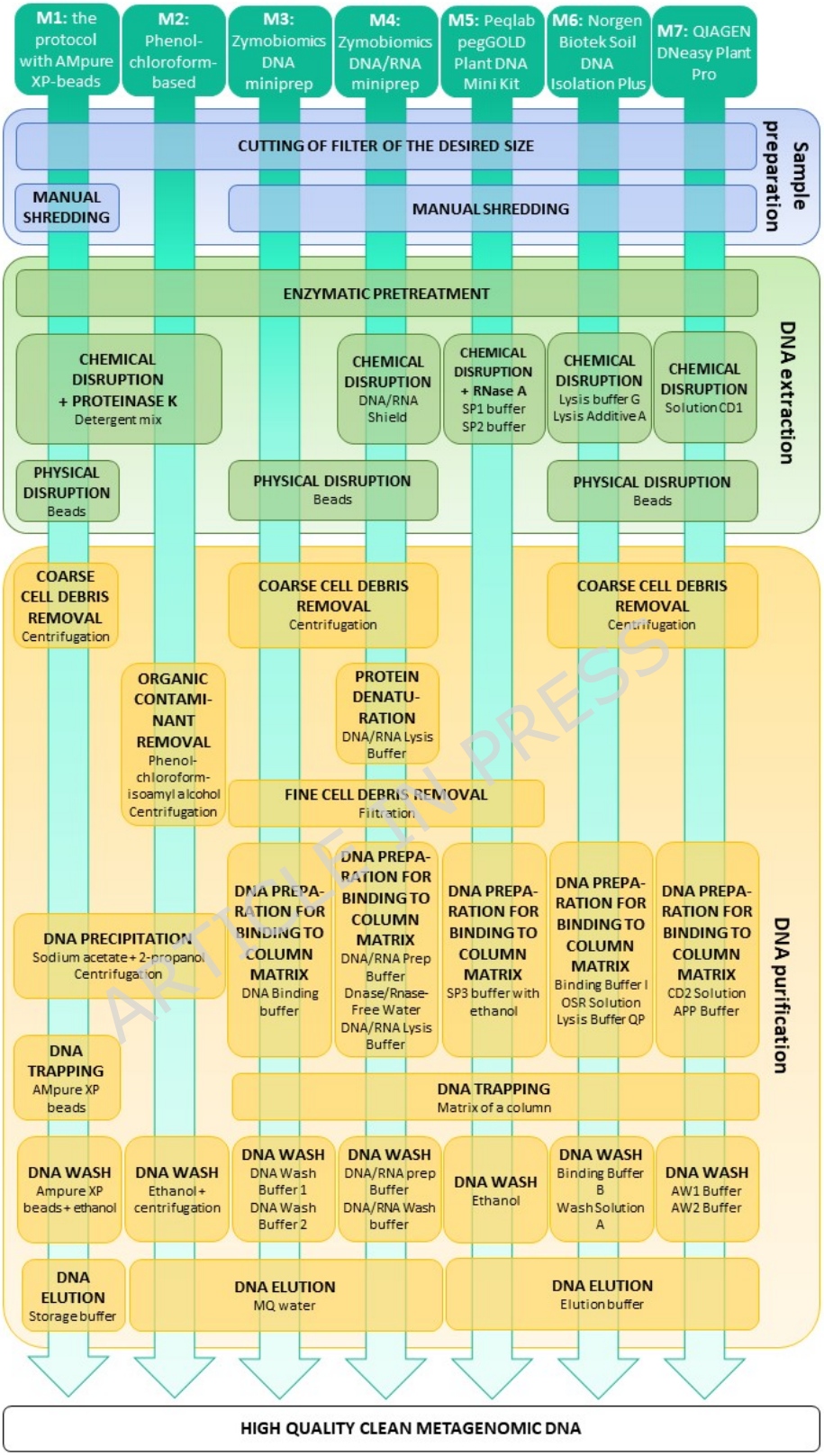

Figure 2. *Workflows of different DNA extraction methods. The methods used are as follows: M1 is the paramagnetic bead-based extraction method presented in this manuscript; M2 is the phenol–chloroform-based method developed and tested in the 2021 Bioaerosol campaign[11]; M3 is the Zymobiomics DNA miniprep kit; M4 is the Zymobiomics DNA/RNA miniprep kit; M5 is the Peqlab PegGOLD Plant DNA Mini Kit; M6 is the Norgen Biotek Soil DNA Isolation Plus Kit; and M7 is the QIAGEN DNeasy Plant Pro Kit.*

The pellet on the magnet was washed with 200 µl of 80% EtOH without any disturbance. This washing process was repeated four times. The pellet was then centrifuged to remove any remaining ethanol. After allowing the pellet to air dry for 5 minutes, it was resuspended in 50 µl of 10 mM Tris buffer (pH 8) containing 50 mM NaCl. After incubating for 10 minutes at room temperature, the elution buffer was removed, and the mixture was separated through magnetic separation of the bead pellet. Finally, the eluate was transferred to a new 1.5 ml tube for further application.

Controls were required throughout the process, and we recommend at least the following: an unused filter subjected to the same treatment sequence and an empty tube with only added buffers/chemicals subjected through the same treatment sequence. In case of contamination, the AMpure XP beads need to be tested separately on MQ, buffers, and chemicals to determine the source of the contamination.

More detailed information on the reagents, devices, protocols, and potential troubleshooting methods is presented in the Supplementary Materials.

3.3. Sequencing, sequence analysis and quality control

Sequencing was performed with an ONT GridION platform with a Rapid Barcoding Kit. The adapters were removed while sequencing. Upon the sequencing completion, the sequences were processed bioinformatically in a similar manner as in the Bioaerosol campaign[12]. Briefly, successful removal of adapters was confirmed with Porechop. The barcodes were also confirmed and trimmed off with Porechop.[33] The per base quality at the ends of the reads was confirmed to be >5 with Cutadapt[34]. Low-quality ends were trimmed off. Preprocessed reads were assembled using SPAdes[35,] and the quality of the assembly was evaluated with FastQC and MultiQC[36,37]. Visualization of the results was done with Chipster platform[38].

Preprocessed reads were classified with Kraken2 using sequential classification against unmasked plant, bacterial, fungal, human, and nt-database[39] sequences with a confidence level of 0.01. The abundance of each library was estimated with Bracken[40]. Unless otherwise specified, default setup values were used.

4. Conclusions

The updated DNA extraction protocol, M1, has been shown to be efficient and repeatable for treating high-yield glass-fiber filter samples collected from the ambient atmosphere (more than 12 µg of DNA per sample are required). For lower-yield samples, M1 is not suitable, and only the initial version of the method, M2, which includes the phenol–chloroform stage, turned out to be effective. Application of these methods allowed direct sequencing of the DNA extract with the 3rd generation Oxford Nanopore sequencing platform, without the PCR multiplication step.

Methods based on commercially available DNA extraction kits have not shown sufficient efficiency in processing the atmospheric samples.

For targeting rigid structures, such as pollen, M1 was more efficient, while M2 was preferable for structures that are more prone to degradation, such as bacteria. Both M1 and M2 methods are very cost-efficient, and the flexibility of the Nanopore sequencing platform tolerates high variation in GC content in DNA.

The minimum amount of DNA necessary for the successful application of the quicker and safer method M1, with subsequent 3rd generation sequencing, poses significant

requirements on the sampling equipment. Based on experience of the Bioaerosol campaign and subsequent studies, the recommended air flow rate is: 300 L min^{-1} for ⌀ 47mm filter, and 3500 L min^{-1} for ⌀ 240mm filter, both for daily collection period.

Author contribution statements

All authors participated in writing and editing of the manuscript and approved its final version. J.S. and S.S. jointly designed the experiment, developed the initial and updated protocol versions and applied them. M.S. participated in the study design. J.P., E.A., A.K., and M.S. provided material, financial, and organizational support for the study.

ORCID for corresponding author

Mikhail Sofiev: https://orcid.org/0000-0001-9542-5746

Funding

The study was performed within the Horizon project SYLVA (grant 101086109) and supported by the Academy of Finland project PS4A (grant 318194), SPORELIFE (grant 355851), DASI (grant 479507) and ACCC Flagship funding (grant 337552).

Competing interests

The authors declare no competing interests.

Data availability statement

The DNA sequences obtained within the study have been uploaded to the ENA database as trimmed read data and are publicly available from the project PRJEB76246.

References

1. Fröhlich-Nowoisky, J. *et al.* Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. *Atmospheric Research* **182**, 346–376 (2016).
2. Sofiev, Mi. & Bergmann, K. C. *Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts*. (Springer Netherlands, Netherlands, 2012).
3. Walser, S. M. *et al.* Evaluation of exposure–response relationships for health effects of microbial bioaerosols – A systematic review. *International Journal of Hygiene and Environmental Health* **218**, 577–589 (2015).
4. Kim, K.-H., Kabir, E. & Jahan, S. A. Airborne bioaerosols and their impact on human health. *Journal of Environmental Sciences* **67**, 23–35 (2018).
5. Goldman, D. L. & Huffnagle, G. B. Potential contribution of fungal infection and colonization to the development of allergy. *Med Mycol* **47**, 445–456 (2009).

6. Damialis, A., Halley, J. M., Gioulekas, D. & Vokou, D. Long-term trends in atmospheric pollen levels in the city of Thessaloniki, Greece. *Atmospheric Environment* **41**, 7011–7021 (2007).
7. Buters, J. T. M. *et al.* Pollen and spore monitoring in the world. *Clin Transl Allergy* **8**, 9 (2018).
8. Stephens, B. What Have We Learned about the Microbiomes of Indoor Environments? *mSystems* **1**, e00083-16 (2016).
9. Shin, S.-K. *et al.* Metagenomic Insights into the Bioaerosols in the Indoor and Outdoor Environments of Childcare Facilities. *PLoS ONE* **10**, e0126960 (2015).
10. Serrano-Silva, N. & Calderón-Ezquerro, M. C. Metagenomic survey of bacterial diversity in the atmosphere of Mexico City using different sampling methods. *Environmental Pollution* **235**, 20–29 (2018).
11. Spring, A. M. *et al.* A Method for Collecting Atmospheric Microbial Samples From Set Altitudes for Use With Next-Generation Sequencing Techniques to Characterize Communities. *Air, Soil and Water Research* **11**, 117862211878887 (2018).
12. Sofiev, M. *et al.* Bioaerosols in the atmosphere at two sites in Northern Europe in spring 2021: Outline of an experimental campaign. *Environmental Research* **214**, 113798 (2022).
13. Jiang, W. *et al.* Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat Protoc* **10**, 768–779 (2015).
14. Wydro, U. Soil Microbiome Study Based on DNA Extraction: A Review. *Water* **14**, 3999 (2022).
15. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**, 1348–1365 (2021).
16. Bock, D. G., Liu, J., Novikova, P. & Rieseberg, L. H. Long-read sequencing in ecology and evolution: Understanding how complex genetic and epigenetic variants shape biodiversity. *Molecular Ecology* **32**, 1229–1235 (2023).

17. Kim, C., Pongpanich, M. & Porntaveetus, T. Unraveling metagenomics through long-read sequencing: a comprehensive review. *J Transl Med* **22**, 111 (2024).

18. Abdel-Latif, A. & Osman, G. Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. *Plant Methods* **13**, 1 (2017).

19. Dauphin, L. A., Stephens, K. W., Eufinger, S. C. & Bowen, M. D. Comparison of five commercial DNA extraction kits for the recovery of Yersinia pestis DNA from bacterial suspensions and spiked environmental samples: DNA extraction from Y.Â pestis. *Journal of Applied Microbiology* **108**, 163–172 (2010).

20. Rostami, H., Firoozeh, F., Zibaei, M. & Salahshoorifar, I. Evaluation of Genomic DNA Extraction Using Monolayer and Bilayer Magnetic Nanoparticles. *Int J Enteric Pathog* **8**, 51–54 (2020).

21. Ma, C. *et al.* Magnetic Nanoparticles-Based Extraction and Verification of Nucleic Acids from Different Sources. *Journal of Biomedical Nanotechnology* **9**, 703–709 (2013).

22. Kumaishi, K. *et al.* High throughput method of 16S rRNA gene sequencing library preparation for plant root microbial community profiling. *Sci Rep* **12**, 19289 (2022).

23. Kumaishi, K. *et al. Simple Amplicon Sequencing Library Preparation for Plant Root Microbial Community Profiling*. http://biorxiv.org/lookup/doi/10.1101/2021.04.14.439905 (2021) doi:10.1101/2021.04.14.439905.

24. Tilak, M.-K., Botero-Castro, F., Galtier, N. & Nabholz, B. Illumina Library Preparation for Sequencing the GC-Rich Fraction of Heterogeneous Genomic DNA. *Genome Biology and Evolution* **10**, 616–622 (2018).

25. Van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research* **322**, 12–20 (2014).

26. Butler, C., Matsumoto, A., Rutherford, C. & Lima, H. K. Comparison of the Effectiveness of Four Commercial DNA Extraction Kits on Fresh and Frozen Human Milk Samples. *MPs* **5**, 63 (2022).

27. Esfandani-Bozchaloyi, S., Sheidai, M. & Kalalegh, M. H. Comparison of DNA extraction methods from Geranium (Geraniaceae). *Acta Botanica Hungarica* **61**, 251–266 (2019).

28. Rohland, N., Siedel, H. & Hofreiter, M. A rapid column-based ancient DNA extraction method for increased sample throughput. *Molecular Ecology Resources* **10**, 677–683 (2010).

29. Wu, L. H., Liu, J. L., Zeng, J. & Zhao, J. Comparison of DNA Extraction and Purification Methods from Different Soils for Metagenomic Sequencing. *AMR* **955–959**, 306–309 (2014).

30. Kalisa, E., Kuuire, V. & Adams, M. A preliminary investigation comparing high-volume and low-volume air samplers for measurement of PAHs, NPAHs and airborne bacterial communities in atmospheric particulate matter. *Environ. Sci.: Atmos.* **2**, 1120–1131 (2022).

31. Paatero, J., Hatakka, J. & Viisanen, Yrjö. *Concurrent Measurements of Airborne Radon-222, Lead-210 and Beryllium-7 at the Pallas-Sodankylä Gaw Station, Northern Finland*. (Ilmatieteen laitos, Helsinki, 1998).

32. Dunn, O. J. Multiple Comparisons Using Rank Sums. *Technometrics* **6**, 241–252 (1964).

33. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics* **3**, (2017).

34. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).

35. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).

36. FastQC. (2015).

37. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

38. Kallio, M. A. *et al.* Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics* **12**, 507 (2011).

39. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).

40. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* **3**, e104 (2017).